Fall 2022

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

Summary

Despite of the current trends in data-driven studies of power systems, and in particular renewable energy resources such as Photovoltaics (PVs), having access to real-world PV data still remains a difficult task. This report aims to summarize the existing challenges in accessing real-world PV generation data in the first step and accordingly, recommends the potential approaches to synthetically generate the PV generation data from other publicly available databases like meteorological data. Additionally, the pros and cons of the understudy methods are discussed. In the second step, the PV related data sets that are publicly available are introduced and classified in this report.

1. Introduction

To mitigate the impact of climate change and global warming, the use of renewable energy is increasing drastic. Compared to other types of renewable energy resources, PV has undergone faster development, and are now the mature and economically competitive technology [1]. Although the utilization of solar irradiation to generate electricity is currently at a fast deployment pace, its natural variability still imposes a crucial barrier to overcome. On the other hand, the growing use of data-driven methods in power system studies which involves PVs require access to PV datasets with high fidelity.

The importance of privacy in contemporary globalized information societies has been widely discussed and is undisputed. In other words, data may be of little value to the average user, but the data of masses of users put together is of immense value to big businesses. As it is said, “the world’s most valuable resource is no longer oil, but data”. In this regard, accessing to real-world PV data is a big deal. Thus, due to the privacy concerns, one should not expect too many real-world PV generation dataset to be publicly available. While the privacy concern on PV generation data might be the major obstacle, there are other sources of data which can play the alternative role and pave the way towards having access to the PV generation data. Among these alternatives are the meteorological datasets, which are available for different site locations, at different times of the year, and more importantly, with a variety of time resolutions. The concern on the meteorological data is that it needs to be processed before it can be used in power system studies. Therefore, learning about the methods which convert the meteorological data into PV generation data is necessary.

In this study, we categorize the existing conversion methods into two classes. Then, each class will further be discussed in detail.

2. PV Generation Estimation Using Physics-Based Method

Physics-based models are designed based on the PV power plant characteristics, such as location, meteorological variables, and system sizing parameters like brand of the PV modules and inverters, etc. These methods are somewhat easy to use, as they do not need many complicated calculations. However, the associated models are highly sensitive to weather forecasting data. As a result, the desired model should be developed specifically for a particular location and time of the year.

There are multiple commercial software available, e.g., HOMER [2] and etap [3] and plenty of online free photovoltaic software, e.g., Global Solar Atlas [4], DIAFEM [5], SISIFO [6], EASY-PV [7], and PVWATTS free solar calculator [8], which takes the meteorological and geographical data as the input and predicts the PV generation with respect to the brand and capacity of the PV unit. Here, we briefly introduce a couple of the most commonly used tools to estimate PV production which do not need any historical data.

2.1 PV*SOL [9]

PV*SOL is a well-known software for simulation of PV systems developed by Valentin Software. The climate data supplied by PV*SOL is based on Meteonorm database. It is also possible for users to load their own climate data to the project. The temporal resolution for the simulation can be selected to be either hourly or minutely. The software also has an extensive database of PV modules, inverters, storage systems, and other devices. Moreover, new devices can be defined for inclusion in the project. Wiring losses can also be considered in simulation. The output data includes the predicted 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. Furthermore, a report containing the configuration of the system, its electrical diagrams, and the results can be generated. PV*SOL needs subscription and has a 30-day trial version.

2.2 PVsyst [10]

This software is a European-based PV system predictor developed by the University of Geneva for the European Energy Center. It is suitable for studying stand-alone and grid-connected PV systems. Meteorological and global irradiation data in PVsyst are imported from Meteonorm and NASA-SSE. It is also possible for users to load their own climate data to the project.

PVsyst has got a very extensive database of available PV panels, inverters, and batteries which is directly updated by the manufacturers, but can also be imported from the PHOTON database and even permits the definition of "ad hoc" panels of different technologies: mono-Si, poli-Si, and thin film. The required inputs consist of the design of the system, distribution of PV panels, connection with the inverter, tilted angle, and sun-tracking system specification. Also, the outputs are annual PV production, as well as the performance factor. A 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). This software needs subscription, but a complete version is free for one month. Consequently, the Demo mode of the software has got limited options.

2.3 PVGIS [11]

The Photovoltaic Geographical Information System (PVGIS) provides a map-based inventory of solar energy resource and an assessment of electricity generation from PV systems in Europe, Asia, Africa, and South America. The database includes monthly and yearly average values of the global irradiation on horizontal and inclined surfaces, as well as climatic parameters needed for an assessment of the potential PV electricity generation [12]. The main outputs of PVGIS consists of daily and monthly average global irradiation data, as well as daily and monthly average PV power production.

The databases are widely implemented in Europe, Africa, and Asia with regular updates, and their reliability has been demonstrated in a large number of scientific papers [13] and [14].

2.4 System Advisor Model (SAM) [15]

System Advisor Model (SAM) is provided by the National Renewable Energy Laboratory (NREL) of the U.S. Department of Energy (DOE). SAM is suitable for comparing power system costs and performance across the range of solar technologies and markets, from PV systems for residential and commercial markets to concentrating solar power and large PV systems for utility markets. It provides hourly simulation model with performance, cost, and finance models to calculate energy output, and energy costs. SAM input consists of hourly weather data and the software can read files in either TMY2, TMY3 (typical meteorological year [16]), and EPW (energy plus weather data file) formats from over 20 sources [17] and data that the user may introduce. SAM includes a wide database of PV panels and inverters, and it is possible to define new elements. SAM is a free program.

2.5 RETScreen [18]

This software is managed by the research center of Natural Resources Canada (NRCan). It can be used to evaluate the energy production and savings, costs, emission reductions, financial viability, and risk for various types of renewable energy resources and energy efficient technologies (RETs) [19], [20]. Meteorological database includes both the ground-based meteorological data and NASA’s satellite-derived meteorological data sets. The required input consist of: type of panel, total PV installed power of facility, efficiency, estimated electrical losses, losses caused by shadows. Moreover, the inverter data includes input power, performances, and estimated losses. Finally, the output are: different tables and graphs including 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. Although this software was previously a free software, it has become becme a paid program recently.

3. PV Generation Estimation Using Historical Data

Suppose we have access to PV generation data for a few sites in a region, but we seek for PV production in nearby sites; or PV output data is available for a short period of time, but data at a larger time span is required. Under this circumstance, we can generate PV output data synthetically from the historical data as well as the meteorological data. To this end, we can categorize the existing approaches into four groups.

3.1 Persistent Method

Suppose PV output data is available for a particular day, and we seek to estimate the PV generation for the latter day. In the persistent model, the estimated PV power output is equal to the actual power output of the previous day at a similar time. In this method, only the historical PV power output data are required to estimate the PV power generation. In this model, the estimated PV power output is assumed to remain the same at the same time of the previous or following day. This model is generally used for the short-term forecasting. The estimation error of this model depends on the stability of the weather condition. If the weather condition does not frequently change, then the PV power output of the previous day is a realistically a good indicator of the power output of the next day. However, as the time range of estimation increases, the accuracy of this model is lowered [21].

3.2 Statistical Methods

In the statistical methods, the PV power generation is estimated by the statistical analysis of the different input variables. In general, these approaches are adopted for the short-term estimation of the PV power generation. Some of well-known statistical approaches are as follow:

3.2.1 Regression method

The regression method is a statistical method used to establish a relationship between the explanatory and dependent variables. In case of the PV power estimation, PV output is considered as a dependent variable, and the historical data and the meteorological variables are considered as explanatory variables. The regression model which uses temperature in addition to solar irradiance as the input, leads to better results compared to the case in which only one of them is considered as the input. As the weakness of this method one can refer to the fact that a mathematical model and several explanatory variables are required to design a regression-based estimation model.

3.2.2 Autoregressive Moving Average (ARMA) Model

The ARMA model consists of the combination of the two basic models: AR and MA models. ARMA is suitable for estimating the PV power generation from the specified time-series data. The main reason of the popularity of the ARMA model is its capability to extract the statistical properties. The main shortcoming of the ARMA model is that the utilized time series data must be stationary [22]. An extension of this model is widely used for the different estimation models with an acceptable accuracy which is known as the AR integrated MA (ARIMA) [23]. In ARIMA model, an integrated part removes any non-stationarity from the data [24].

3.3 Machine Learning Methods

The machine learning model is an intelligent technique, and it can handle linear, non-linear, and non-stationary data patterns. It is worth to mention that to have an acceptable result, machine-learning methods require a large dataset to estimate the PV power generation. Some of the popular approaches in machine learning method are as follows:

3.3.1 Artificial Neural Network (ANN)

This method has been used in different prediction applications, including the estimation of PV power generation. ANN is widely used in forecasting the PV power generation in many research, because of non-linearity that exists in meteorological data. ANN is more suitable compared with the statistical methods when a non-linear and complicated bonding exists between the data without any prior assumption.

3.3.2 Support Vector Machine (SVM)

SVM is a supervised machine-learning method based on structural risk minimization (SRM) principle. SRM minimizes an upper bound of the expected risk. Therefore, SVM can minimize the error of the training data. However, SVM was recently extended to the domain of regression problems. The application of SVM in time series regression is known as support vector regression (SVR). The estimation of PV power generation is a typical time series analysis problem; thus, SVR is a suitable method in this case [25]. To develop the SVR-based model for forecasting PV power generation, three parameters dominate the performance of the model [26]. These parameters are penalty (C), which determines the penalties for estimation errors, tube radius (ε), which determines the data inside the tube to be ignored in regression, and the kernel function's parameter. The suitable values of C, ε, and kernel function's parameter must be selected to develop the appropriate forecasting model. The performance of this model depends largely on the selection of the three parameters, which is a limitation of this method. [27]

3.4 Hybrid Methods

The combination of two or more techniques is used to design an estimation model known as the hybrid model. The hybrid model shows better results than a single model for different forecasting problems by combining the advantages of each individual technique. Fuzzy inference model with recurrent neural network has been successfully applied in PV power forecasting [28]. In this case, fuzzy inference model has been utilized to smoothen the meteorological data, which is used for estimating PV power generation.

4. Available Datasets

While due to privacy issues, not many PV generation datasets might be publicly available, but there are still resources which provide such information for free. Meanwhile, as discussed in previous section, lots of meteorological datasets are available where the PV generation data can be synthetically constructed from them. In this section, we have provided the list of available datasets for both PV generation data, and meteorological data and compared them from several aspects as in Table 1.

Table 1. Publicly available PV generation and meteorological databases

Name	Type	Description	No. Locations	Temporal Resolution	Time	total size	Format	Link	Data URL	User Manual
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	–
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 Forcasting Data – Belgium	PV Power Generation Forcasting Data	Solar PV Power Generation 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	–

Name	Type	Description	No. Locations	Temporal Resolution	Time	total size	Format	Link	Data URL	User Manual
Solar-to-Grid Public Data File for Photovoltaics Generation,	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	XLS	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	XLS	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	monthly basis	1 Jan 2007 - 31 Dec 2014	13 MB	XLS	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

Name	Type	Description	No. Locations	Temporal Resolution	Time	total size	Format	Link	Data URL	User Manual
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	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	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
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	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	–

Name	Type	Description	No. Locations	Temporal Resolution	Time	total size	Format	Link	Data URL	User Manual
		available resources based on meteorological or satellite data.								
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 subset of other locations	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	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-	https://www.nrel.gov/grid/solar-resource/semrts.html	–

Name	Type	Description	No. Locations	Temporal Resolution	Time	total size	Format	Link	Data URL	User Manual
and Training Sites Data Set								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	–
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.				4.68 GB	.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	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	–

Name	Type	Description	No. Locations	Temporal Resolution	Time	total size	Format	Link	Data URL	User Manual
		increments in ISO 8601 date and time. The timestamp is striped and data is averaged daily.								
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.	about 1.3 million PVs, (81% of U.S. residential & non-residential PVs	-	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/	–
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	–

Name	Type	Description	No. Locations	Temporal Resolution	Time	total size	Format	Link	Data URL	User Manual
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	–

5. Acknowledgement

This work was funded by an INTERN supplemental grant from National Science Foundation to the University of California, Riverside to Grant Number 1711944. This work was also done by supervisory of GridBright and BetterGrids.org.

References

[1] IEA (Internaional Energy Agency). World Energy Outlook-2017; IEA (Internaional Energy Agency): Paris, France, 2017.
[2] https://www.homerenergy.com/
[3] https://etap.com/
[4] https://globalsolaratlas.info/
[5] https://www.agenciaandaluzadelaenergia.es/Diafem/
[6] https://www.sisifo.info/en/default
[7] http://www.easy-pv.co.uk/
[8] https://pvwatts.nrel.gov/
[9] https://pvsol.software/en/
[10] https://www.pvsyst.com/
[11] https://re.jrc.ec.europa.eu/pvg_tools/en/tools.html#PVP
[12] TamizhMani, G.; Ishioye, J.-P.; Voropayev, A.; Kang, Y. Photovoltaic performance models: An evaluation with actual field data. In Proceedings of the Reliability of Photovoltaic Cells, Modules, Components, and Systems SPIE, San Diego, CA, USA, 11–13 August 2008; Volume 7048, p. 70480.
[13] Conceição, R.; Silva, H.G.; Fialho, L.; Lopes, F.M.; Collares-Pereira, M. PV system design with the effect of soiling on the optimum tilt angle. Renew. Energy 2019, 133, 787–796.
[14] Ozdemir, S.; Sahin, G. Multi-criteria decision-making in the location selection for a solar PV power plant using AHP. Meas. J. Int. Meas. Confed. 2018, 129, 218–226.
[15] https://sam.nrel.gov/
[16] Marion, W.; Urban, K. User's Manual for TMY2s (Typical Meteorological Years)—Derived from the 1961–1990 National Solar Radiation Data Base; National Renewable Energy Lab. (NREL): Golden, CO, USA, 1995.
[17] Weather Data for Simulation | EnergyPlus. Available online: https://energyplus.net/weather/simulation
[18] https://www.nrcan.gc.ca/maps-tools-and-publications/tools/modelling-tools/retscreen/7465
[19] Zandi, M.; Bahrami, M.; Eslami, S.; Gavagsaz-Ghoachani, R.; Payman, A.; Phattanasak, M.; Nahid-Mobarakeh, B.; Pierfederici, S. Evaluation and comparison of economic policies to increase distributed generation capacity in the Iranian household consumption sector using photovoltaic systems and RETScreen software. Renew. Energy 2017, 107, 215–222.
[20] Rehman, S.; Ahmed, M.A.; Mohamed, M.H.; Al-Sulaiman, F.A. Feasibility study of the grid connected 10 MW installed capacity PV power plants in Saudi Arabia. Renew. Sustain. Energy Rev. 2017, 80, 319–329.
[21] Perez R, Kivalov S, Schlemmer J, Hemker K, Renné D, Hoff TE. Validation of short and medium term operational solar radiation forecasts in the US. Sol Energy 84, no. 12 (2010): 2161-2172.
[22] Diagne, Maimouna, Mathieu David, Philippe Lauret, John Boland, and Nicolas Schmutz. "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids." *Renewable and Sustainable Energy Reviews* 27 (2013): 65-76.
[23] Box, George EP, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. Time series analysis: forecasting and control. John Wiley & Sons, 2015.
[24] Wan Ahmad WKA, Ahmad S, Ishak A, Hashim I, Ismail ES, Nazar R. Arima model and exponential smoothing method: a comparison. AIP Conference Proceedings: AIP; 2013. p. 1312–21.
[25] Hu J, Gao P, Yao Y, Xie X. Traffic flow forecasting with particle swarm optimization and support vector regression. In: Proceedings of the IEEE 17th international conference on intelligent transportation systems (ITSC); 2014. p. 2267–2268.
[26] Cherkassky, Vladimir, and Yunqian Ma. "Practical selection of SVM parameters and noise estimation for SVM regression." *Neural networks* 17, no. 1 (2004): 113-126.
[27] Das, Utpal Kumar, Kok Soon Tey, Mehdi Seyedmahmoudian, Saad Mekhilef, Moh Yamani Idna Idris, Willem Van Deventer, Bend Horan, and Alex Stojcevski. "Forecasting of photovoltaic power generation and model optimization: A review." *Renewable and Sustainable Energy Reviews* 81 (2018): 912-928.
[28] Yona, Atsushi, Tomonobu Senjyu, Toshihisa Funabashi, and Chul-Hwan Kim. "Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead PV power output correction." *IEEE Transactions on Sustainable Energy* 4, no. 2 (2013): 527-533.