Springboard Capstone Project 1 – Data Storytelling
Predicting Short Term Solar Energy Production

Connor McAnuff

July 22, 2019

1. Overview

This report discusses the investigation into the total daily solar energy availability (referred to as “energy” in this report) and how it relates to Mesonet Station location, elevation, time of year, and the weather forecast variables. For the sake of brevity, the investigative visualizations are not included in this document. This report should be read in conjunction with the accompanying Jupyter Notebook.

2. Investigation

Please refer to the document “Capstone Project 1 – Data Wrangling – Connor McAnuff.pdf” for an explanation of the raw data and data processing. The raw data is available for download from Kaggle. The raw data includes time-series and spatial data, thus providing many opportunities for investigation and visualization. The following questions were asked of the data:

- What are the distributions of Mesonet station location (lat/long) and elevation?
- How do the energy measurements vary over time (month-to-month, year-to-year) for a single station and for all stations combined? For all stations combined, how variable is the data for a given month/year?
- What is the distribution of total energy by station location?
- What are the differences in the energy data of stations in the east of Oklahoma vs the west of Oklahoma?
- How do the total/average of weather forecast variables vary in space and time and how do they relate to the energy?

3. Results and Discussion

3.1 Mesonet stations

The stations are spread evenly across Oklahoma (Figure 1). The state spans approximately 8.5° of latitude and 3° of longitude. The difference in elevation from the lowest (110 m) to highest (1322 m) station is 1212 m. The elevation of the stations steadily increases from east to west. This trend is indicative of Oklahoma’s variable geography. The eastern side of Oklahoma is lower and contains extensive forested areas, central Oklahoma contains a transition from forest to prairies, and in the north-west there are flat grasslands.

3.2 Energy over time and space

Visualizing the daily energy data for a single station from 1994-2007 (alongside a 14-day rolling average) on a time-series scatterplot reveals that it is cyclical (Figure 2). Energy peaks yearly during June/July and troughs during December/January. This trend is expected, as Oklahoma is significantly north of the equator, thus from December 21 to June 21 of each year, the sun’s time and height above the horizon increase (and vice versa). Year-to-year, the data does not appear to vary significantly. The energy in the winter months has a consistent maximum but also appears to be more variable than the summer months. Additionally, there appears to be significant variability in the data within short periods of time along the entire 14-year span.

Comparing the yearly and monthly energy totals for all stations combined, it is seen that there is little variation year-to-year in terms of yearly totals and monthly trends (Figure 3). The combined station monthly data shows the same cyclical trend as the single station data. Yearly distributions shown by box and whisker plots and violin plots show that the yearly distributions are very similar year-to-year (Figure 4). 1994 appears to have a greater spread than any other year – this is due to a single datapoint from 1994-04-07 at IDAB Mesonet station.

The monthly distributions (1994-2007) allow the cyclical trend to be further explored (Figure 5). The median value peaks in June/July and troughs in December/January. June-September have fewer low values than the remaining months - low values are shown as outliers as opposed to the box whiskers extending to 0. Overall, the summer months have a distribution heavily concentrated at the peak, with a long tail extending to 0, while the winter months are bimodal, having a peak at the top and a smaller peak at the bottom of the distribution. The bimodal distribution is most prevalent in November-January.

The total energy available at each station from 1994-2007 has been plotted on the Oklahoma map to analyze the distribution of energy across the state (Figure 6). There is a clear trend of increasing energy from east to west. As discussed above, the geography (and therefore, likely climate) changes from east to west.

Stations were sorted into two groups, those east of 98°W, and those west of 98°W. The stations in the west had greater average energy for all years (Figure 7). The differences in total energy and energy distribution between the two groups appears to be consistent for all years (Figure 8). The violin plots show that there are concentrations of energy values at 2.75e7 J/m^2 (summer peak), 1,25e7 J/m^2 (winter peak), and 0.25e7 J/m^2 (winter minimum).

3.3 Weather forecast variables

Weather forecast variables have been visualized by summing or averaging the forecast values from 1994-2007, using the median value of the 11 predictive models, and contouring over a map of Oklahoma. On the same plot, the total energy for each station is shown to explore relationships between the weather forecast variables and energy.

The total forecast precipitation contours show that it increases from west to east (Figure 9). The east side of Oklahoma had an order of magnitude greater precipitation forecast than the west side from 1994-2007. Unsurprisingly, the stations to the east had less energy available in the same time frame relative to the stations in the west. The same trend is seen with average forecast cloud cover percentage (Figure 10).

Total forecast downward short-wave radiative flux at surface generally increases from east to west, with a slight rotation to increase to the south (Figure 11). The radiative flux at surface is likely a function of other weather variables such as precipitation and cloud cover. The radiative flux appears to correlate with energy.

Lastly, average forecast surface temperature increases from north to south (Figure 12). It does not appear to be strongly correlate with energy.

Total forecast precipitation varies year to year (Figure 13). 2004 and 2007 had the most forecast precipitation – these two years had the lowest combined energy, signalling a correlation. Totalling the monthly forecast precipitation (1994-2007), May and June have had the most forecast precipitation (Figure 14).

3.4 Conclusions

The Mesonet stations are spread evenly across the state of Oklahoma. The elevation and other geographical/topographical factors vary significantly in the east and west of the state. Stations in the west are over 1 km higher than stations in the east.

The energy data does not appear to vary significantly year-to-year. It is cyclical, as expected, peaking in summer months and troughing in winter months. Within the overall cyclical trend, there are significant short time span variations, with the summer months having less variability. Stations in the west have higher energy availability than stations in the east.

Total forecast precipitation and cloud coverage percentage increase from west to east, showing a correlation between higher forecast precipitation/cloud coverage and lower energy. The short-wave radiative flux has the opposite relationship, showing a correlation between higher short-wave radiative flux and higher energy. Forecast surface temperature does not appear to have strong correlation with energy. This initial investigation into the weather forecast variables shows promising prospects of predicting energy availability using these features.