Increasing Grid Flexibility Through Improved Electricity Demand Prediction in Nicaragua

Stephen Suffian
Electrical and Computer Engineering
Villanova University
Villanova, USA
ssuffian@villanova.edu

Diego Ponce de Leon Barido
The Energy and Resources Group
University of California
Berkeley, USA
diego.leon@berkeley.edu

Dr. Madhura Ingalhalikar,
Dr. Pritpal Singh
Electrical and Computer Engineering
Villanova University
Villanova, USA
pritpal.singh@villanova.edu

Abstract— **Renewable energy provides an increasingly significant contribution to power production around the globe. The variable and uncertain nature of certain renewable energy sources, however, requires increased grid flexibility to reliably match electricity supply with demand. On average, wind energy accounts for 20% of Nicaragua's total generation, and can produce up to 50% within a given hour. Under the renewable energy regime fuel-oil generators are the main source of grid flexibility. Information-driven flexibility, such as improved demand prediction, can be used to reduce the need of fuel-oil based flexibility without affecting reliability. This paper evaluates and compares the use of multiple linear regression (MLR) and support vector regression (SVR) in their ability to minimize electricity demand forecast error in Nicaragua. We find SVR reduces the mean absolute percent error of prediction to 3.8%, compared with MLR (7.7%). SVR further performs a prediction with 21% less error than the current prediction mechanism employed by the utility. Finally, we discuss how improved prediction algorithms can be used to reduce Nicaragua's dependency on fuel-oil for flexibility, while also reducing costs for the utility.**

Keywords—grid flexibility; STLF; SVR, MLR, Nicaragua

I. INTRODUCTION

In 2013 approximately 22% of global electricity was produced using renewable energy, representing a 5% increase from the prior year [1]. The International Energy Agency suggests that renewable energy will reach about one-third of total generation by 2040 [2]. Wind energy alone contributed almost 4% of total global electricity generation in 2013 [3], with a 12% increase in global installed capacity from 2012 [4].

While an increase of renewable energy production in the global energy matrix provides multiple and immediate co-benefits (i.e., carbon abatement, energy security, economic efficiency), it also requires more advanced grid management in order to maintain an acceptable level of electricity reliability. Specifically, grids with high penetrations of wind and solar generation are required to perform under the same reliability and operational constraints despite their inherent uncertainty and variability. Greater flexibility is addressed through a variety of measures depending upon a grid's availability of different energy sources [5]. Certain types of plants can be used for flexibility by being quickly dispatched or ramped, such as hydroelectricity, combined-cycle gas turbines, and fuel oil plants. Regional power interconnections can also be used to provide cost-effective integration of renewable energy [6], and demand response programs can help adjust the supply-demand imbalance by enabling demand-side resources [7].

It is important to note, however, that the feasibility of the above solutions is dependent on a country's geography and level of development. For example, in 2014 Nicaragua produced up to 50% of total generation from wind on an hourly basis, and generated 20% wind on average per year [8]. Nicaragua is the 2nd poorest country in the Western Hemisphere [9], yet it is expected to reach 90% renewable generation by 2020 [10], and could achieve this goal using a diverse set of renewable resources and integration strategies [11]. Despite Nicaragua's great renewable energy potential, grid flexibility options are limited. Fuel oil and hydropower constitute the most important sources of flexibility (2014: 47% and 9% of total generation respectively), but both resources prove problematic in the long run due to hydropower's vulnerability to hydroclimatological variability [11] [12], and fuel oil's vulnerability to price fluctuations, localized air pollution, and its contribution to climate change. A dependence on fuel-oil for grid flexibility could prove both expensive and detrimental to Nicaragua achieving its mutual goals of grid cost-effectiveness and energy independence.

This paper develops prediction models for Nicaraguan electricity demand using openly available data. The goal of the study is to evaluate the potential for improving demand prediction both as a method for reducing the utility's demand prediction error and as a means for creating a more time-resolved ancillary services market. The background provided in Section II discusses algorithms that have been used in the literature for electricity demand prediction. Section III describes our data collection and preparation methodology followed by a discussion (Section IV) on the methodology used for hyper-parameter optimization, cross-validation, and comparison between algorithms. Sections V and VI present our results and a discussion regarding the importance of improving demand prediction accuracy.

II. BACKGROUND

With increased penetration of uncertain and variable energy into the grid, cost-effective load-matching has become a central concern of global power industries. Short-term load forecasts (STLFs) are vital inputs to the planning and operational process and have been used for decades to solve unit commitment and economic dispatch problems [13]. Day-ahead and spot markets, demand response programs, development and adjustment of bidding strategies, and reliability concerns such as avoiding overloading and the occurrence of blackouts all rely heavily on accurate load forecasting [14].

The literature identifies two types of forecasting methods: 1) statistical learning methodologies including ARIMA [15], Kalman filtering [16], and multiple linear regressions [17], and 2) artificial intelligence-related models such as fuzzy expert systems [18], support vector regression (SVR), artificial neural networks (ANN) [19], and self-organizing networks [20]. Neural networks are a popular approach to demand prediction, but their hidden layer causes difficulties when interpreting the reasoning behind results. More interpretable models such as SVR are capable of exceeding the accuracy of neural networks in STLF [21]. Further, SVR applies a structural risk minimization principle thus minimizing the upper bound of generalization error through regularization, rather than relying purely on minimizing training error through empirical risk minimization, which is used by ANNs. SVRs therefore reduce overfitting by balancing model complexity with training success. However, this balance and thus the accuracy of an SVR are dictated by hyper-parameters such as the choice of kernel and the values of C, epsilon, and gamma. The selection of these hyper-parameters can be done by searching the whole space, or optimized through methods such as simulated annealing (SA), genetic algorithms (GA), and particle swarm optimization (PSO) [22].

III. DATA COLLECTION AND PREPARATION

Hourly demand and generation data was obtained from Nicaragua's National Dispatch Center [23] for January 1st 2012 to June 23rd 2015, and weather data (three hour intervals) was obtained for the same time period [24]. The weather data was interpolated to hourly values in order to maintain a single index, with the assumption that weather (rainfall, temperature, etc.) changes linearly within each 3-hour timeframe.

Hourly demand profiles depicted bimodal distributions represented by peak and off-peak times. The presence of this underlying structure in our data led to the creation of sub-groups of data within the larger dataset in order to find two unimodal distributions. K-means clustering was used to develop the sub-groups and construct more targeted models. Clustering was iteratively applied to three temporal variables (months, days, and hours). Average demand over each time interval was used for clustering, and K was varied across all acceptable values. For example, when clustering by month the average demand for each month was used as the input to the K-means algorithm, K was varied from 2-11, and the K with the best silhouette score was chosen. A silhouette score was used as the evaluation metric, which compared the average distance of each point *i* within its own cluster to its average distance from the points in its next nearest cluster.

Clustering was first applied to the 'months of the year' in order to create sub-groups based on seasonality, then applied to 'day of week' and finally 'hour of day.' Based on silhouette scores, K=2 was chosen in each case. The sub-groups that were found through clustering are listed in Table 1, and Figure 1 displays the division of the bimodal distribution into two unimodal distributions based on the sub-groups.

TABLE I. NICARAGUAN NATIONAL ELECTRICITY DEMAND CLUSTERS

Peak Sub-groups			
March-May		***June-Feb***	
Mon-Fri	***Sa-Su***	***Mon-Fri***	***Sa-Su***
8AM-10PM	7PM-9PM	8AM-9PM	6PM-9PM
Off-Peak Sub-groups			
March-May		***June-Feb***	
Mon-Fri	***Sa-Su***	***Mon-Fri***	***Sa-Su***
11PM-7AM	10PM-6AM	10PM-7AM	10PM-5PM

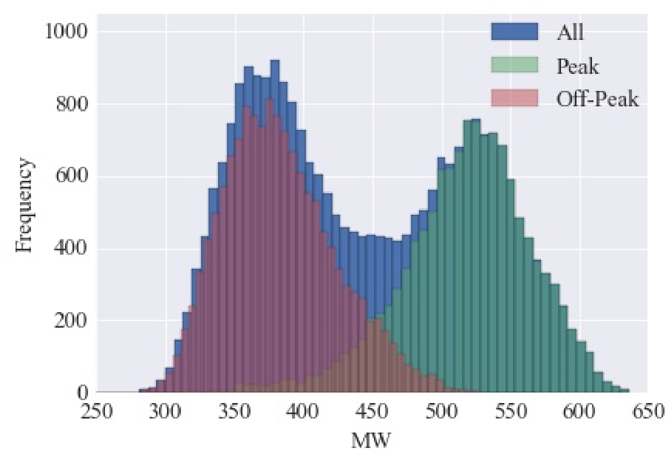

Figure 1. Histogram of Total, Peak, and Off-Peak Nicaragua Hourly National Demand (2012-2015)

IV. METHODOLOGY

We compared the accuracy of MLR and SVR against that of the Nicaraguan grid operator for day-ahead posted net demand predictions. It is assumed that current predictions are posted at 6 PM for the next day from midnight to 11 PM, and therefore contain 6 through 29 hour-ahead predictions. The data used for this comparison spanned from January 1st 2014 to June 23rd 2015. The data were then split into a training and validation subset (January 1st 2014 to May 31st 2015) and a testing subset (June 1st to June 23rd 2015). The testing subset was chosen because it was the only interval in which utility predictions were available. The data was divided into the 7 sub-groups according to the k-means analysis described above.

Both MLR and SVR used the following independent variables: time-lagged (t-1,2,3…24), weather (temperature, pressure, humidity, etc.) and demand, as well as dummy variables including hour (0-23), day of month (0-30), day of week (0-6), hour in week (0-167), year (2014-2015), and a binary variable for holiday (0,1). The SVR also used additional independent variables for future (t+1,2,3…24) weather. For the test data, future weather variables were sampled from hourly distributions of weather variables in the training and validation subset to avoid using perfectly forecasted weather.

MLR used a least squares regression to estimate coefficients based on all available data up to the hour of prediction. Optimal model fit was determined through a step-wise procedure where variables were iteratively added. For SVR, 5-folds cross validation was used with the training and validation subset to find the optimal hyper-parameters C, epsilon, and gamma. A radial basis function was used for the kernel as it is best suited for smooth data, which is an appropriate assumption for hourly electricity demand.

A. Comparison to Current Prediction

In both the MLR and SVR cases, n-hour (n=1,2,3,…,29) ahead models were developed and used together in order to calculate an entire day-ahead prediction. After the regression or SVR model at n hours was constructed, the output was used as an additional independent variable for the n+l regression or SVR model. For each day in the test set, the MLR coefficients and SVR model were updated using the previous days demand data. We used the mean absolute percent error (MAPE) as the evaluation metric for prediction accuracy. Within the timeframe of the test set, the utility achieved a MAPE of 4.8%.

V. RESULTS

Across the 7 sub-groups, MLR had the best fit for March-May weekdays at 11PM-7AM, and the worst fit for June-February weekdays at 10PM-7AM. The regression received a MAPE of 7.7% with the June 2015 test set using the previously described method for day-ahead prediction.

The optimal hyper-parameters for SVR were C=1000, epsilon=.0001, gamma = 10e-6. The SVR model that performed best during cross validation achieved a MAPE of 2.3% for 24-hour-ahead prediction. This model had the lowest MAPE (1.7%) for March-May weekdays from 11PM-7AM, and the highest MAPE (2.7%) for June-February weekdays from 10PM-7AM. SVR achieved a MAPE of 3.8% with the test set for day-ahead prediction.

SVR outperformed MLR by 38% and the current utility prediction method by 21%. SVR over-predicted electricity demand an average of 223 MWh per day and under-predicted demand by an average of 227 MWh per day. MLR over-predicted electricity demand by an average of 413 MWh per day and under-predicted demand an average by of 487 MWh per day. The utility over-predicted demand an average of 446 MWh per day and under-predicted demand an average of 177 MWh per day. The results of MLR, SVR, and current prediction for a subset of the test data are shown below during one week (June 5^{th} to June 12^{th}) in Figure 2 and for an average weekday in Figure 3.

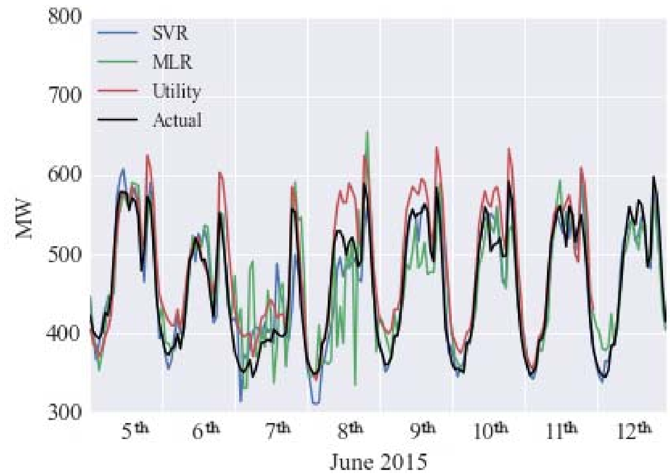

Figure 2. One week of Electricity Demand Prediction (June 5^{th} to June 12^{th})

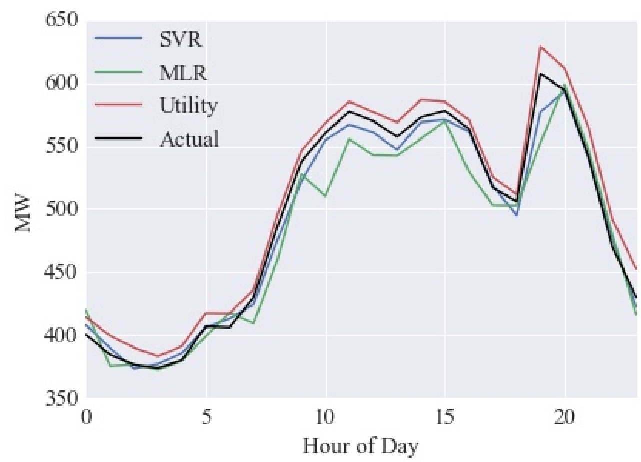

Figure 3. Average Weekday Electricity Demand Prediction

VI. DISCUSSION

The SVR algorithm proposed in this article offers a reduced day-ahead prediction error using openly available data. Our results suggest that more accurate day-ahead electricity demand predictions could lead to a reduced dependence on fuel-oil (the primary source of current flexibility). On average, SVR over-predicted 50% less MWh than the utility per day and only under-predicted by an additional 28%. Over-prediction leads utilities into purchasing more energy than they actually need in the day-ahead market. The average cost of energy in June was approximately \$105/MWh, meaning that the utility could have saved an average of \$23,000 per day by using a statistical method with similar performance to the authors' SVR. Further, flexible power plants that are used to correct for this error must cycle their equipment, which can add thermal and pressure stresses that decrease plant life expectancy and increase operation and maintenance costs [25].

Day-ahead demand prediction could be further improved by incorporating anomaly detection and more accurate methods for weather forecasting. Future research will incorporate wind generation forecasting and seek to conduct a similar analysis to this one for national net electricity demand prediction, as well as quantifying the environmental and economic impacts that could be mitigated through improved prediction accuracy.

VII. Conclusion

Using open-access electricity demand and weather data and consistent daily model updates, day-ahead electricity demand for June 1st to June 23rd 2015 was predicted with a MAPE of 3.8% with SVR. SVR resulted in higher prediction accuracy than both the current utility prediction method (4.8% MAPE) and MLR (7.7% MAPE). Reduced prediction error saves the utility money when purchasing energy in day-ahead markets as well as improves plant lifetimes with less frequent power cycling. Overall, information-driven grid flexibility is an important part of Nicaragua's pursuit for greater renewable energy penetration and reduced dependence on fuel-oil.

VIII. References

[1] IEA, "Medium-Term Renewable Energy Market Report," International Energy Agency, Paris, 2014.

[2] IEA, "World Energy Outlook Executive Summary," International Energy Agency, Paris, 2014.

[3] WWEA, "Half-Year Report," World Wind Energy Association, Bonn, Germany, 2014.

[4] IEA Wind, "2013 Annual Report," International Energy Agency, Boulder, Colorado, 2014.

[5] J. Cochran, M. Miller, O. Zinaman, M. Milligan, D. Arent, B. Palmintier, M. O' Malley, S. Mueller, E. Lannoye and A. Tuohy, "Flexibility in the 21st Century," National Renewable Energy Lab, Golden, Colorado, 2014.

[6] M. Huber, D. Dimkova and T. Hamacher, "Integration of wind and solar power in Europe: Assessment of flexibility requirements," *Energy,* vol. 69, pp. 236-246, 2014.

[7] K. Kornbluth, J. Cook and P. Erickson, "A More Renewable-Friendly Electrical Grid: Thermal Storage Refrigeration for Demand Response in California and Denmark," *Smart Grid and Renewable Energy,* vol. 3, no. 4, pp. 1-8, 2012.

[8] D. P. d. L. Barido, "Beyond Oil: the transition to a low-carbon grid in Nicaragua," 2014.

[9] IMF, "World Economic Outlook," International Monetary Fund, 2014.

[10] Estrategia y Negocios, "Nicaragua quiere tener 90% de energyas renovables en 2020," 29 4 2015.

[11] D. P. d. L. Barido, J. Johnston, M. V. Moncada, D. Callaway and D. M. Kammen, "Evidence and Future Scenarios of a Low-Carbon Energy Transition in Central America: A Case Study in Nicaragua," *Environmental Research Letters,* vol. 10, no. 10, p. 104002, 2015.

[12] S. Gourdji, P. Läderach, A. M. Valle, C. Z. Martinez and D. B. Lobell, "Historical climate trends, deforestation, and maize and bean yields in Nicaragua," *Agricultural Forest and Meteorology,* vol. 200, pp. 270-281, 2015.

[13] M. Abu-El-Maged and N. Sinha, "Short-term load demand modeling and forecasting: a review," *IEEE Transactions for System, Man and Cybernetics,* vol. 12, no. 3, pp. 370-382, 1982.

[14] E. Kyriakides and M. Polycarpou, "Short term electric load forecasting: A tutorial," *Trends in Neural Computation,* vol. 35, pp. 391-418, 2007.

[15] C.-M. Leea and C.-N. Ko, "Short-term load forecasting using lifting scheme and ARIMA models," *Expert Systems with Applications,* vol. 38, no. 5, pp. 5902-5911, 2011.

[16] H. Al-Hamadi and S. Soliman, "Short-term electric load forecasting based on Kalman filtering algorithm with moving window weather and load model," *Electric Power Systems Research,* vol. 68, no. 1, pp. 47-59, 2004.

[17] N. Amral, C. Ozveren and D. King, "Short term load forecasting using Multiple Linear Regression," in *42nd International Universities Power Engineering Conference*, Brighton, UK, 2007.

[18] A. Khosravi, S. Nahavandi, D. Creighton and D. Srinivasan, "Interval Type-2 Fuzzy Logic Systems for Load Forecasting: A Comparative Study," *IEEE Transactions on Power Systems,* vol. 27, no. 3, pp. 1274-1282, 2012.

[19] S. S. Reddy and J. A. Momoh, "Short Term Electrical Load Forecasting Using Back Propogation Neural Networks," in *North American Power Symposium (NAPS)*, Pullman, Washington, 2014.

[20] A. F. Huseynov, A. Hashimov and N. Yusifbeyli, "Electrical System Load Forecasting with Polynomial Neural Networks," in *Modern Electric Power Systems*, Wroclaw, Poland, 2010.

[21] B. Wang, B. X. Leng, X. H. Zhang, C. H. Shan and Z. Cong, "Application Profiles of Support Vector Machine in Short-term Load Forecasting," *Proceedings of the CSU-EPSA,* vol. 23, no. 4, pp. 115-121, 2011.

[22] G. Li, C. T. Cheng, J. Y. Lin and Y. Zeng, "Short-term load forecasting using support vector machine with SCE-UA algorithm," in *Third International Conference on Natural Computation*, Haikou, China, 2007.

[23] CNDC, "Centro Nacional de Despacho de Carga," ENATREL, 2013. [Online]. Available: www.cndc.org.ni. [Accessed 23 6 2015].

[24] World Weather Online, "World Weather Online Historical Weather API," World Weather Online, 2015. [Online]. Available: developers.worldweatheronline.com. [Accessed 23 6 2015].

[25] N. Kuma, P. Besuner, S. Lefton and D. Agan, "Power Plant Cycling Costs," Intertek APTECH, Sunnyvale, California, 2012.