Analysis and Demand Forecasting of Residential Energy Consumption at Multiple Time Scales

Poojitha Amin
Arm Research
San Jose, CA 95134, USA
poojitha.amin@arm.com

Ludmila Cherkasova
Arm Research
San Jose, CA 95134, USA
lucy.cherkasova@arm.com

Rob Aitken
Arm Research
San Jose, CA 95134, USA
rob.aitken@arm.com

Vikas Kache
Arm Inc, ADIM Dept
San Jose, CA 95134, USA
vikas.kache@arm.com

***Abstract*—The Internet of Things (IoT) allows us to connect and monitor devices from virtually anywhere. Electric utility companies have been replacing the outdated analog meters with the new smart meter versions to automatically capture information about electricity consumption at a fine time granularity and transmitting it back to the utility provider. Energy demand forecasting is essential for Smart Grid operations. Ability to perform data analytics on the collected smart meter measurements and then predicting the electricity demands plays an important role in the utility companies' decision making for their system planning and operations. While fine granularity measurements could be useful for getting deeper insights into electricity usage patterns of different households, they might be not optimal for energy demand forecasting. We demonstrate the importance of considering and selecting different time scales in performing data analytics and demand forecasting of residential buildings. Using smart meters measurements collected from 114 residential apartments at 1 minute granularity over one year, and weather information for the same period, we design an automated process for building an efficient ensemble of linear regression models to forecast the future energy demands. This process identifies the linear portions of the daily usage patterns and creates the apartment clusters with similar usage profiles to optimize the forecasting accuracy of the designed linear regression models. It could be applied to different residential areas and geographical regions to produce the customized ensembles of fast and efficient linear regression models. Experimental results demonstrate that the proposed approach and designed performance models achieve good accuracy with 7%-19% of prediction error, and therefore could be used for optimizing the future energy distribution across the utility grid and making related critical pricing decisions.**

***Index Terms*—Smart Grid, smart meters, data analytics, demand forecast, linear regression models.**

I. Introduction

The Smart Grid initiative [1], [2] represents an unprecedented opportunity to modernize the energy industry and its infrastructure for more efficient and reliable generation and transmission of electricity, while reducing the operation and management cost. One of the enabling components of Smart Grid is the Advanced Metering Infrastructure (AMI) [3]. AMI is an integrated system of smart meters [4], [5], communications networks, and data management systems, that supports two-way communication between utilities and customers. The system enables new useful functions that were not previously possible or had to be performed manually, such as the ability to remotely measure electricity use, connect and disconnect service, detect tampering, identify and isolate outages, etc.

In 2017, about 39% of total U.S. energy consumption was due to the residential and commercial buildings [6] (with residential sector energy usage being at approximately 54% [7]). Installations of smart meters have more than doubled since 2010 — almost half of all U.S. electricity customer accounts now have smart meters. By the end of 2016, U.S. electric utilities had installed about 71 million advanced metering infrastructure (AMI) smart meters, covering 47% of the 150 million electricity customers in the United States [8].

978-3-903176-15-7 © 2019 IFIP

Smart meters are a critical component of the Smart Grid. They support automated collection of fine-grained energy consumption data. This data provides invaluable insights in the electricity usage patterns of different households over time. This collected data may enable utility companies to offer new time-based rate programs and incentives that encourage customers to reduce peak demand and better manage energy consumption and costs. For example, energy providers may offer electricity pricing schemes, where the consumers are charged higher prices during the peak hours. This will stimulate the customers to shift some of non-critical time activities (e.g., laundry, dishwashing) to other hours. In such a way, the utility companies can manage and reduce their peak demand. In recent years, the energy industry is witnessing increased research efforts and initiatives along smart meter data analytics (see a detailed survey in [9]). The United States' National Science Foundation (NSF) provides a standard grant for cross-disciplinary research on smart grid big data analytics [10].

Prediction of energy consumption in both residential and commercial buildings is an increasing area of research in recent years [11]–[20]. The methods for predicting the building energy consumption can be categorized into *engineering, statistical*, and *artificial intelligence* approaches.

Engineering methods [11], [15] are based on complex modeling of structural and thermal parameters of buildings and require comprehensive engineering methods and detailed building description, that are not always easily available. *Statistical methods* use historical data to predict energy consumption as a function of most significant variables. These models require less physical buildings understanding and offer models with a smaller number of variables. In many cases, auto regressive moving average models like ARMA and ARIMA have produced good results [17], [18]. However, the quality of the models designed by using statistical methods critically depends on the quality and quantity of historical data, the measurements granularity, and the collection of related important data, e.g., weather information. Finally, *artificial intelligence* methods based on neural networks, support vector machine, and fuzzy logic were applied to capture complex non-linear relationships between inputs and outputs [19], [20].

Understanding patterns and trends of energy consumption in the residential sector is crucial for utility companies to properly support and provision their current and future services. The accuracy of demand prediction depends on the ability to characterize energy demand patterns and recognize trends for expected changes in future demands.

In this paper, we consider the issues of workload analysis, performance modeling, and demand forecasting based on collected historical data. We analyze the *UMass Apartments dataset* – smart meters measurements collected over 1 year at a fine granularity of 1 minute [21]. We demonstrate the effect of different time scales on average energy consumption and motivate the importance of incorporating these multiple time scales in building the prediction model.

The *UMass Apartments dataset* is augmented with detailed weather temperature reports, which makes it possible to analyze and model the energy consumption as a function of temperature. By modeling the dataset along critical time scales and clustering the "similar-usage" apartments, we build an efficient ensemble of linear regression models to forecast future power demands. Experimental results demonstrate that the proposed approach and designed performance models achieve good accuracy with 7%-19% of prediction error, and therefore could be used for optimizing the future power distribution across the utility grid.

The rest of this paper is organized as follows. Section II describes the dataset used in our study, analyzes the effect of different time scales on energy consumption, and identifies critical weather features and energy usage patterns essential for modeling. Section III presents the energy forecasting model design and provides its formal definition. Section IV assesses the quality of the forecasting results to evaluate the accuracy and effectiveness of our approach. It also discusses the computing requirements of the proposed solution. Section V outlines related work. Finally, Section VI presents conclusion and future work directions.

II. Data Analysis: Effect of Different Time Scales

In this section, we describe the experimental dataset used in the study, analyze the effect of different time scales on average energy consumption, and identify the weather features most significantly associated with the energy consumption.

A. Experimental Dataset

The dataset used in our study is based on the *Apartments dataset* released as a part of the UMass Trace repository [21]. The created data collection infrastructure (built as part of the Smart* project) records the smart meter data from **real** *homes in Western Massachusetts*. It contains one minute granularity electricity usage from *114 residential apartments* collected over a period of 2016. Moreover, the *UMass Apartments dataset* includes the weather station data reports with key weather variables such as humidity, pressure, temperature, wind speed, rainfall, and visibility.

B. Importance of Different Time Scales

First, we aim to analyze the effect of different time scales on average energy consumption. The energy consumption is known to vary significantly from minute to minute, hour to hour, day to day, and month to month. Figure 1 shows the energy consumption of an arbitrarily selected apartment at a minute granularity during a day. We observe that at a minute

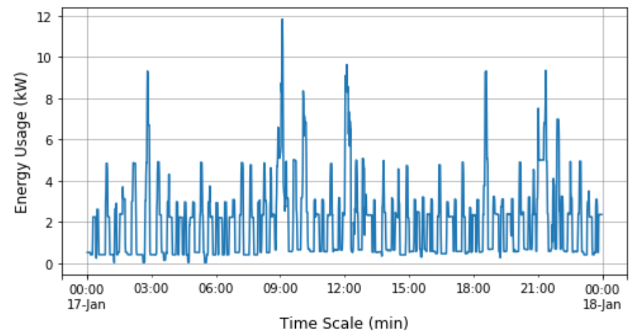

Fig. 1. Energy consumption of an apartment over a day at 1-minute scale.

granularity the energy usage is very bursty: the proportion of intervals with high and low energy values is very high, making it practically impossible to accurately forecast future energy usage values at this scale.

Figure 2 shows the energy consumption for the same apartment aggregated at an hourly scale.

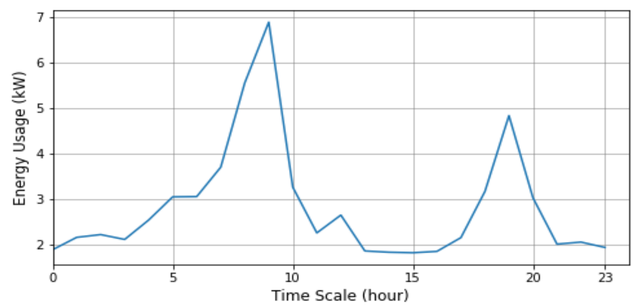

Fig. 2. Mean energy consumption of an apartment per hour over a day.

The plot is much smoother and exhibits more predictable daily usage patterns. One can see an increased energy consumption during the morning and evening hours (which are typical to "before" and "after" work daily activities).

Figure 3 shows the energy consumption aggregated for all 114 apartments at an hourly scale.

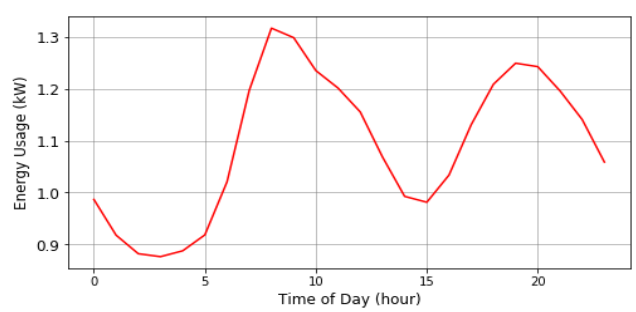

Fig. 3. Mean energy consumption across all 114 apartments on an hourly scale for a day.

While the aggregate consumption plot is smoother, it has pronounced usage peaks around the same time periods compared with the individual apartments' usage peaks (shown in Figure 2). It suggests that many apartments in the collected dataset have similar daily energy usage patterns. Most people spend active time at home during the mornings (before they leave for work) and in the evenings (after they get back home from work or school). Therefore, they tend to use the electrical appliances more often during the morning/evening ramp resulting in the increase in energy usage between 5:00 am and 8:00 am followed by an early evening time window of 3:00 pm to 8:00 pm. Two peaks shown in Figure 3 demonstrate this pattern.

Additionally, the energy consumption curve has different seasonal usage patterns, i.e., it changes with the season of the year. The climate in Massachusetts is a humid continental climate, usually warm during the summer and snowy during the winter. Figure 4 shows the mean energy consumption across 114 apartments plotted for different months.

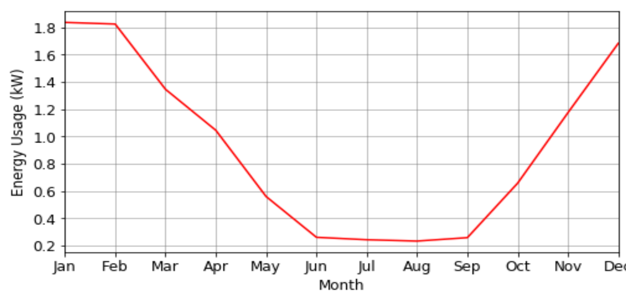

Fig. 4. Mean energy consumption for 114 apartments on monthly scale.

We clearly see the reflection of seasonal changes in the energy usage patterns across the year. There is a high energy usage during winter months, as people tend to consume more energy in a winter with a shorter daytime, to stay warm using electric blankets and room heaters. As the daylight gets longer and the outdoor temperature rises during the summer, the energy demand becomes smaller, given that it is a moderately warm state. (However, this usage pattern might be different for a different apartment complex. If apartments have an air conditioning system then the energy consumption pattern during the hot summer periods might be higher).

Another set of useful insights can be derived from the cumulative distribution of energy consumption shown in Figure 5. From this graph, we can see that there is a significant range of annual energy consumption values (up to 9 times difference) across the apartments in the set, i.e., some apartments use very little energy, while the other ones being "high-spenders".

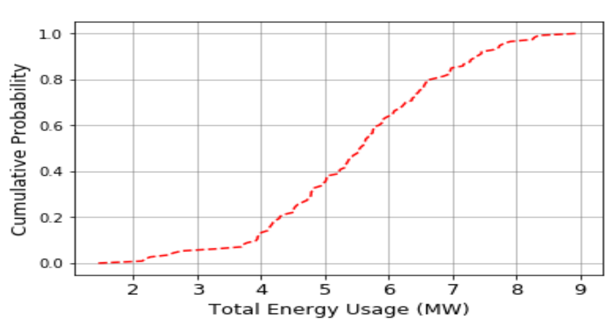

Fig. 5. CDF of energy usage.

It is an interesting observation: the apartments in the set have very different energy consumption values, in spite of having similar usage patterns over time.

C. Analyzing Weather Related Features

The goal of the study is to predict the energy demand with respect to weather temperature, season of the year, and daily time intervals. However, the weather data provided by UMass trace repository includes variables like precipitation, humidity, pressure, wind speed, etc. To identify the variables most significantly associated with the energy consumption, we plot the correlation matrix to measure the degree of correlation between them.

Python Data Analysis Library Pandas [22] provides a method $corr()$ to derive the correlation coefficients and establishes the degree of pairwise correlation between the variables. We plot the matrix of weather and energy features' comparison in Figure 6. This figure shows how the selected variables are linearly related. Correlation coefficient r assumes value in the range between -1 to $+1$ with the interpretation as follows:

- $r = +1$ indicates strongest possible linear correlation (shown as yellow in the color bar).
- $r = -1$ indicates strongest possible inverse linear correlation (shown as navy blue in the color bar).
- $r = 0$ represents no linear relationship (a light blue and green color ranges in the color bar).

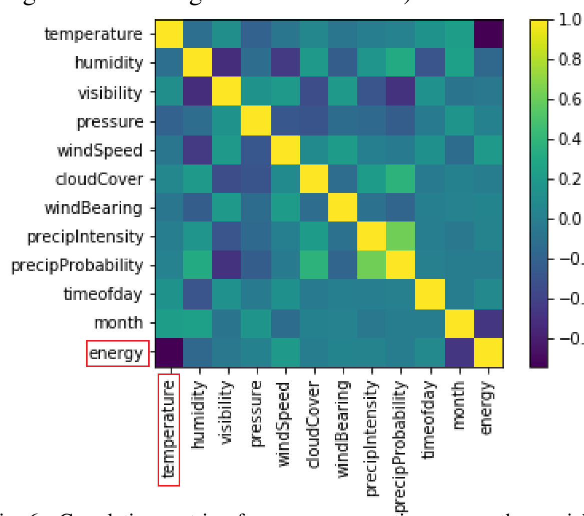

Fig. 6. Correlation matrix of energy consumption vs. weather variables.

Using analysis shown in Figure 6, we can conclude that there exists high negative correlation between temperature and energy consumption. However, the other variables like pressure, wind speed, precipitation, etc., do not correlate well with energy. Therefore, we include the weather temperature as a critical feature in our demand forecasting model.

Figure 7 shows the association between the contributing factors, i.e., a weather temperature during the month of January (12 am to 5 am time interval) and the mean hourly energy consumption during this period. The regression line running down the cloud of data points shows a good linear relationship between the weather temperature and the energy consumption.

III. An Ensemble of Linear Regression Models for Energy Demand Forecasting

This section describes the problem definition, critical features of interest, and our linear regression-based approach for modeling and predicting the energy demands as a function of a given weather temperature, season, and time of the day.

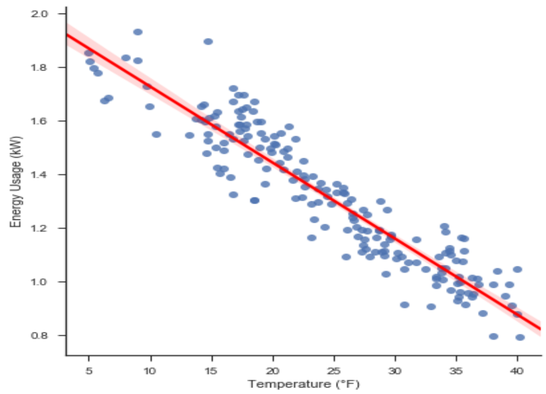

Fig. 7. Linear regression model fit for the mean hourly energy consumption data during the month of January, 12 am to 5 am time interval.

A. Energy Forecasting Model Design

Our goal is to predict the mean future energy demand per apartment per hour. In Section II, we have shown the importance of modeling the energy demands at different time scales. Therefore, we aim to build an ensemble of linear regression models characterizing the energy demand as a function of weather temperature in specific time intervals defined by a month of a season and time of the day. We design this ensemble of the models in the following way:

1) *Building the season specific models:* Residential energy sector shows the high seasonal variation in energy use, with significant demand spikes during the late autumn, winter, early spring. To capture this seasonal variation, we build 12 models, one for each month of the year.

2) *Clustering apartments based on the energy usage profile:* Energy consumption is mainly triggered by the user activities. Every home is different and the household's energy usage profiles differ from each other based on their energy demand characteristics. The variation depends on the types and number of appliances used, the frequency of use, family members working or staying at home. Hence, we cluster the apartments based on their usage pattern[1]. In this way, the model captures the differences across the demand profiles of individual homes. In our case of 114 apartments, we broadly cluster the apartments into k ($0 \leq k \leq n$) groups using k-means clustering. K-means provides an easy method of grouping together a set of apartments in a way that apartments in the same cluster are more like each other than to apartments in other clusters. By automatically iterating the model construction for different values of k, we select the number of clusters k with the best modeling accuracy.

3) *Building the models based on time of the day energy usage patterns:* As shown in Section II there are specific energy usage patterns (peaks and lows) at a daily scale which require careful modeling. We use an automated way to detect peaks and lows of energy consumption over 24 hours of a typical day (obtained by averaging the hourly demands across the same hour in the dataset). In the UMass apartment dataset, we identified five daily classes according to peaks and lows shown in Figure 3 (see Section II): 1) 12:00 am - 05:00 am, 2) 06:00 am - 08:00 am, 3) 09:00 am - 03:00 pm, 4) 04:00 pm - 08:00 pm, and 5) 09:00 pm - 12:00 am. For each of the five daily intervals, we build a prediction model reflecting the energy usage in the given time interval.

Therefore, we aim to build (in automated way) an ensemble of linear regression models for different time scales:

$$M_{i,j,k}(month_i, hours\_of\_day_j, apts\_cluster_k)$$

where $1 \leq i \leq 12$, $1 \leq j \leq 5$, $1 \leq k \leq 2$ based on the data properties of the UMass apartment dataset.

Note, that each model $M_{i,j,k}$ will be built by using the corresponding subset $Data_{i,j,k}$ of time series data from the original dataset.

B. Regression Model

Each model $M_{i,j,k}$ is built as a function of energy consumption and weather temperature using the subset of time series data $Data_{i,j,k}$ from the original dataset. Note, that each data point in this subset has two values: average hourly weather temperature and mean hourly energy usage for a corresponding clustered subset of apartments.

Therefore, we can form the following set of equations:

$$E_n^{i,j,k} = c_0^{i,j,k} + c_1^{i,j,k} \times T_n^{i,j,k}, \text{where}$$

- $E_n^{i,j,k}$ is the energy usage for hour n in $Data_{i,j,k}$;
- $T_n^{i,j,k}$ is the weather temperature for the same hour n;
- $c_0^{i,j,k}$ and $c_1^{i,j,k}$ are the regression coefficients.

To solve for $(c_0^{i,j,k}, c_1^{i,j,k})$, one can choose a regression method from a variety of known methods in the literature (a popular method for solving such a set of equations is a Least Squares Regression). Once the model is trained and deployed in a real world scenario, its forecasting accuracy might be dependent on the weather temperature coverage (ranges) available in the training data. As we gather more data over the years, and have a broader coverage of weather conditions, the model should get better at predicting energy usage even under odd weather conditions.

IV. Model Evaluation and Result Analysis

When building our prediction model, the primary goal is to predict the future energy usage. We train the model with data from the initial 25 days of the month and predict the energy usage for the last 5 days.

In order to formally evaluate the prediction accuracy of each generated model $M_{i,j,k}$ we compute for each data point in our 5 days test dataset a *prediction error*. That is, for each hourly measured mean energy consumption value E_n^{measrd}, we compare it with the predicted value E_n^{pred}. The relative error is defined as follows:

$$prediction\_error_n = \frac{|E_n^{measrd} - E_n^{pred}|}{E_n^{measrd}}$$

We calculate the relative prediction error for all the data points in the test dataset.

To assess the quality of the built model (i.e., how well the built model fits the training or test data), we use RMSE

[1]We believe that the proposed clustering step might be especially useful in case of diverse energy usage patterns due to building type specifics. For example, the apartment complex with air conditioning vs without it, residential buildings vs commercial ones, etc. These different clusters will have customized models built for them. The energy usage profiles of residential and commercial buildings are very different, and therefore, would require separately trained and built models.

measure. RMSE is the square root of the variance of the residuals. Lower values of RMSE indicate a better fit. The RMSE can be obtained by using the equation:

$$RMSE = \sqrt{\frac{1}{N}\Sigma_{n=1}^{N}\left(E_n^{measrd} - E_n^{pred}\right)^2}, \text{where}$$

N is the total number of observations used to build a model.

Table I shows the prediction error of the regression models (using the test dataset) and the 95% confidence interval as well as the goodness of the regression model fit with RMSE (using both training and testing data).

TABLE I
ERROR ESTIMATION

Month	Model Performance Metrics			
	Mean Prediction Error	95% Confidence Interval	Test RMSE	Train RMSE
Jan	0.07	0.0096	0.18	0.17
Feb	0.13	0.0186	0.22	0.21
Mar	0.10	0.0122	0.21	0.21
Apr	0.11	0.0260	0.25	0.17
May	0.16	0.0288	0.18	0.15
Jun	0.17	0.0239	0.11	0.11
Jul	0.18	0.0209	0.10	0.09
Aug	0.17	0.0240	0.09	0.09
Sep	0.18	0.0273	0.10	0.09
Oct	0.16	0.0191	0.11	0.12
Nov	0.16	0.0157	0.30	0.15
Dec	0.19	0.0223	0.43	0.30

The prediction errors are between 7%-19% which reflect a high quality of the regression results. Note, that the RMSE measure highly depends on the numerical values of the results. For months with colder temperature the absolute energy consumption values are higher, resulting in a slightly higher RMSE values.

Note that the modeling and prediction errors are higher for warmer months (because of weather temperature being less correlated with energy usage). While the relative prediction errors are higher for these months, the RMSE values are smaller (indicating that the absolute value of modeling errors are small).

Figure 8 shows the actual vs predicted energy usage for last 5 days of January, 2016. The predicted values of energy

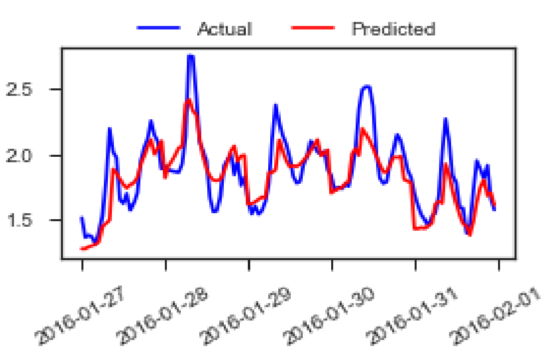

Fig. 8. Actual vs predicted energy usage for last 5 days of January, 2016.

consumption closely follow the real measurements (with 7% of mean error).

The data analysis, regression modeling, and the experiments are performed using the following computing environment:

- **Lenovo ThinkPad T440S** laptop, with on Intel(R) Core(TM) i7-4600U CPU @ 2.10GHz, 2 Core(s), 4 Logical Processor(s), 12 GB RAM, running Microsoft Windows.
  - The piecewise linear regression model was implemented using Python 3.4.3 and the related libraries/tools.

Solving linear regression equations on the modern hardware is fast. The obtained models are simple and explainable. We have performed the modeling described in the paper in less than 2 minutes time. In addition to energy demand forecasting, the designed regression models can be also used for *what-if* analysis. For example, one can model the energy consumption in winter months assuming colder or warmer than usual temperature in order to project possible ranges for energy demands. Utility companies might use this analysis across different regions to design possible management scenarios across the Grid for optimizing the desirable outcome and cost.

V. RELATED WORK

Energy demand forecasting has been broadly studied due to the problem importance and its significance for utility companies. The methods for predicting the building energy consumption can be categorized into three main categories: *engineering, statistical*, and *artificial intelligence* approaches.

Engineering methods [11], [15] are based on complex modeling of structural and thermal parameters of buildings and require comprehensive engineering methods and detailed building description that are not always easily available. To reduce the complexity of detailed engineering methods, some simplified, approximation approaches have been proposed [23].

Statistical methods use historical data to predict energy consumption as a function of most significant variables. These models require less physical buildings understanding and offer models with a smaller number of variables.

The detailed survey on regression analysis for prediction of residential energy consumption is offered in [24]. The authors believe that among statistical models, linear regression analysis has shown promising results because of satisfactory accuracy and relatively simple implementation compared to other methods. The authors discuss *top-down* and *bottom-up* approaches to modeling energy consumption. Top-down approach identifies factors defining changes in energy consumption of residential sector in the long-term. As an example, a multilevel regression (MR) model [25] is used to calculate the magnitude and significance of household features on residential energy consumption, such as housing type, house size, usage of space heating equipment, household size, and use of air-conditioning, etc. Bottom-up approach [27] aims to characterize the energy consumption at the house level and then apply this model for a segment of residential sector with similar characteristics. In many cases, the choice of the framework and the modeling efforts are driven by the specifics of the problem formulation. In [26], linear regression is used to predict the country annual energy use as a function of GDP, GDP per capita, population, population growth, and industrial growth rate.

Our modeling approach with linear regression differs from the described above: we automatically identify the time-related (linear) daily usage patterns in the overall energy usage as well as apply seasonal sub-modeling. Additionally, we optimize this process by forming similar *by energy usage* clusters of apartments.

In general, the quality of the models designed by using statistical methods critically depends on the quality and quantity of historical data, the measurements granularity, and the collection of related important data, e.g., weather information.

Finally, *machine learning* and *artificial intelligence* methods based on neural networks [19], [20], support vector machines (SVM) [28], [29], fuzzy logic [30], and Decision Trees (DT) [31] were applied to capture complex non-linear relationships between inputs and outputs. While overall, as a black-box approach, these models might produce an acceptable forecast, they lack the "explainability" of the results. Moreover, computing requirements of many artificial intelligence methods are much higher compared to regression-based models. The training phase of model building might require the cluster of GPU-based computers and could take hours. Therefore, the efficient application of these methods is still a challenge for demand forecasting problems.

VI. Conclusion and Future Work

In this work, we consider the hourly energy demand modeling and forecasting for the residential sector. Due to progress in smart meter technology, fine granularity measurements are possible and available. We show that for modeling and forecasting of energy consumption a coarser time scale is required: aggregated, hourly data lead to more accurate and reproducible results. The energy consumption in the residential sector exhibits pronounced seasonal and daily patterns. We designed an automated method for grouping the apartments with similar usage profiles and finding "peaks" and "lows" in daily data patterns to partition the data into a set of segments, which can be modeled with linear regression. Using the strong correlation of weather temperature and the energy usage, we design an ensemble of linear regression models which are built for different seasons (month) and time of the day. The prediction errors are between 7%-19%, reflecting a high quality of the regression results. We believe that by considering additional calendar events (e.g., weekends and holidays) the accuracy could be further improved. In our future work, we plan to compare performance of our linear regression models to forecasting models built using ARIMA and LSTM approaches. We aim to apply our modeling approaches to commercial buildings and analyze the differences in the energy consumption patterns and forecasting accuracy. We also plan to include the pricing information as one of the features in the model building process. This will help in designing the management strategies to optimize the energy efficiency of commercial buildings.

References

[1] What is the Smart Grid? https://www.smartgrid.gov/the_smart_grid/smart_grid.html

[2] Smart Grid, Wikipedia, https://en.wikipedia.org/wiki/Smart_grid

[3] Advanced metering infrastructure (AMI) https://www.smartgrid.gov/recovery_act/deployment_status/sdgp_ami_systems.html

[4] Smart meter. https://en.wikipedia.org/wiki/Smart_meter

[5] Smart meters. https:/www.whatissmartgrid.org/smart-grid-101/smart-meters

[6] How much energy is consumed in U.S. residential and commercial buildings? https://www.eia.gov/tools/faqs/faq.php?id=86&t=1

[7] US Energy Information Administration, Monthly Energy Review, June 2018, https://www.eia.gov/totalenergy/data/monthly/pdf/mer.pdf

[8] Nearly half of all U.S. electricity customers have smart meters. https://www.eia.gov/todayinenergy/detail.php?id=34012

[9] Y. Wang, Q. Chen, T. Hong, C. Kang: Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges. Will appear in IEEE Transactions on Smart Grid, 2018.

[10] National Science Foundation, Smart Grids Big Data Analytics. https://www.nsf.gov/awardsearch/simpleSearchResult?queryText=smart+grid+big+data+analytics

[11] H.-X. Zhao, F. Magoulès: A review on the prediction of building energy consumption. Renew. Sustain. Energy Rev., vol.16, Elsevier, 2012.

[12] Foucquier, A.; Robert, S.; Suard, F.; Stephan, L.; Jay, A.: State of the art in building modeling and energy performances prediction: A review. J. Renew. Sustain. Energy Rev., Elsevier, vol. 23, 2013.

[13] Ahmad, A.S.; Hassan, M.Y.; Abdullah, M.P.; Rahman, H.A.; Hussin, F.; Abdullah, H.; Saidur, R.: A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renew. Sustain. Energy Rev., vol. 33, Elsevier, 2014.

[14] Fumo, N.: A review on the basics of building energy estimation. J. Renew. Sustain. Energy Rev., vol. 31, Elsevier, 2014.

[15] Mena, R.; R.; Rodríguez, F.; Castilla, M.; Arahal, M.R. A prediction model based on neural networks for the energy consumption of a bioclimatic building. J. Energy Build., vol. 82, Elsevier, 2014.

[16] Lu, X.; Lu, T.; Kibert, C.J.; Viljanen, M. Modeling and forecasting energy consumption for heterogeneous buildings using a physical-statistical approach. J. Appl. Energy, vol. 144, Elsevier, 2015.

[17] J. Nowicka-Zagrajeka, R. Weron. Modeling electricity loads in California: ARMA models with hyperbolic noise. Signarl Processing, vol 82, Elsevier, 2002.

[18] C. Nichiforov, I. Stamatescu, I. Făgărăşan, G. Stamatescu: Energy consumption forecasting using ARIMA and neural network models. 5th International Symposium on Electrical and Electronics Engineering (ISEEE), 2017.

[19] S.M. Sulaiman, P. Aruna Jeyanthy, D. Devaraj: Artificial neural network based day ahead load forecasting using Smart Meter data. Biennial Intl. Conf. on Power and Energy Systems: Towards Sustainable Energy (PESTSE), 2016.

[20] Hu Y.-C.: Electricity consumption prediction using a neural-network-based grey forecasting approach. J. of the Operational Research Society, v. 68, 2017.

[21] UMass Trace Repository, Apartment dataset. http://traces.cs.umass.edu/index.php/Smart/Smart.

[22] Python Data Analysis Library https://pandas.pydata.org/

[23] S. Wang, X.Xu: Simplified building model for transient thermal performance estimation using GA-based parameter identification. Intl. J. on Thermal Sciences, vol 47, Elsevier, 2015.

[24] N. Fumo, M. A. Rafe Biswas: Regression analysis for prediction of residential energy consumption. J. Renew. Sustain. Energy Rev., vol. 31, Elsevier, 2014.

[25] G. Tso, J. Guan A multilevel regression approach to understand effects of environment indicators and household features on residential energy consumption. J. Energy, vol. 66, Elsevier, 2014.

[26] S. Tabazi, A. Aslani, H. Forotan: Prediction of Energy Consumption by Using Regression Model. Computational Research Progress in Applied Science & Engineering, vol.2, July 2016.

[27] L. Swan, V.I. Ugursal: Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. J. Renew. Sustain. Energy Rev., vol. 13, Elsevier, 2009.

[28] Son, H., Kim, C.: Forecasting short-term electricity demand in residential sector based on support vector regression and fuzzy-rough feature selection with particle swarm optimization. Procedia Engineering, vol. 118, Elsevier, 2015.

[29] Yu, W., An, D., Griffith, D., Yang, Q., Xu, G.: Towards statistical modeling and machine learning based energy usage forecasting in smart grid. ACM SIGAPP Applied Computing Review, 15(1), 2015.

[30] K. Song, Y. Baek, D. H. Hong, G. Jang: Short-term load forecasting for the holidays using fuzzy linear regression method. IEEE Transactions on Power Systems, vol.20, no.1, 2005.

[31] Yu, Z., Haghighat, F., Fung, B.C., Yoshino, H.: A decision tree method for building energy demand modeling. J. Energy and Buildings, 42(10), Elsevier, 2010.