Berkeley Lab Solar Data Sets: Data Commons

Naïm Darghouth, Joachim Seel, Dev Millstein, Joe Rand

Lawrence Berkeley National Laboratory

Kenneth Gillingham

Yale University

Webinar

June 9, 2022

This work was funded by the U.S. Department of Energy Solar Energy Technology Office, under Contract No. DE-AC02-05CH11231.

Agenda

- Introduction
- Distributed solar data: Tracking the Sun and solar demographics
- Utility Scale Solar data
- Solar to Grid: System impacts, reliability and market value
- Wholesale electricity market data
- Interconnection Queues
- Yale-LBNL Collaborations
- Q&A

Introduction

- Berkeley Lab collects, purchases, collates, cleans, manages, or otherwise curates a number of authoritative and unique datasets used in SETO-funded research
- When possible, data is shared directly through the Electricity Market & Policy department web site
- When data is protected by NDAs:
  - Online data visualizations – customizable maps, charts, graphics
  - Share aggregated data
- Berkeley Lab Solar Data Commons
  - Goal is to make LBNL solar-related data accessible
  - Focus is solar, but relevant beyond solar

Webinar Objective:
To describe the data sets we manage, what can shared directly, and the tools used to interact with the data

Tracking the Sun and Solar Adopter Income & Demographics Data

trackingthesun.lbl.gov

solardemographics.lbl.gov

Tracking the Sun Data Description

Tracking the Sun relies on project-level data for distributed solar

- Datasets provided by state agencies, utilities, and other organizations for systems participating in PV incentive programs, renewable energy credit registration systems, interconnection processes, and net metering programs
- Some of these data already exist in the public domain though LBNL may receive supplementary fields, often under non-disclosure agreements

61 entities spanning 29 states contributed data to upcoming 2022 TTS report

2.5 million systems through 2021 in full sample (77% of U.S. market)

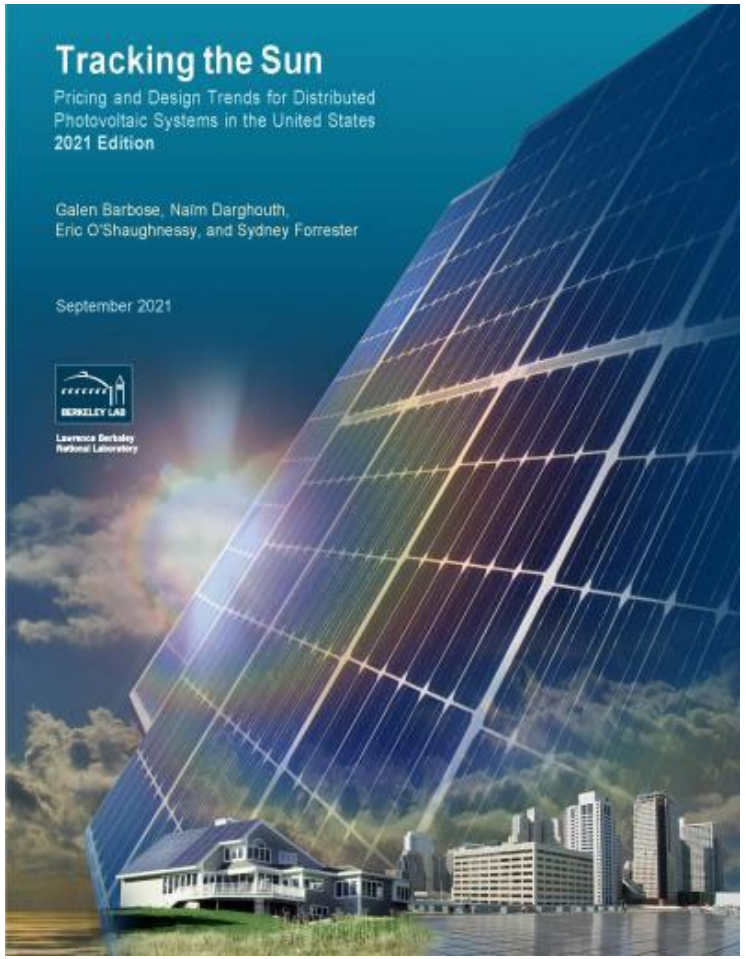
- Date, zip code, system size, install date at minimum
- >80 other PV system characteristics variables available

Data is cleaned and standardized

- Standardize component spellings, integrate data from spec sheets

Tracking the Sun: Data Sharing

- TTS public data file
  - Data not protected by NDAs is shared on https://trackingthesun.lbl.gov
  - Updated once a year
  - Some programs prevent us from sharing any of their data, other only ask to exclude certain variables
  - Does not include any addresses
- TTS data visualization
  - Allows user to customize figures and maps on topics covered in report
  - https://emp.lbl.gov/tracking-sun-tool
  - Demo
- Though addresses cannot be shared, some US census tract data can
  - Currently, approval for CA only
  - Working with lawyers to increase number of states



Augmenting the TTS data: Address, income, real estate and demographic data

Buildzoom

- Provides solar addresses extracted from municipal permitting data
- Increased coverage in 44 states (233k systems) in 2020 report

Ohm Analytics

- Similar to Buildzoom, a few additional fields extracted
- Increased coverage in 16 states (169k systems) in 2020 report

Experian

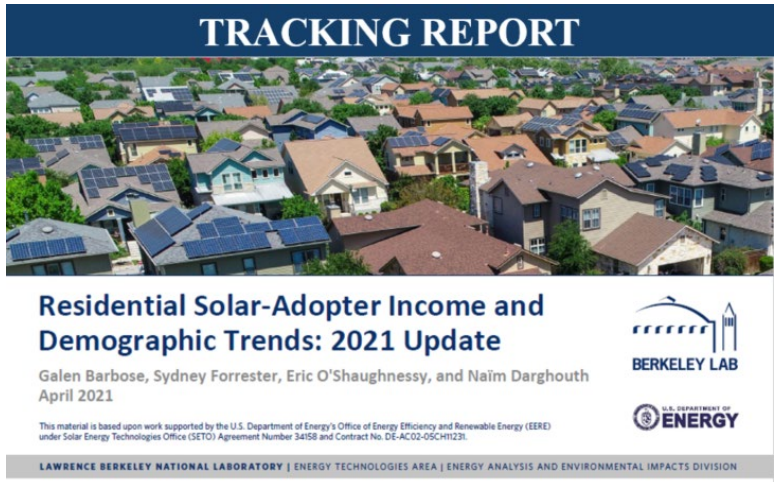
- Modeled household level incomes and select demographic data for PV adopters with addresses

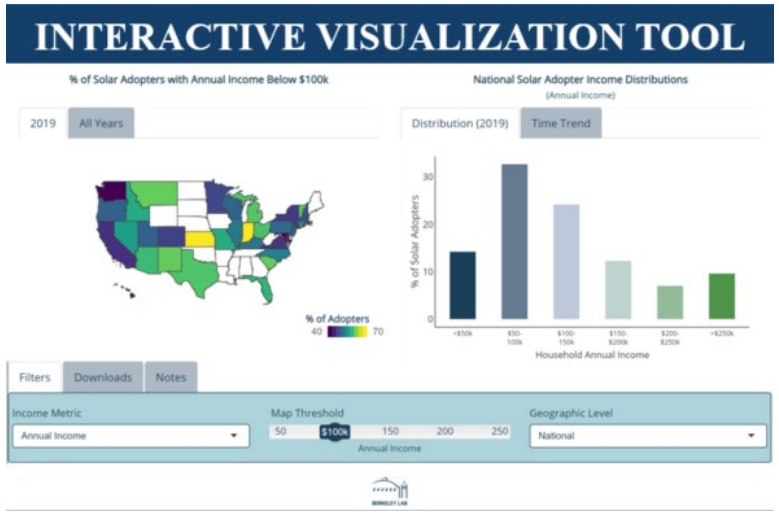
CoreLogic

- Real estate characteristics of properties where residential solar is installed
- Can share description of data (data dictionary requires a NDA)

Solar Adopter Income Trends, Data, and Visualizations

- We describe income and other socio-economic trends of residential solar adopters over time and across geographies using these data sets in a tracking report
- Many of the charts in the report can be customized to different geographies and income thresholds in our data visualization
  - Underlying data set cannot be shared due to NDAs but we do share data aggregated at the US census tract-level
- Single web page for all solar demographics trends, analysis, and data programs
  - https://solardemographics.lbl.gov
- Demo





Utility Scale Solar Data

utilityscalesolar.lbl.gov

“Utility-Scale Solar” Data Description

“Utility-Scale Solar” synthesizes and analyzes diverse data from all ground-mounted solar (and solar+storage) plants in the United States that are >5 MW_{AC}

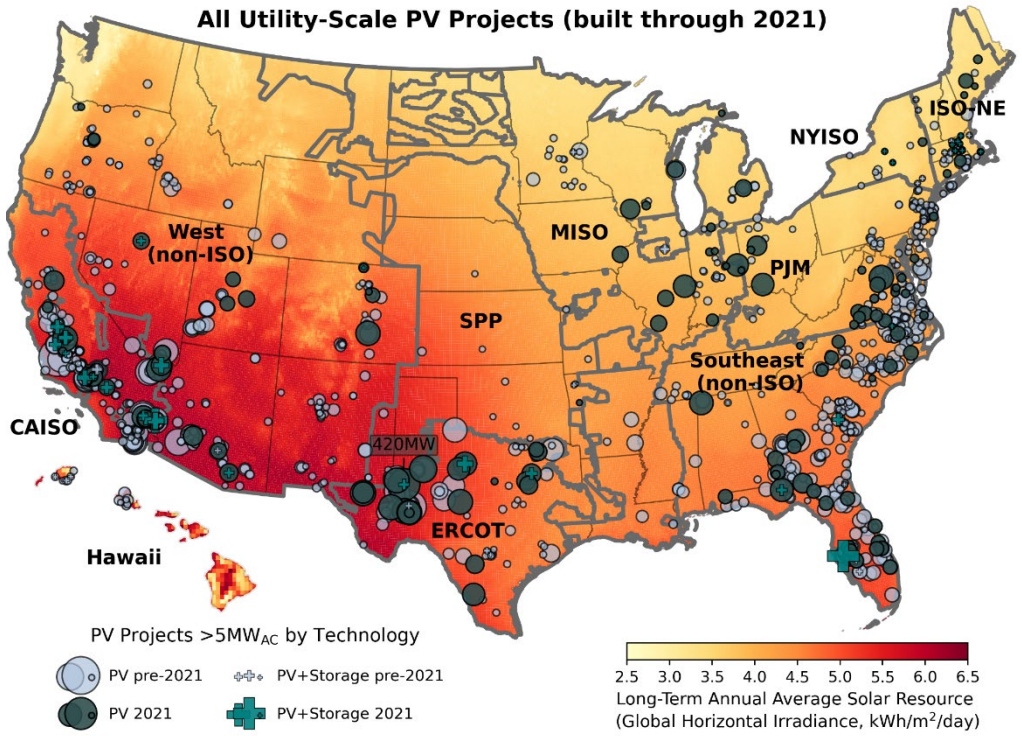
We mostly work with publicly available data - the primary value add is in synthesizing and augmenting disparate data and the creation of derivative metrics. We use:

- **Plant deployment and characteristics (metadata):** EIA-860, FERC Form 556, regulatory filings, developer/owners interviews, trade press, satellite imagery
- **CapEx:** Regulatory filings, FERC Form 1, developer/owner interviews, trade press, confidential data from EIA-860 (NDA)
- **OpEx:** FERC Form 1, informal industry survey
- **Capacity factor:** EIA-923, EIA-860, FERC Form 1, FERC Electric Quarterly Reports, regulatory filings
- **LCOE:** Calculated from empirical CapEx, OpEx, capacity factor data, plus financing costs
- **PPA prices:** Regulatory filings, FERC Electric Quarterly Reports, FERC Form 1, EIA-923, trade press
- **Wholesale market value:** NSRDB and HRRR for generation modeling, ISO-reported data (aggregate solar generation, solar curtailment, nodal energy prices, capacity credit, capacity prices), EIA-923

“Utility-Scale Solar” Data Coverage

Through 2021, we tracked 1126 plants totaling 52.5 GW_{AC}:

- **Plant deployment and characteristics (metadata):** ~100% coverage
- **CapEx:** ~88% coverage
- **OpEx:** ~9% coverage (projects by investor-owned utilities)
- **Capacity factor:** >96% coverage
- **LCOE:** ~87% (requires CapEx and capacity factor data)
- **PPA prices:** ~43% of “total market” (some plants don’t have PPAs, no reports to FERC in ERCOT)
- **Wholesale market value:** ~90% of capacity >1MW_{AC} (wholesale markets+10 balancing areas)



“Utility-Scale Solar” Data Sharing

- **“Utility-Scale Solar” public data file**
  - Excel workbook with >40 worksheets/tabs that provide data (that are not restricted by an NDA) behind report graphs
  - In addition, the last worksheet/tab in the workbook synthesizes non-NDA data at the individual plant level
    - The only NDA data are plant-level CapEx from EIA-860; contact EIA-860 team (Glenn McGrath) if you’re interested in exploring NDA
  - Available under the “Files” tab at the bottom of utilityscalesolar.lbl.gov
  - Updated once a year (annual data resolution, for the most part)

- **“Utility-Scale Solar” data visualizations**
  - Six interactive (and in some cases multi-window) Tableau visualizations spanning the main topics covered in report:
    - Deployment and technology trends (metadata)
    - Capacity and generation over time, by state
    - CapEx, LCOE, and PPA prices over time, by region or plant
    - Capacity factors
  - Available under the “Visualizations” tab at the bottom of utilityscalesolar.lbl.gov

- **Data requests**
  - Always happy to entertain specific requests for data that are not already disclosed via public data file or visualizations

Solar-to-Grid: System Impacts, Reliability and Market Value

emp.lbl.gov/renewable-grid-insights

Solar-to-Grid: Data Description

Solar Deployment and Generation

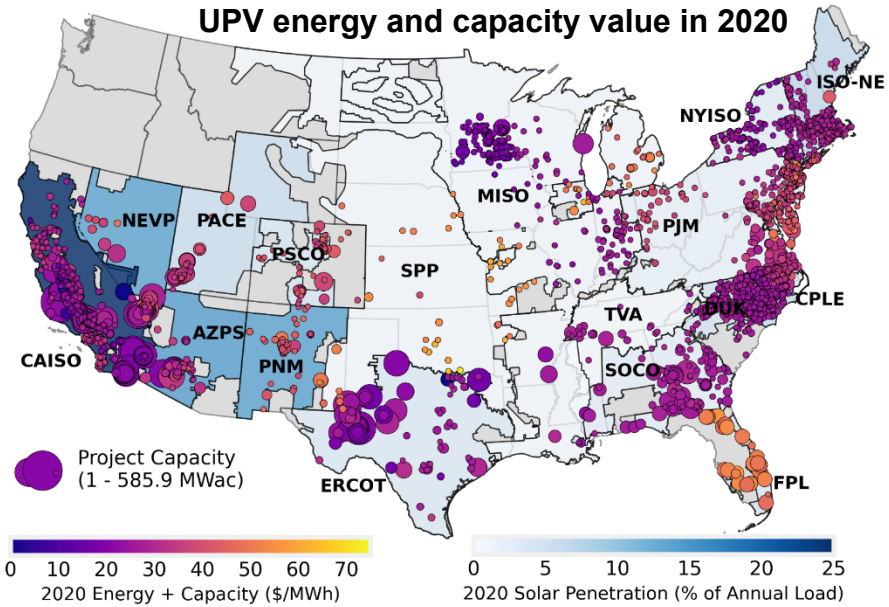
- Utility-scale (>1MW$_{AC}$, EIA 860) and distributed PV (EIA 861) in 7 ISO/RTOs and 10 additional balancing areas, 2012-2020. Focus on stand-alone solar (not solar-storage hybrids).
- Hourly project-level (UPV) / county-level (DPV) generation estimates based on plant-specific technical characteristics, local weather data (NSRDB), debiased using hourly ISO generation reports, monthly/annual plant generation records (EIA 923), ERCOT 60-day reports and curtailment data

Capacity Credit of Solar

- Plant/county-level estimates of capacity credit based on prevalent ISO rules. Contribution of top 100 net-load hours in BAs outside of ISOs.

Wholesale Market Value

- Plant/county-level annual wholesale market value based on nearest node's hourly LMPs and zonal/seasonal capacity prices
- Plant/county-level value factor (relative to flat block of power)



Solar-to-Grid: Data Sharing

- **Annual summaries of plant/county-level data**
  - Generation, curtailment, energy value, capacity credit, capacity value ($/MWh + $/kW-yr), and value factors
  - Available for download at project website https://emp.lbl.gov/renewable-grid-insights

- **Hourly plant/county-level data**
  - Modeled and debiased generation and curtailment data
  - Available for download at the *Open Energy Data Initiative*: **data.openei.org/submissions/4503**

- **Visualizations**:
  - Regional annual sectoral capacity and generation summaries: https://emp.lbl.gov/solar-capacity-and-generation
  - Mapping of annual plant-level energy and capacity value of UPV projects by year, balancing area, and state: https://emp.lbl.gov/utility-scale-solar-value

For more details see our data user guide:
https://emp.lbl.gov/sites/default/files/s2g_2021_user_guide_for_data_file.pdf

We will continue to generate these data for large-scale solar projects in our annual Utility-Scale Solar report series: https://utilityscalesolar.lbl.gov

Wholesale Market Electricity Price Data

“Wholesale Market Electricity Price” Data Description

We use wholesale market price data across multiple projects to help estimate the market value of wind, solar, and hybrid generation

We also use this data to investigate how generation profiles interact with system value, and how infrastructure development, such as transmission, impacts electricity markets

- **Energy Prices**
  - We primarily collect prices at market nodes (~50,000 nodes established by the 7 major ISO/RTOs)
  - Data includes hourly real-time and day-ahead prices
  - At a subset of nodes, real-time prices can be broken into components energy, losses, and congestion
  - In Texas we also collect ORDC prices (Operating Reserve Demand Curve)

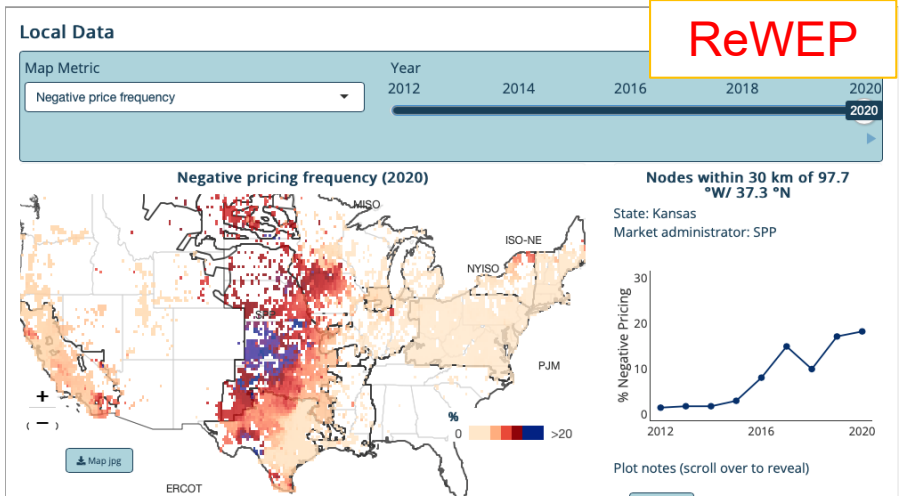
- **Capacity Prices**
  - We collect capacity prices by ISO and by zone within ISO (where relevant)

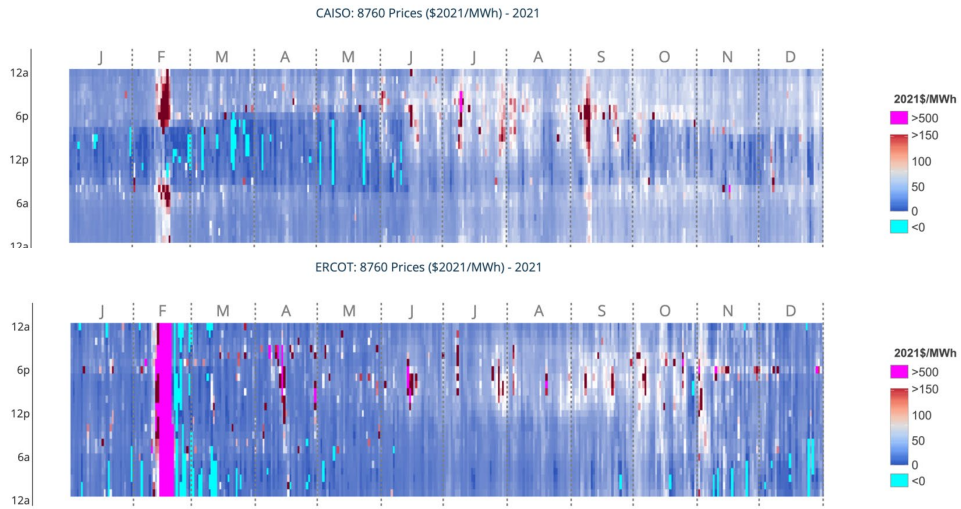
“Wholesale Market Electricity Price” Data Sharing

- Nodal price data is purchased through the ABB Velocity Suite – i.e., we can not share the nodal price series directly

- We share aggregated data through the **The Renewables and Wholesale Electricity Prices (ReWEP) Tool**
  - https://*emp.lbl.gov/renewables-and-wholesale-electricity-prices-rewep*
  - Updated this spring with 2021 data
  - Associated journal article
    - https://dx.doi.org/10.1016/j.adapen.2021.100073

- We share data on solar and wind value in our annual reports and associated visualizations:
  - https://utilityscalesolar.lbl.gov
  - https://emp.lbl.gov/publications/land-based-wind-market-report-2021
  - https://emp.lbl.gov/wind-energy-market-value





“Wholesale Market Electricity Price” Data Sharing

- **Also see:**
  - Solar and wind value analysis
    - https://emp.lbl.gov/publications/solar-and-wind-grid-system-value
  - Solar-to-grid 2020 report:
    - https://emp.lbl.gov/publications/solar-grid-trends-system-impacts-0
  - Capacity credit evaluation for solar+storage
    - https://dx.doi.org/10.1016/j.energy.2020.118587
  - ‘Price Drivers’ report
    - https://emp.lbl.gov/publications/impact-wind-solar-and-other-factors

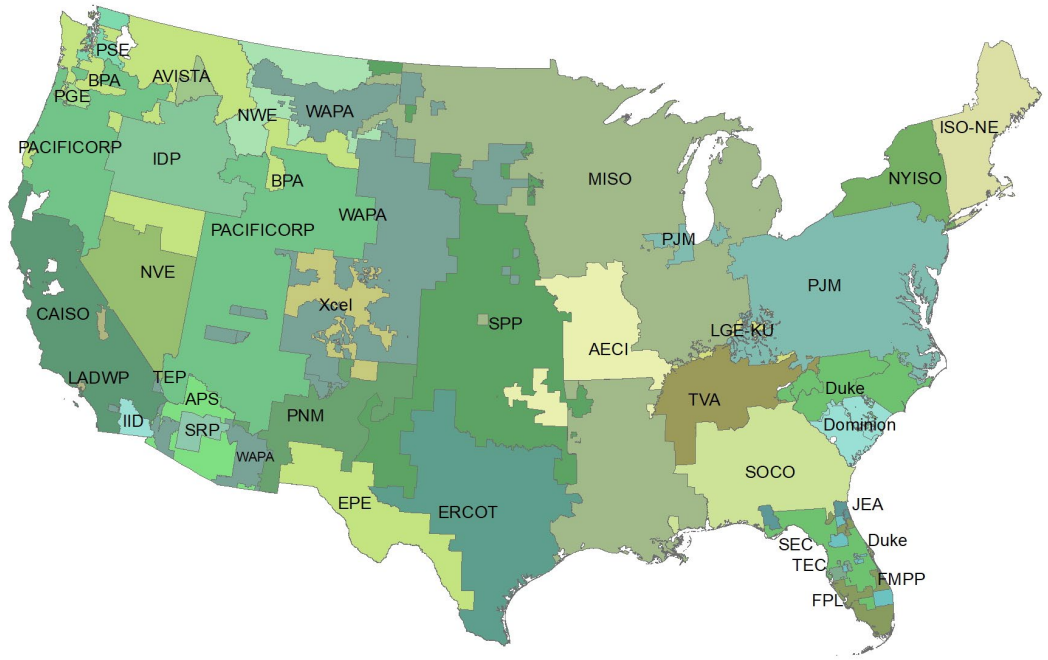
- **Data requests**
  - Always happy to entertain specific requests for data that are not already disclosed via public data file or visualizations
  - We can also connect you with the account managers that we work with at ABB for Velocity Suite data

Interconnection Queue Data

emp.lbl.gov/queues

Interconnection Queues: Data Description

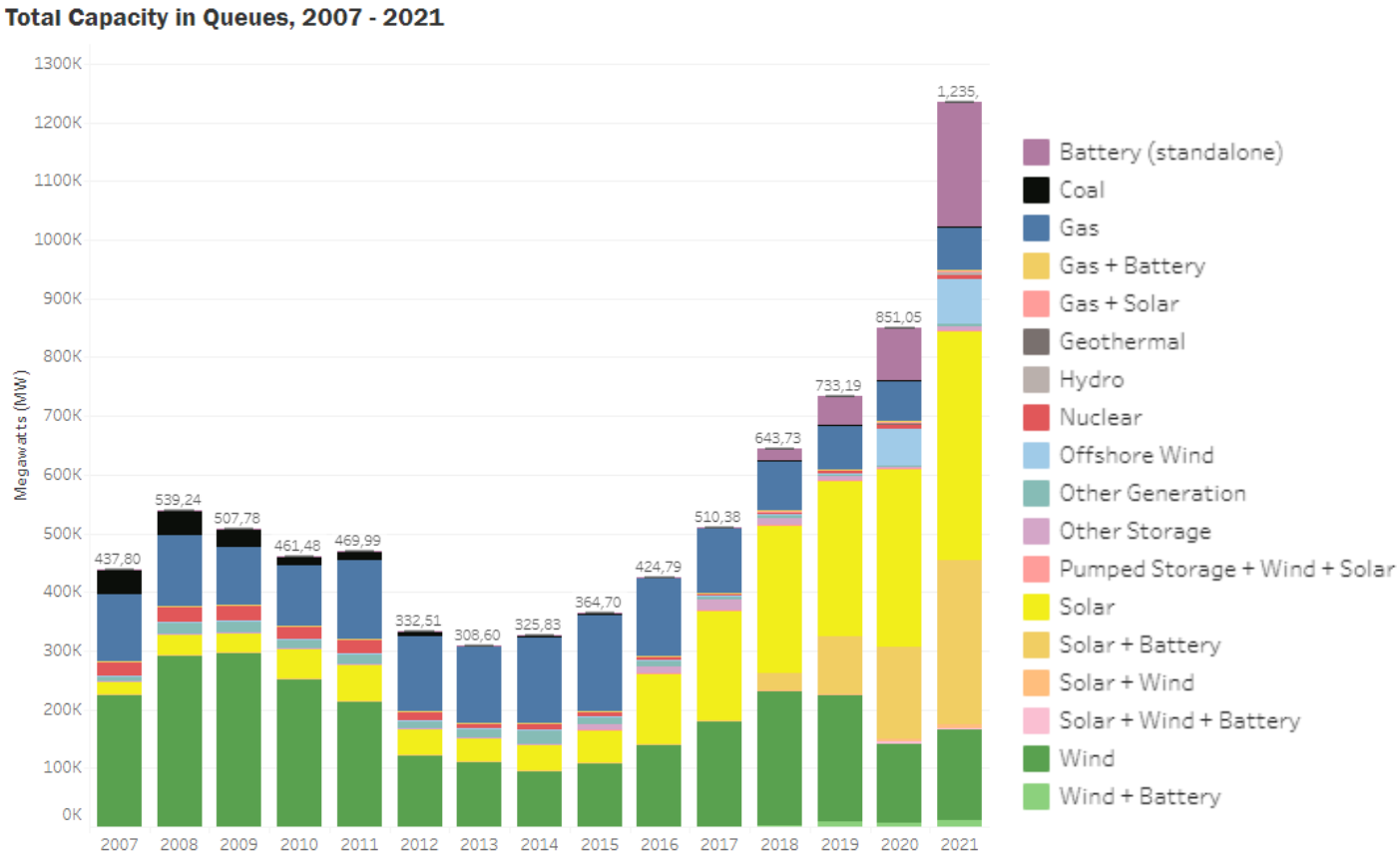
- Data collected from interconnection queues for 7 ISOs / RTOs and 35 utilities
  - Collectively represent >85% of U.S. electricity load
  - Projects that connect to the bulk power system: not behind-the-meter
  - Includes all projects in queues through the end of 2021
  - The full sample includes:
    - 8,133 “active” projects
    - 12,585 “withdrawn” projects
    - 3,439 “operational” projects
    - 229 “suspended” projects
- Hybrid / co-located projects were identified and categorized
  - Storage capacity in hybrids (separate from generator capacity) was estimated based on available data for some projects
- Note that being in an interconnection queue *does not guarantee* ultimate construction: majority of plants are not subsequently built



Coverage area of entities for which data was collected
Data source: Homeland Infrastructure Foundation-Level Data (HIFLD)
A full list of included balancing areas can be found in the Appendix
Note that service areas can overlap
No data collected for Hawaii or Alaska

Interconnection Queues: Data Sharing

- Cleaned queue data file and briefing:
  - https://emp.lbl.gov/queues
  - The data have undergone substantial cleaning, standardizing, and QC efforts by Berkeley Lab
  - Updated annually (last update: through 2021)
- Queues data visualization
  - https://emp.lbl.gov/generation-storage-and-hybrid-capacity
- Additional data collection from ISOs:
  - LBNL is currently collecting additional interconnection data on study durations and project-level interconnection costs
  - Much of this data is protected under CEII or NDA agreements; contact LBNL for more information



Yale-LBNL Collaborations

A series of papers exploring the rooftop solar market

- Deconstructing Solar Photovoltaic Pricing – Energy Journal

Deconstructing Solar Photovoltaic Pricing

Kenneth Gillingham, Hao Deng, Ryan Wiser, Naim Darghouth, Gregory Nemet, Galen Barbose, Varun Rai, Changgui Dong

Free Article

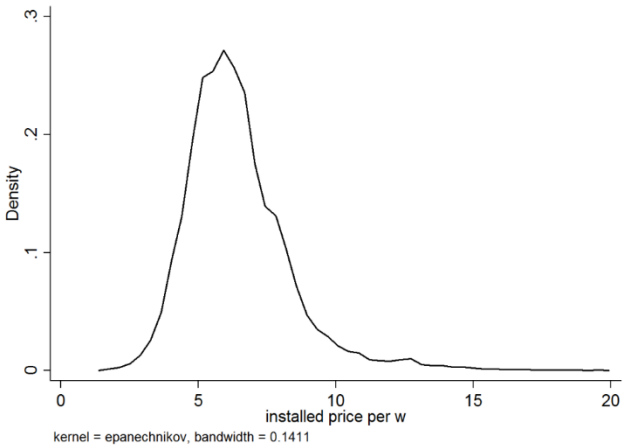
Abstract:
Solar photovoltaic (PV) system prices in the United States display considerable heterogeneity both across geographic locations and within a given location. Such heterogeneity may arise due to state and federal policies, differences in market structure, and other factors that influence demand and costs. This paper examines the relative importance of such factors on equilibrium solar PV system prices in the United States using a detailed dataset of roughly 100,000 recent residential and small commercial installations. As expected, we find that PV system prices differ based on characteristics of the systems. More interestingly, we find evidence suggesting that search costs and imperfect competition affect solar PV pricing. Installer density substantially lowers prices, while regions with relatively generous financial incentives for solar PV are associated with higher prices.

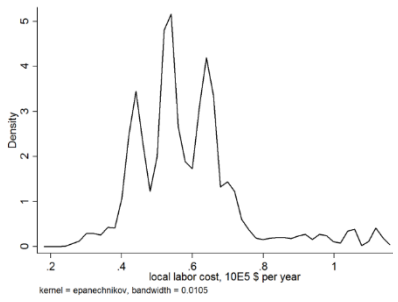
Published in Volume 37, Number 3 of the bi-monthly journal of the IAEE's Energy Economics Education Foundation.

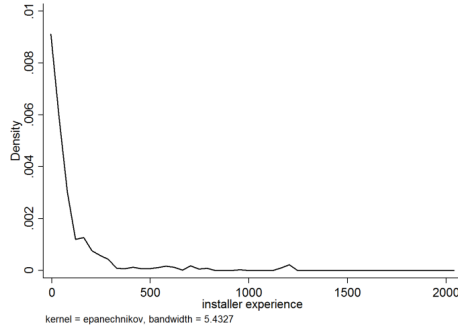
Energy Journal Articles
Index of Volumes and Issues
Recent Issues
Forthcoming Issues and Preprints
My Bookmarks
Search
OpenURL Linking
Most Downloaded Articles

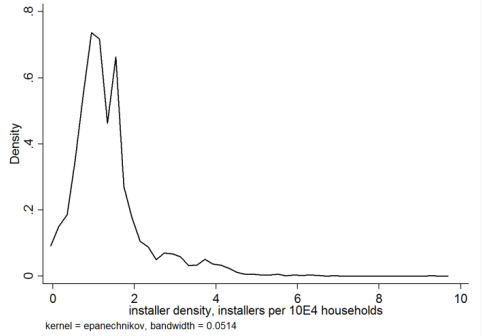
Publication Information
Aims & Scope
Editorial Board
Audience

Table of Contents
Get Permission
Learn About Permissions
Export Citation
Order Physical Print
Article Cited By
Citation Statistics







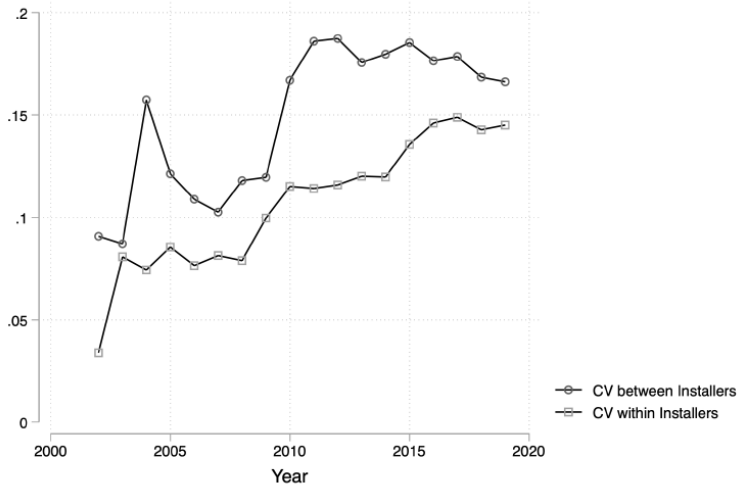


Currently Ongoing Yale-LBNL Collaborations

Further exploring solar and related technologies

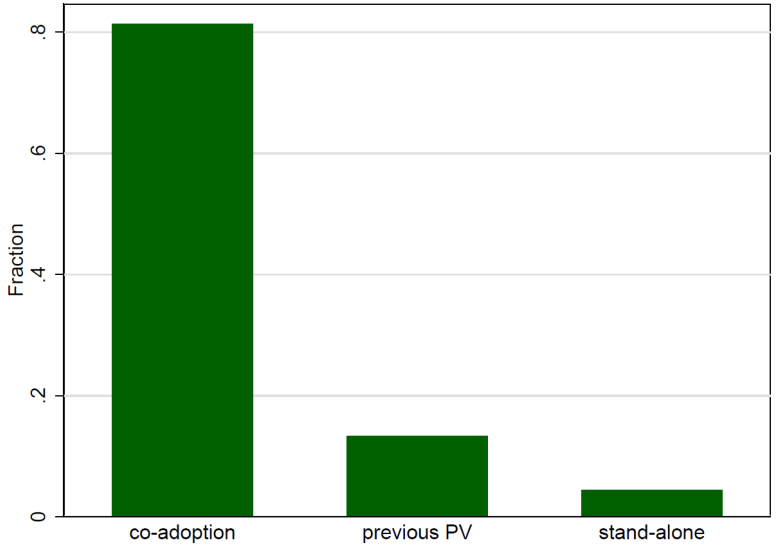
- What explains the variability of solar prices even within a geography or within an installer?
- What are the factors explaining the coadoption of solar and energy storage?

Figure 5: Within-installer and Between-installer CVs



CV = coefficient of variation = std dev/mean

Some figures of
preliminary results



Battery+solar coadoption status

ELECTRICITY MARKETS & POLICY

Contacts

Naïm Darghouth, Tracking the Sun and demographics data, ndarghouth@gmail.com
Joachim Seel, Utility Scale Solar report data, jseel@lbl.gov
Dev Millstein, Wholesale market electricity price data, dmillstein@lbl.gov
Joe Rand, Interconnection Queue data, jrand@lbl.gov
Kenneth Gillingham, Yale-LBNL Collaborations, kenneth.gillingham@yale.edu

Acknowledgements

This work was funded by the U.S. Department of Energy Solar Energy Technology Office, under Contract No. DE-AC02-05CH11231.

ELECTRICITY MARKETS & POLICY

Questions?

Disclaimer

Copyright Notice