Long Term Forecasting using Machine Learning Methods

Hossein Sangrody

Long Term Forecasting using Machine Learning Methods

Hossein Sangrody[1], *Student Member, IEEE,* Ning Zhou[1], *Senior Member, IEEE*, Salih Tutun[2], Benyamin Khorramdel[3], *Student Member, IEEE*, Mahdi Motalleb[4], *Student Member, IEEE*, Morteza Sarailoo[1]

[1] Electrical and Computer Engineering Department, State University of New York at Binghamton, NY, USA {habdoll1, ningzhou, msarail1}@binghamton.edu

[2] Industrial Engineering Department, State University of New York at Binghamton, NY, USA, stutun1@binghamton.edu

[3] Department of Electrical and Computer Engineering, University of Saskatchewan, Canada, bek067@mail.usask.ca

[4]Department of Mechanical Engineering, Massachusetts Institute of Technology (MIT), MA, USA, motalleb@mit.edu

***Abstract*— A robust model for power system load forecasting covering different horizons of time from short-term to long-term is an indispensable tool to have a better management of the system. However, as the horizon of time in load forecasting increases, it will be more challenging to have an accurate forecast. Machine learning methods have got more attention as efficient methods in dealing with the stochastic load pattern and resulting in accurate forecasting. In this study, the problem of long-term load forecasting for the case study of New England Network is studied using several commonly used machine learning methods such as feedforward artificial neural network, support vector machine, recurrent neural network, generalized regression neural network, k-nearest neighbors, and Gaussian Process Regression. The results of these methods are compared with mean absolute percentage error (MAPE).**

I. Introduction

Distributed energy resources (DERs) are penetrating in power system more and more and a lot of studies have focused on implementation of them [1-3]. Although DERs have benefits for a power system, their indispatchability, intermittency, and uncertainty have presented unprecedented challenges to power grid operation and planning [4]. At this condition, load forecasting (LF) plays a critical role in the operation and planning of a power system. Depending on the purposes of LF, the lead times of LF can vary from seconds to years. Very short-term load forecasting (VSTLF) [5] and short-term load forecasting (STLF) [6] usually have lead times of seconds to weeks and are often used for control and operation purposes. In contrast, medium-term load forecasting (MTLF) [7] and long-term load forecasting (LTLF) [8] have lead times of month(s), years, even decades and are often used for scheduling and planning purposes [9].

The driving inputs of a forecasting model are important factors to yield an efficient forecast model. The inputs of forecasting model depend on the purpose of forecasting and its term (very short to long term). In [10] historical data of price are applied to predict the hourly prices in the California Independent System Operator (CAISO)'s day-ahead electricity market. Along with historical data of load in load forecasting model, temperature is also one of the most common input variables. However, since load pattern is a nonlinear function of temperature, heating degree days (HDD) and cooling degree days (CDD) are applied as weather indicators in load forecasting modeling [11].

The methodologies applied in LF can be classified in three main categories of statistical analysis, machine learning, and hybrid methods. Among them, machine learning methods have got more attentions in recent years [12]. Despite of the benefits of hybrid methods, their parameters need to be adjusted well to achieve accurate forecasting [7]. Among all machine learning methods, feedforward artificial neural network (ANN), support vector regression (SVR), recurrent neural network (RNN), k-nearest neighbors (KNN), generalized regression neural network, and Gaussian Process Regression (GPR) are the most common methods in load forecasting. In this study, the LTLF for monthly load forecasting in ISO New England Network is applied using the aforementioned machine learning methods and their results are quantified and compared by mean absolute percentage error (MAPE).

The rest of the paper is organized as follows. Section II elaborates the forecasting inputs, output, the methodology of applying them, and forecasting models. Section III represents the simulation results for the case study using different methods. The conclusion is drawn in Section V.

II. Forecast Model

Obtaining accurate forecast results depend on various factors. Generally, the horizon of LF, certainty of the inputs, and efficiency of forecasting models are the major influential factors on the accuracy of forecasting results. As the horizon of load forecast increases, having accurate prediction with highly time resolution will be more challenging. The main reason is the high uncertainty in the inputs of forecasting model in a long term forecasting. Accordingly, in VSTL and STLF, since the horizon of forecasting spans only in the time frame of seconds to weeks, the weather indicators are more accurate inputs while for the MTLF and LTLF which cover the lead time of months to years, the prediction relying on inaccurate weather indicators.

A. *Input and Output of the Models*

As mentioned the most common variables in LF are weather indicators. However, the relationship of weather indicators and energy usage is not linear. Thus, two other weather indicators, HDD and CDD are applied in forecasting model. HDD is a criterion showing whether a unit requires to be heated and it is obtained by the number of degrees which average temperature of the day is below 65° F. On the other hand, CDD showing that a unit needs to be colder is the number of degrees that the average temperature of a day is above 65° F. Such variables yield linear relationships between energy and weather indicators. In this study, the forecasting model is applied to predict monthly energy for the New England Network and the total HDDs and CDDs for each month are applied as inputs of the model.

To improve the accuracy of forecasting model, the historical record of energy is also used as input using moving average method. As shown in (1), to have prediction of energy in a target month, an 11-month average energy corresponding to the target month is also fed to the model along with weather indicators. Here T is historical data time which is 11 for this case and t represents target variable.

$$x_t = \frac{1}{T}\sum_{i=1}^{T} x_{t-i} \tag{1}$$

Fig. 1 illustrates the overall structure of inputs and outputs in the forecasting model. In this figure, for each month of forecasted energy as target, the model uses corresponding total HDD, CDD, and average energy of 11 corresponding months of the past years.

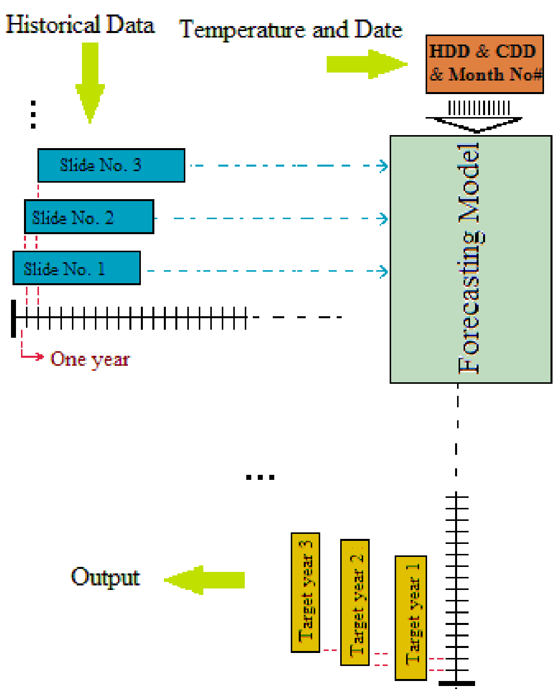

Fig. 1. Forecasting structure

B. Forecasting Models

In this sections, a brief review on the commonly used machine learning methods in forecasting are discussed.

1) Feed forward Artificial Neural Network Model

The ANN method provides an efficient way to address modelling of a complex nonlinear system. In other words, in forecasting model using ANN, there is no need for a forecaster to have a clear understanding of the complex relationship between inputs and outputs.

Fig. 2 depicts a typical neural network normally consists of three layers of input, hidden and output layers. Each layer consists of several neurons which are connected to other layer's neuron(s) with weighted connections. As shown in this figure, the arrowheads of the connections indicate that all data propagate in the direction from inputs to the output. Such a structure is entitled feedforward ANN model. The number of neurons in the input and output layers are the number of inputs and output, respectively. The hidden layer located between the input layer and output layer has an arbitrary number of neurons which are defined by the forecaster. For many problems in forecasting, one or two hidden layers often give good results.

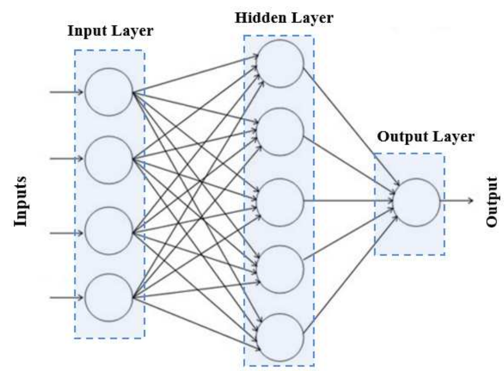

Fig. 2 Typical structure of feedforward ANN model

The training algorithm for this study is the Levenberg-Marquardt algorithm which takes more training time but gives better results.

2) Support Vector Machine

Support vector regression is the version of the support vector machine method that are applying for forecasting model. Assuming $\boldsymbol{X}$ as the input variable vector and $\boldsymbol{Y}$ as the output variables, the SVR solution can be obtained by minimizing the sum of training error $\sum_{i=1}^{N}(\xi_i + \xi_i^*)$ and regularization term $\frac{1}{2}\|w\|^2$ in (2) subjected to constraints (3).

$$Min \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{2}$$

$$\begin{cases} y_i - (w^T\phi(x_i) + b) \le \varepsilon + \xi_i \\ (w^T\phi(x_i) + b) - y_i \le \varepsilon + \xi_i^* \end{cases} \quad \xi_i, \xi_i^* \ge 0, \quad i = 1, \dots, N \tag{3}$$

Here, N is the total number of observation sets, ξ_i and ξ_i^* are upper and lower training errors associated to ε (margin of tolerance) and ϕ is the kernel function, which transforms x_i to higher dimensional space.

3) Recurrent neural network

The structure of Recurrent neural network(RNN) includes an input layer, hidden layer, a context layer, and output layer [13, 14]. (4) and (5) show the calculations in RNN model for training data points $x_i$and a target values y_i.

$$h_i = \delta_h(W_h x_i + U_h h_{i-1} + b_h) \tag{4}$$

$$y_i = \delta_y(W_y h_i + b_y) \tag{5}$$

Where $h_i$denotes the hidden layer vector, b is bias vector, W and U are weigh matrices, and δ represents the activation function.

4) Generalized Regression Neural Network

Generalized regression neural network (GRNN) is nonparametric model whose structure includes input layer, output layer, radial basis layer, and a special linear layer. The GRNN model derives prediction of a target value corresponding to a given data point X by calculating the weighted average of target values in the training data points in the vicinity of the data point X [15]. As shown in (6), a target point ($\hat{y}$) corresponding to the data point x is predicted by the average of the target points and assigning weights using a kernel function considering the distance of predictors in training set to the data point X. In this case, the kernel function (K) is a standard Gaussian kernel.

$$\hat{y} = \sum_{i=1}^{N} w_m \, y_i \tag{6}$$

$$w_m = \frac{K(\frac{\|x - x_i\|}{h})}{\sum_{i=1}^{N} K(\frac{\|x - x_i\|}{h})} \tag{7}$$

5) K nearest neighbors Regression

K nearest neighbors method (KNN) is a nonparametric method applied for both regression and classification. In this method, the prediction is yielded based on the target values of the K nearest neighbors in the given point. In other words, given a data point, the K nearest data points in the training data set are selected and the average of their target values are considered as predicted target value as follows.

$$\hat{y} = \frac{1}{K}\sum_{i=1}^{N} y_i \tag{8}$$

6) Gaussian Process Regression

Gaussian Process Regression (GPR) is a nonparametric method which is modeled based on considering a priori distribution (multivariate joint Gaussian distribution) for any subset of target values of different data points in training set [16]. In GPR, if two input vectors are close, the correlation between their function value is higher. The posterior distribution for a predicted value is derived using the prior distribution. The covariance two inputs x_i and x_j in training inputs can be model as follows [15].

$$K(x_i, x_j) = \delta_f^2 e^{-\frac{\|x_i - x_j\|^2}{2\gamma^2}} \tag{9}$$

So, the vector function f(consisting function values f_i for training point i) follows multivariate Gaussian density function as follows.

$$f \sim N_f(0, K(X, X)) \tag{10}$$

Where, N_f denotes multivariate normal density function and $K(X, X)$ denotes covariance matrix whose $(i, j)^{th}$ element is $K(x_i, x_j)$. Considering the target values y as (11), the predicted posterior function $\hat{f}*$ for an input value x_* is derived as (12).

$$y = f + \epsilon \tag{11}$$

$$\hat{f} *= E(f_* | X, y, x_*) = K(x_*, X)[K(X, X) + \delta_n^2 \, I]^{-1} y \tag{12}$$

Where ϵ denotes the vector of noise with standard deviation of δ_n^2.

C. Accuracy metric

To quantify the results of forecasting models, there are various statics metrics. Some common metrics are mean absolute percentage error (MAPE) [17] defined by (13), mean absolute error (MAE) [18] defined by (14), mean squared error (MSE) [19] defined by (15), and root-mean-square error (RMSE) [20] defined by (16).

$$\text{MAPE} = \frac{1}{N}\sum_{i=0}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{13}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=0}^{N}|y_i - \hat{y}_i \;| \tag{14}$$

$$\text{MSE} = \frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i \;)^2 \tag{15}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i \;)^2} \tag{16}$$

Where, N is the total number of time instants, y_i is the target value at instance of i and $\hat{y}_i$ is the corresponding forecasted target.

Between aforementioned error metrics, MAPE is the most common metrics. As shown in (13) the MAPE gives relative errors in percentage, which does not depend on the scale of forecasted variables. Therefore, the MAPE has been widely used to compare forecasting accuracy under different scenarios. Accordingly, in this study, the MAPE is used as the criterion to measure the proficiency of the models.

III. Simulation Results

In this section, the long-term load forecasting models are applied for the case study of the New England Network. The load profile of the network during 2000 to March 2016 is shown in Fig. 3.

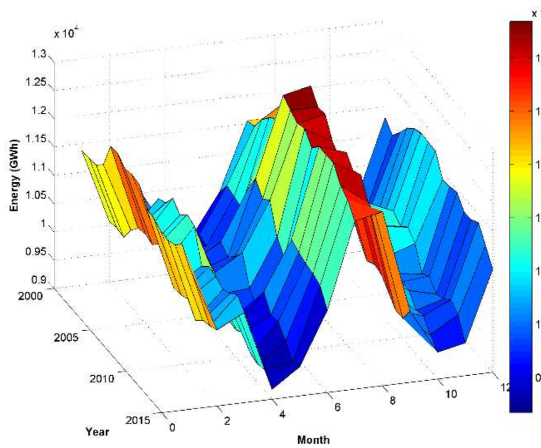

Fig. 3. New England Network Energy during 2000 to March 2016

As mentioned before, the input variables of each load forecast model depend on the case study and horizon of forecasting. In long term forecasting, one of the likely inputs is population growing rate which has a positive exponential rate. However, Fig. 3 illustrates that the load did not grow exponentially during long term of 17 years which dismisses the influence of growing population as a input for this case.

In Fig. 4, the energy usage is represented in month and year axes. As shown the load in July and August peaks dramatically. Such a stochastic load behavior will make it more difficult for the forecasting model to have accurate prediction if the model is supposed to rely on only weather variables. Note that although residential load pattern depends on whether indicators like temperature, industrial and commercial loads do not correlate strongly on weather temperature. However, in this case since the dramatic changes of load in the aforementioned months as well as January are repeated at the same time during all years, using historical value of load data can be a good solution to deal with such a problem. In addition, considering the number of month as an input is one of the inputs which improves the results of prediction. Accordingly, each target value in a load forecasting model corresponds to weather indicators, historical average, and dummy variable of the month number.

The monthly energy data are divided in three categories of historical data (which is applied in forecasting model as one of the inputs), training set, and cross validation data set. As shown in Fig. 5, the green color are historical data from 2000 to 2011 and sixty percent of the rest of data which is the monthly energy during January 2011 to March 2016 are used for training of (blue color) the forecast model and 40 percent are applied for validation (red color).

The data before 2011 are implemented as one of inputs corresponding to each target value. In other words, each month of target gets benefit from the corresponding monthly historical data of during 11 years ago and the historical data moves ahead as the target value of monthly energy moves ahead in training and validation processes.

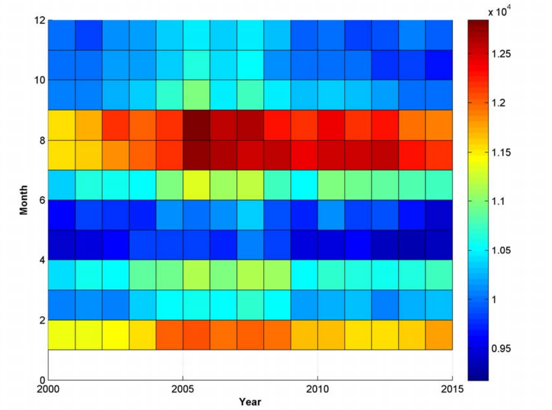

Fig. 4. 2D representation of energy usage for New England Network

As an example, to forecast energy in January 2011, as the first target, the inputs are total HDD and CDD of January 2011, dummy variable for the month number which is one for January and zero for other months, and historical monthly energy of the January during 2000 to 2010. For the next slide, for forecasting the energy in February 2011, the energy usage in February 2000 to 2010, the month variable (which is 1 for February), as well as HDD and CDD of February 2011 are applied in the model.

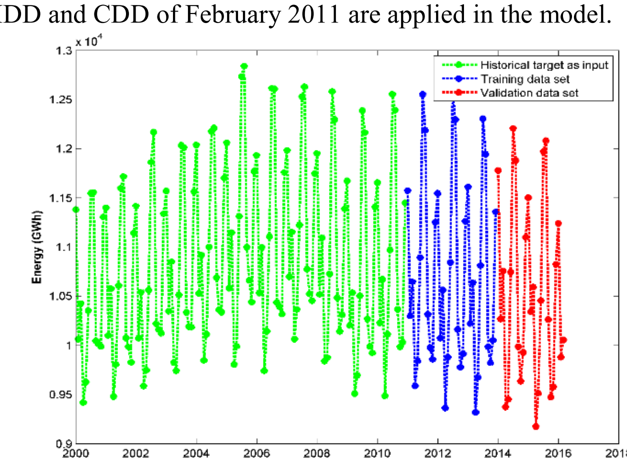

Fig. 5. Training and validation set of data

As mentioned before, the most common machine learning methods in load forecasting are ANN, SVR, RNN, GRNN, KNN, and GPR. All LF models are implemented using MATLAB® and the results for both training and validation set are compared with MAPE metric.

For the feedforward ANN, by trying with different hidden layers, one hidden layer and 3 neurons results in low forecasting errors for both training and validation data sets. The SVR method is applied for LF using LIBSVM [21].

Table I represents the results of the 6 forecasting models. In this table, the results are resented in MAPEs for both training and validation data sets.

As seen in the table, although the results of LF for all methods are close to each other, the feedforward ANN represents better results than other methods for the validation set while it also has decent result for the training set. Note that the MAPE for training set in the KNN method is zero since the same data for model training is applied in testing of training data set.

TABLE I. RESULTS OF LF MODELS FOR TRAINING AND VALIDATION DATA SETS

Method	Training	Validation
ANN	0.7	**1.5**
SVR	0.9	1.7
RNN	0.7	1.9
GRNN	0.7	2.3
KNN	0.0	2.3
GPR	0.6	2.0

The result of ANN is also represented in Fig. 6. In this figure, the green color graph exhibits the actual energy and the blue and red ones demonstrate the training and validation set, respectively. Note that since the ANN method, random process is applied, each running of the simulation may result in slightly different values. In this regard, the average of several simulation running is considered as ANN's final result.

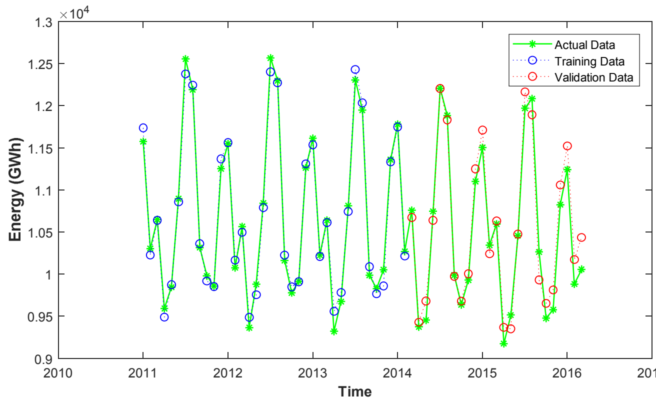

Fig. 6. LF using ANN model

IV. CONCLUSION

In this study, the performance of the most commonly used machine learning methods in load forecasting has been studied. These methods are feedforward artificial neural network (ANN), support vector regression (SVR), recurrent neural network (RNN), k-nearest neighbors (KNN), generalized regression neural network (GRNN), and Gaussian Process Regression (GPR). The case study is New England Network and its monthly energy usage during 2000 to May 2016 is considered for training and validation of the load forecasting models. The inputs of the load forecasting models are weather indicators (HDD and CDD), dummy variable of month number, and the moving average of the target variable before 2011. The results of forecasting models which are represented by MAPE indicate that although for both training and validation data set, all LF methods depict proficient performance, the feedforward ANN method shows better results than the other forecasting methods.

REFERENCES

[1] H. Sadeghian and Z. Wang, "Decentralized Demand Side Management with Rooftop PV in Residential Distribution Network," in *Innovative Smart Grid Technologies Conference (ISGT), 2018 IEEE Power & Energy Society*, 2017, pp. 1-5.

[2] M. Ghorbaniparvar, X. Li, and N. Zhou, "Demand side management with a human behavior model for energy cost optimization in smart grids," in *Signal and Information Processing (GlobalSIP), 2015 IEEE Global Conference on*, 2015, pp. 503-507.

[3] M. Motalleb, A. Eshraghi, E. Reihani, H. Sangrody, and R. Ghorbani, "A Game-Theoretic Demand Response Market with Networked Competition Model," in *49th North American Power Symposium (NAPS)*, 2017.

[4] M. H. Athari and Z. Wang, "Impacts of Wind Power Uncertainty on Grid Vulnerability to Cascading Overload Failures," *IEEE Transactions on Sustainable Energy,* vol. 9, pp. 128-137, 2018.

[5] J. W. Taylor, "An evaluation of methods for very short-term load forecasting using minute-by-minute British data," *International Journal of Forecasting,* vol. 24, pp. 645-658, 2008.

[6] T. Hong, *Short term electric load forecasting*: North Carolina State University, 2010.

[7] N. Amjady and A. Daraeepour, "Midterm demand prediction of electrical power systems using a new hybrid forecast technique," *Power Systems, IEEE Transactions on,* vol. 26, pp. 755-765, 2011.

[8] R. J. Hyndman and S. Fan, "Density forecasting for long-term peak electricity demand," *Power Systems, IEEE Transactions on,* vol. 25, pp. 1142-1153, 2010.

[9] H. Sangrody and N. Zhou, "An initial study on load forecasting considering economic factors," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, Boston, MA, 2016, pp. 1-5.

[10] A. Sadeghi-Mobarakeh, M. Kohansal, E. E. Papalexakis, and H. Mohsenian-Rad, "Data mining based on random forest model to predict the California ISO day-ahead market prices," in *Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2017 IEEE*, 2017, pp. 1-5.

[11] C.-L. Hor, S. J. Watson, and S. Majithia, "Analyzing the impact of weather variables on monthly electricity demand," *Power Systems, IEEE Transactions on,* vol. 20, pp. 2078-2085, 2005.

[12] E. Foruzan, S. D. Scott, and J. Lin, "A comparative study of different machine learning methods for electricity prices forecasting of an electricity market," in *North American Power Symposium (NAPS), 2015*, 2015, pp. 1-6.

[13] H. Liu, X.-w. Mi, and Y.-f. Li, "Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network," *Energy Conversion and Management,* vol. 156, pp. 498-514, 2018.

[14] X. Chen, X. Chen, J. She, and M. Wu, "A hybrid time series prediction model based on recurrent neural network and double joint linear–nonlinear extreme learning network for prediction of carbon efficiency in iron ore sintering process," *Neurocomputing,* vol. 249, pp. 128-139, 2017.

[15] N. K. Ahmed, A. F. Atiya, N. E. Gayar, and H. El-Shishiny, "An empirical comparison of machine learning models for time series forecasting," *Econometric Reviews,* vol. 29, pp. 594-621, 2010.

[16] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning* vol. 1: MIT press Cambridge, 2006.

[17] J. W. Taylor and R. Buizza, "Neural network load forecasting with weather ensemble predictions," *Power Systems, IEEE Transactions on,* vol. 17, pp. 626-632, 2002.

[18] H. Shih-Che, L. Chan-Nan, and L. Yuan-Liang, "Evaluation of AMI and SCADA Data Synergy for Distribution Feeder Modeling," *Smart Grid, IEEE Transactions on,* vol. 6, pp. 1639-1647, 2015.

[19] K. Siwek, S. Osowski, and R. Szupiluk, "Ensemble neural network approach for accurate load forecasting in a power system," *International Journal of Applied Mathematics and Computer Science,* vol. 19, pp. 303-315, 2009.

[20] T. Chen and Y.-C. Wang, "Long-term load forecasting by a collaborative fuzzy-neural approach," *International Journal of Electrical Power & Energy Systems,* vol. 43, pp. 454-464, 2012.

[21] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST),* vol. 2, p. 27, 2011.