PERFORMANCE ASSESSMENT WITHOUT PYRANOMETERS: PREDICTING ENERGY OUTPUT BASED ON HISTORICAL CORRELATION

Anastasios Golnas[1], Joseph Bryan[1], Robert Wimbrow[2], Clifford Hansen[3], and Steve Voss[1]
[1]SunEdison, Beltsville, Maryland, United States
[2]San Clemente, California, United States
[3]Sandia National Laboratories, Albuquerque, New Mexico, United States

ABSTRACT

Estimating the energy that should be generated by a PV system under the prevailing conditions of irradiance and temperature is very important for system or fleet investors, energy customers, and operators. Traditionally this has been achieved by measuring irradiance and temperature in the proximity of the PV array in order to calculate the expected energy output either by using an appropriate model or by making assumptions about the system's Performance Ratio. However, this method requires the accurate installation, maintenance, and continuous monitoring of sensors thereby increasing the system's capital and maintenance costs. In this work we present an alternative methodology which can calculate the expected output of one or more systems in a regional fleet based on the measured power output from a subset of the total fleet. This can be achieved thanks to high accuracy energy measurements and the ability to correlate historical performance records.

INTRODUCTION

One key characteristic of power and energy generation from PV systems is the dependence on external quantities like irradiance and temperature which tend to vary over a wide range of values at many time scales (from seconds to years). This dependence poses a challenge in the prediction of a system's performance and constitutes an issue for all the parties having a vested interest in the optimal performance of a system, including the energy consumers, the system owners or investors, and the system operators.

To overcome this challenge the PV community has devised various metrics among which the Performance Ratio (the energy generated by a system over a specific time interval normalized by the power rating of the system at standard test conditions (STC) and the insolation incident over the same time interval) is the most widely accepted.

Once the Performance Ratio of a hypothetical system is calculated based on a long-term historical weather profile and certain assumptions about energy losses in the system, then one can estimate the energy that can be expected from such a system under a specific insolation input. This method of assessing performance depends on the accurate and continuous monitoring of irradiance via a pyranometer that is mounted in the plane of the array. There are numerous challenges associated with scaling this method over a very large number of systems: the cost of accurate instrumentation, the cost of optimal placement and maintenance, the reliability of the instruments and the data acquisition system.

This work presents an alternative method that can be used to estimate the performance of PV systems belonging to a regional fleet when local insolation data are unreliable or non-existent.

The basic premise behind the proposed method is that longer-term (daily, weekly, monthly) performance of neighboring systems – absent outages – is sufficiently correlated to permit the prediction of a system's output using information from neighboring systems. This can be achieved in principle by building an NxN matrix that describes the correlation of the generated energy among N different systems belonging to a regional fleet, about which we only know their size and geometry (tilt and azimuth). The coefficients of the matrix can be calculated starting from constant-frequency meter reads over the course of a 'training' period. Subsequently, the expected generation from any given system over a specific period can be calculated by multiplying the matrix with the energy generated from each of the remaining (N-1) systems over the same period and averaging the results.

In this paper we present results over an eight-month period (April-November 2010) based on training periods of 1-7 months. We analyzed (N=55) PV systems operated by SunEdison in the state of New Jersey with a geographical distribution shown in Figure 1 and with nameplate sizes ranging from 30 to 500 kWp.

The success of the various implementations of the method is measured by the deviation of the predicted energy for each of the systems from the measured energy: a most successful implementation predicts the energy within an arbitrary accuracy threshold (e.g. ±5% or ±10%) for the largest possible fraction of the fleet.

One of the most interesting aspects of our study has been the relative influence of the historical correlation of the systems' performance versus the geographical proximity. If we start from a baseline defined by the arithmetic mean of the Bird Performance Index (BPI – defined in the Methodology section) of the N-1 systems as a predictor for the N^{th} system, we can then assume that any additional information about these systems will increase the success rate of the prediction, i.e. a larger fraction of *N* will fall within the desired accuracy threshold.

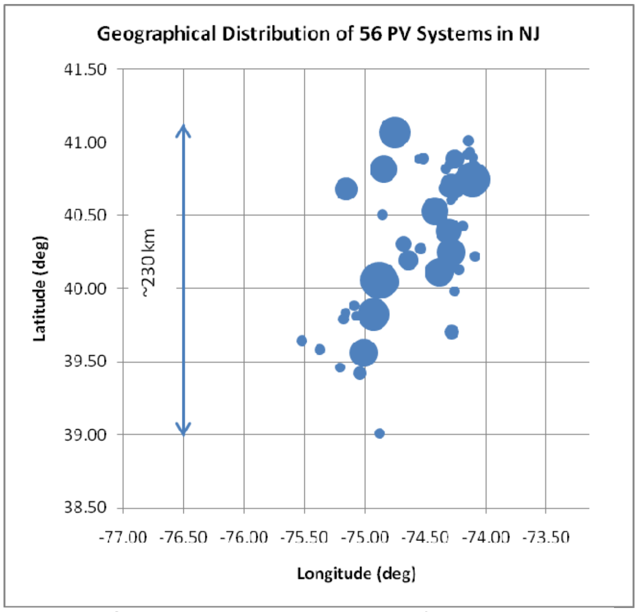

Figure 1: Geographical distribution of 55 PV systems in the state of New Jersey. The area of the circles is proportional to the nameplate rating of the systems.

METHODOLOGY

The algorithm is fed with two data streams: (a) the 15-minute energy values from each system's generation meter, and (b) the 15-minute clear-sky plane-of-array irradiance values as estimated by the Bird Clear Sky Model. We then calculate the daily sum of each stream. The daily energy total is divided by the potential total energy at nameplate power (e.g. a clear summer day may produce 7 nameplate-hours of energy); the ratio is denoted here by $E_i(d)$, where i indexes the systems and d denotes the day. The clear-sky irradiance values are integrated to find the potential total insolation incident on the array, $I_i(d)$, expressed in suns (1000 W/m^2).

For each day and each location, we compute a Bird Performance Index (BPI):

$$BPI_i(d) = E_i(d) / I_i(d) \qquad (0)$$

i.e., the fraction of the clear-sky energy actually produced each day.

A set D of training days is selected by filtering out days containing any system outages. From the valid training days, we define at each location a clear sky BPI as the 93rd percentile of the set of BPI values:

$$CSBPI_i = Q_{0.93}\left\{BPI_i(d) : d \in D\right\} \qquad (0)$$

and regard this value as typical of the system's performance on a clear day.

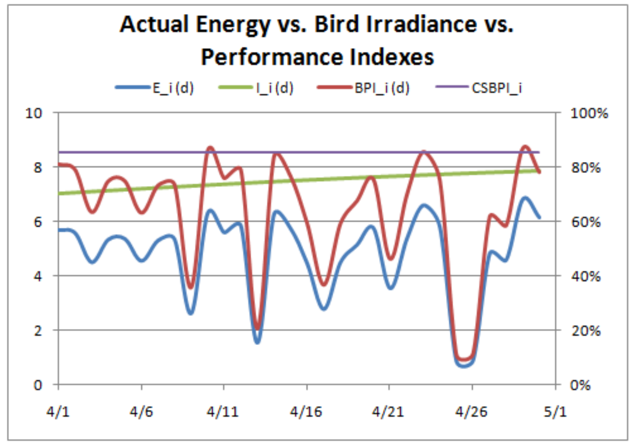

Figure 2: For sample system *i*, compare (a) daily energy vs. clear sky daily insolation, and (b) BPI vs. clear sky BPI.

Three different relationships between the *N* systems are then used to compute the estimated energy. First, a Conversion Profile $CP_{i,j}$ is computed as the ratio of CSBPI values for each pair of systems:

$$CP_{i,j} = CSBPI_i / CSBPI_j \qquad (3)$$

The conversion profiles are used to adjust for differences in system performance ratios. Second, the coefficient of determination (i.e., R^2) is calculated over the valid training interval for each pair of sites:

$$R^2_{i,j} = 1 - \frac{\sum_{d \in D}\left(BPI_i(d) - BPI_j(d)\right)^2}{SS_i \times SS_j}, \qquad (0)$$

$$SS_i = \sum_{d \in D}\left(BPI_i(d) - \overline{BPI_i}\right)^2$$

The coefficient of determination measures the potential for the BPI of a given site *i* to serve as a predictor for the BPI at a different site *j*. Third, we compute the distance between each pair of systems, $S_{i,j}$.

To compute the estimated energy at a given system, each other system's BPI is multiplied by *N*-1 clear-sky insolation values to obtain *N*-1 estimates of local energy. At this stage we consider four different algorithms to combine the *N*-1 estimates into a single energy estimate. We may (a) calculate the arithmetic mean of these *N*-1 estimates; (b) calculate an average weighted by the inter-system distance raised to a negative exponent; (c) multiply the estimates by the Conversion Profile and then calculate an average weighted by the R^2 coefficients; or (d) scale the *N*-1 estimates with the Conversion Profile and then calculate an average weighted by the product of the R^2 value and the inter-system distance raised to a negative exponent.

The method outlined above can be applied to energy and potential insolation summed over periods other than one day. We evaluated each of the four algorithms for energy and insolation summed over periods ranging from 15 minute to one month.

RESULTS

The period at which the predicted generation is compared to the actual generation (aggregation period) depends mostly on the geographical dispersion of the systems and the local climate characteristics. In general, except for clear or overcast days, systems which are 10s to a few 100s of kilometers apart may have significant differences in total output for a given day due to differing weather conditions; there would be even less correlation between the hourly or sub-hourly output.

Because each of the four methods for predicting performance involves averaging, we expect that the predicted energy at hourly or sub-hourly periods will be less variable than the actual energy. Consequently we do not expect our predictions to compare favorably over short time intervals, and we compare instead predictions and observations of total weekly and monthly energy.

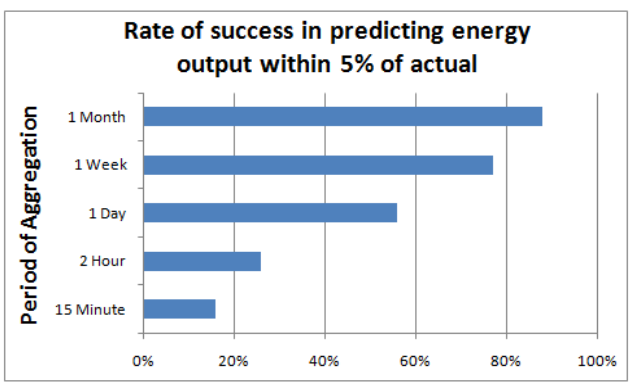

Figure 3: Rate of success in predicting energy output given by the arithmetic mean of *N*-1 energy estimates. Results based on 15-minute measurements and compared over 7 months (May-November 2010).

Aggregating energy over weeks or months will smooth shorter-term temporal variations in the observed energy and produce quantities that may be compared to predictions obtained by averaging. From Figure 3, we see that the algorithm is best suited for accurately predicting energy on the longer time scales.

Two additional critical parameters are the interval at which energy is measured in order to quantify the correlation of historical data and the methodology of establishing the correlation. To determine the optimal choices we compared the accuracy of the four predictive algorithms described in the Methodology section using all the measurement intervals that were shorter than the two aggregation periods of choice. The results are summarized in Figure 4.

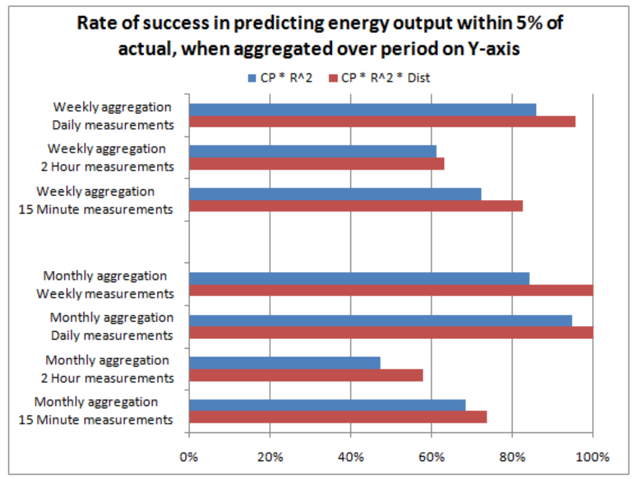

Figure 4: Rate of success in predicting weekly and monthly energy output during June 2010 with two different methods. The R^2 coefficients were calculated based on the intervals noted (15-min, 2-hour, 1-day, 1-week) for the period of May 2010.

From Figure 4 we observe that more accurate estimates are obtained when correlation is determined over daily energy values compared to values from shorter periods. The reduction in accuracy when energy is measured over shorter time intervals may signify that, at that temporal resolution, the correlation between nearby systems is not a high-quality predictor of another system's performance. This finding is consistent with studies of the correlations in irradiance as a function of distance between sensors (e.g., [1]), where the correlations decrease as either distance increases or as time intervals decrease.

Based on the preliminary results presented in Figures 3 and 4 we focused our analysis on the predicted weekly and monthly output, using the mean weighted by both R^2 and distance, where R^2 values are determined from daily total energy.

The results show the rate at which we were able to predict the actual weekly or monthly output of the systems during each month between May and November. The rate of success is associated with an accuracy threshold (±5% or ±10%). A higher success rate corresponds to an increased confidence in the accuracy of the prediction.

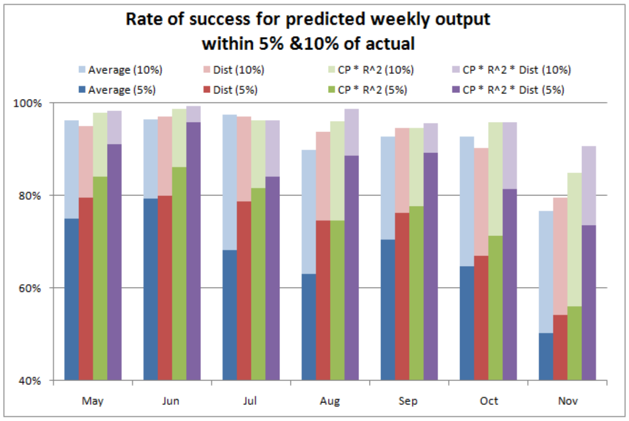

Figure 5: Rates of successfully predicted weekly output within 5% (dark colors) and 10% (light colors) of actual generation. The Conversion Profile and the R^2 values were calculated from daily energy data measured during the preceding month.

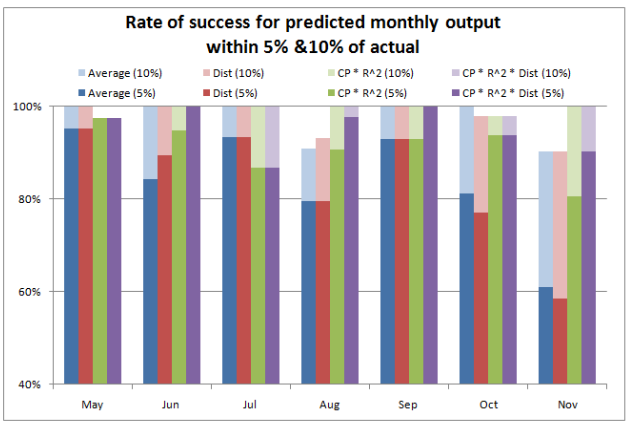

Figure 6: Rates of successfully predicted monthly output within 5% (dark colors) and 10% (light colors) of actual generation. The Conversion Profile and the R^2 values were calculated from daily energy data measured during the preceding month.

DISCUSSION

In Figures 5 and 6 we have presented the effect of increased information on the success of prediction. The results marked as "Average" only rely on the concurrent energy measurements and the system sizes. Each successive series relies on more information: the "Dist" series takes into account inter-system distance information; the "CP * R^2" series takes into account the correlation of the data during clear days (through the Conversion Profile) that occurred in the preceding month, and the correlation of the historical data during any day in the preceding month (through the coefficient of determination); and the "CP * R^2 * Dist" series combines all available historical correlation (from the preceding month) and explicit inter-system distance information.

In general, it is clear that additional information increases the success rate for a given aggregation period and accuracy threshold. However, the weekly output predictions subject to a ±5% accuracy show considerable spread: more than 20% difference exists between the results for June and November (Figure 5). This variation restricts the general confidence in the prediction results to those of the worst case.

From an operational perspective a consistently high success rate allows the owner or operator to be more confident in the accuracy of an alert based on the difference between predicted and actual measurements. For example, with a success rate of 95% at a 10% threshold one can be confident that when the measured energy is less than 90% of the predicted, then the probability of this event being an actual underperformance is better than 19-in-20.

It is important to note that this methodology can be used to identify "partial" outages which may occur due to factors such as extreme soiling or severe degradation at an individual panel. So called "full" outages can be identified immediately by a meter reading that stays constant during daytime hours.

Another interesting result is the relative effect of the individual classes of information. In general, the use of the historical information via the Conversion Profile and the R^2 values provides slightly larger increase in the success rate of the predictions. The similar magnitude of improvement could be attributed to the fact that the historical performance correlation provides information that is implicitly related to distance: it is expected that, all else being equal, nearby systems will exhibit similar performance during periods of sufficient aggregation length because they share similar weather input.

When we include information about both distance and historical data correlation then we may predict the energy output with higher success rates (purple columns in Figures 5 and 6). We attribute this improvement to the stricter selection of systems with similar performance that is afforded by the compounded weighting of the average values.

An obvious suggestion for increasing the success rate of the weekly predictions, which is important from an operation and ownership perspective, is to use a longer historical data record instead of just the month prior to the one of interest. However, our analysis shows that including information from the distant past degrades the quality of the prediction (Figure 7). We opine that the degradation results because the information most relevant to performance during a given month is performance during the previous month. Calculating the Conversion Profile and R^2 over several preceding months will drive these terms toward values representative of average annual performance, rather than values useful to predict month-ahead energy.

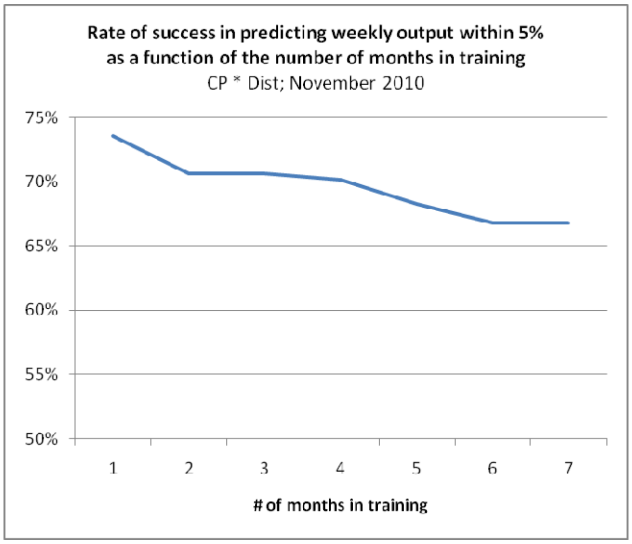

Figure 7: Success rate for weekly output prediction during the month of November. The predictions were based on the distance-weighted Conversion Profile.

Another source of potential improvement lies with the selection of the appropriate exponent of the inter-system distance. By varying the exponent between -0.5 and -4 we see no sharp maximum in the resulting success rates but the best results appear to occur for exponents between -1 and -2. In Figures 5-7 we used an exponent equal to -2

CONCLUSIONS

We have developed a heuristic method that can predict the energy output of a single PV system that belongs to a regional fleet by using information from other systems in that fleet. The input information consists of historical performance correlation and inter-system distance information. The benefit of this method is to provide the capability of monitoring system performance based solely on two sets of information: system location and readings from the generation meter.

We have tested the method on 55 systems from SunEdison's fleet in New Jersey, predicting weekly and monthly energy generation from May to November 2010. The best performing algorithm successfully predicted the weekly generation within 5% of the measured value between 73% and 96% of the time. When the accuracy threshold is relaxed to 10% the success rate varies between 91% and 99%. The monthly generation was predicted successfully at even higher rates – between 85% and 100% for a 5% threshold, and between 96% and 100% for a 10% threshold.

We have determined that the best results are obtained when the correlation of historical data (from both clear and non-clear days) is based on recent daily energy values and weighted with the inverse of the square of the inter-system distance.

ACKNOWLEDGEMENTS

Sandia National Laboratories is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

REFERENCES

[1] A. Mills, R. Wiser, "Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power", LBNL-3884E, Lawrence Berkeley National Laboratory, 2010.