Optimizing the Geographic Location of Photovoltaic Panels in the Contiguous US

New Mexico
Supercomputing Challenge
Final Report
April 06, 2022

Team Number: 4

Santa Fe Preparatory School
Capital High School
Santa Fe High School

Team Members
Lucas Blakeslee
Ian Olson
Val Ornelas
Shrey Poshiya

Teachers
Irina Cislaru
Jocelyn Comstock

Mentors
Drew Einhorn
Leon Viveros

Contents

1 Executive Summary

Photovoltaic energy has been at the forefront of our transition to renewable energy. It has the potential to be the primary alternative to fossil fuels, and with improving energy panel technology it has the potential to produce far more energy than it currently does. The problem is that photovoltaic farms only work when the conditions are favorable, and if there is a demand for energy that cannot be met due to overcast weather, it could result in a large problem. Because of this, the project was designed around the concept that there must be a spot in the contiguous US where solar panels would be the most reliable and produce the most possible energy. The model that resulted ultimately predicted the energy generation of a 300 watt solar panel in a certain location based off of weather data; the following variables were used: Global Horizontal Irradiance (Watts/Meter Squared), Temperature (Celsius), Relative Humidity (Percent), Pressure (Millibar), and Wind Speed (Meters/Second). Using this it was possible to find the latitude and longitude of the point where a singular photovoltaic panel would produce the maximum energy, as well as an overall visual representation in the form of a heat map of the most efficient placement of solar panels.

2 Background

2.1 Multiple Linear Regression

Multiple linear regression is a method which can be used to find the predicted output of two or more independent variables. Effectively, it finds the correlation between the independent and dependent variables. This relationship can be modeled by the equation:

$$(b_1 * x_1) + (b_2 * x_2) + ...(b_n * x_n) + b = y$$

3 Introduction

3.1 The Problem

The root problem of this project is that despite the potential of solar panels to provide a lasting source of renewable energy, they ultimately are unreliable in their energy production due to the weather variability. For photovoltaic energy to have the potential of replacing fossil fuels, they ultimately need to have both the reliability and power output needed to meet the current energy consumption in the US.

3.2 The Question

Because of the necessary requirements to make solar energy a viable option, this project looked to answer the question: Where is the best geographic location for a photovoltaic panel such that it produces the greatest net energy output?

4 Methods

4.1 Data Acquisition

Data on power generation was acquired for two sites in the contiguous USA; one is located in Lakeside,CA and the other in Shady Hollow,TX. This data was sourced from pvout.org [5].

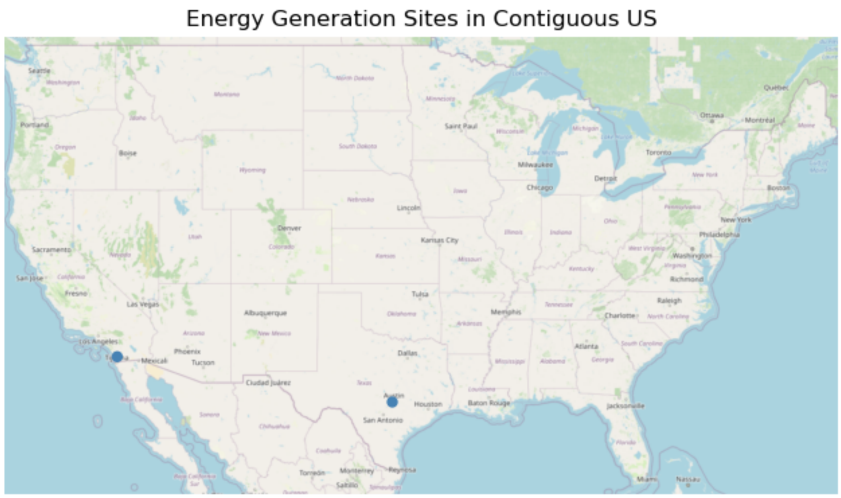

Fig. 4.1 The locations of the sites that were chosen for data on energy generation

Each site had a different configuration of photovoltaic systems with each system having a different amount of panels and panel capacities. To standardize each of the systems, the energy was calculated with the assumption that the system contained 300 W panels. The compiled data contained energy for a 300 W panel in that location. The following equations were used to standardize the energy generation for each of the systems:

(Total System capacity) / (300 W) = # of Panels
(Energy Generation in Watt Hours) / (# of Panels) = Energy Generated by 300 W panel

Energy generation data from the first site in Lakeside,CA used data from May, 2020; the second site in Shady Hollows,TX used data from June,2020. For each month, weather data from a specific day was matched with energy generation data for each of the respective locations and dates. Weather data was sourced from the National Solar Radiation Database (NSRDB) [1][4]. This data returned hourly weather data for a specified location over the course of a year. Therefore, when compiling the data, daily averages were taken for the specified dates were taken and added to the compiled data. The following meteorological variables were sourced from NSRDB and added to the data:

- Global Horizontal Irradiance (Watts/Meter Squared)
- Direct Normal Irradiance (Watts/Meter Squared)
- Diffuse Horizontal Irradiance (Watts/Meter Squared)
- Temperature (Celsius)
- Relative Humidity (Percent)
- Wind Speed (Meters/Second)

The final dataset contains energy generation from a certain site for a 300 watt panel matched with weather data from that specific day and location.

4.2 Linear Regression Model

Using the compiled data set of energy generation and weather data, a multiple linear regression model was calculated to predict energy generation of a 300 W solar panel under certain weather conditions. This model was calculated using the scikit-learn library.

The correlation of between the meteorological variables and the energy generation was first plotted to determine which of variables had the most linear relationship with output. The correlation can be represented with the following plot:

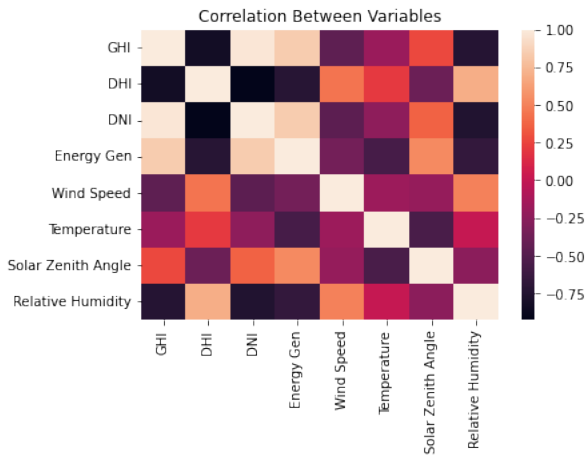

Fig. 4.2

Using the correlation data, it was determined that the following variables were dropped from the multiple linear regression model:

- Relative Humidity
- Diffuse Horizontal Irradiance
- Temperature

Using the remaining variables, the following linear equation was produced by the model:

$$((2.58338107 * \mathrm{GHI}) + (1.35516597 * \mathrm{DNI}) + (17.3974214 * \mathrm{Wind\ Speed}) + 165.16409487008036 = \mathrm{PV\ OUT(Watt\ Hr\ /\ day)}$$

Further weather data was collected from NSRDB for 5681 locations in the contiguous USA. The NSRDB data returned hourly data for the entirety of 2020; for each of the chosen points, yearly averages of each of the metrics was taken. This compiled weather data was then fed into our regression model in order to predict energy output in different locations. The Prediction gives the energy

5 Results

The multiple linear regression model that was calculated had a correspondence or R^2 of 0.7181800646348857 to our energy output data.

The predictions made by the regression model can be modelled by the following figure:

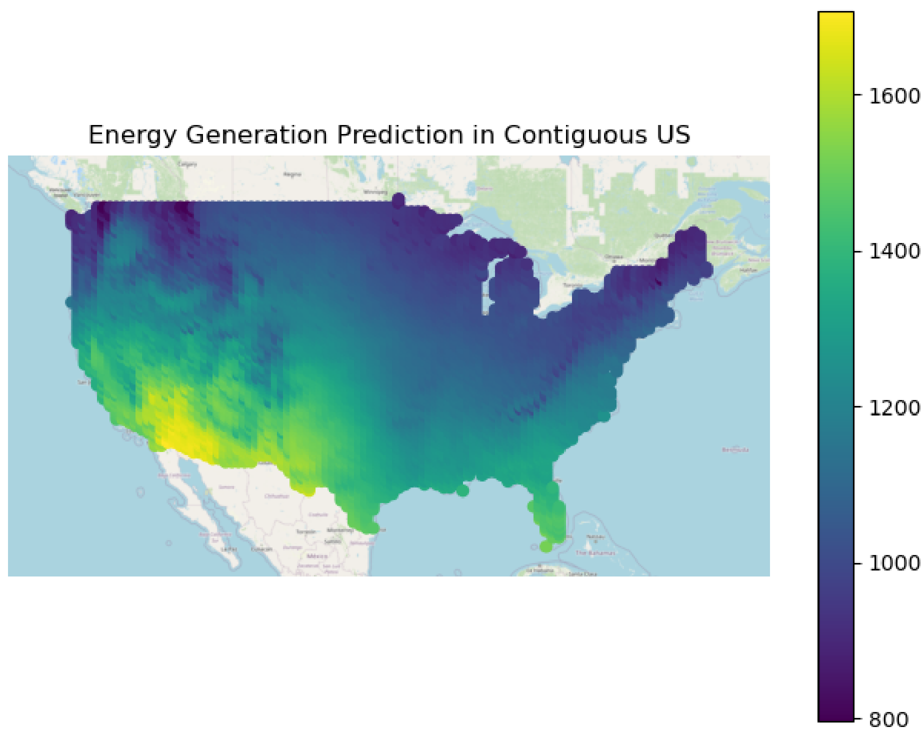

Fig 5.1

According to our model, the location with highest energy generation is located at (33.10023, -115.802026).

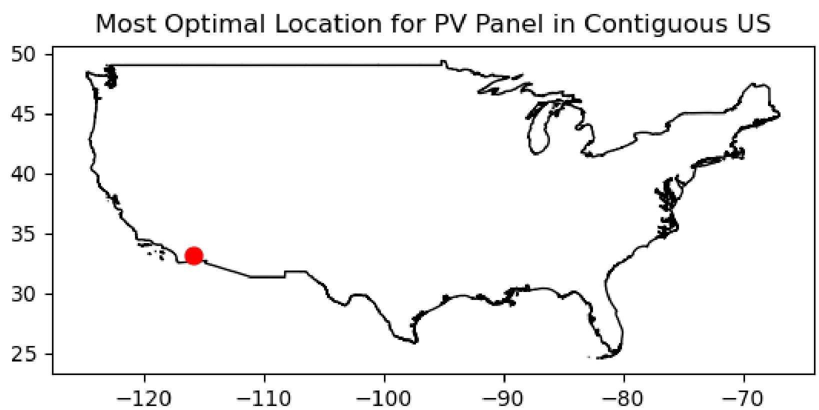

Fig 5.2

6 Conclusion

Our results conclude that the best region to maximize PV power output is in the Southwest United States. We also can observe a general trend that the power output of PV panels decreases with increasing latitude. Ultimately, this model will not be completely accurate, as it is possible to include more variables which impact PV power output. More specifically, it is possible to include temperature, humidity, and air pressure into the linear regression model. Although it is possible to include more variables, the results should ultimately still follow general trends, as the most important variables have been included. In the future, to expand on this project, we would like to upgrade the linear regression model to accommodate more variables. We would also like to account for the transportation of the energy, which would mean evaluating the cost to transport the energy to the places where it would be utilized. To do this, we would analyze the cost of using existing electrical infrastructure and evaluate both the feasibility of using such infrastructure as well as the cost of building new power distribution.

To access all code, data, and plots, visit the Team 4 Github repository : `https://github.com/sposhiy33/SCC_team04`

7 Bibliography

[1] [Data/information/map] obtained from the "Global Solar Atlas 2.0, a free, web-based application is developed and operated by the company Solargis s.r.o. on behalf of the World Bank Group, utilizing Solargis data, with funding provided by the Energy Sector Management Assistance Program (ESMAP). For additional information: https://globalsolaratlas.info

[2] Clack, C. T. M. (2017). Modeling Solar Irradiance and Solar PV Power Output to Create a Resource Assessment Using Linear Multiple Multivariate Regression, Journal of Applied Meteorology and Climatology, 56(1), 109-125. Retrieved Apr 6, 2022, from https://journals.ametsoc.org/view/journals/apme/56/1/jamc-d-16-0175.1.xml

[3] Tarek AlSkaif, Soumyabrata Dev, Lennard Visser, Murhaf Hossari, Wilfried van Sark, A systematic analysis of meteorological variables for PV output power estimation, Renewable Energy, Volume 153, 2020, Pages 12-22, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2020.01.150. (https://www.sciencedirect.com/sc

[4] Sengupta, M., Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby. 2018. "The National Solar Radiation Data Base (NSRDB)." Renew-

able and Sustainable Energy Reviews 89 (June): 51-60.

[5] https://pvoutput.org/

8 Acknowledgements

- Thank you Drew Einhorn for mentoring us and guiding us through the project
- Thank you Celia Einhorn for helping us network and find people who could assist us in research
- Thank you Jocelyne Comstock for helping us meet deadlines and get our project done
- Thank you to all the coordinators and judges for the Supercomputing Challenge for providing us the opportunity to participate