CEE322: Data Analytics for Urban Systems

Lab 2: Exploratory Data Analysis: Time Series

Author: Davyd Tamrazov
Due date: Oct 5, 2020
Stanford University

1 Introduction

The purpose of this report is to perform exploratory data analysis for the utility serving parts of the northwestern United States. Datasets used in the following study contain hourly weather information for the sites in Ashton, Idaho and Deer Lodge, Montana. The primary aim of of the analysis is to assess the potential of each site for wind and solar generation, based on the temporal weather characteristics as well as their predictability and correlation.

2 Dataset

The dataset for each site contains 8758 entries corresponding to the hourly weather information recorded through the year of 2014. From here onward, Ashton, Idaho and Deer Lodge, Montana are referred to as Site 1 and Site 2 respectively. Each entry in the dataset is uniquely identified with the date and time stamp. Table 1 outlines hourly average weather condition variables included in each entry and the summary of their distributions.

Variable Name	*Site*	*Mean*	*St. Dev.*	*Min*	*Median*	*Max*
Air Temperature, °F	1	42.69	19.24	-19.89	41.50	89.40
	2	41.67	20.43	-28.05	41.92	92.50
Relative Humidity, %	1	68.33	23.11	9.750	73.98	100.1
	2	63.39	21.56	8.620	66.69	99.80
Wind Direction, °azimuth	1	152.3	89.23	0.000	114.5	359.9
	2	181.1	109.9	0.000	193.0	360.0
Peak Wind Gust, mph	1	11.25	7.424	0.000	9.210	49.10
	2	11.18	7.941	0.000	8.550	55.46
Wind Speed, mph	1	7.098	5.025	0.000	5.730	33.69
	2	6.643	5.171	0.000	4.880	31.92
Solar Radiation, Lang/hour	1	15.83	23.61	0.000	0.450	154.3
	2	15.24	22.60	0.000	0.490	102.2

Table 1: *Summary of the key weather characteristics variables for both sites*

As a part of the pre-processing, it was noticed that the small percentage of data is missing, namely an entry missing on March 9th, 2014, corresponding to the date of daylight savings time change. On the other hand, there is only one entry for when the clock was moved back on November 2nd, in total constituting a negligible data loss of less than 0.1%. Otherwise, dataset appears to be already pre-processed by the distributor and no further inconsistencies were identified.

3 Analysis

3.1 Potential for Solar Generation

To begin with, from the basic overview of Table 1, it may be observed that two sites have very similar yearly weather conditions. In particular, air temperature appears to be around 42°F on average with the large variance, covering the range between -30°F and 93°F throughout the year. On the other hand, relative humidity tends to be slightly higher in Site 1, also with the high yearly variation, although both sites tend to average around 65%. Since generally heat and high humidity reduce the efficiency of solar panels, low average temperatures and relative humidity at both sites create a suitable environment for solar panels to operate at full efficiency.

For solar generation, solar radiation is the key parameter. From the yearly statistics, it is clear that average values are similar for both sites at around 15 Lang/hour, equivalent to approximately 175W/m^2. Although seemingly low, it is important to recognize that the mean value includes night hours, which drive the average down. To get more insight into this characteristic, seasonal decomposition was performed on the solar radiation time-series data for both sites, illustrated on Figures 1 and 2. A period of $24 * 7$, equivalent to 7 days, was used to produce these graphs, as the solar radiation is a slow changing process and weekly seasonality is justified.

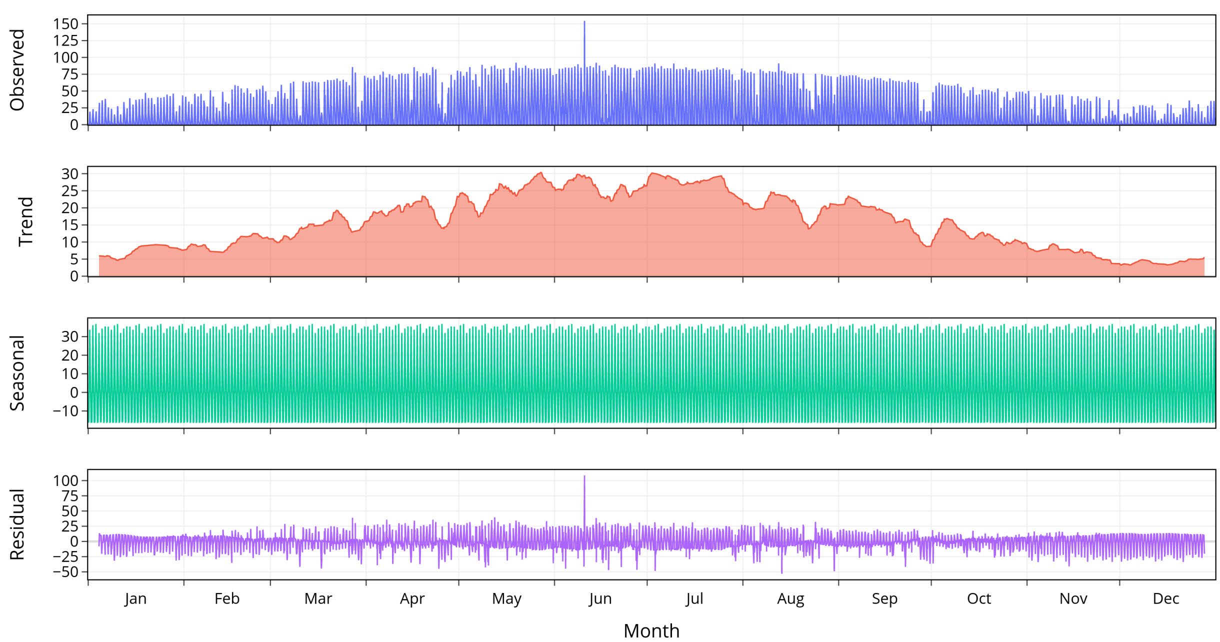


Figure 1: *Site 1: Ashton, ID, seasonal decomposition of solar radiation*

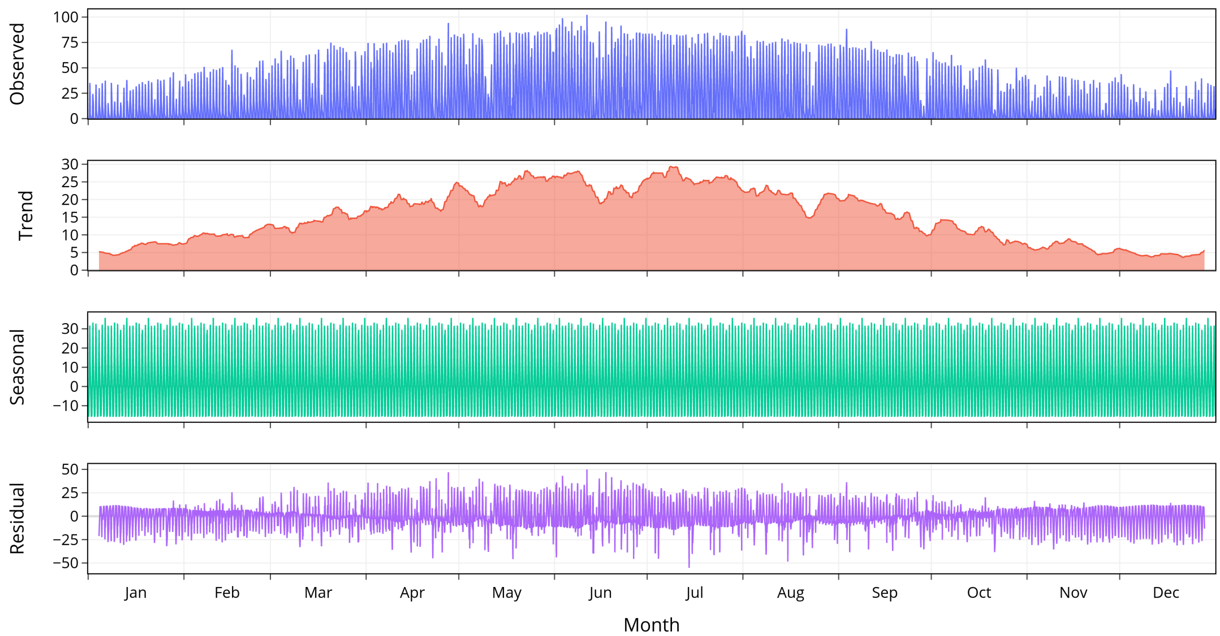


Figure 2: *Site 2: Deer Lodge, MT, seasonal decomposition of solar radiation*

From these graphs, it is again evident that both sites have very similar climate and the solar radiation trend appears almost identical. Further look into the seasonal components reveals that the daily solar radiation oscillates by more than 40 Lang/hour. Additionally, by taking the sum of the trend and seasonal components, it can be determined that at daytime, during winter months, solar radiance tends to reach 35-40 Lang/hour during the day; while during summer months it rises to 55-60 Lang/hour on average. From the residual component, it can be seen that there the residuals are generally very large, with daily peaks significantly deviating from the trend and seasonal components, particularly during summer months.

To corroborate the above findings, Figure 3 shows the distribution of peak daily solar radiation over the year. is very similar for both sites with the mean being around 60 Lang/hour. Nevertheless, Site 1 appears to have more days when solar radiation exceeds 75 Lang/hour, while Site 2 has more days with lower solar radiation (<40 Lang/hour). Based on this aspect, location of Site 1 provides a better choice for generating power with solar panels.

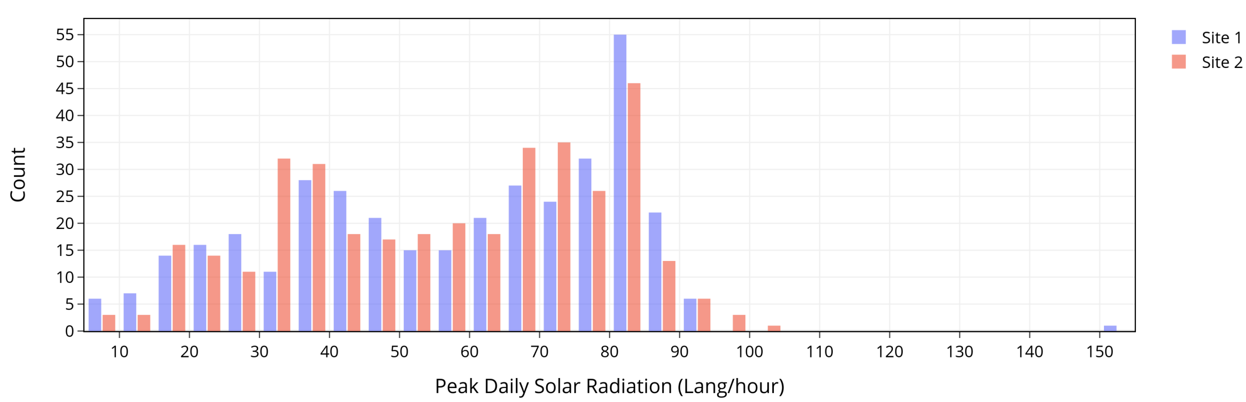


Figure 3: *Histogram showing distributions of the peak daily solar radiation for both sites*

Another important aspect to consider is the predictability of the solar radiation, as it directly affects the cost of the power delivery. In order to determine this, autocorrelation was performed with each lag corresponding to one day (i.e. step size of 24), as shown on Figure 4.

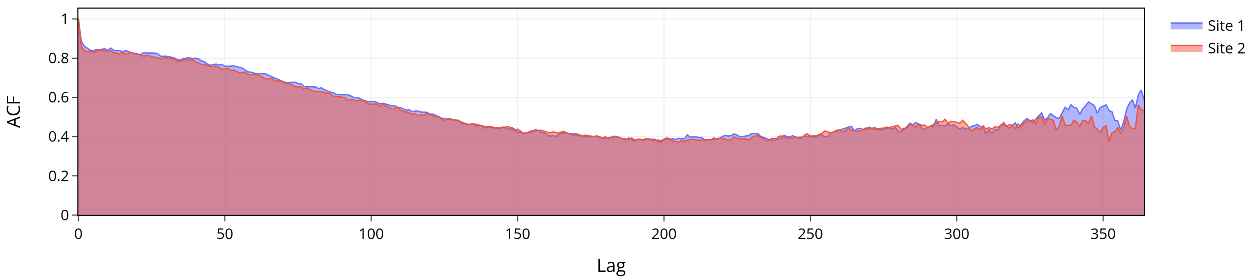


Figure 4: *Autocorrelation of solar radiation with the lag window size of 24 for both sites*

The above plot displays how well one can predict next day's solar radiation based on day 0 (January 1st), what suggests how strong is the seasonality of this time-series. Thus, from the autocorrelation plot it is clear that, as previously noted, solar radiation seasonality is very similar for both sites. Moreover, lags below 50 show very strong correlation as the lag is still not significantly far away from the winter season. With the further lags, correlation reduces to approximately 0.4, suggesting weaker correlation between winter and summer seasons, as can be expected. Overall, despite large daily changes throughout the year, solar radiation remains reasonably predictable for sites' locations, thereby making solar panels a good choice for both sites.

Looking for further ways to predict solar radiation, its correlation with the temperature and relative humidity was determined, as summarized in Table 2. As can be seen, strong either positive or negative correlation exists between all of the variables, suggesting that all of them have strong seasonality and have a well defined yearly trend that roughly matches that of the solar radiation. Accounting for all of the above weather characteristics would make prediction of the solar power even more accurate and solar power generation more appealing at both sites.

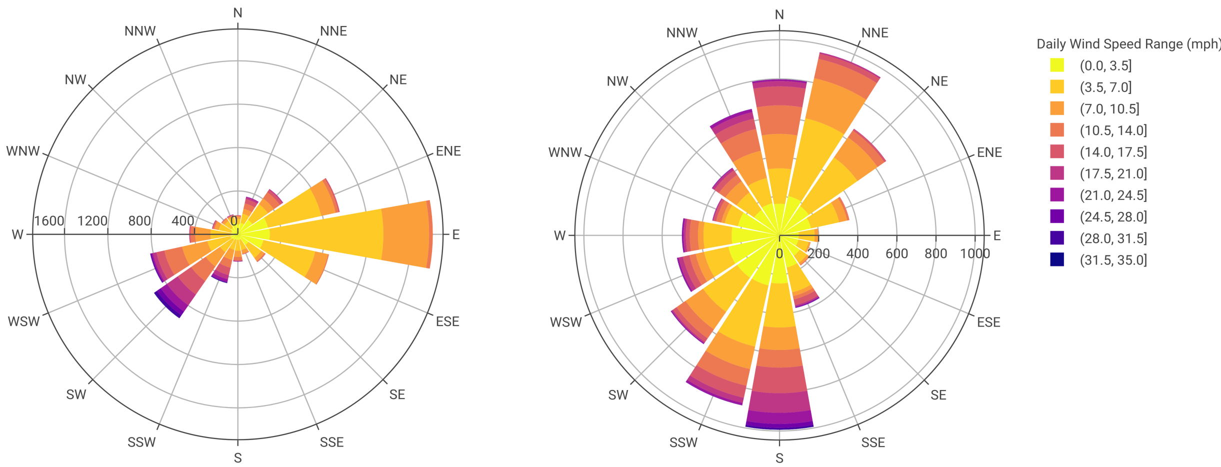
	Site 1			Site 2		
	Solar Rad.	Temp.	Rel. Humidity	Solar Rad.	Temp.	Rel. Humidity
Solar Rad.	1.00	0.51	-0.62	1.00	0.47	-0.60
Temp.	0.51	1.00	-0.63	0.47	1.00	-0.54
Rel. Humidity	-0.62	-0.63	1.00	-0.60	-0.54	1.00

Table 2: *Correlation between key characteristics for both sites (wind is not included for its lack of seasonality)*

3.2 Potential for Wind Generation

As per Table 1, yearly average wind speed is around 7mph for both sites, reaching 11mph in gusts. To further analyze wind characteristics of the sites, wind speed distribution was plotted in a form of "wind rose", as shown on Figure 5. Such plot describes frequency of each daily wind speed range in one of the 16 compass directions, thereby identifying dominant wind directions.

There, it can be observed that the prevalent wind direction for Site 1 is East and it is dominated



(a) Site 1: Ashton, ID

(b) Site 2: Deer Lodge, MT

Figure 5: *Wind speed distribution for two sites*

by the mild wind speeds. However, the consistently stronger winds are usually experienced from the South-West direction, with the average speed throughout the year exceeding 13mph. On the other hand, Site 2 appears to have a wider distribution of daily wind directions with the highest frequency coming from either South or North-North-East directions. As such, Site 2 does not display a direction with the consistently strong winds as lower daily wind speeds tend to dominate in every direction. This results in the strongest wind direction having a significantly lower average speed of 10mph. Judging from the above evidence, it is clear that Site 1 constitutes a more suitable location for generating power from the wind turbines all year round.

However, similarly to the solar radiation, predictability of the wind speed and direction is an important factor in determining suitability of the site for wind energy generation. Figure 6 and **??** reveal that there is very little seasonality in the wind speed, as it is likely dependent on factors other than the time of the of the year, such as cyclones or other local atmospheric changes.

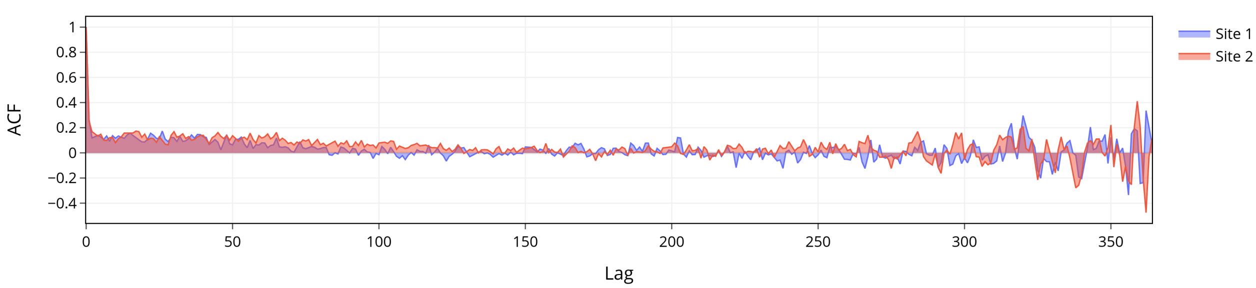


Figure 6: *Autocorrelation of wind speed with the lag window size of 24 for both sites*

Nevertheless, as was shown on the wind speed distribution plot, for Site 1, wind speed is correlated with the direction its coming from, particularly for the higher ranges. To assess this, the correlation table was generated for each site for the wind characteristics. From Table 3 it may be observed that correlation between wind speed and wind direction is significantly higher for Site 1, although in both cases correlation being rather weak. On the other hand, as could be expected, correlation between wind gusts and wind speed is almost perfect, since wind gust speed will always be higher with higher average wind speeds. However, for the purpose of wind power generation, wind gusts do not play as much of a role and so the relationship between two variables is not useful for predicting future wind speeds.

	Site 1			Site 2		
	Wind Dir.	Wind Gust	Wind Speed	Wind Dir.	Wind Gust	Wind Speed
Wind Dir.	1.00	0.29	0.23	1.00	0.07	0.08
Wind Gust	0.29	1.00	0.97	0.07	1.00	0.96
Wind Speed	0.23	0.97	1.00	0.08	0.96	1.00

Table 3: *Correlation between key wind characteristics for both sites*

4 Conclusion

To summarize, from the performed analysis, it was found that weather characteristics are very similar for both sites. Generally low temperatures and relative humidity throughout the year would increase the efficiency of the solar panels at both locations. At the same time, slightly higher daily peaks at Site 1 make it a better location for solar power generation, although both sites have a good potential for this power source with plenty of high solar radiation days. On the other hand, wind characteristics vary more between sites, with Site 1 having a better defined prevalent direction of the wind, what makes it a better choice for wind generation than Site 2. Additionally, yearly average speed in the prevalent direction is 13mph, what would likely generate more power than the equivalent solar power generator. Thus, it was concluded that the preferred technology for Site 1 is wind, while for Site 2 other factors would determine if solar or wind should be selected, as both of these renewable sources would suit the site reasonably well.