Commercial Building Electricity Load Forecasting using ARIMA and Linear Regression

Kevin Chalhoub, Justin Appleby

Problem Statement and Data Processing

Predicting a building's electricity load, or the instantaneous amount of electric power a building is consuming, is important to modernizing today's electric power grid. The ability to accurately predict a building's power consumption will facilitate optimal performance from modern green power systems, such as tandem rooftop solar and battery storage systems. Performance affects degradation of the system conversion efficiency of solar energy to useful electricity, and ultimately costs to the electricity producer. These improvements in efficiency can be scaled to groups of buildings on commercial and academic campuses, or neighborhoods. Our goal is to implement a load forecasting model that makes accurate predictions over the next hour to several hours, given current load, time of day and time of year.

The data we use to predict a building's energy demand is structured as follows:
For every hour of every day of the year (A total of 8760 hours), we have the electric power (kW) used by every system of that building. As such, we may predict the load for the next hour of the day, using the previous hour's data. We used, in turn, every month as the test set in order to test our forecasting model for every month of the year.

Methodology

We were interested in looking at ways in which load forecasting was used in industry.[1, 2] Hence, we compared two different machine learning models to forecast the loads of the building: An AutoRegressive Integrated Moving Average (ARIMA) and a linear regression.

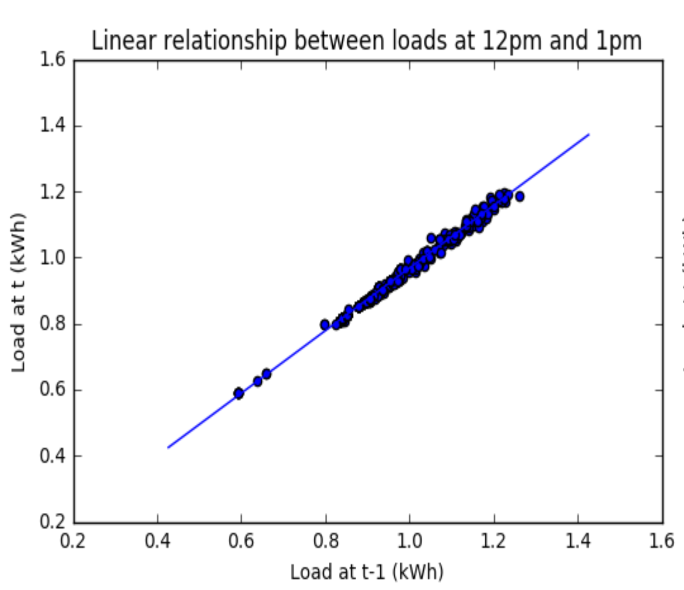
ARIMAs consist of three main parts[3]:

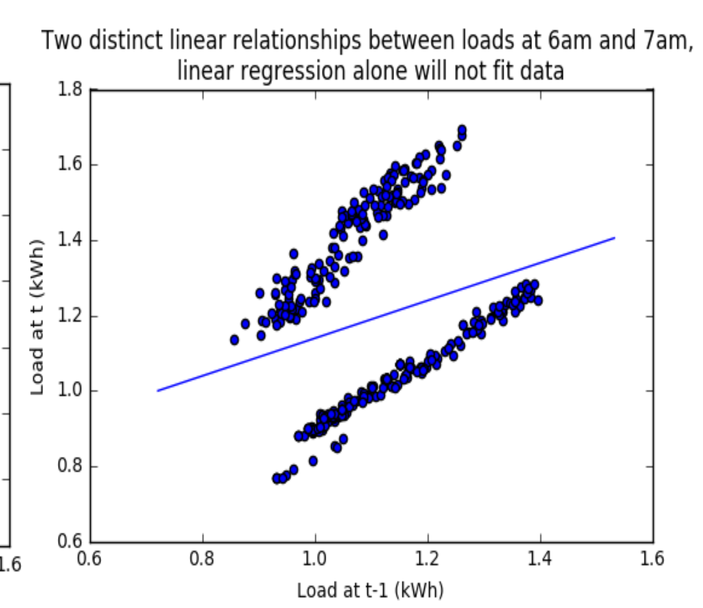
1. The AutoRegressive (AR) part indicates that the evolving variable of interest is evolved on its own part. Parameter p indicates the number of autoregressive terms. If p = 2,it means that predictors of x(t) will be x(t-1) and x(t-2).
2. The Integrated part (I) differences non-stationary data to remove any seasonality trends it may have. Parameter d indicates the degree of differencing.
3. The Moving Average part (MA) indicates that the regression error is actually a linear combination of error terms. Parameter q indicates the number of moving average terms there are. If q = 2, the predictors for x(t) will be e(t-1) and e(t-2) where e(i) is the difference between the moving average at i-th instant and the actual value.[4]

All in all, we set ARIMA(p,d,q) where we varied the parameters. The model uses the following equation for its forecast:

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$$

In linear regression, used a different regression model for each hour-to-hour change in load, using data from the previous hour to predict our energy use for the next hour. Our six features are the loads used from the different systems in the building: Overall Facility, HVAC, Lighting, Exterior, Appliances, Miscellaneous. We thought using linear regression would be a good idea to compare to ARIMA is that, when plotting the overall facility load at time t-1 and the load at time t, we would either get a linear relationship between the data or two distinct linear relationships, as shown in the image below:

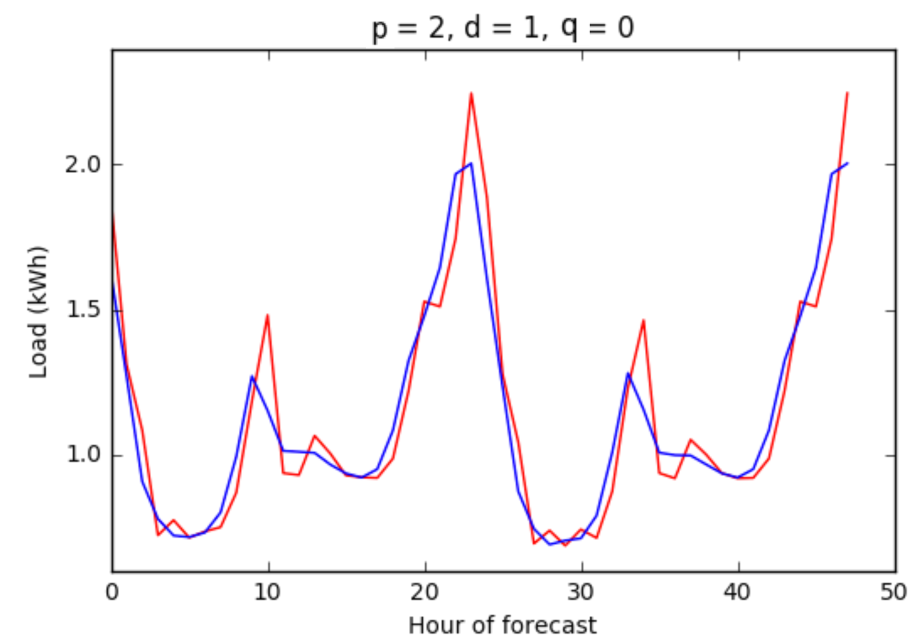




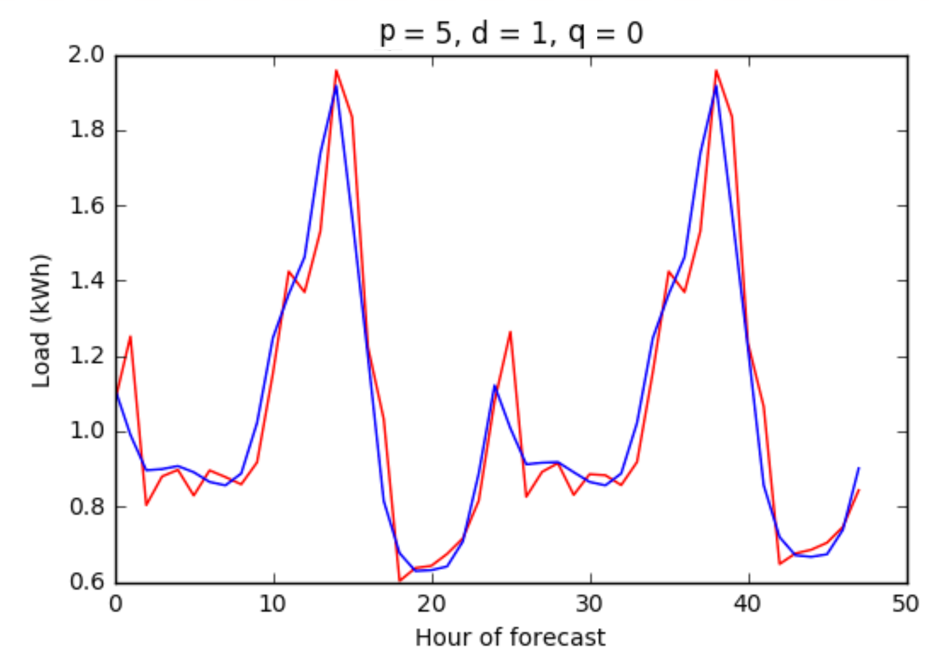
As we ran both models, we chose to select random variables from the data set as our test data. We therefore first train each of the models before testing these random variables and outputting the resulting graphs, mean squared-error and percent error to compare their accuracy.

ARIMA Results

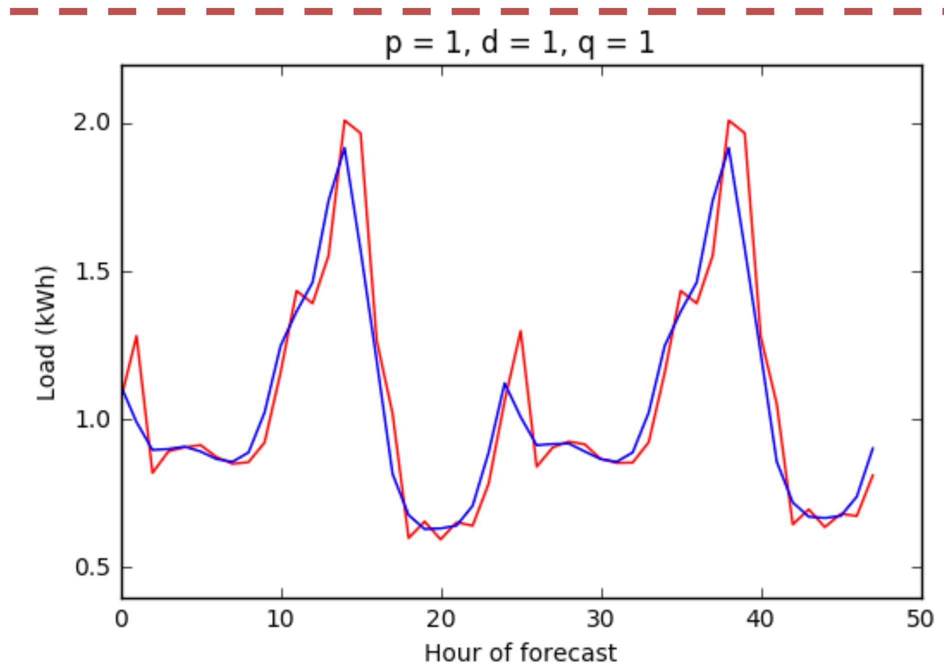
1



4

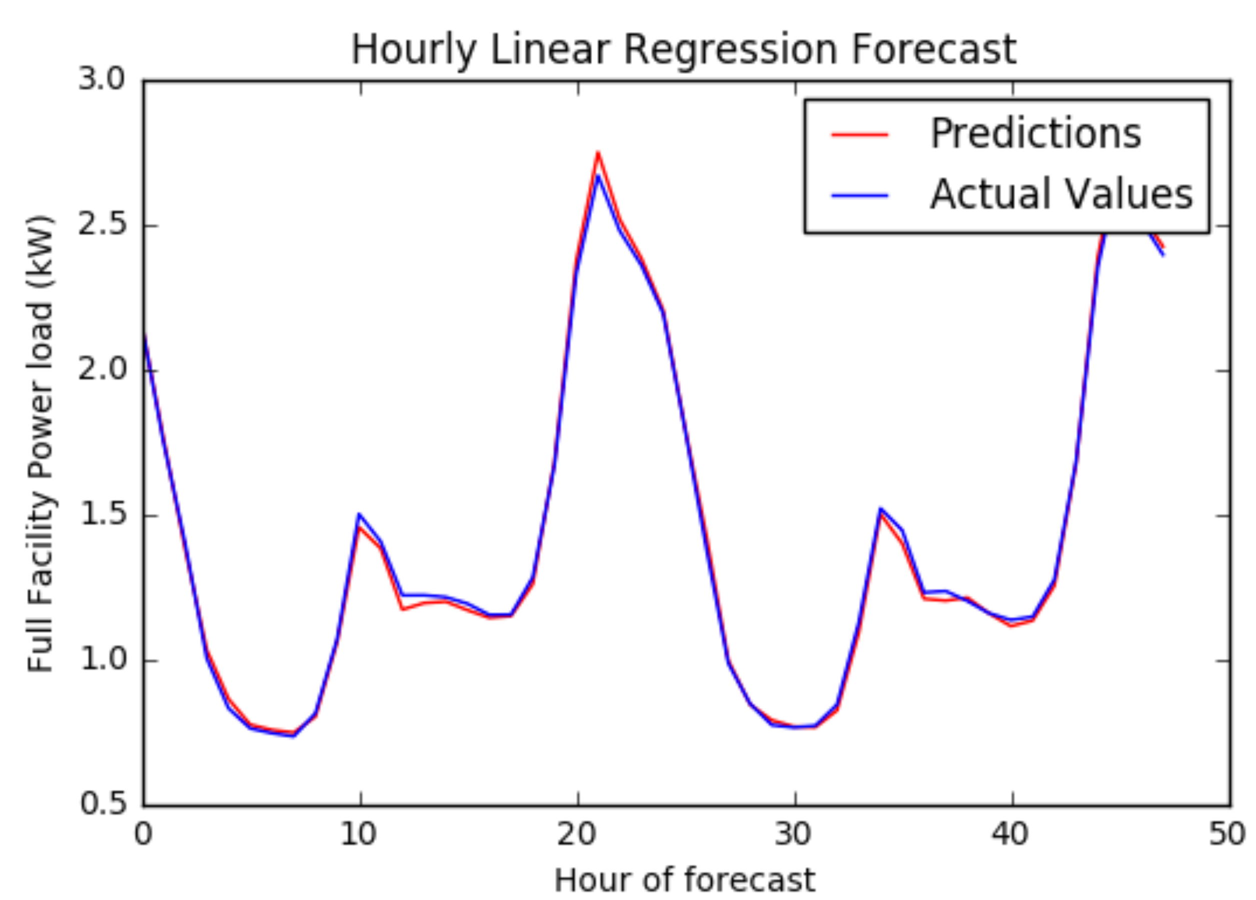


5



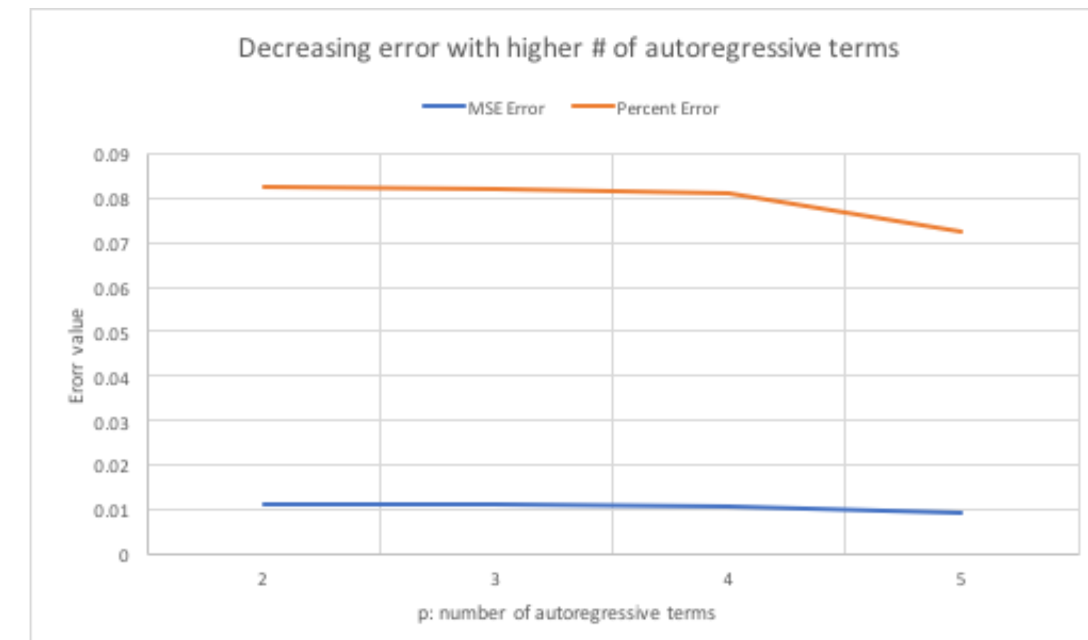
Linear Regression Results

6



Scenarios and Comparison

Scenario	p	d	q	Mean Squared Error AVG	Mean Squared Error STD	Percentage Error AVG	Percentage Error STD
1	2	1	0	0.0110	0.00148	8.26%	0.562%
2	3	1	0	0.0112	0.00148	8.22%	0.56%
3	4	1	0	0.0104	0.001544	8.09%	0.07%
4	5	1	0	0.00895	0.000803	7.26%	0.328%
5	1	1	1	0.0112	0.00148	8.26%	0.48%
6	Linear	Regression		0.00117	0.00215	2.67%	0.285%



Future Works

Our current model uses two types of regression to predict a building's load forecast. There are several ways we could improve our future works:

- Since our data is a load forecast, we could use a Long Short Term Memory (LSTM), which would consist of a neural network / deep learning algorithm, to predict our data more accurately.
- As our model only uses hourly data from one building, over one year, we could incorporate data from different buildings over that year in order to better test our current 3-D regression and ARIMA algorithms.
- As our data is only over one year, we could try to incorporate data from additional years in order to have a larger training data set and even, potentially, a larger test data set.
- In our linear regression, we could incorporate data from previous days or time t-2 in order to better predict our next hour load forecast.
- Determine the amount of data required to begin producing meaningful predictions. As in, how much of one year must the building have experienced before a model is making good predictions?

References

[1] J.C. Lu, D.X. Niu, Z.Y. Jia, A study of short-term load forecasting based on ARIMA-ANN http://ieeexplore.ieee.org/document/1378583/

[2] L.C. De Andrade, I.N. Da Silva, Very Short-Term Load Forecasting Based on ARIMA Model and Intelligent Systems http://ieeexplore.ieee.org/document/1378583/

[3] Autoregressive Integrated Moving Average, Wikipedia Online Pagehttps://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average

[4] A. Shah, Forecast a time series with ARIMA using Python, IBM Data Science Experience https://datascience.ibm.com/exchange/public/entry/view/815137c868b916821dec777bdc23013c

Acknowledgements

Andrew Ng
Dan Boneh
Kian Katanforoosh