Solar Power Forecast Component Description for New York State

3002025153

Solar Power Forecast Component Description for New York State

3002025153

Technical Update, November 2022

EPRI Project Manager

M. Bello

EPRI
3420 Hillview Avenue, Palo Alto, California 94304-1338 ▪ PO Box 10412, Palo Alto, California 94303-0813 ▪ USA
800.313.3774 ▪ 650.855.2121 ▪ askepri@epri.com ▪ www.epri.com

DISCLAIMER OF WARRANTIES AND LIMITATION OF LIABILITIES

THE FOLLOWING ORGANIZATION, UNDER CONTRACT TO EPRI, PREPARED THIS REPORT:

NATIONAL CENTER FOR ATMOSPHERIC RESEARCH (NCAR)

This is an EPRI Technical Update report. A Technical Update report is intended as an informal report of continuing research, a meeting, or a topical study. It is not a final EPRI technical report.

NOTE

For further information about EPRI, call the EPRI Customer Assistance Center at 800.313.3774 or e-mail askepri@epri.com.

Together…Shaping the Future of Energy®

ACKNOWLEDGMENTS

The following organization, under contract to EPRI, prepared this report:

National Center for Atmospheric Research
P.O. Box 3000
Boulder, CO 80307

Principal Investigator
J. A. Lee

This report describes research sponsored by EPRI. EPRI would like to acknowledge the New York Power Authority (NYPA), the New York State Energy Research and Development Authority (NYSERDA) and the DOE Solar Energy Technologies Office (DOE SETO) for their roles in co-sponsoring this effort.

This publication is a corporate document that should be cited in the literature in the following manner:

Solar Power Forecast Component Description for New York State. EPRI, Palo Alto, CA: 2022. 3002025153.

ABSTRACT

The Solar Power Forecast System is one of the key components of the Solar Forecasting for New York State (NYSolarCast) system that is currently under development by the National Center for Atmospheric Research (NCAR). This Technical Update describes the elements of the Solar Power Forecast System, including data used in model training and testing, configuration files, C++ applications, and Python scripts needed to run this component of NYSolarCast. Figures illustrating sample power forecast output are also provided.

Keywords

Machine learning
New York State
Power conversion
Renewable energy
Solar forecasting

Deliverable Number: 3002025153

Product Type: Technical Update

Product Title: Solar Power Forecast Component Description for New York State

---

PRIMARY AUDIENCE: Providers of renewable energy forecasting for utility and system operator end users

SECONDARY AUDIENCE: Users of renewable energy forecasting in bulk power system operations

KEY RESEARCH QUESTION

What is the best way to configure a quasi-operational numerical weather prediction system to predict solar power generation at high-spatial resolution over New York State (NYS) at nowcast (0–6 hour) and day-ahead lead times. The gridded solar power forecasts described here will ingest tuned global horizontal irradiance (GHI) forecasts and leverage machine learning models to produce percent capacity and total power output at selected photovoltaic (PV) farms and distributed PV sites across NYS, aggregated to regional totals.

RESEARCH OVERVIEW

To facilitate the provision of quasi-operational solar power forecasts across NYS, the work described here focused on configuring machine learning models to convert GHI forecasts (and other atmospheric variables) into percent power forecasts. This configuration produces gridded forecasts of percent power at 3-km horizontal grid spacing and 15-min output across NYS at both nowcasting (0–6 hour) and day-ahead lead times. Percent power forecasts are also configured for selected farms. The nowcasting simulations are initialized every 15 min between 1100–1945 universal time coordinated (UTC), running "ahead of the clock" so that the entire 6-hour forecast is usable as a true nowcast. The day-ahead simulations are run once daily, starting at 0600 UTC, going out 42 hours through the following day's daylight period. Python scripts run each step of the Solar Power Forecast system automatically from a crontab entry.

KEY FINDINGS

- The final element of the in-development NYSolarCast system is the prediction of power at PV farms and regionally aggregated distributed PV from blended GHI forecasts.
- Key configuration files are included and described.
- Python scripts that run the Solar Power Forecast system for NYSolarCast are described.
- Crontab entry examples are included.

WHY THIS MATTERS

Renewable energy forecasting used in operations requires forecasts of power production, not just predicted solar irradiance. This report describes the solar power forecasting components of the NYSolarCast system, how it is configured, and how to run it quasi-operationally. This Solar Power Forecasting System can be run on a single processor and does not require high-performance computing (HPC) resources, provided it has access to the gridded GHI forecasts produced by the Blended Solar Irradiance Forecast sub-system of NYSolarCast.

HOW TO APPLY RESULTS

The software comprising the Solar Power Forecast System runs on NCAR's Cheyenne HPC system. The HPC resources are not required, however, as the Solar Power Forecast System requires only a single processor (such as a head node or any other standard Linux server), provided it can also read the gridded GHI forecasts produced by earlier processes in NYSolarCast. This report assumes the user has a basic understanding of the Python programming language and familiarity with the Linux operating system. All code and configuration files listed in this report will be available via a private GitHub repository (https://github.com/NCAR/NYSolarCast_delivery).

LEARNING AND ENGAGEMENT OPPORTUNITIES

- This work is being conducted as part of a New York State Energy Research and Development Authority project, in parallel (and with co-funding from) the U.S. Department of Energy Solar Energy Technologies Office (DOE SETO) Systems Integration Program: (https://www.energy.gov/eere/solar/systems-integration-team). Further reports from this project will describe other parts of the forecasting system, culminating in a roadmap for solar forecasting across the state. Such information could also support solar forecasting deployment in other regions by providing useful underlying information and models that could be applied.
- EPRI continues to collaborate with members and other industry stakeholders to improve the development of and integration of renewable forecasting. Similar application of forecasting to other regions is possible and could provide benefit. For those in the New York area, engaging with this effort would allow for use of these forecasts in operations. Please contact the people listed below for more details.

EPRI CONTACTS: Aidan Tuohy, Sr. Program Manager, atuohy@epri.com; Mobolaji Bello, Technical Leader, mbello@epri.com

PROGRAM: Bulk System Integration of Renewables and Distributed Energy Resources, P173

EPRI
3420 Hillview Avenue, Palo Alto, California 94304-1338 • PO Box 10412, Palo Alto, California 94303-0813 USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1
INTRODUCTION

This report describes the development, installation, configuration, and operation of the Solar Power Forecast component of the Solar Forecasting for New York (NYSolarCast) system. This sub-system produces both distributed and utility-scale photovoltaic (PV) power forecasts.

This report is the third in a series of three technical updates describing aspects of NYSolarCast. A flowchart illustrating the overall design of the NYSolarCast system is presented in Figure 1-1. The first report [1] described the configuration and operation of the Weather Research and Forecasting model configured for solar forecasting applications (WRF-Solar®). (WRF-Solar is a registered trademark of the University Corporation for Atmospheric Research.) The second report [2] described the Blended Solar Irradiance Forecast Component Description system and StatCast, which is driven by the Cubist machine learning algorithm. This report describes the Solar Power Forecast component of the system, which is depicted as the three red boxes in Figure 1-1.

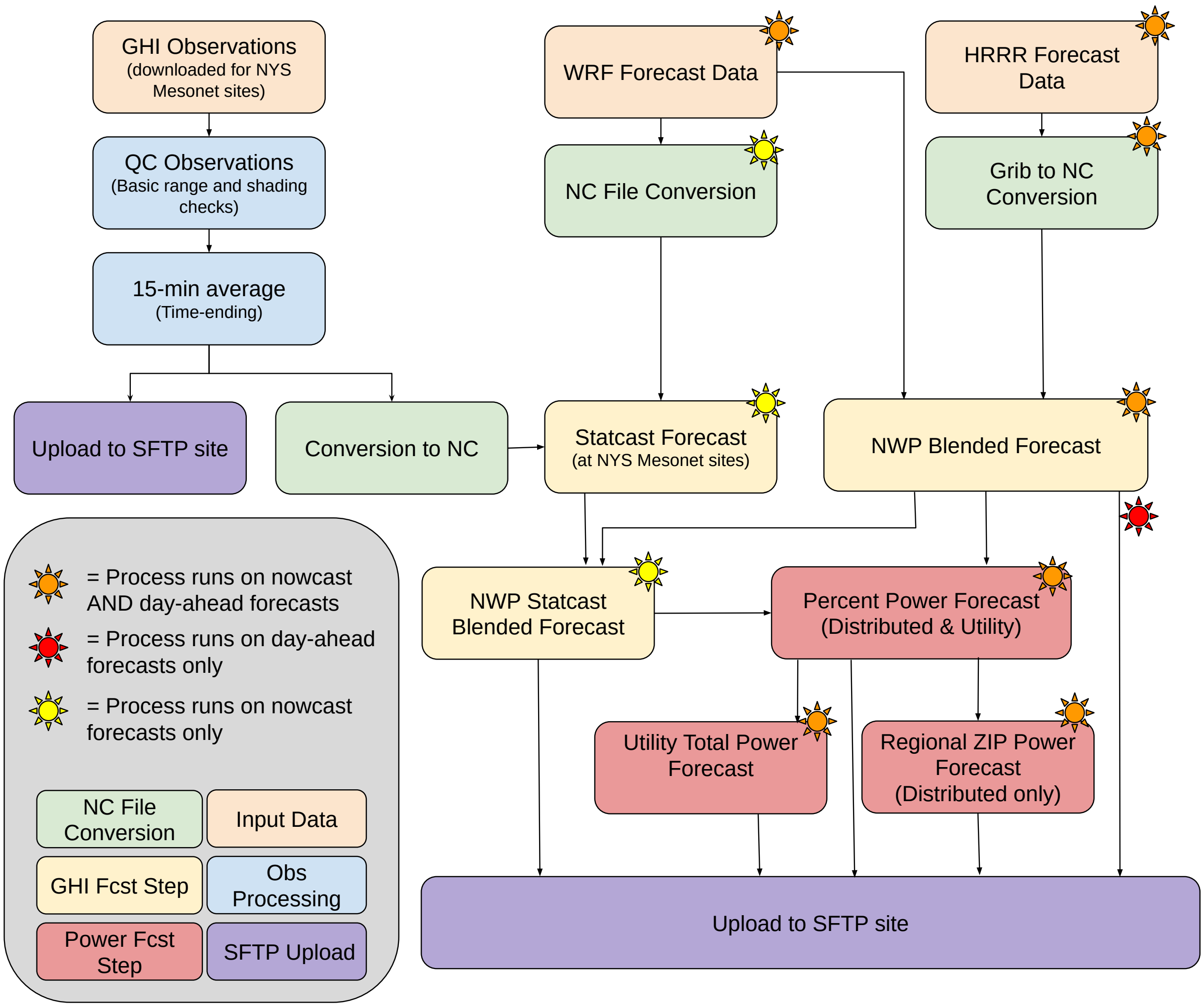

Figure 1-1
Flowchart for the NYSolarCast system. This report focuses on the three red boxes, which comprise the Solar Power Forecast component of NYSolarCast.

This report includes descriptions of the data and quality control (QC) methods applied to them in order to create the machine learning models that predict percent power capacity for both distributed and utility scales. It also includes descriptions of system configuration files, C++ applications, and Python scripts used to run the system either from the command line or in an automated manner using crontab entries on a high-performance computing (HPC) system. The scripts described here are configured for running the Solar Power Forecast on NCAR's Cheyenne HPC but can be easily modified to allow for installation and operation on other HPC systems, or standard Linux servers.

There are two forecast horizons for which the NYSolarCast system makes forecasts, nowcast (0–6 hour) and day-ahead. The WRF-Solar and Blended Solar Irradiance Forecast Technical Reports mentioned above describe the forecast timing in more detail, but a schematic illustrating the schedule of these nowcast and day-ahead processes is illustrated in Figure 1-2. For both the nowcast and day-ahead NYSolarCast forecasts, WRF-Solar output and blended StatCast output are generated on a 3-km gridded domain every 15 min. These gridded forecasts across New York State (NYS) serve as input to the Solar Power Forecast System.

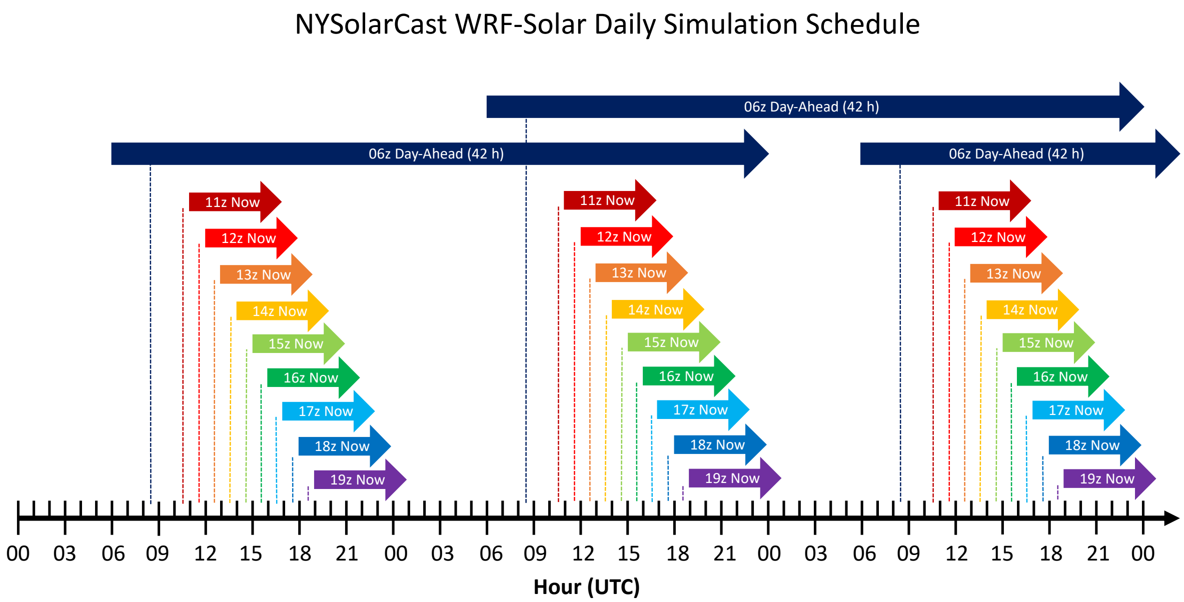

Figure 1-2
Schematic showing the time periods covered by each simulation (thick arrows) in the day-ahead and nowcast WRF-Solar simulations, and their typical quasi-operational availability times (dashed lines) on NCAR's Cheyenne HPC at the time of the writing of this technical report

There are, however, different time scales and geographic extent for the distributed and utility-scale PV power forecasts. These are described as follows:

- **Distributed PV**
  - These forecasts are made at an hourly forecast resolution due to having only hourly data available for model training. Intermediate forecasts of percent capacity (based on nameplate capacity) are created for all domain grid points within NYS. Then, final solar power forecasts are generated on a regional level based on all sites listed in the master PV installation configuration file ("pv_match_grid_within_NYS.csv"). These regional forecasts are currently defined by the regional (first-three-digit) ZIP code and output total expected power for the region. This will be discussed more below.
- **Utility-Scale PV**
  - These forecasts are made at a 15-minute (min) forecast resolution for 10 specific utility-scale PV farm locations. Forecasts for total power are based on a configurable parameter that is the estimated maximum power output of the farm. This parameter is located in a configuration file ("farm_capacity_list.csv"). This will be discussed more below.

A schematic showing the timing of the various power forecasts issued every 15 min in the nowcast time horizon is shown in Figure 1-3. The dashed lines indicate that the forecast is delivered "ahead of the clock," about four minutes before the first forecast (lead time 15 min after the issue time of the nowcast).

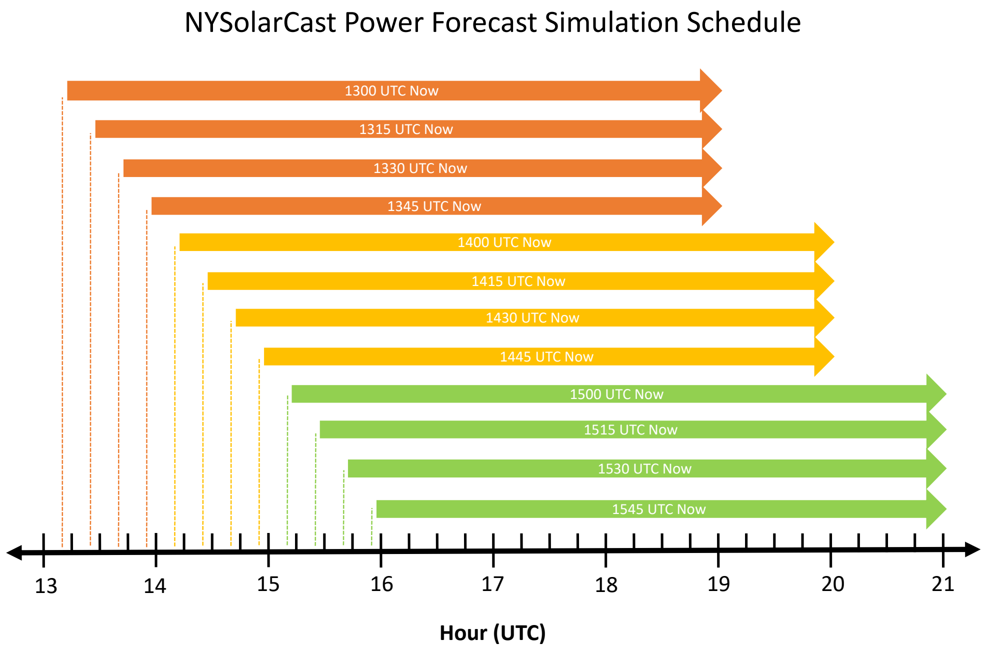

Figure 1-3
Schematic showing the time periods covered by each nowcast simulation from 1300–1545 universal time coordinated (UTC) (thick arrows), and their typical quasi-operational availability times (dashed lines). Similar timing patterns apply for all other nowcast forecasts, with issue times from 1100–1945 UTC daily.

An example set of normalized power forecasts and the corresponding observed normalized power from one of the anonymous PV farms for which we are providing forecasts, Farm B2, for one day in July 2021 is shown in Figure 1-4.

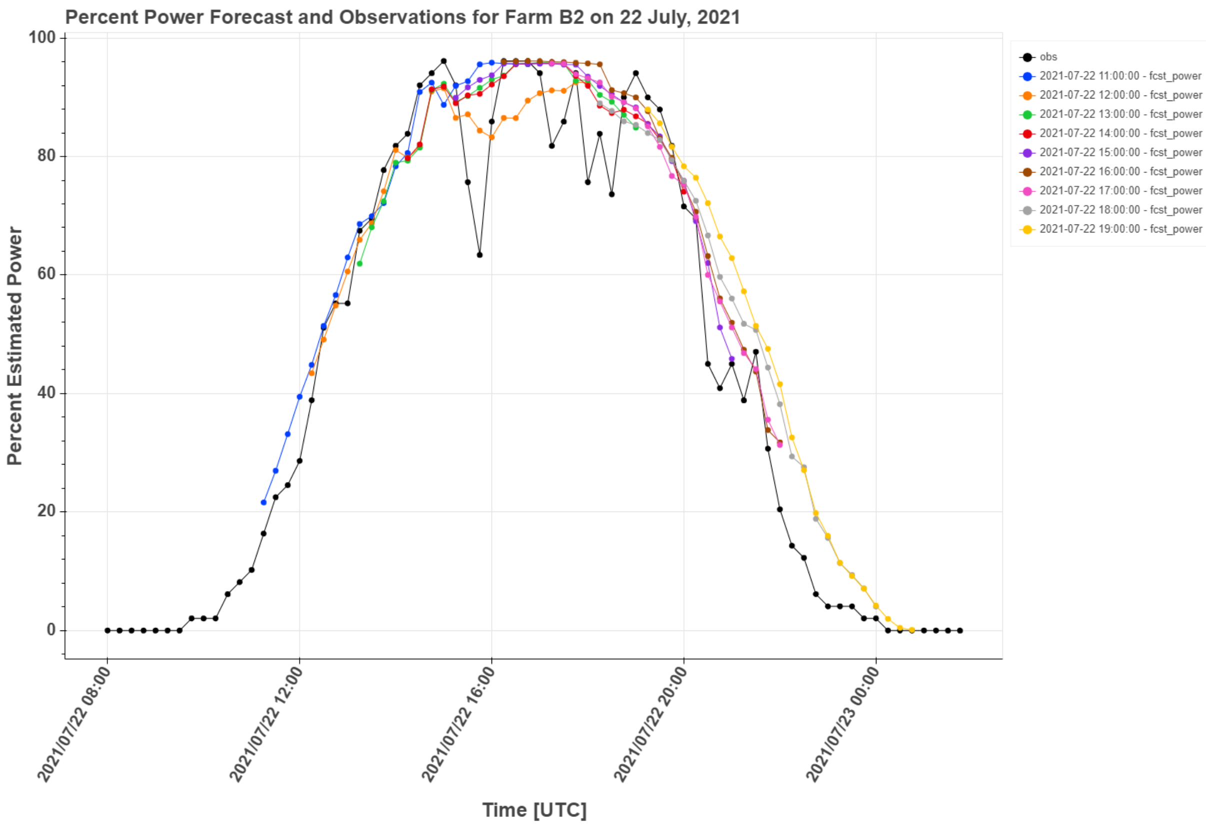

Figure 1-4
Example set of observed normalized power (black) and forecasted normalized power from nowcasts issued at the top of every hour (other colors; see legend), for PV Farm B2 on July 22, 2021

2
DATA AND QUALITY CONTROL

In this section we describe the data and QC procedures applied to them in order to create the percent power capacity models for both the distributed and utility scale solar installations. A description of the data used to create the master PV installation configuration file ("pv_match_grid_within_NYS.csv") with all known distributed solar forecast site locations across the state is also included.

Distributed Solar

A list of all known distributed solar PV site installations across NYS was obtained from the New York State Energy Research and Development Authority (NYSERDA) website (https://data.ny.gov/widgets/3x8r-34rs), downloaded on December 1, 2020. This dataset contains information important to the power forecast component of the NYSolarCast system, specifically site nameplate capacity, ZIP code, and the latitude and longitude of each installation. There are over 100,000 installations included in this NYSERDA dataset, and thus power forecasts for all these sites are made on an aggregated basis. Figure 2-1 shows the locations of all the PV installations from this dataset.

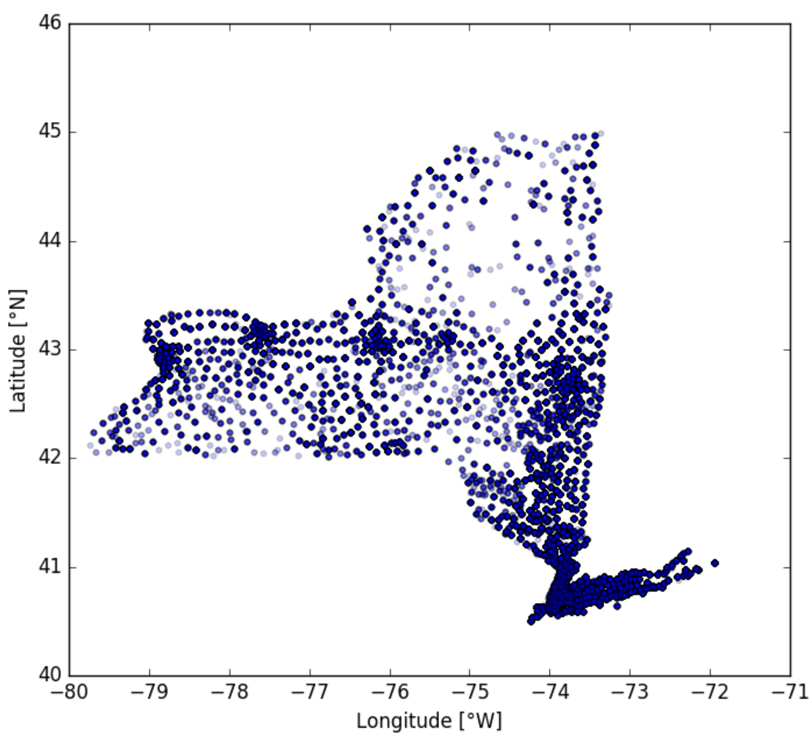

Figure 2-1
Distributed PV installation site locations

Six sites were removed from the NYSERDA master installation configuration file due to issues of incorrect location or ZIP code. Table 2-1 lists the installations that were removed and the reason for removal.

Table 2-1
PV installations removed from the NYSERDA master configuration file

Project Number	ZIP Code	Nameplate Capacity (kW DC)	Location (Latitude, Longitude)	Reason for Removal
190068	22310	6.62	(38.784528°N, -77.124306°E)	Incorrect ZIP and location outside NYS
187120	15043	5.58	(40.554339°N, -80.491646°E)	Incorrect ZIP and location outside NYS
5143	12573	5.01	(42.0047°N, -74.785301°E)	Incorrect ZIP code for Lew Beach
128419	12573	7.80	41.580799°N, -73.927498°E)	Incorrect ZIP code for Hughsonville
72620	12800	7.29	43.308932°N, -73.6434°E)	Incorrect ZIP code for Glen Falls
148415	12800	10.20	(43.299187°N, -73.635289°E)	Incorrect ZIP code for Glen Falls

Distributed power production data were acquired for approximately 490 PV installation sites across NYS from the publicly available NYSERDA website https://der.nyserda.ny.gov/. These data were used as representative sites for all distributed solar across the state. Figure 2-2 shows the location of the representative sites across the state. The locations of the sites are regionally diverse but are more concentrated in urban areas, such as New York City, Albany, Buffalo, and Rochester. They are similarly distributed across the state as all PV installations seen in Figure 2-1. Power data used from this dataset was reported as capacity factor, which represents the percentage of the nameplate capacity output. This percentage, on average, reaches an approximate maximum of 85% of the nameplate capacity, but varies by site. All data requested was specified as quality controlled prior to download. The starting time period of this data varied by site depending on when the site was installed, but the end date for all sites was April 1, 2020.

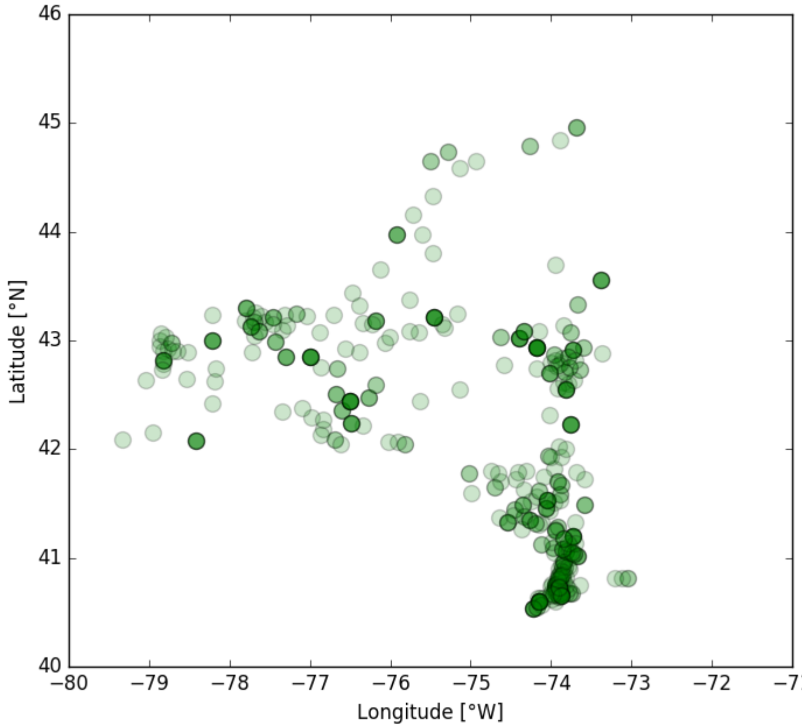

Figure 2-2
Locations of the 490 representative distributed PV sites

In order to obtain meteorological observations matched with the distributed solar power data from NYSERDA, New York State Mesonet (NYSM) data were obtained over the period of January 1, 2018 to April 1, 2020 to correspond with the time period covered by data from the representative sites. Specific meteorological variables used from the NYSM data are GHI [W/m^2], temperature [K], and relative humidity [%]. In order to obtain matched observations, each representative solar power PV site was matched with the closest NYSM site. A histogram of the distance between each representative site and the closest NYSM site is shown in Figure 2-3. Most matched sites are fairly close, within 10–15 km of each other, with the farthest matched sites being just under 30 km apart. The 5-min resolution NYSM data were converted to hourly time-ending averages to match the solar power data resolution. Approximately 17 NYSM sites were observed to have shading or other data quality issues. Junhong Wang from the University of Albany provided a list of times for these sites that were considered to have erroneous data and these data were removed from the historical dataset.

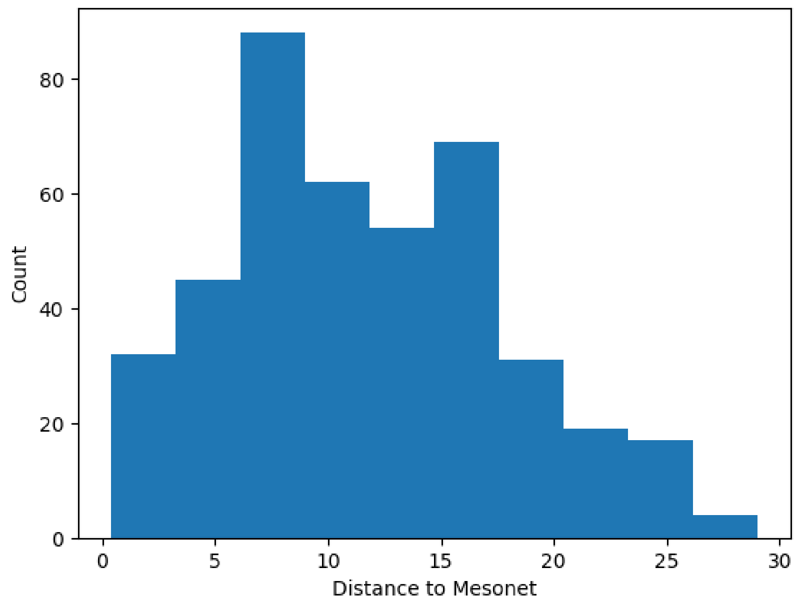

Figure 2-3
Histogram of the distance (km) between representative PV installation sites and the closest NYS Mesonet site

Once PV installations and NYSM sites were matched, GHI (also called solar insolation) and percent power capacity data were obtained, and additional QC methods were applied. First, QC by site was applied. Specifically, as follows:

- All data from certain sites were removed for which:
  - The Pearson r correlation coefficient between the capacity factor and matched GHI for the site was less than 0.75 (43 sites).
  - The 99th percentile of the capacity factor data was greater than 100% or less than 50% of the nameplate capacity (14 additional sites).
- Matched data pairs were also removed for which (as follows):
  - GHI was greater than 1200 W/m^2 and capacity factor was greater than 150%.
  - Both GHI and capacity factor were zero.
  - Capacity factor was zero, but GHI > 240 W/m^2 (20% of 1200 W/m^2).
  - Missing or NaN values for GHI or capacity factor.

Lastly, data were aggregated into climate regions across the state. Climate division polygons for NYS are available in the NCAR netCDF file "climdiv_polygons.nc" (available at https://www.ncl.ucar.edu/Applications/Data). These boundaries are defined by the National Oceanographic and Atmospheric Administration NOAA https://psl.noaa.gov/data/usclimdivs/boundaries.html). Figure 2-4 shows the 10 climate regions for NYS. After data were aggregated by climate region, one final QC procedure was performed. This involved binning matched GHI data by capacity factor data into bins of 5% capacity and then removing data that fell outside the 1st–99th percentiles of the distribution of GHI values within each capacity factor bin.

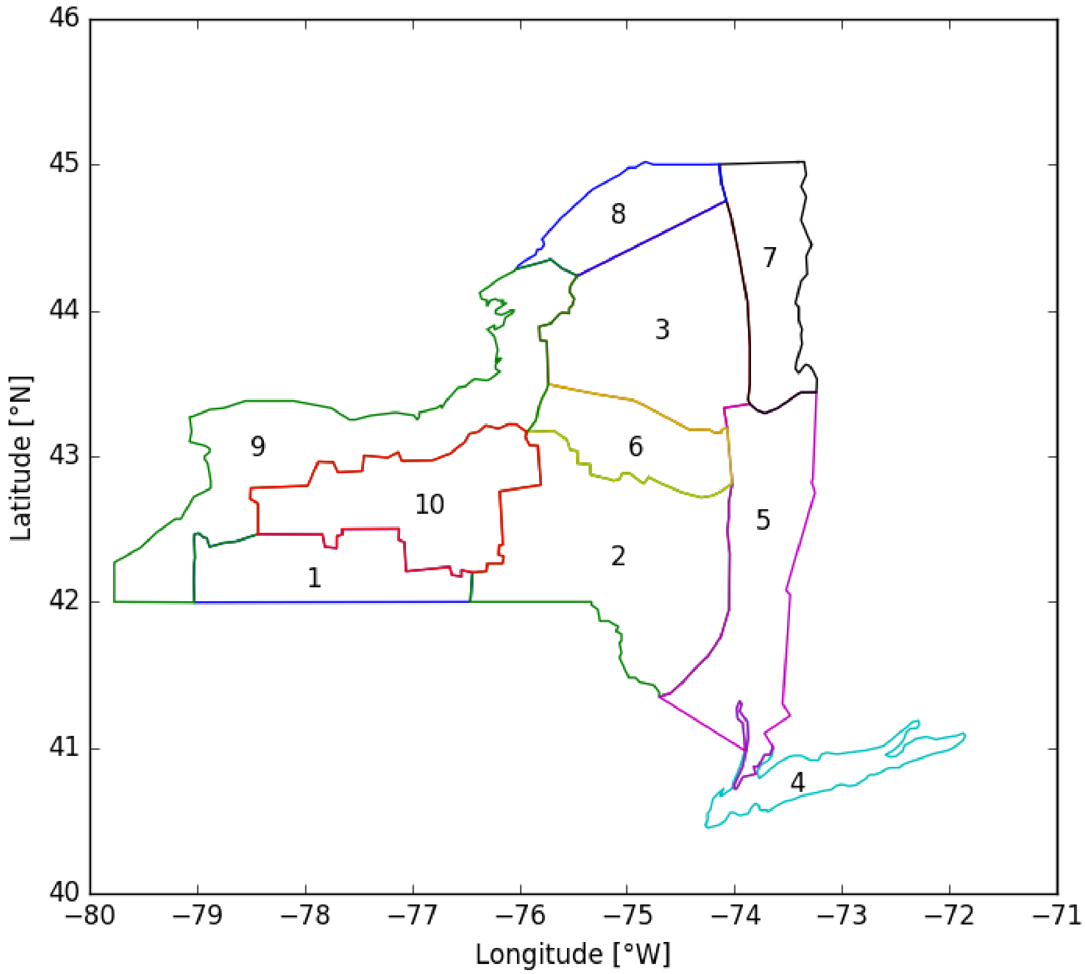

Figure 2-4
NOAA-defined climate regions for New York State

Figure 2-5 shows an example of the QC that was applied to the matched distributed PV power and GHI data that is located within climate region 4. All three panels show a density scatter plot of matched GHI by capacity factor. The left-hand panel shows all of the matched data prior to any QC. The middle panel shows the matched data after preliminary QC was applied. Note that quite a bit of data that fell above the 100% capacity factor was removed. The right-hand panel shows the matched data after the removal of data outside the 1st–99th percentile of GHI values within each 5% capacity factor bin.

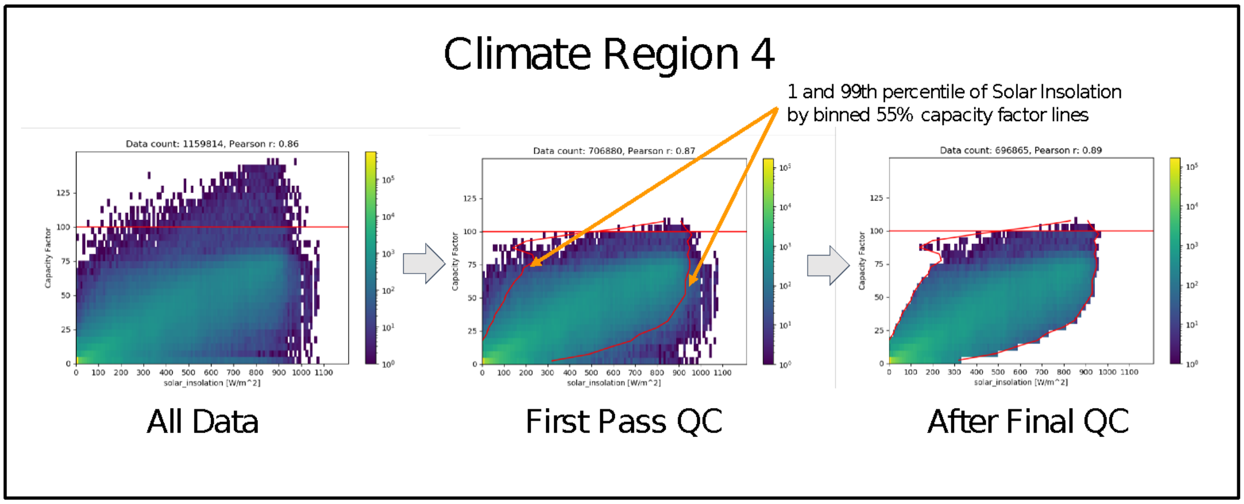

Figure 2-5
Example of data quality control for distributed solar PV installation data located within climate region 4. Left panel shows all data, and right panel shows the final quality-controlled dataset.

Utility-Scale Solar

NCAR received solar power and co-located weather station data for 10 utility-scale solar farms within NYS from two providers (Providers A and B). Provider A has four farms (Farms A1–A4), while Provider B has six farms (Farms B1–B6). Data from Farms A1–A4 contained a mix of 15-min and 5-min resolution power, GHI, and temperature values, while the data from Farms B1–B6 all contained 15-min resolution power, GHI, and temperature values. For consistency with the forecast time horizon, the 5-min resolution data were converted into 15-min time-ending averaged values. Data from the closest NYSM site to each farm were matched with the PV farm to provide observations of relative humidity, and also to test whether GHI from the nearest NYSM site could be used in place of the farm-based measurements. The 5-min resolution NYSM data were converted to 15-min resolution to match the PV farm data temporal resolution. The time periods covered by the historical data varied by farm, with the oldest data starting in January 2018. The latest data used for Farms A1–A4 was April 30, 2021, and for Farms B1–B6 was January 31, 2021. Note that throughout this project, data from each farm was emailed to NCAR in monthly or quarterly deliveries from the providers, so real-time or near-real-time validation was not possible.

QC measures and data rescaling were applied based on individual farm. The general process for Farms B1–B6 was as follows:

- Remove data whose power values are over 120% of the stated nameplate capacity.
- Remove data for which GHI is very low (< 15 W/m^2) but power is greater than 1.5% nameplate capacity.
- Rescale power data to percent capacity by dividing the power output by an estimated maximum output set to be the 99.9th percentile of the power distribution, multiplied by 100.
- Remove data with a rescaled capacity value greater than 101%.

This process was straightforward for four of the six farms from Provider B. However, upon closer inspection of the data from Farms B1 and B4, erroneous data spikes were observed in the power output. For both of those farms, the data spikes occurred over a continuous time interval at the start of the data period, leaving data from a second time interval without spikes and covering the remainder of the historical period. Figure 2-6 shows an example of these data spikes in the power data for Farm B4 during July 24–27, 2019. In order to remove these erroneous data spikes from the datasets, first the estimated maximum output was recalculated for these farms over the second time interval for which the spikes did not occur. The data were then rescaled using the refined maximum output estimate and data with a percent capacity greater than 101% were removed.

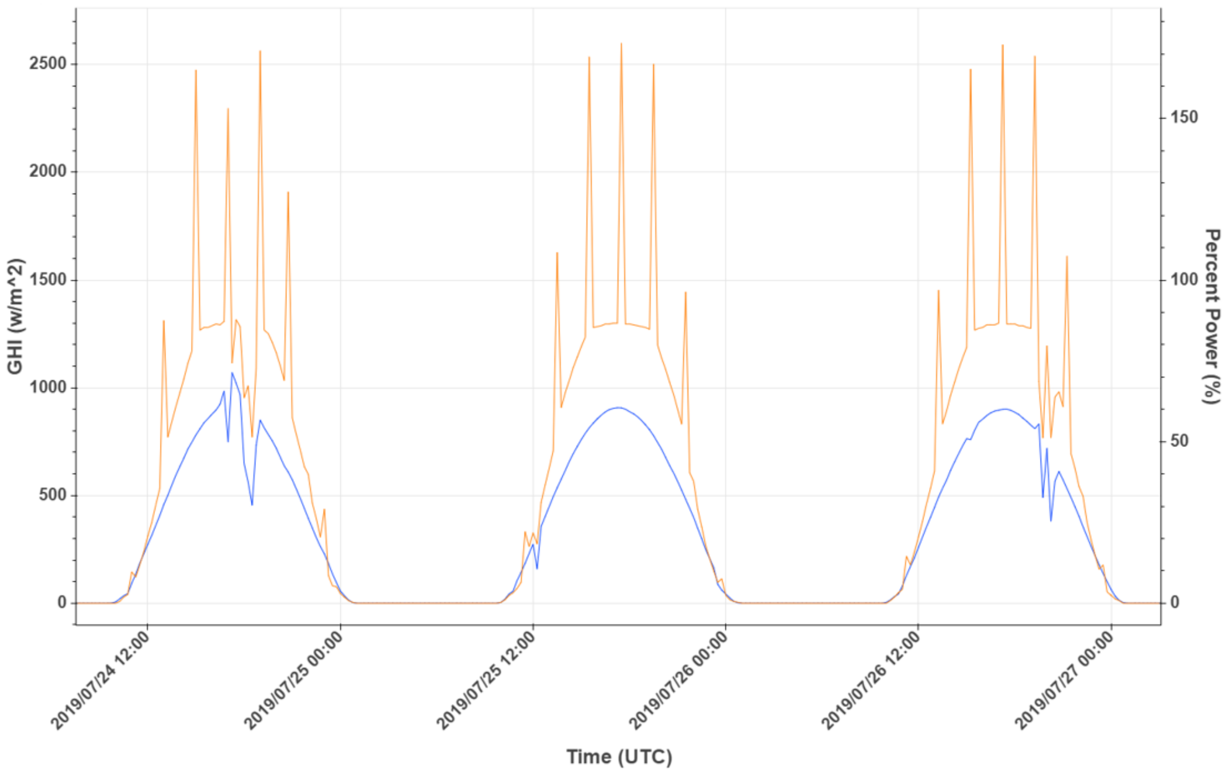

Figure 2-6
Time series of GHI (blue, left axis) and percent power output (orange, right axis) from Farm B4 showing erroneous spikes in power data from July 24–27, 2019

Data from Farms A1–A4 also required QC. Each of these farms appeared to have periods of time during which the maximum power output was curtailed. The data for these time periods were removed. The data were then rescaled similarly to the Farm B1–B6 data, using the estimated maximum power output over the remaining data for each farm set to the 99.9th percentile. Next, data were removed for which GHI values were very low (< 10 W/m^2) but power was greater than 10% capacity. Lastly, each farm's data were further quality controlled so that cases with low-power but high-GHI or high-power and low-GHI were removed. Figure 2-7 shows an example of non-quality-controlled (left panel) and quality-controlled data (right panel) for Farm A3.

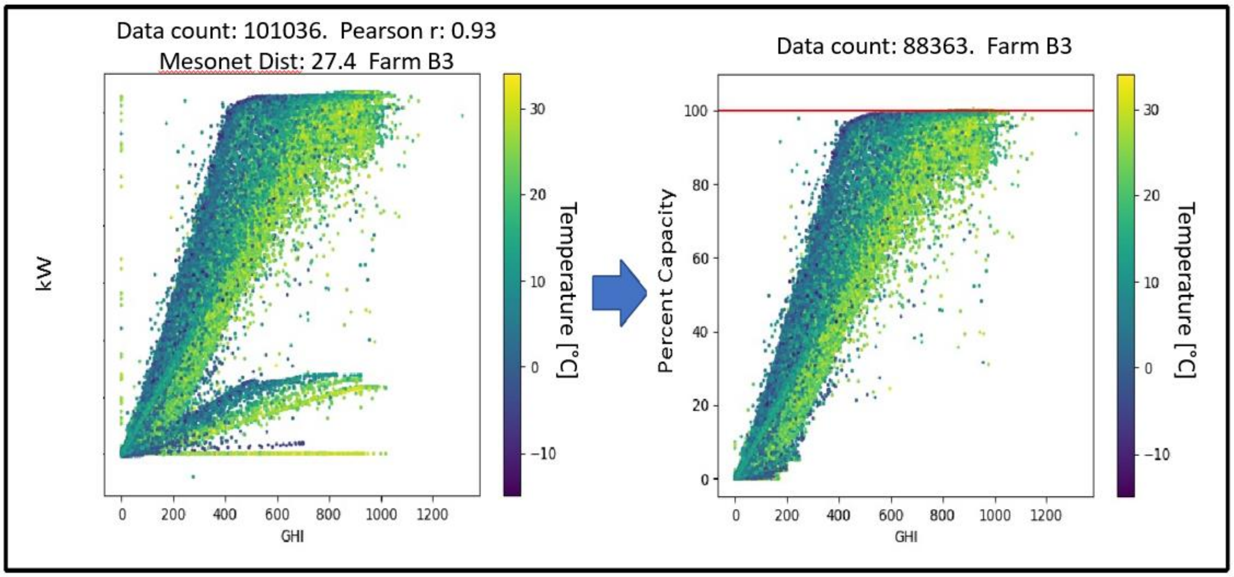

Figure 2-7
Scatter plots of GHI vs. power output for Farm A3, color-coded to observed temperature, showing non-quality-controlled (left panel) and quality-controlled data (right panel)

3
POWER FORECASTING

The power forecasting component of the NYSolarCast system uses forecasts from the blended GHI forecasting task [2], configuration files, and percent power capacity models created using machine learning (ML) to forecast total power for distributed PV solar at configurable regions as well as at 10 individual utility-scale PV solar farms. In this section we present the creation of the ML models that predict an intermediary output of percent capacity, and then describe the procedure for how total power forecasts are produced.

Percent Power Capacity Model Creation

The ML models were created to predict percent capacity output for distributed PV solar and each of the 10 individual utility-scale PV solar farms using the quality-controlled data described in Section 2 as training data. These percent power capacity models are the basis for the Solar Power Forecasting component of the NYSolarCast system. NCAR investigated several different techniques for generating power forecasts including random forest, k-nearest neighbors, and regression trees (for example, [3–6]). In this and previous work, the Cubist regression trees algorithm was found to provide the best performance and thus was chosen as the ML technique for the NYSolarCast system.

For the distributed PV solar percent power capacity model, the input variables include GHI (W/m^2), surface temperature (K), relative humidity (%), all from the nearest NYSM site to the representative PV installation, along with the climate region (1–10) where the site is located. The predictand is capacity factor from the representative PV site, which represents the expected percentage of the nameplate capacity. Recall from Section 2 that this percentage, on average, achieves a maximum of approximately 85% of the nameplate capacity. The model used in the NYSolarCast system was created using quality-controlled data as described in Section 2; however, for initial variable selection and evaluation, a model was created using a two-thirds/one-third split of the data into training and testing datasets, by removing every third day from the training dataset for use in the testing dataset. Performance of the percent power capacity model by hour of day (UTC) and by climate region for the test dataset is shown in Figure 3-1.

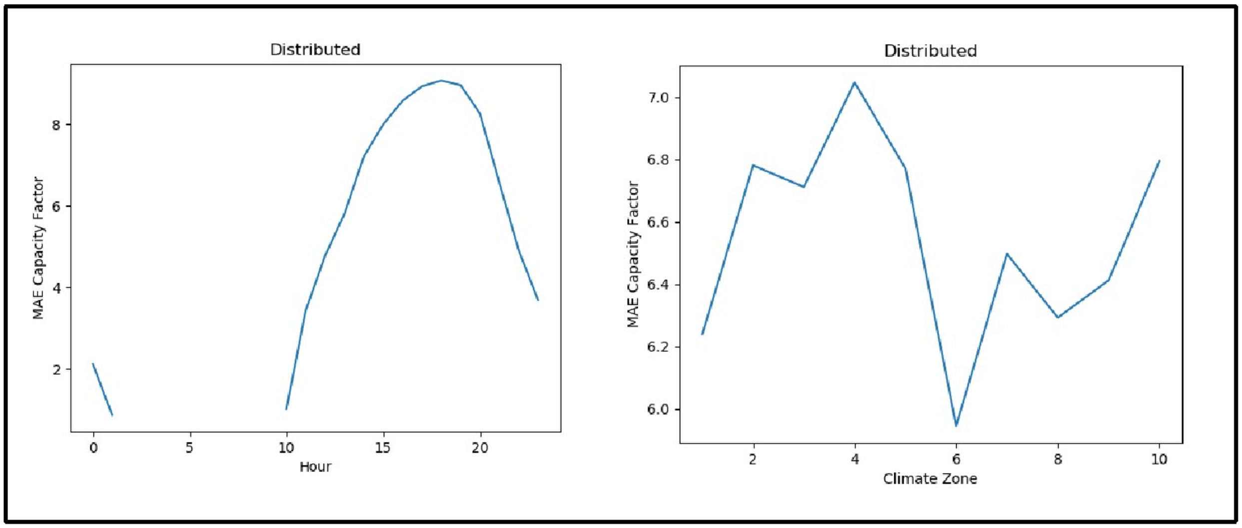

Figure 3-1
Power conversion model performance for the distributed PV solar Cubist model, as evaluated on the test dataset showing percent capacity mean absolute error (MAE) by hour of the day (UTC) (left panel) and by climate region (right panel)

For the utility-scale farms, the Cubist percent power capacity models were created using the quality-controlled dataset described in Section 2. The potential inputs to the models included GHI (W/m^2) observations on-site, surface temperature (K) observations on-site, and relative humidity (%) observations from the closest NYSM site. The predictand is percent capacity, which, recalling from Section 2, is a scaled value based on the estimated maximum quality-controlled output at the farm. Similar to the model created for the distributed solar application, models for each farm were created and evaluated using a two-thirds/one-third split of the quality-controlled datasets. Analysis showed that for all farms, except for Farm A2, relative humidity did not boost performance. For Farm A2 relative humidity did improve performance, but temperature did not appear to improve performance. For the final power models used in the NYSolarCast system, all farm models except for Farm A2 use GHI (W/m^2) and surface temperature (K), while the model for Farm A2 uses GHI (W/m^2) and relative humidity (%).

Power Forecast Procedure

In this section, the procedures for forecasting distributed solar power and utility-scale solar power are described.

Distributed Solar Power Forecast

The distributed solar power forecast predicts the total power output for all sites specified in the configuration file (“pv_match_grid_within_NYS.csv”) by 3-digit ZIP code region. Figure 3-2 shows a map of the 3-digit ZIP code regions for NYS. For each distributed PV solar installation, this configuration file must contain the location of the PV installation site (latitude and longitude), the nameplate capacity of the site, and the ZIP code associated with the site for its regional designation. Percent power forecasts are produced for every grid point within NYS. Power forecasts at the PV installations are then created by using the percent power prediction at the closest grid point, and then multiplying that value by the site’s nameplate capacity to get a total power forecast value. All sites' forecasted total power values are then aggregated by a 3-digit ZIP code region to produce the total power forecast. Certain regional ZIP codes have been combined in order to make the regional ZIP codes more contiguous and sensible. The combined regional ZIP codes are as follows:

- 101 and 102
- 105, 106, 107, and 108
- 117 and 118
- 120 and 121
- 130 and 131
- 133 and 134
- 137 and 138
- 140, 141, and 143
- 144 and 145

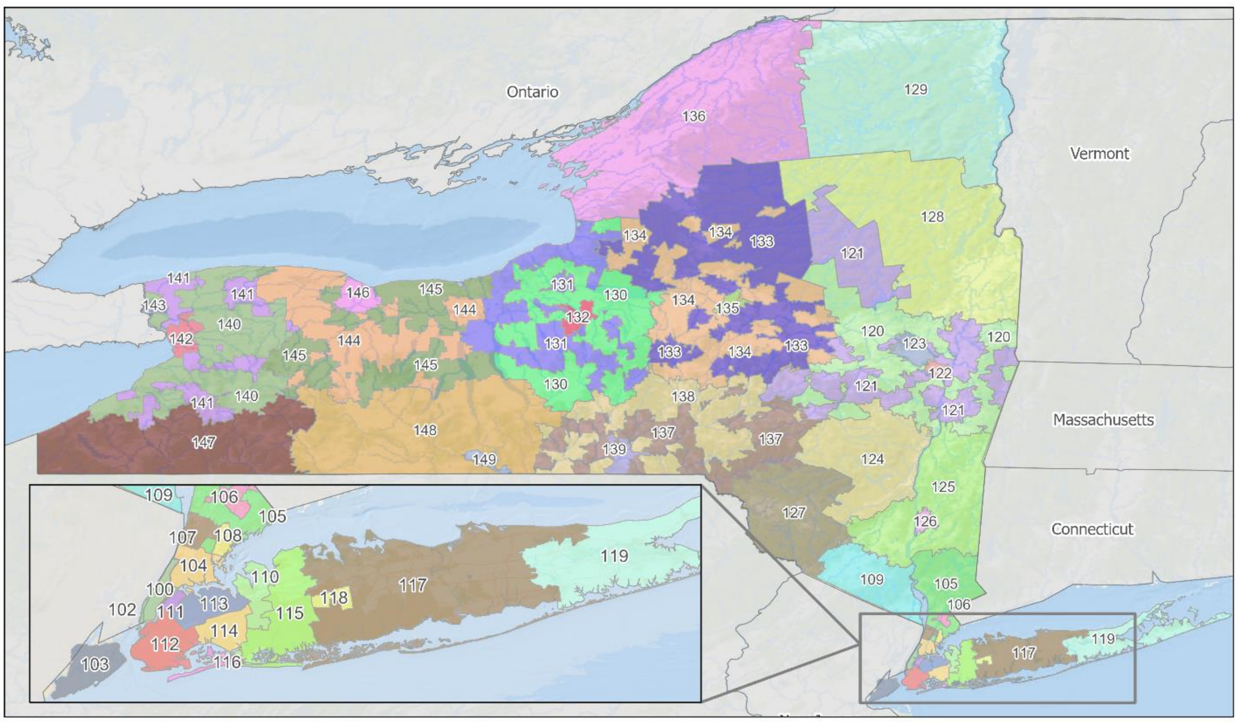

Figure 3-2
Map showing 3-digit ZIP code regions for New York State
Courtesy of Jennifer Boehnert, NCAR

Utility-Scale Power Forecast

For each utility-scale PV farm, a percent power capacity model is applied using data from the grid point located closest to the farm. The forecasted percent capacity is multiplied by the estimated maximum power output value to get that farm's total power forecast. The estimated maximum power output for each farm is designated in a configuration file ("farm_capacity_list.csv"). If the capacity or estimated maximum power output at the farm changes, the power forecast can easily be updated by modifying the maximum power value in this configuration file. Utility-scale farm forecasts differ from the distributed solar power forecast in that they are issued every 15 min with 15-min lead time resolution, rather than being issued hourly at 1-hour lead time resolution.

4

POWER FORECAST CONFIGURATION FILES

Site lists, cdl files and models exist for each “model” in {“distributed,” “Farm_A1,” “Farm_A2,” “Farm_A3,” “Farm_A4,” “Farm_B1,” “Farm_B2,” “Farm_B3,” “Farm_B4,” “Farm_B5,” “Farm_B6”}, where each anonymized farm name in this report corresponds to the actual name of the farm.

Site lists

- <model>_site_list.csv
  - This file contains the grid point(s) over which to produce the percent power forecast for each individual model.
- pv_match_grid_within_NYS.csv
  - This file maps the solar installations to the nearest WRF-Solar grid point. It specifies location, capacity, and ZIP code of the installations, as well as the name of the nearest grid point and distance between the grid point and PV installation.
- farm_capacity_list.csv
  - This file maps Farms A1–A4 and Farms B1–B6 to the nearest WRF-Solar grid point. It specifies location and capacity (installed and estimated as described above) of the farm, as well as the name of the nearest WRF grid point and distance between the grid point and farm.

CDL

- <model>.cdl
  - This file defines the format of the power forecast netCDF file for each model.
- <model>_day_ahead.cdl
  - This file defines the format of the day-ahead power forecast netCDF file for each model.

Models

- <model>.model
  - This is the actual Cubist model file.
- <model>.names
  - This file specifies the variables used in each Cubist model.

5

SYSTEM SOFTWARE AND PYTHON SCRIPTS

All the software listed here can be downloaded directly from the private GitHub repository (https://github.com/NCAR/NYSolarCast_delivery).

NYSC_sys_path.py

- This script outlines the directory and file locations referenced by the rest of the system.

run_power_fcst.py

- This script calls "power_fcst.py" for the nowcast system for each model, specifying the date, model name, number of forecasts, and forecast resolution.

run_power_fcst_day_ahead.py

- This script calls "power_fcst.py" for the day-ahead system for each model, specifying the date, model name, number of forecasts, and forecast resolution.

power_fcst.py

- This script gathers input files and creates and executes the command for running the percent power forecasts (C++ application called "pct_power_fcst") for the given date, model, and forecast cycle (nowcast or day-ahead).

pct_power_fcst

- This is the C++ application that ingests the data and runs the solar power forecast models to produce percent power forecasts at the grid points within NYS.

run_pct_power_rollup.py

- This script calls "pct_power_rollup.py," specifying the forecast date, input directory, site list, and output file. This script runs the code to produce the regional ZIP code total power forecasts.

pct_power_rollup.py

- This script reads the percent power forecasts at the distributed PV sites and combines them into regional ZIP code total power forecasts. The script outputs the total installed capacity, forecasted total power, and forecasted percent capacity at each regional ZIP code.

run_pct_power_tot_total_power.py

- This script calls "pct_power_to_total_power.py" for each PV farm, specifying the input file, output file, and estimated farm capacity (to calculate total forecasted power).

pct_power_to_total_power.py

- This script ingests the percent power forecasts and multiplies it by the estimated farm capacity to create a total power forecast.

upload_total_power_fcsts_to_sftp.py

- This script is responsible for uploading the nowcast total power forecast files to the SFTP site. The script is designed to look over files from the past hour and upload anything that does not exist in the SFTP site.

upload_day_ahead_total_power_fcsts_to_sftp.py

- This script is responsible for uploading the day-ahead total power forecast files to the SFTP site. The script is designed to look over files from the past day and upload anything that does not exist in the SFTP site.

upload_power_fcsts_to_sftp.py

- This script is responsible for uploading the nowcast percent power forecast files to the SFTP site. The script is designed to look over files from the past hour upload anything that does not already exist in the SFTP site.

upload_day_ahead_power_fcsts_to_sftp.py

- This script is responsible for uploading the day-ahead percent power forecast files to the SFTP site. The script is designed to look over files from the past day upload anything that does not already exist in the SFTP site.

6
CRONTAB EXAMPLE

The file crontab.NYSC contains an example set of crontab statements to run the Solar Power Forecast components.

Note that on Cheyenne, the system time is in Mountain Local Time, meaning that the crontab scheduling must be manually updated at the switches between Mountain Standard Time (UTC-7) and Mountain Daylight Time (UTC-6). The times that are provided in the examples below are for the MDT crontab on Cheyenne.

 # Run the nowcast power conversion at distributed PV sites, perform the regional ZIP power forecast and upload the files to the SFTP site, as follows:

10 5,6,7,8,9,10,11,12,13 * * * run_power_fcst.py distributed 60 >& /dev/null ; run_pct_power_rollup.py >& /dev/null ; upload_total_power_fcsts_to_sftp.py >& /dev/null

 # Run the nowcast power conversion at <model>, as follows:

10,25,40,55 5,6,7,8,9,10,11,12,13 * * * run_power_fcst.py <model> 15 >& /dev/null

 # Run the day-ahead power conversion at distributed PV sites, perform the regional ZIP power forecast, and upload the files to the SFTP site, as follows:

50 3 * * * run_power_fcst_day_ahead.py distributed 60 >& /dev/null ; run_pct_power_rollup.py >& /dev/null ; upload_day_ahead_total_power_fcsts_to_sftp.py >& /dev/null

 # Run the day-ahead power conversion at <model>, as follows:

50 3 * * * run_power_fcst_day_ahead.py <model> 15 >& /dev/null

 # Run the nowcast percent power to total power conversion and upload it to SFTP, as follows:

11,26,41,56 5,6,7,8,9,10,11,12,13 * * * run_chain.py run_pct_power_to_total_power.py upload_power_fcsts_to_sftp.py >& /dev/null

 # Run the day-ahead percent power to total power conversion and upload it to SFTP, as follows:

51 3 * * * run_chain.py run_pct_power_to_total_power.py upload_day_ahead_power_fcsts_to_sftp.py >& /dev/null

7
REFERENCES

1. *WRF-Solar® System Description for New York State*. EPRI, Palo Alto, CA: 2022. 3002025151.
2. *Blended Solar Irradiance Forecast Component Description for New York State*. EPRI, Palo Alto, CA: 2022. 3002025152.
3. S. Alessandrini, L. Delle Monache, S. Sperati, and J. N. Nissen, "A Novel Application of an Analog Ensemble for Short-term Wind Power Forecasting," *Renewable Energy*, vol. 76, pp. 768–781, Apr. 2015, doi: 10.1016/j.renene.2014.11.061.
4. T. C. McCandless, S. E. Haupt, and G. S. Young, "A Model Tree Approach to Forecasting Solar Irradiance Variability," *Solar Energy*, vol. 120, pp. 514–524, Oct. 2015, doi: 10.1016/j.solener.2015.07.020.
5. S. E. Haupt *et al*., "Building the Sun4Cast System: Improvements in Solar Power Forecasting," *Bull. Amer. Meteor. Soc*., vol. 99, no. 1, pp. 121–136, Jun. 2017, doi: 10.1175/BAMS-D-16-0221.1.
6. S. E. Haupt *et al*., "Combining Artificial Intelligence with Physics-based Methods for Probabilistic Renewable Energy Forecasting," *Energies*, vol. 13, no. 8, Art. no. 8, Jan. 2020, doi: 10.3390/en13081979.

About EPRI

Founded in 1972, EPRI is the world's preeminent independent, non-profit energy research and development organization, with offices around the world. EPRI's trusted experts collaborate with more than 450 companies in 45 countries, driving innovation to ensure the public has clean, safe, reliable, affordable, and equitable access to electricity across the globe. Together, we are shaping the future of energy.

3002025153

EPRI
3420 Hillview Avenue, Palo Alto, California 94304-1338 • PO Box 10412, Palo Alto, California 94303-0813 • USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com