User Behaviour Study and Energy Demand Prediction from Disaggregated Electricity Consumption Data

Shubham Saini, Jonathan Cervantes, Vyom Shah, Kenneth Vuong
University of California, San Diego
s1saini@ucsd.edu, jcervant@ucsd.edu, vyshah@ucsd.edu, kqvuong@ucsd.edu

ABSTRACT
Electricity Consumption and Occupancy Data is important in this day and age as more and more people are consuming energy. Motivated to take this data and find useful information, we focused in providing useful and informative information about the user and forecast power consumption. We make use of an electricity consumption data set from Switzerland with appliance level consumption information and occupancy details. We analyze the data to profile the residents' activities like watching television and cooking. We also find insights into users occupancy information. Finally, we use the data for day ahead electricity demand prediction based on historical consumption curves by making use of time series forecasting models.

1. INTRODUCTION
Energy consumption is at an all-time high and still increasing. Economic development and industrialization has put an enormous pressure on the producers of energy. One of the major problems faced by DISCOMS is addressing the peak demand energy consumption due to mismatch in the supply and the demand of energy.

Initially we intended to study a bitcoin dataset. We researched different datasets and thought about some predictive analysis we could do. We thought of utilizing the data to predict optimal trading for the real valued quantity of the price of Bitcoin. We also thought of focusing on Bitcoin's popularity and if it is gaining or losing traction. However after studying the dataset more, we realized a majority of the data is encrypted. It was difficult to identify returning users by their id and to find useful features in the dataset. We switched our focus on to a different dataset with more useful features.

We have been working on a Electricity Consumption and Occupancy (ECO) Dataset collected by ETH Zurich. The data is collected from 6 random Swiss households of different appliance consumption and occupancy data over a period of 10 months. The data collected has 86,400 measurements per day, measuring the power for each second. The ECO dataset contain 1 Hertz (Hz) plug level data, 1Hz aggregate consumption data, and occupancy information.

We also chose to combine this data with datasets from Zurich City Weather Report over the same 10 months. We studied to find if any correlation existed between weather and certain appliance consumption for occupants.

We focused on using the data to:

- Understand user behavior based on power consumption
- Forecast next day power consumption based on historical data
- Provide efficient use of available resources

Using real world data, consisting of both plug level data from appliances and aggregate consumption data, we evaluate the impact and applicability of the above practices in providing predictive measures and useful information of regular power consumers.

2. PREDICTION OBJECTIVES
For our first objective, we use power consumption profiles from the stoves and television sets from each household to detect patterns in the time people spend cooking or watching TV. The analysis for these tasks was mainly exploratory and no complicated methods were used. For power forecasting, we used four different techniques:

- Baseline: We use average power consumption from a 10-day window preceding any given day to forecast consumption for that day. This is our baseline model.
- SVR: We use the RBF as the kernel and the outside weather along with the current hour as features.
- ARIMA: A standard ARIMA model which assumes linearity.
- ARIMA/ANN Hybrid: This model feeds the residues from the previous model into an Artificial Neural Network. This allows the model to account for non-linearity.

We use the mean average percentage error (MAPE) to measure the accuracy of our models. This is a common error metric for forecasting tasks. We considered MSE and RMSE but decided not to use them as they are scale dependent.

3. DATA SETS USED

3.1 ECO Data set

ECO data set [1, 2] has 8 months data from 6 Swiss households sampled at 1 Hz. Released primarily for Non-Intrusive Load Monitoring research, ECO data set has aggregate consumption data, plug level data of selected appliances and occupancy data for 5 out of 6 households. Occupancy data is collected through tablet computers during summer and winter season. The appliance level data and occupancy data available within this data set makes it beneficial for extracting behavioural information of the residents'.

The average power consumption in the six houses for the month of December and January is presented in Table 1. The power consumption for House 4 and House 5 is significantly higher than other houses. One possible reason for this behavior is that these houses have the least hours of vacancy, as shown in subsequent sections.

Households	Appliances
House 1	332.803
House 2	231.2986
House 3	401.341
House 4	1162.81
House 5	995.6338
House 6	335.8318

Table 1: Appliances monitored in each house

Apart from the total aggregated power consumption data for each house, power consumption profiles of several appliances were also monitored for the duration of data collection. Table 2 summarizes the applications that were monitored:

Households	Appliances
House 1	Fridge, Dryer, Coffee machine, Kettle, Washing machine, PC, Freezer
House 2	Tablet, Dishwasher, Air exhaust, Fridge, Entertainment, Freezer, Kettle, Lamp, Laptops, Stove, TV, Stereo
House 3	Tablet, Freezer, Coffee machine, PC, Fridge, Kettle, Entertainment
House 4	Fridge, Kitchen appliances, Lamp, Stereo and laptop, Freezer, Tablet, Entertainment, Microwave
House 5	Tablet, Coffee machine, Fountain, Microwave, Fridge, Entertainment, PC, Kettle
House 6	Lamp, Laptop, Router, Coffee machine, Entertainment, Fridge, Kettle

Table 2: Appliances monitored in each house

The occupancy data is available for summer and winter. Occupants specified presence/absence through a tablet computer.

3.2 Weather Data

Apart from data supplied within the ECO data set, we also fetched weather information from Weather Underground[1]. Weather Underground provides an API for fetching historical weather information. Since the location of the homes is not released, weather data for Zurich, Switzerland (where the research group is location) was used.

[1] `http://www.wunderground.com/`

4. USER BEHAVIORAL ANALYSIS

Our focus started in finding similarity between consumption patterns of different houses. Some of our findings reveal when residents most likely watch TV or cook food. The data was useful in discovering what times people were mainly home based on appliances used. We worked with different features to forecast next day power consumption based on this historical data.

4.1 TV Power Consumption

Our first approach was to find any correlations with TV power consumption. We tried to calculate averages between the 6 households and see if any correlation or similarities existed with weekdays and weekends.

Things to Note:

- We are including Friday in the set of data for weekends. Fridays are a part of the weekend culture. Most adults are done for work for the week and students are done with school. Occupants of households are free from most obligations until the upcoming Monday.
- The data measured is the power from the plug each second of each day for a span of about 10 months. The data set did contain some missing measurements or faulty data, which was necessary to remove since they held a value of -1.
- House 1 preferred to use a laptop for most entertainment, instead of TV and this data is not factored in with the other houses.
- House 3 holds low averages because data coverage was weak due to the concrete ceiling in the basement, which disturbed the radio connection between our data set's gateway and the plugs.
- No occupancy data set had been available for household 6.

Households	Average TV Power	
	Weekdays	Weekends
House2	35.60	39.84
House3	3.42	6.15
House4	33.61	32.05
House5	27.5	26.75
House6	23.55	23.26
All Houses	**24.71**	**26.01**

Table 3: Houses TV Power (Hz) Consumption

These results show that people tend to watch a couple more hours of television on the weekends than on weekdays. We were expecting more people to watch television on weekends because of the amount of free time available. But with that in mind, a lot of people most likely do other things than stay at home and watch television all weekend.

Seeing how many home occupants watch just as much tv on weekdays versus weekends, we studied the features deeper and focused on the hours a day. Below are the charts of data pulled from the houses 2 through 6 over 10 months of the hours of each house.

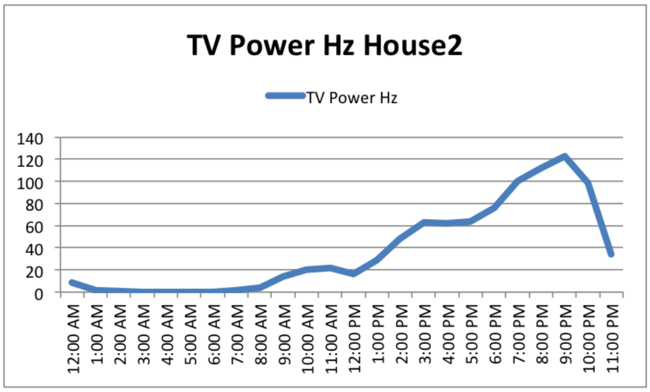

Figure 1: House 2 TV Power Avg Hourly Trend

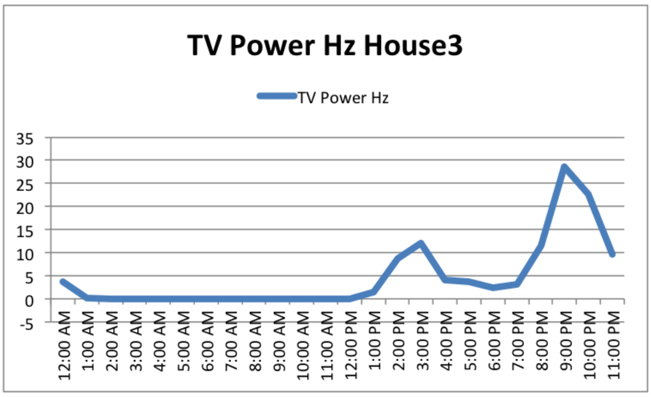

Figure 2: House 3 TV Power Avg Hourly Trend

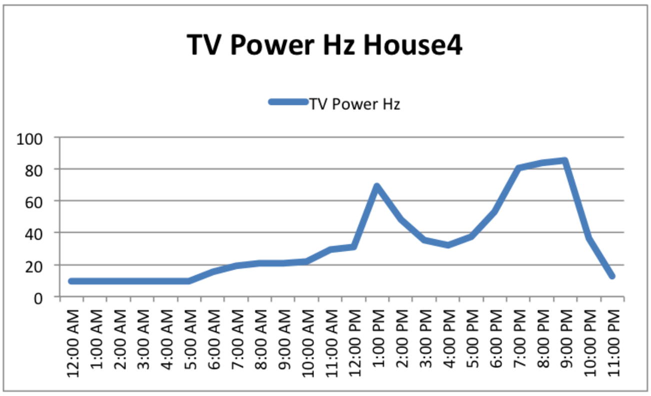

Figure 3: House 4 TV Power Avg Hourly Trend

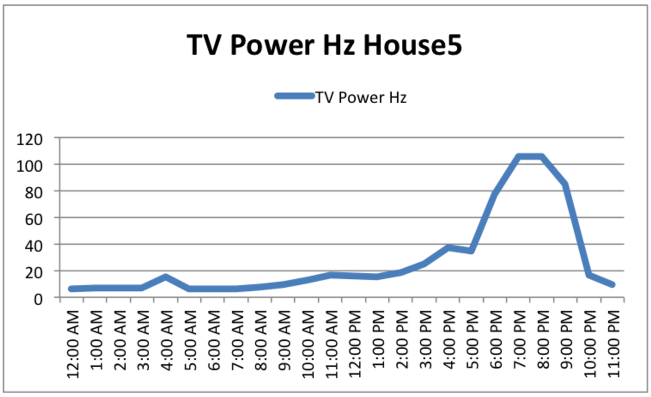

Figure 4: House 5 TV Power Avg Hourly Trend

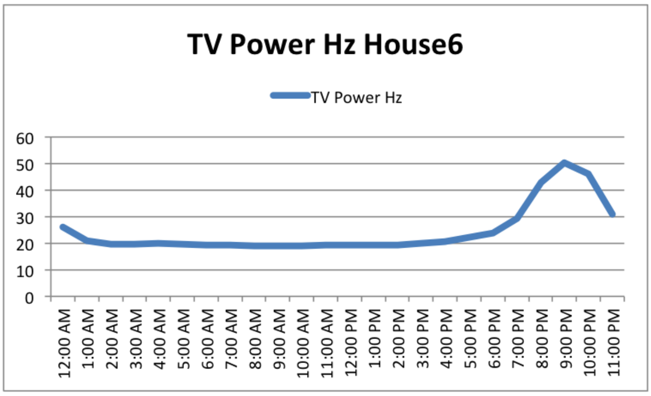

Figure 5: House 6 TV Power Avg Hourly Trend

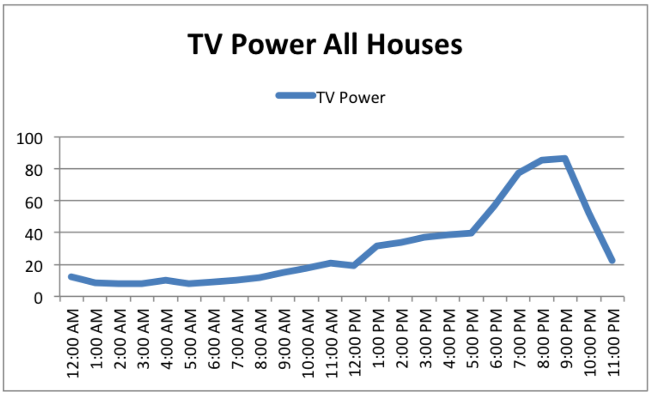

Figure 6: Aggregate Houses TV Power Avg Hourly Trend

The results including weekdays and weekends, that no matter the day, people watch television later at night, with 9pm being the time that most people will be watching television. Each house shared the about same graph curve structure. The later it gets, and as people get off work and begin to relax from their daily activities, the TV is more in use.

4.2 Cooking Power Consumption

We decided to focus in on plug data for cooking appliances. We focused mainly on those appliances of a stove, microwave, and/or oven. We decided to leave out things like bread maker, coffee maker, tea kettle, and toaster out of this category. We left out these appliances because when they are in use, they are used for a short duration and most times only once a day.

Households	Average Cooking Power	
	Weekdays	Weekends
House2 stove	15.49	10.57
House4 microwave	18.02	13.64
House4 stove	10.25	8.81
House5 microwave	9.67	7.839
All of above	**13.36**	**10.22**

Table 4: Houses Cooking Power (Hz) Consumption

These results show how during weekends people tend to cook much less. During weekends since most people usually have some free time during the day, they tend to go out for food to take a break from cooking at home. During the weekday people already are use to routine and usually eat at home. Cooking at home can be cost beneficial, while eating out can be a luxury that is done only on weekends. Using hours as a feature provided us with some more useful information. Below are the cooking power data charts we compiled from Houses 2, 4 and 5. These include the plug data from microwaves and stoves.

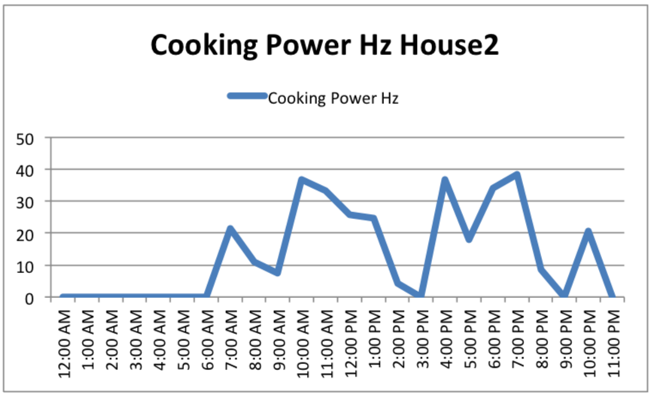

Figure 7: House 2 Stove Power Consumption

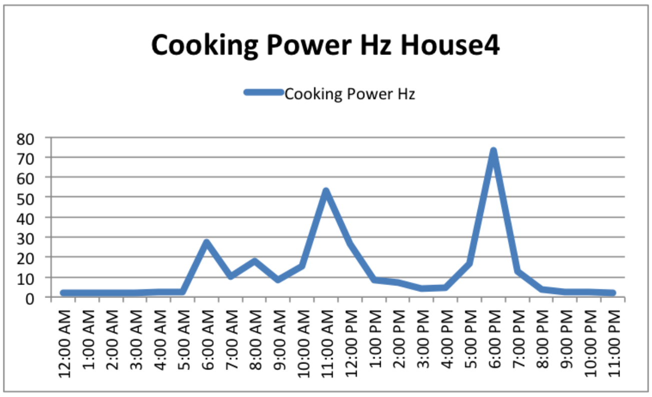

Figure 8: House 4 Stove/Microwave Power Consumption

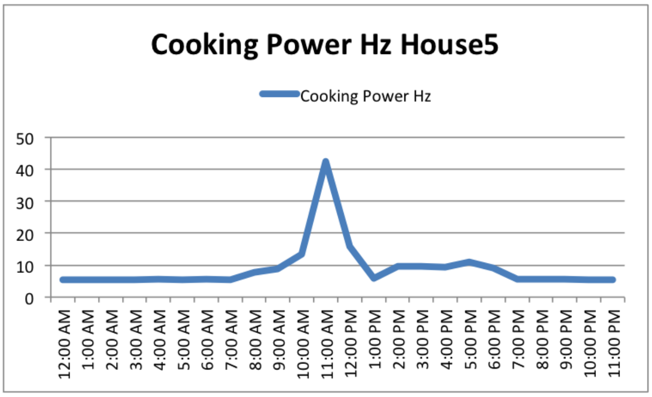

Figure 9: House 5 Microwave Power Consumption

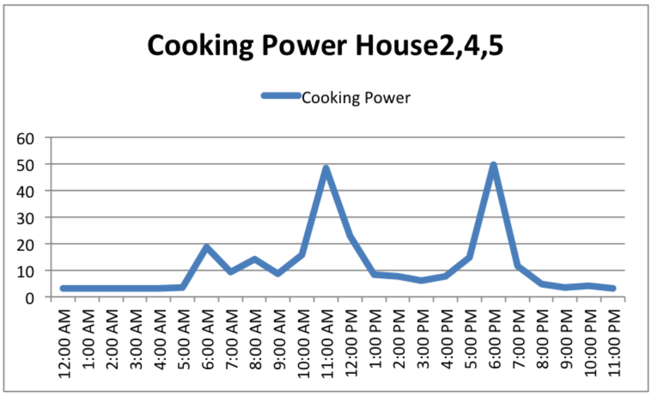

Figure 10: Aggregate Houses Cooking Power Consumption

Household 5 only happened to contain microwave data, while House 4 has both microwave and stove data, and House 2 has only stove data. This could skew some of our results, but there seemed to be some correlation and similarities. The three main peaks seen in the aggregate graphs containing all 3 houses, exemplify the cooking appliance power that is in use for breakfast, lunch and dinner. But there is some interesting information we can observe.

For those who must leave and eat breakfast early or on the go in the morning, we notice low power being consumed. This could occur since breakfast, that early, is usually small, quick, and cold food. But for those eating breakfast from 10-12 or eating an early lunch from 11-1, there is a spike in power data. Here the meal could be assumed to be warm and more hearty. The next big spike is definitely people cooking for dinner and it is pretty consistent 5-7pm. The dinner and lunch spikes represent the times most power is consumed, with more power being consumed during the dinner time. We can assume this pattern will continue for other households.

4.3 Occupancy Analysis

Our findings are reflected in Table 6 below. After analytic analysis, these findings show how families with children tend to leave the house much more on the weekends, than those without children. Parents usually have free time on weekends to spend time outside of the house with their kids. Also with no work and no school, the weekends provide most people the time to get out of the house, as well as trips and

Model	Household					
	House 1	House 2	House 3	House 4	House 5	House 6
Basline	31.21	39.7	30.87	39.9	33.76	57.63
ARIMA	44.48	42.49	44.13	35.95	41.1	62.56
ARIMA+ANN	41.36	38.26	44.1	50.85	38.9	68.06
SVR	31.2	25.56	28.45	37.28	29.54	41.97

Table 5: Performance of different models

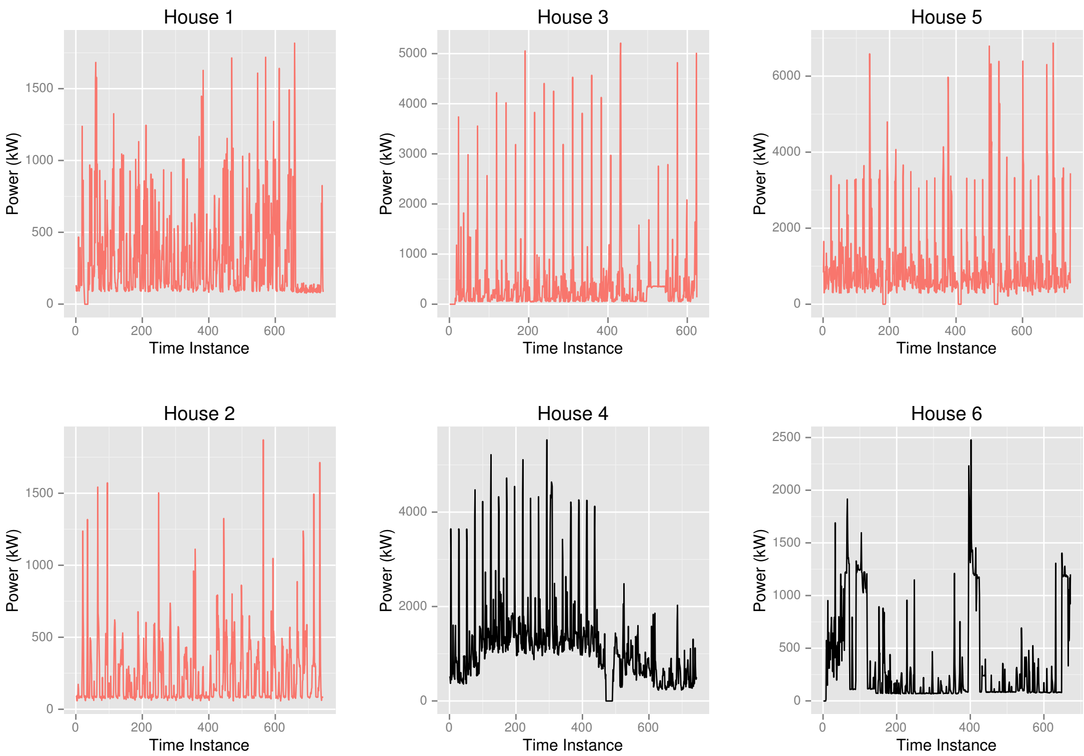

Figure 11: Forecasting Houses

House	Occupants		Hours	
	Adults	Kids	Weekdays	Weekends
House1	2	2	3	4
House2	2	0	5	5
House3	2	0	3	2
House4	2	2	2	7
House5	2	0	1	2
House6	2	0	N/A	N/A

Table 6: No. of kids and Average Hours Vacant per Household

vacations. These hours are rounded and represent when the household members are not home nor consuming any power. No occupancy data set had been available for household 6.

5. ENERGY DEMAND FORECASTING

Using the forecasting methods mentioned in earlier sections, we conducted forecasting experiments for 10 different randomly selection dates. The forecasting errors for these dates were averaged to find the final error. We compared three different forecasting methods with the baseline. While the ARIMA and hybrid ARIMA+ANN model did not improve the error rate, SVR beat the baseline across all the homes. A summary of the performance of the models is presented in Table 5.

5.1 Clustering Households

Using the temporal features of the electricity consumption curves, we extracted a set of global features as used by [3]. These features were used as the similarity measure between the different time series to cluster them with K-Means algorithm. This approach of clustering electricity consumption profiles was done previously in [4]. Houses 1, 2, 3, and 5 were clustered into one cluster; and Houses 4 and 6 were clustered into another - as shown by Figure 11. This corroborates why we get higher prediction error for Houses 4 and 6.

5.2 Results

As apparent in Table 5, the model that performed best was SVR. SVR had the advantage of non-linear multivariate modeling with two strong features: weather and hour of day. Day of the week was another potential feature, but similar averages between weekdays and weekends provided little information for SVR. After that, the baseline remained the top performer. This aligns with intuition; a household is unlikely to change their power consumption drastically from the past 10 days. Both of these models outperformed ARIMA and ARIMA + ANN, which failed because of the lack of flexibility in their linear univariate structure. They were unable to incorporate contextual features that the other models could.

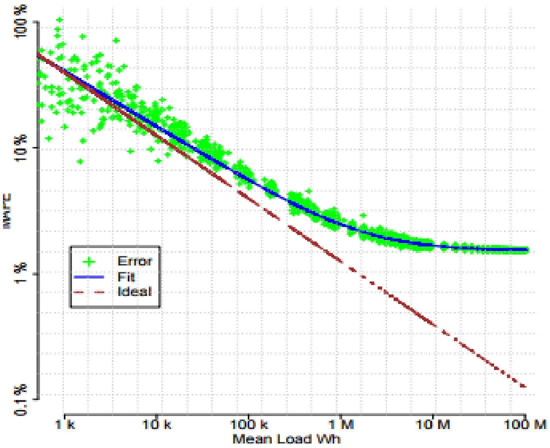

Figure 12: Aggregation MAPE

We also experimented with aggregated demand prediction, i.e. predicting demand for more than one house at once. In this approach, a group of houses are considered as a single entity. [5] shows that such an approach reduces the average MAPE as the number of houses are increased. Ideally, we should have obtained a curves like Figure 12. But since we did not have sufficient houses in the dataset, this approach failed to give favorable results.

6. RELEVANT LITERATURE

The data set we decided to narrow in on and analyze is an existing set of data sets from the university in Zurich, Switzerland: ETH. Twenty-one Nobel Laureates have studied, taught or conducted research at ETH Zurich, underlining the excellent reputation of the university. The research focused on is the needs of society, that makes a valuable contribution to the economy, politics and society in general. The data sets we used came from the Pervasive Computing Distributed Systems research. This data consisted of a research group of 6 households in Zurich and their Electricity Consumption and Occupancy (ECO) data. This data set is used in the University for research on certain distributed systems groups. We used the aggregate power data, single appliance plug data, and occupancy information data of each household to find similarities and any consumption patterns.

Other data sets exemplify similar studies on consumption analysis. Several sample data reports and studies exist on DataPort Pecan Street with yearly electrical and water data on Solar Power Performance and Maintenance. For example, this example data set had been used to study and predict water use consumption from flow rate. Other similar data sets include Smart Energy and open metering.

Electricity Demand prediction have been widely studied in the recent years. Univariate Time series forecasting that involve analyzing the historical observations has gained popularity in electricity consumption forecasting [6, 7]. The AutoRegressive Integrated Moving Average (ARIMA) model is one of the most widely used time series models.

However, ARIMA model assumes linearity and can not model non-linear patterns. Artificial Neural Networks (ANN) have also gained popularity because of their flexible non-linear modeling [8]. Certain hybrid approaches for time series forecasting have also been proposed, the most prominent being hybrid ARIMA-ANN forecasting model proposed in [9].

Another famous non-linear modeling algorithm is the Support Vector Regression, based on Support Vector Machine (SVM). SVR can be configured by using different kernel functions, cost, and loss function. We use SVR with RBF kernel for our experiments.

7. CONCLUSION

Studying the TV and cooking power consumption between the houses provided us useful insights. Time-series analysis presented us with the most popular and least popular hours and days people cook and watch TV. Analyzing the data further, we were able to use the occupancy data to find when people are usually home or not based on their power consumption. Finding commonalities of power consumption between several households can prove useful to users as well as electric companies. This analysis helped provide us with useful features in our approaches for a predictive model.

Since our data was limited to only a set amount of months, our MAPE was not as low as we had wanted. But if we increased our data and added more than 6 houses and we had cycle year data, than our MAPE would have decreased. A study done, shows how as more houses are included, our models would perform better predictive analysis.

We saw that power consumption tends to be strongly correlated with weather/seasonal changes. This could be a large part of the reason why our baseline performed better than the ARIMA models. The weather is unlikely to change drastically within a 10 day window, so the average power consumption from that period is naturally a good estimator for the following day.

8. REFERENCES

[1] Wilhelm Kleiminger, Christian Beckel, and Silvia Santini. Household occupancy monitoring using electricity meters. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2015)*, Osaka, Japan, September 2015.

[2] Christian Beckel, Wilhelm Kleiminger, Romano Cicchetti, Thorsten Staake, and Silvia Santini. The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the 1st ACM*

International Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys 2014). Memphis, TN, USA, pages 80–89. ACM, November 2014.

[3] Xiaozhe Wang, Kate Smith, and Rob Hyndman. Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery*, 13(3):335–364, 2006.

[4] Shubham Saini, Pandarasamy Arjunan, Amarjeet Singh, and Ullas Nambiar. E-adivino: A novel framework for electricity consumption prediction based on historical trends. In *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems*, pages 213–214. ACM, 2015.

[5] Raffi Sevlian and Ram Rajagopal. Short term electricity load forecasting on varying levels of aggregation. *arXiv preprint arXiv:1404.0058*, 2014.

[6] Volkan Ş Ediger and Sertac Akar. Arima forecasting of primary energy demand by fuel in turkey. *Energy Policy*, 35(3):1701–1708, 2007.

[7] Gürhan Küçük and Can Basaran. Reducing energy consumption of wireless sensor networks through processor optimizations. *Journal of computers*, 2(5), 2007.

[8] Hsiao-Tien Pao. Forecasting electricity market pricing using artificial neural networks. *Energy Conversion and Management*, 48(3):907–912, 2007.

[9] G Peter Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003.