Short-Term Electricity Demand Forecasting Using Independent Component Analysis

Jeff Chern, Matthew Ho, Edwin Tay

Introduction

Utilities providers like Pacific Gas & Electric (PG&E) have a vested interest in being able to effectively forecast electricity demand. This allows the utility to manage the supply-side of the electricity wholesale market more efficiently, as well as influence the demand-side using dynamic pricing schemes.

Load forecasting is usually done at the aggregate level. However, with more and more households installing smart meters (and with that, an increase in the availability of data), being able to forecast at the individual customer level is becoming of interest as well.

We worked from a data set containing usage history for 61 high-income customers from a 2003-2004 study performed by the California Energy Commission.

Our goal was to investigate the use of ICA (independent component analysis) for time series forecasting, particularly on this data set. We compare those results with results from ARIMA, a standard algorithm for conducting load forecasting.

Multiplicative Seasonal ARIMA Model

There is a variety of univariate methods that can be applied to online time series analysis. In particular, one method that has been popular for use in predicting short-term electricity load, and appears in many papers as a benchmark approach, is multiplicative seasonal ARIMA modeling[1].

The multiplicative seasonal ARIMA model, for a series X_t, with one seasonal pattern can be written as

$$(1-L)^d \left(1 + \sum_{i=1}^{p} a_i L^i\right)(1-L^s)^D \left(1 + \sum_{i=1}^{P} A_i L^{is}\right) X_t = \left(1 + \sum_{i=1}^{q} b_i L^i\right)\left(1 + \sum_{i=1}^{Q} B_i L^{is}\right) e_t$$

where L is the lag operator ($L^k X_t = X_{t-k}$) and s is the number of periods in a seasonal cycle. The model is often expressed as $ARIMA(p, d, q) \times (P, D, Q)_s$. It is multiplicative in the sense that the amplitude of the seasonal adjustment is proportional to the amplitude of the actual series (as compared to additive seasonal adjustment which is independent of the amplitude of the actual series).

Independent Component Analysis

Independent component analysis (ICA) is a statistical method used to find a linear representation of non-Gaussian data so that the components are as independent as possible. Such a representation would then capture only the essential structure of the data in these independent components.[2]

We conjecture that there is a common set of activities amongst all the customers that will be reflected to various degrees in the usage patterns of each of the individual customers.

The purpose of ICA will then be to discover these underlying "components."

Methodology

Preprocessing

The data was presented to us in a MySQL database. We wrote scripts to organize and import the data into MATLAB for further analysis.

Some customers had associated metadata about their geographic region, so we hand-selected 12 customers from the Bay Area that had sufficient overlapping date ranges. We chose a single 5 week slice from all of these customers. As the existing literature and our observations showed that weekend behaviors differed significantly from that of weekdays, we chose to focus on only the weekdays, Monday through Friday, for our predictions.

Finally, the original data was presented to us at a granularity of 96 data points a day, corresponding to 15 minute time intervals. We modified this to 30 minute time intervals to improve running time.

Error Metric

$$Mean\ APE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad Max\ APE = \max_{t}\left|\frac{A_t - F_t}{A_t}\right|$$

In load forecasting, the utility company is interested in the peak demand so that it can anticipate the actual infrastructure needed, which is why we have included the max APE as a metric as well.

SARIMA

By examining the autocorrelation and partial autocorrelation plots of customer electricity usage, we tried out several different values for the parameters (p,d,q) x (P,D,Q) as shown in Figure 1.

ARIMA parameters	Mean APE	Max APE
(0,1,1) x (0,1,1) $_{336}$	**0.5592**	**4.1508**
(0,1,1) x (0,1,2) $_{336}$	0.5530	4.1419
(0,1,1) x (1,1,1) $_{336}$	0.5526	4.1440
(1,1,1) x (0,1,1) $_{336}$	**0.2683**	**2.1529**
(1,1,1) x (0,1,2) $_{336}$	0.2566	2.1621
(1,1,1) x (1,1,1) $_{336}$	0.2535	2.1599
(1,0,1) x (1,1,1) $_{336}$	0.2575	2.1619

Figure 1. Mean and maximum absolute percentage error for different combinations of ARIMA parameters.

Other combinations of (p,d,q) x (P,D,Q), e.g. (1,0,1) x (1,1,1) $_{336}$ and (0,1,1) x (1,1,1) $_{336}$, resulted in failure to fit the parameters in the SARIMA training phase for customer 722. Moreover, these stability and convergence issues were even more severe for the other customers, which constrained us to use ARIMA(0,1,1) x (1,1,1)$_{336}$ and ARIMA(0,1,1) x (0,1,1)$_{336}$ for the rest of our predictions.

ICA

We treat the electricity usage series of each customer as an observation, and run ICA to obtain the time series for each independent component. We then attempt to predict the next day's usage pattern of each component using SARIMA, and combine the components to give each customer's usage pattern for the next day using the mixing matrix. This algorithm based off Popescu.[3]

Algorithm:

For each sliding window of 5 days {

1. Arrange the customer data as follows: $X = \begin{bmatrix} -cust_1- \\ -cust_2- \\ . \\ -cust_n- \end{bmatrix}$

2. Run ICA on X to obtain a mixing matrix A and the components $S = \begin{bmatrix} -s_1- \\ -s_2- \\ . \\ -s_m- \end{bmatrix}$

3. Predict (next day) on the components (via SARIMA) to obtain $P = \begin{bmatrix} -p_1- \\ -p_2- \\ . \\ -p_m- \end{bmatrix}$

4. Use the mixing matrix A to obtain the predictions in the original domain: compute AP

   where $n = 12$ is the number of customers and $m = 10$ is the number of components we chose

}

Tools

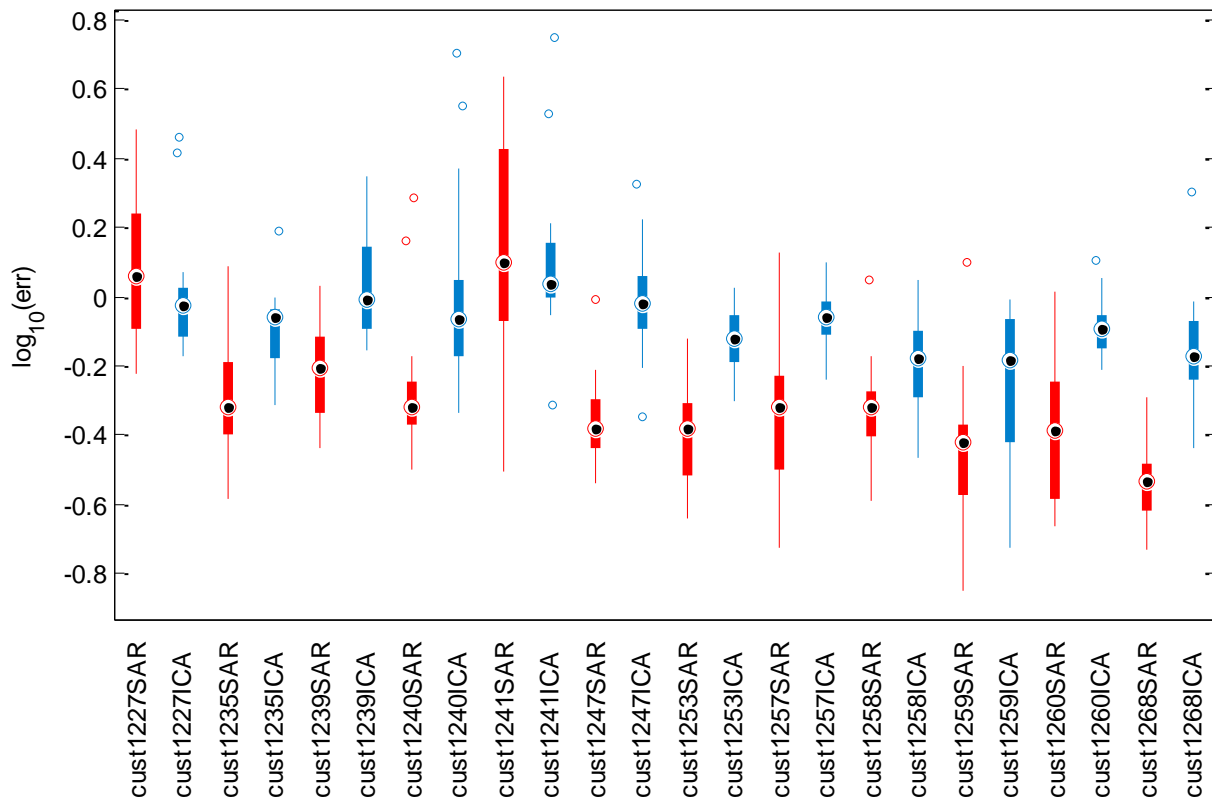SARIMA was run using an R package[4]. ICA was run using FastICA[5].

Results

Figure 2. Box Plot of error metric for ten-component ICA with $ARIMA(1,1,1) \times (0,1,1)$.
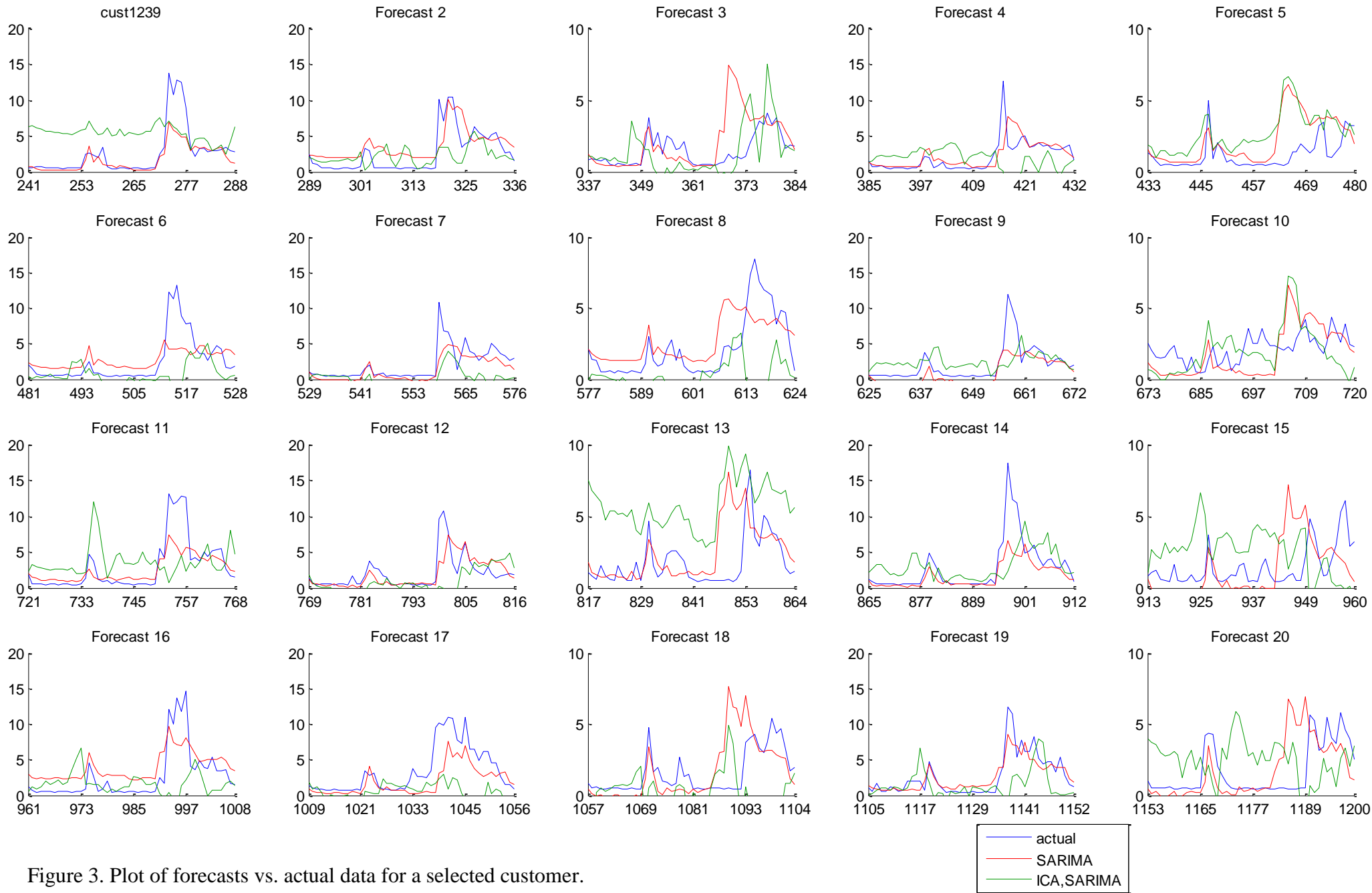(Red = SARIMA, Blue = ICA)

Figure 3. Plot of forecasts vs. actual data for a selected customer.

Figure 2 shows the APE distribution across forecasted days for each customer. Errors were computed for both the SARIMA and ICA forecasting approaches and are displayed in pairs for each customer. We see that in general, the ICA APE is higher, but with a lower spread than the SARIMA APE's. However, for customers whose usage patterns are more erratic (customers 1227 and 1241) ICA appears to provide more reliable results.

Figure 3 shows plots of the forecasts for each day for customer 1239, chosen as ICA forecasts seem the most reasonable. In general, neither SARIMA or ICA tracks the peaks well, although SARIMA seems to do a slightly better job.

Conclusion

Overall, performing ICA on the time series for each consumer and predicting on the resulting components resulted in poorer forecasts than predicting on the original time series.

We think these are some contributing factors:

1. The data used for ICA was too short (only 5 weeks). This might have resulted in longer term trends not showing up.
2. We did not include external influences likely to have an impact on the electricity usage, e.g. weather and income levels.
3. We did not vary the SARIMA seasonality to account for the particular periodicity of an ICA component.

Although applying ICA did not result in better forecasts in general, the fact that it better handles more erratic customers suggests that if the three factors listed above were addressed, ICA could be useful as a preprocessing stage.

Acknowlegements

We would like to thank Prof. Amit Narayan (Consulting Professor, EE), for providing the data and for support and guidance throughout the project.

References

1. Taylor, J.W., "Short-term electricity demand forecasting using double seasonal exponential smoothing," *Journal of the Operational Research Society*, 54, pp. 799-805, Mar. 2003

2. Oja, E., Kiviluoto, K., Malaroiu, S., "Independent component analysis for financial time series," *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, vol., no., pp.111-116, 2000

3. Popescu, Theodore D., "Time Series Forecasting using Independent Component Analysis," *World Academy of Science, Engineering, and Technology*, 49, 2009

4. Shumway, R.H., and D.S. Stoffer. "Time Series Analysis and Its Applications: With R Examples." N.p., 2006. Web. 07 Dec 2010. <http://www.stat.pitt.edu/stoffer/tsa2/Examples.htm>.

5. "FastICA." Helsinki University of Technology, 17 Oct 2007. Web. 07 Dec 2010. <http://www.cis.hut.fi/projects/ica/fastica/>.