

Solar Energy Disaggregation using Whole-House Consumption Signals

Rahul Mohan, Tim Cheng, Abhay Gupta, Vivek Garud, Ye He
 Bidgely Inc.
 Sunnyvale, CA 94086
 {rahul,tim,abhay,vivek,ye}@bidgely.com

Abstract—This paper presents a novel concept called solar disaggregation. In homes with solar panels, electric utilities are able to meter the *net consumption* power signals, which measures the difference between whole-home consumption and solar production. Solar disaggregation aims to separate, or “disaggregate”, the production of solar panels from a home’s net consumption. We present an algorithm to accurately forecast solar panel production given just net consumption and weather forecast data. The results presented in this paper greatly outperforms prior work in forecasting solar power. In addition, this novel approach can be used by electric utilities at scale, for better demand planning and insights into their solar customers.

I. INTRODUCTION

Renewable energy sources, such as solar, offer many environmental advantages over fossil fuels for electricity generation, but the energy produced by them fluctuate with changing weather conditions. Electrical utility companies need accurate forecasts of solar energy production in order to have the right balance of renewable and fossil fuels available [9]. Errors in the forecast could lead to large expenses for the utility from excess fuel consumption or emergency purchases of electricity from neighboring utilities. Currently, most utilities are only able to obtain net power readings, which is the production from the solar panels subtracted from electric consumption, from a home [5]. However, the inability to differentiate production from consumption leads to uncertainty on how much electricity to put into the grid and unnecessary costs for the utility. This method describes methods to solve exactly this problem. Using hourly interval data for a solar home, we provide the output of that home’s solar panels for a given time period. This allows utilities to forecast solar in their territory, better understand their solar customers, and optimize their costs [10]. We collected solar panel and historical weather data from hundreds of homes and each of their local weather stations. First, as a preprocessing technique, spiky solar production signals are removed by looking at net consumption data as described. Furthermore, a method to predict solar panel generation with a Radial Basis Function (RBF) [6] and Wavelet [7] kernel Support Vector Machine (SVM) [8] ensemble is presented. We demonstrate the superiority and practicability of our solution on real life solar panel data from across across the world, and how this approach can provide significant cost savings to utilities.

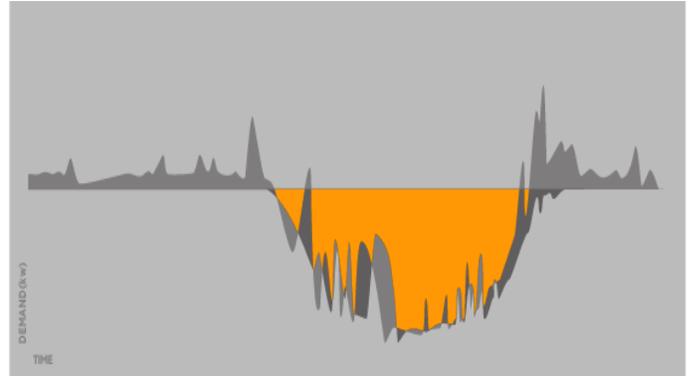


Fig. 1. An example of solar disaggregation

II. METHODS

A. Problem Definition

Figure 2, 3, and 4 describe a situation on sample hourly net power data where disaggregation is not applied, on three different days for the same house during April:

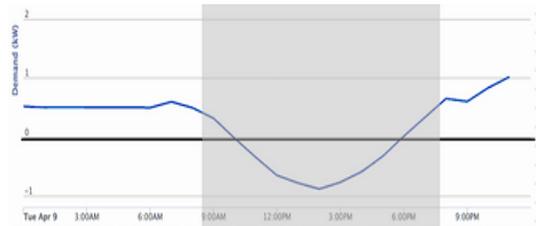


Fig. 2. Day 1: A nice sunny day with excess generation during the day

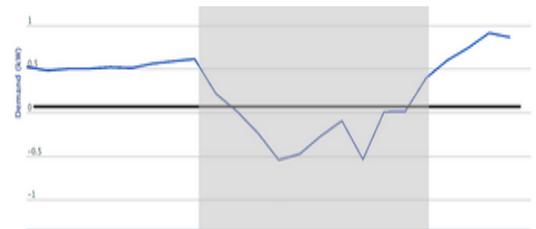


Fig. 3. Day 2: Without disaggregation, it’s not possible to understand whether it was a cloudy day or user was using load during the day



Fig. 4. Day 3: While there was generation, the net power is positive. User was consuming more energy than producing during the day

As seen in the figures above, it is not possible to directly extract PV generation from the net power signal based on “signatures” in the data such as transitions, steady-state features, and harmonics (as in other energy disaggregation tasks). Therefore, a method to *forecast* the PV generation is required.

B. Data Collection

Ground truth PV system generation and net power was collected through CT clamps at one-minute intervals and was provided by Bidgely. Data was collected from 137 homes from all across the world including San Francisco, Boston, New York, Australia, Canada, Kansas, and New Jersey. We used various weather API’s to extract the following weather parameters for each hour: skycover (%), dewpoint (%), temperature (Fahrenheit), windspeed (miles/hour). These API’s were also used to get the sunrise and sunset times for each day.



Fig. 5. A map of the locations where data was collected from

C. Algorithms: Solar Spike Removal

Although, the ground truth PV generation and net power data was originally obtained at the minutely intervals, weather data can only be obtained at hourly intervals. Therefore, in practice, the minutely level data was downsampled to the hourly level which is then used by a forecasting model. However before this, the sudden spikes in noisy one-minute solar signals must be removed as a preprocessing steps. Sudden and short spikes in the data are caused by random fluctuations in the weather which are impossible to capture at one-hour intervals; removing these spikes improve the quality of the downsampled signal and the overall quality of the forecasting model in training and testing. However, “spikes” need to be carefully defined as we need to be careful not to remove appliances, whose transitions can also be “spiky”. In this section, we describe an algorithm to detect and remove spikes

caused by random fluctuations in the weather from solar signals, using the *net power* signal. Net power at the minutely level is used to identify solar spikes. First, net power and solar generation were plotted against each to make sure that random spikes in solar are represented in net. We were able to conclude that the signatures in net are noticeably different when solar is spiky than when solar is smooth and developed an algorithm based on these observations. This is represented in Figure 6 and 7:

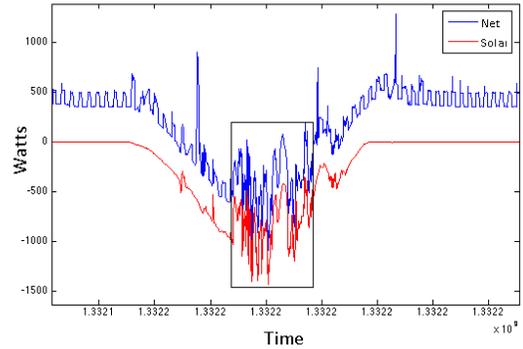


Fig. 6. The spikes in solar are perfectly represented in the shape of the net signal

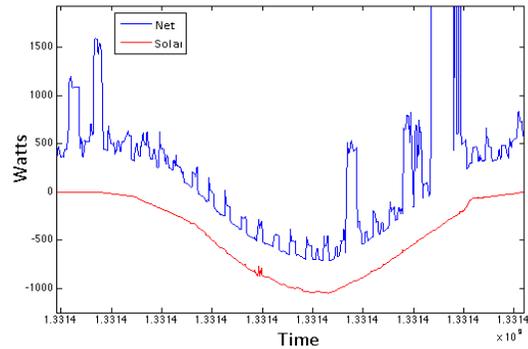


Fig. 7. Here, solar is perfectly smooth. Likewise, instead of random spikes, the net signal is filled with steady-state peaks which represent appliances.

D. Algorithms: Solar Generation Prediction

A regression model was trained with weather and the number of hours from sunrise as the independent variables and solar intensity as the dependent variable. Specifically, we used an ensemble of Radial Basis Function (RBF) Support Vector Machines (SVMs) and Wavelet SVMs which are machine learning algorithms that create highly complex non-linear models and are ideal for solar prediction. RBF [1] and Wavelet [2] SVMs have been proven to perform very well on time series forecasting tasks. The model was further improved through obtaining optimal model parameters, 10-fold cross validation, and regularization. Yearlong data for 106 homes were trained on and the remaining 31 homes were used

for testing. Figure 8 shows a summarized flowchart of the algorithm:

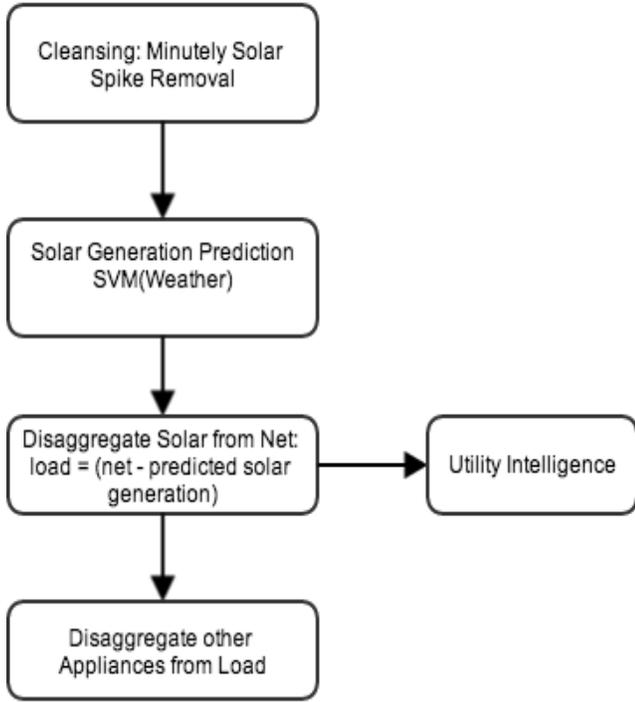


Fig. 8. A flowchart of the overall algorithm

III. EXPERIMENTAL RESULTS

We achieved solar prediction accuracy higher than the reported accuracies from previous work, despite training and testing on different homes from across the world. The previous work [3] [4] have trained and tested their algorithms on the same home. While not only achieving better accuracy, our model is applicable to other places than where it was trained. Table 1, Figure 9, Figure 10, Figure 11, Figure 12, and Figure 13 summarize our results:

TABLE I. RESULTS

Method	Error (%)	RMS Error (watts/m ²)
This Paper	5.5	70
Sharma et. al [3]	15.88	127
Chakraborty et. al [4]	8.13	101

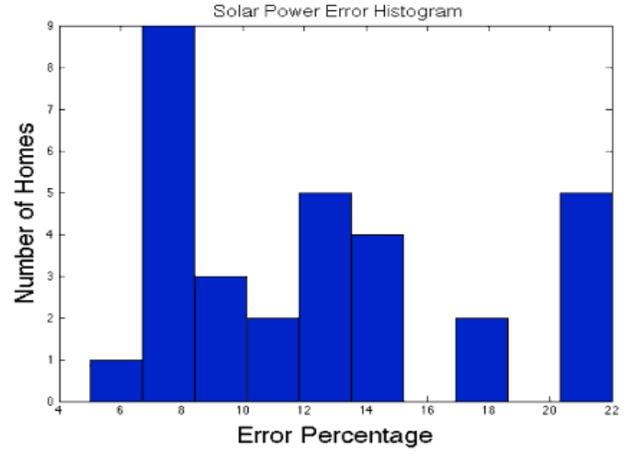


Fig. 9. A histogram of the prediction error percentages of our model

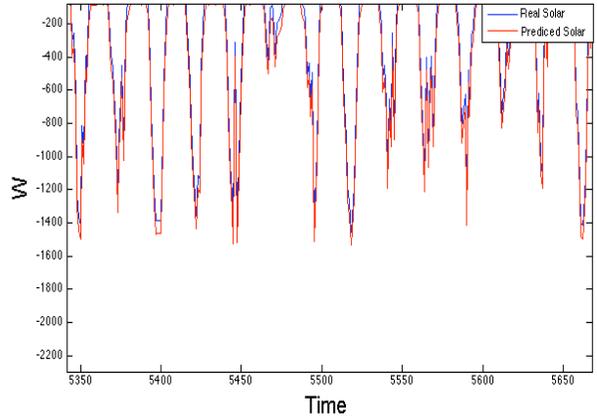


Fig. 10. Plot 1: Actual vs Predicted Solar Output from one of the homes

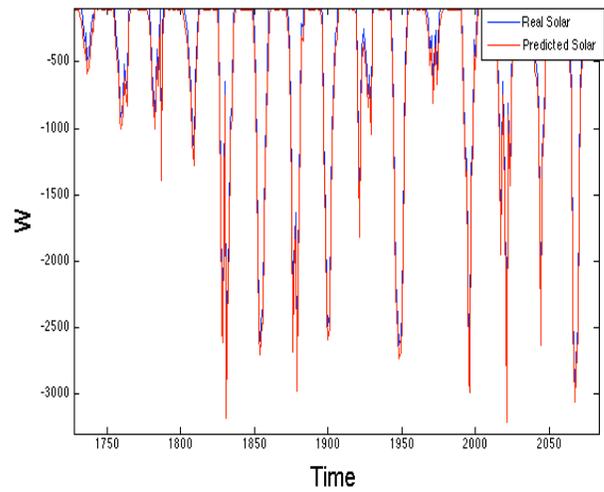


Fig. 11. Plot 2: Actual vs Predicted Solar Output from one of the homes

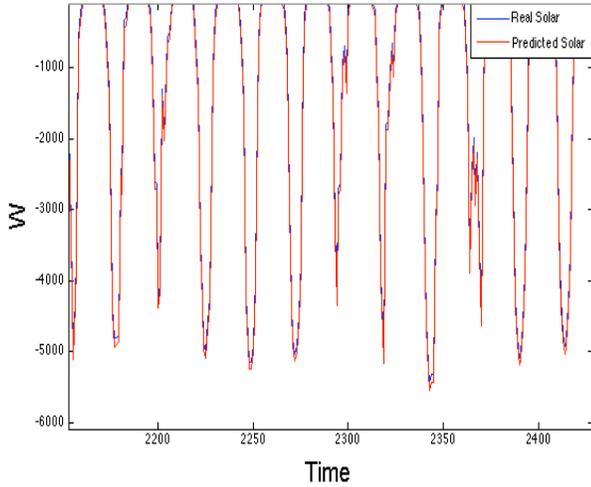


Fig. 12. Plot 3: Actual vs Predicted Solar Output from one of the homes

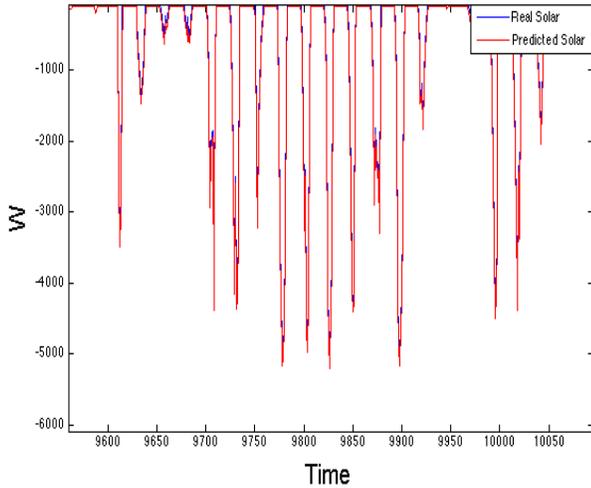


Fig. 13. Plot 4: Actual vs Predicted Solar Output from one of the homes

Our model is more accurate than the bayesian ensemble approach used by Chakraborty et. al by approximately 2.88%. Their approach was tested on one building and contained 267 days of data. Our model was tested on about 31 years of data points (31 homes).

IV. CONCLUSION

In this paper, we presented solar disaggregation, a new approach to extract solar power generation from whole-home net power signals. We demonstrated a RBF and Wavelet SVM based forecasting model to predict solar generation using various weather parameters. Our results show that this algorithm is more accurate and practical than previous approaches to forecast solar power. Utilities can harness this model to increase their intelligence and save costs by having an accurate estimate on

the amount of electricity needed to put into the grid.

V. REFERENCES

1. J.M. Hutchinson: A Radial Basis Function Approach to Financial Time Series Analysis. MIT AI. (1994).
2. H. Chao, H. Li, H. Ting: Financial time series forecasting based on wavelet kernel support vector machine. IEEE Transactions, ICNC. (2012).
3. N. Sharma, P. Sharma, D. Irwin, and P. Shenoy: Predicting Solar Generation from Weather Forecasts Using Machine Learning. SmartGridComm. (2012).
4. P. Chakraborty, M. Marwah, M. Arlitt, N. Ramakrishnan: Fine-grained Photovoltaic Output Prediction using a Bayesian Ensemble. AAAI. (2012).
5. D.R. Baker: Solar “net metering” extended by California regulators. SFGate. (2014).
6. K.K. Chin: Using a Radial Basis Function as kernel. (1998).
7. L. Zhang, W. Zhou, and L. Jiao: Wavelet Support Vector Machine. IEEE Transactions Volume 34. (2004).
8. V.Vapnik: The Nature of Statistical Learning Theory. Springer. (1995).
9. J. Postelwait: Solar Forecasting: The Next Big Thing For Solar Power?. UtilityPower Volume 16, Issue 1. (2013).
10. M.D. Lemonick: Solar Forecasts Could Help Electric Utilities and Climate. ClimateCentral. (2013).