Conference Proceedings

ASES National Solar Conference 2016

San Francisco, CA, USA 10 – 14 July 2016

Some consequences to grid operation due to high penetration of distributed PV systems

Umnouy Ponsukcharoen[1] and Walter Murray[1]

[1] Stanford University, Stanford CA (The United States)

Abstract

Traditional forms of power generation could on their own be the sole supplier of power. This is not true of PV, which raises the question of what the limits are and how best to operate and modify the grid to maximize the limits. Here we analyze distributed PV since it is the most problematic and currently has the fastest growth. We study bounds on limits to distributed PV based on its generation profile compared with the grid demand profile. We use data collected over two years from two postal codes, one in New Jersey and one in California.

Keywords: *distributed PV generation, daily maximum PV power curves*

---

1. Introduction

We concern ourselves with residential PV, which is the fastest growing sector (SEIA, 2014) due in part to the fact it does not face protracted planning permission. The ideas presented are easily extended to other sectors. PV suffers from variation due to time of the year, time of the day, environment and weather. It is the later that causes the great concern since it makes PV generation volatile. The great virtue of the current means of generation is it enormous inertia, which in turn makes dynamical issue rare. High volatility clearly makes matching demand to generation hard.

As noted PV cannot in itself be the only form of generation. We address the issue of what the maximum possible limit is to the penetration of PV. We deal only with residential PV but the ideas we present can be extended to commercial sites and sites operated by the utility. The hard case is caused by distributed PV. Right now almost anyone who wants to install a system may do so. A point may be reach where permission may be withheld or come with restrictions. We investigate where that point may be. Our primary tool is extrapolate the generation of PV and see when a point is reached where it exceeds demand. That gives an upper bound. However, that bound would be lowered by any amount a utility was not prepared to turn off such as their own PV systems. Surplus power could be stored but doing so raises the cost of PV as does the inability to sell surplus generation.

To determine when PV generation exceeds demand we need to model current generation and extrapolate. To do that we need to a maximum power curve for each day of the year. On means of doing that for a set of PV installations is analyzing the features of the installations along with the irradiation at the locations. This is complex and to be done accurately would need to take into account shadows, state of the systems, etc. Also it is constantly changing as more installations are made. We call this a structural approach. Modeling platforms by PVMPC (Hansen et al., 2014) and PVSyst may be used to compute the PV power output from quite a number of inconstant inputs. Estimates of the maximum penetration limit from various studies (Lopez et al.,

doi:10.18086/solar.2016.01.18 Available at http://proceedings.ises.org

2012; Energy and Inc. Environmental Economics, 2012) relies on this approach. We propose a reduced-form approach for constructing maximum PV power curves, which describes the maximum power from a group of PV systems at any time instance. To generate such a curve we use the *output* of a sample set of systems. Since this sample is fixed regardless of how many systems are installed this is a manageable set of information for the grid operator to collect. We first define and validate the concept of a normalized power curve, which is needed to construct a maximum power curve. We then construct a maximum power curve from a normalized power curve for a group of PV systems in two different geographical areas. Finally, we illustrate how to use the maximum power curve of a group of PV systems in two applications: constructing the expected PV power curve finding the maximum PV penetration limit.

An obvious issue is the size of sample and how that can be used to meet the required accuracy. We analyzed the impact of varying sample size and showed that for California 50 sites was sufficient and that New Jersey required 100 sites. The reason the sample is so small is the generation for two sites is highly correlated. We expect sample size will vary a little with the area over which the sites are distributed and with the variability of the weather between locations of sets of sites.

2. Data

2.1 Data set

The time series of power outputs from 2 groups of PV systems were collected at a 15-minute frequency by a solar PV monitoring company. Each time series of a PV system starts on either January 1st , 2014 or on its installation date. The locations of the PV systems in each group are from the same zip code. A grid operator's perspective may prefer a group of PV systems linked to the same substation rather than the same zip code. Still we assume the same characteristics. We name the 2 data sets as `CA' for the group in one particular zip code of California, and `NJ' for the group in New Jersey. The installed capacity of each PV system is reported as a range of 0-1 kW, 1-2 kW, 2-3 kW, 3-5 kW, 5-10 kW, 10-20 kW, 20-50 kW, 50-100 kW, or 'NaN' since the exact installed capacity is confidential.

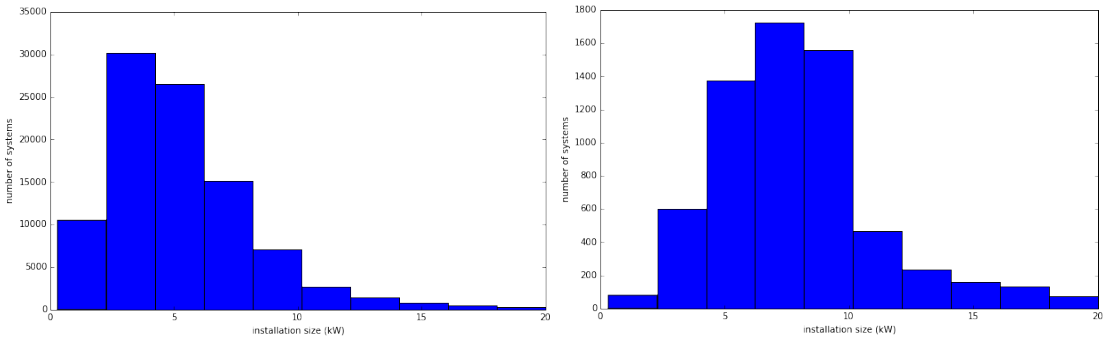
2.2 Data cleaning

In order to construct normalized maximum power curves correctly in an organized manner, we clean data by first removing data from systems with an installed capacity greater than 20 kW as we are only interested in residential-size PV systems. Each time stamp in the time series is adjusted so that it ends with 00:00, 15:00, 30:00, or 45:00.

2.3 Installed Capacity Estimates

In order to define a normalized power curve, the installed capacity of each PV individual system should be known. However, due to confidentiality this information is not available. Instead, the installed capacity of a PV system is given as a range. If we assume that the installed capacity is uniformly distributed in a range, then we can estimate any installed capacities in the range to be equal to the midpoint of the range. We still aggregate data from several sites so it is not necessary for the data from a specific site be known accurately.

In order to find a better estimate of the installed capacity, we investigated other sources of data where information about the installed capacities of PV systems is available. NREL's Open PV project has a data set of installed capacities of PV systems over time and region. Figure 1 shows histograms of installed capacities of PV systems in California and New Jersey from 2008 and 2015.



a) California b) New Jersey

Fig. 1: Histograms of installed capacities up to 20 kW in 2008-2015 from NREL's Open PV project

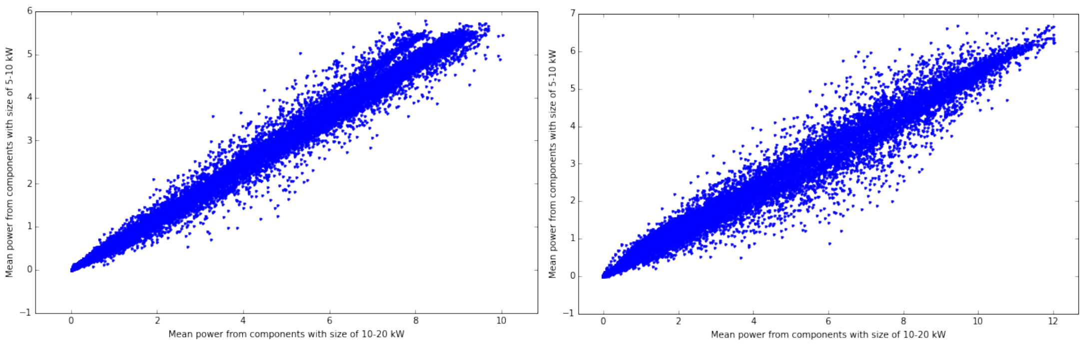
From the data in this histogram, we computed the mean value in each bin and use it as an estimate for the installed capacity in our data set. We show in Section 3.1 that this estimate is better than using the midpoint of the range.

3. Methods and results

3.1 Definition and validation the concept of a normalized power curve

A normalized power curve for a group of PV systems is defined at each time instance as the total power from the group divided by its total installed capacity. The notion of normalization is important because scalability by the capacity of the PV installation is an essential feature of a maximum power curve model. The current PV generation capacity is low compared to both the total load and the potential for new PV installations. Consequently, it is possible that the maximum power curve for a current group of current PV systems may not be applicable to a system with high PV penetration. Hence, we establishes the consistency in defining a normalized power curve. With this normalized power curve, we can scale it in order to estimate future limitations on the capacity of PV installations.

To ensure that the definition of a normalized power curve is consistent, we first demonstrate that the power generated by a group of PV systems is proportional to its total installed capacity. We compute time series of the mean power by the number of systems from PV systems with a size of 5-10 kW and another time series from PV systems with a size of 10-20 kW. The scatter plot of the mean power from two groups as shown in Figure 2 shows proportional relationships from both `CA' and `NJ' data sets. The slopes of the relationships match the ratios of ratio of our installed capacity estimates for PV systems in a 5-10 kW bin and a 10-20 kW bin from Open PV project data rather than the estimates from the midpoints of the bins. This means the estimates from Open PV project is better than the estimates from the midpoints of the bin.

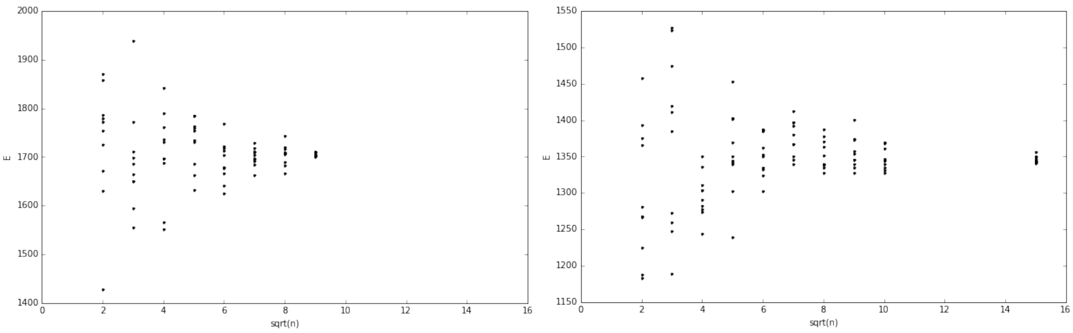


a) 'CA' data set b) 'NJ' data set

Fig. 2: Scatter plots of the mean power by PV systems from 2 bins: 10-20 kW and 5-10 kW

Next we observed that the deviation of energy generation from a group of PV systems from different years of data collection is small. For 2013 and 2014 it is about 1-2%. We confirmed that the deviation of energy generation of groups of PV systems with different installation dates is also acceptable. The deviation of yearly energy generation from a group of PV systems installed in different quarters of 2013 is about 4-8%.

We identified a sufficiently large number of PV systems for consistent normalized power curves. As shown in the Figure 3, the average absolute deviation in yearly energy generations decreases as the number of PV systems increases. We need only 49 systems to achieve a deviation of less than 1 percent for the `CA' data set while we need 100 systems to obtain the deviation of less than 1 percent for the `NJ' data set.



a) 'CA' data set b) 'NJ' data set

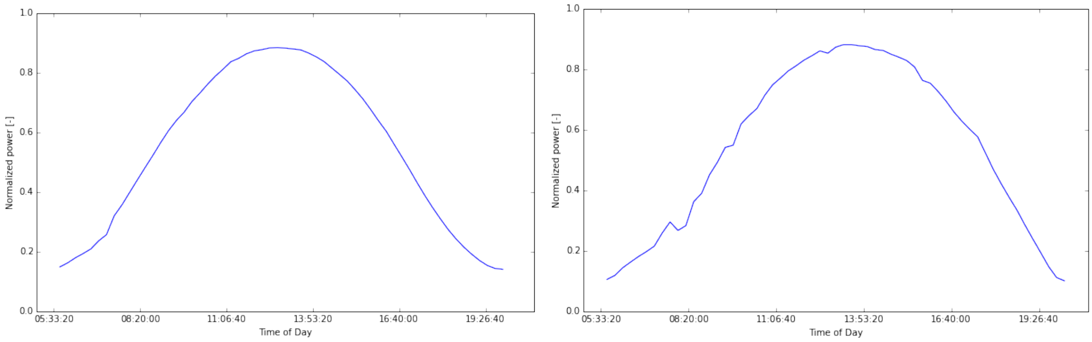
Fig. 3: Yearly energy generations derived from a different number of PV systems. Each number has 10 samples. The horizontal axis shows the square root of the number of PV systems

Resulting from these consistency checks, the heuristics to produce consistent normalized power curves is:

1. Clean the data so that time series of power outputs from PV systems are aligned. Make sure that the installed capacity of each PV system is known or well estimated.
2. Determine a point on a normalized power curve by summing all power outputs and dividing them by the sum of installed capacities associated with the power readings at that instance. As a rule of thumb, the number of PV systems should be at least 50
3. Check if there is a need to distinguish data from different years of collection or from PV systems with various installation dates. However, due to a limited amount of data, we may treat them equally and accept a possible deviation of about 4-8%.

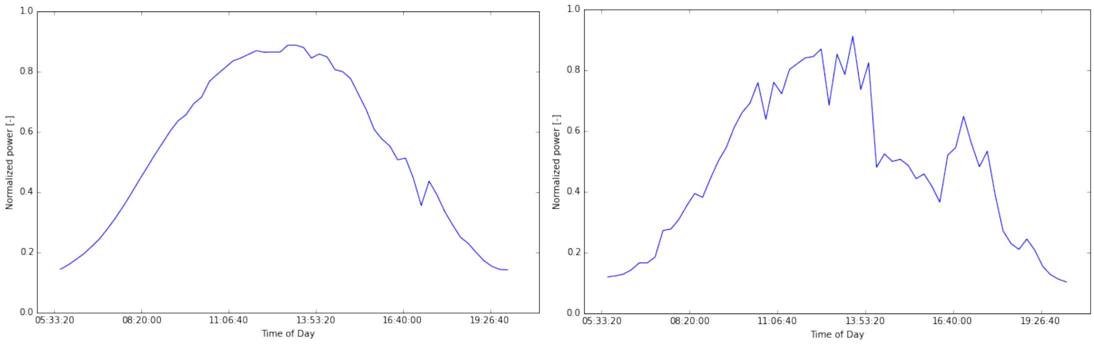
3.2 Construction of a normalized maximum power curve

To get an insight of how to construct a normalized maximum power curve for every day in a year, we first consider several daily normalized power curves generated from the `CA' data set and the `NJ' data set using heuristics described in the previous section. Some examples are shown in Figure 4 and 5. We found some daily normalized power curves that resembles a simple bell shape with a high daily energy generation relative to neighboring days. This leads to an idea that we may model these well-behaved power curves (Figure 4), which are likely to be the maximum normalized power curves, and interpolate models to all the remaining days in a year. A function that models maximum normalized power curves should be non-negative and similar to a bell shape with the domain of time interval between sun rise and sun set.



a) ,CA‘ data set b) ,NJ‘ data set

Fig 4: Well-behaved daily normalized power curves on 2014/6/24



a) ,CA‘ data set b) ,NJ‘ data set

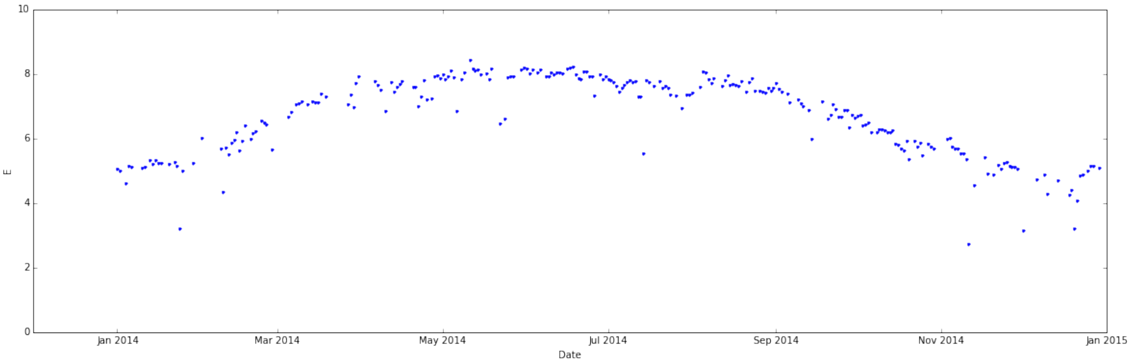
Fig 5: Typical daily normaliized curves on 2014/6/21

In order to construct models to neighboring days in which the sun rises and sets at different times, we transform the domain of daily power curves so that they are common to all days. We define s as a negative consine of an angle on a facial plane of an observer facing the south. Such a value can be computed from

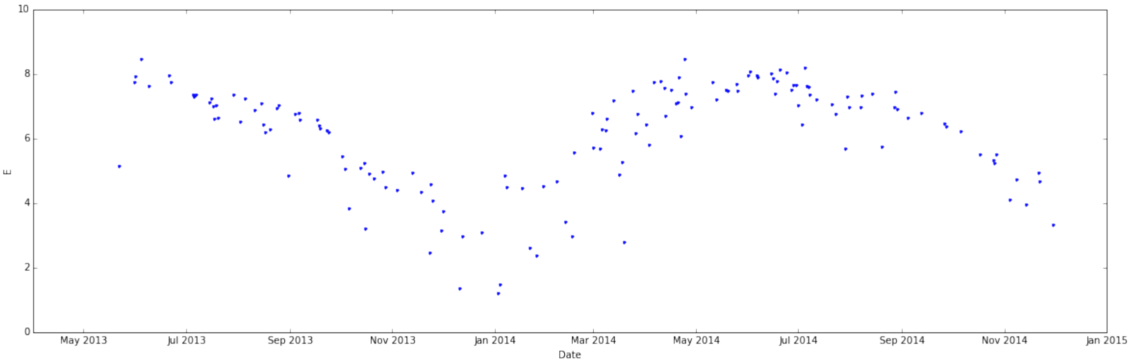
$$s = \frac{y}{\sqrt{y^2 + z^2}} \qquad \text{(eq. 1)}$$

, where (x,y,z) is a cartesian coordinate of the sun in which the positive x-axis points to the south and the postive y-axis points to the east. With this transformation s = -1 corresponds to a sun rise and s = 1 corresponds to a sun set.

Next, we automate the selection of well-behaved power curve selection. In the first stage, we perform regression on each daily normalized power curve and filter out the power curves that either fit poorly (R-squared < 0.95) or has more than one critical point. In the second stage, we filter out the power curves that have significantly low daily energy generations relative to the neighboring days. In order to do such a task, we transform the plot of daily energy generations of all daily powers that pass the first stage (Figure 6) into a plot against a variable called Day Before a Winter Solstice (DBW) a single year window (Figure 7). The DBW for each day is defined such that DBW for a Winter Solstice of each year is zero and DBW for any other day is negative integers up to -365. From Figure 7 to Figure 8, we remove a point below the envelope if there exists a higher point in the plot such that the line segment connecting them has a slope higher than a threshold. Daily power curves associated with remaining points in Figure xx are claimed as well-behaved power curves.

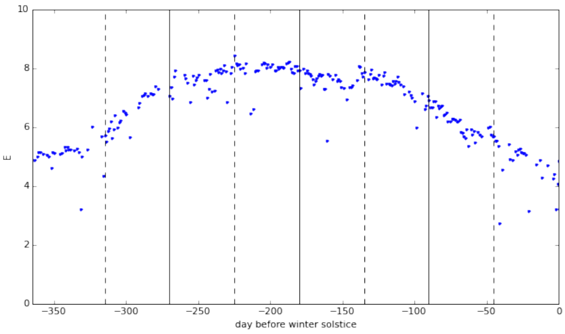


a) 'CA' data set

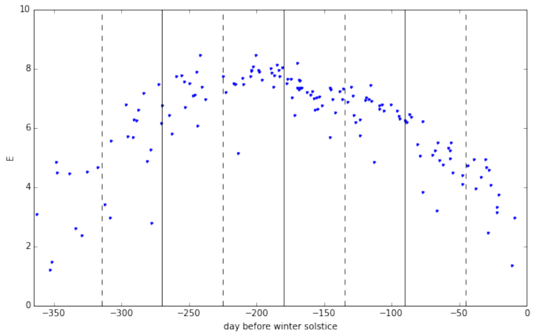


b) 'NJ' data set

Fig 6: The daily energy generations of all daily power curves that pass the 1st stage.

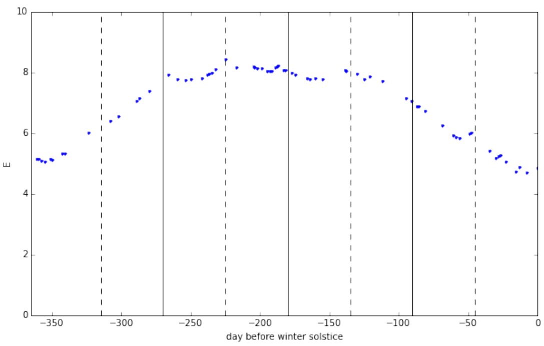


a) 'CA' data set

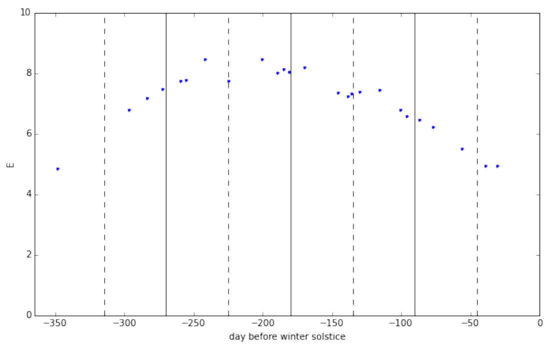


b) 'NJ" data set

Fig 7: The daily generations of daily power curves that pass the 1st stage with a new variable DBW and overlay.



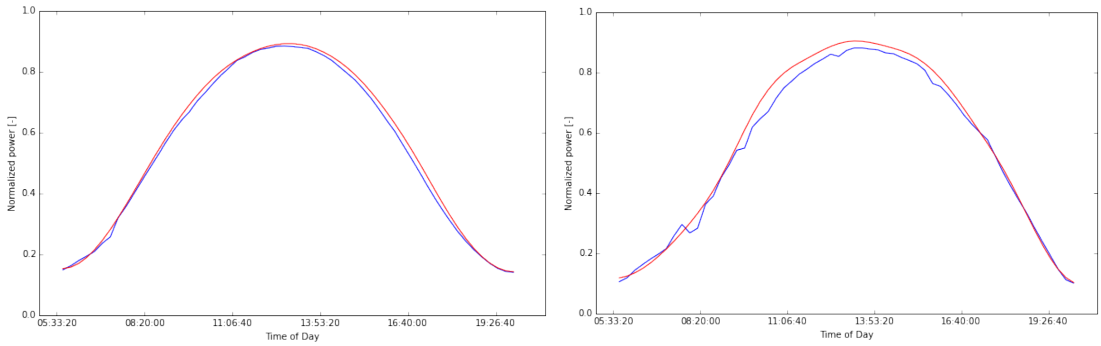
a) 'CA' data set



b) 'NJ" data set

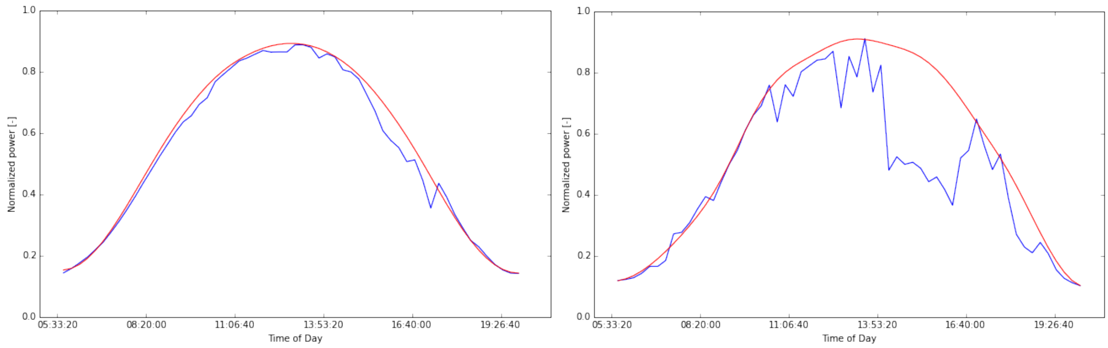
Fig 8: The daily generations of daily power curves that pass the 2nd stage with a new variable DBW and overlay.

After that, we generalize models by peforming quantile regression with $\tau = 0.5$ for coefficients of B-splines from well-behaved power curves. A family of Fourier series is used in the regression as the periodicity of coefficients over a year is required. Once all coefficients are found for all the days in a year, we construct maximum power curves for all days. Figure 9 and 10 are examples of daily maximum normalized power curves. Figure 11 shows daily energy generations of maximum power curves in comparison wit all daily power curves.



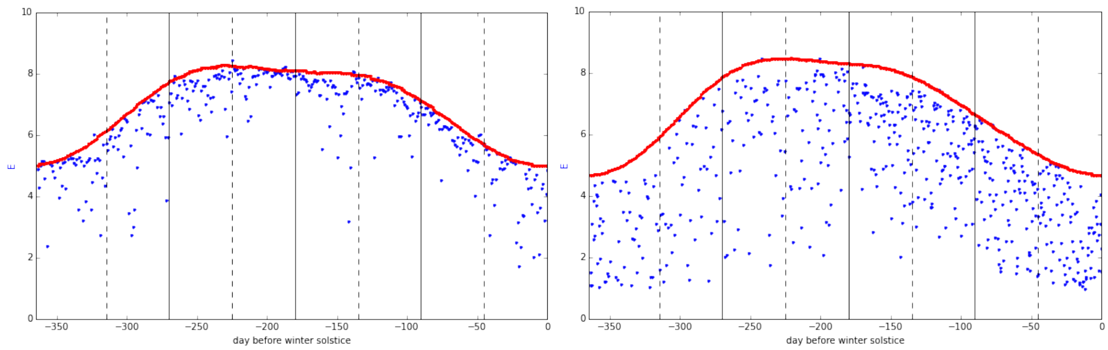
a) ‚CA‘ data set b) ‚NJ‘ data set

Fig 9: Well-behaved daily normalized power curves and proposed maximum power curves on 2014/6/24



a) ‚CA‘ data set b) ‚NJ‘ data set

Fig 10: Typical daily normaliized curves and proposed maximum power curves on 2014/6/21



a) 'CA' data set b) 'NJ" data set

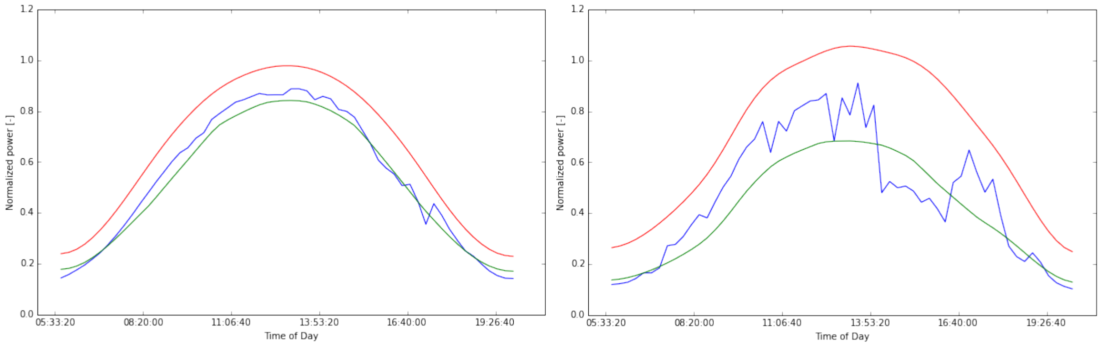
Fig 11: Daily generations of proposed maximum power curves in comparison with all daily power curves

Note that we may adjust the generated maximum power curves by scaling and shifting so that they are always higher than any normalized power curves from the data.

4. Determining the maximum PV generation

4.1 Expected PV power curves

An expected power curves is defined as a graphical illustration of an average power output from a group of PV system over a day. The expected PV power curve should reflect a PV power's variability but not volatility. Such a curve is useful in a day-ahead planning as the grid operator needs to supplement extra energy generation to match the demand. A simple expected PV power curve can be constructed by scaling a maximum PV power curve by a factor. A suitable factor is the mean of performance ratios in a year. One can refine the expected PV power curve by scaling each part of the maximum PV power curve with a different factor depends on a time of a day and a day of a year. Such a factor is the mean of performance ratios in a bucket of (s, DBW) pairs. With a simple linear interpolation and periodic boundary conditions, we generate a continuous function of the scaling factor and multiply it with the maximum power curve to obtain the expected PV power curve. Figure 12 shows examples of daily expected power curves (green) in a comparison with maximum power curves (red) and actual power curves (blue).



a) ‚CA' data set b) ‚NJ' data set

Fig 12: Typical daily normaliized curves and expected power curves on 2014/6/21

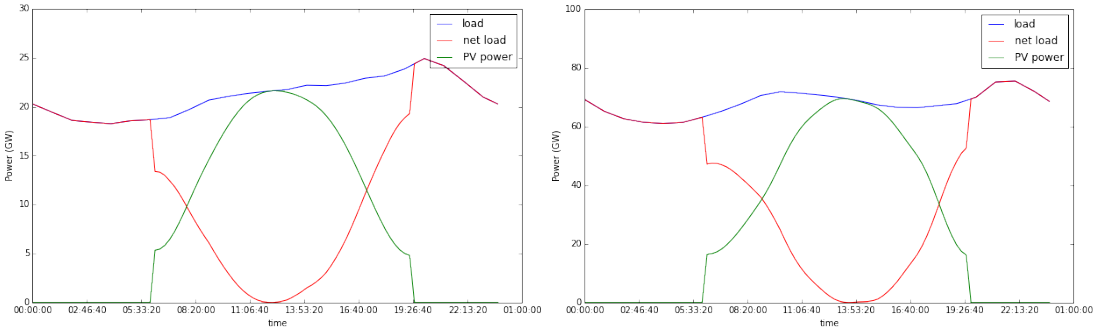
4.2 Finding the maximum penetration limit

We define the maximum penetration limit to be the maximum generation such that at no point is supply greater than demand. Here the whole grid is assume to have no storage capacity. On the other hand, the grid operator can transfer the power from one point to another in the grid without any constraints. All PV systems in the grid behaves similar to normalized power curves from our sample PV systems.

It implies that, if the ideal generation of PV systems is assumed, the maximum installed capacity is equal to the minimum of ratios between the load curves and the normalized maximum power curves over a year. In the case of realistic generation, however, the maximum installed capacity is equal to the minimum of ratios between the load curves and the normalized expected power curves over a year. The numerical values of contribution from PV generation are summarized in Table 1.

Tab. 1: A contribution of PV to annual peak load and annual demand in 2014

Quantities	CAISO	PJM
% Peak load (GW)	29 (at 5:00 pm August 1^{st})	26 (at 6:00 pm June 17^{th})
% Demand	28	24



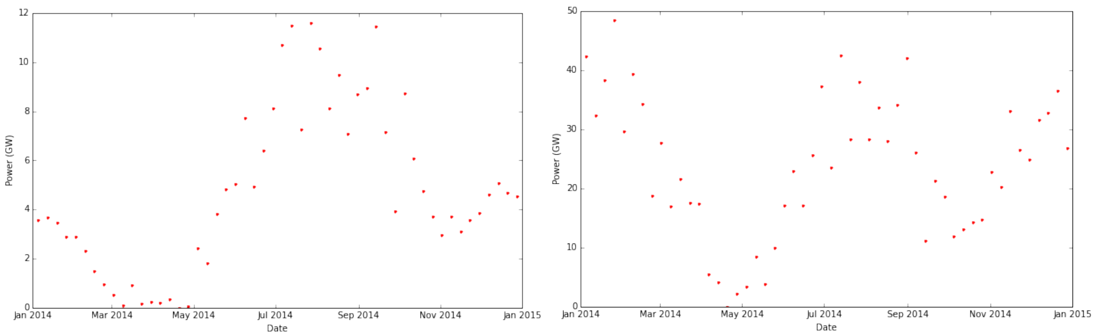
a) California b) New Jersey

Fig 13: Load, PV power and the net load curves under the maximum penetration limit on 2014/4/20

Recall this assumes we have the maximum PV generation. Obviously the real contribution will be smaller. On average the real contribution for CASIO is about 80% and for PJM 55%. Note that the behavior of PV generation near sun rise and sun set is problematic to extrapolate but plays no role in determining PV limits. Surprisingly the critical day for both Cal data and the NJ data was the same despite the disparity in locations. However, the critical time in the day is significantly different due to the NJ profile although being similar to CA has an additional dip in demand. We had envisaged that each zone would require its own analysis but it could be the differences are sufficiently close that a good enough global solution can be obtained.

We have made a number of assumptions the most obvious of which is assuming that we have perfect weather. The limit could be raised by applying the same methodology, but using daily normalized expected generation curves (see Fig 12). To match the normalized maximum power curves means raising installation in California about 25% and in New Jersey about 80%. Such curves lead to a very similar contribution to the grid as using the maximum power curve. However, there will be a 50% probability of either exceeding or not meeting load at the critical point and a significant probability elsewhere. As a consequence it is highly likely the contribution to the grid will be lower than the figures for the maximum power curve. When we have a bad weather day we are below the expectation but when it is a good day we do not get the full benefit if we exceed the load. For real generation the outlook is bleaker since volatility will exacerbate both the degree of mismatching load and the frequency it arises.

A serious cause for concern is the rapidly changing generation the utility now has to provide. Note this is not caused by volatility but from giving preference to generation from distributed PV. Almost all other considerations such as the assumption of being able to transport energy to any part of the grid will lower the limit. The exception is storage but this has financial implications and also physical limits on how fast it can absorb and release power. Raising the PV curve in fig xx above demand by a small amount would not be an issue but anything significant raises the rate surplus power needs to be absorbed and released. All this would be exacerbated by the volatility of PV.



a) California b) New Jersey

Fig 14: The minimum net load for Sundays in 2014

5. Conclusion

By analyzing extensive residential PV generation data from two disparate zones we were able to construct scalable models of the growth of residential PV contribution to the grid. We then assumed we have perfect circumstances for PV generation and determined the point at which PV generation exceeds load. At such a point the utility has to shut down all its generation. Growth of PV installations beyond this point implies that there will be progressively less generation as a percentage of installation

It is unlikely a utility will decrease generation to accommodate distribute PV generation to the point that we computed the limit. Any PV generation the utility owns is similar in behavior to distributed PV and as such reduces the level that distributed PV is attractive. Other utility renewal generation will also have priority along with base load generators such as nuclear power. PV generation distorts the load profile that other generation needs to match. Rather than flattening the load it distorts it further from the ideal of a flat line. This is not something that can be addressed easily by suitable pricing models. Making electricity very cheap from 11am to 1pm would undermine the financial benefit of distributed PV to the home owner. The great hope would be storage but it would need in total to be utility scale and be able to have fast charge and discharge capability.

6. References

Solar Energy Industry Association (SEIA), 2014. Solar Market Insight report 2014 Q4. http://www.seia.org/research-resources/solar-market-insight-report-2014-q4

Anthony Lopez et al., 2012. US Renewable Energy Potential Analysis: A GIS-based analysis, NREL technical report. http://www.nrel.gov/docs/fy12osti/51946.pdf

Energy and Inc. Environmental Economics, 2012. Technical Potential for Local distributed Photovoltaics in California http://www.cpuc.ca.gov/NR/rdonlyres/8A822C08-A56C-4674-A5D2-099E48B41160/0/LDPVPotentialReportMarch2012.pdf

Clifford Hansen, Robert W. Andrews, Joshua S. Stein and Daniel Riley, 2014. Introduction to the open source pv lib for python photovoltaic system modeling package. Sandia National Laboratories.

PVsyst SA. Pvsyst photovoltaic software. http://www.pvsyst.com/en/

Open PV project. National Renewable Energy Laboratory. https://openpv.nrel.gov/