Solar Panel Location Selection Using Power Generation Potential Prediction

Andrew Xavier: ahx2001

December 16, 2022

Abstract

As greenhouse gases from using fossil fuels accelerate global warming and climate change, renewable sources of energy have become a necessity for a sustainable future. A popular source of renewable energy is solar radiation, typically captured using solar cells to generate electricity. The power generated by these solar panels is affected by several environmental and innate cell features. While the innate cell features are determined by the manufacturer, the environmental variables vary over time and vary by location. Global Horizontal Irradiance (GHI) and cell temperature are two features that vary significantly in the environment and that drastically affect how solar panels perform. This project looks to help provide a simplified method to approximate the potential power output over historical data using GHI and ambient temperature measurements, predict the future power output for years in the future, and help determine the best locations to build solar panels.

1 Introduction

Energy generation and consumption is the largest source of greenhouse gas emissions, accounting for 75.6% of all human causes greenhouse gas emissions Ge et al. (2020). In order for countries to abide by the Paris Climate Agree par (2015), they need to reduce our emissions by 45% in the next decade and reach net 0 by 2050. To reach this goal, nations have moved towards using renewable sources of energy given energy generation's massive contribution to greenhouse gas emissions. Solar panels are a powerful tool to capture and transform solar energy for human use. Solar panels are available for commercial use in the form of large-scale farms, such as California's Solar Star Wesoff (2015), as well as for private use by smaller institutions and household augmentation. The power generated from these solar panels, however, is determined by many system and environmental factors. While manufacturers have been able to improve the internal cell capabilities, the generated power still largely depends on factors such as diffuse sky radiation intensity, cloud cover, ambient temperature, daylight hours, snow cover, shading, and other weather features Hasan et al. (2021). This potential power output can, however, be approximated using Global Horizontal Irradiation (GHI) and ambient temperature as derived in this paper.

GHI is defined as the total amount of shortwave terrestrial irradiance received by a surface horizontal to the ground Lorenz and Heinemann while the ambient temperature is the temperature of the air surrounding a device, in this case, solar panels. A visualization taken from Lilienthal et al. is shown below in Figure 1.

Solar panels function best in locations with high GHI and low ambient temperatures Teqo (2022). This is because solar panel cells work best when their cell temperature is lower due to the voltage

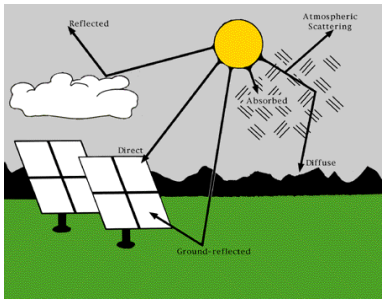

Figure 1: GHI depiction from Homer. Source: https://www.homerenergy.com/products/pro/docs/3.11/global_horizontal_irradiance_ghi.html

difference from electrons activation from solar radiation Sunrun (2021) and colder ambient temperatures help transfer heat from the cells. However, though increased GHI levels increase the power available for capture by solar panels, this increased solar radiation also increases cell temperature above the ambient temperature and thus decreases cell efficiency Tartakovsky et al. (2018). In order to identify good locations to build a solar panel array, one needs to consider the thermodynamics of the system as those interactions determine the final power output. This project develops and demonstrates how to use GHI and ambient temperature for this purpose. Finally, climate change makes the environmental factors that determine solar panel PV output more volatile Crimmins et al. (2016) and makes long-term prediction of solar energy generation essential for creating reliable systems. This project develops a power output approximation by building a highly accurate photovoltaic (PV) output dataset, using the granular data from the National Solar Radiation Data Base (NSRDB) Sengupta et al. (2018) and thermodynamic models from HOMER pro 3.15 Lilienthal et al., to train predictive models for monthly power output for future years.

2 Related Work

Future prediction for solar power generation has been explored from different angles. One common approach is to predict solar radiation as an approximation as shown in Sharma et al. (2011) and Kim (2019). However as noted before, radiation is not the only factor in power generation. Additionally, these predictions are short-sighted and only predict hours or days into the future which is not useful for long-term planning. However, these approaches do find success with variations of linear regression models which we will use in our work as a starting point. Using existing solar panel measurements to help predict future output is possible as shown in siv. However, these methods do not guarantee generalization as only a small number of sites can be used to collect data as setting up additional sites is expensive. The lack of variety in existing solar panel locations is a motivation for using satellite data in this paper. This is also an issue with the National Renewable Energy Laboratory's (NREL) WATTs calculator and their powerful System Advisor Model (SAM) Blair et al.. The online calculator cannot generalize for locations without existing addresses, severely limiting the variety of locations that can be analyzed. The SAM model is incredibly powerful and not dissimilar to the approach in this project but requires precise information about the environment, some of which is not accessible from satellite data such as soiling which limits the areas SAM can be applied to precisely. Still, within the USA, SAM is a good tool for estimating power outputs for residential areas and will be used to validate the approximations in this project.

Year	Month	Day	Hour	Minute	GHI	Temperature	Relative Humidity	Wind Speed
2010	1	1	0	0	0	15	79.37	2.3
2010	1	1	0	30	0	15	79.35	2.5
2010	1	1	1	0	0	15	78.67	2.7
2010	1	1	1	30	0	15	78.65	3.0
2010	1	1	2	0	0	15	77.03	3.3

Table 1: Example data from Los Angeles (33.2164, -118.2437) for the year 2010

3 Data

The data used for this project are the solar resource maps and geospatial data from NSRDB Sengupta et al. (2018). The solar resource maps are in a Tag Image File Format (.tif) and thus require the osgeo python library to process GDAL/OGR contributors (2022). The maps are GHI averages over the months in 2018 and the GHI average over the year. The maps can visualize GHI levels in the Americas to help select promising points for solar panel placement as GHI levels have a significant impact on the PV wattage. An example is shown below in Figure 2 after being scaled to range from 0 to 255 for visualization within graph parameters.

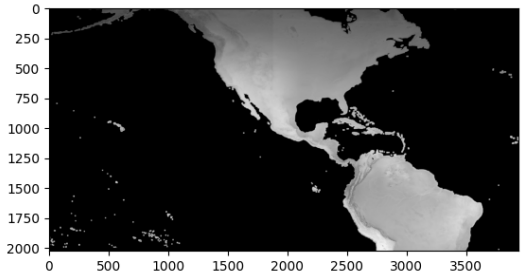

Figure 2: GHI average levels in the upper Americas in 2018 scaled to the range of display output. Lighter is greater GHI

The geospacial data is fetched using the NSRDB API https://developer.nrel.gov/. This data contains various environmental features given the coordinates for the location as well as the year in question. These features include but are not limited to Direct Normal Irradiance (DNI), Diffused Horizontal Irradiance (DHI), dew point, pressure, wind speed, GHI, and ambient temperature. These measurements have a temporal resolution of 30 minutes and a spatial resolution of 4km. For example, using the coordinates for Los Angeles at 33.2164, -118.2437, we can look at the wind speed, relative humidity, ambient temperature, and GHI levels every 30 minutes throughout the year 2010 as shown below in Table 1 and Figure 3.

In this project, I use locations ranging between latitude values ranging from 20.0000 to 27.0000 and longitude values set to -104.0000. The reasoning for these locations is explained in "Map Visualization and Location Selection" 4.1. As the spatial resolution is 4km or about 4.5% of a change in longitude, I space my points by 0.045 degrees in longitude starting at 20.0000 and increasing until 26.9750 while keeping latitude set at -104.0000. This results in 156 locations. In each location, I gather all the available historical data at this time which ranges from the years 1998 to 2020. Data for 2021 is not available at this time. My process results in a total of 62,861,760 data points, each containing temporal and spatial information as well as GHI and ambient temperature. I then leverage the process described in 4.2 to determine an approximation of PV output as well for each 30-minute period. This data is then aggregated down temporally using the process described in 4.2 to 43,057 data points that are used in training the predictive model. This data is then processed using various train-test splits along

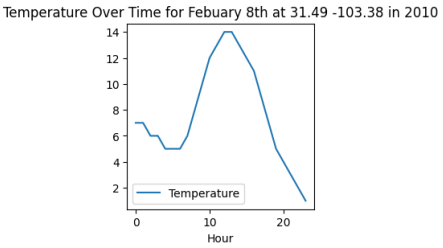

Figure 3: Temperatures over time example

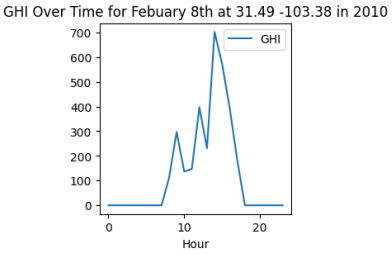

Figure 4: GHI over time example

with different parameters for the time used to predict future values and how far in the future the model should predict.

4 Methods

4.1 Map Visualization and Location Selection

NREL's NSRDB database is massive Sengupta et al. (2018) making virtually any location a candidate for analysis. To reduce my search space, I initially filter locations based on their history of GHI levels. I generate a visual map of GHI averages and highlight locations that have historically high average GHI levels. For this project, I threshold the data to highlight locations with only the top 20% of average GHI level for the USA and America's data set as shown below in Figure 5.

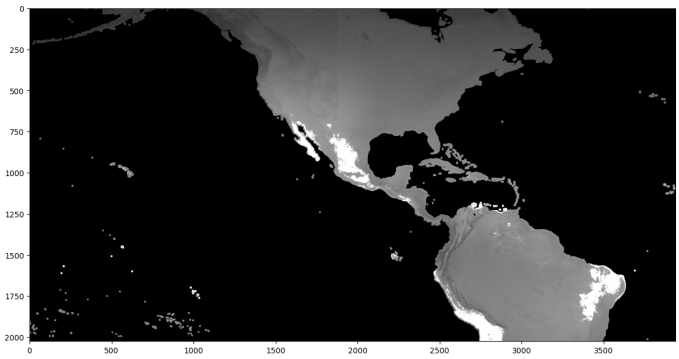

Figure 5: GHI average levels in the upper Americas in 2018 with locations within the top 20% being highlighted in white.

The highlighted central west region of Mexico is a notable location as shown above. The top (northernmost) region of this area has an approximate longitude value of 27.0000 and the lower (southernmost) region has am approximate longitude value of 20.0000. The region's latitude ranges from -104.00 (westernmost) to -100.000 (easternmost) and the region's longitude ranges from 20.0000 to 27.0000. To gather data along the west side where there is a full range in longitude, the sampled data is isolated between lat/lon -104.0000, 20.0000 and -104.000, 27.0000 with a spatial difference of 0.045 in longitude between each point to account for the 4km resolution (4.5% difference in longitude) in the NSRDB's data set. The final result of this process is 156 notable locations (in the top 20% of average annual GHI level in 2018) to gather data on in the next steps.

4.2 Data Processing

Given the GHI and ambient temperature, I look to approximate the potential PV output for a poly-crystalline silicon solar panel cell with a 400W rated capacity. An example of this at the time of writing this report is the "400-Watt Polycrystalline OffGrid Solar Power Kit from HomeDepot" https://tinyurl.com/mr2em98j. In 2022, this is industry standard Ost (2022).

I start with the true equation for calculating PV output from Homer Lilienthal et al. as given below.

$$P = Y_{pv} f_{pv} (\frac{G_T}{G_{T,STC}})[1 + \alpha * (T_c - T_{c,STC})] \quad (1)$$

Y_{pv} = the rated capacity of the PV array, power output under standard test conditions [kW]

f_{pv} = the PV derating factor [%]

G_T = the solar radiation incident on the PV array in the current time step [kW/m2].

$G_{T,STC}$ = the incident radiation at standard test conditions [1 kW/m2]

α = the temperature coefficient of power [%/°C]

T_c = the PV cell temperature in the current time step [°C]

$T_{c,STC}$ = the PV cell temperature under standard test conditions [25°C]

Our next step is to turn this into a function that only depends on GHI levels and ambient temperatures. First, in our case, Y_{pv} is 400W as that is the rating of our panels. $G_{T,STC}$ is given as $1kW/m2$ and $T_{c,STC}$ is given as 25°C. The temperature coefficient of power (α) for a Poly-silicon panel is -0.48[%/°C] Lilienthal et al.. This leaves the derating factor (f_{pv}), the solar radiation incident on the PV array (G_T), and the cell temperature (T_c).

4.2.1 Derating Factor

The derating factor (f_{pv}) is a scaling factor that determines the final real-world output of the solar cell after adjusting for environmental conditions. Note that the largest factor, temperature, is accounted for in the other sections of the equation. The derating is affected by soiling, shading, mismatch, wiring, connection, light-induced degradation, and nameplate rating Masrur et al. (2021). For all of these, the expected loss in efficiency is typically below 3% Masrur et al. (2021) with a total typical loss of 15% annually on the high end. However, mismatch, wiring, light-induced degradation, and nameplate rating are all features of the hardware, not the environment, and as we are simply comparing locations with an approximation, we can ignore these features. If we need a closer approximation to the true output, we can uniformly scale all our outputs at the end with an expected efficient loss for every location. This is because this loss of efficiency will happen no matter where the panels are placed so this would not only help us compare locations. This leaves soiling and shading.

Soiling refers to dust, dirt, snow, animal droppings, bacteria, pollen, and other particles collecting on solar panels and is considered to be reversible with maintenance but only is expected to contribute to 3% PV loss annually, though it can be seasonal. However, in some conditions, this can reach 25% loss if heavily soiled and solar panels are not cleaned regularly Marion et al. (2005). This is the one factor I have not been able to fully account for and would be where I would continue this project given more time. Still, the suggested values for most locations range from 2% to 5%, a 3% difference meaning for most cases, my values would be off by 3% Solar (2022). Cleaning can be used to combat soiling effects as a system that typically had 1.9% loss from soiling was able to sustain 1.2% loss instead with 3 cleanings a year Solar (2022) Walker. Based on documentation for state-of-the-art (SOTA) models, NREL's System Advisor Model (SAM) Blair et al. and HOMER Pro 3.15 Lilienthal et al., to accurately account for

soiling, on-ground sampling is recorded and thus I conclude that this factor cannot be derived simply from satellite data. While the expected difference between locations is small given annual cleaning, it is still a factor that should be precisely accounted for in future work. For the locations selected in this project, based on Rodrigo et al. (2020), soiling loss in this region approximately increments at 0.16% per day up to a maximum total accumulated efficiency loss of 11% if left uncleaned. Given monthly cleaning, we can approximate a 2.5% loss annually from soiling uniformly as the climate in the region is similar Burton (2013).

Finally, shading effects will come down to local logistics as to whether objects such as trees, or tall, shadow-casting obstacles can be removed (if necessary). As this project looks to provide power potential to help select locations of interest, we will assume that no shading is occurring from these causes in the ideal case for our locations. This will leave the logistics of this to be determined when locations can be closely examined after using this method to help narrow down candidates. With the exclusion of soiling and temperature for reasons specified prior, the effects from other features that contribute to the derating factor do not vary dramatically from location to location. In other words, the variation in location is mostly captured by temperature and GHI given proper maintenance of systems hence the terminology of potential PV output.

4.2.2 Solar Radiation Incident and Cell Temperature

Solar radiation incident is dependent on the orientation of the solar panel. However, if we install panels horizontally, we can approximate the radiation as GHI levels Richardson et al. (2019) Ost (2022). Cell temperature is dynamic as it warms from radiation, and cools from ambient air, wind, rain, and snow. Luckily, HOMER derives an intermediary equation that lets us approximate this heat transfer process. First, cell temperature is given below as

$$T_c = T_a + G_T * \frac{\tau\alpha}{U_L}(1 - \frac{\eta}{\tau\alpha})$$

Lilienthal et al.. We can simplify this using the study from HOMER on nominal operating cell temperature Lilienthal et al. and our previous assumption of a horizontal panel to get:

$$\frac{\tau\alpha}{U_L} = \frac{T_{c,NOCT} - T_{a,NOCT}}{G_{T,NOCT}} = (45 - 20)/0.8 = \frac{392.25}{kW/m2°C}$$

. From values recommend by Duffie and Beckman (2013) and HOMER we get

$$\frac{\eta}{\tau\alpha} = 0.20/0.90$$

giving

$$T_c = T_a + (GHI * (45 - 20)/0.8) * (1 - 20/90)°C. \quad (2)$$

Now, our cell is defined as a function of GHI and ambient temperature. We add this back into our final equation to get our approximation for PV output P_r.

$$P_r = Y_{pv}GHI(1 - (0.48\%/°C)[T_a + GHI(\frac{T_{c,NOCT} - T_{a,NOCT}}{G_{T,NOCT}})(1 - \frac{20}{90}) - 25°C]) \quad (3)$$

Note that we leave out the derating factor's soiling effects due to the reasons described in 4.2.1. P_r is a relative approximation of the true PV output without factoring in spatially invariant, innate system loss.

I first validate my approximation using eq. (3) under standard conditions: GHI levels of 1kW and T_a such that the cell temperature is $25°C$. Applying $T_a = 0°C$ to reach a $25°C$ cell temperature under 1kW of GHI from eq 2, we get $P_r = 400W * 1.000 = 400W$. As the cell's rating by definition is its expected output under standard conditions, this validates that the

Month	PV Output	Units
Jan	45.16	kWh/Month
Feb	53.22	kWh/Month
March	80.38	kWh/Month
April	81.85	kWh/Month
May	67.50	kWh/Month
June	58.96	kWh/Month
July	82.56	kWh/Month
Aug	78.98	kWh/Month
Sept	53.87	kWh/Month
Oct	50.45	kWh/Month
Nov	45.98	kWh/Month
Dec	44.31	kWh/Month

Table 2: Predictions for power output PV (kWh/month) for each month in 3 year(s) after 2020 in location lat long = [26.975 -104.] using 6 years for input

results match the manufacturer's measurements Ost (2022). Next, I validate against NREL's SOTA SAM model Blair et al.. Using a cell array of 4000W with 0 tilt (4.2.2) and 0 system loss from cell hardware (as explained in 4.2.1), SAM predicts an annual output of 6,968 kWh/Year with an expected range between 6,639 and 7,114 kWh. Using 3, my average approximation over the last 21 years results in 7023.43 kWh/Year. My results are well within the expected range with a 0.79547% error, well under 1% of SAM's projection. Note that soiling loss in this area in California is expected to be only 0.051% Kleissl and Mejia (2013) thus had we included the derating factor, all expected loss would be from the hardware system which does not vary between locations and merely scales both the SAM predictions and our values uniformly.

Additional validation is also done by comparing intermediate temperature-based power loss and expected cell temperature for (2) against Niclas (2022). Finally, I use this to generate the P_r values for every hour for every day within my data set.

4.3 Initial Modeling

I first use linear regression models and random forest models to build my predictor models where each has 2 hyper-parameters: the number of years ahead to predict the PV output for (y values), and the number of years to base these predictions on (x values). The predictions are monthly kWh PV output levels for each month during the target year. I use a train-test split of 0.8 and 0.2.

In order to reduce the number of features, I reduce the temporal granularity from hours to months. As such, each year in the training set has 7 features (Year, Month, Average Temperature, Average GHI, total PV output, Latitude, Longitude) for each month for a total of 84 features. Each output/prediction in the target year has 12 values, one per month as shown below in Table 2.

5 Results

The model is sensitive to the years used to train and the years to predict ahead. When given many number years of data to base predictions on, the model started to over-fit both due to an excess of features and a reduction in the size of the training data. The size of the dataset is the biggest factor. These are reflected in Table 3 and Table 4. What was surprising is that predicting farther ahead seemed to do better than the years closer for the case where only 1 year was given,

Train R2 Score	1 year ahead	3 year ahead	5 years ahead	10 years ahead	15 years ahead
Given 1 year prior	0.8397	0.85798	0.88339	0.9081	0.9368
Given 3 years prior	0.9597	0.9609	0.9626	0.9705	0.9778
Given 5 years prior	0.9765	0.9753	0.9764	0.9813	0.9933
Given 10 years prior	0.9905	0.9914	0.9925	1.0	N/A
Given 15 years prior	1.0	1.0	1.0	N/A	N/A

Table 3: R2 scores on the training set for different parameters in Linear Regression Model

Test R2 Score	1 year ahead	3 year ahead	5 years ahead	10 years ahead	15 years ahead
Given 1 year prior	0.84434	0.84703	0.8658	0.9052	0.9332
Given 3 years prior	0.9532	0.9556	0.9515	0.9555	0.9560
Given 5 years prior	0.9666	0.9640	0.9605	0.9550	0.9106
Given 10 years prior	0.9671	0.9575	0.9386	0.0090	N/A
Given 15 years prior	0.6006	0.4668	0.9055	N/A	N/A

Table 4: R2 scores on test set for different parameters in Linear Regression model

and then the opposite tended to occur as more years prior were given (I later realized why as addressed in the Error Analysis section).

However, by using a random forest model, this variation vanishes and my model's train score nearly reaches 1 and the test score nearly reaches 0.95 no matter how I vary the years in the future when only 1 year's data is given. However, the random forest model takes significantly longer. For other points, the performance between the two models is comparable. Thus, it seems better to use a random forest model for predicting the future when only a few historical records are available, but better to use a linear regression model when more historical data is available due to its faster run time.

6 Error Analysis and Adjustments

While the scores for both training and testing were very promising (suspiciously so), this ended up revealing an issue in my initial approach. By shuffling all points, I had included points that were in every year which while good for training diversity, also meant that the model could gather information on how similar locations would perform in the future before making predictions. During predictive modeling, that prior information would not be available. This model can, however, be a powerful tool for filling in missing data points where there might be gaps or erroneous anomalies in satellite data thus I retain it in the results of this project.

To adjust my approach for the purposes of future prediction, I trained my models only on the first 21 years, and then looked at how they performed for the remaining data. While the models fit well during the training phase, they did very poorly in future prediction. The R2 score for testing dropped down to negative values for every case, meaning the model does worse than just guessing the mean of the data.

The model is not able to predict future PV output levels better than the null hypothesis. However, by observing the outputs themselves, we can see that the predictions are not terribly different from the real values as shown below in Figure 6. The negative R2 score simply means using the mean of the data as our guess would have been a better choice (the null hypothesis). Given the predictions are not terribly off the mark, it is likely the overall trend, or bias, in the data simply is not deviating significantly. In other words, a flat line is a strong approximation for the trends in the calculated PV over the last 2 decades and I am over-fitting even with simple models. While disappointing for the model, this is not a bad result to get! Had there been a

statistically significant change in this output over the last 20 years, that would indicate a drastic change in our climate or solar radiation levels. According to **?**, since 1981, the average increase in global temperature per decade is 0.18°C, or 0.36° C in the time frame of this project's data. Given fluctuations from year to year have relatively higher magnitude, decoupling just this minor average trend in temperature is extremely difficult given the time span of the data.

While I had hoped the model had picked up spatial information to help aid its predictions, it is not unexpected that it did not to the degree that it would have significantly aided my predictions as I only had 156 points and I did not sample a diverse set of locations. As it stands now, choosing locations with the expectation that they will have similar PV output levels as they would have over the last 20 years is a strong prediction. This of course is based on historical data and does not account for the unpredictable directions human activity might take the climate. In this project, the best location is at **21.4400, -104.0000** with an expected annual output of **8,582.73 kWh/Year** using a 4kW system

```
Predictions for power output PV (kWh/month) for each month in 1 year(s) after 2020 in location lat long = [  26.975 -104.   ]
predicted Jan 50.35866639970251 kWh/Month True Jan 48.67770229200006 kWh/Month
predicted Feb 53.03325808964786 kWh/Month True Feb 54.250876825333336 kWh/Month
predicted March 71.78966229769627 kWh/Month True March 65.38398722266666 kWh/Month
predicted April 78.59060085134101 kWh/Month True April 76.48001033733334 kWh/Month
predicted May 85.90976398887449 kWh/Month True May 87.66101766266667 kWh/Month
predicted June 77.59701977443622 kWh/Month True June 80.39550647066667 kWh/Month
predicted July 82.19858242987034 kWh/Month True July 79.77037213866667 kWh/Month
predicted Aug 80.00729611835781 kWh/Month True Aug 82.68266143600002 kWh/Month
predicted Sept 66.62728994120593 kWh/Month True Sept 63.10631262933333 kWh/Month
predicted Oct 59.13842216245745 kWh/Month True Oct 67.40261739733333 kWh/Month
predicted Nov 51.85420590463039 kWh/Month True Nov 54.57293636 kWh/Month
predicted Dec 49.928496228793 kWh/Month True Dec 52.731597918666665 kWh/Month
```

Figure 6: Prediction outputs vs true outputs for 2020 example

7 Conclusion

During this project, I was able to create an approximation of PV output potential for hours, days, months, and years by using only two channels of satellite data, train a strong model for filling in missing data points in this PV potential, and demonstrate how trends in the last two decades have not pointed towards a significant change in expected PV output levels from year to year. While the modeling section was unable to beat the null hypothesis, its close predictions paired with negative R2 scores indicate that at least based on the last 2 decades of data, we can closely predict future PV output levels by simply taking the mean of our historical PV output approximations. To improve, I would look to better account for soiling factors and gather data on more diverse locations and gather a great history of recorded data when available.

References

(2015). Paris agreement.

Blair, N., Dobos, A. P., Freeman, J., Neises, T., Wagner, M., Ferguson, T., Gilman, P., and Janzou, S. System advisor model, sam 2014.1.14: General description.

Burton, T. (2013). Mexico's seven climate regions: Geo-mexico, the geography of mexico.

Crimmins, A., Kolian, M., Bacanskas, L., and Rosseel, K. (2016). Climate change indicators in the united states 2016 (fourth edition).

Duffie, J. A. and Beckman, W. A. (2013). Solar engineering of thermal processes, 4th edition.

GDAL/OGR contributors (2022). *GDAL/OGR Geospatial Data Abstraction software Library*. Open Source Geospatial Foundation.

Ge, M., Friedrich, J., and Vigna, L. (2020). 4 charts explain greenhouse gas emissions by countries and sectors.

Hasan, K., Yousuf, S. B., Tushar, M. S., Das, B. K., Das, P., and Islam, M. S. (2021). Effects of different environmental and operational factors on the pv performance: A comprehensive review. *Energy Science amp; Engineering*, 10(2):656–675.

Kim, A. (2019). Machine learning - projects spring 2019.

Kleissl, J. and Mejia, F. (2013). Soiling losses for solar photovoltaic systems in california. *Solar Energy*, 95:357–363.

Lilienthal, P., Flowers, L., and Rossmann, C. Homer: The hybrid optimization model for electric renewable.

Lorenz, E. and Heinemann, D. Global horizontal irradiance.

Marion, B., Adelstein, J., Boyle, K., Hayden, H., Hammond, B., Fletcher, T., Canada, B., Narang, D., Kimber, A., Mitchell, L., Rich, G., and Townsend, T. (2005). Performance parameters for grid-connected pv systems. In *Conference Record of the Thirty-first IEEE Photovoltaic Specialists Conference, 2005.*, pages 1601–1606.

Masrur, H., Konneh, K. V., Ahmadi, M., Khan, K. R., Othman, M. L., and Senjyu, T. (2021). Assessing the techno-economic impact of derating factors on optimally tilted grid-tied photovoltaic systems. *Energies*, 14(4).

Niclas, N. (2022). Measuring the temperature coefficient of a pv module.

Ost, I. (2022). How much energy does a solar panel produce?

Richardson, W., Cañadillas Ramallo, D., Moncada, A., Guerrero-Lemus, R., Shephard, L., Vega-Avila, R., and Krishnaswami, H. (2019). Validation of all-sky imager technology and solar irradiance forecasting at three locations: Nrel, san antonio, texas, and the canary islands, spain. *Applied Sciences*, 9:684.

Rodrigo, P., Gutiérrez, S., Micheli, L., Fernández, E., and Almonacid, F. (2020). Optimum cleaning schedule of photovoltaic systems based on levelised cost of energy and case study in central mexico. *Solar Energy*, 209:11–20.

Sengupta, M., Xie, Y., Lopez, A., Habte, A., Maclaurin, G., and Shelby, J. (2018). The national solar radiation data base (nsrdb). *Renewable and Sustainable Energy Reviews*, 89:51–60.

Sharma, N., Sharma, P., Irwin, D., and Shenoy, P. (2011). Predicting solar generation from weather forecasts using machine learning. In *2011 IEEE International Conference on Smart Grid Communications SmartGridComm)*, pages 528–533.

Solar, A. (2022). Understanding pv system losses, part 3: Soiling, snow, system degradation.

Sunrun, S. T. (2021). Do solar panels work in cold weather?

Tartakovsky, V. A., Cheredko, N. N., Maximov, V. G., and Volkov, Y. V. (2018). How does the sun affect the surface temperature? *IOP Conference Series: Earth and Environmental Science*, 211(1):012002.

Teqo, W. b. M. (2022). Impact of temperature on solar panels efficiency.

Walker, H. A. Best practices for operation and maintenance of photovoltaic and energy storage systems; 3rd edition.

Wesoff, E. (2015). Solar star, largest pv power plant in the world, now operational.