1 October 2018

Methodology for Wind/Solar Electricity Forecaster

by

Christopher Chu and Eric Larson
Climate Central, Inc., Princeton, NJ

with contributions from
John Zack and Tim Melino
MESO, Inc., Albany, NY

Contents

Overview

Climate Central's wind/solar electricity forecaster provides hourly estimates of the amount of electricity generated by solar-PV and wind-turbine installations yesterday, today, and tomorrow across each Nielson Designated Market Area (DMA) in the U.S. The estimates are generated using spatially-resolved hourly actual-observed (for yesterday) or forecast (for today and tomorrow) wind speeds and solar irradiance, together with spatially-resolved estimates of installed wind turbine and solar PV capacities. The weather data and installed generating capacities are estimated for each cell of a uniform 0.05° latitude by 0.05° longitude grid across the continental U.S. and Alaska and Hawaii. The forecasting tool makes electricity generation estimates at this cell resolution and then aggregates the individual cell estimates into totals for larger regions of interest. In the current release of the tool (v1), aggregation is by DMA and (for wind) by the EPA's eGRID regions. Results for other aggregations of cells (e.g., county, state, congressional district, etc.) may be added in a future version of the tool. To put the electricity generation estimates in perspective, several metrics are calculated and reported in the v1 tool, including the equivalent number of households in a DMA that could be powered each day by the estimated solar or wind generated electricity. Additionally, a "Wind Power Index" (WPI) and "Solar Power Index" (SPI) are calculated. The value of these indices reflect how "good" the day was (yesterday) or is expected to be (today and tomorrow) in terms of solar or wind electricity generation. This document details the methodologies, assumptions, and input data sources used to construct the forecaster.

Weather Parameters at 0.05° x 0.05° Grid Cell Resolution

Hourly wind speed and solar irradiance estimates on a 0.05° lat x 0.05° lon grid are provided by MESO.[1] Specifically, MESO provides observed (historical) and forecast wind speed at 80-meter elevation (m/s) and global horizontal irradiance (GHI) at the earth's surface plus its direct and diffuse components and a clear-sky GHI (all in W/m^2). Historical observed values are estimated for the continental U.S. from the hourly analysis (i.e. the 0-hour forecast) dataset from the 3-km grid of the High-Resolution Rapid Refresh (HRRR) weather forecasting model[2] and from the 13-km grid of the Rapid Refresh (RAP) model[3] for Alaska and Hawaii. Forecast values are from the hourly 0.25° degree pressure level dataset from the Global Forecast System (GFS) weather forecasting model.[4] The values from the HRRR, RAP and GFS datasets are interpolated to the 0.05° grid by the bilinear interpolation scheme contained in NCEP's "degrib" software package.[5] The data from MESO are downloaded and processed at Climate Central once each day starting at 6 a.m. local time (Eastern U.S. standard or daylight savings time). Forecast values included with each download cover 120 hours, beginning from 00:00 UTC (or

GMT) of the download day.* In the v1 release of the tool, only results for yesterday, today, and tomorrow are reported to the tool user.† An additional 3 days of forecasts are generated by the tool but are not reported to the user in the v1 release of the tool.

Estimating Solar and Wind Electricity Generation at the Grid-Cell Level

The weather parameter values at the 0.05° grid cell level are used together with estimates of the installed solar-PV and wind-turbine generating capacity (as of late 2017) in each grid cell to estimate the solar and wind electricity generation in each grid cell across each hour for which estimated or forecasted values are provided. Hourly estimates are summed to get daily values, and the daily values for different aggregations of grid cells, e.g., DMAs, are then reported as outputs.

Installed Solar-PV Generating Capacity by Grid Cell

Estimates of installed solar-PV generating capacity at the grid cell level are based on three data sources.‡

First, all electricity generators in the U.S. with installed capacity of 1 MW or more, including solar-PV generators, submit information about their generating facilities to the Energy Information Administration, which reports these collectively on form EIA860.[6] For each solar installation included on form EIA860, we extract the following data: the EIA-assigned plant ID number, installed AC capacity, installed DC capacity, type of tracking (fixed, east-west fixed,[§,7] single-axis, or dual axis), tilt angle, and plant location (latitude, longitude, and zip code

* GFS forecasts are released by NOAA four times each day (at approximately 02:20, 08:20, 14:20, and 20:20 UTC). The data downloaded by Climate Central each morning at 06:00 (ET) incorporates the 08:20 (UTC) release. The first hour for which a forecast is included in the 08:20 release is 00:00 (UTC) of that day.

† Yesterday = 24 hours beginning at 04:00 UTC of the day prior to the download day. Today = 24 hours starting from 12:00 UTC of the download day. Tomorrow = 24 hours starting from 12:00 of the day after the download day.

‡ The forecasting tool does not include estimates for concentrating solar power generation (CSP). There are 16 currently operational CSP plants in the US with a total installed capacity of 1771 MW (https://www.nrel.gov/csp/solarpaces/index.cfm). In 2017 solar-PV and solar-CSP generation totals across the U.S. were 73,828 GWh, and 3,269 GWh, respectively (Energy Information Administration).

§ East-west oriented arrays will generate less total electricity annually than south-facing arrays, but may produce more electricity per day at certain times of the year.[7] Given the small number of east-west oriented installations in the EIA860 database (5 of 2,794 installations report having an east-west orientation), we assume for purposes of the forecasting tool that all fixed and single-axis tracking installations have south-facing orientations.

tabulation area).[**] Using the latitude and longitude coordinates we associated each solar installation with a unique 0.05° grid cell by simply rounding the lat/lon coordinates for the installation to the nearest 0.05° values.

Second, for solar-PV installations with less than 1 MW of installed capacity, we use data reported by the Open PV project.[8] Characteristics of the majority (~80%) of the more than 1 million installations included in the Open PV dataset are provided by the Lawrence Berkeley National Laboratory (LBL), which annually produces the *Tracking the Sun* report.[9] The remaining data are provided to the Open PV dataset on a voluntary basis directly by PV installers, businesses, and residential owners. From the Open PV dataset we extract for each reported solar installation its DC capacity and the zip code in which it is located. (Latitude and longitude are not reported.)[††]

We assigned installations from the Open PV data set to specific 0.05° grid cells using the following methodology. Open PV installations report their location by zip code, but zip codes are, strictly speaking, associated with postal delivery routes rather than geography and so cannot be used directly to determine precise physical locations. However, the U.S. Census groups zip codes into geographic regions called zip code tabulation areas (ZCTA, Figure 1).[10] We used an online tool, UDS Mapper,[11] to determine which ZCTA each zip code is associated with. Thus, by assigning zip codes to ZCTAs and then summing the capacities of all Open PV solar installations in a given ZCTA, we are able to estimate the total Open PV solar capacity installed in each ZCTA.

ZCTAs vary in spatial extent, but can span multiple 0.05° grid cells. As well, some grid cells will overlap more than one ZCTA. Thus, in a further step we distribute the solar capacity of a given ZCTA across grid cells inside or overlapping a ZCTA. For a given grid cell *X*,

** For the v1 release of our forecasting tool, we used EIA860 2017 Early Release data. The revised/final 2017 data were released in August 2018. We found some errors in the 2017 early release data, e.g., some values for zip code tabulation areas that do not exist in practice. We corrected these manually, but the errors we identified in the early release data raise the question of whether there are other errors, e.g., in the installed generator characteristics. In a future release of the forecasting tool, revised/final EIA860 data should be incorporated, if possible.

†† Some of the Open PV installations are larger than 1 MW and so there is the possibility of double-counting if the specifics of an installation are included in both the EIA860 and Open PV datasets. To minimize double counting, we searched the two datasets manually for installations located in the same zip code and having identical installed DC capacities. We discarded data for installations in the Open PV dataset that duplicated ones found in the EIA860 dataset. The total capacity of discarded entries was 921 MW_{DC}.

$$Solar\ capacity\ in\ grid\ cell\ X = \sum_{i=1}^{i=N} \left\{ \frac{Population\ of\ grid\ cell\ X}{Sum\ of\ populations\ of\ all\ grid\ cells\ intersecting\ ZCTA_i} \cdot Solar\ Capcity\ in\ ZCTA_i \right\} \qquad \text{Eqn. 1}$$

where the summation is overall all ZCTAs intersected by grid cell X.

We estimated the population in each grid cell for use in the above equation using a similar approach as for estimating solar capacity in a grid cell:

$$Population\ of\ grid\ cell\ X = \sum_{i=1}^{i=N} \frac{spatial\ area\ of\ grid\ cell\ X}{spatial\ area\ of\ all\ grid\ cells\ in\ Block\ Group_i} \cdot population\ Block\ Group_i \qquad \text{Eqn. 2}$$

where the summation is overall all block groups intersected by grid cell X. Block Groups are U.S. Census Block Groups (Figure 2),[12,‡‡] the smallest spatial unit for which population data are made available by the U.S. Census. The population of each Block Group was obtained through the socialexplorer database.[13] The spatial area of each grid cell was determined from its latitude and longitude. The length corresponding to 0.05° of latitude is essentially the same everywhere on earth: 3.46 miles. The length corresponding to 0.05° longitude varies with latitude, so each grid cell is trapezoidally shaped, with the longer side being the southern end and the shorter side being the northern end. The length corresponding to 0.05° longitude is calculated as

$$0.05^{\circ}\ longitude = \cos(latitude) \cdot 3.46\ miles \qquad \text{Eqn. 3}$$

And the area of the grid cell is the product of its latitude length and the average of its northern and southern longitude lengths:

$$Area\ of\ 0.05^{\circ}\ grid\ cell\ (in\ sq.miles) = 3.46 \cdot \frac{(N_{long} + S_{long})}{2} \qquad \text{Eqn. 4}$$

Where N_{long} and S_{long} are the lengths of the northern and southern borders of the cell in miles.

A third data source was used to make a final adjustment in the installed solar capacities per 0.05°grid cell based on the Open PV dataset [calculated using Eqn. 1] and based on the EIA860 dataset. The third data source is the Solar Energy Industry Association's (SEIA) state-level

‡‡ Geographic information system maps of block groups are only available by individual state. We compiled a national map by combining the individual state maps using QGIS 3.0.3 software.

installed solar PV capacity estimates.[14] We found that the sum of EIA860 and Open PV installed solar PV capacities that we calculated for a given state were sometimes significantly different from the total installed PV capacity for that state that SEIA reported. (In these cases the SEIA value was usually higher than ours.) We subsequently learned from SEIA that their in-house database relies on EIA860 and Open PV data, but is additionally informed by direct communications with a number of utilities and other entities involved with solar electricity generation.[15] Thus, to arrive at final estimates of installed solar capacity for our forecasting tool, we multiplied the capacity of each installation for which we had data from the EIA860 and the Open PV datasets by a ratio ***A/B***, where ***A*** is the total installed DC capacity as reported by SEIA for a given state and ***B*** is the total installed DC capacity for that state as given by our sum of EIA860 and Open PV data for that state.[§§]

Installed Wind Turbine Capacity by Grid Cell

Determining the installed wind turbine capacity at each 0.05° grid cell was a much simpler task than determining installed solar capacity. Data relating to wind electricity generators in the U.S. are publically available in a dataset maintained by the Lawrence Berkeley National Laboratory, the US Geological Survey, and the American Wind Energy Association.[16] The latest release includes data on 57,646 turbines, including some that became operational as recently as the first half of 2018. The oldest turbines in the data set were installed prior to 1990. Technical specifications of the turbines in the dataset include make, model and other information. The latitude and longitude of each turbine is also provided. For each individual wind turbine in the database, we extract a number of different parameter values,[***] the most salient of which for the v1 forecasting tool are the turbine's latitude, longitude, and installed AC generating capacity. To determine which 0.05° grid cell a given turbine resides in, we simply rounded the lat/lon coordinates for the turbine to the nearest 0.05° values.

§§ SEIA reports state capacities in MW_{DC}, so in calculating the ratio, ***A/B***, we had to first ensure that our EIA860 and Open PV capacities were all expressed in MW_{DC}. Open PV reports capacities in MW_{DC}. The EIA860 reports both MW_{DC} and MW_{AC} for most installations. For those reporting only AC, we calculated the value of DC capacity assuming the ratio between the DC and AC capacity to be the same as the average of the ratio for all installations in the EIA860 database that reported both values. We found the average MW_{DC}/MW_{AC} to be 1.25.

*** For each turbine, the parameter values that we extract and have in our database for possible future use are location related: state, county, fips number, and longitude and latitude coordinates, and turbine related: manufacturer, model, capacity (MW_{AC}), hub height, rotor diameter, rotor swept area, total height.

Electricity Generation Calculations at Grid Cell Level

We use the hourly weather parameter values and the installed solar and wind generating capacity values, both at the 0.05° grid cell level, to calculate the estimated actual (previous day) and forecasted (current and next day) hourly electricity generation for each 0.05° grid cell. We sum the hourly values for each 24-hour period to determine daily electricity generation. Details of the electricity generation calculations are described next.

Solar Electricity Generation

We calculate solar electricity generation separately for installations from the Open PV dataset and from the EIA860 dataset.

For installations reported in the Open PV database, we assume these are all distributed roof-top installations, with compass-direction orientations and tilt angles varying from installation to installation in a given grid cell. For a set of such distributed PV installations, the collective electricity generation can be correlated with the global horizontal irradiance (GHI) and expressed as a fraction of the collective installed capacity (Figure 3[17]). The following equation describes the curve in Figure 3:

$$y = -7.0778 \times 10^{-10} GHI^3 + 7.1347 \times 10^{-7} GHI^2 + 8.7895 \times 10^{-4} GHI + 7.9739 \times 10^{-3} \qquad \text{Eqn. 5}$$

Where *GHI* is in watts/m^2 and *y* (dimensionless) is the fraction of installed AC capacity that would be generating power at the given *GHI*. To calculate hourly electricity generation in a given grid cell, the forecast *GHI* for that cell in that hour is used in Eqn. 5. The resulting fraction is multiplied by the installed kW_{AC} capacity to get the kWh generated in that hour.

For solar installations in the EIA860 dataset, we assumed that these are all utility-scale installations, and that most include tracking or have a fixed tilt angle designed to provide more optimal solar exposure than the more-random angles found with distributed PV systems discussed in the previous paragraph. To calculate the power generated from each installation in the EIA860 dataset, we used the relationship plotted in Figure 4 and described by the following expression:[17]

$$GEN = -7.3045 \times 10^{-10} \times POA^3 + 7.2772 \times 10^{-7} \times POA^2 + 9.7863 \times 10^{-4} \times POA + 0.01503$$

Eqn. 6

Where *GEN* is the fractional AC generation, and *POA* is the Plane of Array irradiance in watts/m^2. POA is calculated from the direct (DHI) and diffuse (DIFFI) components of the GHI. Different models can be used for estimating DHI and DIFFI from GHI,[18] but differences in the results from different models are small compared with other uncertainties in our forecasting tool's results. For the forecasting tool, DHI and DIFFI are calculated by MESO using an internally developed polynomial curve fit to measured direct and diffuse data that employs the clear sky GHI (CS-GHI) and the clear sky index (CSI) as input variables. The CS-GHI is calculated using the "pvlib-python" software module from Sandia National Laboratory.[19] CS-GHI represents the global solar irradiance that would reach a horizontal plane at a specified geographical location on the earth's surface under cloudless conditions and an assumed set of reference clear sky atmospheric conditions (e.g. optical path length due to aerosols and water vapor). The CSI is the ratio of the actual GHI to CS-GHI. The DHI and DIFFI values are transmitted along with the CS-GHI with the daily download of hourly weather parameters for each grid cell from MESO. We then calculate the POA irradiance from the following sequence of relationships:[20]

$$POA = (DNI \cdot corfac) + DIFFI \qquad \text{Eqn. 7}$$

where,

$$DNI = Direct\ Normal\ Irradiance = \frac{DHI}{\cos(solar\ zenith\ angle)}$$

$$solar\ zenith\ angle = 90 - solar\ elevation\ angle$$

$$solar\ elevation\ angle = \sin^{-1}[\sin(dec)\cdot\sin(lat) + \cos(dec)\cdot\cos(lat)\cdot\cos(HRA)]$$

$$dec = solar\ declination\ angle = \text{-}\ 23.45\cdot\cos\left[\frac{360}{365}\cdot(d+10)\right]$$

$$d = day\ number\ \text{(January 1=1)}$$

$$lat = latitude\ of\ PV\ array\ location$$

$$HRA = hour\ angle = 15^o\cdot(LST - 12)$$

$$LST = local\ solar\ time = LT + \frac{TC}{60}$$

$$LT = local\ time$$

$$TC = time\ correction\ factor = 4\cdot(long - LSTM) + EoT$$

$$long = longitude\ of\ PV\ array$$

$$LSTM = local\ standard\ time\ meridian = 15^o\cdot\Delta T_{UTC}$$

$$\Delta T_{UTC} = (LT - UTC),\ \text{in hours}$$

$$EoT = Equation\ of\ time$$
$$= 9.87 \cdot \sin(2B) - 7.53 \cdot \cos(B) - 1.5 \cdot \sin(B)$$
$$B = \frac{360}{365} \cdot (d - 81)$$

and

$$corfac = \sin(dec) \cdot \sin(lat) \cdot \cos(tilt) - \sin(dec) \cdot \cos(lat) \cdot \sin(tilt)$$
$$+ \cos(dec) \cdot \cos(lat) \cdot \cos(tilt) \cdot \cos(HRA) + \cos(dec) \cdot \sin(lat) \cdot \sin(tilt) \cdot \cos(HRA)$$

where *tilt* is the tilt angle of the PV array. For a PV array with dual-axis tracking, *corfac* = 1.0.

Eqn. 6 (Figure 4) was derived from a regression analysis of actual measured data from utility-scale solar generation facilities in California. Figure 5 lends confidence to our approach for utility-scale calculations. In that figure, we compare results from Eqn. 6 with actual performance results (measurements every 15 minutes) from one specific facility in California. For that facility, the POA irradiance was measured using an onsite pyranometer, which gives a point value, while the reported generation corresponds to an area-weighted value. This accounts for some of the scatter - especially the points well below the main band. In those cases the pyranometer was recording cloud effects much greater than the cloud impacts on the larger facility area.

Wind Electricity Generation

To calculate wind electricity generation, we assume each turbine performs according to the power curve shown in Figure 6. This is a composite facility-scale curve developed by analysts at MESO.[17] The curve is not as sharp as a power curve for a specific individual turbine because it accounts for the partially uncorrelated behavior of the generation among a set of turbines (due to wind speed differences from wakes and other factors, as well as performance variations). Also, the maximum generation is less than 100% which represents the fact that, on average, a turbine or two is offline for maintenance or other issues. Furthermore, the high speed shut down is gradual (from 20 to 25 m/s) to reflect variations in wind speed experienced by individual turbines (some will shut down sooner than others). The curve in Figure 6 can be represented analytically as follows, where ***x*** is wind speed (m/s) and ***y*** is wind power output expressed as fraction of installed AC generating capacity:

$0 < x < 2.5$ m/s: $y = 0$

2.501 < x < 13.5 m/s:

$$y = 3.0693 \times 10^{-6}x^6 - 1.138 \times 10^{-4}x^5 + 0.0014x^4 - 0.0077x^3 + 0.0318x^2 - 0.0734x + 0.0596$$

13.501 < x < 20 m/s: $y = 0.9646$

20.001 < x < 25 m/s: $y = -0.1929x - 4.8232$

x > 25.001 m/s: $y = 0$

Eqn. 8

To calculate hourly wind electricity generation in a given grid cell, the forecast wind speed for that cell in that hour is used in Eqn. 8. The resulting fraction is multiplied by the installed AC MW capacity in the cell to give MWh generated in that hour.

Spatial Aggregation of Grid Cell Level Results

Calculated solar and wind electricity generation at the 0.05° grid cell level were grouped into different aggregations of grid cells corresponding to larger regions of interest. In v1 of the forecasting tool, results are reported for two aggregations of grid cells: Nielsen Designated Market Areas (DMA, Figure 7)[21] and EPA Emissions and Generation Resource Integrated Database (eGRID) regions (Figure 8).[22,23] eGRID is a database maintained by the Environmental Protection Agency of environmental characteristics of almost all electric power generators in the U.S. Among other applications, the eGRID database is widely used for estimating greenhouse gas emissions associated with the consumption of electricity, i.e., emissions that occurred when the consumed electricity was originally generated. In practice, the source of a consumed electron on the grid is difficult to determine precisely. The eGRID database assigns individual power plants to an eGRID region within which the electricity generated is judged likely to also be consumed in that region.[22]

To aggregate the 0.05° grid cells across the continental U.S. and Alaska and Hawaii into DMA and eGRID, as well as time-zones,[24,†††] we used a python program provided by Scott Kulp, Senior Computational Scientist and Developer at Climate Central, that grouped cells according to their center-point coordinate.[25] For aggregating grid cells by block groups and ZCTAs, we used another program that grouped cells according to their boundary contours.[26]

††† Time zones are needed for grid cells for solar electricity generation calculations (Eqn. 7).

Each grid cell was assigned to a single DMA, eGRID, and time zone. Because each of these aggregate regions include very large numbers of grid cells, errors in aggregated results introduced by the few grid cells that overlap more than one DMA or more than one eGRID region are negligibly small. More precise aggregation of grid cells was made for ZCTAs and block groups, as described earlier, since the size of these is typically of the same order of magnitude as the size of grid cells.

In some instances along coastlines and near national borders, there was the possibility that the geospatial sorting algorithm we used to assign most grid cells to one or another aggregated region would fail to assign the cell to any region. To address this issue, additional adjustments to the sorting algorithm were made such that grid cells adjacent to coastlines and national borders were assigned to the same aggregate region as the nearest cell that the algorithm assigned automatically to the region. This adjustment ensured that wind turbines or installed solar capacity located near national borders or coastlines was properly included in aggregate-region totals. In one corner case, the ZCTA 95314 was listed in the census shape file but in no other documents. Further research indicated that this ZCTA is uninhabited, and so it was assigned to the nearest populated ZCTA (95634).

In addition to DMA and eGRID aggregations, additional groupings of grid cells are also planned, but not yet implemented: states, counties, and federal congressional districts.

Wind/Solar Forecasting Tool v1 Reporting Metrics

The v1 release of the forecasting tool reports several solar and wind electricity metrics for “yesterday”, “today”, and “tomorrow” for each DMA and in the case of wind generation also for the eGRID region associated with that DMA.[‡‡‡] (When the geographic area of a DMA overlaps multiple eGRID regions, we associate the DMA with the eGRID region that contains the largest portion of the DMA.) The reason for reporting eGRID aggregation results only for wind generation is that, unlike solar, wind generators are generally located remotely from urban centers around which DMAs are concentrated. Such urban centers, if located within an eGRID region with a significant amount of wind generation, are likely consuming some of those wind-

[‡‡‡] Metrics for “yesterday” are calculated from different weather models (HRRR and RAP) than for “today” and “tomorrow” (GFS), which may result in some inconsistencies in metric values from yesterday to today and tomorrow.

generated electrons. The methodologies and data sources used to calculate the solar and wind metrics are described next.

Solar Electricity Metrics

Table 1 lists the metrics calculated for solar electricity generation, considering all solar-PV generators in the forecasting tool's database. Metric #1 is simply the 24-hour sum of grid-cell electricity generation estimates for all grid cells in the DMA of interest. The units are megawatt-hours per day.

Table 1. Solar electricity metrics reported by v1 of the wind/solar electricity forecasting tool. Metrics are calculated for each DMA.

SOLAR ELECTRICITY	Yesterday	Today	Tomorrow
1. Electricity Generated			
2. Percent of Homes Powered			
3. Percent of Daily Cost Saved			
4. Solar Power Index			

Metric #2 is the equivalent number of households that could be served in the DMA by the solar generation divided to the total number of households in the DMA. The denominator (total number of households in each DMA) was determined by identifying the counties[27] that make up a DMA and summing U.S. Census estimates of the number of households in each of those counties.[28,§§§] The equivalent number of households that could be served in the DMA is calculated as the total solar electricity generated in the DMA divided by HH_{elec}. The latter is the average electricity consumption per day per residential customer in the state in which the DMA is located. If the DMA overlays multiple states, the average of HH_{elec} values across those states is used. The HH_{elec} values were based on annual (2016) household electricity consumption data reported by the Energy Information Administration (EIA).[29,****] Daily consumption by the

§§§ All except one of the more than 200 DMAs include multiple counties, and DMA boundaries correspond to county boundaries. The Palm Springs DMA is unique in being wholly contained within a single county and not even covering the whole county. For this DMA, the number of households was estimated to be the number of television-owning households in the DMA. (personal communication from Sean Sublette, Climate Central, August 2018.)

**** 2016 data were used because EIA had not yet released data for 2017 as of mid-August 2018.

household (HH_{elec}) was assumed to be the same each day throughout the year for purposes of calculating this metric.

Metric #3 is an estimate of the fraction of a hypothetical household's electricity expenditures that would have been saved if the household had a PV-based generation system operating that day. For this calculation, we assume an average household PV array has a generating capacity of 5.5 kW_{DC}, which several sources suggest is a reasonable estimate for the current U.S. residential roof-top solar PV fleet.[30,31,32] We calculate the hourly generation from such a hypothetical 5.5 kW_{DC} array in each 0.05° grid cell, sum the hourly values to get daily generation, and then find the average of the daily generation across all grid cells in the DMA. This average value is divided by the average daily household electricity consumption in the DMA (HH_{elec}) to arrive at the fraction of the electricity bill saved (metric #3).

Metric #4, the Solar Power Index (SPI), indicates how the solar electricity that was (or is forecast to be) generated compares with the maximum amount of solar generation that could be expected with a clear sky. In the v1 release of the forecasting tool, we first calculate the maximum possible daily generation (under clear sky conditions) in the DMA for the "Today" time period for the installations in the Open PV database by calculating maximum values for each 0.05° grid cell for each hour using Eqn. 5 by replacing GHI with the clear-sky GHI (CS-GHI), which MESO provides with hourly resolution at the 0.05° grid cell level. We sum the maximum hourly generation values at each grid cell for the 24-hours constituting "Today" and then sum the resulting daily totals for all grid cells in the DMA to arrive at the daily maximum generation for the DMA for "Today", considering only the Open PV installations. The SPI value for "Today" is ten times the ratio of the actual solar generation calculated for all Open PV installations for "Today" divided by the maximum value calculated for the same DMA. The factor of ten is included such that the range in possible SPI values is 0 to 10.[††††] To calculate SPI values for "Yesterday" and for "Tomorrow", we first divide the total estimated actual solar MWh generated for "Today" (including MWh generated by the Open PV and EIA860 installations) by the calculated SPI value for "Today" to estimate the total maximum clear-sky

†††† The estimate of CS-GHI assumes reference clear sky conditions, with typical amounts and vertical profiles of aerosols, water vapor and other atmospheric gases that determine the degree of transmission of solar radiation through a cloud-free atmosphere. The reference atmospheric conditions for estimates of CS-GHI are assumed the same at all locations and dates in v1 of the tool. Thus, the actual GHI could exceed CS-GHI if the actual atmospheric conditions are more favorable for irradiance at the earth's surface than the assumed reference clear sky conditions. This would produce SPI values greater than 10. In such cases, the tool reports the SPI value as 10.

generation from all installed solar capacity in the DMA. The SPI values for "yesterday" and for "tomorrow" are then calculated as the total estimated actual solar MWh generated for each of these days divided by the total maximum clear-sky generation from all installed solar capacity (for "Today") in the DMA.[‡‡‡‡]

Wind Electricity Metrics

The wind electricity metrics are shown in Table 2. Metric #1 and #2 are calculated in the same fashion as the corresponding solar metrics. However, when a DMA does not have any wind generation facilities within its boundaries (which will often be the case because wind turbines are typically sited remotely from population centers), these metrics are not meaningful. Metrics #3 and #4 are analogous to #1 and #2, but are calculated for eGRID regions associated with DMAs, rather than for the DMA itself. The number of households in eGRID regions was determined in the same way as the number of households in each DMA: the number of households (US Census estimates) in each county that is part of a given eGRID region were summed to estimate the total number of households in that eGRID region.

Metric #5, the wind power index, is ten times the ratio of actual wind generation in the eGRID region for the 24-hour period divided by the maximum possible wind generation if all turbines in the region were operating at their rated capacity at all times. The factor of ten is included such that the possible range of WPI is 0 to 10. The maximum generation is calculated at the grid cell level using Eqn. 8 assuming a wind speed of 14 m/s at all times, the speed at which the wind power curve (Figure 6) indicates turbine output is maximized. The maximum output for all grid cells in the eGRID region are summed to give total maximum possible generation in the region.

‡‡‡‡ Because the reference atmospheric conditions for estimates of CS-GHI are assumed to be the same at all locations and dates in v1 of the tool (see prior footnote), the total maximum clear-sky solar generation potential will be nearly the same "yesterday" and "tomorrow" as "today".

Table 2. Wind electricity metrics reported by v1 of the wind/solar electricity forecasting tool. Metrics are calculated for each DMA and for each eGRID region associated with that DMA.

WIND ELECTRICITY	Yesterday	Today	Tomorrow
By DMA			
1. Electricity Generated			
2. Percent of Homes Powered			
By eGRID region			
3. Electricity Generated			
4. Percent of Homes Powered			
5. Wind Power Index			

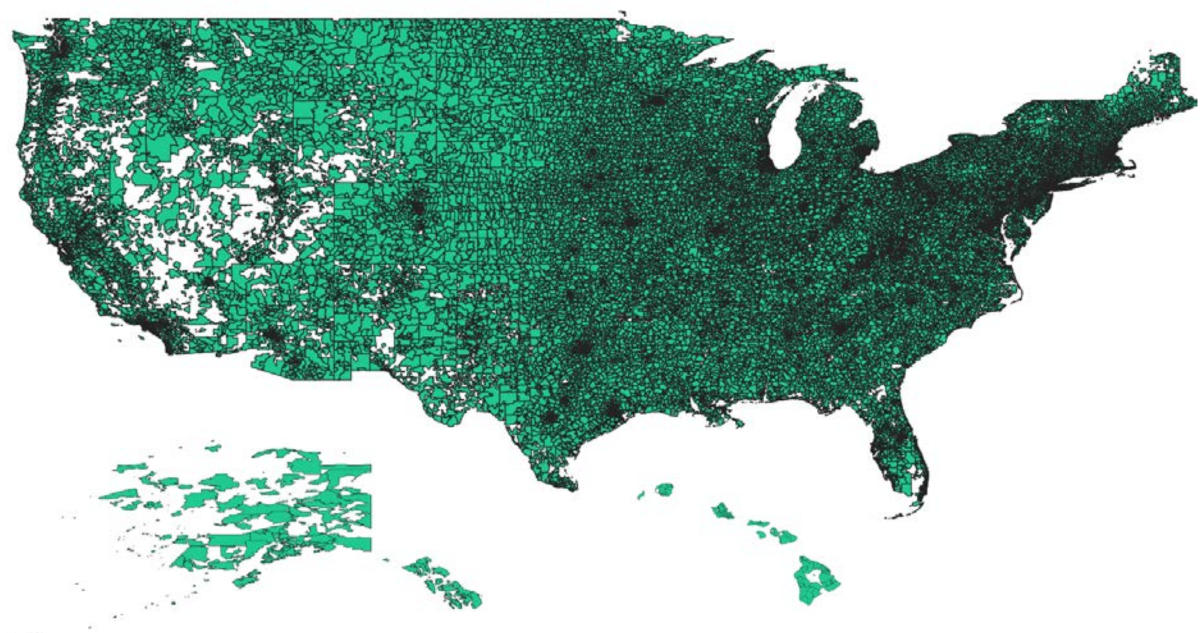

Figure 1. ZCTA regions as represented by GIS shapefile.[10]

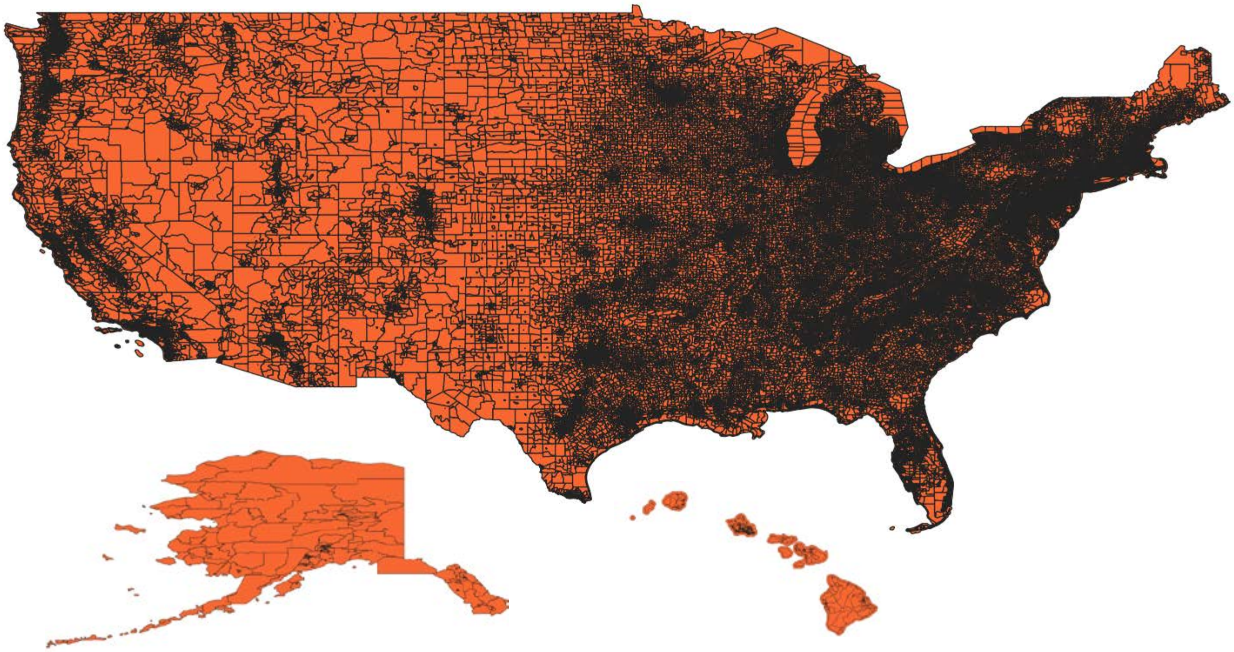

Figure 2. Census block groups, as represented by GIS shapefile.[12]

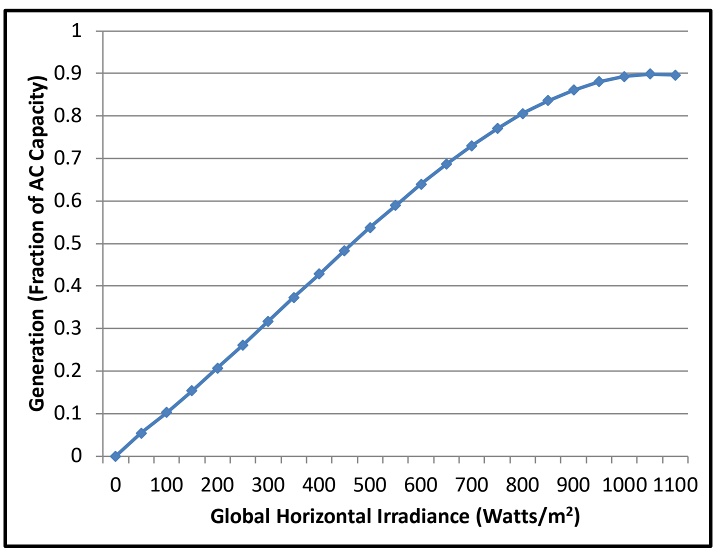

Figure 3. Aggregated distributed solar-PV curve.[17] The curve is the result of regression analysis of actual measurement data from individual roof-top systems and measured solar radiation data in an urban area of Hawaii. The measurements were for a substation-scale aggregate of residential and commercial (predominantly roof-top) PV systems.

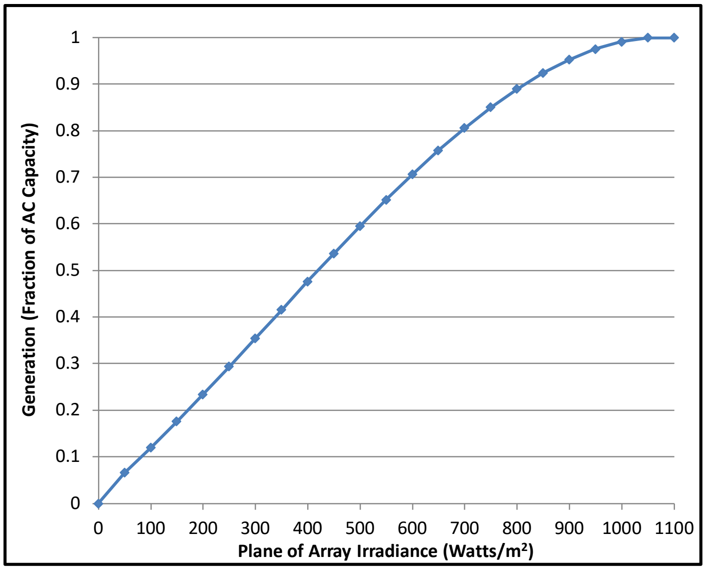

Figure 4. Power curve for utility-scale solar facilities. This curve is from a regression analysis of actual measured data from solar generation facilities in California. The input is plane of array (POA) irradiance in watts/m^2, and the output is power production as a fraction of AC capacity. The curve assumes a DC/AC capacity ratio of 1.3, which is typical.

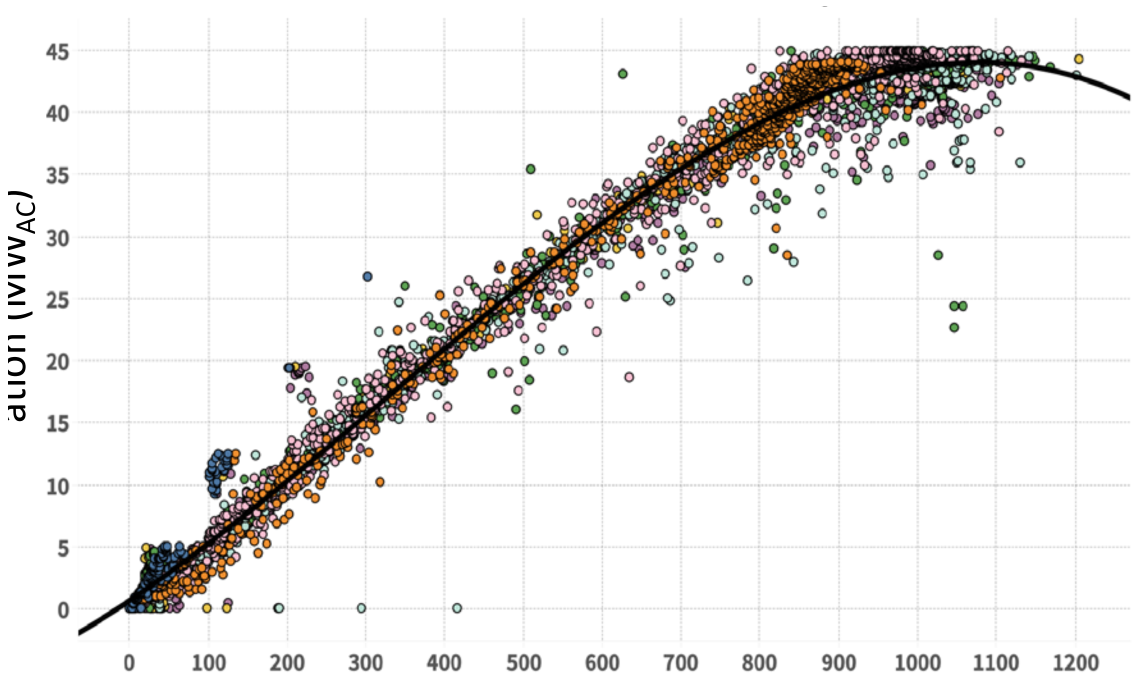

Figure 5. Comparison of generic power curve for utility-scale solar facilities (Figure 4) with actual data for a 45 MWAC capacity dual-axis tracking solar facility in California with a DC/AC ratio of about 1.33. The data points are 15-minute generation data, some of which have been adjusted to account for curtailments or equipment outages. [Actual production is impacted by weather conditions as well as outages and curtailments. The outages (number of panels or inverters offline) and curtailment (max generation allowed by the grid operator) are reported by the facility, and this information was used to scale the measured (reported) generation to an estimate of what would have occurred if there were no outages or curtailments.]

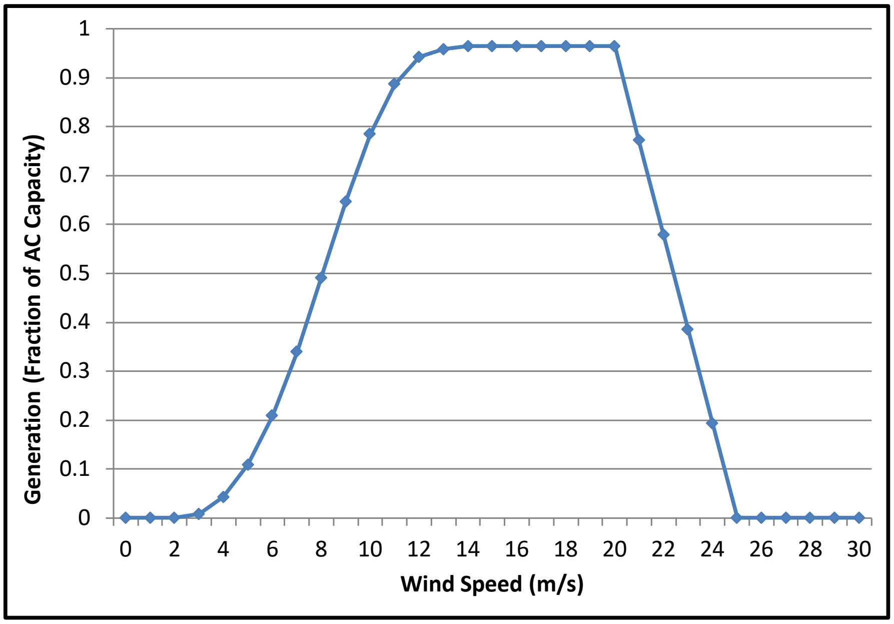

Figure 6. Representative facility-scale wind turbine electricity generation curve.[17]

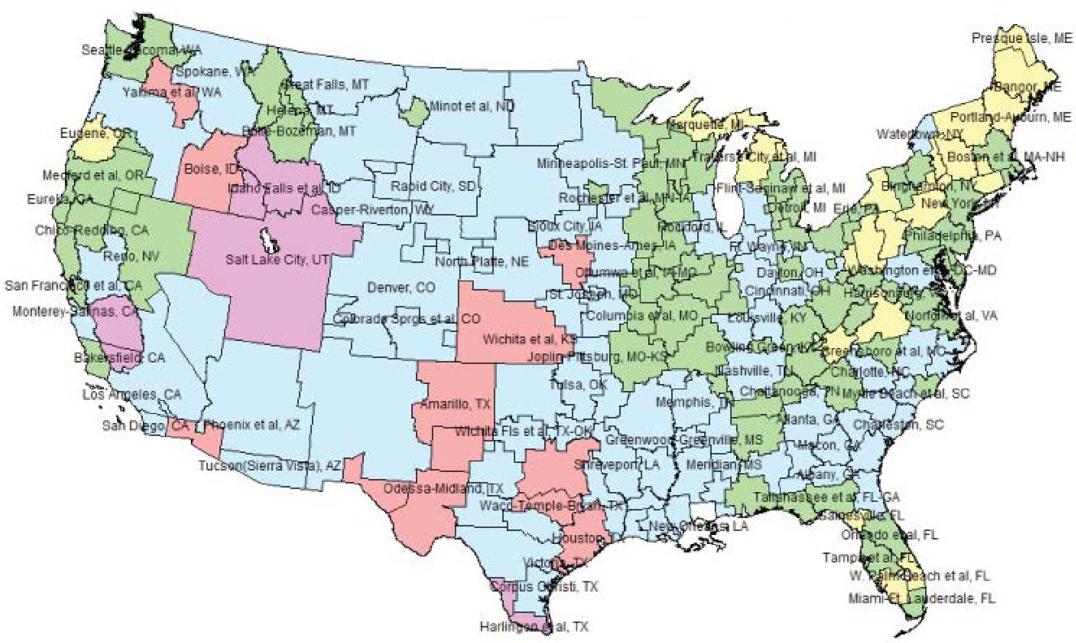

Figure 7. Nielsen Designated Market Areas (DMA).[33]

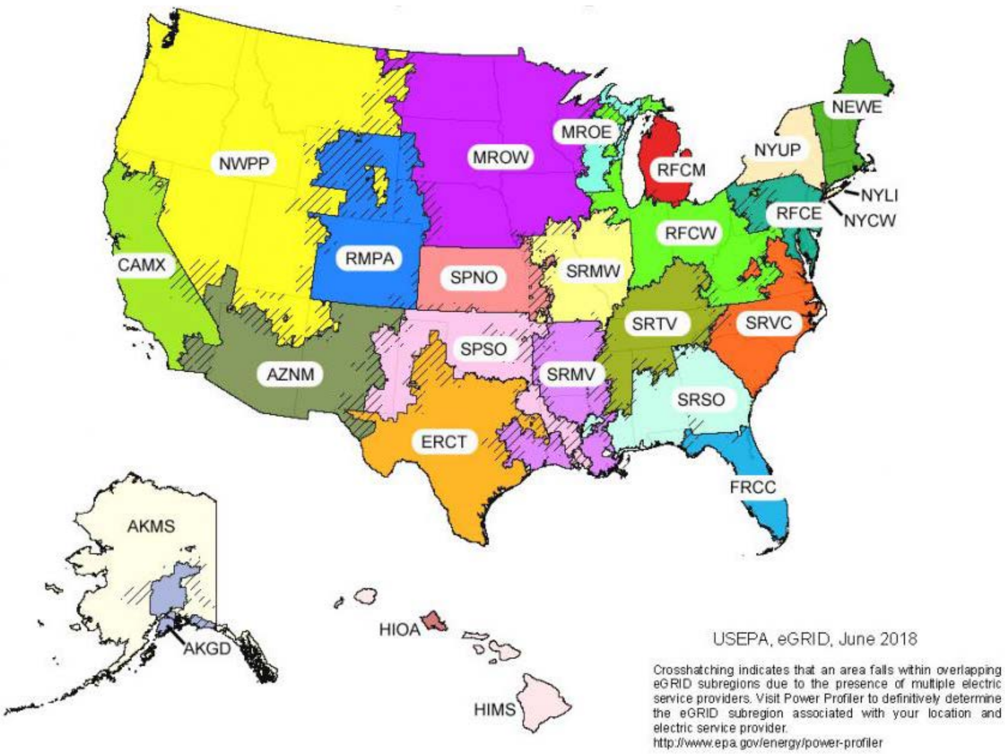

Figure 8. eGRID regions.[22]

References

1. http://www.meso.com/
2. https://rapidrefresh.noaa.gov/hrrr/
3. https://rapidrefresh.noaa.gov/
4. https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs
5. https://www.weather.gov/mdl/degrib_home
6. https://www.eia.gov/electricity/data/eia860/
7. K. Vickery and R.J. Baron, "Solar photovoltaic energy production comparison of east, west, south-facing and tracked arrays," Paper CSBE15070, CSBE/SCGAB 2015 Annual Conference, Edmonton, Alberta, July 2015.
8. National Renewable Energy Laboratory, Open PV Project, https://openpv.nrel.gov/.
9. G.L. Barbose, N.R. Darghouth, D. Millstein, K.H. LaCommare, N. DiSanti, and R. Widiss, "Tracking the Sun 10: The Installed Price of Residential and Non-Residential Photovoltaic Systems in the United States," LBNL-2001062, Lawrence Berkeley National Laboratory, September 2017.
10. https://www.census.gov/geo/maps-data/data/cbf/cbf_zcta.html
11. https://www.udsmapper.org/zcta-crosswalk.cfm
12. U.S. Census Bureau, "Cartographic Boundary Shapefiles - Block Groups," https://www.census.gov/geo/maps-data/data/tiger-line.html.
13. U.S. Census Bureau, "Total Population All Block Group," 2010. Prepared by Social Explorer (July, 2018). https://www.socialexplorer.com/.
14. https://www.seia.org/states-map
15. Aaron Holm (Data Analyst), personal communication, Solar Energy Industries Association, June 21, 2018.
16. USGS, LBNL, and AWEA, "The U.S. Wind Turbine Database," version uswtdb_v1_0_20180419 (https://eerscmap.usgs.gov/uswtdb/).
17. John Zack, President and Senior Research Scientist, MESO, Inc. (http://www.meso.com/), personal communication, July 2018.
18. M. Lave, W. Hayes, A. Pohl, and C.W. Hansen, "Evaluation of Global Horizontal Irradiance to Plane of Array Irradiance Models at Locations across the United States," SAND2014-16367J, *IEEE Journal of Photovoltaics*, 5(2): 597 – 606, 2015.
19. https://pvlib-python.readthedocs.io/en/latest/
20. https://www.pveducation.org/pvcdrom/properties-of-sunlight/solar-time
21. Environmental Systems Research Institute (ESRI), "U.S. Designated Market Areas, 2016," California: ESRI, 2016.
22. Abt Associates, "The Emissions and Generation Resource Integrated Database: Technical Support Document for eGRID with 2016 data," Office of Atmospheric Programs, Clean Air

Markets Division, US Environmental Protection Agency, February 2018. https://www.epa.gov/energy/egrid-technical-support-document

23. https://www.epa.gov/energy/download-egrid2016-shapefiles
24. US Geological Survey, "National Atlas of the United States: Time Zones of the United States, 2005," USGS, Washington, DC, 2005 (https://maps.princeton.edu/catalog/stanford-nt016bt0491).
25. S. Kulp (Climate Central), personal communication, June 2018 (python code: "ShpCointainsPoints.py").
26. S. Kulp (Climate Central), personal communication, June 2018 (python code: "ShpCointainsShp.py").
27. U.S. Census Bureau, "Cartographic Boundary Shapefiles – Counties," https://www.census.gov/geo/maps-data/data/cbf/cbf_counties.html
28. U.S. Census Bureau, "Household by Household Type All Counties," 2010. Prepared by Social Explorer (August, 2018). https://www.socialexplorer.com/.
29. Energy Information Administration, "2016 Average Monthly Bill - Residential," https://www.eia.gov/electricity/sales_revenue_price/.
30. https://www.nrel.gov/docs/fy17osti/68925.pdf
31. https://www.seia.org/research-resources/solar-photovoltaic-technology
32. https://news.energysage.com/average-solar-panel-size-weight/
33. https://www.multihousingnews.com/post/young-markets-2/