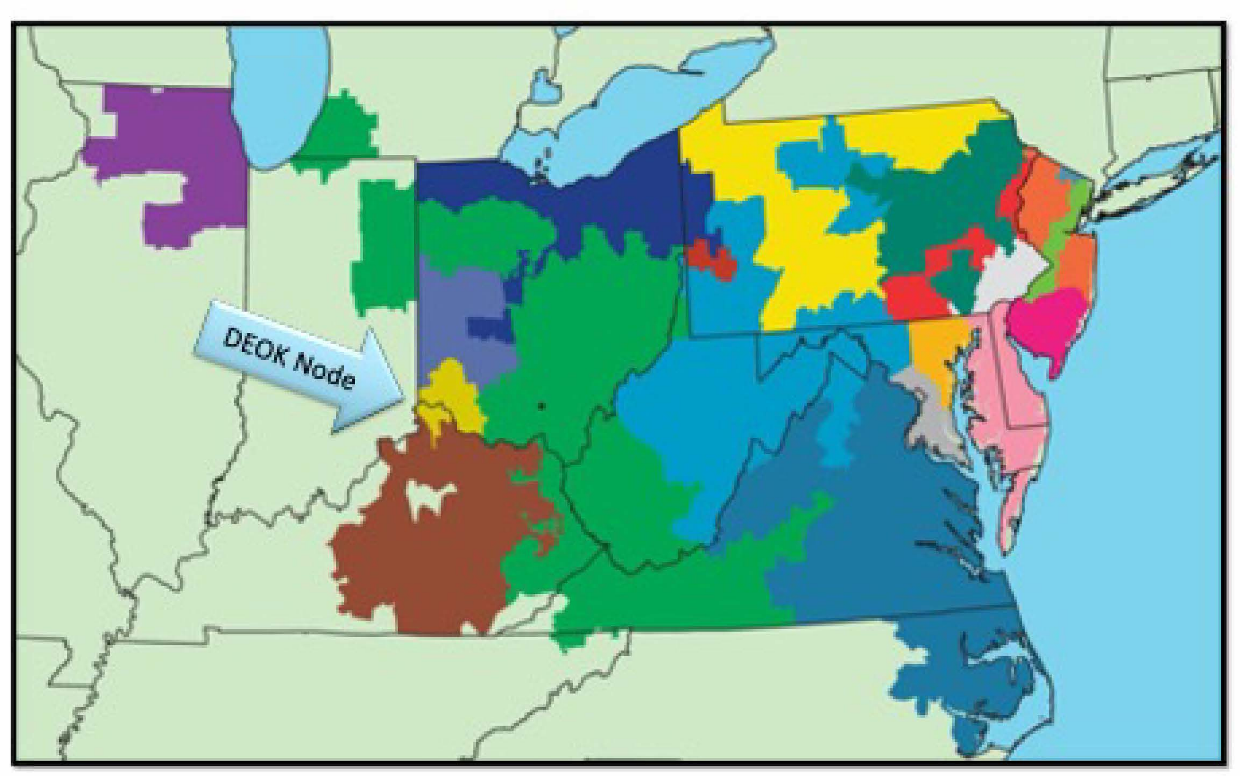
Modeling Real Time Prices for Energy Markets

Erin Byrd, Kyle Cohen, Nhi Nguyen, Khanh Vy Nguyen
University of North Florida, Coggin College of Business Jacksonville, FL USA 32224

INTRODUCTION

Electricity is a unique commodity because supply and demand must be continuously balanced to have a safe and reliable grid. Electricity generation, transmission, and consumption are nearly instantaneous. Wholesale energy prices are expressed as the price for a megawatt of power to flow for one hour. The two main prices in energy markets are day-ahead prices, quoted a day before the desired time of energy flow, and real-time prices, quoted an hour before the desired time of energy flow. Real time prices have more updated information about market factors. We developed and compared models to predict real time prices at the Duke Energy Ohio Kentucky (DEOK) price node of the Pennsylvania, Jersey, Maryland Power Pool Region (PJM). The ability to predict the next hour of real time price allows power traders and grid operators to effectively control costs and optimize resources.



A map of the PJM service area, highlighting the Duke Energy Ohio Kentucky (DEOK) region that is our area of study.

OBJECTIVES

The objective of this project is to forecast the real-time price of electricity in the wholesale market using a variety of regression modeling approaches to see which model has the best fit to the data and the least amount of error. The project also examines different combinations of influencing factors to maximize the accuracy of the models.

MODELS & DISCUSSION

We gathered raw data from the PJM Region for the years 2013 – 2018 containing real time prices, day ahead prices, and variables that could influence those prices, such as: system price, congestion price, marginal loss price, temperature, and megawatt load and generation. We cleaned and aggregated the dataset and created lag values to use in our models.

Model 1: Neural Network Regression - Data prior to 2018 was used to train the model. Variables selected were system, congestion, loss price, temperature and megawatt load, with one lag value.

Model 2: Boosted Decision Tree Regression – Built on the same set of contributing factors as the Neural Network.

Model 3: Linear Regression A - Data prior to 2018 was used to train the model. Variables selected were temperature, fuel type generation, megawatt load, and system, congestion, and loss prices to predict real time price, with two lag values.

Model 4: Linear Regression B – The real-time price was regressed on the lag of the day ahead price, current temperature, current MW load, and current generation by fuel type of the PJM region.

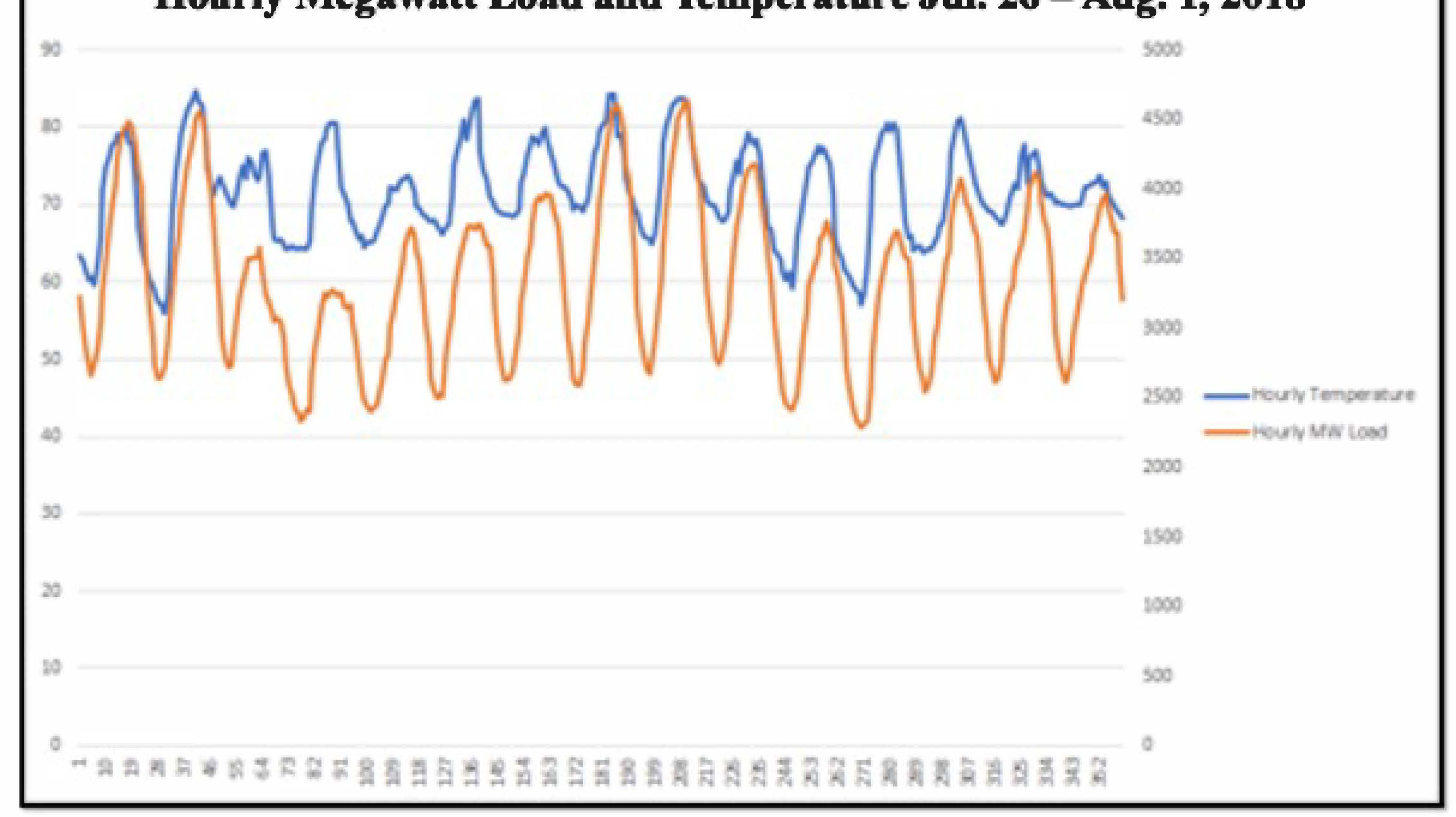
Model 5: Seasonal ARIMA(5,1,2)(0,0,2)[24] - This ARIMA was the best fit as found by the auto.arima function in R. It includes slight adjustments for seasonality defined as 24 hours per seasonal period.

In the chart to the right, we have shown temperature and megawatt (MW) load over the latest available week of data. This shows that temperature and MW load follow a similar pattern. Temperature (in blue) is shown on the left axis. MW Load (in orange) is shown on the right axis.

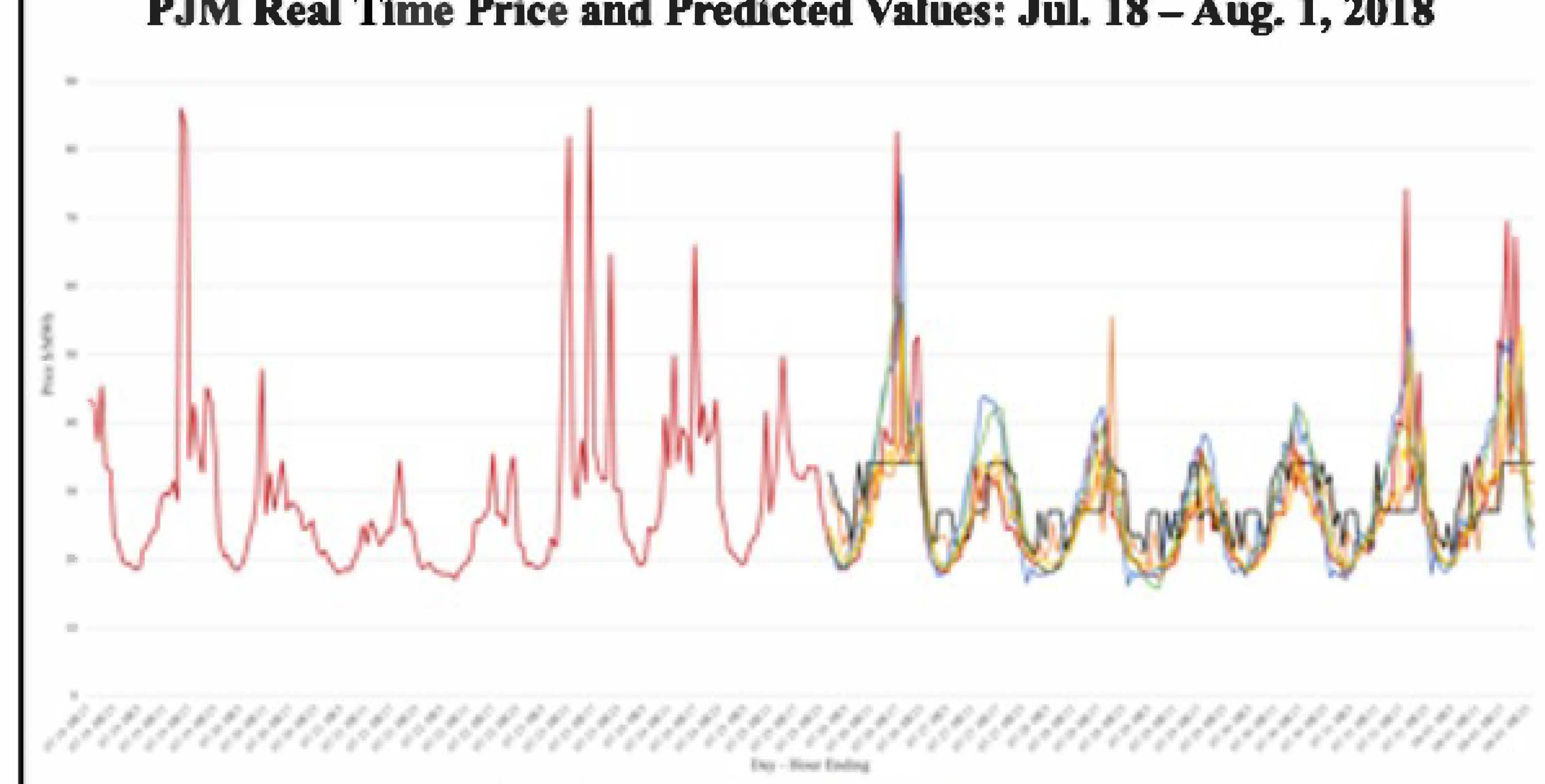
The next chart shows two week of data. The first week shows only the actual values of real time price (RTP), in red. The following week shows actual RTP as well as the predictions from each model.

After constructing our models, we looked at the weights or coefficient given to each variable. Marginal generation regardless of fuel type puts downward pressure on price, but marginal oil generation shows a significant positive correlation with price. As temperature decreases, price increases because people tend to use more power when it is cold. Megawatt load has a strong positive influence on price.

Hourly Megawatt Load and Temperature Jul. 26 – Aug. 1, 2018



PJM Real Time Price and Predicted Values: Jul. 18 – Aug. 1, 2018



RESULTS

The results shown below from each model were calculated on the errors of last week of data available, July 26 to August 1, 2018.

Model	1. Neural Network Regression	2. Boosted Decision Tree Regression	3. Linear Regression A	4. Linear Regression B	5. Seasonal ARIMA(5,1,2)(0,0,2)[24]
Mean Absolute Error	5.11	6.44	4.96	4.12	3.67
Root Mean Squared Error	8.29	9.35	7.80	6.84	7.24
Coefficient of Determination	0.372	0.201	0.444	0.594	N/A

CONCLUSION

To compare the models, we looked at measures of fit and measurements of error as shown in the results table. Of all the models, the lowest Mean Absolute Error was 3.67 found in the Seasonal ARIMA model. The lowest Root Mean Squared Error was 6.84 found in Linear Regression B. Pairing our findings with business knowledge from the industry, we believe that Linear Regression B will be the best model to implement and utilize when predicting real time energy prices. Although we focused our study on the DEOK price node in the PJM region, we believe this model can be useful for other regions as well to project prices to best conserve costs and allocate resources.

DATA & RESOURCES

1. https://www.pjm.com
2. https://www.ncdc.noaa.gov/cdo-web/datatools/lcd
3. https://gist.github.com/nk773/af9bb2caf92bb23b95bac87 3312a5269
4. https://www.pjm.com/-/media/committees-groups/committees/mc/20190425-somr/20190425-2018-imm-state-of-the-market-report-presentation.ashx
5. https://otexts.com/fpp2/arima-forecasting.html

UNIVERSITY *of* NORTH FLORIDA