



Model Trees for Electric Load Forecasting

Abby Martin and Richard Povinelli

GasDay Lab and Marquette MSCS REU Program funded by the National Science Foundation

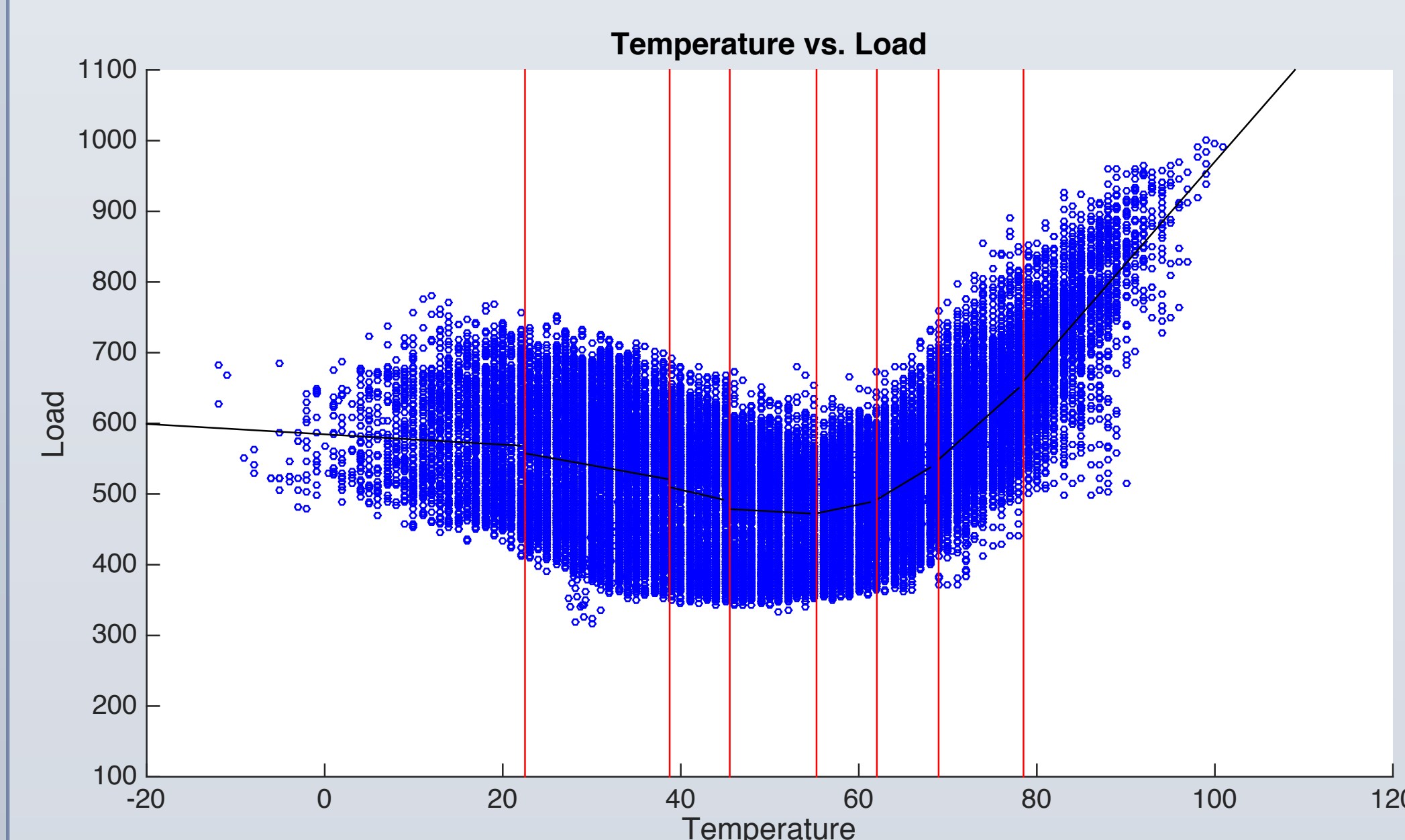


Introduction and Objective

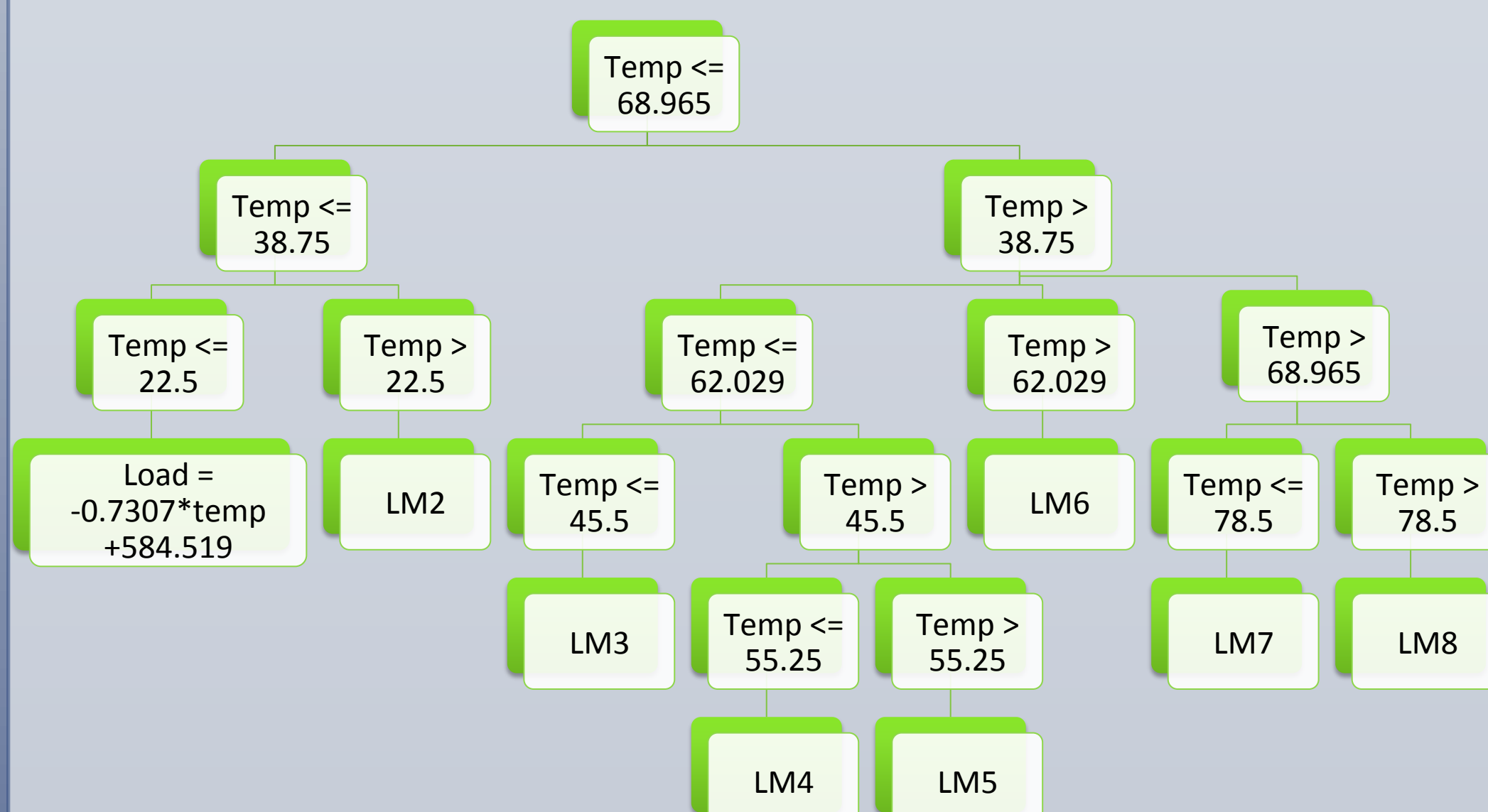
Predicting electric loads is vitally important to reduce overloading, blackouts, and power failures. The goal of this research is to develop a forecasting method using the model tree that is as good or better than the artificial neural network currently being used. The methods will be compared using the statistical measurement of MAPE; the artificial neural network has a MAPE of approximately 2.4%.

Methods

A model tree is a type of decision tree which splits large amounts of data into smaller categories to create several linear regression models.

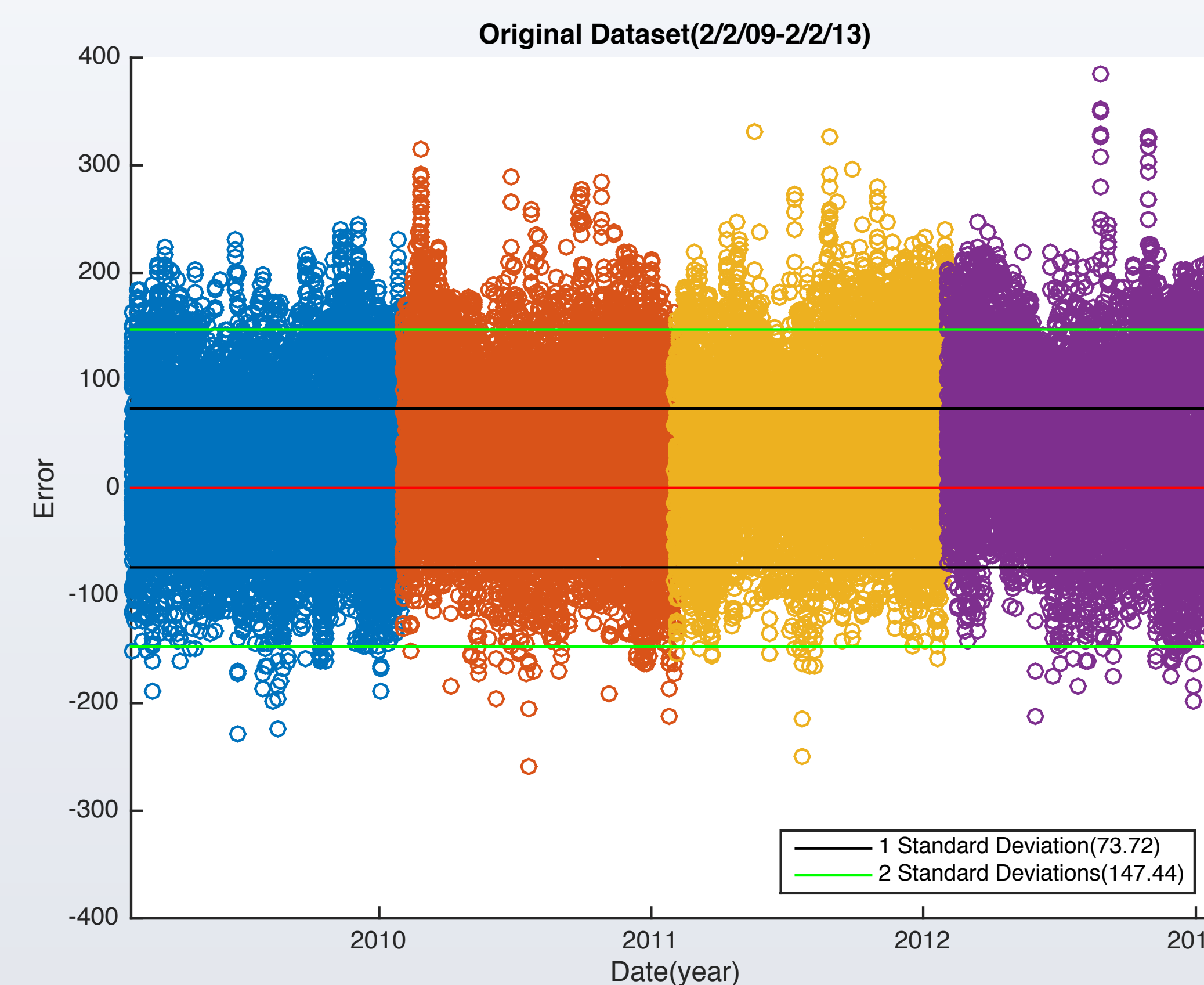


Splits effectively create smaller subsections of data that are better modeled linearly than a linear model of the entire dataset.



Trees consist of nodes, where the data is split into different categories based on a particular value, and leaves, which contain the linear regression model used for predictions.

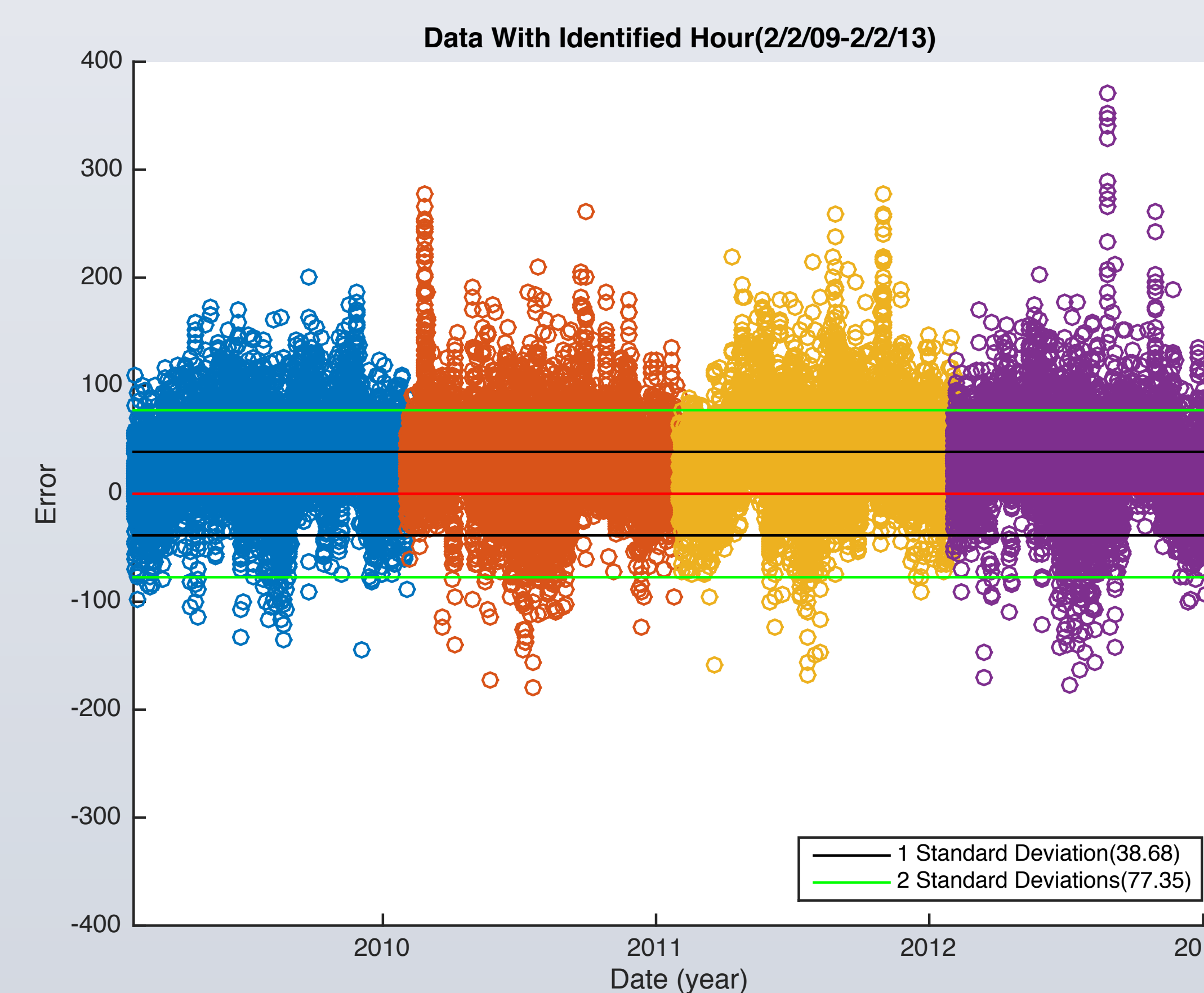
Results



Factors for original dataset:

- Day of Week (function of cosine)
- Day of Year (function of cosine)
- Day of Week (function of sine)
- Day of Year (function of sine)
- Temperature
- Wind Speed
- Wind Direction
- Dew Point
- Cloud Cover
- Precipitation

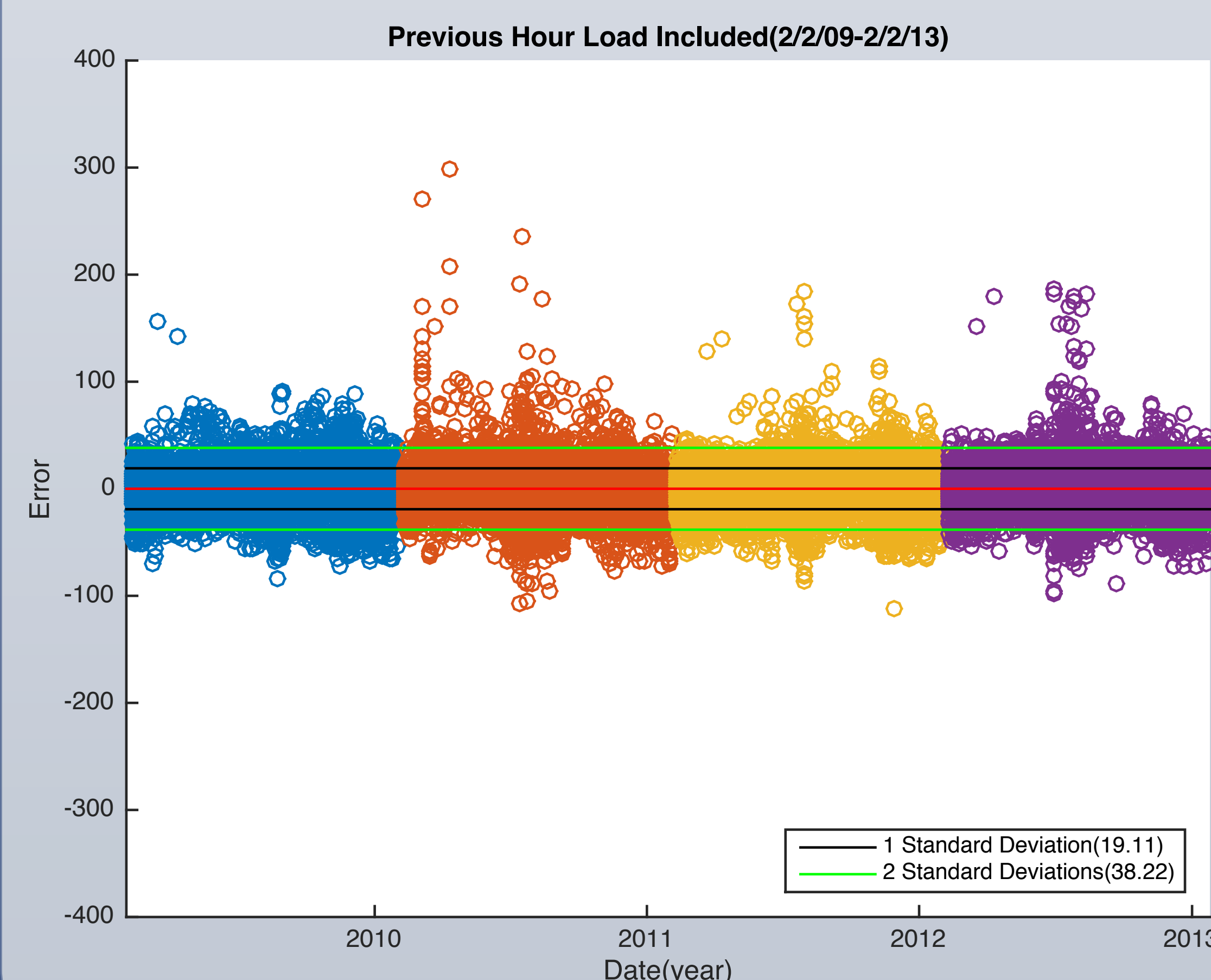
Four Year MAPE: 11.29%



Factors for hourly dataset:

- Hour (0-23)
- Day of Week (function of cosine)
- Day of Year (function of cosine)
- Day of Week (function of sine)
- Day of Year (function of sine)
- Temperature
- Wind Speed
- Wind Direction
- Dew Point
- Cloud Cover
- Precipitation

Four Year MAPE: 6.66%



Factors for previous hour dataset:

- Previous Hour Load
- Hour (0-23)
- Weekend (0 or 1)
- Day of Week (function of cosine)
- Day of Year (function of cosine)
- Day of Week (function of sine)
- Day of Year (function of sine)
- Temperature
- Wind Speed
- Wind Direction
- Dew Point
- Cloud Cover
- Precipitation

Four Year MAPE: 2.07%

Conclusions

Using the previous hour's load as an input reduces the average MAPE to 2.09%. A 95% confidence interval for the MAPE of this dataset states that the actual MAPE is between 1.873% and 2.312%, which is better than the 2.4% given by the artificial neural network.

First Test (Average MAPE)	Second Test (Average MAPE)	T-Statistic	P-Statistic	Conclusion
Original Dataset (11.36%)	Hourly Dataset (6.65%)	18.5	0.000	The datasets have significantly different means
Hourly Dataset (6.65%)	Previous Hour Dataset (2.09%)	31.5	0.000	The datasets have significantly different means

References

- [1] Genethliou, D. (2005). *Statistical approaches to electric load forecasting* (Order No. 3189368). Available from ProQuest Dissertations & Theses Global. (305374302). Retrieved from <http://search.proquest.com/docview/305374302?accountid=100>
- [2] Brierley, P. D. (1998). *Some practical applications of neural networks in the electricity industry*. Cranfield University.
- [3] Wang, Y., Witten, I. H., & University of Waikato. (1996). *Induction of model trees for predicting continuous classes*. Hamilton, N.Z: Dept. of Computer Science, University of Waikato.
- [4] Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. Amsterdam: Elsevier.

Acknowledgements

Thank you to Dr. Povinelli for his help and guidance, as well as the rest of the members of the GasDay lab. I would also like to thank Dr. Brylow and Dr. Factor for their work to organize the REU program and my fellow REU students and the Nation Science Foundation for its support. This research has been funded by the National Science Foundation grant.