Justin Appleby
Kevin Chalhoub

Building Electricity Load Forecasting with ARIMA and Sequential Linear Regression

Abstract
Load forecasting is an essential tool for buildings and residential facilities as it allows for a smarter use of a building's appliances. Our project therefore aims to forecast the load from different systems in the building and of the overall building by using two commonly used time series models, Autoregressive Integrated Moving Average (ARIMA) and Sequential Linear Regression. Both in-sample and out-of-sample forecasting are evaluated using root mean squared error and percent error for obth ARIMA and sequential linear regression. Of the parameters tested, ARIMA works best when using the set of parameters (p,d,q) = (16,1,2). Since the relationship between the loads for any two consecutive hours is linear, sequential linear regression outperforms ARIMA.

Introduction
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. These improvements in efficiency can be scaled to groups of buildings on commercial and academic campuses, neighborhoods, or even utilities' grids. We implemented a load forecasting model that makes accurate predictions over the next hour to day on a hourly resolution.[1,2] We used two models: Autoregressive Integrated Moving Average (ARIMA), and a set of 24 linear regressions, one for each of the 24 hour-to-hour changes, which we call sequential linear regression. We determined that ARIMA reaches reasonably good predictions, but with large amounts of data available, sequential linear regression can significantly outperform ARIMA.

Dataset and Features
Our dataset is one commercial building's hourly average total power, for an entire year. In addition to the full facility's electricity consumption, the file contains the hourly consumption of five categories: HVAC, interior lighting, exterior lighting, appliances and miscellaneous. This is 52,560 data points. Consistently, the sum of the five components is exactly 0.0207kW less than the facility's total load. The reason for the offset is not known. For the ARIMA model, only the facility's total load is used. For sequential linear regression, the full set of categories is used.

Methods & Procedure
ARIMA
AutoRegressive Integrated Moving Average (ARIMA) is commonly used in forecasting time series in industry.[1, 2] We implemented an ARIMA forecast. By convention, different ARIMA forecasts are written in the form ARIMA(p, d, q).[3] The parameters p, d, and q represent, respectively, the number of previous terms in the time series used to predict (AR), the order of differencing used (I), and the number of error terms, based on noise from previous data points (MA).[4] For in-sample forecasting, we varied these parameters according to table 1 in results.

ACF and PACF
The Autocorrelation Factor (ACF) and the Partial Autocorrelation Factor (PACF) dictate our choice for p and q. The ACF(p) of a time series measures a time series' correlation with itself shifted by p, and the PACF(q) measures a time series' correlation with itself shifted by q after eliminating variations explained by comparisons with shifts by all integers between 0 and q. Plotting the ACF and PACF as functions of p and q allows us to select p and q before training the ARIMA model.

In-sample forecasting

The ARIMA model is trained on 363 days of data and tested on the final 2 days. Errors are recorded and forecasts are plotted together with the real values. Note that these forecasts are in-sample forecasts, where each prediction is made, then the model is restrained with the prediction's corresponding real value. This is in contrast to out-of-sample forecasting, where future forecasts are based on prior forecasts. We used lower q values for in-sample forecasting because each time-step requires retraining the model, and training the model with high q values is costly timewise.

Out-of-sample forecasting

We now use the optimal p and q values found during ACF and PACF analysis. ARIMA(16, 1, 2) and ARIMA(3, 1, 2) are used to make forecasts on one day, having been trained on the other 364 days of the year. Forecasting is tested on 10 random days and errors are recorded.

Sequential Linear Regression

After spending time with the data, we noticed that the relationship between one hour's load and the previous hour's load is linear, as shown in figure 1a. The linear relationship is singular for most hour-to-hour changes, but clustered in others, as shown in figure 1b and 1c. Each hour-to-hour change (we'll call these 'jumps' from now on) has its own linear relationship.For each jump, there is a separate linear regression model. Thus, we have the name "sequential linear regression." Figure 1c shows that some jumps have two distinct patterns, *if you only look at the total load*. Including the other five categories in the feature set helps distinguish between these two "types" of jumps.

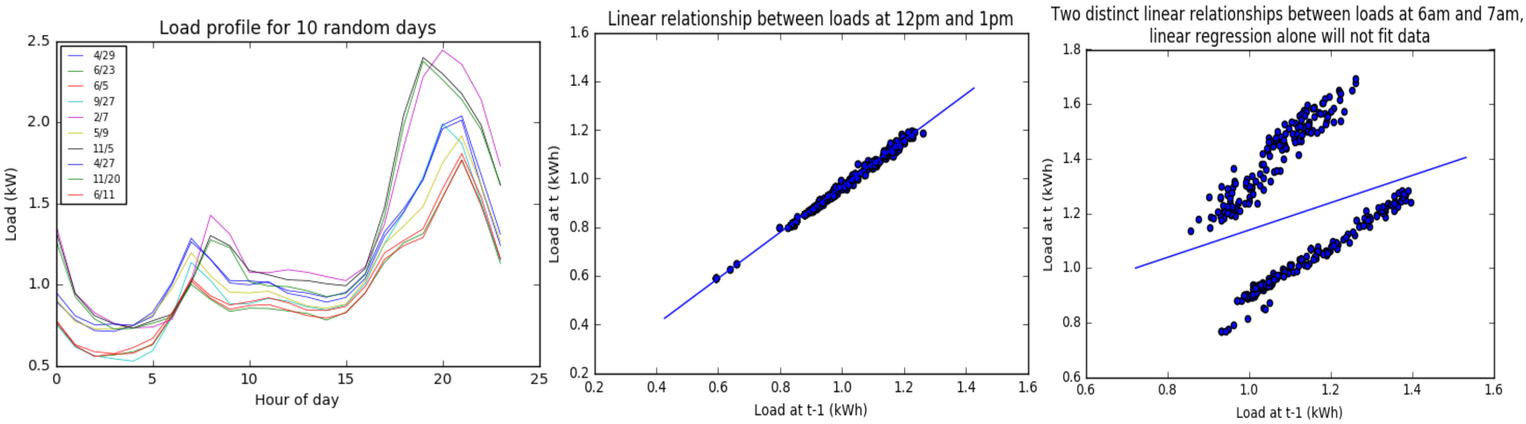


Figure 1a (left): 10 random days' load profiles. Figures 2b (center) and 2c (right): Scatter plots of full facility loads at time t-1 and time t, overlayed with sequential linear regression of the single variable full load.

In-sample forecasting

The features for the model at time t are the total load and each category from time t-1. This set of 24 linear models, each representing one jump, is trained on 335 jumps and tested on the other 10 jumps experienced by the building that year. Errors are recorded and forecasts are plotted. These forecasts represent in-sample forecasting. That is, after a prediction for one jump is made, the prediction for the next jump is based on real values, not the predicted values. Given the strong performance using sequential linear regression, we decided to evaluate the method further.

Out-of-sample forecasting

We then evaluate the out-of-sample forecasting, where each prediction is based on the previous prediction. This is in contrast to in-sample forecasting, where the forecast reverts to the true value before making the next prediction. This we needed to predict not only the total load at a given hour, but also the other categorical loads for that hour. These six predicted values are then fed to the sequential linear regression model for the next hour. The model is trained on 364 days and tested on 1 day. The process is repeated until every single day of the year has been the test day. This is then repeated for 9 different regularization terms, including zero.

Importance of training set size

Last, we try to understand the importance of training set size. For a building with no prior data, it is important to quickly develop a working model based on the data being acquired in real time. We train the 24 linear models starting with 5 days of history. Incrementally we increase the number of days "seen" by the building's sensors, and evaluate the model's accuracy on the remainder. To make sure the building still has data to predict even after "seeing" most of the year, we added a second year, a repeat of the first year with a very small amount of Gaussian noise. All forecasts are in-sample. Overfitting was extreme small training sets, so we repeated this process testing several regularization terms.

Results and Discussion

ARIMA

ACF and PACF

The best values to pick for p and q are values at which their ACF and PACF cross from positive to negative.[5] For this reason, we chose p = 3 and q = 2 for our initial analysis. We also chose p = 16 as it is another value at which ACF(p) passes from positive to negative. Selection of parameters is crucial as model-fitting is time-intensive. ACF and PACF graphs are in figure 2b.

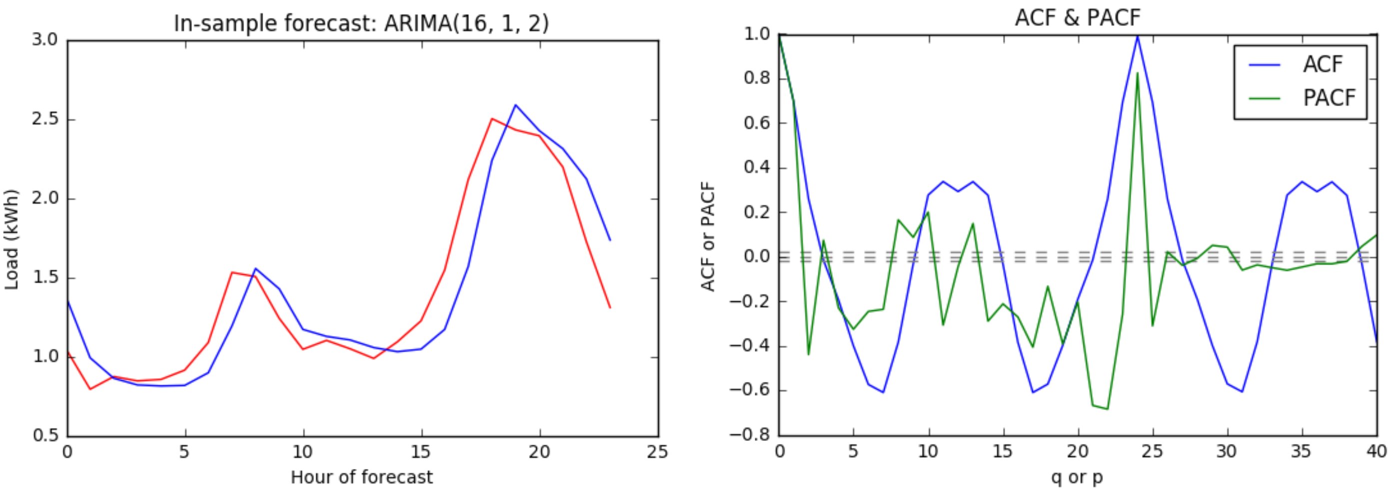


Figure 2a (left): In-sample forecast for ARIMA(16,1,2). Figure 2b: ACF and PACF for load time series. Used for choosing p and q parameters for ARIMA(p,1,q).

In-sample forecasting

In ARIMA, we train the model to 363 days of the year, and perform in-sample forecasting on 2 test days and record the errors. This is repeated for 10 different iterations using random pairs of two consecutive test days. We try several different sets of ARIMA parameters. An example is in figure 2a. Notice that ARIMA(16,1,2), the best of the ARIMA models chosen, overestimates the jumps in load.

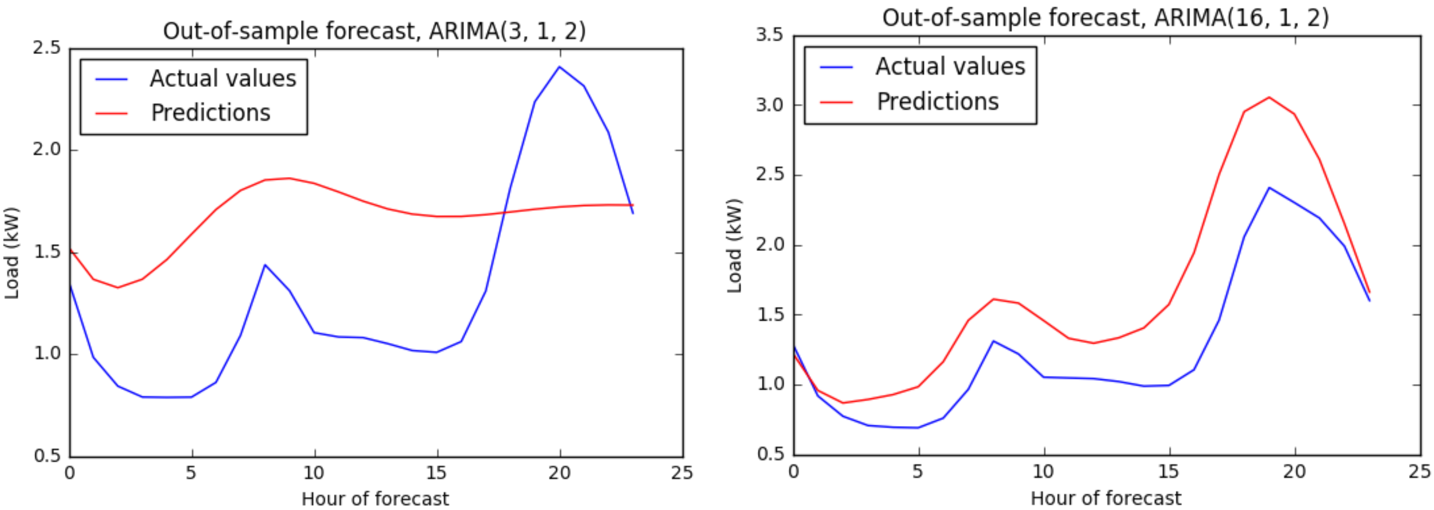


Figure 3a (left) and Figure 3b(right): ARIMA(3,1,2) and ARIMA(16,1,2) out-of-sample forecasts.

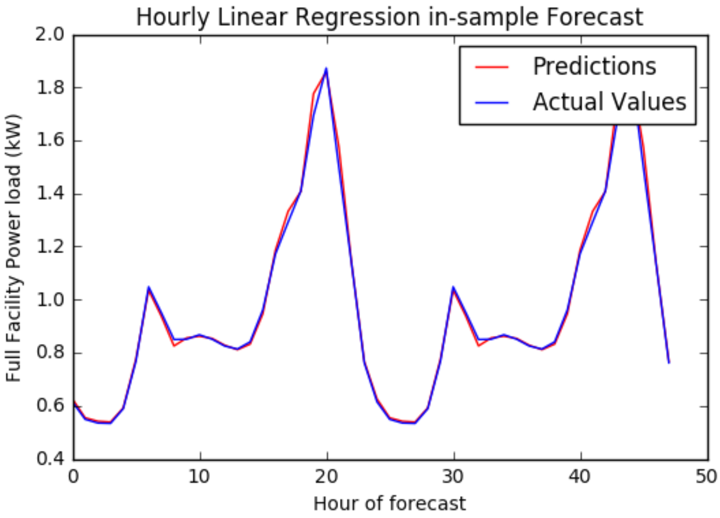
Out-of-sample forecasting

For out-of-sample forecasting, we use ARIMA(3,1,2) and ARIMA(16,1,2). As expected from the ACF and PACF analysis, they performed the best on in-sample forecasting. For ARIMA(3,1,2), the model was unable to replicate the daily fluctuations, shown in Figure 3a. ARIMA(16,1,2) was able to somewhat mimic the daily rise-and-fall in load. Intuitively, this makes sense, as ARIMA(16,1,2) keeps track of more previous data points, shown in Figure 3b. ARIMA(16,1,2) achieved an average error of 32.7%.

Sequential Linear Regression

In-sample forecasting

With nearly a full year of training data, the model achieves very good accuracy on in-sample forecasting of the remainder of the year. Comparisons for in-sample forecasting between ARIMA and Sequential Linear Regression are in table 1. This accuracy was replicated regardless of which two consecutive days were used as the test set. While every day's load and categorical load breakdown are slightly different, the model's use of HVAC, lighting and appliances is effective in detecting these minor differences while preserving daily trends. A visual comparison of the forecast to real values is shown in figure 4a. Sequential Linear regression does not overestimate the big jumps in load like ARIMA does.



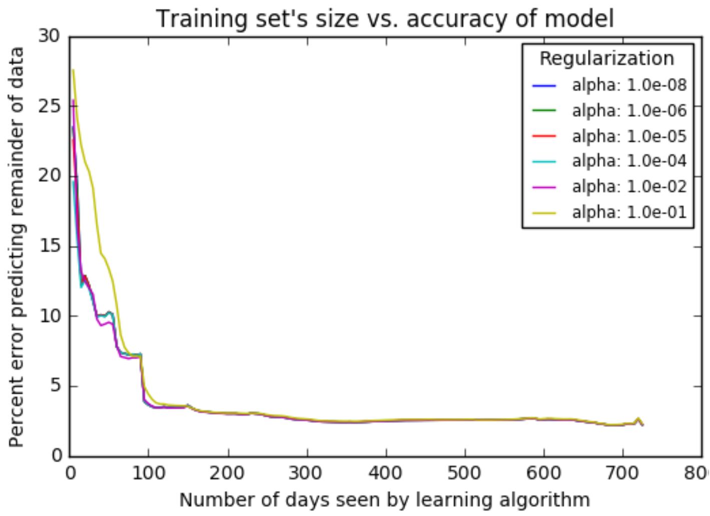


Figure 4a (left): Comparison of real values to predicted values for in-sample forecasting. Two days of load predicted with 363 days of training. Figure 4b (right): Average test accuracy versus number of days seen by model. Model achieves most improvement in the first 100 days.

Scenario	p	d	q	Root Mean Squared Error		Percentage Error	
				AVG	STD	AVG	STD
ARIMA1	2	1	0	0.139	0.0126	8.39%	0.501%
ARIMA2	3	1	2	0.118	0.0117	6.99%	0.530%
ARIMA3	16	1	2	0.0592	0.0104	4.31%	0.620%
ARIMA4	5	1	0	0.120	0.0157	7.26%	0.328%
ARIMA5	1	1	1	0.140	0.0.0144	8.17%	0.369%
Linear Regression				0.00117	0.00215	2.67%	0.285%

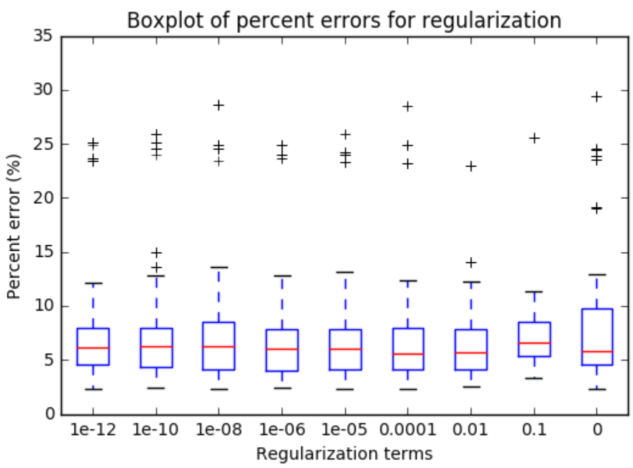
Table 1: Errors for in-sample forecasting between ARIMA(p,d,q) and sequential linear regression.

Importance of training set size

Initially, sequential linear regression without regularization produced erratic predictions until about 100 days in. The percent errors were orders of magnitude off the charts. The percent error quickly reduced to 2-3% after 100 days and stayed that low for the remainder of the two years of training. To avoid extremely erratic predictions in the first 100 days terms, we imposed regularization terms on the sequential linear regression models. Learning curves for these training iterations are in figure 4b.

Out-of-sample forecasting

In out-of-sample forecasting, previous predictions are used as input for the following predictions. Each day of the year is iteratively used as the test day and percent errors are stored. The distributions of these percent errors are shown in figure 5a. The sequential linear regression model achieves averages of about 7% for every regularization term (alpha) tested. Alpha equal to 10e-3, or 0.01, has the best performance with an average error of 6.4% and a standard deviation on that error of 2.8%. A numerical table of these accuracies is available in the appendix. A side-by-side comparison between the real load and a forecast from alpha equal to 10e-3 is shown in figure 5b. The model successfully fits the data for a full 24 hour period.



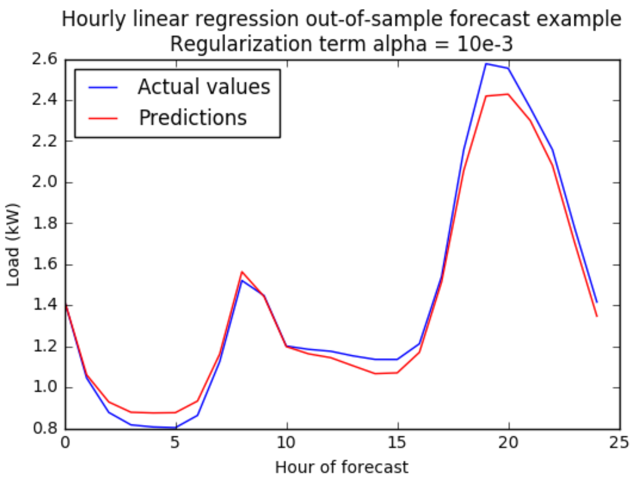


Figure 5a (left): Box plot with distribution of errors for each alpha, some values, especially for alpha equal to 0, are off the chart and very poor predictions. Each datapoint is an error for a test day. Figure 5b (right): Out-of-sample predicted load profile compared to real load for a randomly selected test day.

Conclusion & Future Works

The method of sequential linear regression outperforms the baseline ARIMA. It is possible to achieve under 3% error on in-sample forecasting of a full day with sequential linear regression, whereas ARIMA achieves about 8% error. In out-of-sample forecasting, the linear regression reaches below 7% average error if the regularization term is carefully selected. In addition, linear regression trains much faster.

Future work on ARIMA can include: making separate ARIMA models for each categorical load, more robust parameter (p, d, q) testing, requiring more computing power, and evaluating ARIMA on out-of-sample forecasting with better calibrated parameters.

Future work on sequential linear regression can include: expanding the feature set, and narrowing the most useful ones, including loads at the same time from the day before, loads from two hours before, and weather data. In addition, we would evaluate the quality of categorical load forecasts, in addition to total load, and determine the algorithm's performance on different buildings

In addition to improving and testing our ARIMA and sequential linear regression, we could evaluate our train-test procedure on other algorithms, such as Long Short Term Memory (LSTM). LSTM is a deep learning algorithm in the category of recurrent neural networks. The best benefit of sequential linear regression, however, is its speed of computation. The higher resolution we desire, however, the higher number of regressions we need, and those regressions would be subject to more noise. This caveat may mean LSTM or ARIMA could be more effective in high resolution (5 minutes or less).

Acknowledgements
Andrew Ng
Dan Boneh
Kian Katanforoosh

References / Bibliography

[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 Page https://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
[5] MathWorks, Autocorrelation and Partial Autocorrelation https://www.mathworks.com/help/econ/autocorrelation-and-partial-autocorrelation.html