Forecasting Methods 6-638-18A
Group Project – Team 5.

Felipe Gonzalez 11246645
Junyuan Lin 11253404
Blosher Brar 11258419
Miao Hu 11255049

Part 1: Forecasting North Illinois Regional Electricity Demand

Introduction

In today's modern era, electricity has become a necessity for everyday life. Everything from watching the latest Netflix release to launching the next NASA mission requires the use of the electricity grid. Moreover, with a rising population, a transition to electric automobiles, and an ever-evolving economy, being able to accurately forecast electricity demand can play a vital role in helping utilities create better service for their customers. Although one cannot expect forecasters to predict with exact precision, having a reasonably accurate forecast and managing risk can help avoid costly instances of rolling blackouts. In fact, this is a daily occurrence in certain countries like Pakistan, where millions of people are left without power for as long as 10 hours a day (Jillani, 2011). Even in Canada and the United States, rolling-blackouts can occur due to severe weather causing unexpected power shortages (Krishnan, 2016). On the other hand, having higher than needed electricity output can cause wasteful expenditure on unnecessary power generation projects.

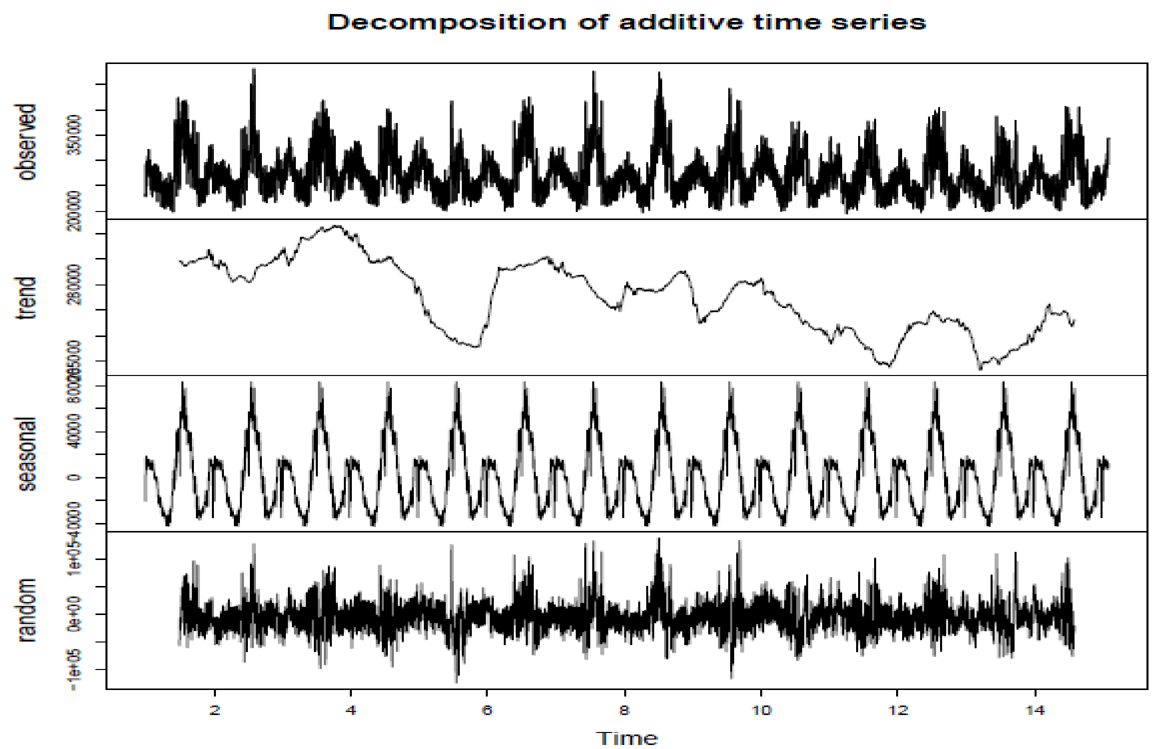
This paper seeks to investigate electricity forecasting using data from PJM (Pennsylvania-New-Jersey-Maryland Interconnection), which is an RTO (Regional Transmission Organization) that manages the electricity grid in the Northeast Region of the United States. It is a non-government organization, albeit still regulated by the Federal US government, that seeks to promote economic efficiency and reliability of the wholesale electricity market in its region. This paper will explicitly deal with one region within PJM's management, designated as ComEd (Commonwealth Edison) and located in North Illinois. The mandate of this paper will be to forecast future electricity demand with a lag h, equal to one day. In other words, we wish to use historical data to predict the electricity demand for the next day. This data can include both electricity demand as well as other related variables such as temperature (HDD and CDD), humidity, or wind speed. By studying previous research on electricity demand forecasting, we found that as temperature increases, electricity demand also increases due to the use of Air Conditioning and other appliances. Similarly, when temperature decreases, electricity demand also increases due to an increase in heating. Both Fall and Spring have decreased electricity consumption due to minimal heating and cooling needs. Our expectation is that our forecasting will generally become more accurate as the complexity of our model increases, however, we will be using an AIC (Akaike Information Criterion) to find the best trade off between accuracy and complexity.

Consequently, to begin our investigation, we began by aggregating the hourly electricity data from 2012 to 2018 into daily electricity data to make the data relevant towards our end goal. After carrying out some exploratory data analysis, we discovered some seasonality and found the general trend to be similar for most of our dataset. After 'cleaning' our dataset, we used a naïve method to get a benchmark evaluation to compare with other more complex forecasting methods to come in Part 2 of this paper.

Exploratory data analysis

In the dataset downloaded from PJM, the time used is EPT (Eastern Prevailing Time), however, the north Illinois area that ComEd operates are using Central Time Zone (CT) which is one hour late behind EPT. We have adjusted it by making all our data one hours earlier. Also, since EPT has Daylight Saving time, so there will be one hour repeated in the second half year each year and one hour vacant for the first half of each year. We remove the second repeated hour in the

Winter and estimate the vacant hour in Spring using linear method, which consider more about the most recent hours near the vacant one. Since people would treat the hour before the extra hour normally, but the extra hour would be treated like extra sleeping time.

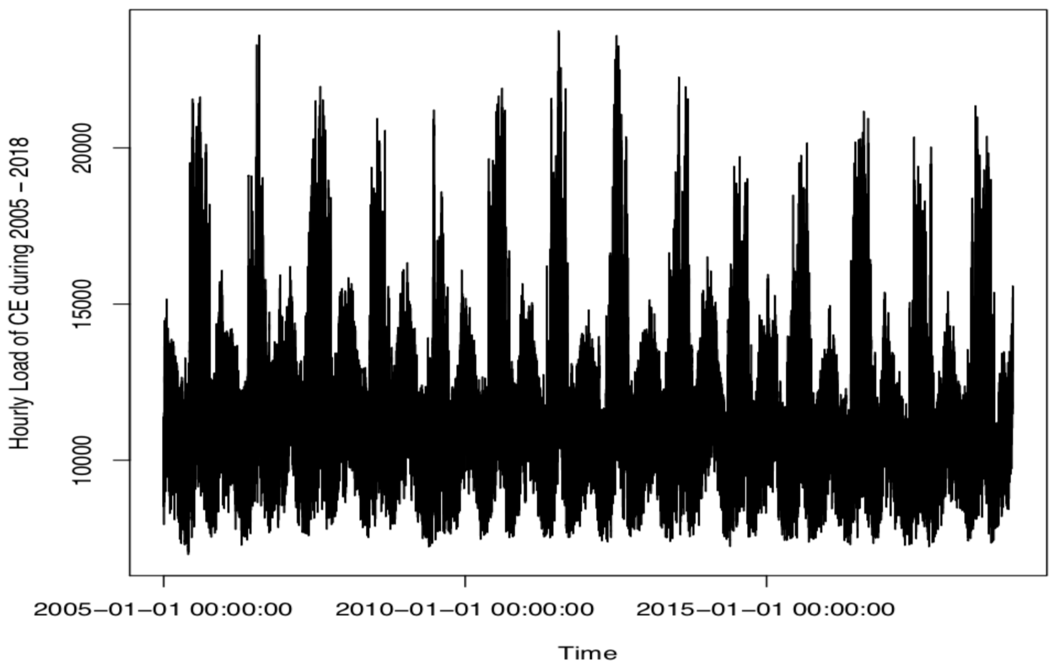


We do decomposition of additive time series based on one year frequency. The graphic is very typically. Graphic line ups and downs are regularly searchable. Besides, the graphic shows a declining trend.

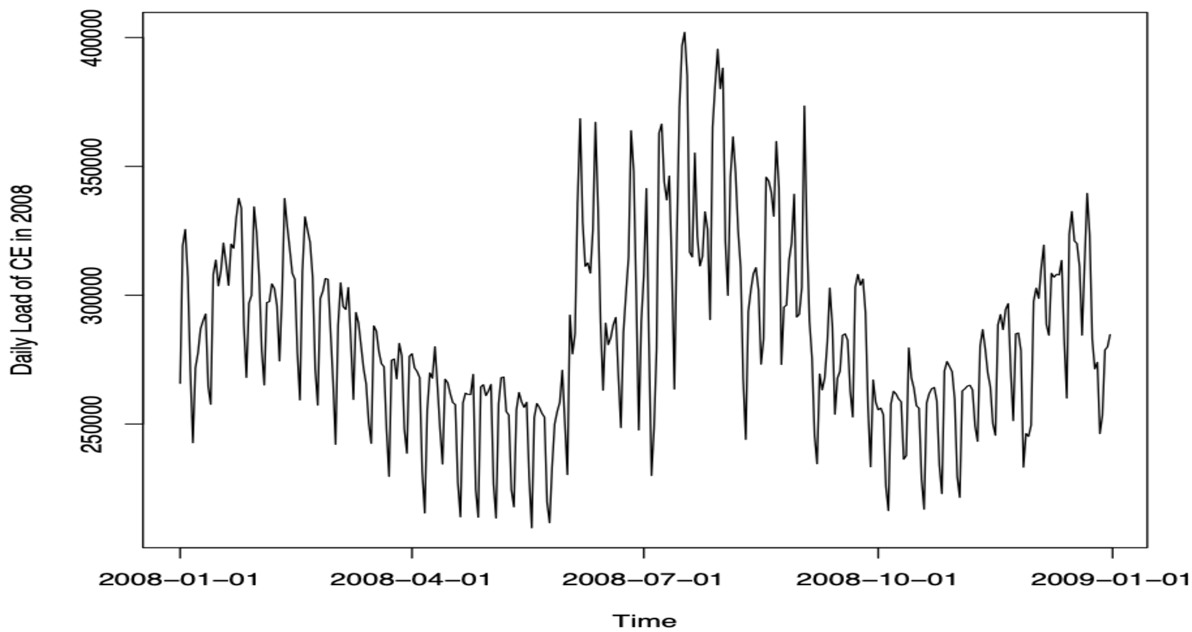
There are seven outliers in our data separated in three different groups. Matching the indexes to the dates we observed a similar phenomenon that increased the demand for energy:

- July 31,2006, August 1, 2006, August 2, 2006: There was an increase in the temperature with maximums reaching 99 °F combined with very few rain.

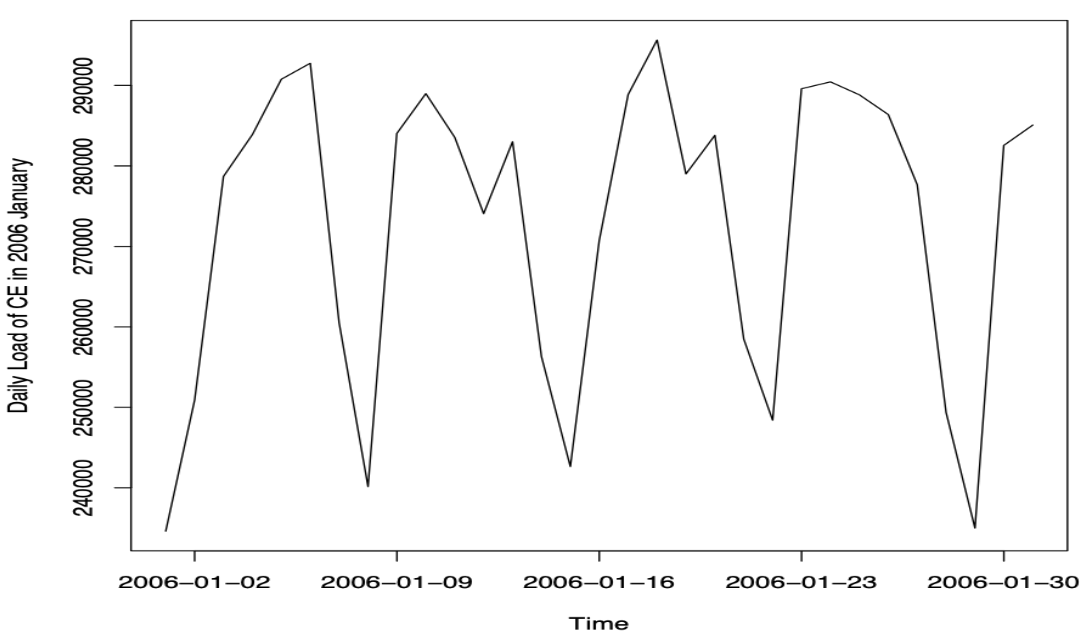
- July 19, 2011, July 20, 2011, July 21, 2011: During these three days there was an increase in the temperature with maximums reaching 99 °F, causing a urban island effect, especially in the city of Chicago.
- July 4, 2012, July 5, 2012, July 16, 2012: There was sustained increase in the temperature with maximums rounding the 100 °F for these three days and the last 6 preceding days.

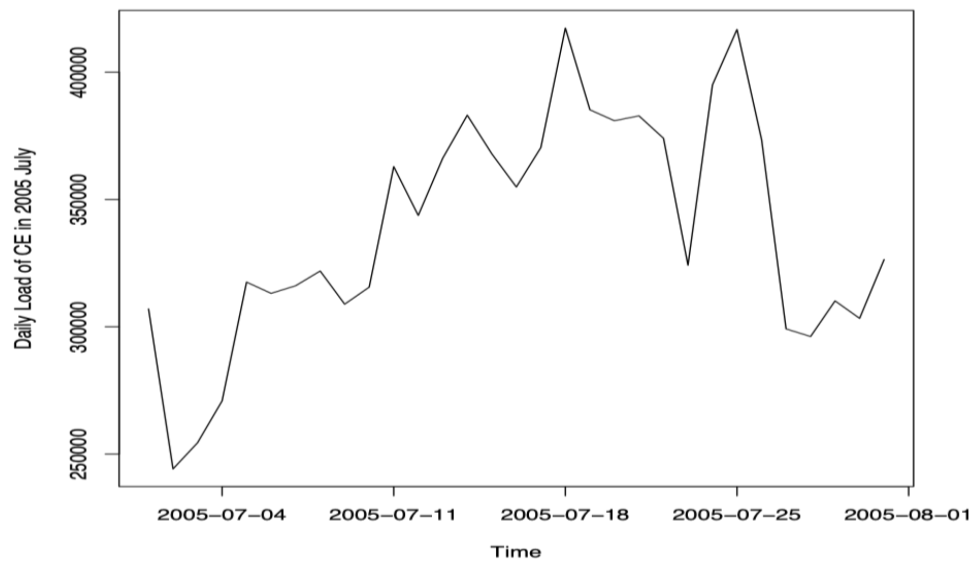


We did daily and hourly load of CE during 2005-2018. The seasonal and trend are not quite clearly on this graphic. Graphic lines are more evenly undulating and there is no extremely outstanding undulating through those years.



Then, we did daily load of CE for each year. It shows clearly trends that every summer and winter, it has a upward trend and go down when it turns to spring and autumn. For example, Daily Load of CE in 2008. The electricity usage start increasing from at the end of May and reaching peak in August. Then, it goes down in the autumn and goes up in the winter.





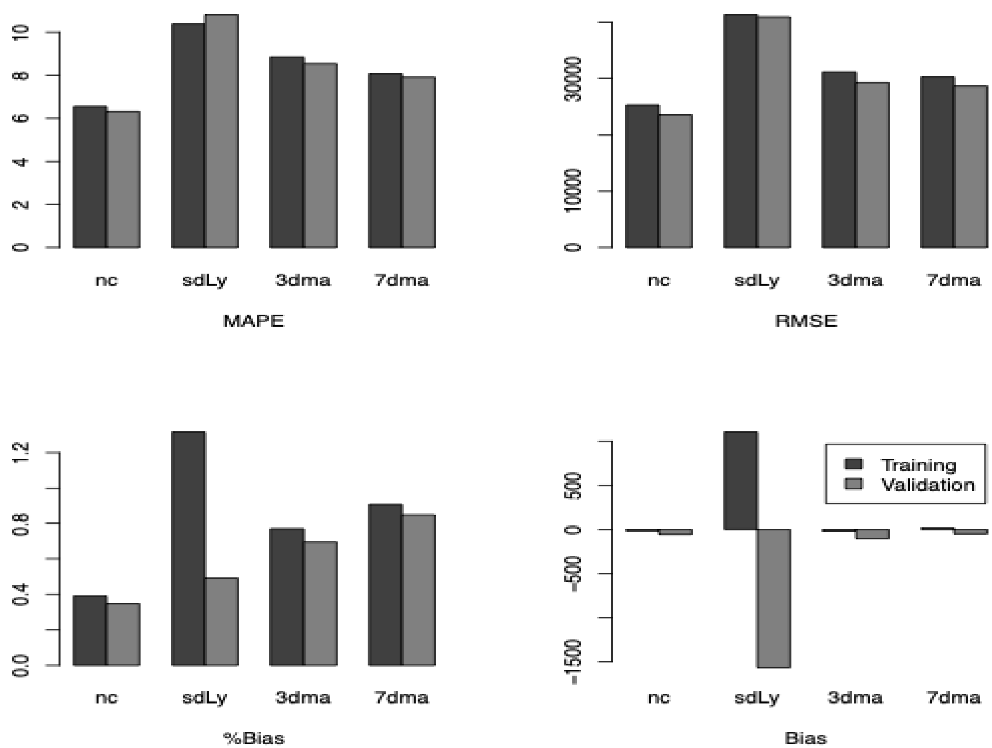
To have a deeper look of that trend, we randomly pick daily load of CE in several years (2005, 2006, 2011, 2012, 2017, 2018). We use July as the representative for summer and January as the representative for winter. We find that in winter, the seasonality is one week. Usually, it has a peak in the middle of a week. In summer, the highest electricity usage is in the middle of the month.

Naïve method

To make a better understanding of our data, we divided the dataset into 2 parts: the first 10 years (2005-2014) as training data, and the latter 4 years and a month (2015-2019/1) as validation data. In our project, we select several naïve methods ("nc": no change naïve, "smLy": same day Last year, "3dma": 3-day Moving average, "7dma": 7-day Moving average.) as the baselines to evaluate more complex methods. In order to obtain the performance of those methods, we used MASE, RMSE, Percentage bias and bias as the measurement on the two different datasets that we split: "Training" and "Validation", the result is given in the graph and table below. As we can see that No Change Naïve Method performs the best overall. As for same day last year, we can see that the bias for it are too large.

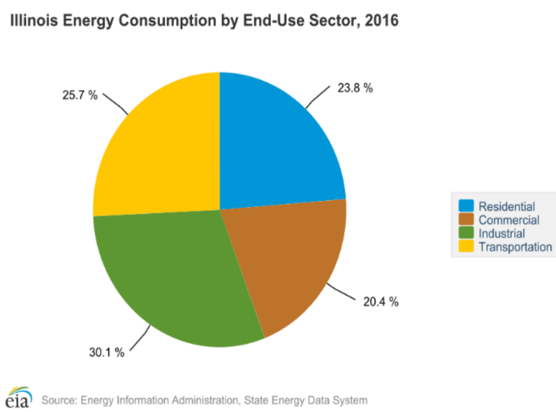
Training	MAPE	RMSE	%Bias	Bias
nc	6.56	25317.58	0.39%	-15.52
smLy	10.39	41338.43	1.32%	1100.76
3dma	8.85	31132.74	0.77%	-19.59
7dma	8.07	30273.47	0.91%	16.36

Validation	MAPE	RMSE	%Bias	Bias
nc	6.32	23559.87	0.35%	-56.48
smLy	10.82	40930.96	0.49%	-1567.14
3dma	8.55	29242.62	0.70%	-104.92
7dma	7.91	28642.66	0.85%	-49.83



Explanatory variables

Our region is the ComEd region located in Northeast Illinois which includes Chicago, the third largest city in U.S with a population of nearly three million inhabitants [1]. Chicago and its surroundings are a very industrialized area with high population density, consequently, the state ranks in the 5th position of electricity generation among all states[2]. In order to get a sense of the electricity consumption across the state we can observe the following graph from the US Electricity Information System.



As we can see on the graph, there exists four major sectors that consume electricity in a somewhat evenly way. It is worth noting that the industrial sector is the one that consumes the largest amount.

Off course this distribution changes depending on the area that one looks at, for instances, Chicago is much more residential than industrial. However, we consider that the consumption across the whole state is much more representative of the consumption across the ComEd region.

The electricity demand for a determined area whether it is the city or outside, results from a complex process determined by the interactions of multiple variables. Typically, the electricity demand is influenced by the climate component, seasonal and socio-economic conditions [3].

To see how the climate component is influencing the demand we are considering variables related to the weather such as temperature, rainfall and humidity. Although some of these

---

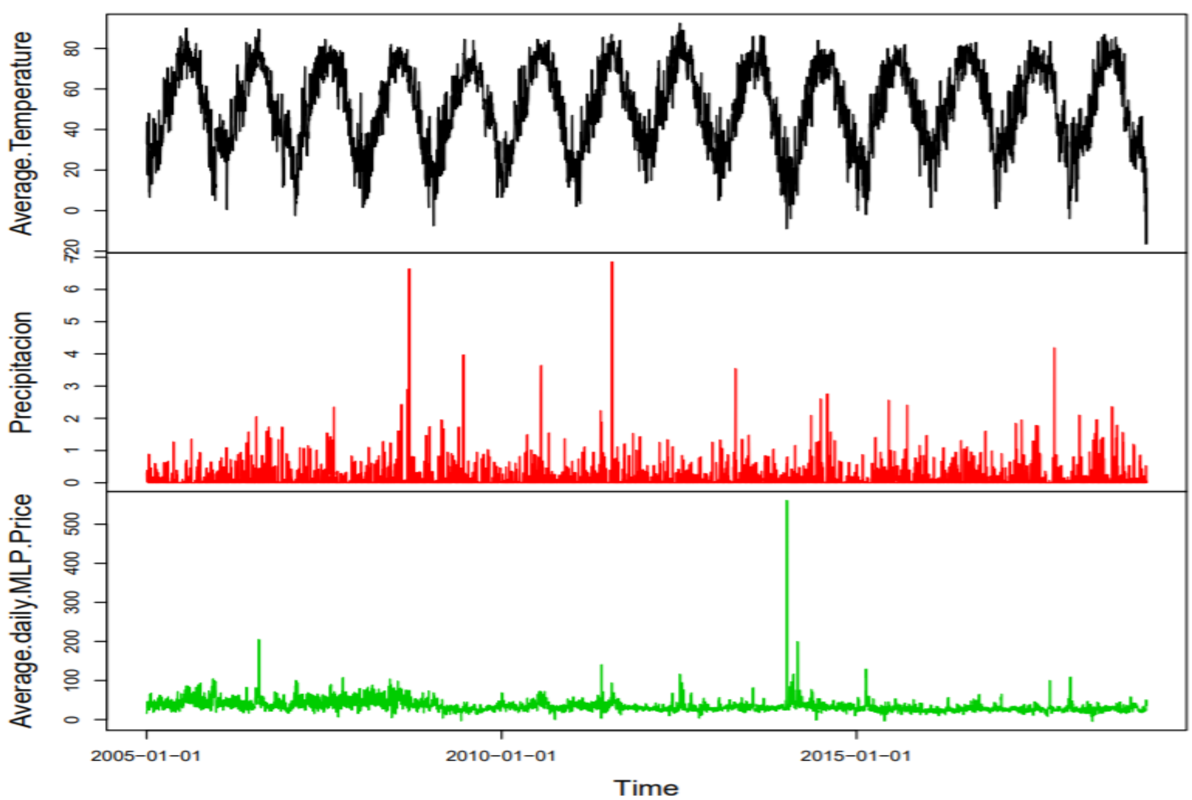
[1] City of Chicago
[2] U.S Energy Information Administration
[3] Schellong, W. (2011). Energy Demand Analysis and Forecast

explanatory variables might appear obvious, the interactions of this variables with economic variables such as price, GDP growth and population size are rather complex. Even more, a given zone can be influenced as well by calendar data that contains special dates such as holidays and different types of cultural and religious events.

As is to expect, there is no perfect model that can consider all possible variables and their combined effects. For the purpose of this project, we are going to limit ourselves to the explanatory variables that can be measured on a daily basis. That means data that can take the form of a time series and that can be useful when implementing a multiple linear regression. Thus, we won't be considering economic and demographic data.

Climate variables.

Taken from the National Weather Service Forecast Office for the period (Jan.2005 to Jan.2019) and the Weather Underground website.



Temperature is probably the most correlated variable that might exist with electricity demand. Given that we are dealing with a zone located on the mid-west USA at a considerable latitude to the north, annual weather has very marked seasons. The temperature is going to be a fundamental variable to explain the main seasonal component. On residential areas for example, the needs for cooling and heating can vary significantly from one geographical location to another.

Included in this category, we will also be taking into account the CDD and HDD, these are measures that reflect the number of celsius degrees that a day's average temperature is above or below a reference temperature (65 F) and can explain easily the needs for cooling and heating.

Max °F	103	July 6, 2012
Min °F	-23	January 30,2019
Average °F	50.68	
SD	20.1908	

Precipitation, humidity and wind speed: These variables play an important role as well since they have a direct effect that can substantially affect energy demand; The precipitation can have a cooling effect which is especially notable during summer. Given the proximity to the lake Michigan, wind speed and humidity also play an important role, they can have a cooling or a heating effect in the areas near the lakeshore.

Calendar data is an important source of information when making forecasts. We have seen in the time series graph that the weekdays have a different behavior than weekends, the later usually has a lower demand given the absence of many industrial and business users. Also, the consumption varies greatly when we look at special dates like Christmas and NYE's or public holidays. For this reason, we decided to include calendar data as one of the explanatory variables.

Economic variables - Taken from PJM Energy Market. ComEd Region

Energy price is one of the economic variables that has the closest relation with demand. Given that the energy market is deregulated, the price to the final consumer is variable; This price changes on an hourly basis for each of the main distribution nodes across the ComEd territory. In our zone, we identified three main nodes: Wilton, Plano and Collins. However, given that our purpose is to measure the changes of the price across time, we decided to take only one of these nodes as reference for the price (Wilton). It could be the case that large businesses and industrial consumers have a large price elasticity demand if they have access to other sources for energy generation.

The price that we used corresponds to the real-time hourly PJM price which is a price used to price energy purchases and sales in the PJM market. The LMP accounts for three components which are

i) System Marginal Price: Represents the price of electric energy.
ii) Congestion Component: Represents the price of congestion across the system.
iii) Marginal Loss Component: Represents the price of transmission losses.

Thus, we get that LMP = System Marginal Price + Congestion Component + Marginal Loss Component. These factors will give us a final price expressed in Dollars per MWh.[4]

[4] PJM. Location Marginal Pricing Components.

References

City of Chicago. (2019). *The city of Chicago's official site*. Retrieved from Facts and Statistics.

National Weather Service Forecast Office. (2019). Retrieved from Chicago, IL: https://w2.weather.gov/climate/xmacis.php?wfo=lot

PJM. (2017). *Locational Marginal Pricing.*

PJM. (2019). *Energy Market*. Retrieved from Real-Time Hourly LMPs: https://www.pjm.com/markets-and-operations/energy.aspx

Schellong, W. (2011). Energy Demand Analysis and Forecast. *Energy Management Systems. University of Cologne, Germany.*

U.S Energy Information Administration. (2019). *Illinois, state profile and energy estimates*.

Weather Undergroung. (2019). *Chicago O'hare International Station*. Retrieved from Historical: https://www.wunderground.com/history/monthly/us/il/chicago/KORD/date/2019-2

Wikipedia. (2019). Retrieved from Climate of Chicago.

Jillani, S. (2015, June 15). *Power cuts hit Pakistan's economy*. Retrieved from https://www.bbc.com/news/av/world-asia-pacific-14164825/power-cuts-hit-pakistan-s-economy

Krishnan, S (2016, August 17). *Rolling blackouts in Montreal as Hydro-Québec recovers from storm damage.* Retrieved from https://www.cbc.ca/news/canada/montreal/hydro-outages-montreal-rolling-blackouts-1.3724370