Weather’s Impact on PV Power Output

Nebojsa, Ajdarevic (N10434348)  
IFN704 Assessment 3, due 11:59pm Friday 30th October 2020

# Executive Summary

This project aims at assessing the impact changes in local weather can have on the power output of consumer grade Photovoltaics (PV). While many studies have been conducted on industrial grade PV systems, studies assessing multiple weather factors on consumer grade PV’s are few and far between. Data were gathered from three main sources. Solar Analytics for the energy data, world weather online API for the weather variables and Solcast API for the solar irradiation variables. A linear regression model was then built determining that for the variables considered only GHI, DHI, windspeed and cloud cover were relevant variables. These results seem to suggest that a sunny windy place would be ideal for consumer grade solar panels. However, as the data was only gathered for the Brisbane location this result is limited in its generalisation.

# Introduction

There is a wide level of agreement amongst the scientific community that the emission of greenhouse gases from energy and energy productions must be reduced, or eliminated entirely to avoid the extreme dangers that climate change will pose to society if it is left unchecked [1]. Even sources like the WHO are predicting 250,000 additional death each year between 2030 and 2050 [2]. Therefore, to combat this, sunlight based photovoltaic (PV) frameworks are being used as one of the main sources of renewable energy [3]. Furthermore, Solar power has just recently become the cheapest form of energy gathering in the world according to the International Energy Agency’s World Energy Outlook for 2020.

While PV testing has been strenuous over standard operating conditions, it is also of the upmost importance to understand their effectiveness when deployed in the real world. This will allow for accurate prediction. Currently there are many studies that look at the impact various weather conditions have on commercial grade solar systems. These studies have repeatedly shown that simple linear regression can capture the impact a single weather condition has on PV output, for example temperature will decrease output while windspeed will increase it [4]. However, few studies aim to assess the impact of various weather conditions in a holistic view of many weather variables and fewer still explore these on consumer grade PV systems. Therefore, this study will aim to explore the real-world impact that meteorological data can have on consumer grade PVs in the field.

**Literature Review**

Solar power remains one of the most con free energy sources, it is easy to install, has no noise, emissions or moving parts. This plus the potential energy production has added greatly to its popularity increase over the last few years [4]. The main barrier to entry is the expensive onset cost of buying PV. Despite this, PV capacity has surpassed 303 giga watts as of 2016. Solar panels work by converting sunlight directly into electricity using PV, thus having many applications including, off-grid industrial, off-grid residential, on-grid and consumer [4]. however, with its exposure to the elements comes some limitations in terms of the weather conditions present.

The weather can have an immense impact on the energy output of PV systems [5], [6]. It is known that solar irradiance has the greatest impact on energy output as PV systems directly convert this into solar energy [5]. However, there are also other types of weather conditions that can have an impact on the energy output of PV systems. For example, ambient temperature [4], [5], rain and humidity [3], and dust particles [7] which decrease the energy output of PV systems. Or alternatively windspeed which increases the energy output as it cools the PV systems [4], [5], [8], [9]. However, aside from the [3] study, most of these studies looked at variables in isolation from one another. When they were compared multiple variables simultaneously, [7] found that rain increased the output of PV in dusty climates. In terms of modelling these relationships most studies have observed simple linear relationships between variables with the only exception being that [7] found that there was an interaction effect between the amount of dust and rainfall.

There are many issues that arise when variables are considered in isolation. For example, when rain is considered alone it decreases the PV output, but if dust is considered it has and interaction effect [7], or when cloud cover is considered it increases the output as a result of cooling [3]. Additionally, these studies only focused on commercial grade PV systems and as a result may not be generalisable to consumer grade PV systems as they are of lower quality and have smaller power generation ability.

As a result of these shortcomings, my analysis will aim to investigate all of these covariates together in the way they would happen in nature. A linear relationship will be assessed first but will be adjusted based on the data exploration as this will allow the project to avoid type one and two errors, as well as breaking any underlying assumptions [10]. It is now more important than ever to understand the varying impacts of these weather conditions as a more unstable environment caused by climate change will force more extreme weather conditions more regularly [1]. This will allow more informed decisions to be made regarding the potential output of solar panels to be installed as well as informing more efficient locations for such panels. Manufactures could even target locations more likely to have optimal conditions by informing consumers of the benefits of their location on top of other benefits provided by the PVs.

# Approach

The project undertaken had three major deliverables. These were; The collection of data for the construction of a linear regression relating weather data to the given power output of PV systems, the construction of this model and the validation of its effectiveness and the potential uses for this model.

## Data Collection

The PV generation data was provided by Solar Analytics. This dataset contained the power generation (in watt hours), the minimum and maximum voltage all in 5-minute intervals. The data provided was de-identified due to privacy concerns, but postcodes were provided for each site ID.

The other weather data was gathered from multiple sources. Due to the high resolution of solar generation data provided by Solar Analytics obtaining weather data was a difficult task, however weather data was obtained in hourly resolution via a 60-day free trial from the World Weather Online API. The data gathered was limited to a few pings of the API and as such weather data was gathered only for Brisbane over 2019. The variables and their descriptions can be seen in appendix one below. However, the main variable for solar power generation (solar irradiation) was not included in this dataset and as such needed to be collected from an additional dataset,

Solar irradiation data was collected from the Solcast API, it was in hourly resolution and contained the variables as can be seen in appendix 2 below. However, it needs to be noted that as this data was obtained with a student/publication licence only, it may not be used for any commercial purposes.

These datasets were then combined. First sites located close to Brisbane were chosen based on the postcodes provided by Solar Analytics. Then PV generation data was resampled to hourly and the rest of the datasets were then merged based on the date and time. Although it is not necessarily optimal to change the data to a lower resolution it was decided that this was more optimal than attempting to factor lower resolution data into higher resolution data due to the large amount of variables that had no logical way of being resampled into a higher resolution. Furthermore, it is unlikely that this would have produced accurate or meaningful results to the underlying uncertainty of making so many assumptions about each potential predictor variable.

## Model Construction

Data exploration played a large role in the eventual model construction. As the power data was resampled to be hourly, there are times of day where there is no sunlight and as such would impact the analysis negatively, as a result all rows where there was no power production were dropped from the dataset.

Outliers were present in the dataset according to an IQR assessment. However, it was determined that then majority of these were not impacting the results negatively, nor were they necessarily unlikely to occur. However, there were a few energy values that were clearly skewing the results of the analysis and as such were removed. They can be seen in figure 1 below.

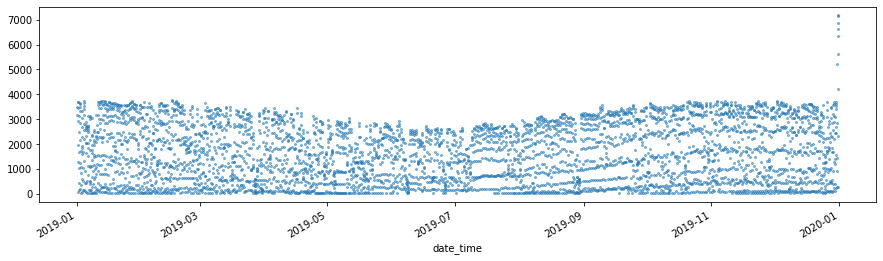


Figure 1 Energy output over time with outliers

After these outliers were removed the trends in energy production could be seen more clearly, as evident in figure 2 below.

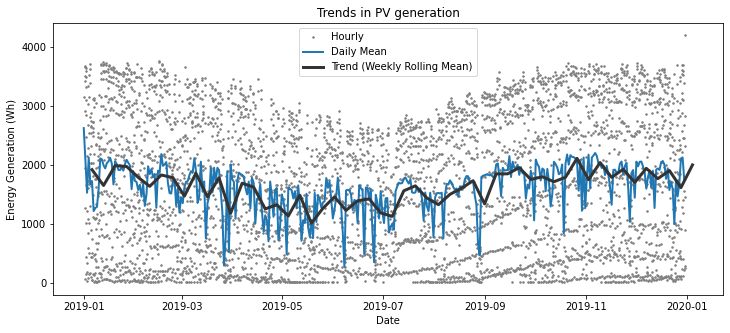


Figure 2 Energy output over time without outliers

Energy values seemed to have a trend, and a s such were coloured based on the time of day in figure 3 below. As you can see there is a clear effect on energy output as a result of the time of day. Furthermore, adjusting by the cloud coverage can expose this trend even more as seen in figure 4 below

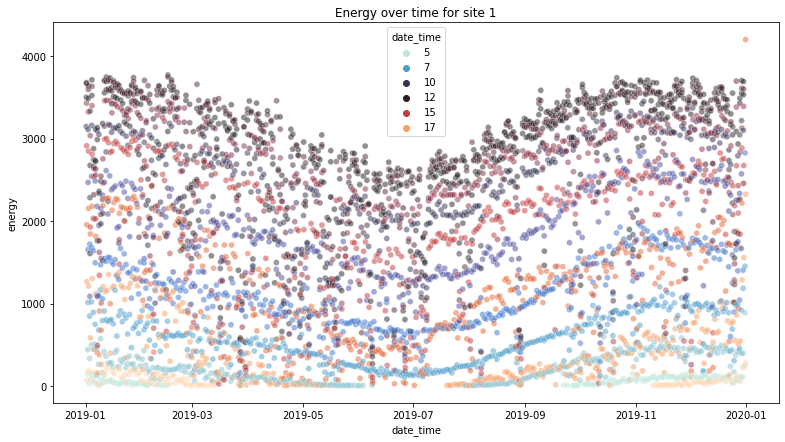


Figure 3 Energy output over time with hour of day as colour scale

The cloud cover of 36% was chosen as it is the mean of the dataset. A comparison of the two graphs shows that cloud cover has a serious effect on the chaotic-ness of energy outputs of consumer grade PV systems

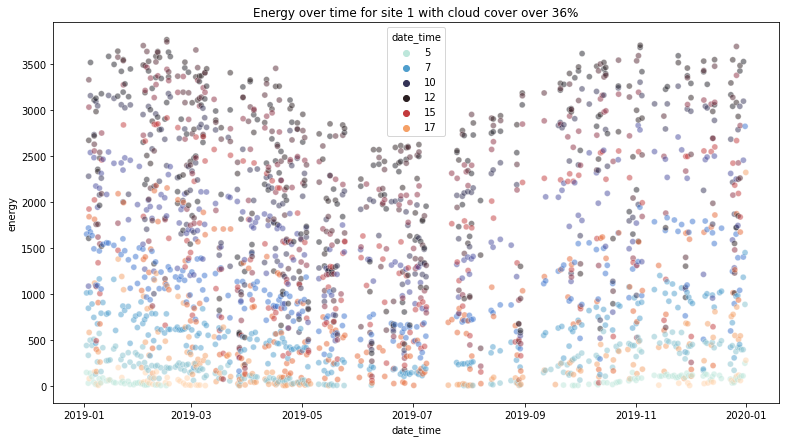
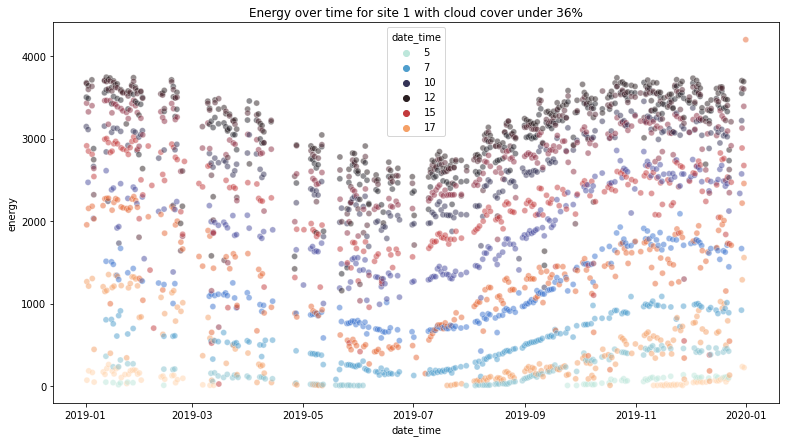
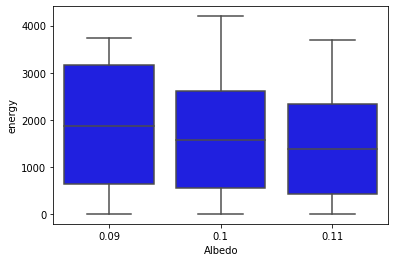
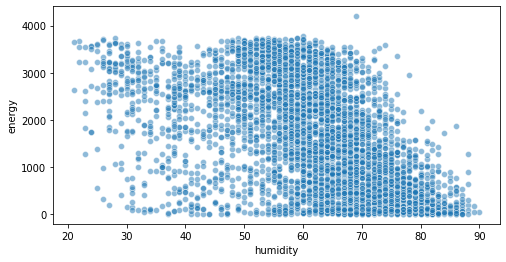
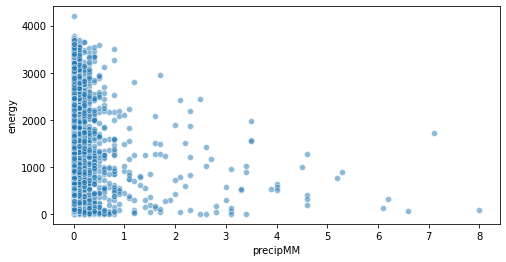
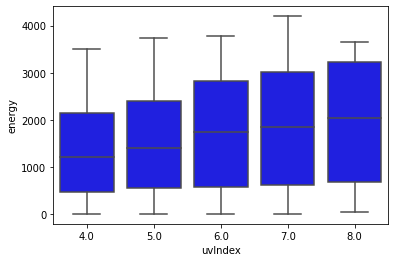


Figure 4 Energy output over time with above mean and below mean cloud cover split

Energy was plotted against all the variables in the dataset and a linear trend was clear for most of the variables, this can be seen in figure 5 below.

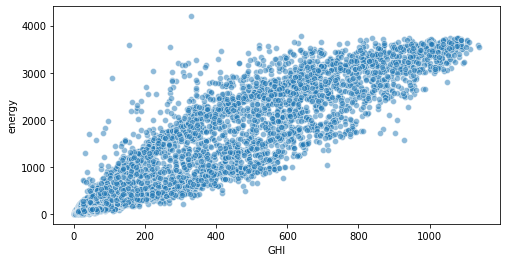
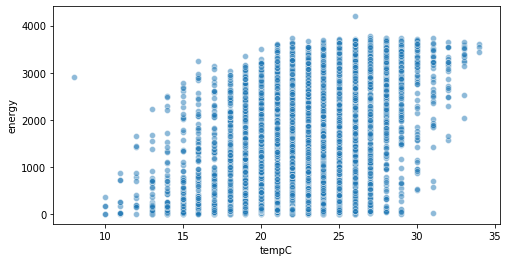


Figure 5 Data exploration of x-variables vs energy output

A big challenge in this dataset was the multicollinearity of the predictor variables. This was especially evident in the solar irradiance variables. DNI, EBH, GtiFixedTilt and GtiTracking were removed before model construction as GHI had the highest correlation with energy while GHI and the rest of these variables had a higher correlation with each other than with energy. DHI was left in for the time being as it seemed to have less of a correlation with the other solar irradiation variables.

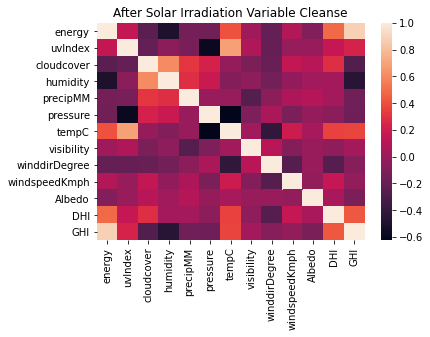
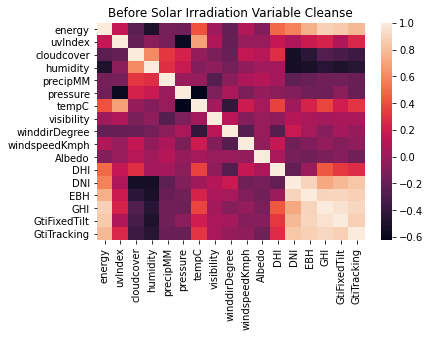


Figure 6 Variable correlation comparison

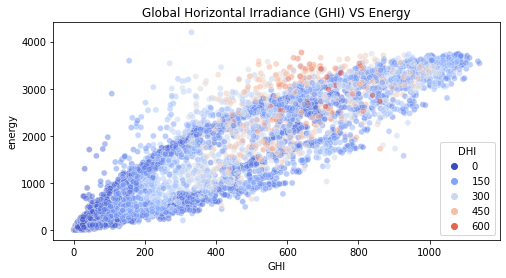
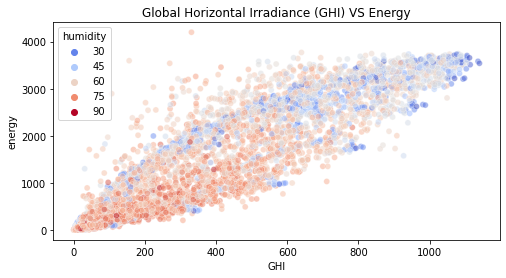
After the initial cleansing of these variables the intended direction of this research was to explore an interaction effect, as such some more exploratory data analysis was undertaken as can be seen in figure 7 below. However, it became clear that having only one site to train on this model would not be generalisable and as such a different way forward was taken.

Figure 7 Interaction exploration

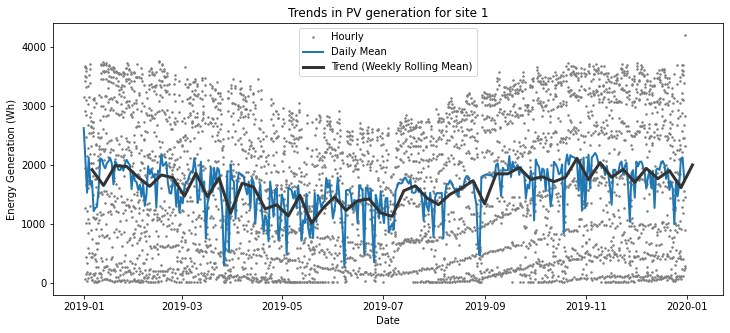
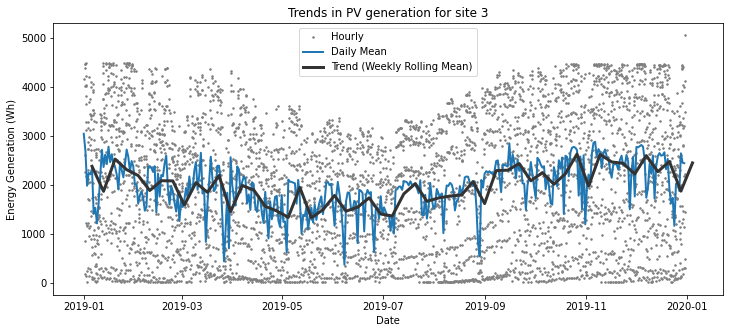
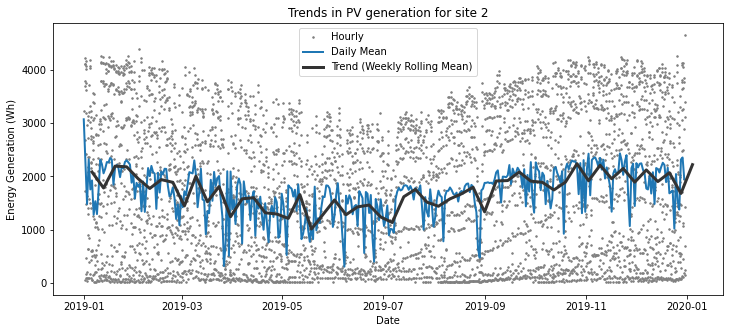
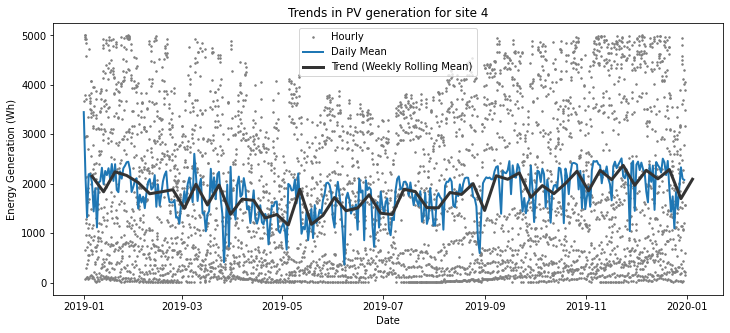
As model generalisability was important 3 more sites were added for the model to learn on. It is noted however that due to API limitations all the sites had to be within the Brisbane region as I could only access weather data for Brisbane. The energy production of all sites can be seen in figure 8 below.

Figure 8 Energy output comparison among all sites

Furthermore, a comparison of the final model with and without outliers was conducted and it concluded that there was no significant difference between the model with outliers and the model without them, therefore outliers were kept in the analysis.

Finally, to deal with the high multicollinearity in the dataset, variance inflation index with a cut-off of 5 was used to get rid of some variables. In this procedure the following variables were eliminated; "visibility", "uvIndex", "humidity", "tempC", "Albedo" and "pressure".

## Findings

The final model produced an output with the predictor variables as per table two below. The variables were selected based on an AIC score which got rid of the “precipMM” variable.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Coefficients | STD. Error | p-value |
| Intercept | -125.10 | 16.69 | 0.000 |
| GHI | 3.62 | 0.02 | 0.000 |
| windspeedKmph | 20.10 | 0.85 | 0.000 |
| DHI | 0.97 | 0.05 | 0.000 |
| cloudcover | -2.87 | 0.19 | 0.000 |

**Table 2 Final Exogenous Variables**

The variables within the model accounted for 75% of the variance in energy output. These results are not surprising as they are often found significant in the literature [3], [4]. However, some of the interesting results here are for the variables that were not deemed to be useful in model prediction. These are; precipitation, temperature and humidity are not statistically significant variables when GHI, DHI, cloud cover and wind speed are considered. This is the case as these variables were dropped from the analysis during the multicollinearity cleaning of this dataset.

These results were mostly as expected as GHI and DHI caused an increase in energy output. Wind speed also cause an increase in the energy output of a solar panel, this is most likely due to the cooling effect wind speed had on the panels [4], [5], [8], [9]. Furthermore, cloud cover caused a decrease in energy output as they blocked out the sun’s rays [3]. Therefore, these results would suggest that a windy cloudless place near the equator would be most beneficial for consumer grade PV’s, however there are limitations to this study that will be discussed later in the report.

The QQ plot suggests that linear regression may not be the way to go as the assumption of normally distributed residuals is broken as can be seen in figure 9 below. However, these don’t look too bad all things considered as it only errors of towards the edges.

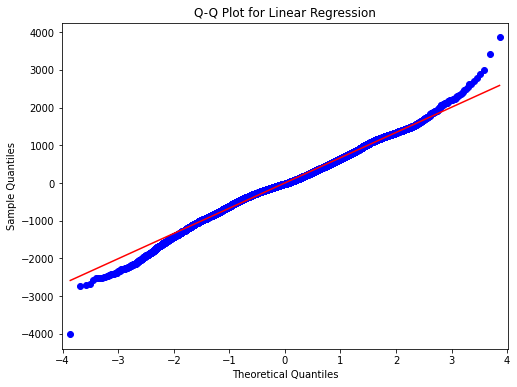


Figure 9 QQ plot for final regression

The residuals are homogenously distributed as can be seen below in figure 10.

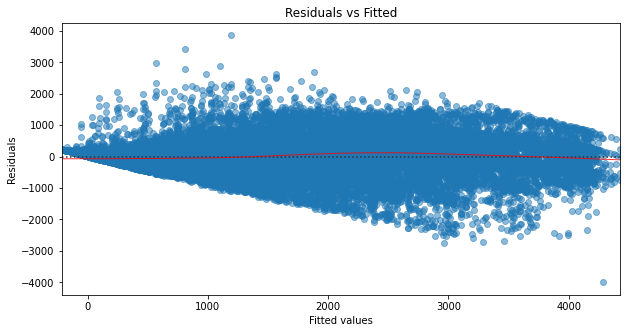


Figure 10 Residual plot for final regression

There are some limitations with the study. This study doesn’t have a real understanding of the size of the PV installations. Nor does it have any onsite data but is instead relying on “local” data for the weather station in Brisbane which is near these sites. Furthermore, the only location assessed within this study was Brisbane, as such this study will not be generalizable as it does not explore the effects of latitude and longitude. So, while it may suggest that a place with high solar irradiation, windspeed and low cloud cover is optimal. This study hasn’t had the opportunity to test if such extreme heat would impact the performance of consumer grade PV as suggested in the literature [3].

# Reflection

There are many ways in which future work could improve on this report. For example, any future work should explore sites in different areas so that the model has a chance to get a more complete understanding of the effects different latitude and longitude and weather effects can have on solar panels, for example, there was no examples of snowy or dusty weather, nor was there any way to compare the effects latitude and longitude could have on the energy output of consumer PV systems. Furthermore, an interaction effect between weather variables should also be explored in future research as this has been shown to be significant in commercial grade solar panels [3], [9]. It would also be of a great benefit if future studies had access to actual onsite data instead of relying on local data that may not be accurate at representing data for each site.

Aside from the various future work, the experience working with a real-world complicated dataset has provided me with valuable experience as to what it takes to make a real-world model using the resources available to me. The presentation has allowed me to develop my presenting skill (a personal weakness) and therefore allowed me to grow as an analyst. As this is my second advanced project unit, I had the opportunity to implement the necessary time scheduling I learnt from the previous time around to allow myself to complete this analysis to the best of my ability.

# References

[1] Edenhofer, O. 2015. *Climate change 2014: mitigation of climate change* (Vol. 3). Cambridge University Press.

[2] World Health Organization. 2020. *Quantitative Risk Assessment Of The Effects Of Climate Change On Selected Causes Of Death, 2030S And 2050S*. [online] Available at: <https://www.who.int/globalchange/publications/quantitative-risk-assessment/en/> [Accessed 16 August 2020].

[3] Kazem, H. and Chaichan, M., 2016. Effect of environmental variables on photovoltaic performance-based on experimental studies. *International Journal of Civil, Mechanical and Energy Science (IJCMES)*, 2(4), pp.1-8.

[4] Hoffmann, W., 2006. PV solar electricity industry: Market growth and perspective. *Solar energy materials and solar cells*, 90(18-19), pp.3285-3311.

[5] Kaldellis, J. Kapsali, M. and Kavadias, K., 2014. Temperature and wind speed impact on the efficiency of PV installations. Experience obtained from outdoor measurements in Greece. *Renewable Energy*, 66, pp.612-624.

[6] Diaf, S., Notton, G., Belhamel, M., Haddadi, M., and Louche, A., 2008. Design and techno-economical optimization for hybrid PV/wind system under various meteorological conditions. *Applied Energy*, *85*(10), pp.968-987.

[7] Hachicha, A. Al-Sawafta, I., and Said, Z., 2019. Impact of dust on the performance of solar photovoltaic (PV) systems under United Arab Emirates weather conditions. *Renewable Energy*, *141*, pp.287-297.

[8] Gökmen, N., Hu, W., Hou, P., Chen, Z., Sera, D. and Spataru, S., 2016. Investigation of wind speed cooling effect on PV panels in windy locations. *Renewable Energy*, 90, pp.283-290.

[9] Abiola-Ogedengbe, A., Hangan, H., and Siddiqui, K., 2015. Experimental investigation of wind effects on a standalone photovoltaic (PV) module. *Renewable Energy*, *78*, pp.657-665.

[10] Zuur, A., Ieno, E., & Elphick, C., 2010. A protocol for data exploration to avoid common statistical problems. *Methods in ecology and evolution*, *1*(1), pp.3-14.

# Appendix 1: World Weather Online API Data

|  |  |
| --- | --- |
| Value | Description |
| time | Local time |
| tempC | Temperature in degrees Celsius. |
| FeelsLikeC | Feels like temperature in degrees Celsius |
| HeatIndexC | Heat index temperature in degrees Celsius |
| DewPointC | Dew point temperature in degrees Celsius |
| WindChillC | Wind chill temperature in degrees Celsius |
| windspeedKmph | Wind speed in kilometers per hour |
| WindGustKmph | Wind gust in kilometers per hour |
| winddirDegree | Wind direction in degrees |
| weatherCode | Weather condition code |
| weatherDesc | Weather condition description |
| precipMM | Precipitation in millimeters |
| humidity | Humidity in percentage (%) |
| visibility | Visibility in kilometers |
| pressure | Atmospheric pressure in millibars (mb) |
| cloudcover | Cloud cover amount in percentage (%) |

# Appendix 2: Solcast API Data

|  |  |
| --- | --- |
| Value | Description |
| Global Horizontal Irradiance (GHI, W/m2) | The total irradiance received on a horizontal surface |
| Direct Normal Irradiance (DNI, W/m2) | Solar irradiance arriving directly from the sun as measured from a surface perpendicular from the sunlight |
| Direct (Beam) Horizontal Irradiance (EBH, W/m2) | The horizontal component of DNI |
| Diffuse Horizontal Irradiance (DIF, DHI, W/m2) | The horizontal component of irradiance scattered through the atmosphere |
| Albedo | the measure of the diffuse reflection of solar radiation out of the total solar radiation |

# Appendix 3: Abridged Data Analysis