PV Real-World Data, Challenges and Approaches to Generate Synthetic Data

Availability of Real-World PV data

- Major obstacle: Privacy concern on PV generation data

- Other sources of data which can play the alternative role to have access to the PV generation data:
  - Meteorological datasets,
    - ✓ Available for different site locations, at different times of the year, and with a variety of time resolutions.
    - Concern: It needs to be processed before it can be used in power system studies.

➢ Learning about the methods which convert the meteorological data to PV generation data is necessary.

Different Approaches to Generate PV Output Synthetically

- **Physical Methods (without using historical PV output data):**
  - Based on the PV power plant characteristics, such as location, different meteorological data
  - Are considerably simple (they are only based on the global solar irradiance)
  - The accuracy is higher when the weather conditions are stable
  - These models are highly sensitive to weather forecasting and they should be designed specifically for a particular PV plant and location.

- **Methods using historical PV output data**
  - Synthetically generates PV output for either another period of time OR for nearby PV plants.

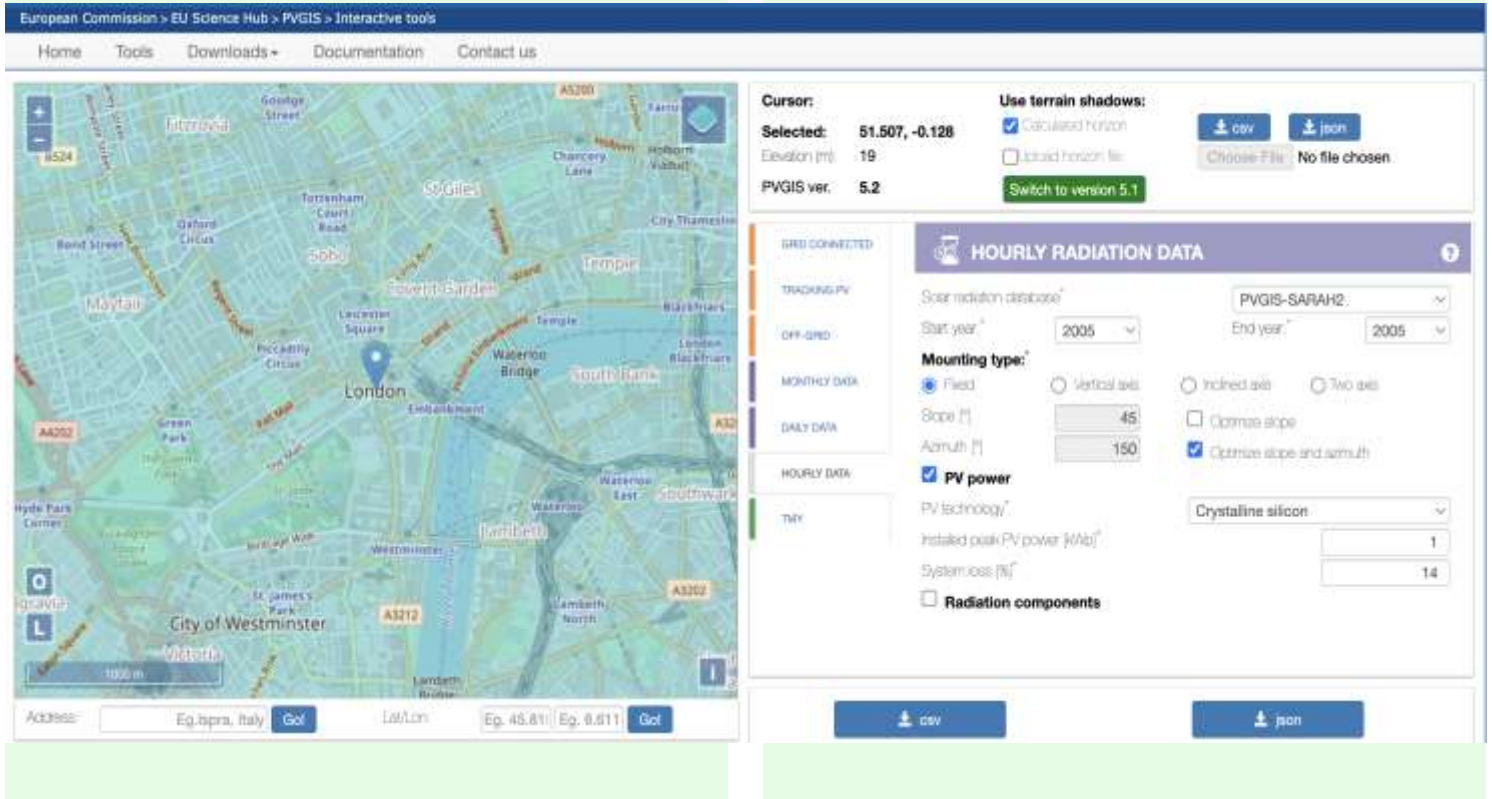
Physical Methods

- Popular software and website for synthetically generate PV output:
  - PVGIS [1]
  - PVSYST [2]
  - SAM [3]
  - RETSCREEN [4]
  - PV*SOL [5]

- Other useful software :
  - HOMER [6], etap [7], etc.
- Online free photovoltaic software :
  - Global Solar Atlas [8], DIAFEM [9], SISIFO [10], EASY-PV [11], and PVWATTS free solar calculator [12], etc.
  - ✓ Provides hourly or monthly PV generation.

European Commission

PVGIS



- PVGIS is a web site that gives you information about solar radiation and PV system performance.

- Outputs:
  - Daily PV output in each month.
  - It also shows the yearly average production per day and gives statistics data.

- PVSYST software is a European-based PV system predictor developed by the University of Geneva for the European Energy Center.

- Input:
  - meteorological and global irradiation data from Meteonorm and NASA-SSE. (other external databases can be introduced)
  - It has got an extensive database of available PV panels, inverters, and batteries which is directly updated by the manufacturers

- Outputs:
  - PV production
    - Monthly diagram of standard production per installed is generated, including the following information: power loss, total energy produced, the effective power to the output of the generator, the energy injected into the network, the electrical efficiency of the facility, and a table with the solar irradiation data (horizontal global irradiation, incident irradiation at the receptor).

NREL NATIONAL RENEWABLE ENERGY LABORATORY System Advisor Model (SAM)

- System Advisor Model (SAM) is provided by the National Renewable Energy Laboratory (NREL) of the U.S. Department of Energy.

- Inputs:
  - hourly weather data
  - data that the user may introduce.(SAM includes a wide database of PV panels and inverters, and it is possible to define new elements)

- Outputs:
  - hourly PV simulation model with performance, cost, and finance models to calculate energy output, energy costs, and cash flows.

- RETSCREEN is managed by the research center of Natural Resources Canada (NRCan).

- Inputs:
  - Type of panel (poly-Si, mono-Si, etc.),
  - Total PV installed power of facility,
  - Efficiency, estimated electrical losses, losses caused by shadows.
  - Inverter data, defined by the user, are input power, performance and estimated losses.

- Outputs:
  - Daily solar horizontal irradiation by month,
  - The estimation of annual electricity production (MWh),
  - Environmental benefits (CO_2 emissions savings, equivalent in barrels of petroleum), return on investment period, evolution of the benefits over the operating life of the installation, internal rate of return (IRR), net present value (NPV), and unit cost of energy.

- PV*SOL is software for planning, design, and simulation of PV systems developed by Valentin Software.

- Inputs:
  - Meteorological data which is based on Meteonorm database. It is also possible for users to load their own climate data to the project.
  - Panels and inverters and wiring information

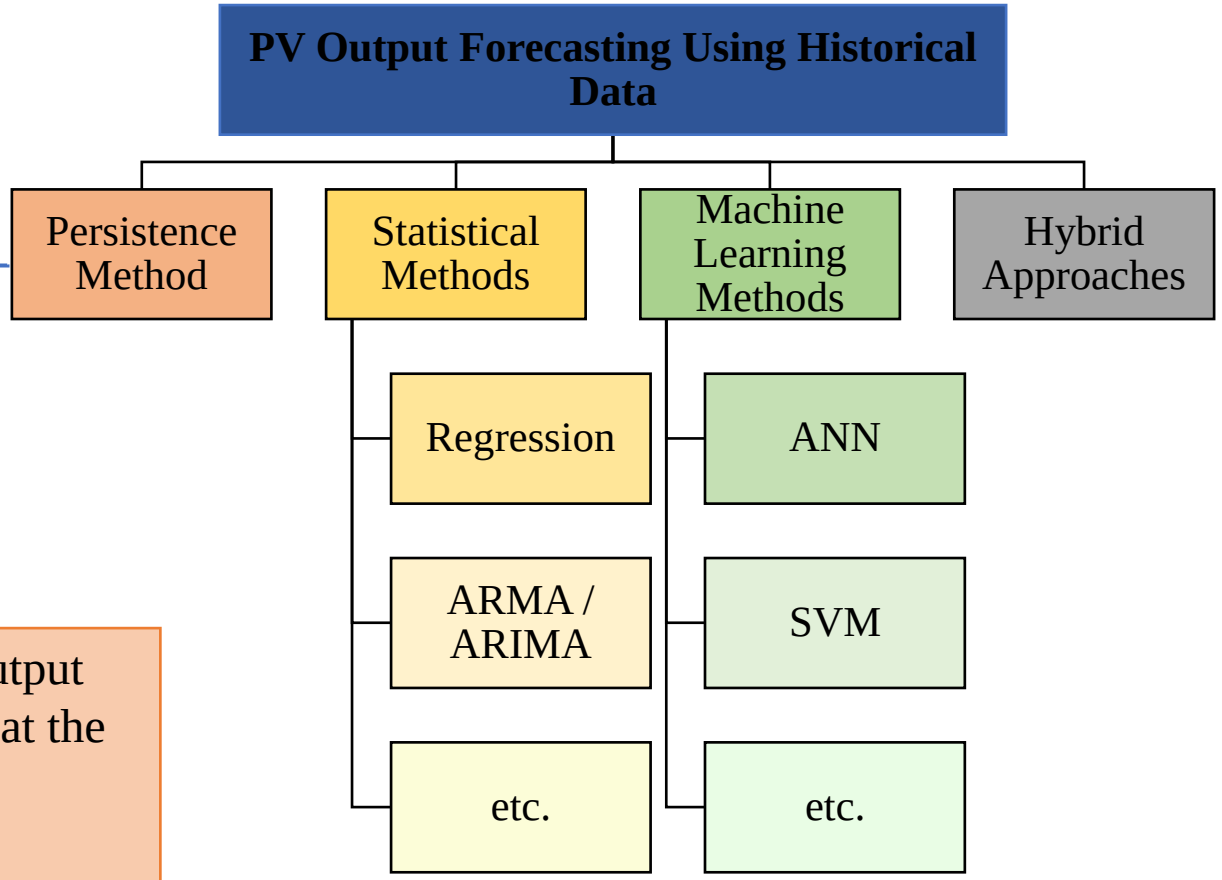
- Outputs:
  - The predicted monthly electrical production of the system, its performance ratio, the irradiance received on the module's plane and its temperature, a general energy balance of the system, and a financial analysis of the installation.

PV*SOL

Methods Using Historical PV Output Data

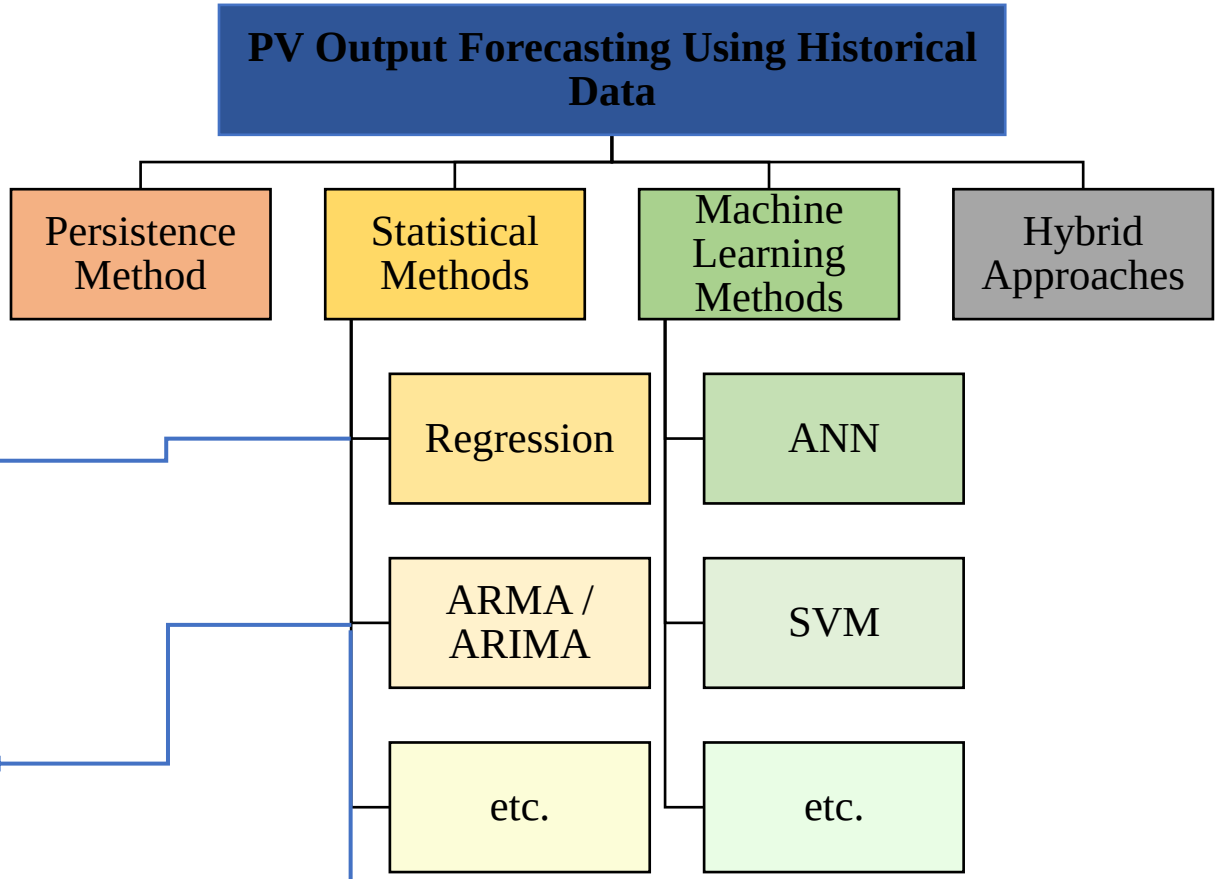
- Having access to PV data for a few sites in a region, but we seek for PV production in nearby sites.
- Availability of PV data for a short time interval, but we require data for longer span of time.

✓ We need to synthetically generate data using historical data and other available meteorological data.



For DAH: Synthetic PV output equals to the previous day at the similar time.

- Suitable for short time forecasting



- It stablishes a relation between the explanatory variable and dependent variables.
  PV output: dependent var.
  Meteorological & historical data: explanatory var.
- **Weakness**: A mathematical model and several explanatory variables are required to design a good regression-based estimation model.

- **A**uto **R**egressive **M**oving **A**verage
- Popular, because of its capability to extract the statistical properties.

- ARIMA, i.e., an extension of ARMA, an integrated part removes any non-stationarity from the data.

$$X(t) = \sum_{i=1}^{p} \alpha_i X(t-i) + \sum_{j=1}^{q} \beta_j e(t-j)$$

PV output — Regression part — White noise

PV Output Forecasting Using Historical Data

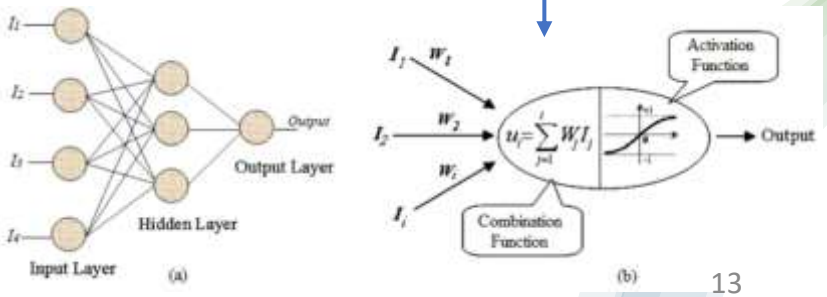
- Persistence Method
- Statistical Methods
  - Regression
  - ARMA / ARIMA
  - etc.
- Machine Learning Methods
  - ANN
  - SVM
  - etc.
- Hybrid Approaches

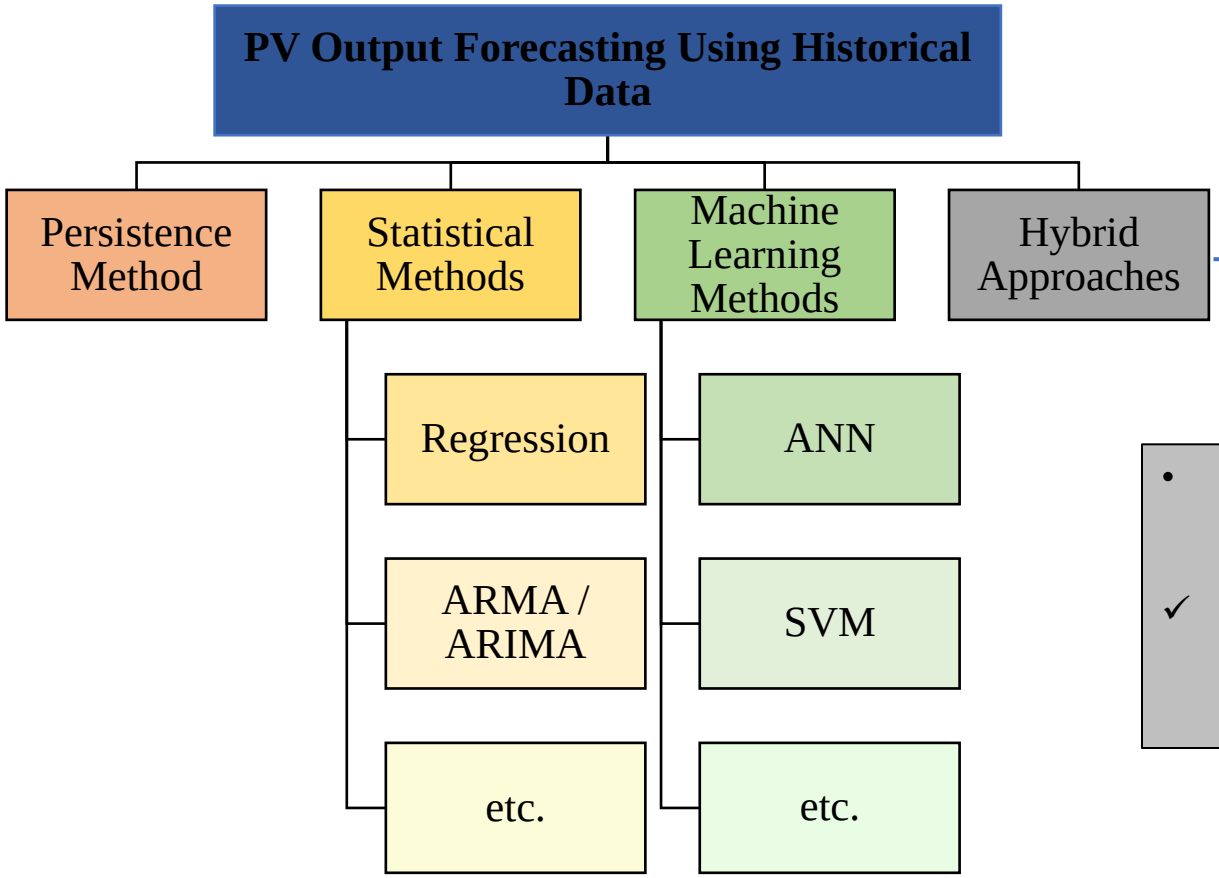
Weakness: large data is needed to have good ML method results.

✓ An intelligent technique which can handle linear and nonlinear data patterns.

- **A**rtificial **N**eural **N**etwork is highly used in forecasting PV power generation because of the nonlinearity in meteorological data.

- **S**upport **V**ector **M**achine is a supervised ML method.
- **Weakness**: The performance of the SVM-based model highly depends on the selection of three parameters,
  (i.e.,penalty (C), tube radius (ε), and the kernel function's parameter.)





- The combination of two or more of the previous approaches.
- ✓ Usually shows better results than a single model.

Name	Type	Description	No. Locations	Temporal Resolution	Time Duration	total size	Format	Link	Data URL	user manual / contributer Email
Solar Power Data for Integration Studies	synthetic PV power plant data	1 year of 5-minute solar power and hourly day-ahead forecasts for approximately 6,000 simulated PV plants.	51 location in US	5 minute and Hourly	2006	3 GB	csv	https://www.nrel.gov/grid/solar-power-data.html	https://www.nrel.gov/grid/solar-power-data.html	ningzhou@binghamton.edu
ARPA-E PERFORM datasets	Time-coincident load, wind, and solar data including actual and probabilistic forecast	Time-coincident load, wind, and solar data including actual and probabilistic forecast datasets at 5-min resolution for ERCOT, MISO, NYISO, and SPP. For ERCOT actuals are provided for 2017 and 2018 and forecasts for 2018, and for the remaining ISOs actuals are provided for 2018 and 2019 and forecasts for 2019.	ERCOT, MISO, NYISO, and SPP.	5-minute	2017-2019		.h5	https://data.openei.org/submissions/5772	AWS S3 Explorer for the Open Energy Data Initiative (openei.org)	–
Solar PV Power Generation measurement & Forcasting Data – Belgium	PV Power Generation measurement & Forcasting Data	Solar PV Power Generation Measurement & Forecasting Data	Belqium (country wide)	15- minutes	2012-present	N/A	XLS	https://bigdata.seas.gwu.edu/data-set-10-belgium-solar-pv-power-generation-data-set-belgium/	https://www.elia.be/en/grid-data/power-generation	https://www.elia.be/en/grid-data/power-generation
PV Rooftop Dataset	PV Rooftop Generation	NREL PV Rooftop Database (PVRDB) is a lidar-derived, geospatially-resolved dataset of suitable roof surfaces and their PV technical potential for 128 metropolitan regions in the United States. The PVRDB is downloadable at the AWS S3 Bucket by city and year of lidar collection. Five geospatial layers are available for each city and year.	128 metropolitan regions in the United States		2006-2013	318.64 GB		https://data.openei.org/submissions/4	https://data.openei.org/s3_viewer?bucket=oedi-data-lake&prefix=pv-rooftop%2F	–

Actual_30.65_-87.65_2006_DPV_36MW_5_Min.csv

DA_30.45_-88.25_2006_UPV_70MW_60_Min.csv

HA4_30.55_-87.75_2006_DPV_36MW_60_Min.csv

Name	Type	Description	No. Locations	Temporal Resolution	Time Duration	total size	Format	Link	Data URL	user manual / Contributer Email
Solar-to-Grid Public Data File for Utility-scale and Distributed Photovoltaics Generation, Capacity Credit, and Value	hourly project-level generation data	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.		hourly	2012- 2020	8.69 GB		https://catalog.data.gov/dataset/solar-to-grid-public-data-file-for-utility-scale-upv-and-distributed-photovoltaics-dpv-gen	https://data.openei.org/submissions/4503	–
Data-HM	PV generation, Wind Turbine, Active load, Reactive load	data for a microgrid containing hourly data for a year		hourly	1 year	4 MB	excel	https://github.com/ZepLiang/Supplements-HMs	https://github.com/ZepLiang/Supplements-HMs/blob/main/Data-HM.xlsx	–
Rooftop Solar Data Set – Australia	PV generation	The half-hour electricity data is for 300 homes with rooftop solar systems that are measured by a gross meter that records the total amount of solar power generated every 30 minutes.	300 homes in Ausgrid's electricity network area	half- hour	July 2010- June 2013	70 MB	excel	https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data	https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data	–
Solar home monthly data	PV generation	The monthly electricity data is for 2,657 solar homes with rooftop solar systems that have a gross metering configuration. In addition, a dataset of 4,064 non-solar homes is provided over the same time period in order to compare electricity usage patterns between the two datasets.	2,657 solar homes with rooftop solar systems that have a gross metering configuration. Ausgrid's electricity network area	monthly basis	1 Jan 2007- 31 Dec 2014	13 MB	excel	https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data	https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data	–
DKASC – Australia	Solar Power Data	Desert Knowledge Australia Solar Centre (DKASC)	Australia	10second	2008-Present	N/A	CSV	https://bigdata.seas.gwu.edu/data-set-24-dkasc-solar-power-data-set-australia/	https://dkasolarcentre.com.au/locations/alice-springs	https://dkasolarcentre.com.au/locations/alice-springs
Solar Generation Data Set – Worldwide	Solar Generation Data	Solar Generation Data	world-wide	5- minutes	N/A	N/A	HTML	https://bigdata.seas.gwu.edu/data-set-29-solar-generation-data-set-worldwide/	https://pvoutput.org/outputs.jsp?p=2&df=20180319&dt=20180319&o=date&d=desc	https://pvoutput.org/outputs.jsp?p=2&df=20180319&dt=20180319&o=date&d=desc
PV GECAD LASIE	PV Generation Data;	PV Generation Data; Solar Irradiance		5 minutes	2013		XLS	https://bigdata.seas.gwu.edu/data-set-61-summer-pv-generation-data-set-brazil/	https://site.ieee.org/pes-iss/data-sets/#canizes2015	https://site.ieee.org/pes-iss/data-sets/#canizes2015
PV Generation Data Set	PV Generation Data; Solar Irradiance	PV Generation Data; Solar Irradiance		1 minute	2016		XLS	https://bigdata.seas.gwu.edu/data-set-63-pv-generation-data-set/	https://site.ieee.org/pes-iss/data-sets/#canizes2015	https://site.ieee.org/pes-iss/data-sets/#canizes2015
PV Generation Data Set	PV Generation Data	PV Generation Data		1 minute	2013	255 MB	CSV	https://pages.nist.gov/netzero/data.html#download_data	https://pages.nist.gov/netzero/data.html#download_data	–
PV Generation Data Set	PV Generation Data	PV Generation Data		1 minute	2014	219 MB	CSV	https://pages.nist.gov/netzero/data.html#download_data	https://pages.nist.gov/netzero/data.html#download_data	–
V GECAD N	Photovoltaic generation and temperature	Photovoltaic generation and temperature	PV generation and temperature	5 minutes	2019	4 MB	XLS	https://site.ieee.org/pes-iss/data-sets/#canizes2015	https://site.ieee.org/pes-iss/data-sets/#canizes2015	https://site.ieee.org/pes-iss/data-sets/#canizes2015

Name	Type	Description	No. Locations	Temporal Resolution	Time Duration	total size	Format	Link	Data URL	user manual / Contributer Email
Cambium	Emission Reduction	Cambium data sets contain hourly emission, cost, and operational data for modeled futures of the U.S. electric sector with metrics designed to be useful for long-term decision-making.	-	Yearly basis	N/A	N/A	.CSV	https://scenarioviewer.nrel.gov/	https://scenarioviewer.nrel.gov/	_
Circumsolar Radiation Data:	Solar Radiation	Contains detailed intensity profiles of the solar and circumsolar region, direct normal radiation data, and total hemispherical solar radiation data for 11 U.S. locations from 1976 to 1981.	11 US locations	-	1976-1981		.TXT	https://www.nrel.gov/grid/solar-resource/circumsolar.html	https://www.nrel.gov/grid/solar-resource/circumsolar.html	–
CONFRRM	Irradiance Study network.	Provides high-quality data for determining site-specific resources as well as data for the validation and testing of models to predict available resources based on meteorological or satellite data.	12 US locations	monthly basis	1997-2012		.CSV	https://www.nrel.gov/grid/solar-resource/confrrm.html	https://www.nrel.gov/grid/solar-resource/confrrm.html	–
NREL Solar Radiation Research Laboratory – USA	Solar Radiation Data	The SRRL is an outdoor laboratory located on South Table Mountain, a mesa providing excellent solar access throughout the year, overlooking Denver.	Colorado (United States)	1- minute	1981- present	N/A	.TXT	https://bigdata.seas.gwu.edu/data-set-9-nrel-solar-radiation-data-set/	https://midcdmz.nrel.gov/apps/day.pl?BMS	–
MIDC	irradiance and meteorological data	Provides irradiance and meteorological data from stations throughout the United States	35 stations in US and	monthly basis	from 1989		HTML	https://midcdmz.nrel.gov/	https://midcdmz.nrel.gov/	_
NASA Remote Sensing Validation Data: Saudi Arabia	solar radiation monitoring network	The data were made available to support validation of satellite data products related to the NASA Earth Observing System project to evaluate long-term climate trends based on measurements from EOS Terra Platforms.	12 stations across Saudi Arabia	monthly basis	1998-2003	N/A	excel	https://www.nrel.gov/grid/solar-resource/saudi-arabia.html	https://www.nrel.gov/grid/solar-resource/saudi-arabia.html	–
National Oceanic and Atmospheric Administration Solar Data	solar radiation and other weather elements required by solar energy technology users	(NOAA) solar radiation monitoring network has existed for about 75 years, with varying numbers of stations. With Department of Energy help, some of the data were archived and published by the National Climatic Center.	39 locations acroos US		1947-2022		.TXT	https://www.nrel.gov/grid/solar-resource/noaa.html	https://www.nrel.gov/grid/solar-resource/noaa.html	–
National Solar Radiation Database	Solar data , and the meteorological data	The National Solar Radiation Database (NSRDB) is a serially complete collection of hourly and half-hourly values of meteorological data and the three most common measurements of solar radiation: global horizontal, direct normal and diffuse horizontal irradiance.	US + a growing subset of international locations including india	5-10-30-60 minutes	1998-2021			https://nsrdb.nrel.gov/	https://nsrdb.nrel.gov/data-sets/how-to-access-data	–
Solar Energy Meteorological Research and Training Sites Data Set	Meteorological data	Solar Energy Meteorological Research and Training Sites hourly data cover four locations across the United States between 1979 and 1984.	4 location inside US	monthly and hourly	1979 -1984		.TXT	https://www.nrel.gov/grid/solar-resource/semrts.html	https://www.nrel.gov/grid/solar-resource/semrts.html	–
Solar Resource Variability	spatial and temoral data and maps	The Solar Resource Variability collection includes data and maps that demonstrate the variability in time and space of the solar resource across the United States from 1998 to 2005.			1998-2005	100 MB	.CSV	https://www.nrel.gov/grid/solar-resource/variability.html	https://www.nrel.gov/grid/solar-resource/variability.html	–
Spectral Solar Radiation Data Base	radiation	The spectral solar radiation data base represents a range of atmospheric conditions (or climates) and is applicable to several types of solar collectors.	3 loactions inside US	monthly basis	1997, 1998, 1986, 1988		.TXT	https://www.nrel.gov/grid/solar-resource/spectral-solar.html	https://www.nrel.gov/grid/solar-resource/spectral-solar.html	–

Name	Type	Description	No. Locations	Temporal Resolution	Time Duration	total size	Format	Link	Data URL	user manual / Contributer Email
WEST Associates Solar Monitoring Network	horizontal and direct normal solar irradiances and dry-bulb temperatures	In the mid-1970s, Southern California Edison and Western Energy Supply and Transmission (WEST) Associates created the WEST Solar Monitoring network, which included 52 stations in six Western states (Arizona, California, Colorado, Nevada, New Mexico, and Wyoming)	52 station inside US	15-minute	1976- 1980		TXT	https://www.nrel.gov/grid/solar-resource/west.html	https://www.nrel.gov/grid/solar-resource/west.html	–
Puerto Rico Grid Resilience and Transition to 100% Renewable Energy	boundaries, habitats, hazards, infrastructure, and topography throughout Puerto Rico	PR100 is a comprehensive analysis of stakeholder-driven pathways for Puerto Rico to achieve its goal of 100% renewable energy by 2050. The data includes boundaries, habitats, hazards, infrastructure, and topography throughout Puerto Rico. Most of the data is in geospatial and json formats. Links to project background, history, and planning are also included along with the data.				4.68GB	.json	https://data.openei.org/submissions/5749	https://data.openei.org/s3_viewer?bucket=oedi-data-lake&prefix=PR100%2F	–
PVDAQ Public Datasets	PV performance raw data & enviromental data	The Photovoltaic field array (PVDAQ) data is composed of time-series, raw performance data taken through a variety of sensors connected to a PV array. NREL source data is typically aggregated into the main database every 24 hours. Data is then processed to the NREL PVDAQ data lake on a monthly basis. The PVDAQ data is partitioned by system_id, year, month and day. Raw data is reported at 15 minute increments in ISO 8601 date and time. The timestamp is striped and data is averaged daily.	157 PV systems	15 sec (vary between systems, e.g., hourly)	2000-present	N/A	.CSV and .json	https://data.openei.org/submissions/4568	https://data.openei.org/s3_viewer?bucket=oedi-data-lake&prefix=pvdaq%2Fcsv%2F	–
Tracking the Sun	PV Pricing	For more than 1.3 million individual PV systems, representing 81% of U.S. residential and non-residential PV systems installed through 2017. The analysis of installed pricing trends is based on a subset of roughly 770,000 systems with available installed price data.	more than 1.3 million individual PV systems, representing 81% of U.S. residential and non-residential PV systems	-	2017			https://data.openei.org/submissions/3		–
PVRDB-PR		NREL PV Rooftop Database for Puerto Rico (PVRDB-PR) is a lidar-derived, geospatially-resolved dataset of suitable roof surfaces and their PV technical potential for virtually all buildings in Puerto Rico.	78 Counties in Puerto Rico	-	-			https://data.openei.org/submissions/2862		–
NYSERDA Distributed Energy Resources (DER) Data – USA	Distributed Energy Resources (DER) Data	Data can be viewed as aggregated summaries or as granular performance data from individual projects.	New York State, USA	1 hour	2016-2017		csv	https://bigdata.seas.gwu.edu/data-set-3-nyserda-distributed-energy-resources-der-data/	https://der.nyserda.ny.gov/	ningzhou@binghamton.edu
High-Resolution Solar Radiation Data – CANADA	solar radiation	Files are named by date and the identifier of the unit from which the data is collected. The location of each unit is saved in KML files, as described below. Data was taken on the date corresponding to the folder name containing the file (using the format yyyymmdd).	Ontario and Quebec (Canada)	1 sec	several days in 2014	1 GB	csv	https://bigdata.seas.gwu.edu/data-set-7-high-resolution-solar-radiation-data-set/	https://www.nrcan.gc.ca/energy/renewable-electricity/solar-photovoltaic/18409	https://www.nrcan.gc.ca/energy/renewable-electricity/solar-photovoltaic/18409
Humboldt State University (SoRMS) Solar Radiation Data – USA	Solar irradiance and meteorological data	Solar irradiance and meteorological data	Humboldt State University (USA) (Campus wide)	1- minute	2007-present		TXT	https://bigdata.seas.gwu.edu/data-set-13-humboldt-state-university-sorms-radiation-data-set/	https://midcdmz.nrel.gov/apps/daily.pl?site=HSU&start=20070502&yr=2019&mo=4&dy=22	h.sangrody@gmail.com
University of Oregon (SRML) Solar Radiation Data – USA	Solar irradiance and meteorological data	Solar irradiance and meteorological data	University of Oregon (USA) (Campus wide)	1- minute	2016-present	N/A	TXT	https://bigdata.seas.gwu.edu/data-set-14-university-of-oregon-srml-radiation-data-set/	https://midcdmz.nrel.gov/apps/daily.pl?site=UOSMRL&start=20160819&yr=2019&mo=4&dy=22	https://midcdmz.nrel.gov/apps/daily.pl?site=UOSMRL&start=20160819&yr=2019&mo=4&dy=22
Solar Radiation Data – USA	solar radiation	Hourly solar irradiance, US, 10 km squre resolution	US, 10 km squre resolution	hourly	1998- present	N/A	CSV	https://bigdata.seas.gwu.edu/data-set-22-solar-radiation-data-set/	https://www.solaranywhere.com/	https://www.solaranywhere.com/
Weather Data Set for Load and Generation Forecast – Worldwide	Weather Data	Weather Data	world-wide	hourly			CSV	https://bigdata.seas.gwu.edu/data-set-28-weather-data-set-worldwide/	https://mesonet.agron.iastate.edu/request/download.phtml?network=NY_ASOS	https://mesonet.agron.iastate.edu/request/download.phtml?network=NY_ASOS
MesoWest Weather Data Set for Wind and Solar Integration – USA	Weather Data,Wind Speed Data,Temperature Data	Weather Data,Wind Speed Data,Temperature Data	US	5 minutes	1997-present	N/A	CSV	https://bigdata.seas.gwu.edu/data-set-35-mesowest-weather-data-set-for-wind-and-solar-integration/	https://mesowest.utah.edu/cgi-bin/droman/download_api2.cgi?stn=KSLC	–

References

[1] https://re.jrc.ec.europa.eu/pvg_tools/en/tools.html#PVP

[2] https://www.pvsyst.com/

[3] https://sam.nrel.gov/

[4] https://www.nrcan.gc.ca/maps-tools-and-publications/tools/modelling-tools/retscreen/7465

[5] https://pvsol.software/en/

[6] https://www.homerenergy.com/

[7] https://etap.com/

[8] https://globalsolaratlas.info/

[9] https://www.agenciaandaluzadelaenergia.es/Diafem/

[10] https://www.sisifo.info/en/default

[11] http://www.easy-pv.co.uk/

[12] https://pvwatts.nrel.gov/

Thank you!