

12-9-2015

Predicting energy demand peak using M5 model trees

Sara S. Abdelkader

Western University, sabdelka@uwo.ca

Katarina Grolinger

Western University, kgroling@uwo.ca

Miriam AM Capretz

Western University

Follow this and additional works at: <https://ir.lib.uwo.ca/electricalpub>



Part of the [Artificial Intelligence and Robotics Commons](#), and the [Software Engineering Commons](#)

Citation of this paper:

Abdelkader, Sara S.; Grolinger, Katarina; and Capretz, Miriam AM, "Predicting energy demand peak using M5 model trees" (2015). *Electrical and Computer Engineering Publications*. 72.

<https://ir.lib.uwo.ca/electricalpub/72>

Predicting energy demand peak using M5 model trees

Sara S. Abdelkader, Katarina Grolinger, Miriam A. M. Capretz

Department of Electrical and Computer Engineering
Western University, London, ON, Canada, N6A 5B9
{sabelka, kgroling, mcapretz}@uwo.ca

Abstract—Predicting energy demand peak is a key factor for reducing energy demand and electricity bills for commercial customers. Features influencing energy demand are many and complex, such as occupant behaviours and temperature. Feature selection can decrease prediction model complexity without sacrificing performance. In this paper, features were selected based on their multiple linear regression correlation coefficients. This paper discusses the capabilities of M5 model trees in energy demand prediction for commercial buildings. M5 model trees are similar to regression trees; however they are more suitable for continuous prediction problems. The M5 model tree prediction was developed based on a selected feature set including sensor energy demand readings, day of the week, season, humidity, and weather conditions (sunny, rain, etc.). The performance of the M5 model tree was evaluated by comparing it to the support vector regression (SVR) and artificial neural networks (ANN) models. The M5 model tree outperformed the SVR and ANN models with a mean absolute error (MAE) of 8.94 compared to 10.02 and 12.04 for the SVR and ANN models respectively.

Keywords: *M5 model trees; support vector regression; multiple linear regression; artificial neural networks; feature selection; predicting energy demand peak.*

I. INTRODUCTION

One of the common ways to measure energy use in a building is to determine total consumption over a certain period of time [1]. Energy consumption is the number of kilowatt-hours consumed in a system. However, the total consumption over a long time does not give an accurate description of energy usage trends for commercial buildings. For instance, typical office buildings consume most of their energy during weekday office hours and use relatively little energy at night and on weekends. This pattern requires the network to supply these buildings with a large amount of energy at certain times instead of an average amount constantly. For this reason, the rate of energy consumption, defined as demand, is the focus of this study. In some countries, such as the United States and Canada, monitoring energy demand for commercial buildings is beneficial because electricity bills are dependent on the value of the highest demand peak [2].

The predicted energy demand peak for a certain interval can be compared to the actual current demand values. If the current demand value approaches this predicted peak, the energy system can send a warning or alert to the system monitors. Thus, energy consumption can be reduced or balanced to avoid reaching the peak, hence reducing electricity bills. Buildings energy demand is influenced by many different

features, such as weather conditions, building structures, and occupant behaviours [2]. Choosing the proper feature set to train the prediction model is a key issue for machine learning methods in this area. This study assumes a sensor-based approach; it uses historical data captured from sensors and meters to predict energy demand peaks.

The contributions of this paper are first to analyze how features such as seasonal change and work load can influence the performance of prediction models. Multiple Linear Regression algorithm is used to study the impact of these features on prediction model and select proper features. Features are selected based on their correlation coefficients [3].

The second contribution is to show how M5 model trees can be used to predict energy demand for commercial buildings. The M5 model tree is not typically used for energy demand peak prediction. Support vector regression (SVR) and artificial neural network (ANN) models are commonly used for energy demand peak prediction [4]. A comparison of the results of the M5 tree and of both the SVR and ANN models indicate the suitability of M5 model trees for predicting energy demand.

The rest of the paper is organized as follows: Section II reviews related work. Section III introduces the theoretical principles of the multiple linear regression model, the SVR algorithm, the ANN algorithm, and the M5 model trees. Section IV describes the methodology, and Section V discusses the experiments and compares the results. Finally, Section VI concludes the paper.

II. RELATED WORK

Various models for predicting energy demand have been proposed in recent years [4]. These models can be divided into two categories. The first contains traditional algorithms, including time-series analysis, regression, and gray models. The second includes soft computing algorithms such as genetic algorithms, fuzzy logic, and other machine learning methods. This research has focussed on machine learning strategies; hence, the second category approaches are reviewed.

Artificial neural networks (ANNs) have attracted much attention in demand prediction [5]. Bashir and El-Hawary proposed an adaptive ANN with particle swarm optimization (PSO) to adjust network weights; which obtained good prediction precision [6]. A Bayesian neural network has been used for short-term demand prediction, and resulted in better performance than conventional neural networks [7].

Support vector regression (SVR) showed good generalization performance that outperformed other non-linear forecasting techniques [8]. A locally weighted SVR for short-term demand forecasting achieved higher accuracy than an ANN model [9]. An SVR-based model combined with fuzzy c-means (FCM) and particle swarm optimization (PSO) was proposed to forecast the short-term demand [10]. One of the popular models used in demand prediction is an autoregressive integrated moving average (ARIMA) model, which is a generalization of the autoregressive moving average (ARMA) model. ARIMA and seasonal ARIMA (SARIMA) have been used to estimate future energy demand [11]. A hybrid technique combining Takagi-Sugeno fuzzy inference system and fuzzy regression was used to predict short-term energy demand variations [12]. A hybrid algorithm combining a genetic algorithm and a local search algorithm was developed to overcome problems of premature convergence to local optima problems [13].

Each of these soft computing algorithms has its advantages, but also has certain weaknesses. ANNs are powerful, but suffer from opacity problems because it has a large number of parameters to be tuned and suffer from the danger of overfitting [5]. The generalization ability of SVRs depends strongly on adequate setting of parameters; such as the penalty coefficient, the kernel parameters, and the width of the loss function [9].

This study proposes a novel demand prediction approach with feature selection using multiple linear regression and M5 model trees. The goal of this study is to closely inspect the effect of humidity, seasonal change, weather conditions, and workload on demand prediction. An M5 model tree is proposed for demand peak prediction, a model which is specific for non-linear continuous problems.

III. REGRESSION MODELS

A. Multiple linear regression models

A regression model is a mathematical model of the relationship between the target variable and the input variables in a given design space [14]. Multiple linear regression models are widely used to provide estimates of parameter impact as well as predictions of the response variable at arbitrary points [15]. Often, the regression variables interact, i.e., the effect of a change in x_i on y depends on the value of x_j . In such cases, the model is expressed as:

$$y = \alpha + \sum_{i=1}^m \beta_i x_i + \sum_{i=1}^m \sum_{j=i+1}^m \beta_{i,j} x_i x_j + \varepsilon, \quad (1)$$

where α is the intercept of the response surface with the y -axis and the $\{\beta_i | 1 \leq i \leq m\}$ are known as the effects or coefficients. ε is the error due to lack of fit. Correlation coefficient R is the partial derivative of the target dependent variable y with respect to the various independent input variables x_i and x_j and is calculated by:

$$R_{y,x_i x_j} = \frac{\sqrt{r^2 y x_i + r^2 y x_j - 2 r y x_i r y x_j r x_i x_j}}{1 - r^2 x_i x_j} \quad (2)$$

where y is the prediction variable and x_i, x_j are the input features. $r y x_i$ is the correlation coefficient between y and x_i , $r y x_j$ is the coefficient between y and x_j , and $r x_i x_j$ is the coefficient between x_i and x_j . $r x y$ is defined by:

$$\text{Cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

where \bar{x} is the mean of x and \bar{y} is the mean of y [15].

Correlation coefficient is an important indicator of the impact of each feature on the regression curve and the prediction error. The sign and the absolute value of the correlation coefficient describe the direction and the magnitude of the relationship between two variables. The greater the absolute value of a correlation coefficient, the stronger is the relationship between the input and target variables [14].

B. Support vector regression (SVR)

The basic idea of support vector regression (SVR) is to find a regression model function representing the relationship between the input parameters and the target [16]. SVR is firmly grounded in the framework of statistical learning theory, known as the Vapnik and Chervonenkis (VC) theory [17]. Suppose that we are given training data $\{(x_1, y_1), \dots, (x_n, y_n)\} \in X \times \mathbb{R}$, where X denotes the space of the input set (e.g., $X = \mathbb{R}^d$). These might be, for instance, sensor and meter readings. In SVR, the goal is to find a function that has at most ε deviation from the actually obtained targets y for all the training data and at the same time is as linear as possible. In many cases, the optimization problem is non-linear. Therefore, slack variables ξ are introduced to provide a *soft margin* loss function which can be defined as:

$$\begin{aligned} & \text{minimize} \quad \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi_i + \xi_i^* \\ & \text{subject to} \quad \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon, \\ \xi_i, \xi_i^* \end{cases} \quad (4) \end{aligned}$$

where w is the coefficients and $\langle w, x \rangle$ denotes the dot product in X , b is a constant, ε is the precision and the constant $c > 0$ determines the trade-off between the linearity and the extent up to which deviations larger than ε are tolerated.

C. Artificial neural networks (ANN)

Artificial neural networks (ANN) are related to biological neural networks. They consist of neurons which execute functions cooperatively and in parallel [18]. Artificial neural networks (ANN) consist of an input layer, hidden layer(s), and an output layer. The input layer has one neuron corresponding to each input parameter. The number of hidden layers depends on the problem to be solved. The result of an ANN depends on the number of neurons in the hidden layer. Optimizing the number of hidden-layer neurons leads to a result closer to the optimum. Hidden-layer neurons can be selected using optimization techniques or trial-and-error methods. The output layer has one neuron for each output. Three types of networks are commonly used in ANN applications: feedforward

networks, competitive networks, and recurrent associative memory networks. The backpropagation (BP) algorithm is one of the dominant ANN learning algorithms and is convenient for feedforward networks [18].

D. M5 model trees

The “model tree” is a technique for dealing with continuous class learning problems. It was developed by Quinlan [19] and was exemplified in a learning algorithm known as the “M5 model tree”. A model tree is like a regression tree, but it builds trees whose leaves are associated with a multivariate linear model. The nodes are then chosen over the attributes that maximize the expected error reduction as a function of the standard deviation of the output parameters. Building the model tree consists of three steps:

- i) Building the initial tree: A decision-tree induction algorithm is introduced to create a tree. Instead of maximizing the information gain at each interior node, a splitting criterion is presented that minimizes the intra-subset variation in the class values down each branch.
- ii) Pruning the tree: this is based on minimizing the estimated absolute error of the multiple linear regression models. It starts from each leaf by using the regression plane rather than a constant value [20].
- iii) Smoothing the tree: this is done to compensate for severe discontinuities that cannot be avoided between adjacent linear models at the leaves of the pruned tree.

IV. METHODOLOGY

The methodology is divided into two phases, as illustrated in Figure 1. Phase I first processes the data set and generates new features that express the impact of seasonal changes and workload. Multiple linear regression is used to evaluate the impact of these features on the prediction model. Features are selected based on their regression correlation coefficients. In Phase II, M5 model trees are used to predict energy demand. The impact of data set normalization on the results of the M5 model tree is explored. The following subsections provide details of the research methodology.

A. Phase I

1) Data processing

To conduct this study, data including energy demand values with their related temporal features were selected and retrieved from meter and sensor readings.

Weather history data such as temperature, humidity, and weather conditions (sunny, rainy, etc.) were obtained. To obtain a better idea of the influence of both seasonal changes and workload on demand values, the data were expressed in new forms such as date as season, day in week, and day in month:

- a) Date as season: has a discrete value in the set {1, 2, 3, 4}, where 1 is winter and 4 is fall. It provides a more comprehensive understanding of the effect of the seasonal changes on demand values.
- b) Day in week: indicates the effect of work load on the demand peak regardless of the weather. Day in week is assigned one of seven values: 1 for Sunday, 2 for Monday, and so on to 7 for Saturday.
- c) Day in month: is assigned a discrete value according to the order of a given day in the month.

2) Feature selection with multiple linear regression

The aim of feature selection is to select the most valuable features to establish an efficient predictor for the learning algorithm. Irrelevant features are discarded to minimize model dimensionality. Selecting the features used in prediction streamlines the calculation while increasing the possibility of refining the model accuracy. Multiple linear regression was used to study the impact of each feature on demand values and was calculated by:

$$Y = X\beta + \epsilon, \quad (5)$$

where β is the vector of coefficients, X is the model matrix, and ϵ is the error due to fit. The model matrix consisted of the collection of all features chosen to test the prediction model. The correlation coefficients R were calculated using equations 2 and 3.

The second step was to arrange the features in descending order according to their correlation coefficients. To find the data set with the optimum features number, this ordered list was divided into smaller subsets. The first data subset started with the first two features on this list. The next data subset included the previous data subset and added the next feature on the list. This process continued until the final data subset with all features on the list. Finally all these subsets were tested using the prediction model to decide on the optimum number of features. The optimum number was selected based on the prediction model with the lowest error.

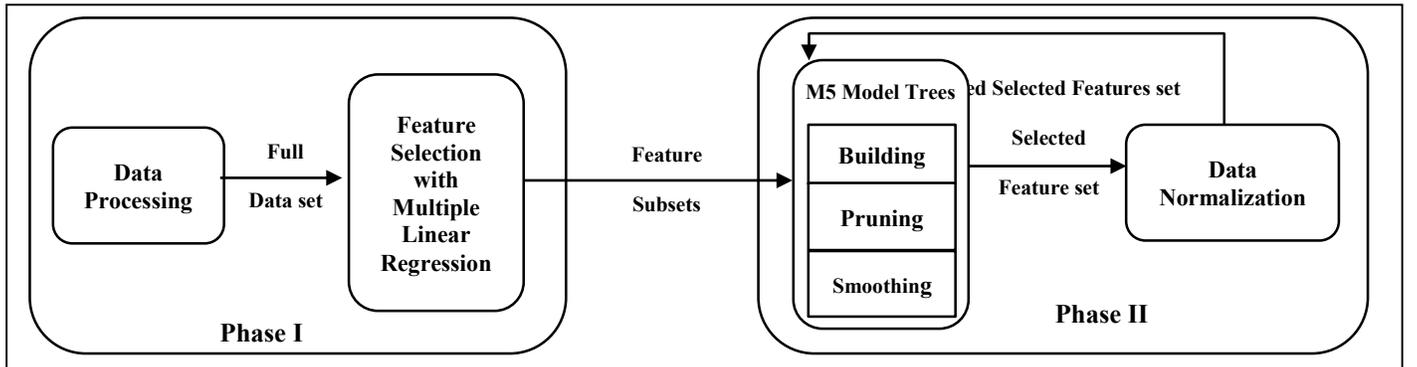


Fig. 1: Methodology framework

B. Phase II

1) M5 model trees

The feature subsets obtained from phase I were tested using M5 model trees. The M5 model trees were created using a three-step procedure.

Building the initial tree

The features used in building the tree are precisely those that contribute to decisions in nodes subordinate to the current one. These attributes are feature subsets generated from the feature selection process. The standard deviation reduction (SDR) used as the splitting criterion was defined as:

$$\text{SDR} = \text{sd}(T) - \sum_i \frac{T_i}{T} \times \text{sd}(T_i), \quad (6)$$

where T_1, T_2, \dots, T_n are the sets defined by splitting data at the node.

Pruning the tree

First, the absolute difference between the predicted demand value and the actual value was averaged for each of the training examples that reached the node. Then this average was multiplied by the factor $(n + v) / (n - v)$, where n is the number of training examples that reached the node and v is the number of features in the model at that node. Pruning was applied from the bottom up; the error for the regression model at an internal node was compared to the error for its sub-tree.

Smoothing the tree

The smoothing procedure first used the leaf model to compute the predicted value and filtered that value along the path back to the root while smoothing each node by combining it with the value predicted by the linear regression model for that node. The smoothing formula was defined as:

$$p' = \frac{np + kq}{n + k}, \quad (7)$$

where p' is the prediction passed up to the next higher node, p is the prediction passed to this node from its children, q is the value predicted by the model at this node, n is the number of training instances that reach the child node. k is a constant.

2) Data normalization

Normalization, which is adjusting values measured on different scales to a notionally common scale, was used to test the impact of data non-formality. M5 model trees are typically used with un-normalized data; however, this work explores the effect of data normalization on the accuracy of the prediction model. Normalization was applied to the final feature subset selected from the M5 model tree prediction. The impact of data normalization was tested by comparing the M5 model tree prediction model accuracy with and without normalization.

V. EXPERIMENTS AND EVALUATION

This study was conducted in collaboration with Powersmiths, a company that aims to help in building a sustainable future by supplying products and services that support reducing electricity waste [21]. Powersmiths is located in Brampton, approximately 26 km from Toronto International Airport. They have developed a sustainability management

system called Windows on the World (WOW). WOW provided this study with historical data assembled from building meters and sensors. Below, a detailed description of the implementation and evaluation of the methodology are provided.

A. Phase I

1) Data processing

Readings for energy demand measured in kilowatts (KW) were captured on a time scale varying between five-minute to fifteen-minute intervals. Weather history information was obtained from Weather Underground [22]. Data from Toronto International Airport were added to the data set. The data consisted of temperature, humidity, conditions, visibility, Dewpoint and SeaLevelPressure. The original data set consisted of 13,208 tuples representing readings of electricity demand over a three-year interval from 2011 to 2013. The original data set needed to be cleaned and processed. Processing and preparing the data set included the following steps:

- i) Eliminating tuples with missing and invalid values because they significantly affected the prediction.
- ii) Aggregating the demand readings recorded for each day by finding the maximum readings of the demand from the given readings for each day. This aggregation is convenient for understanding the daily peak.
- iii) Aggregating each feature readings in the dataset by calculating its mean value.

This process resulted in a cleaned and consistent data set of 2000 tuples. The data set was then split into two sets, with 85% of old data assigned as the training set and the other recent 15% of data used for testing. Each of the training and test sets was divided into two subsets, the Input and Target sets. The training set was fed to the prediction model for learning. The test set was used to evaluate the prediction model and to calculate the prediction error. The predicted demand vector was calculated from feeding the training set to the model. This vector was compared to the target test set, which have the actual demand values.

2) Feature selection

The multiple linear regression model was implemented in R, which is a free open-source language and environment for statistical computing and graphics developed at Bell Laboratories [23]. Figure 2 shows the result of features correlation coefficients values.

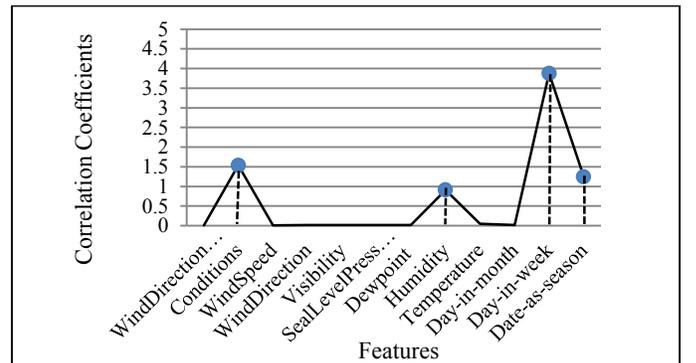


Fig. 2: Multiple linear correlation coefficients of the attributes.

SeaLevelPressure, Dewpoint, and WindDirection scored the lowest correlation coefficients. These low values indicate the weak impact of these features on the demand value. Date-as-season, Day-in-week, Humidity, and Conditions had the highest coefficients. These high values indicate the strong impact of weather and workload on demand. Features were ordered in a descending order according to their correlation coefficients and were divided into feature sets to be used in training the prediction model. In phase II the M5 model tree is used to select the optimum number of features according to the lowest error value of the prediction model.

B. Phase II

1) M5 model trees

The M5 model tree was implemented in R using the Weka package that enhances the original Quinlan M5 algorithm [24]. Root mean square error (RMSE) and mean absolute error (MAE) were used to compare the feature subsets used and evaluate prediction accuracy. Figure 3 shows the experimental results for M5 model tree performance using the feature subsets. The subset including the features Date-as-season, Day-in-week, Humidity, and Conditions resulted in the lowest errors in the prediction model.

2) Data normalization

Values of the selected feature set varied in scale from integers belonging to the interval [1,4] to other scales that belong to the real numbers and have different ranges. Feature such as weather conditions had string values. This kind of non-conformity in feature value scales could have a significant impact on the predicted values. Therefore, all values were scaled to (0, 1). Weather condition values were mapped to integer values in order to match with other features domain. Figure 4 shows the comparison between the model accuracy before and after normalization. The normalized feature set performed better than that of the non-normalized set.

C. Experimental results

The “e1071” package was used to implement SVR in R. The statistical model was trained using the normalized selected feature set obtained from the M5 model tree experiments. Tuning was applied to achieve the best fit to the SVR model. The ANN model was implemented using the Stuttgart Neural Network Simulator (SNNS) library. The “RSNNS” package was used to wrap the SNNS functionality and make it accessible in R. The same normalized selected feature set was used to test the ANN model.

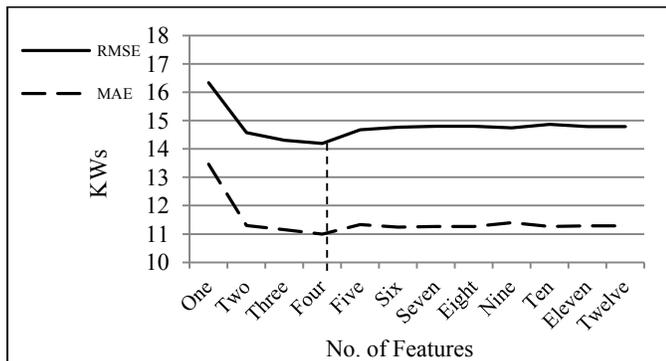


Fig. 3: Results of testing M5 model trees on feature subsets.

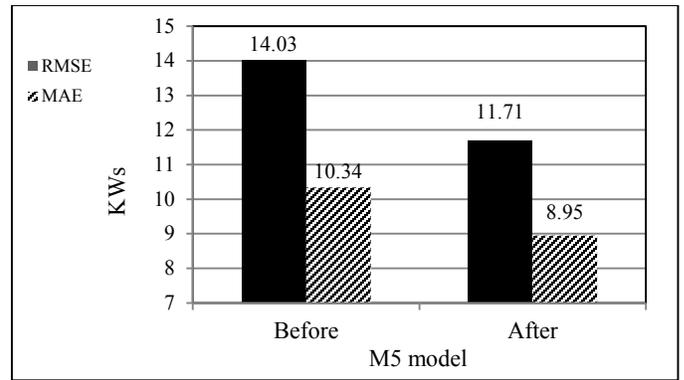


Fig. 4: M5 model tree performance before and after normalization.

The comparison between M5 model tree performance and both SVR and ANN models is shown in Figure 5. M5 model tree outperformed SVR and ANN models with MAE of 8.95, compared to 10.03 and 12.05 KWs for SVR and ANN models respectively. M5 model tree scored 11.71 KWS in RMSE compared to 14.13 and 18.16 for SVR and ANN respectively. Finally, the M5 model tree was used to predict both daily and monthly demand peaks. The most recent data for year 2013 was used for testing the model. August 2013 was selected to demonstrate the daily peaks. Figure 6 shows the comparison between the actual and predicted daily peaks in this month. The entire year data were selected to test monthly peaks prediction. Figure 7 shows the comparison between the actual and predicted peaks for each month in year 2013. The dashed line represents the predicted monthly peak values. The dots in the figure show that the M5 model tree predicted the peak in the same day as the actual one, but with a difference in the values according to the MAE of 8.95 KWs.

VI. CONCLUSIONS

Predicting energy demand peaks can be beneficial to balance the energy demand and reduce electricity bills. This paper discussed the capabilities of M5 model trees in predicting energy demand peaks for commercial buildings. The proposed framework uses multiple linear regression for selecting features based on their correlation coefficients. Experimental results showed that the selected feature set can improve prediction model efficiency. The performance of the M5 prediction model improved remarkably after normalizing the data set values. The data set used in this study was provided from sensors and meters of Powersmiths, a green energy company located in Canada.

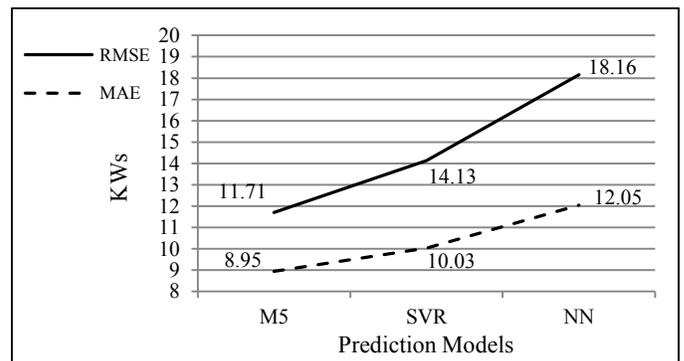


Fig. 5: Comparison of M5, SVR, and NN models

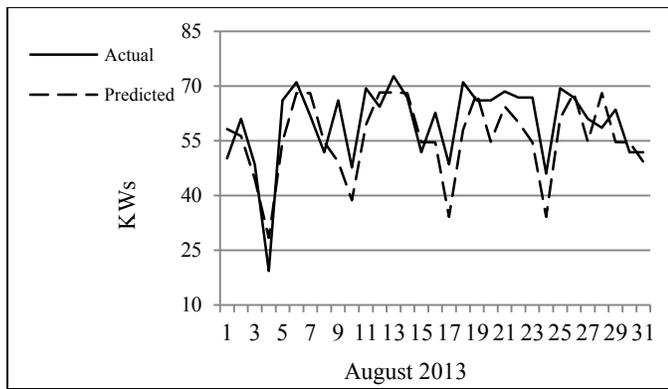


Fig. 6: Actual vs. predicted daily demand peaks.

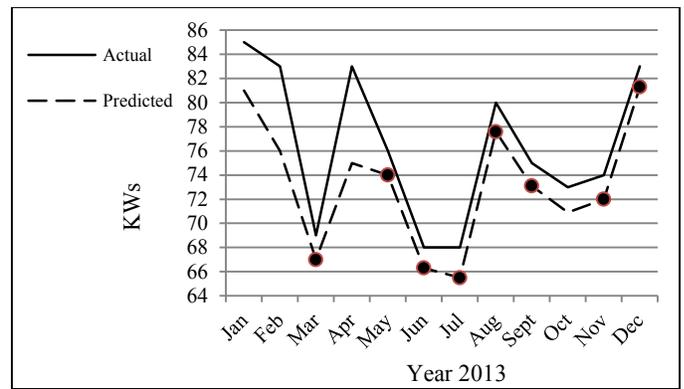


Fig. 7: Actual vs. predicted monthly demand peaks.

The results showed that M5 prediction model outperformed both SVR and ANN with an MAE of 8.94 compared to 10.02 and 12.04 for the SVR and ANN models respectively.

Further examination could be pursued to improve the accuracy of this model. Eliminating weekends from the data set to avoid weekly cyclist nature could be considered. Increasing the size of training set and different feature selection methods will be considered.

ACKNOWLEDGMENT

This research was supported in part by an NSERC CRD at Western University (CRDPJ 453294-13). Additionally, the authors would like to acknowledge the support provided by Powersmiths.

REFERENCES

- [1] U.S. Energy Information Administration (EIA) :<http://www.eia.gov>.
- [2] M. Albadi, and E. El-Saadany, "A summary of demand response in electricity markets." *Electric power systems research*, vol. 78, no. 11, pp.1989-1996, 2008.
- [3] F. Harrell, "Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis," Springer Science and Business Media, 2013.
- [4] L. Suganthi and A. Samuel, "Energy models for demand forecasting—A review," *Renewable and Sustainable Energy Reviews*, vol.16, no. 2, pp. 1223-1240, 2012.
- [5] A. Ahmad, "A review of applications of ANN and SVM for building electrical energy consumption forecasting", *Renewable and Sustainable Energy Reviews* , vol. 33, pp. 102-109, 2014.
- [6] Z. A. Bashir and M. E. El-Hawary, "Applying wavelets to short-term load forecasting using PSO-based neural networks," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 20–27, 2009.
- [7] P. Lauret, E. Fock, R. N. Randrianarivony, and J-F. Manicom-Ramsamy, "Bayesian neural network approach to short-time load Energy Conversion and Management, vol. 49, no. 5, pp. 1156–1166,
- [8] E. E. Elattar, J. Goulermas, and Q. H. Wu, "Electric load forecasting based on locally weighted support vector regression," *IEEE Transactions on Systems, Man, and Cybernetics C*, vol. 40, no. 4, pp. 438–447, 2010.
- [9] J. Wang, W. Zhu, W. Zhang, and D. Sun, "A trend fixed on firstly and seasonal adjustment model combined with the ϵ -SVR for short-term forecasting of electricity demand," *Energy Policy*, vol. 3, no.11, pp. 4901-4909, 2009.
- [10] G. Atsalakis, "New technology product demand forecasting using a fuzzy inference system," *Operational Research*, vol. 14, no. 2, pp. 225-236, 2014.
- [11] C. Moreno-Chaparro, "State of the art of electricity demand forecasting based on wavelet analysis and a nonlinear autoregressive model NAR," *Engineering Applications (WEA)*, IEEE, pp. 1-6, 2012.
- [12] K. Li, H. Su, and J. Chu, "Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: a comparative study," *Energy and Buildings*, vol. 43, no. 10, pp. 2893-2899, 2011.
- [13] P. Joseph, V. Kapil, and J. Matthew, "Construction and use of linear regression models for processor performance analysis," *High-Performance Computer Architecture*, IEEE Twelfth International Symposium on, pp. 99-108, 2006.
- [14] C. Halim, and H. Ozturk, "Estimating energy demand of Turkey based on economic indicators using genetic algorithm approach," *Energy Conversion and Management*, Vol.45, no. 15, pp. 2525-2537, 2005.
- [15] J. Cohen, C. Patricia, S. G. West, and L. S. Aiken, "Applied multiple regression/correlation analysis for the behavioral sciences," Routledge, 2013.
- [16] D. Basak, P. Srimanta, and C. Dipak, "Support vector regression," *Neural Information Processing-Letters and Reviews*, vol. 11, no. 10, pp. 203-224, 2007.
- [17] V. Vapnik, and A. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory of Probability and its Applications*, vol. 16, no. 2, pp. 264–281, 1971.
- [18] E. Guresen and G. Kayakutlu, "Definition of artificial neural networks with comparison to other networks," *Procedia Computer Science*, vo. 3, pp. 426-433, 2011.
- [19] J. Quinlan, "Learning with continuous classes," *Proceedings, 5th Australian Joint Conference on Artificial Intelligence*, vol. 92, pp. 343-348. 1992.
- [20] D. Solomatine and X. Yunpeng, "M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China," *Journal of Hydrologic Engineering*, vo. 9, no. 6, pp. 491-501, 2004.
- [21] Powersmiths, "Powersmiths: Power for the Future," 2010, <http://ww2.powersmiths.com>
- [22] Weather Underground: Toronto-Pearson International Airport Forecast, <http://www.wunderground.com/weather-forecast/Canada/Ontario/Toronto.html>
- [23] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [24] Machine Learning Group, "Weka 3 - Data Mining with Open Source Machine Learning Software in Java," 2012, <http://www.cs.waikato.ac.nz/ml/weka>.