Empirical Comparison of Prediction Methods for Electricity Consumption Forecasting

Saima Aman, *Student Member, IEEE,* Marc Frincu, *Member, IEEE,* Charalampos Chelmis, *Member, IEEE,* Muhammad Usman Noor, *Student Member, IEEE,* Yogesh Simmhan, *Senior Member, IEEE,* Viktor Prasanna, *Fellow, IEEE*

Abstract

Recent years have seen an increasing interest in providing accurate prediction models for electrical energy consumption. In Smart Grids, energy consumption optimization is critical to enhance power grid reliability, and avoid supply-demand mismatches. Utilities rely on real-time power consumption data from individual customers in their service area to forecast the future demand and initiate energy curtailment programs. Currently however, little is known about the differences in consumption characteristics of various customer types, and their impact on the prediction method's accuracy. While many studies have concentrated on aggregate loads, showing that accurate consumption prediction at the building level can be achieved, there is a lack of results regarding individual customers consumption prediction. In this study, we perform an empirical quantitative evaluation of various prediction methods of kWh energy consumption of two distinct customer types: 1) small, highly variable individual customers, and 2) aggregated, more stable consumption at the building level. We show that prediction accuracy heavily depends on customer type. Contrary to previous studies, we consider the consumption data granularity to be very small (i.e., 15-min interval), and focus on very short term predictions (next few hours). As Smart Grids move closer to dynamic curtailment programs, which enables demand response (DR) events not only on weekdays, but also during weekends, existing DR strategies prove to be inadequate. Here, we relax the constraint of workdays, and include weekends, where ISO models consistently under perform. Nonetheless, we show that simple ISO baselines, and short-term Time Series, which only depend on recent historical data, achieve superior prediction accuracy. This result suggests that large amounts of historical training data are not required, rather they should be avoided.

Index Terms

Smart Grid, Prediction Modeling, Time Series Models, Regression Tree Models, ISO Baseline Models.

I. Introduction

Recent years have seen an increasing interest in developing accurate prediction models for electrical energy consumption. This is motivated by both energy markets and environmental aspects. The growing prevalence and accuracy of sensors, uptake in renewables and energy storage technology, and availability of scalable processing solutions have facilitated the advent of the Smart Grid, in which energy consumption is optimized based on utilities' and clients' goals, to enhance the reliability of the power grid. For the utility provider, it is important to efficiently deal with peak demands (e.g., by Demand Response Optimization – DR) in order to avoid exceeding its generation capacity and having to buy energy from the spot market at high rates. Smarter energy consumption also benefits clients as they save money by turning off equipment or temporarily adjusting their comfort parameters (e.g., raising the thermostat level on hot summer days during peak hours).

One of the key challenges in Smart Grid is to reliably, accurately and efficiently predict future supply and demand trends [1], [2] to allow utility providers and consumers to prepare for peak periods as well as to plan for DR activities. Currently, US utilities have deployed millions of smart meters that collect energy consumption data from residential and commercial customers [3]. Such growing availability of energy consumption data offers unique opportunities in applying forecasting models and evaluating their efficacy for Smart Grid applications. However, to date, there has been little study on the differences in consumption characteristics of various customer types and their impact on the prediction model's accuracy. Existing studies have shown that consumption prediction's accuracy is high for aggregated consumers [4], [5], [6], i.e., at the substation-, microgrid- or building- level, however, there is a lack of results regarding individual customers' consumption prediction. Techniques that work well for large commercial customers with smaller consumption variability over time, could be less efficient for small residential customers, whose consumption pattern fluctuates significantly.

Our study differs from previous work (cf. §II), in that: a) it deals with both small, highly variable individual customer consumption, and aggregated, more uniform building consumption; b) the consumption data granularity is small (i.e., 15-min interval); c) it focuses on very short-term predictions (hours ahead), as opposed to most prior work on day-ahead predictions; and d) it evaluates the relationship between prediction accuracy and day type, i.e., workday versus weekends or holidays.

Saima Aman is with the Department of Computer Science, University of Southern California, Los Angeles, CA, 90089
E-mail: saman@usc.edu
the other authors are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA, 90089

We investigate two real-world Smart Grid applications: for the aggregated consumption at building level we rely on data from the University of Southern California (USC) campus microgrid [7], [8], while for the fine grained consumption at the customer level, we use data from a major California power utility. By comparing results from these two environments, we can show the extent to which the findings from building-level predictions within USC can be generalized to more variable usage patterns of residential customers within a utility.

In this paper, we address *consumption prediction*, which is different from demand (or load) prediction. The former deals with electrical energy consumption (measured in kWh) – we use the term energy to refer to electrical energy – sampled at regular intervals of every 15-mins. The latter deals with predicting instantaneous power measured in kW, and is beyond the scope of this paper. The motivation for 15-min energy consumption prediction originates from the rate at which Advanced Metering Infrastructure (AMI) acquire consumption data. This level of granularity is used for two types of applications [9]:
1. DR: Under DR programs, customers are encouraged to curtail – on-demand by the utility – their electricity consumption during peak electricity usage periods to reduce chances of black-outs. 15-min granularity predictions enable the utility to appropriately time curtailment requests and plan pricing mechanisms;
2. Customer Education: Providing customers with fine-grained information about electricity consumption can enhance their participation in efficient consumption [10] and active curtailment during DR periods. In our project, we are providing 15-min predictions about campus buildings to customers through web and mobile applications[1].

The rest of the paper is structured as follows: §II presents the main achievements in energy consumption prediction. §III presents the prediction models used in this study. We consider several prediction models: historical averaging, ISO, time series and regression tree. The experimental set-up is described in §IV and the result analysis in §V. §VI concludes with a discussion of the implications of our findings, and directions of future work.

II. Related Work

Research on electricity demand forecasting considers long-term and medium-term prediction for utility planning and maintenance purposes, and short-term forecast for economic scheduling [11]. As utilities move towards Dynamic Demand Response (D2R), very short-term predictions are required for near real-time control. Research on energy consumption prediction can be divided into three groups [11]: simple averaging models; statistical models (e.g., regression and time series); and artificial intelligence (e.g., Artificial Neural Networks (ANNs) and pattern matching [5]) approaches. Next, we briefly discuss the most related work in this area.

a) Averaging models: Utilities and Independent Service Operators (ISOs) use averaging models [12][13][14] based on recent consumption [15], due to their simplicity. Such models make predictions based on linear combinations of consumption values from "similar" days. Some of these methods used in practice are described in §III-A.

b) Regression models: Regression models combine several independent features to form a linear function. This helps interpret the relationship between various factors more easily. Our prior work [16] builds regression tree models using weather and schedule data for energy prediction. It evaluates the effect of different feature combinations on the prediction accuracy.

Probabilistic linear regression and Gaussian process regression models for predicting the total kWh consumption as a function of building features, were proposed in [17]. A hybrid method for probabilistic short term load prediction was presented in [18], where a regression tree is used to cluster similar data, and then, a relevance vector machine is constructed for each cluster using Bayesian Inference. Support Vector Machines (SVM) were used for load forecasting in [19]. An on-line least-square SVM based method is introduced by Aung et al. [20]. In their small scale evaluation (only 2 smart meters were used), they argue their method outperforms the least-square SVM based method introduced in [21].

A multiple linear regression model for load prediction, where affecting factors are iteratively analyzed, was presented in [22]. A nonlinear and nonparametric regression model for next day half-hourly load prediction was employed in [23] for stochastic planning and operations decision making. The model contains a combination of maximum, minimum and average demand and temperature from the last 1-hour, 24-hours, and 48-hours. Day of the week, day of the year and holiday effects are also incorporated.

c) Time series: An overview of time series forecasting approaches for electricity price prediction was presented in [1]. In order to capture the market fundamentals in multiple granularities (e.g., short, medium, and long-term), the price vector was split into components, which can be separately solved on different time horizons. One of the early reviews for time series based methods for load forecasting was given in [24]. Later, a time series method for short term load forecasting (few hours to few weeks ahead) of hourly loads was proposed in [25]. A comparison of time series methods for load forecasting with other methods was presented in [26]. Seasonal time series were investigated in [27].

d) Artificial Intelligence (AI) Approaches: Many common AI techniques, such as ANNs, expert systems and pattern matching techniques can be beneficial to demand forecasting [2]. An overview of AI methods applied to short-term electric load forecasting was provided in [28], [29]. Hourly electricity consumption forecasting for day-ahead prediction based on pattern sequence similarity was performed in [5]. An ensemble model based on this work, was later presented in [6].

[1]USC SmartGrid Portal, smartgrid.usc.edu

All aforementioned methods give accurate predictions in limited scenarios according to the requirements we set forth in §I, i.e., small consumption data granularity, very short-term prediction horizon, and diverse customer types. In this work, we evaluate multiple prediction methods from the literature, while considering all requirements simultaneously.

III. Prediction Models

In this section we present the prediction models used in our experiments. We focus on simple methods used by utilities and ISOs in the US, and statistical models which are both prevalent and easy to interpret. From our study, we exclude black-box approaches such as ANNs, as our intention is to uncover correlations between features and their impact on energy consumption forecasting.

A. Historical Averaging Models

Historical averaging models are the simplest prediction methods used by utility providers [15]. Here, we consider models used by two Independent Service Operators (ISOs), a utility, as well as a Time of the Week (ToW) averaging model.

1) New York ISO (NYISO): The next day's prediction is calculated by taking the hourly average of the five most recent days with the highest average load [12]. Days are chosen from a pool of ten previous days, starting from two days prior to prediction. Selected days exclude weekends, holidays, past DR event days or days on which there was a sharp drop in the energy consumption. Also, a day is selected only if the average consumption on that day is more than 25% of the last selected day. For a DR event day, an optional morning adjustment factor can be used.

2) California ISO (CAISO): The next day's prediction is calculated by taking hourly averages across three most recent days with highest average consumption value [13] out of a pool of ten previous days. Weekends, holidays, and past DR event days are excluded. Similarly to NYISO, an optional morning adjustment factor can be used [15].

3) Southern California Edison (CASCE): The next day's prediction is calculated by taking hourly averages across the past ten immediate or similar days [15], [14]. These days exclude weekends, holidays or past DR event days. A morning adjustment factor can be applied.

4) Time of the Week (ToW) Average Model: We define the Time of the Week (ToW) average for each 15-min interval in a week, as the kWh value for that interval, averaged over all weeks in the training dataset. As electricity consumption is closely tied to human schedules and activities, time related features are important for demand prediction [30]. ToW captures consumption variations over the duration of a day , i.e., from day to night, and across different days of the week. To capture seasonality, we further considered Time of Year (ToY) average, and total Annual average, which were consistently out-performed by ToW in our experiments. Hence we refrain from discussing them any further. We explain this result as an effect of low seasonal weather variability in the Southern California region, were we conducted our experiments.

B. Time Series

A Time Series (TS) model predicts future values based on previous observations. The commonly used Auto-Regressive Integrated Moving Average (ARIMA) [24] is defined in terms of three parameters: d, the number of times a time series needs to be differenced to make it stationary; p, the auto-regressive order, that denotes the number of past observations included in the model; and q, the moving average order that denotes the number of past white noise error terms included in the model. These parameters are derived from the Box-Jenkins test [31]. Advantages of ARIMA include the fact that it does not require domain knowledge, nor does it depend on other features. However, even if computational methods have been proposed to automatically estimate its parameters [32], human expertise to examine the partial correlogram of the time series is still required.

C. Regression Tree

Regression Tree (RT) models combine several independent features into a linear function, to represent complex data that cannot be effectively modeled by linear regression [33]. A RT recursively partitions data into smaller regions, based on ordinal or numeric features, until they can be represented by a constant or a linear regression model. This confers the model with several advantages [16]. Its flowchart/tree representation enables domain users to interpret the impact of different features on predicted values. Further, predictions, unlike training, are fast to compute, as the process only involves a tree look-up [9].

IV. Experimental Setup

A. Dataset Description

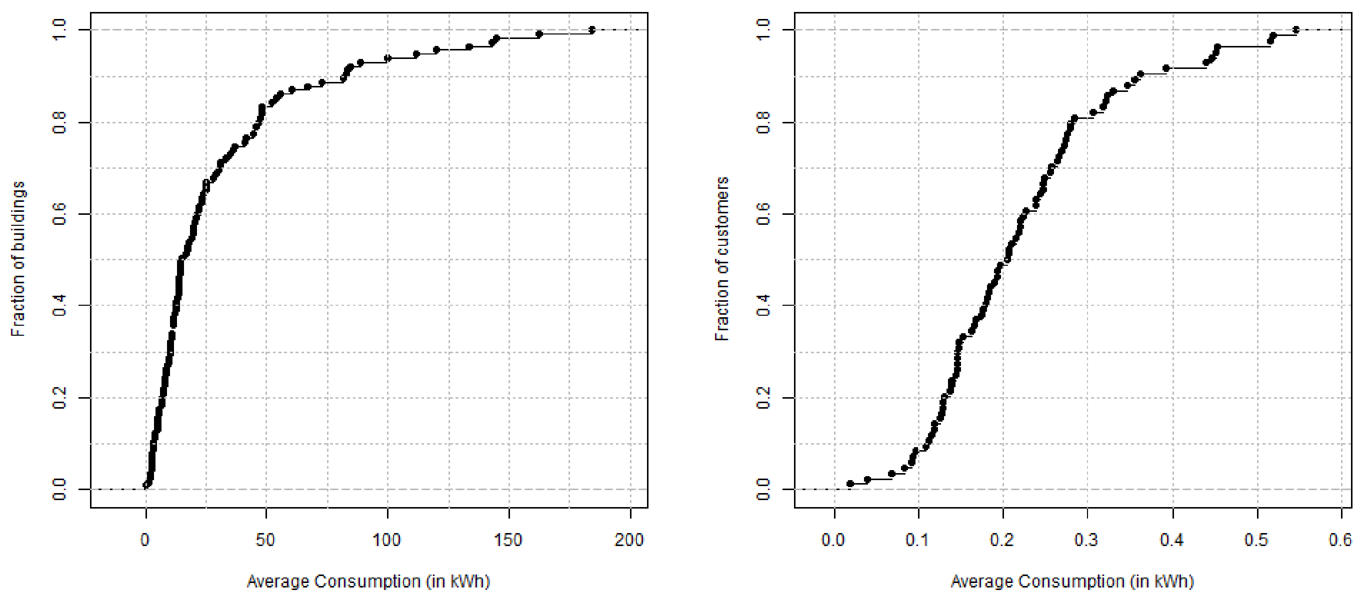
For our experiments, we consider two real-world Smart Grid datasets (see Table I).

Our building-level electricity consumption dataset from the **University of Southern California campus microgrid**[2] [7], [8] consists of observed, 15-min granularity electricity consumption values (in kWh) from 170 campus buildings, collected between Jul 2009 and Jun 2013. Out of the total 170 buildings, we retained 115 buildings, after omitting those with spurious

[2]The dataset is available upon request for academic use from the USC Facility Management Services (FMS).

TABLE I: Description of microgrid and utility datasets.

	Campus Microgrid	Utility Dataset
Number of participants	170	89
Data collection period	4 years	3 months
Data points	4 years × 365 days × 96 intervals ≈ $140 * 10^3$ points per building ≈ $16 * 10^6$ points total	3 months × 30 days × 96 intervals ≈ $8.5 * 10^3$ points per customer ≈ $0.73 * 10^6$ points total
Client type	buildings	households
Mean consumption (kWh)	large (30.52±7.65)	small (0.22±0.15)
Average variance (kWh)	122.56	0.026



(a)

(b)

Fig. 1: Cumulative Distribution Function (CDF) of average kWh consumption per 15-min interval of: (a) individual campus buildings, and (b) utility area customers.

data. We linearly interpolated missing values ($< 1\%$). The dataset comprises of a diverse set of academic buildings with teaching and office space, residential dormitories, and administrative buildings. Due to the broad spectrum of building intended use, electricity consumption values among buildings vary significantly. Figure 1a shows the cumulative distribution of 15-min average kWh values of individual buildings. 80% of all buildings consume, on average, less than 50 kWh per 15-min interval, whereas only a small number of buildings exceed 100 kWh per 15-min interval. For reference, we provide the mean and variance of average electricity consumption at the building level in Figure 5 in the Appendix.

Our fine grained electricity consumption at the customer level dataset from a **major California power utility** contains 15-min kWh consumption data from 89 household customers, collected between Feb 2013 and Apr 2013. Out of the 89 customers, 5 were filtered out, because of spurious data and high number of missing values. For the remaining customers, the data was cleaned through linear interpolation ($< 2\%$ of intervals). Figure 1b shows the cumulative distribution of 15-min average kWh values of individual customers. According to Figure 1b, 80% of all customers exhibit values of less than 0.3 15-min average kWh. For reference, Figure 7a in the Appendix, shows individual customers' mean and variance average electricity consumption.

Weather data: Two commonly used sources for weather data are NOAA [34] and Weather Underground [35]. We chose NOAA, as its data is curated [36] and processed under a quality control system [34]. The hourly temperature observations are interpolated to obtain 15-min values, in alignment with the granularity of our electricity consumption datasets.

Schedule Data: The academic schedule affects the activities in campus buildings and consequently the energy consumption patterns. The schedule data consists of information related to semester periods, working days, and holidays. The RT model uses this information for the campus dataset experiments, gaining a small advantage over the utility dataset, which lacks customers' daily schedule information.

B. Models Configuration

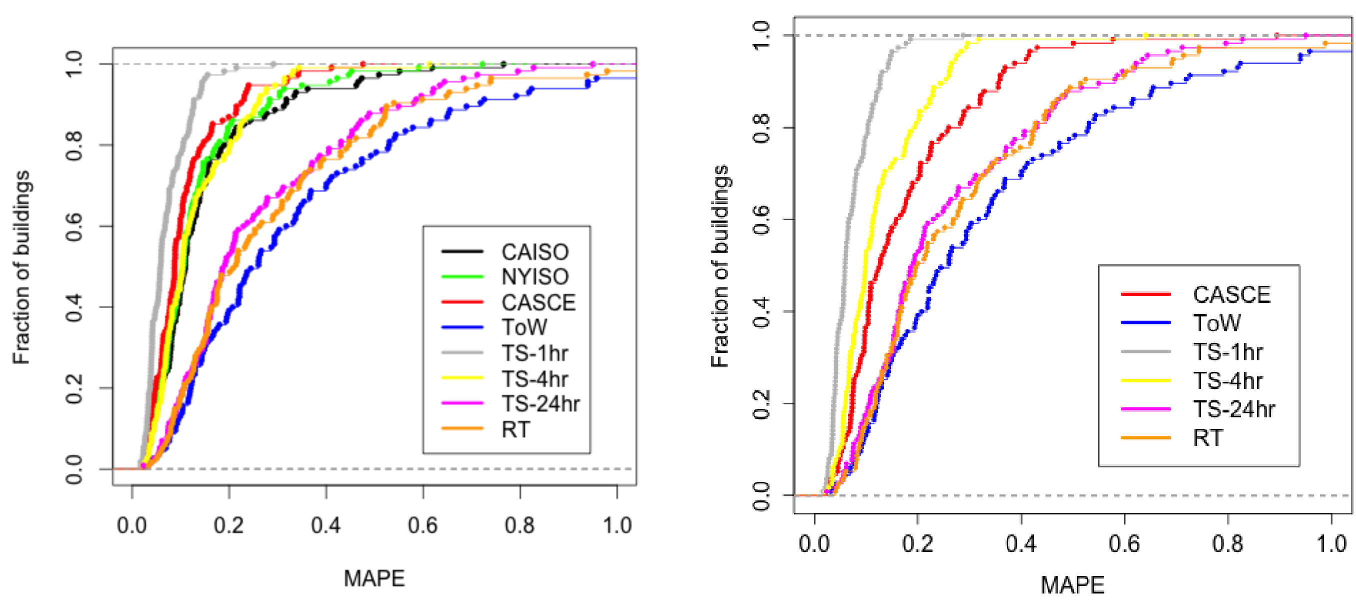
For our **campus dataset**, we conduct out-of-sample experiments for two testing periods of 12 months each: Jul 2011 - Jun 2012, and Jul 2012 - Jun 2013. For brevity, we report our observations for the first period only. Analysis of our experiments for the second period yielded similar results, hence we believe our conclusions to be robust. We build one prediction model per building. The averaging and RT models are trained over a period of two years: Jul 2009 - Jun 2011 for the first experiment, and Jul 2010 - Jun 2012 for the second. Since RT is a data-rich model, it can be costly to train as more features are included. For this reason, we picked the feature combination that offers the best prediction accuracy, based on our previous work [16]: day of the week, semester, temperature, and holiday/working day flag. TS is trained using a sliding window of 8 weeks preceding the prediction period, and predictions are made for three horizons: 1, 4, and 24 hours. In the first case, the model is retrained every hour, while in the last two cases, the model is retrained every 4 hours. The objective of using three prediction horizons is to examine how performance changes as a function of how far ahead in time the predictions are made. After a sweep test, we determined the optimal TS parameters ($(p, d, q) = (8, 1, 8)$) for both testing periods.

The **utility dataset** has a shorter span of historical data, so we conduct a single experiment. The averaging and RT models are trained over a period of two months from Feb 1 - Mar 31, 2013. Predictions are made for the month of April 2013. The TS model is trained using the same sliding window and prediction intervals as for the campus data. In this case, the optimal TS parameters were found to be $(p, d, q) = (4, 1, 4)$.

We test the prediction methods' accuracy in terms of Mean Absolute Percentage Error (MAPE). We calculate MAPE values considering either *weekdays only* or *all-days*. ISO and utility models are defined for weekdays only (cf. §III-A). For a meaningful comparison when all days are considered we modified CASCE to include all days in its averaging. We selected CASCE, because, as we show next, it outperforms the rest of the ISO models.

V. Results Analysis

A. Campus Dataset Results



(a) Workdays

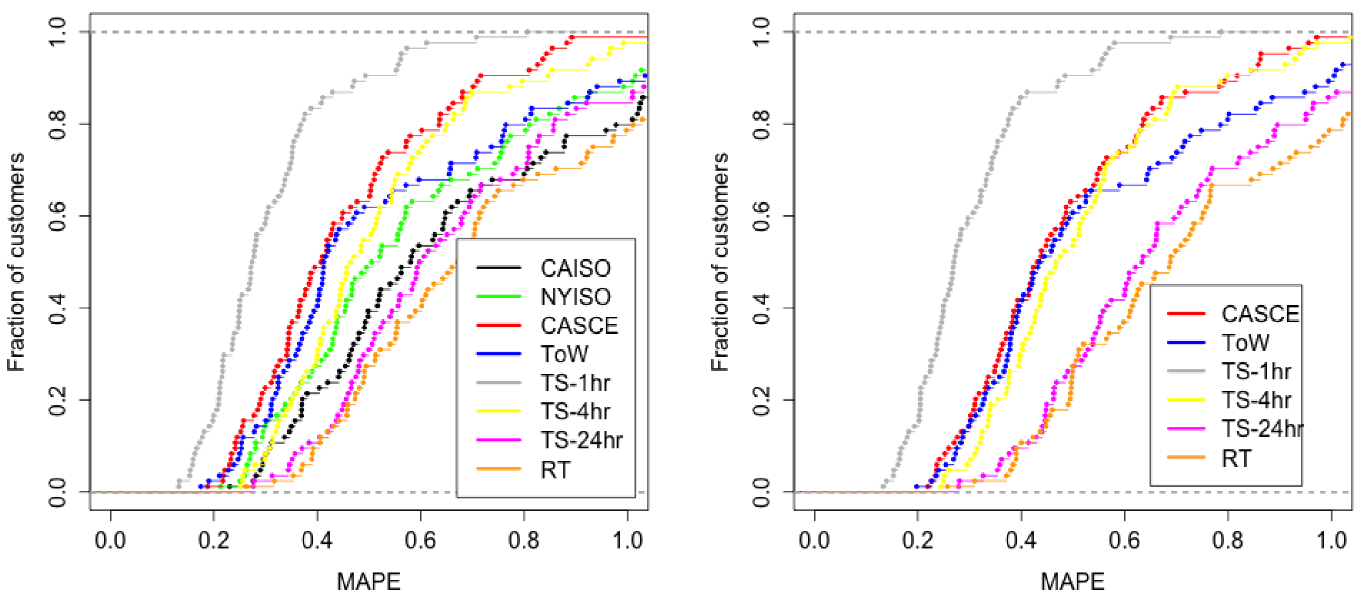
(b) All days

Fig. 2: CDF of MAPE values for campus (Jul 2011 - Jun 2012 test period).

Figure 2 shows the performance of the various models in our campus dataset. When only workdays are considered (Figure 2a), two distinct groups of models emerge: 1) TS-1hr, TS-4hr & the ISO models, and 2) TS-24hr, RT and ToW. TS-1hr model performs best among all models, with an average MAPE of $7.05 \pm 4.4\%$ for all buildings. As we increase the prediction horizon, TS-4hr (13.05% average MAPE) becomes comparable to ISO models. For larger horizons TS performance deteriorates further, i.e., 26.38% average MAPE for TS-24hr. In fact, ISO models outperform TS-24hr, achieving less than 20% MAPE for about 80% of buildings. CASCE is the best among ISO model, with an average MAPE of 10.93%. We conclude that TS is suitable for very short-term predictions only. TS-1hr and TS-4hr capture temporal locality, while ISO models capture recent, repetitive patterns. Instead, TS-24hr fails due to inadequate temporal locality at the 24hr scale, given that there is a big phase shift in usage every every 12hrs. Both ToW and RT try to capture global recurrence over longer periods but that seems to be less effective. We conclude that recent consumption observations are more accurate predictors than long term historical data.

While presently, weekends are not peak load days, since commercial activity is naturally reduced at that time, as we move towards real time DR, predictions round the clock will become important, We address this gap by considering all days, hence evaluating weekend prediction accuracy as well. Figure 2b shows the results. TS-1hr remains the best model among all, even though slightly deteriorates (average MAPE increases to 7.13% from to 7.05% for weekdays). This is due to the fact that TS benefits from temporal locality, and thus does not distinguish between workdays and non-workdays. TS-4hr and TS-24hr are also consistent between workdays and non-workdays. RT and ToW remain almost the same for workdays and all days. RT uses the day of the week and workday/holidays as features, whereas ToW is calculated by taking averages for each day of the week, thus inherently capturing the workday/non-workday information. Finally, CASCE is affected the most, as its average MAPE increases from 10.93% to 17.29%. Since our modified CASCE is trained on all days of the week, not just workdays, we explain the degradation to its performance as a result of campus buildings' weekend loads being dramatically lower (approaching base load) than weekdays. Such difference is less acute for residential customers, hence we anticipate modified ISOs not to be affected when considering the utility dataset.

B. Utility Dataset Results



(a) Workdays

(b) All days

Fig. 3: CDF of MAPE values for utility customers.

Figure 3 shows the performance of the various models in our utility dataset. Unlike campus buildings, which have larger electricity consumption, we don't observe a clear grouping into two piles, for utility customers. This indicates that there is more variability for utility customers, hence the models perform variably. TS-1hr outperforms other prediction techniques for workdays, achieving MAPE below 30% for 80% of the customers (Figure 3a). Overall, TS-1hr achieves 30.74% average MAPE (median: 27.56%). As the prediction horizon increases, CASCE outperforms TS-4hr, whereas TS-24hr becomes comparable to RT. In fact, CASCE performs considerably better than other ISO models, whereas RT performs worst than all other methods. ToW achieves comparable performance to CASCE for about 50% of the customers, but but its predictive ability worsens for the remaining customers.

Figure 3b shows models' predictive ability when all days are considered including weekend results. TS-1hr remains the best model among all. As in the campus experiment, CASCE's average (median) MAPE increases from 45.91% to 49.18% (40.07% to 43.45%).

Overall, while ISO models provide a fast way for consumption forecasting, they have limited very short-term prediction accuracy, compared to TS-1hr. In fact, TS-1hr achieves 33.04% average MAPE improvement over CASCE, the best performing

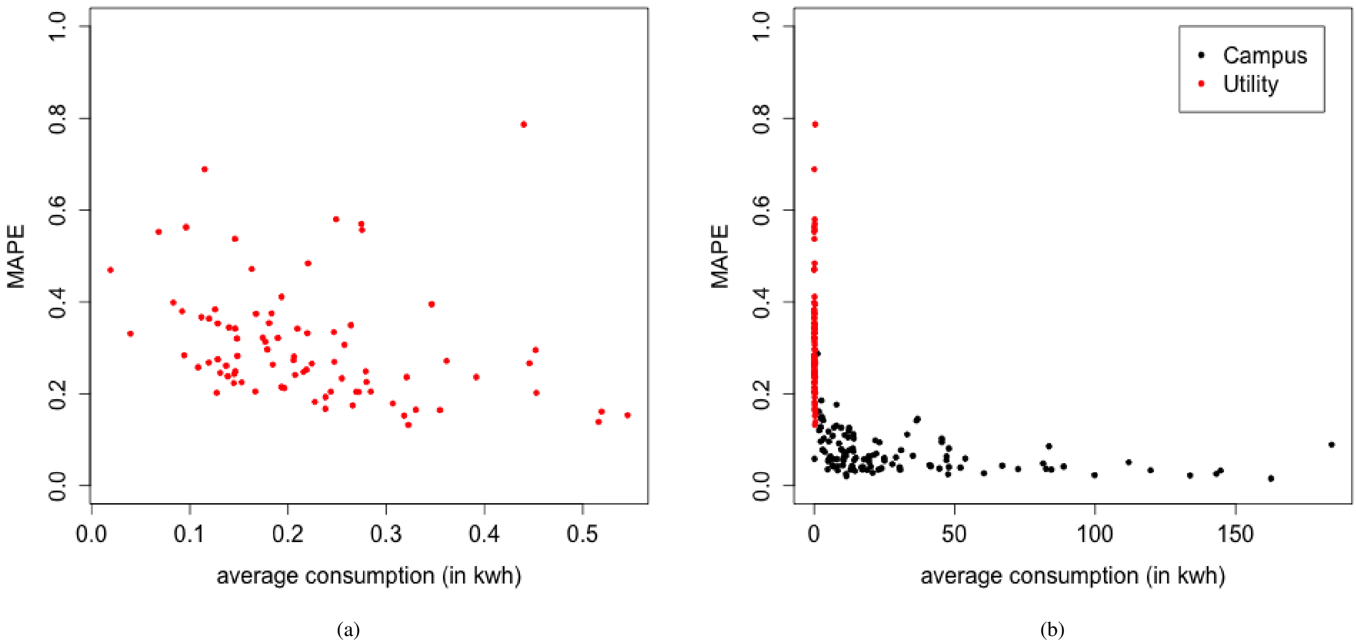


Fig. 4: TS-1hr MAPE as a function of average kWh of: (a) utility customers; and (b) utility customers and campus buildings (MAPE errors for campus buildings refer to Jul 2011 - Jun 2012 test period).

TABLE II: Average MAPE (with standard deviation) for TS-1hr for groups of buildings/customers.

avg. kWh	avg. MAPE (campus)		avg. MAPE (utility)	
	workdays	all days	workdays	all days
kWh ≤ 5	0.1204 ± 0.0649	0.1186 ± 0.0623	0.3075 ± 0.1309	0.3054 ± 0.1269
$5 <$ kWh ≤ 15	0.0743 ± 0.0382	0.0750 ± 0.0355	-	-
$15 <$ kWh ≤ 50	0.0610 ± 0.0308	0.0617 ± 0.0293	-	-
kWh > 50	0.0392 ± 0.0147	0.0416 ± 0.0199	-	-

ISO model. For medium or long-term predictions however, CASCE is the most efficient. For reference, the average (median) MAPE for CASCE is 45.91% (40.00%), for NYISO 59.86% (51.27%), and for CAISO 68.15% (58.11%)

C. From the Campus Microgrid to Utility: Lessons Learned

1) Prediction error as a function of average consumption: Directly comparing electricity consumption of utility customers to that of campus buildings (Figure 7a and Figure 5 in Appendix respectively) makes clear the vast difference in scale and variability. The maximum average 15-min kWh value for residential customers is less than 1.5, whereas for campus buildings, it exceeds 250. Next, we measure how aggregate, i.e., campus building, versus individual, i.e., residential customer in the utility area, consumption affects the models' prediction accuracy.

Figure 4 shows that TS-1hr prediction error is inversely proportional to average kWh consumption both for the utility customers and for campus buildings. For reference, we provide the same analysis for CASCE in Figure 8 in the Appendix. In fact, the higher the average consumption of buildings, the less the variability, hence the more improved performance in prediction accuracy. Note that the slope drops sharply in the 0-1 kWh region, i.e., utility customers, and is more gentle for campus buildings. For **small** electricity consumption customers, even turning on/off a light bulb causes a big change, whereas for **medium customers**, a AC unit or washing machine has to be turned on or off to have an impact on MAPE. For **large customers**, there is a lot of inertia built in. TS-1hr *better captures the skews in small customers* due to its temporal locality; the observed load spike in one interval is included when predicting the next interval. CASCE performs worse for small customers since consumption spikes may not recur at the same time across multiple days. For larger customers however, i.e., average consumption greater than 50kWh, CASCE performs similarly to TS-1hr, due to the fact that large campus buildings are more predictable in terms of consumption across days.

We explore the correlation between average consumption and MAPE further, by grouping campus buildings according to their average consumption, and calculating average MAPE for the whole group. Table II summarizes the results. For reference, Figure 6 in the Appendix presents how the absolute TS-1hr percent error varies for individual campus buildings as a function of kWh consumption. While average MAPE for large buildings (average 15-min consumption > 50kWh) is only 3.92%, it increases to 12.04% for buildings with average 15-min consumption ≤ 5kWh. The difference between average MAPE values for buildings with average 15-min consumption ≤ 5kWh, and these for small utility customers (kWh≤ 5) can be attributed to higher variability in the consumption patterns of utility customers and less for campus buildings where activities are mostly governed by pre-defined schedules.

2) ToW model: ToW performs poorly for campus buildings (Figure 2), while it performs as good as other models for about 60% of utility customers (Figure 3). A possible explanation for this difference is that the ToW model is built based on more recent dataset (past 2 months) for the utility, in contrast to 2 years worth of data for campus buildings. This timespan stretches over multiple seasons (e.g. summer/fall), which skew the overall consumption, as usage patterns on campus buildings vary significantly between summer and the academic year (fall and spring semesters). Nonetheless, ToW appears to be consistent across campus and utility customers, without any major swing in MAPE/fraction of customers, indicating resilience in its performance.

3) RT model: For campus buildings, RT outperforms the ToW baseline and is similar to TS-24hr, but performs worse than all methods in the case of the utility area. Although it is possible that RT's bad performance might be due to the absence of features (the only features used are temperature and day of the week), we conclude that RT is insufficient for very short-term predictions; it should be instead used for long-term predictions only [16], [9].

VI. Conclusions

Most previous studies on electricity consumption prediction focus on aggregated energy consumption of multiple customers, at the level of buildings, feeders or substations. we argue that prediction errors in this case are lower due to less data variability. In this paper we focus on very short-term prediction of both aggregated consumption at the building level, and most importantly, individual consumer consumption. We perform a quantitative empirical comparison of several prediction models in both cases. Our work differs from previous studies in that we use a small data granularity (15-min interval), and focus on very short-term predictions (next few hours). We argue that this approach is suited for near real-time forecasting in the context of dynamic DR. In anticipation of automated and dynamic DR, we take into consideration all days of the week, exhibiting the deficiencies of ISO models in this context.

We showed that TS-1hr is the best performing method for both building size and individual customer consumption prediction. We argued that even though ISO models perform considerably well, with CASCE being the second best in 3/4 cases, they experience a significant decline in MAPE, when considering all days of the week instead of the workdays. This indicates that for the majority of customers there is no need for large historical records for reliable prediction. Few weeks worth of data is enough for CASCE and TS-1hr. This has direct implications on predicting the consumption of new buildings/customers, and on data storage requirements. While our results are promising, due to the large errors for individual utility customers ($\approx 30\%$) as compared to buildings ($\approx 7\%$), there is a need for novel techniques for electric consumption forecasting at the individual customer level. Our future work will focus on more complex models that will take into consideration the particular characteristics of utility customers. We also plan on validating our models on larger customer datasets. One of the possible directions for our future work includes building an ensemble of the best performing models from this study. The motivation for this work is provided by the conclusions from this study that show that no single model gives the best performance in all cases and different models may perform differently for different times of the day, week, or year. To validate this further, we conducted a pilot experiment to build an *oracle* model that picks the best performing model for each prediction interval for the campus buildings. The results show an improvement for all buildings with average MAPE errors reducing to 2.64%. The results indicate potential gains from intelligently combining these models.

Acknowledgment

This material is based upon work supported by the United States Department of Energy under Award Number number DE-OE0000192, and the Los Angeles Department of Water and Power (LA DWP). The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, the LA DWP, nor any of their employees.

References

[1] S. Chan, K. M. Tsui, H. C. Wu, Y. Hou, Y.-C. Wu, and F. Wu, "Load/price forecasting and managing demand response for smart grids: Methodologies and challenges," *Signal Processing Magazine, IEEE*, vol. 29, no. 5, pp. 68–85, 2012.

[2] S. D. Ramchurn, P. Vytelingum, A. Rogers, and N. R. Jennings, "Putting the 'Smarts' Into the Smart Grid: A Grand Challenge for Artificial Intelligence," *Communications of the ACM*, vol. 55, no. 4, 2012.

[3] J. Shishido, "Smart meter data quality insights," in *ACEEE Summer Study on Energy Efficiency in Buildings*, 2012.

[4] Y. Simmhan and M. Noor, "Scalable prediction of energy consumption using incremental time series clustering," in *Workshop on Big Data and Smarter Cities, 2013 IEEE International Conference on Big Data*, 2013.

[5] F. Martinez-Alvarez, A. Troncoso, J. Riquelme, and J. A. Ruiz, "Energy time series forecasting based on pattern sequence similarity," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 8, 2011.

[6] W. Shen, V. Babushkin, Z. Aung, and W. L. Woon, “An ensemble model for day-ahead electricity demand time series forecasting,” in *International Conference on Future Energy Systems (ACM e-Energy)*, 2013.
[7] Y. Simmhan, V. Prasanna, S. Aman, S. Natarajan, W. Yin, and Q. Zhou, “Towards data-driven demand-response optimization in a campus microgrid,” in *ACM Workshop On Embedded Sensing Systems For Energy-Efficiency In Buildings (BuildSys)*, 2011.
[8] Y. Simmhan, S. Aman, A. Kumbhare, R. Liu, S. Stevens, Q. Zhou, and V. Prasanna, “Cloud-based software platform for data-driven smart grid management,” *IEEE/AIP Computing in Science and Engineering*, 2013.
[9] S. Aman, Y. Simmhan, and V. K. Prasanna, “Holistic measures for evaluating prediction models in smart grids,” *(under review)*, 2013.
[10] S. Aman, Y. Simmhan, and V. Prasanna, “Energy Management Systems: State of the Art and Emerging Trends,” *IEEE Communications Magazine, Ultimate Technologies and Advances for Future Smart Grid (UTASG)*, 2013.
[11] H. K. Alfares and M. Nazeeruddin, “Electric load forecasting: Literature survey and classification of methods,” *International Journal of Systems Science*, vol. 33, no. 1, 2002.
[12] “Emergency Demand Response Program Manual, Sec 5.2: Calculation of Customer Baseline Load (CBL),” New York Independent System Operator (NYISO), Tech. Rep., 2010.
[13] “CAISO Demand Response User Guide, Guide to Participation in MRTU Release 1, Version 3.0,” California Independent System Operator (CAISO), Tech. Rep., 2007.
[14] “10-Day Average Baseline and Day-Of Adjustment,” Southern California Edison, Tech. Rep., 2011.
[15] K. Coughlin, M. A. Piette, C. Goldman, and S. Kiliccote, “Estimating demand response load impacts: Evaluation of baseline load models for non-residential buildings in California,” Lawrence Berkeley National Lab, Tech. Rep. LBNL-63728, 2008.
[16] S. Aman, Y. Simmhan, and V. K. Prasanna, “Improving energy use forecast for campus micro-grids using indirect indicators,” in *IEEE Workshop on Domain Driven Data Mining*, 2011.
[17] J. Z. Kolter and J. F. Jr, “A large-scale study on predicting and contextualizing building energy usage,” in *AAAI Conference on Artificial Intelligence*, 2011.
[18] H. Mori and A. Takahashi, “Hybrid intelligent method of relevant vector machine and regression tree for probabilistic load forecasting,” in *IEEE International Conference and Exhibition on Innovative Smart Grid Technologies (ISGT Europe)*, 2011.
[19] B.-J. Chen, M.-W. Chang, and C.-J. Lin, “Load forecasting using Support Vector Machines: a study on EUNITE competition 2001,” *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1821–1830, 2004.
[20] Z. Aung, M. Toukhy, J. Williams, A. Sanchez, and S. Herrero, “Towards accurate electricity load forecasting in smart grids,” in *DBKDA 2012, The Fourth International Conference on Advances in Databases, Knowledge, and Data Applications*, 2012, pp. 51–57.
[21] L. Ganyun, X. Wang, and Y. Jin, “Short-term load forecasting in power system using least squares support vector machine,” in *Computational Intelligence, Theory and Applications*. Springer, 2006, pp. 117–126.
[22] T. Hong, M. Gui, M. Baran, and H. Willis, “Modeling and forecasting hourly electric load by multiple linear regression with interactions,” in *Power and Energy Society General Meeting, 2010 IEEE*, 2010, pp. 1–8.
[23] S. Fan and R. Hyndman, “Short-term load forecasting based on a semi-parametric additive model,” *IEEE Transactions on Power Systems*, vol. 27, no. 1, 2012.
[24] M. Hagan and S. M. Behr, “The time series approach to short term load forecasting,” *IEEE Transactions on Power Systems*, vol. 2, no. 3, 1987.
[25] N. Amjady, “Short-term hourly load forecasting using time-series modeling with peak load estimation capability,” *IEEE Transactions on Power Systems*, 2001.
[26] J. W. Taylor, L. M. de Menezes, and P. E. McSharry, “A comparison of univariate methods for forecasting electricity demand up to a day ahead,” *International Journal of Forecasting*, vol. 22, pp. 1–16, 2006.
[27] P. S. Kalekar, “Time series forecasting using holt-winters exponential smoothing,” *Kanwal Rekhi School of Information Technology*, vol. 4329008, pp. 1–13, 2004.
[28] K. Metaxiotis, A. Kagiannas, D. Askounis, and J. Psarras, “Artificial intelligence in short term electric load forecasting: a state-of-the-art survey for the researcher,” *Energy Conversion and Management*, vol. 44, no. 9, 2003.
[29] H. Hippert, C. Pedreira, and R. Souza, “Neural networks for short-term load forecasting: a review and evaluation,” *IEEE Transactions on Power Systems*, vol. 16, no. 1, 2001.
[30] E. A. Feinberg and D. Genethliou, “Load forecasting,” *Applied Mathematics for Restructured Electric Power Systems: Optimization, Control, and Computational Intelligence*, pp. 269–285, 2005.
[31] G. E. P. Box and G. M. Jenkins, *Time series analysis, forecasting and control*. Holden-Day, 1970.
[32] R. J. Hyndman and Y. Khandakar, “Automatic time series for forecasting: The forecast package for R,” Monash University, Tech. Rep., 2007.
[33] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Chapman and Hall, 1984.
[34] NOAA, “Quality Controlled Local Climatological Data Improvements/Differences/Updates.” 2013. [Online]. Available: http://cdo.ncdc.noaa.gov/qclcd/qclcdimprovements.pdf
[35] “Weather underground,” 2013. [Online]. Available: http://www.wunderground.com/history/
[36] N. Addy, S. Kiliccote, J. Mathieu, and D. S. Callaway, “Understanding the effect of baseline modeling implementation choices on analysis of demand response performance,” in *ASME 2012 International Mechanical Engineering Congress and Exposition*, 2013.

Appendix

Figure 5 shows the mean and variance of 2012 electricity consumption at the building level for our campus dataset. The kWh consumption data is recorded for very 15-min. The mean and variance are calculated over all 15-min intervals for one year period. We group the buildings in four categories based on their mean consumption for easier comparison and interpretation.

Figure 6 shows the distribution of absolute percentage errors (for TS-1hr) for individual campus buildings. Overall, high accuracy (lower errors) predictions can be made for buildings with large 15-min average kWh consumption, whereas prediction accuracy drops as the average kWh consumption gets smaller.

Figure 7a shows the kWh distribution of 84 customers in our utility dataset, shorted by their kWh standard deviation. The black line in the boxes shows the mdeian value for each customer. Most customers with high median kWh values, also exhibit high standard deviation.

Figure 7b shows high, low and median TS-1hr MAPE values for utility customers. Averages are computed over a testing period of one month. Overall, TS-1hr achieves 30.74% average MAPE (the median is 27.56%). This constitutes a 33.04% improvement over CASCE.

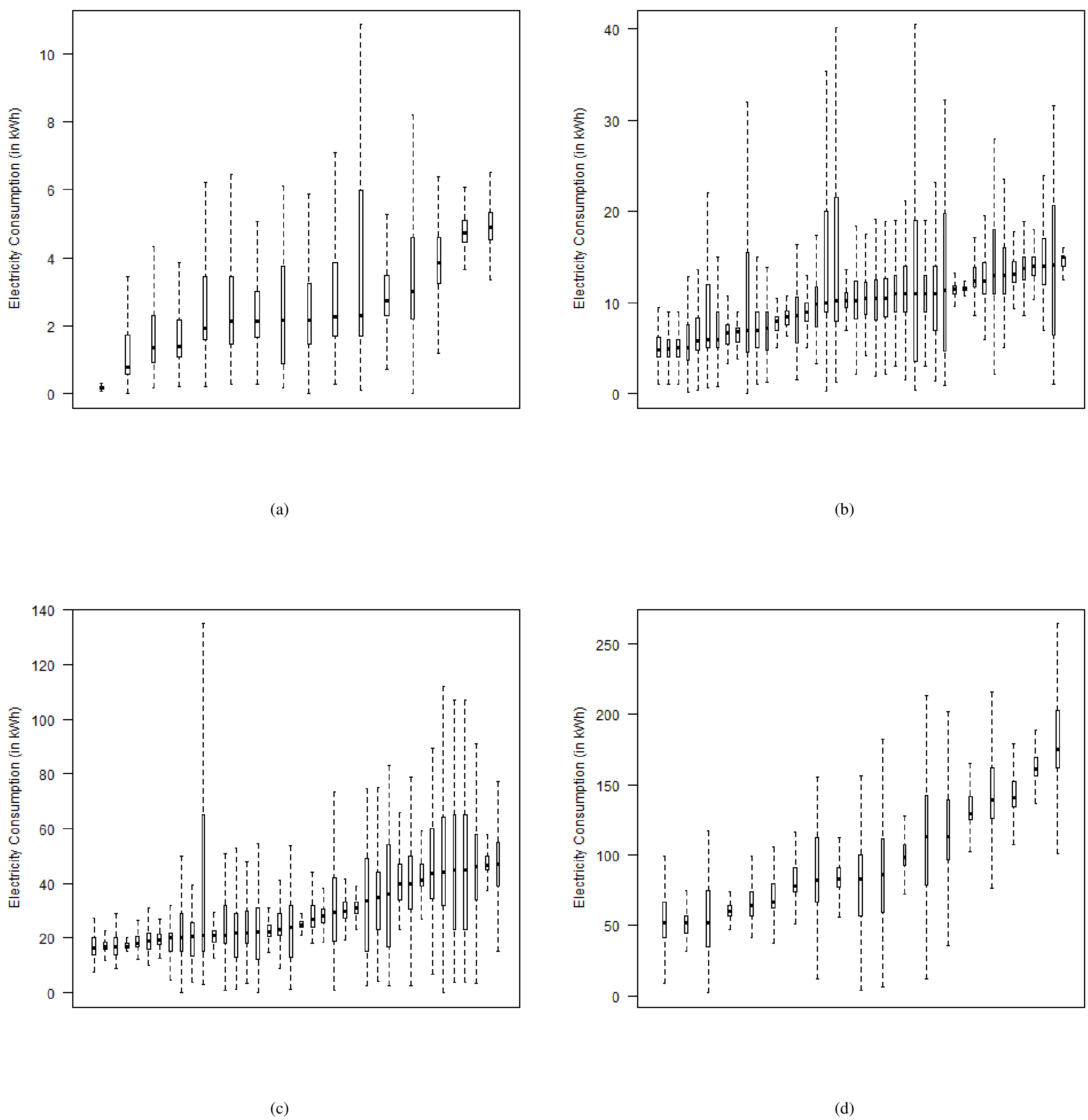
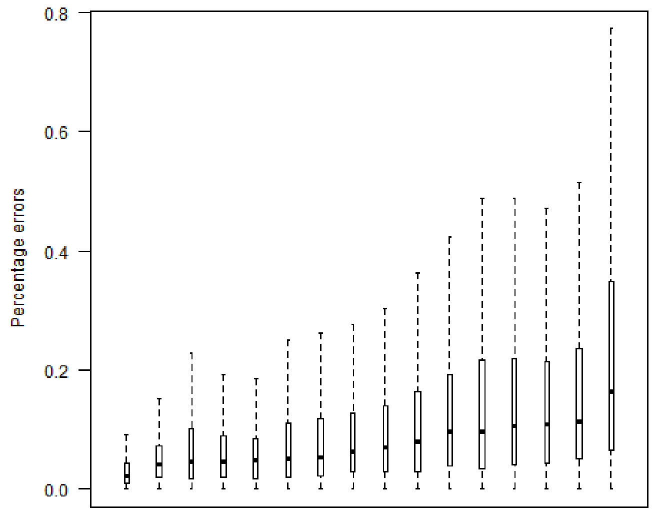
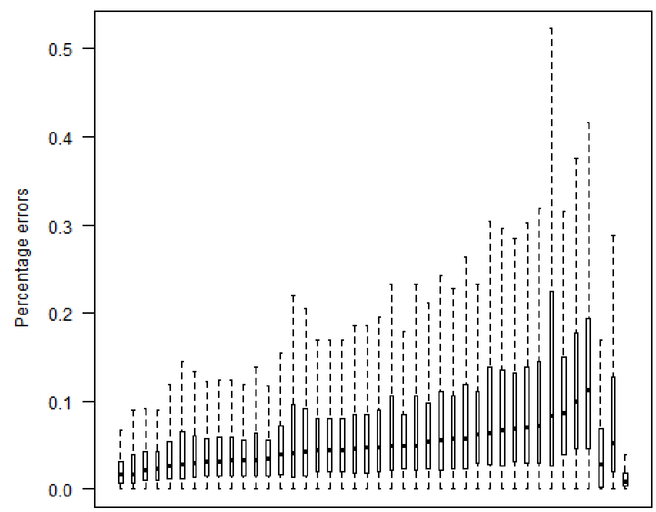


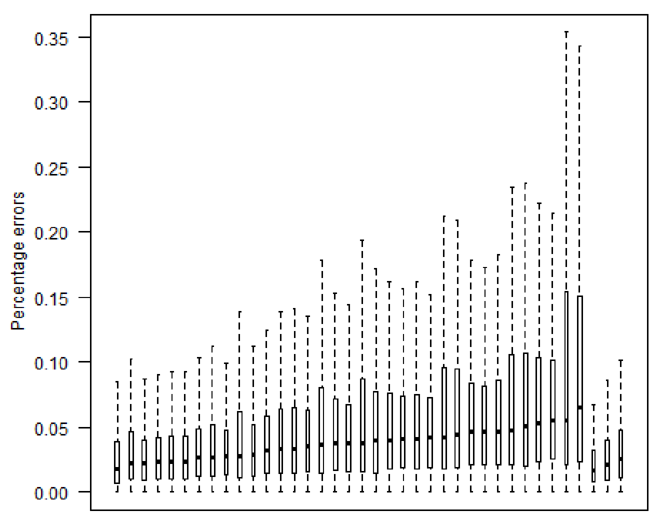
Fig. 5: Average 15-min kWh consumption for 2012 across campus buildings with average consumption of: (a) <5kwh, (b) between 5 and 15 kwh, (c) between 15 and 50 kwh, and (d) >50kwh (plots are sorted by median consumption).



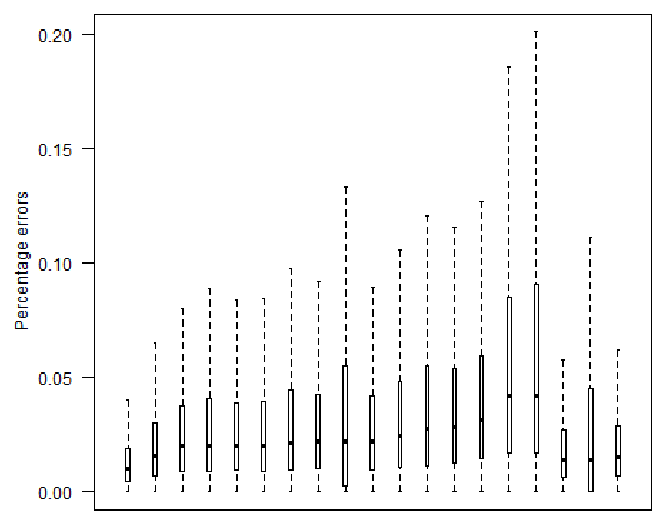
(a)



(b)

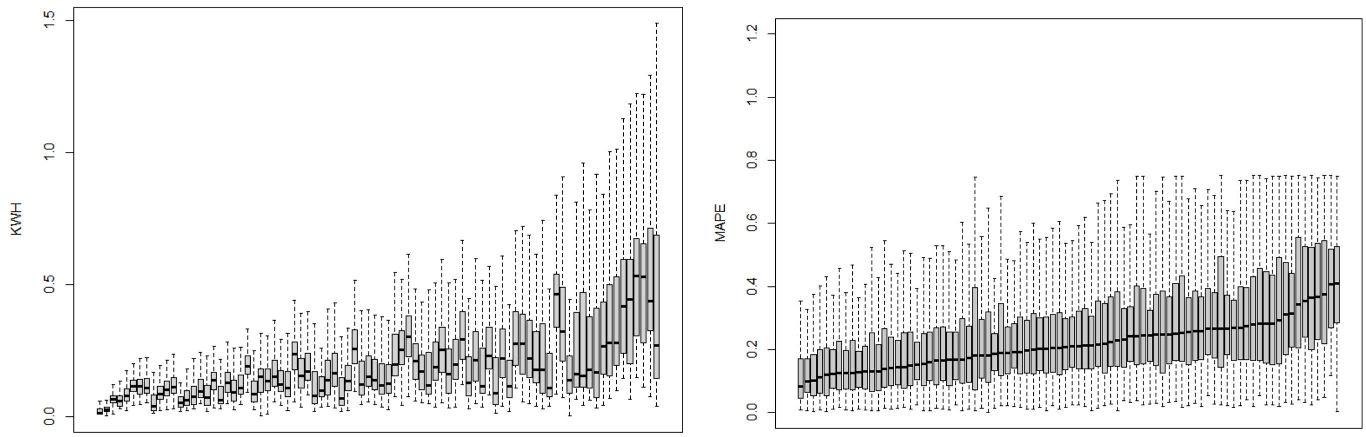


(c)



(d)

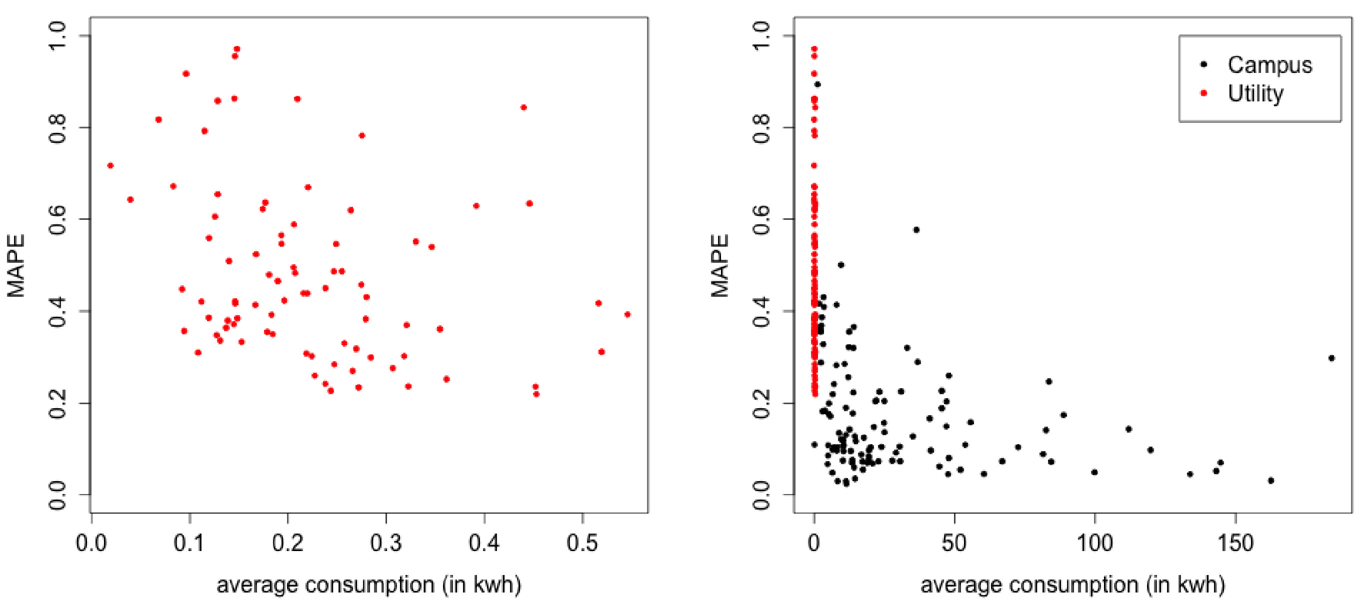
Fig. 6: Variation in absolute percentage errors for the TS 1-hr model for all days for campus buildings (Experiment 1). Results are grouped based on buildings average consumption: (a) <5kwh; (b) between 5 and 15 kWh; (c) between 15 and 50 kWh; and (d) >50kWh.



(a) AVerage 15-min kWh consumption of utility customers

(b) TS-1hr MAPE values for utility customers

Fig. 7: Distribution of kWh consumption and MAPE errors for TS-1hr for utility customers.



(a)

(b)

Fig. 8: CASCE MAPE as a function of average kWh of: (a) utility customers; and (b) utility customers and campus buildings (MAPE errors for campus buildings refer to Jul 2011 - Jun 2012 test period).