Model Selection Criteria for Short-Term Microgrid-Scale Electricity Load Forecasts

Sankaranarayanan Subbayya, Jorjeta G. Jetcheva, and Wei-Peng Chen
Fujitsu Laboratories of America
Sunnyvale, CA, U.S.A.
{ssubbayya, jjetcheva ,wchen}@us.fujitsu.com

***Abstract*—The electricity grid is evolving from a monolithic centralized system to a smart distributed system, composed of distributed and renewable generation resources, where power supply and demand balancing is needed at a microgrid scale. In this paper, we explore model selection criteria for short-term microgrid-level load predictions. To this end, we experiment with five different models in the context of usage traces from six diverse sites collected over a period of eight months. We find that model selection is heavily influenced by the variability in the data and that models which do not use weather forecast information but rely only on historical usage data perform better on sites with highly variable loads.**

***Keywords*-load management, load modeling, smart grids**

I. INTRODUCTION

The proliferation of local and renewable electricity generation is transforming the electric grid from a centralized system where load monitoring, prediction and control are performed by utility companies, to a distributed system where microgrids composed of individual homes, office buildings, or federated groups thereof are becoming active participants in energy generation. Accurate short-term load forecasting at the microgrid level is thus becoming increasingly more important for enabling localized supply-demand balancing of the grid. At the same time, forecasting at a small scale is inherently more challenging than forecasting at large scales due to the higher variability of load relative to its mean.

In this paper, we investigate the effectiveness of three popular univariate models (which use load information) and two multivariate models (which use both load and weather information) in predicting hourly day-ahead electricity usage loads, by evaluating their performance in the context of traces collected at six different sites over an eight-month period. We find that different sites may require different models for accurate forecasting, and that model selection is influenced primarily by the variability of the data both in terms of overall prediction accuracy and optimizing daily peak prediction. Peak under-prediction, in particular, is a critical metric for the effectiveness of a prediction model since it could impact the effectiveness of demand response measures, trigger expensive peaker plant usage, and increase the probability of blackouts. With better prediction of peak demand, proper actions such as balancing demand and supply, informing customers of critical peak events, and issuing attractive demand response pricing, can be taken in advance, before the actual peak demand occurs. We also find that in the context of our sites, models that do not use weather information lead to comparable and often better predictions than models that do utilize such information.

II. RELATED WORK

Electricity load forecasting has been an important area of research over the past several decades. The vast majority of papers in the literature have focused on short-term prediction of aggregate load over large sections of the electricity grid, representing utility service areas or nation-wide electricity usage data (e.g., [1]-[4], [11]-[17]). With the emergence of the smart grid and distributed generation capabilities in recent years, the research community has turned to microgrid-level forecasting (e.g., [5]). Forecasting has proven to be difficult, since microgrid-scale loads exhibit high levels of variability, which we also observe in our study.

Short-term forecasting has been the subject of extensive study as well (e.g., [1], [3]). No single short-term forecasting model has been shown to perform consistently better than all others, which is why we focus on model selection criteria rather than trying to find a single model that works well for all sites.

Univariate models that base their predictions on past usage history alone and do not take into account weather information have been shown to be effective in predicting short-term load forecasts (e.g., [1], [2]), which is in concert with our findings as well.

In this paper we focus on small-scale short-term forecasts, which are becoming increasingly more relevant as renewable generation and microgrids become more widespread. We compare the prediction accuracy of five different models across a diverse set of sites, and come up with practical model selection criteria based on our results.

III. LOAD TRACES

We collected power consumption traces from six buildings over an eight-month period between August 2011 and April 2012. Usage reports were generated at 30-minute intervals, 24 hours a day. We converted the traces into hourly traces and removed weekends, holidays, and off-peak hours (6pm to 9am) since the

loads for those were nearly constant and did not present a modeling challenge.

In the examples included in this paper, site names and locations are not specified, and the absolute values of the loads are not shown, but are normalized relative to the maximum load for each site, in order to protect the privacy of the information as requested by our data source.

Our sites include both office buildings and factory buildings and thus present a diverse mix of load patterns (Fig. 1). Some sites display a strong seasonal behavior, with a marked increase in energy use during the summer and winter months, e.g., Sites B, E, and F, corresponding to heavy use of air conditioning and heating respectively (Fig. 2). Others show partial seasonality with an increase in energy use only during the summer but not during the winter, e.g., Sites A, C, and D.

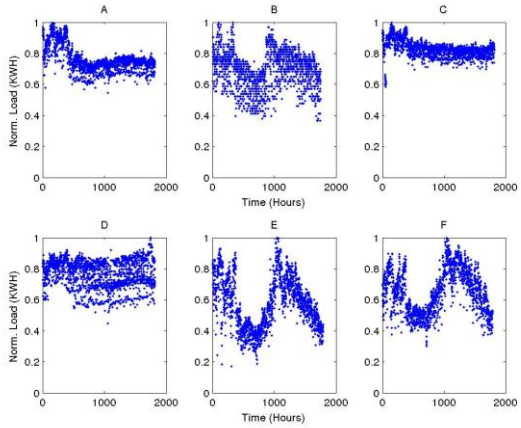

Figure 1 Normalized hourly power consumption for all sites over the period August 2011 – April 2012.

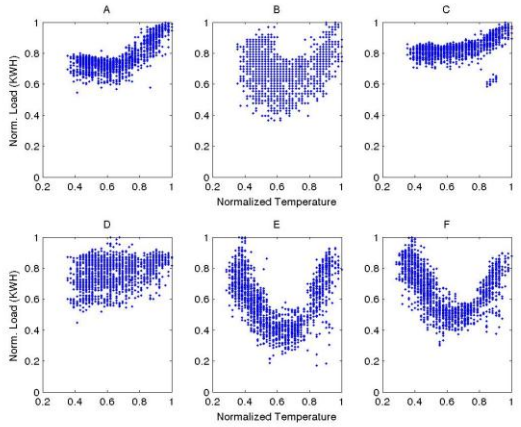

Figure 2 Hourly variations of normalized load vs. temperature for all sites over the period August 2011 – April 2012.

The relationships between load and temperature at different sites indicate that while some sites depend on electricity for both their air conditioning and heating needs, others use other sources of energy for heating (e.g., gas). Furthermore, some sites exhibit a much lower increase in usage due to temperature, e.g., Site C, indicating that air conditioning constitutes a lower portion of their overall electricity usage than at other sites, e.g., Site E.

We also observe that sites differ in the normalized ranges of load they use over time, with Sites B, E and F experiencing ranges that are roughly twice the size of the ranges of the other sites (Fig. 1). Sites B and D on the other hand display the highest variability of load for a given temperature (Fig. 2). In addition, the highest variability in the loads at sites that rely on electricity for air conditioning (e.g., Sites E and F) occurs in late summer and early fall (Fig. 1), which we attribute to varying numbers of employees being on vacation and dealing with the beginning of the school year.

Table 1 shows the ratio of the standard deviation and the mean of the hourly loads at each site, with Site E showing the most pronounced variations, with a standard deviation of 28% of its mean load.

Site	Std/Mean (%)
A	10.5
B	16.7
C	6.4
D	11.3
E	28.0
F	21.5

Table 1 Ratio of the standard deviation and mean of the hourly loads from all six sites for the entire measurement interval.

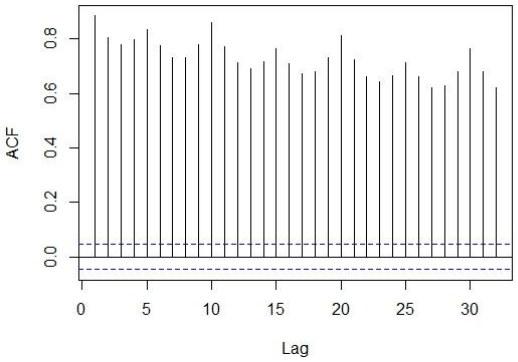

Figure 3 ACF of hourly loads for Site E.

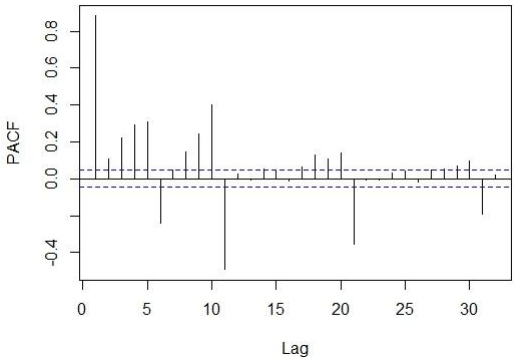

Figure 4 PACF of hourly loads for Site E.

Figs. 3-4 show the Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) of the hourly loads at Site E. High correlation of the load time series at 1, 5 and 10 hour lags can be observed in both ACF and PACF. The ACF and PACF functions for loads for the other sites showed a similar trend. The 1-hour lag indicates the correlation between the loads of consecutive hours, while the 5-hour lag indicates the correlation between morning, mid-day and evening loads. Finally, the 10-hour lag indicates the correlation of loads across days.

IV. Prediction Models

We selected five load forecasting models for our study since they have been shown to perform well for day-ahead hourly load predictions. These are ARMA (Autoregressive Moving Average) ([6]), ARIMA (Autoregressive Integrated Moving Average) ([6]), Multiple Regression ([18]), Double Exponential Smoothing ([1]), and Random Forest ([9]).

The deviations in model-predicted hourly loads and peak daily loads from the observations are quantified in terms of Mean Absolute Percentage Error (MAPE) of hourly loads:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{P_t - \hat{P}_t}{P_t}\right| \times 100, \quad (1)$$

where $P_t, \hat{P}_t$ are the observed and predicted power at time t, and n is the number of observations.

We compute MAPE for peak predictions by locating the actual daily peak load for each day and its time of occurrence, and comparing its magnitude to that of the predicted load at the same date and time, to estimate the percentage daily peak error. The percentage errors in model-predicted daily peaks are then averaged to produce a mean peak error. In particular,

$$\text{MAPE}_{\text{pl}} = \frac{1}{np}\sum_{t=1}^{np}\left|\frac{P_{tpl} - \hat{P}_{tpl}}{P_{tpl}}\right| \times 100, \quad (2)$$

where MAPE_{pl} is the MAPE for peak loads; $P_{tpl}, \hat{P}_{tpl}$ are the observed and predicted loads, respectively, at the time of occurrence of observed daily peak load, *tpl*, and *np* is the number of peaks used for testing the model.

We use the same method to compute under-prediction errors by averaging the errors across all days for which the peak is under-predicted.

A. Multiple Linear Regression

We use a multiple linear regression model to establish the relationship between the load (response variable), temperature and hour of the day (predictor variables). Data analysis of the hourly loads and air temperatures for all sites showed that loads vary with temperatures and hour of the day. Air temperature is treated as a numeric variable, while hour of the day is treated as a categorical variable.

P_i, the power consumption at hour *i*, can be expressed as:

$$P_i = a_0 t_i + a_1 t_i^2 + b_i, \quad (3)$$

where t_i is air temperature at time *i*, and b_i is a constant computed for each hour of the day.

A k-fold cross validation with k=18 is used to test the validity of the model. In the k-fold cross validation, the data is divided into k equal data sets, where the kth set was used for testing and all the remaining sets were used for training, thus testing all k sets. Each set consists of 10 days of hourly loads.

Site	MAPE for training sets (%)	MAPE for test sets (%)
A	3.2	3.4
B	8.1	8.6
C	2.6	2.7
D	5.7	5.9
E	12.9	13.6
F	8.7	9.1

Table 2 Results from k-fold cross validation for the Multiple Regression model.

Table 2 shows the MAPE for the training and test sets from the 18-fold cross-validation of the multiple linear regression model for the different sites. The errors in the model predictions for both the test and training sets are high for Site E, which has the highest standard deviation of hourly loads. The difference between training and test errors is less than 1% for all sites, indicating that the multiple linear regression model used does not overfit the data.

B. SARMA and SARIMA

Linear regression models capture the relationship between predictor and response variables but do not account for all the dynamics seen in the time series. Autoregressive Moving Average models on the other hand use internal structure such as autocorrelation and seasonal variation seen in past data to make future predictions. A general ARMA model can be represented as follows:

$$P_t \phi(B) = \theta(B)\varepsilon_t, \quad (4)$$

where P_t and ε_t are the power usage and noise at time *t*, respectively.

The Autoregressive (AR) polynomial, $\varphi(B)$, is defined as follows:

$$\varphi(B) = (1 - B\phi_1 - B^2\phi_2 - ..... - B^p\phi_p) \quad (5)$$

The Moving Average (MA) polynomial, $\theta(B)$, is defined as follows:

$$\theta(B) = (1 + B\theta_1 + B^2\theta_2 + ... + B^q\theta_q), \quad (6)$$

where p and q represent the orders of the AR and MA polynomials respectively, and B is the backward shift operator ($B^n P_t = P_{n-t}$).

The addition of differencing to the ARMA model leads to the development of ARIMA models that accounts for the non-stationarity in the time series. A general ARIMA model can be defined as follows:

$$P_t \phi(B) \nabla^d = \theta(B) \varepsilon_t, \quad (7)$$

where $\varphi(B)$ and $\theta(B)$ are the same as in (5) and (6) and the differencing operator ∇^d is defined as follows:

$$\nabla^d \mathrm{P}_t = (1-B)^d P_t \quad (8)$$

The dependence on past power loads tends to be more pronounced at multiple seasonal lags, e.g., power consumption varies with hour of the day and day of the week. Multiplicative Seasonal ARIMA models (SARIMA) have autoregressive and moving average polynomials with seasonal lags embedded in them. The SARIMA (*p,q,d,P,Q,D,s*) model can be expressed as follows:

$$P_t \phi(B) \phi(B^s) \nabla^d \nabla_s^D = \theta(B) \theta(B^s) \varepsilon_t \quad (9)$$

and

$$\nabla_s^D \mathrm{P}_t = (1-B^s)^D P_t, \quad (10)$$

where $\phi(B^s)$ and $\theta(B^s)$ are the AR and MA components of the seasonal part, and ∇_s^D is the differencing component of the seasonal part. The multiplicative SARMA can be obtained by removing the differencing parts in (9). The optimal parameters, p and q for ARMA and p, q and d for ARIMA are chosen by comparing values of the Akaike Information Criterion (AIC) ([5],[6]), computed for a range of model parameter values.

The ARIMAX model (ARIMA with Exogenous input) was not included in our analysis since it only allows a linear dependence between load and temperature to be modeled, while their actual relationship is quadratic (Section IVA).

C. Random Forest

Random Forest [9] builds 500 de-correlated regression trees and then averages them. Each regression tree is constructed by classifying data based on temperature and hour of the day; leaf nodes contain predicted values. Random Forests are easy to train and tune and hence implemented in many software packages.

D. Double Exponential Smoothing

Simple Exponential Smoothing uses a weighted average of past observations to make forecasts about the future, where the weights decrease exponentially depending on the coefficient of smoothing parameter. Simple exponential smoothing is widely used in many disciplines due to its simplicity, but performs poorly for forecasts of time series with seasonal variations. The Holt-Winters method ([7], [8]) uses three equations, one each for level, trend and seasonality. Variations of the Holt-Winters method exist depending on whether an additive or multiplicative approach is used for modeling the seasonality ([10]).

The work in [1] extended the standard Holt-Winters exponential smoothing to accommodate two seasonal cycles in the electricity demand series which involved introduction of an additional smoothing equation for the extra seasonal part.

The *k* step-ahead double exponential smoothing load forecast, $P_t(k)$, from the origin time *t* is given by the following equations ([1]):

$$\hat{P}_t(k) = (S_t + kT_t) D_{t-s_1+k} W_{t-s_2+k} + \phi^k (P_t - ((S_{t-1} + T_{t-1}) D_{t-s_1} W_{t-s_2})) \quad (11)$$

$$S_t = \alpha (P_t / (D_{t-s_1} W_{t-s_2})) + (1-\alpha)(S_{t-1} + T_{t-1}) \quad (12)$$

$$T_t = \gamma (S_t - S_{t-1}) + (1-\gamma) T_{t-1} \quad (13)$$

$$D_t = \delta (P_t / (S_t W_{t-s_2})) + (1-\delta) D_{t-s_1} \quad (14)$$

$$W_t = \omega (P_t / (S_t D_{t-s_1})) + (1-\omega) W_{t-s_2}, \quad (15)$$

where S_t and T_t are smoothed level and trend and D_t and W_t are the seasonal indices for the intra-day and intra-week seasonal cycles, respectively. The smoothing parameters α, γ, δ and ω are estimated from the training data and ϕ^k in (11) is the adjustment for the first-order autocorrelation. . We set the two seasonal parameters as follows: s_1=10 hours (one working day) and s_2=50 hours (5 working days) for all sites.

V. Results

In this study, we compare the prediction accuracy of the five models described in Section IV, across all six microgrid-scale sites in terms of MAPE for hourly load predictions, MAPE for daily peak load predictions and MAPE for daily peak load under-predictions.

A. Model Training

We performed a series of experiments to determine the minimum amount of training data needed for achieving good prediction accuracy and determined that for hourly day-ahead forecasts, all five models had negligible improvements in prediction accuracy with more than 40 days of training data. Therefore we decided to use a 40-day rolling window for training data. The day-ahead forecast is computed for the remaining 143 days in our trace starting with the 41st day.

B. Discussion

All the models performed well in terms of MAPE for day-ahead hourly load predictions at the sites with low standard deviations relative to their mean (Table 1), i.e., Sites A, C and D (Fig. 5). Moreover, the performance of all models, with the exception of SARMA (which performed worse than the others) is comparable, indicating that model selection for sites with low

variability is not critical, and thus the simplest and least computationally intensive model should be selected.

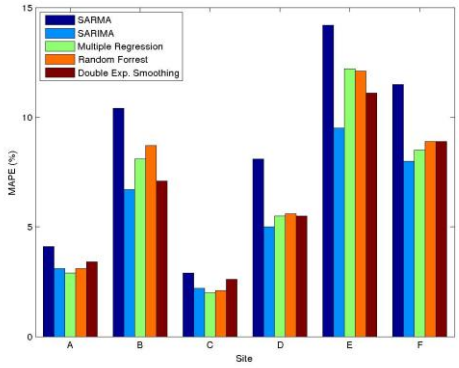

Figure 5 Comparison of MAPE of day-ahead hourly load predictions across all models and sites.

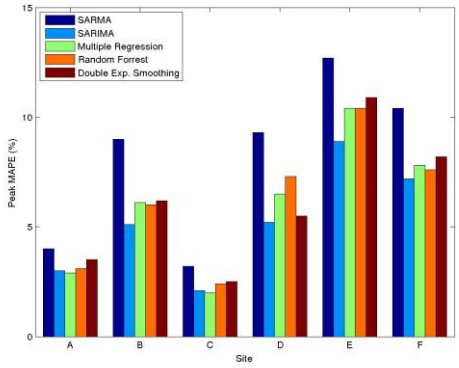

Figure 6 Comparison of MAPE of daily peak load predictions across all models and sites.

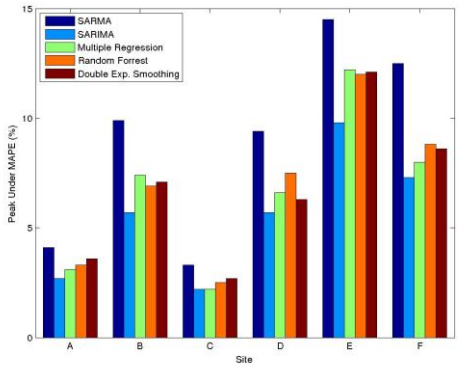

Figure 7 Comparison of MAPE of daily peak under-prediction across all models and sites.

The behavior of the models is quite different for the sites exhibiting high daily and seasonal load variations, i.e., Sites B, E, and F. Prediction errors are higher overall, which is to be expected. In addition, SARIMA performs noticeably better than the other models, and in Sites B and E, all models that do not use weather information (SARIMA and Double Exponential Smoothing) but rely on load level history perform better than those that do. This is not surprising as both of these sites exhibit high variability with respect to temperature (Fig. 2).

To further explore possible differences in performance between the different models, we look at MAPE for daily peak predictions (2). We observe that daily peak errors (Fig. 6) are slightly lower than the overall errors (Fig. 5), and that the difference between temperature-based and load history-based models is no longer present for Sites B and E (as it was with overall error). This is because peak loads are more strongly correlated to temperature than non-peak loads at temperature-sensitive sites, since peak loads typically occur at relatively extreme temperatures. Another difference between the two types of errors can be observed at Site D, where the temperature based models perform noticeably worse in terms of peak error. This is not unexpected as this site displays the weakest relationship between load and temperature across all of our sites and the relationship between peak load and temperature is not as strong (Fig. 2).

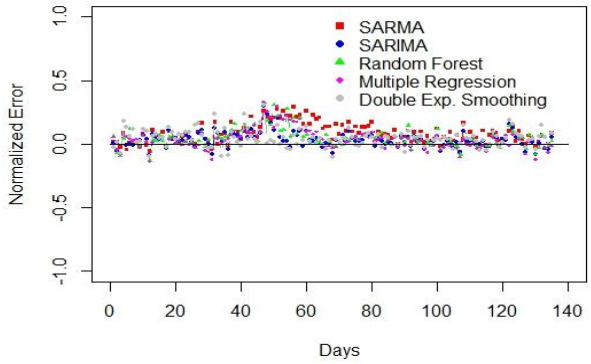

Figure 8 Daily peak error at Site B. Positive errors indicate under-predictions and negative errors indicate over-predictions.

Peak under-prediction is an important aspect of prediction accuracy and is thus a critical model selection criterion. Since unexpected peak demand may lead to the need to use expensive peaker power plants, or potentially to blackouts, peak under-prediction error is even more critical than peak prediction error in evaluating prediction model performance.

We explore the possibility that under-prediction accuracy might help us differentiate between the different models next. The results for peak-under prediction are similar to the ones for peak prediction with respect to the relative performance of the temperature-based and load history-based models (Fig. 7). However, especially in sites with high variability (e.g., Sites B and E), the peak under-prediction errors are noticeably higher than the overall peak prediction errors. Upon closer inspection of the data, we find that under-prediction errors dominate the overall peak prediction accuracy for all sites, with some sites showing an especially strong bias (e.g., Figs. 8-9). This indicates that our models, while optimizing for overall prediction accuracy and peak prediction accuracy, are not aggressive enough in estimating peak usage and tend to underestimate it as a result.

We also observed that under-prediction accuracy has a seasonal component for some sites and models. For example, peak under-prediction errors dominate peak errors at Site E for the

SARIMA model during the cold season but not during the summer and late spring months, when over-prediction is more prevalent (Fig. 9). This could be due to the higher variability in the loads during winter months, for which load variations due to temperature tend to be highest (Fig. 2) thus making it difficult to predict peak loads.

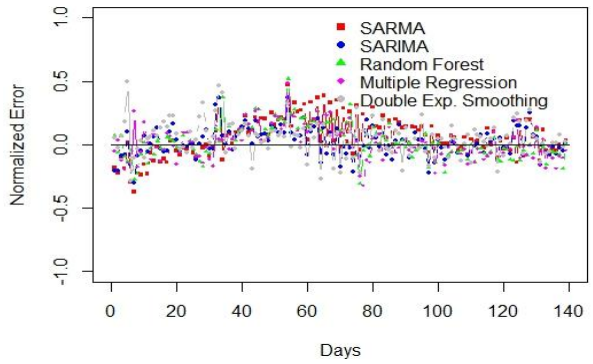

Figure 9 Daily peak error at Site E. Positive errors indicate under-predictions and negative errors indicate over-predictions.

Our observations provide some practical insights for prediction model selection, which we summarize below:

- Given comparable prediction accuracy across a set of models, the simplest model, i.e., the one with the least required inputs (e.g., weather forecast), configuration (e.g., model parameter settings), and computational requirements (e.g., compute resources), should be preferred.
- For sites with low variability, prediction accuracy is high across a variety of different models, and the simplest model should be used (see bullet above).
- Load-history based models should be used for sites and seasons during which:
    - Load does not change in response to temperature.
    - Load changes in response to temperature but there is high variability of load in response to a given temperature.
- Temperature-based models should be used when:
    - Prediction is performed for sites and seasons during which load changes in response to temperature, and the variability of loads in response to temperature is low.
    - Predicting peak-time usage for sites and seasons during which load changes in response to temperature (e.g., a temperature-based model should be used between 3 and 5pm).

VI. Conclusions and Future Work

We used five models to perform day-ahead hourly and daily peak forecasts of electricity loads at six microgrid sites, and derived prediction model selection criteria for load prediction based on our findings.

We also find that SARIMA performs slightly better than the other four models and that models that do not utilize weather-related information perform comparably or better than the models that do use this kind of information. This simplifies the modeling problem and eliminates the reliance on weather forecast information, which is sometimes difficult and costly to obtain, and may also be inaccurate.

Peak under-prediction is an important metric for evaluating model performance as it plays an important role in reducing electricity generation costs. We find that prediction results across all sites and models are dominated by under-prediction errors. This is an important result, which we plan on exploring further in future work.

References

[1] J. W. Taylor, L. M. de Menezes and P. E. McShary, "A Comparison of univariate methods for forecasting electricity demand up to a day ahead", International Journal of Forecasting, vol. 22, pp. 1-16, 2006.

[2] J. W. Taylor, " Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing", Journal of Operational Research Society, 54, 799-805, 2003.

[3] A. Piras and B. Buchenel A tutorial on short term load forecasting, Engineering Intelligent Systems, 1, pp. 41-47, 1999.

[4] H. S. Hippert, C. E. Pedreira and R. Souza, "Neural networks for short term load forecasting: A review and evaluation, " IEEE Transactions on Power Systems,. vol. 16, pp. 44-55, 2001.

[5] M. De Felice and X. Yao, "Neural Networks Ensembles for Short-Term Load Forecasting," in *IEEE Symposium Series in Computational Intelligence 2011 (SSCI 2011)*, 2011.

[6] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, "Time series analysis: Forecasting and control, 3rd edition, EngleWood Cliffs, New Jersey:Prentice Hall, 1994.

[7] P. R. Winters, " Forecasting sales by exponentially weighted moving averages. *Management Science* **6**: 324-342, 1960.

[8] C. Holt, "Forecasting trends and seasonals by exponentially weighted averages, O. W. R. Memorandum, no.52, Carnegie Institute of Technology, 1957.

[9] L. Breiman, "Random Forests", Machine Learning, 45,(1), pp. 5-32, 2001.

[10] R. J. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Synder, "Forecasting with Exponential Smoothing: The state space approach. Springer-Verlag: Berlin Heidelberg, 2008.

[11] Q. Tan, et. al, "Energy Consumption Forecasting Using Support Vector Machines for Beijing", International Conference on E-Product E-Service and E-Entertainment (ICEEE), November 2010.

[12] S. Yang, and Y. Wang, "Applying Support Vector Machine Method to Forecast Electricity Consumption," International Conference on Computational Intelligence and Security, November 2006.

[13] E. E. El-Attar, J. Y. Goulermas, and Q. H.Wu, "Forecasting electric daily peak load based on local prediction," in Proc. IEEE Power Eng. Soc. Gen. Meeting (PESGM 2009), Calgary, Canada, July 2009.

[14] F.A. Razak, A.H. Hashim, I.Z. Abidin and M. Shitan, "Malaysian Peak Daily Load Forecasting ", proceeding of Scored 2009.

[15] J. Xiao, L. Hou, F. Luo, and C. Chen, "Load forecasting and its calibration method considering the influence of temperature," Electric Utility Deregulation and Restructuring and Power Technologies, 2008.

[16] M. Ghomi, M. Goodarzi, and M. Goodarzi, "Peak Load Forecasting of Electric Utilities for West Province of IRAN by Using Neural Network without Weather Information, " In Proceedings of the 2010 12th International Conference on Computer Modelling and Simulation (UKSIM '10).

[17] S. Jafarzadeh, S. Fadali, C.Y. Evrenosoglu, H. Livani, "Hour-ahead wind power prediction for power systems using hidden Markov models and Viterbi algorithm," IEEE PES General Meeting, PES 2010.

[18] W. N. Venables and B. D. Ripley, "Modern Applied Statistics with S-PLUS", 3rd edition, Springer-Verlag, New York, 1999.