Analysis of Profitable Locations for PV Generation in the U.S.A.

Cristian Gómez Peces

***Abstract*—This paper focuses on the analysis of different electricity markets in order to determine whether a potential solar panel farm installation can be profitable. Several datasets from ISOs are taken into consideration to extract market spot prices. In addition, daily solar irradiance profiles are extracted from NREL's database. With these data and some economic assumptions, several simulations are run to estimated potential gains. The purpose of this study is therefore to determine the regions that are attractive to potential solar panel investors that seek domestic solar PV investment opportunities.**

***Index Terms*—PV, Electricity Market, ISO, LCOE, GHI, LMP**

I. Introduction

As solar power is reaching competitive levels with other energy sources in terms of cost, it has become an attractive option of investment not only for the purpose of pushing power systems to be more sustainable but also to actually earn money [1]. Not long ago, large PV installation projects had to be subsidized by public authorities to take off [2], but as large economy scale lands to the silicon-based solar panels industry, more entities are considering the possibility of installing large solar farms [3]. One of the challenges that investors need to deal with is the location where to install the solar panel farm. In the United States, California is chosen [4] for most of the national PV projects due to its high irradiance ratio throughout the year and its electricity prices, one of the highest on average in the country [5].

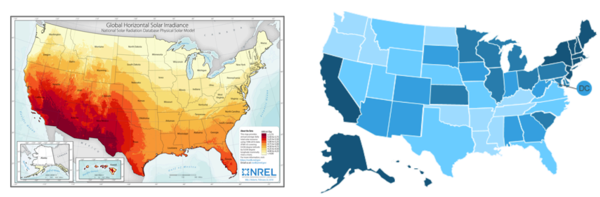
In addition, during the last two decades many other countries have decided to jump on board the boat of solar energy. In 2016, it was announced the largest investment to build a solar panel farm in Dubai [6]. In Europe, Germany has been pushing consistently to move towards a more sustainable power system for two decades, and the number of solar panels installed at the distribution level is increasing at a high rate thanks to the high prices that Germans pay for a kWh, which leave more room for making a profit or saving money compared to the rest of Europeans [7]. Middle East and Asia-Pacific are also areas with great potential ([8], [6]), where investors can find interesting opportunities in a rapid-growing PV market.

There are several variables to take into consideration before choosing a region to build a solar panel farm. Many factors such as socio-political situations may affect to an investment decision, but for the purpose of this study, we will focus on two essential input variables: daily solar irradiance profile based on the area location, and market price of electricity throughout the day. Hence, one should note that some relevant aspects such as public incentives, political stability, local costs, and government intervention are not taken into account in this study. Batteries connected to the solar farms are not considered either, thus power produced should be injected and sold at generation time.

This project constitutes a first approach to determine potentially profitable regions within the United States to install large farms of PV panels, and it is oriented to entities interested in making domestic PV investments. The paper attempts to cover as many territories as possible to provide a wider scope of the available opportunities, as long as electricity market prices are transparent, and power systems' related data are available and disclosed to the public.

II. Background

The PV penetration on the generation mix is taking off nationwide, but at different growth rates depending upon the region. Today, California is leading the expansion of PV systems across the country, constituting the largest PV market among the rest of states, with an installed capacity of nearly 13 GW [9]. The solar energy market growth within the U.S. has been leaded as well by other states such as Arizona, Hawaii, Nevada, New Jersey and Massachusetts, according to the Solar Energy Industries Association [12].



(a) Average Solar Irradiation (b) Average Electricity Prices

Fig. 1. States overview ([3],[10])

There is a correlation between PV power potential based on direct normal irradiation and potential economic gains: private agents are willing to invest because they expect economic gains from it. So in order to estimate projected cash flows, one should not only consider the degree of irradiation over the area but also the electricity market conditions where the PV panel will be installed, since it is ultimately linked to the price at which we can sell the power that is generated. The main goal of this project consists of identifying those markets in which we can obtain highest rates of return. From Figure 1a we extract that the southwestern states would be able to generate a large amount of solar energy based on the solar irradiation, whereas Figure 1b shows the regional electricity prices. Note that in this map there are several hotspots that are localized around California and northeastern states. This suggests that exploring the installation of PV panels in the northeast can be interesting due to its high prices.

In order to explore which areas could be more profitable, we first need to identity potential costs to consider a financial model. These costs are highly dependent of multiple factors –such as location, size of installation, etc.–, and therefore we use the Levelized Cost of Energy (LCOE) in order to evaluate the cost, and ultimately the rate of return of the solar panel farms. The National Renewable Energy Laboratory provides a very detailed model to evaluate LCOE in the United States ([3], [11]). We will to use this methodology and extrapolate it to other locations that are not presented in the mentioned study.

III. Proposed Method

A. Objective

The main goal of this study is to create a metric that let us compare different locations in terms of potential profitability and determine which location could be more suitable to install PV systems under a certain criterion.

The output of the model will be the ratio of benefit-cost for each of hubs analyzed (every state will have a certain number of hubs with specific marginal prices and irradiation data). For each area, it has been collected the GHI (Global Horizontal Solar Irradiance) profile, LMP (locational marginal prices) and LCOE of the state associated to that region. The model will provide the benefit-cost ratio for each area, so every location is easily comparable to others. In the results section, a table is provided with the most profitable regions to install a PV farm, according to following model.

B. Model

The model takes several inputs for each hub, as shown in Figure 2. The locational marginal prices are a set of arrays with the information along 2018, broken down for every hour of the year. The GHI is also a set of time-series data that is used to calculate the estimated power injection. For each location, it has been assumed a PV size of 1 hectare and PV efficiency of 20%. It is also assumed a solar panel lifespan of 25 years with no degradation rate. Then, a 25-years simulation is performed, taking the data from 2018 and applying random deviations and estimated trends simulated period. The LCOE has also been adjusted yearly according to the data from each state.

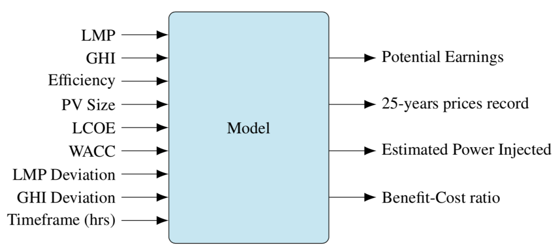


Fig. 2. Model

For every yearly iteration, the total generation and earnings are calculated as the integral of the product of the estimated power injection (P) and the price at which it will be sold at that hub (π). Since the LCOE calculation includes the time value of money, we need to discount every calculation of earnings to get its present value. Hence, the final calculation of the total earning (Ω) results in:

$$\Omega = \sum_{i=0}^{24} \frac{\int_0^H P_i(h) \cdot \pi_i(h)\, dh}{(1 + WACC)^i} \qquad (1)$$

Note that there are 25 summations (from 0 to 24) so the first term present value is the integral itself. In addition, note that the integral covers all the hourly domain of the year, H, adding up a total of 8760 hours. The Weighted Average Cost of Capital (WACC) assumed for this model equals to the treasury bond rate. Within the integral term, we should point out that the integration method applied for this discrete case is the trapezoidal rule, and that $P_i(h)$ and $\pi_i(h)$ correspond to the simulated values of the power injected and the price at which electricity would be sold at year i, respectively. For the power injection, we assume a hectare for all locations, with an efficiency of 20% and power deviation of 10%. The GHI value is provided by NREL database in units W/m^2. Hence, we can compute the injected power as follows:

$$P_i(h) = GHI(h) \cdot A_{PV} \cdot \eta \cdot (1 + \gamma_i) \qquad (2)$$

Where A_{PV} and η correspond to the panel size in m^2 and its efficiency. Both parameters are considered fixed and equal to 1 hectare and 20%, respectively. The coefficient γ_i follows a normal distribution with mean equal to the percentual region's irradiation trend and has a 10% of standard deviation at that year.

For $\pi_i(h)$, we use the locational marginal price that we gathered from the ISO's data, applying a different deviation coefficient for each year:

$$\pi_i(h) = LMP(h) \cdot (1 + \upsilon_i) \qquad (3)$$

Where the coefficient υ_i follows a normal distribution with mean equal to the percentual region's price trend and has a 10% of standard deviation at that year.

C. Methodology

The project simulation is developed in Python, using Pandas, Numpy and Plotly to study approximately 100 million data on hourly market prices and 2 million data on hourly irradiation profiles, gathered for 22 states from different ISOs and NREL databases, respectively. When available, market prices from a hub zone were taken into the model (that is, the average price for a state's region), otherwise the prices for certain load zones of that state are taken into account for the model. The work frame used to process all these data is shown in Figure 3.

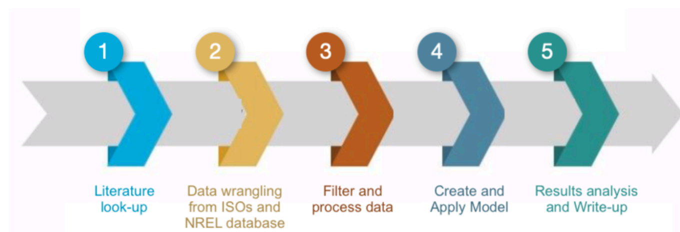


Fig. 3. Project stages

The first step to accomplish is the gathering of useful data. In order to do that, several ISOs databases have been consulted, including CAISO, MISO, PJM, ERCOT, NYISO and ISO-NE. Due to the interesting characteristics gathered from the literature and the relative easy access to data in the northeastern ISOs, they were the starting-point databases for this project. The data extracted corresponds to the spot market records stored in the ISOs servers. The irradiance profiles are obtained from NREL's database by selecting specific locations in the interactive map that is available online. Based on the hub's location, we pick data from the closest location and later on the model matches up hub and locational irradiance.

Once we get all the data for the 365 days of 2018, we filter it and obtain the hourly locational marginal prices for those hubs we are interested in and for all the 8760 hours that compound a year. We do the same for the irradiance profiles, and these files are saved in new files that will be accessed by the model.

The next stage involves the creation of the model and its application over all the data that was wrangled and filtered in the previous steps. With the model, it is straight-forward to plot the earnings evolution along 2018 and also compare the simulations along the 25 years of lifespan, check which hubs have higher LMPs, etc. All these possibilities are explored in the final stage, the results analysis, where insightful plots are analyzed and differences between regions are compared. This analysis, in addition to the inferred conclusions is discussed in the last section.

IV. System Study

A. Target curves

Figure 4 shows the final result obtained: the potential earnings curves over a certain week of 2018 for the hubs analyzed in New York state. Note that the earnings depend on the solar irradiance as no earnings can be made during nighttime. Each curve represents a load zone during a certain hourly period out of the 8760 hours that compound a year.

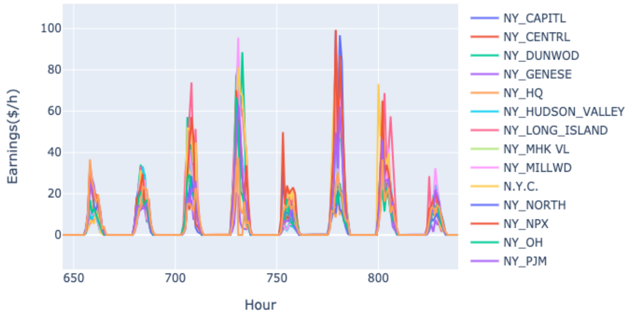


Fig. 4. Simulated Earnings for NY ISO areas

We compute the same curves for the 25 years for each of the load zones. Hence, after the simulation is done, we have all the earnings accumulated along the years, in addition to other variables such as yearly irradiation, price volatility, etc. that we can record to perform analysis.

B. Geographical Results

We use the B/C ratio to evaluate load zones profitability. From the curves presented in part A and the cost assumptions presented in the model section, we evaluate the ratio for all the hubs and display it as bubbles in a national map to see clearly regional performances, as shown in Figure 5.

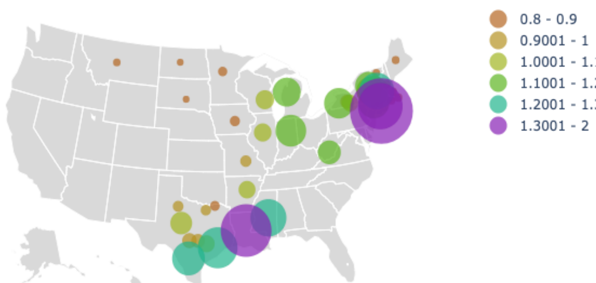


Fig. 5. Estimated B/C ratio locations across the U.S.

As we can observe, there are two main hotspots that output higher ratios. Some regions in the north-east, in particular those close to New York City present higher profits due to the high prices at which the power can be sold. The other hotspot is located in the south, in particular the areas of south of Texas, Houston and the southeastern states of Louisiana and Mississippi. In contrast, most of the hubs under New England ISO control present lower potential earnings. The majority of states located in the middle / north of the country present low B/C ratios that make them appeal less attractive for investments. Lastly, it should be noticed that although the northeastern regions of MISO do not present the highest ratios, some hubs such as Michigan or Indiana present attractive ratios of 1.16 and 1.19, respectively.

C. B/C ratio reliance

As expected, the B/C ratio increases with the overall locational marginal average price. In other words, hubs with higher prices tend to have higher B/C ratios. However, the profitability does not follow a simple positive correlation in case of the price variability and the irradiation. Hence, we plot the bubble chart shown on Figure 6 to explore possible relations between variables. The X axis shows the mean irradiation, and Y axis shows the price volatility at that hub. The size of the bubble represents the value of the B/C ratio.

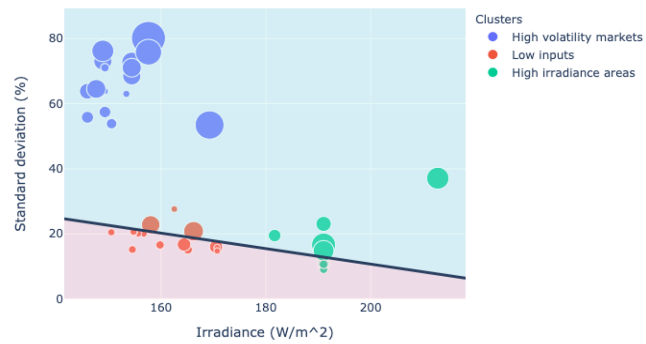


Fig. 6. B/C ratio as function of price volatility and solar irradiance

We observe that the resulting data points distribute themselves in a very particular manner: as clusters. Hence, we applied a K-means clustering algorithm to divide the data points into groups to classify them according to its irradiance and volatility. The elbow rule suggests that the appropriate number of clusters is 3. Therefore, we divide the data points into 3 groups, shown in blue, green and red colors, which represent high volatility prices hubs, areas with high incidence of solar radiation, and hubs with low price volatility and low irradiance, respectively. The first group includes all areas of the NY ISO plus the state of Louisiana, which is the point with higher irradiance than the rest. The second group comprises the NE ISO areas and most of the middle west of the US. Michigan and Indiana stand out from the rest with relative higher profitability. The last group includes all the Texas areas, Arkansas and Mississippi. Mississippi is the state positioned with highest irradiance index.

We notice that the B/C ratio seems to be more sensitive to the price volatility than the mean irradiance. As we want to determine whether an area will be profitable or not, this is a machine learning classification problem. Several classification models have been developed to extract insights for the B/C ratio sensitivities: Logistic Regression, K-nearest Neighbors, Support Vector Machine, Naïve-Bayes and Random Forest Classification. The data was divided randomly into train and test sets over a thousand times in which the classification models were computed. Later on, the confusion matrices were extracted and both False Positive and False Negatives were taken into account to evaluate the performance of the algorithm. Table I shows the final performance that was computed for the 5 algorithms.

TABLE I. Classification Perfomances

Method	***Logistic Regression***	***K-NN***	***SVM***	***N-B***	***Random Forest***
Accuracy	69.23%	63.5%	67.9%	67.14%	68.10%

In most of the iterations, Random Forest and Logistic Regression obtain a higher accuracy than the rest. However, in many cases the Random Forest Classification presents overfitting, leading to a great performance for those data point within the training set but not for the test set. Hence, Logistic Regression was chosen to classify the profitability of the hubs. Note that our goal is not to obtain a very accurate model, but to extract qualitative insights from the classification.

The regression line is drawn as well on Figure 6 on black color. This line divides the chart into 2 areas. The magenta-colored space represents the nodes classified as non-profitable and the turquoise-colored space the profitable ones. As can be observed, price volatilities have a greater impact on the B/C ratio than the average irradiance. That is, the profitability of PV panels is more sensitive upon the electricity market conditions than the average irradiance of that area, since the regression line shows a significant greater sensitivity for the price volatility.

D. *Project Sunroof Study*

In order to set a reference to evaluate this study, the results gathered from the Google Sunroof project were compared to the obtained results. These data take into account potential roofs that can be used to generate PV power. Figure 8 shows the areas with data available.

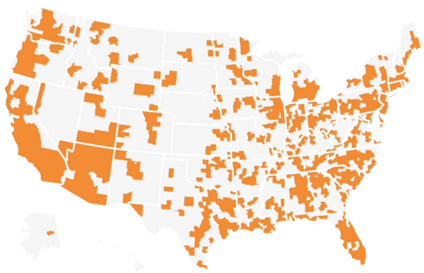


Fig. 7. Regions with data available from Project Sunroof

Every state that we took into account for the study has data available to extract, extrapolate and compare results. Google servers are provided with weather statistics from NREL, Utility electricity rates information from Clear Power Research and the imagery, and 3D modeling and shade calculations from Google. Hence, this seems appropriate to evaluate our model in terms of power generation estimations.

Figure 9 shows the error $\varepsilon = \frac{P_{sunroof}-P_{model}}{P_{model}}$, when we compare our approach with Sunroof's project results. Following Google's methodology in [13], we estimated the power generated from a 1 hectare PV farm based on the number of total panels, yearly sunlight energy, percent of qualified rooftops, DC to AC derate factor and a size of 1.650m x 0.992m for each of the panels accounted in the metadata. As observe in Figure 9, Google's study tends to be more optimistic in terms of power production.

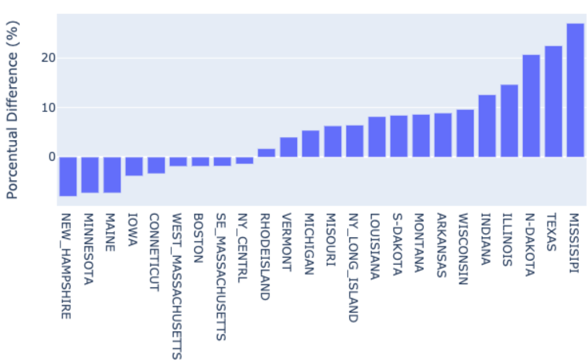


Fig. 8. Porcentual difference in generation estimation

V. Discussion

We observe an overall positive bias of 5.34% between resulting generation when comparing to the Sunroof project study. However, this difference is unequally distributed over the geography of the country. Whereas the New York load zones present an error close to zero, we find that some states in New England were slightly overestimated in terms of power generation, and hubs located in the south such as Texas or Mississippi were greatly underestimated, leading to errors higher than 20%. The rest of hubs located in the middle of country would also expect a higher production of power, but within an error range that varies from 5% to 10%.

The overall positive bias can be explained due to several reasons. First, Google uses a combination of DNI and DHI irradiation indexes to estimate power generation, which are usually higher than GHI. This difference would impact greatly on the generation estimation, especially in those areas of continuous sunlight exposure, such as Texas or Mississippi. Other parameters such as the number of qualified rooftops (sometimes, project Sunroof mistakenly calculates viability for large objects that are not buildings, such as bridges) or the fact that we did not take into account the orientation of the PV panel may have also contributed to the aforementioned bias.

VI. Conclusions

This study has presented a new methodology to evaluate the profitability for PV generation at different locations. The model requires to make some economic assumptions and to be provided with data from electricity market prices and weather statistics from NREL. Hourly locational marginal prices data have been extracted, processed and fit to the model from different ISOs: ERCOT, MISO, NYISO and NEISO. The results obtained from fitting the model with multiple simulations data has yielded to several insights. First, 2 main hotspots have been identified: part of the north-east and some regions in the south of the country. Secondly, 3 different clusters have been identified: high volatility prices hubs, areas with high incidence of solar radiation and hubs with low price volatility and low irradiance. The sensitivities of the profitability for hubs was also explored by performing a Logistic Regression classification, leading to the conclusion that electricity market conditions have a greater impact on the B/C ratio than the solar irradiance. Lastly, the model was compared to Project Sunroof of Google. We find that our approach is more conservative in terms of power generation estimation and we also discussed the reasons that would explain the difference between both methodologies.

References

[1] Rahmatallah Poudineh," Liberalized retail electricity markets", OIES Paper: EL38, Dec. 2019

[2] Reinhard Haas," Market deployment strategies for photovoltaics: an international review", Volume 7, Issue 4, August 2003, Pages 271-315

[3] Ran Fu, David Feldman, Robert Margolis, Mike Woodhouse, and Kristen Ardani, "U.S. Solar Photovoltaic System Cost Benchmark: Q1 2017", National Renewable Energy Laboratory, NREL/TP-6A20-68925 September 2017

[4] Vargas, Carlos and Chesney, Marc," What are You Waiting to Invest? Long-Term Investment in Grid-Connected Residential Solar Energy in California. A Real Options Analysis." May 24, 2019.

[5] U.S. Energy Information Administration; Electric Power Monthly: February 2019, "2018 Average Electricity Retail Prices". Published online at 'https://www.eia.gov/electricity/wholesale/' [Online Resource]

[6] Moritz Borgmann," Dubai Shatters all Records for Cost of Solar with Earth's Largest Solar Power Plant". Published online at 'https://www.apricum-group.com/' [Online Resource]

[7]" Global electricity prices in 2018, by select country". Published online at 'https://www.statista.com/' [Online Resource]

[8] Middle East Solar Industry Association," Solar Outlook Report 2020", January 2020.

[9] "Solar PV Electricity Production, Capacity". Published online at 'https://ww2.energy.ca.gov/almanac/renewables_data/solar/index_cms.php' [Online Resource]

[10] U.S. Energy Information Administration. "2020 Archive", Published online at eia.org. Retrieved from: 'https://www.eia.gov/todayinenergy/detail.php?id=34932#' [Online Resource]

[11] US Department of Energy - Office of Indian Energy" Levelized Cost of Energy (LCOE)". Published online at'https://www.eia.gov/electricity/monthly/' [Online Resource]

[12] 'Top 10 Solar Energy States Per Capita (US)' Published online at 'https://cleantechnica.com/2014/09/03/top-10-solar-energy-states-per-capita-us/.' [Online Resource]

[13] "Project Sunroof data explorer: a description of methodology and inputs". May 2017. Published online at 'https://www.google.com/get/ sunroof/data-explorer/data-explorer-methodology.pdf' [Online Resource]