Solar-to-Grid Public Data File for Utility-scale (UPV) and Distributed Photovoltaics (DPV) Generation and Market Value

Background

Lawrence Berkeley National Laboratory (Berkeley Lab) estimates hourly project-level generation data for utility-scale solar projects and hourly county-level generation data for residential and non-residential distributed photovoltaic (PV) systems in the seven organized wholesale markets and 10 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 http://data.openei.org/submissions/4503 and summary data files at https://emp.lbl.gov/renewable_grid_insights.

Annual solar summary statistics by plant (UPV) and county (DPV)

We provide project-level (UPV) and county-level (DPV) 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_and_county.xlsx*. For more information on methods, data, and validation see Appendix A and C in the technical report.

Capacity credit by plant (UPV) and county (DPV)

We provide project-level (UPV) and county-level (DPV) estimates of the capacity credit in *Capacity_credit_data_by_plant_and_county.xlsx.* For plants in CAISO and ERCOT we use values reported directly by the ISO. See Section 3 in the technical report for more information on the market rules used to calculate the capacity credit.

Hourly generation data

This data set is only available at the OEDI.

UPV

In addition to the annual solar summary statistics Berkeley Lab provides hourly generation estimates for 3404 utility-scale solar projects, starting at the project's commercial operation date (or 2012 for older projects) until the end of 2020. 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 in the context of Berkeley Lab's Utility-Scale Solar Project) and historical irradiance estimates in NREL's National Solar Radiation Database (NSRDB).

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.

DPV

Berkeley Lab provides hourly distributed PV generation estimates for 2469 counties, starting in 2012 until the end of 2020. A separate .csv file is listed for each county, using the county's 5-digit Federal Information Processing Series (FIPS) code as its filename. Records are indexed by UTC-Hour-Beginning datetimes. Here we summarize the data by column:

DPV_res_MW and **DPV_non_res_MW**: Cumulative residential and non-residential PV capacity estimates expressed in MW_{AC} terms. Data is derived from EIA form 861, Wood Mackenzie's *Solar Year in Review* Series, and the Interstate Renewable Energy Council.

CF: Capacity factor estimate derived from NREL's SAM, using historical irradiance estimates of the NSRDB. System characteristics are informed by data collected in the context of Berkeley Lab's Tracking the Sun Project. The capacity factor estimate is the same for residential and non-residential distributed PV installations.

DPV_res_MWh and **DPV_non_res_MWh**: Hourly distributed solar generation estimate for the residential and non-residential sector. It represents the simple product of the hourly cumulative capacity estimate and the coincident capacity factor.

Who to Contact with Questions?

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

Disclaimer