User Guide for *Tracking the Sun* Public Data File

Version Date: September 22, 2017

Background

Lawrence Berkeley National Laboratory (Berkeley Lab) collects project-level data on residential and non-residential photovoltaic (PV) systems for its annual *Tracking the Sun* report. The data are sourced primarily from state agencies and utilities that administer PV incentive programs, solar renewable energy credit registration systems, or interconnection processes. In order to leverage this dataset for broader use, Berkeley Lab has issued a public data file, which can be downloaded via the National Renewable Energy Laboratory's *Open PV Project* data portal.[1] The public data file will be updated once annually with data from the previous calendar year, and may also be updated on additional occasions as improvements to the data cleaning methodology and supplementary data fields are developed.

What is Included in the Public Data File?

The data file includes only grid-connected residential and non-residential PV systems, defined to consist of rooftop systems, regardless of size, and ground-mounted systems up to 5 MW_{AC}. Ground-mounted projects larger than 5 MW_{AC} are considered utility-scale projects and are not included in this dataset.

The current version of the public data file includes more than 1.1 million PV systems installed through year-end 2016. The file includes more than 60 data fields describing key attributes of each system, which are listed and described in the table below. Note, though, that most fields are incomplete for most systems.

What Data Cleaning Operations Are Performed?

The data collected for *Tracking the Sun* undergoes extensive cleaning and quality control. Some elements of those operations are described in the table below. For additional information, please refer to Section 2 of the latest *Tracking the Sun* report ("Data Sources, Methods, and Sample Description") and to Appendix A.

One important convention should be noted: Missing data are coded in the database as -9999. Any operations performed on the data should therefore treat such values accordingly.

Who to Contact with Questions?

Questions or comments specifically about the *Tracking the Sun* public data file may be directed to either Naïm Darghouth (ndarghouth@lbl.gov) or Galen Barbose (glbarbose@lbl.gov).

[1] The public data file excludes any data provided under confidentiality agreements as well as other sensitive information that data providers requested to be withheld.

Data Fields in the Public Data Set

Data Field Name	Units	Description and Key Notes
Data Provider	n/a	The entity that supplied the data, generally a utility or PV incentive program administrator
System ID (from Data Provider)	n/a	This is the system or application ID within the raw data file from the data provider.
System ID (Tracking the Sun)	n/a	This is the system or application ID created within Berkeley Lab's Tracking the Sun database.
Installation Date	date	For some data providers, the installation date may be based on the best available proxy, such as the date that an incentive claim was submitted or when the inspection was performed.
System Size	kW	The total rated direct-current (DC) output of the module arrays at standard test conditions. These data are generally reported directly by the data provider, but in some cases must be estimated, for example, based on the module model and quantity or based on reported alternating-current (AC) capacity.
Total Installed Price	dollars (nominal)	The total installed price for the system, prior to receipt of any incentives, as reported by the installer, host customer, or other incentive applicant. For third-party owned systems, the data may represent one of two things. If the third-party owner procured the system from an independent installation contractor, then the reported installed price likely refers to the intermediate sale price between the installation contractor and the third-party owner. If the third-party owner instead installed the system itself, then the reported installed price likely represents an appraised value.
Appraised Value Flag	n/a	A flag used to indicate whether the reported installed price is likely to represent an appraised value. Caution should be used in relying on appraised values for analysis or benchmarking purposes, as such data do not represent a transaction price.
Module Cost	dollars (nominal)	The reported cost of modules. Conventions may vary in terms of whether installer mark-up is included.
Inverter Cost	dollars (nominal)	The reported cost of inverters. Conventions may vary in terms of whether installer mark-up is included.
Installation Labor Cost	dollars (nominal)	The reported cost of installation labor. Conventions may vary in terms of the scope of labor costs included, and whether those costs are based on only wages or are fully-burdened.

Data Field Name	Units	Description and Key Notes
Permitting Cost	dollars (nominal)	The reported cost of building and/or electrical permit fees.
Balance of Systems Cost	dollars (nominal)	The reported cost of balance of systems. Conventions may vary in terms of the scope of costs included.
Sales Tax Cost	dollars (nominal)	The calculated cost of sales taxes. This is estimated based on average sales tax rates for the given state and year, accounting for any sales tax exemptions that may exist for PV systems. Sales taxes, if applicable, are assumed to be levied only on hardware costs, which are assumed to represent 55% of the total installed price.
Rebate or Grant	dollars (nominal)	The pre-tax value of any up-front rebate or grant provided by the entity supplying the data
Performance-Based Incentive (Annual Payment)	dollars (nominal)	Data reported by data providers generally consists of either the estimated annual PBI payment of the nominal sum of PBI payments over the full incentive term. In some cases, only the PBI rate ($/kWh) and PBI term are available, in which case the annual PBI payment is calculated based on estimated insolation levels and first-year energy production. PBI payment amounts are reported on a pre-tax basis.
Performance-Based Incentives (Duration)	years	Number of years that PBI payments are disbursed.
Feed-in Tariff (Annual Payment)	dollars (nominal)	The estimated pre-tax annual feed-in tariff (FIT) payment received in the first year of the FIT contract term. The calculation procedure mirrors that described above for PBI payments.
Feed-in Tariff (Duration)	years	Contract term length.
Customer Segment	n/a	Data on customer segment is mapped to one of six general types: RES, COM, SCHOOL, GOV, NON-PROFIT, and NON-RES, the last one being used only if more-specific information on non-residential customer type is unavailable.
New Construction	n/a	Indicates if the system was installed at the time of building construction
Tracking	n/a	Indicates if the system includes tracking equipment
Tracking Type	n/a	Indicates if tracking equipment is single-axis or dual-axis
Ground Mounted	n/a	Indicates if the system is ground-mounted (which may include pole-mounted systems). PV systems consisting of a combination of rooftop and ground-mounted arrays are coded as ground-mounted.
Battery System	n/a	Indicates if the system includes batteries

Data Field Name	Units	Description and Key Notes
Zip Code	n/a	Host customer zip code
City	n/a	Host customer city. Spellings have not been corrected or standardized.
County	n/a	Host customer county. Spellings have not been corrected or standardized.
State	n/a	Host customer state
Insolation Rate	kWh/m2/day	Average annual insolation rate for the host-customer zip code, as estimated using the National Renewable Energy Laboratory PVWatts api (https://developer.nrel.gov/docs/solar/pvwatts-v5/). Reported azimuth and tilt values are used as inputs, when available; national-average tilt and azimuth are used otherwise. Default values are used for all other inputs.
Reported Annual PV Generation	kWh/yr	The annual energy production of the PV system, as reported by the data provider.
Estimated Annual PV Generation	kWh/yr	The annual energy production of the PV system, as using the National Renewable Energy Laboratory PVWatts api (https://developer.nrel.gov/docs/solar/pvwatts-v5/). Reported azimuth and tilt values are used as inputs, when available; national-average tilt and azimuth are used otherwise. Default values are used for all other inputs.
Utility Service Territory	n/a	If not reported directly by the data provider, the electric utility service territory is inferred based on the host customer zip code.
Third-Party Owned	n/a	Indicates if the system is third-party owned; that is, owned by an entity other than the site host and either leased or sold under a power purchase agreement to the site host.
Installer Name	n/a	These data have been cleaned and the spellings standardized to the extent feasible; however, there may still be instances where the same company is listed under multiple names.
Self-Installed	n/a	Indicates if the system was installed by the site-host.
Azimuth #1	degrees	The horizontal direction of the array, where 180 degrees defines South facing PV orientation. Azimuth data reported by data providers was, in some cases, modified to adhere to this convention. Data fields are provided for up to three array orientations, though some systems may consist of a larger set of distinct orientations.
Azimuth #2	degrees	
Azimuth #3	degrees	
Tilt #1	degrees	The vertical tilt of the array, where zero degrees corresponds to a flat array. As with the azimuth
Tilt #2	degrees	

Data Field Name	Units	Description and Key Notes
Tilt #3	degrees	data, fields are provided for up to three array orientations, though some systems may consist of a larger set of distinct orientations.
Module Manufacturer #1	n/a	These data have been cleaned and the spellings standardized to the extent feasible.
Module Manufacturer #2	n/a	
Module Manufacturer #3	n/a	
Module Model #1	n/a	
Module Model #2	n/a	
Module Model #3	n/a	
Module Technology #1	n/a	Identifies the module technology type. This is determined by cross-referencing module manufacturer and model names against equipment specification data available through solarhub.com and the California Solar Initiative eligible equipment list.
Module Technology #2	n/a	
Module Technology #3	n/a	
BIPV Module #1	n/a	Indicates if the modules are building integrated photovoltaics (BIPV). This is determined by cross-referencing module manufacturer and model names against equipment specification data available through solarhub.com and the California Solar Initiative eligible equipment list.
BIPV Module #2	n/a	
BIPV Module #3	n/a	
Module Efficiency #1	percent	Identifies the energy conversion efficiency of the modules. This is determined by cross-referencing module manufacturer and model names against equipment specification data available through solarhub.com and the California Solar Initiative eligible equipment list.
Module Efficiency #2	percent	
Module Efficiency #3	percent	
Inverter Manufacturer	n/a	These data have been cleaned and the spellings standardized to the extent feasible.
Inverter Model	n/a	
Microinverter	n/a	Indicates if the system uses micro-inverters. This is determined by cross-referencing inverter manufacturer and model names against equipment specification data available through solarhub.com and the California Solar Initiative eligible equipment list.
DC Optimizer	n/a	Indicates if the system uses DC Optimizers, based on the inverter manufacturer name. All systems using SolarEdge inverters are assumed to also include a DC optimizer. Systems using DC optimizers manufactured by other companies (e.g., Tigo) cannot be identified based on the inverter manufacturer; as such, the DC Optimizer field is coded as unknown for all systems with string inverters manufactured by companies others than SolarEdge.

Disclaimer