Net Interchange Schedule Forecasting Using Bayesian Model Averaging

Maria Vlachopoulou, *Member, IEEE,* Luke Gosink, Trenton Pulsipher, Ryan Hafen, Jeremiah Rounds, Ning Zhou, *Senior Member, IEEE,* and Jianzhong Tong, *Senior Member, IEEE*

***Abstract*—The future power grid will need to incorporate systems and processes with a higher degree of variability and randomness due to the penetration of renewable energy resources and the increase of energy demand. Forecasting variables in a more uncertain environment poses new challenges and revisions of the existing forecasting methodologies will have to be made to maintain forecasting accuracy. This paper investigates an ensemble-based technique called Bayesian Model Averaging (BMA) to improve the performance of Net Interchange Schedule (NIS) forecasts. BMA is used to combine an ensemble of five diverse forecasting methods that each estimate NIS. The results, which examine performance for two separate years of real-world NIS data, demonstrate that BMA's aggregated forecasts reduces forecasting error by 30-55% in comparison to all individual prediction methods. This work illustrates a new possible mechanism for improving NIS forecasting accuracy, as well as other power grid system variables, and lays the foundation for future work on aggregate models that can balance computational cost with prediction accuracy.**

***Index Terms*—Bayesian model averaging, forecasting, interchange schedule, prediction, time series**

I. Introduction

To improve the efficiency and reliability of power grid operations, neighboring RTOs and ISOs often exchange electric power. Net interchange schedule is the sum of the import and export MW transactions between an ISO and its neighbors. One primary task of the ISO is to make the actual net interchange follow the NIS in real time so that the Area Control Error (ACE) stays within an acceptable limit.

Because many efficient generators respond relatively slowly, an ISO often runs security constrained economic dispatch (SCED) to economically dispatch generation resources several hours ahead of the current time. One major constraint of the SCED is to achieve a desired NIS by matching generation and load. If the future NIS can be precisely predicted, this look-ahead approach minimizes the generation cost by dispatching the most efficient generators. Therefore, effective forecasting of the NIS can improve the operation efficiency of the ISO by enabling SCED ahead of the current time. Note that in real time operation, the predicted NIS may be different from its real values. To make interchange follow the real NIS, the automatic generation control (AGC) adjust regulation generation reserves to compensate for the mismatch between real and predicted NIS. Larger prediction errors in NIS call for larger amount of generation reserve and therefore incur the extra cost for purchasing generation reserves.

Variable forecasting has been studied for a long time and forecasting methods range from simple to very complex. NIS forecasting can be categorized either as a time series or a causal forecasting problem, depending on the forecasting approach followed, as explained in greater detail in Section II. Autoregressive (AR) or Autoregressive Integrated Moving Average (ARIMA) methods are commonly used for time series forecasting. They are relatively simple to implement, but choosing the correct model order requires a lot of experimentation. Under the category of causal forecasting, linear regression, Classification And Regression Trees (CARTs), Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and others are commonly applied to derive a variable, like NIS, forecast. Each of these methods has its advantages and shortfalls. For example, linear regression is simple to implement, but can only model linear correlations between variables. Regression trees produce an easy to interpret model, but are susceptible to over-fitting. ANNs can model nonlinearities, but they require a large number of parameter tuning and they are non-intuitive. SVMs have really good modeling capabilities, but require a large amount of data for training.

Recently ensemble-based algorithms have been researched more extensively as a way to form aggregate forecasts. In ensemble approaches, estimates from a collection of forecasting methods are combined (e.g., through a weighted average) to form a single aggregated forecast. The motivation behind ensemble-based approaches is based on two principles: 1) all methods in the ensemble possess some unique, useful information; and, 2) no single method is sufficient to fully account for all uncertainties. Proponents of ensemble-based approaches assert that the best forecasting approach to use for estimation is a combination of all of the methods. The underlying premise behind this tenet is that the information and strengths of individual methods can be combined, and their corresponding weaknesses and biases can be overcome by the strength of the group [1]–[4]. Ensemble-based estimates are therefore expected to be more reliable and potentially more accurate than individual methods, an expectation that has been upheld in numerous examples [1], [2], [5]–[7].

This work leverages a ensemble approach called Bayesian Model Aggregation (BMA) to predict NIS. This paper specifically uses BMA in this work because BMA-based predictions

This work was supported by Laboratory Directed Research and Development funds from the Pacific Northwest National Laboratory (PNNL). The Pacific Northwest National Laboratory is operated by Battelle for the U. S. Department of Energy under Contract DE-AC05-76RL01830.
M. Vlachopoulou, L. Gosink, T. Pulsipher, and R. Hafen are all with the Pacific Northwest National Laboratory, Richland, WA 99352, USA (e-mail: maria.vlachop@gmail.com). N. Zhou is with Binghamton University, J. Rounds is with Purdue University. J. Tong is with PJM Interconnection RTO.

are highly accurate [2], [6], and are known to outperform predictions obtained from other ensemble methods including boosting and bagging [8], as well as a variety of regression techniques. A simple study of the application of BMA for NIS forecasting is described in [9]. As discussed in Section III, BMA's predictive benefits are based on its ability to address uncertainties associated with model specification. These uncertainties are arguably the source of error in predictive modeling [2]–[4], [7], [10].

The remaining of the paper is organized as follows. Section II provides an overview of application examples of forecasting in the power systems domain area. In Section III, the BMA model and the individual forecasting methods of the ensemble are described. Also in Section III, the dataset used for training and estimating of the forecasting algorithms is described. In Section IV, we present a series of results demonstrating the efficacy of applying BMA for NIS forecasting. Finally, some conclusive comments and directions for future work can be found in Section V.

II. Applications of Variable Forecasting

Variable forecasting is an integral part of many applications found in diverse domain areas. From the traditional engineering domains to finance, medical technology, and human behavior analytics, the fundamental methodologies applied are common and have been studied for over 40 years. In the field of power systems engineering, forecasting has been used to forecast load, power output, energy prices, and wind. The accuracy of these forecasts influences the power grid operation and reliability, thus numerous efforts have been made in improving the forecasts.

Typically, variable forecasting is categorized either as time series or as causal forecasting. Forecasting methods are classified as time series forecasting when previous time observations of the dependent variable are used to predict the dependent variable. On the other hand, in causal forecasting a set of explanatory variables that have a causal relationship to the dependent variable are used to predict the dependent variable. Another way of classifying forecasting methods is based on the time horizon used, which can be short, mid or long-term. With reference the power engineering field, a short-term horizon consists of one to a few hours, a mid-term horizon consists of a day to a few days or weeks and a long-term horizon consists of a few months or years' worth of data. Classic forecasting methodologies, like linear regression and artificial neural networks, are still prevalent in many industrial applications, while there is on-going research on refining and improving the classic methods, on deriving new ones and on combining multiple methods.

Load forecasting is probably the most studied forecasting problem in the power systems domain area. Early literature from the 90s mainly revolved around linear regression techniques and some artificial neural networks. With the development of new statistical and machine learning algorithms, engineers started testing their applicability to load forecasting. ANNs have been extensively tested and used, like for example in [11], [12] in load forecasting for short [13] and mid-term time horizons. In more recent years, other methods have been used like SVMs [14], [15], abductive networks [16] and semi-parametric additive models [17]. The methods described above focus on causal modeling of the load in relation to explanatory variables, like weather, that influence the load variation. Other methods like AR and ARIMA derive load forecasts by modeling the load as a time series [18]. As pointed out in [13] load forecasting is difficult because it depends on different levels of seasonality and on many explanatory variables. This explains the different types for forecasting models that need to be derived to capture special cases like holidays and extreme weather conditions.

Another common area of interest where forecasting has an important role is in renewable energy power output prediction. The increasing penetration of renewable energy resources in the power grid accentuates the importance of accurately forecasting their power output. This is not a trivial task because the power output from renewable resources is known to have a high degree of volatility due to its dependence on weather variables, like wind for wind power generation and solar irradiation for photovoltaic (PV) power generation. The statistical and machine learning methods described above are used for forecasting wind power output [19], [20] and PV power output [21].

Applying forecasting algorithms in the area of power system markets is also widely applicable, particularly after the electricity markets deregulation. ANNs have been used for forecasting market clearing prices and also in a hybrid models with ARIMA in [22] to forecast short-term electricity prices. In the deregulated market the locational marginal prices (LMPs) have a significant role in delivering market price signals. [23] describes how the LMPs can be forecasted under load uncertainty and [24] describes how congestion, which is one of the main influencing factors of the LMPs, can be forecasted.

III. Forecasting Model

In this section, the mathematical formulation of the Bayesian model aggregation ensemble approach is explained. Additionally, the individual forecasting methods, whose outputs the ensemble approach aggregates, are briefly described.

A. Bayesian Model Aggregation

For NIS forecasts, a basic BMA approach is to consider a set of forecasting methods as a linear system [2]–[4]. Let y_i for $i = 1, \ldots, N$ be a series of historical NIS observations, and let x_{ij} denote the i^{th} estimate obtained from the j^{th} forecasting method for these observations. Given P forecasting methods, the combination of all x_{ij} forms the numerical ensemble estimate matrix that, along with y_i, defines a linear regression model

$$y_i = \sum_{j=1}^{P} x_{ij}\beta_j + \epsilon_i \tag{1}$$

Here, the parameter vector β_j defines the unknown relationship between the ensemble's P constituents and ϵ_i is the disturbance term that captures all factors (e.g., noise and measurement error) that influence the dependent variable y_i *other* than the regressors x_{ij}.

In evaluating (1), the objective is to estimate the values β_j that will both fit the known NIS data in y_i and facilitate the ability to make inferences on future NIS. Many different regression techniques can estimate β_j [25], [26]; however, these techniques commonly generate estimates that vary in their ability to consistently forecast [2]–[4], [8]. The risk and uncertainty associated with using one of these forecasts over any other is called *statistical model uncertainty*. This uncertainty is arguably one of the greatest sources of error and bias for forecasting [2]–[4], [7], [10].

BMA addresses the challenge of statistical model uncertainty by first evaluating all possible models that can be formed from the P forecasting methods, and then combining each model's estimates for β_j through a weighted average. This aggregation process generates an aggregate-based parameter vector, β_j^{BMA} (2) that can provide more accurate and reliable forecasts than any individual ensemble constituent, and can also outperform other ensemble-based strategies (e.g., stepwise regression) [2], [8].

Formally, there are $k = 1, \ldots, 2^P - 1$ distinct combinations of the P forecasting methods, each with a corresponding statistical model, $M^{(k)}$, and parameter vector, $\beta_j^{(k)}$. BMA combines each $\beta_j^{(k)}$, through a weighted average that weights each $\beta_j^{(k)}$ by the probability that its statistical model, $M^{(k)}$, is the "true" model.

$$\beta_j^{BMA} = E[\,\beta_j \,|\, \mathbf{y}\,] = \sum_{k=1}^{2^P-1} E[\beta_j^{(k)} \,|\, \mathbf{y}, M^{(k)}\,] \Pr(M^{(k)}|\mathbf{y}) \tag{2}$$

In 2, $E[\beta_j^{(k)} \,|\, \mathbf{y}, M^{(k)}\,]$ is the expected value of the posterior distribution of $\beta_j^{(k)}$ that is weighted by the posterior probability $\Pr(M^{(k)}|\mathbf{y})$ (i.e., the probability that $M^{(k)}$ is the *true* statistical model given y_i). The expected posterior distribution of $\beta_j^{(k)}$ is approximated through the linear least squares solution of the given model $M^{(k)}$ and NIS response variable, $\mathbf{y} = [y_1, \ldots, y_N]$. The posterior probability term is estimated from information criteria [4]

$$\Pr(M^{(k)}|\mathbf{y}) \propto \frac{e^{-\frac{1}{2}B^{(k)}}}{\sum_{l=1}^{2^P-1} e^{-\frac{1}{2}B^{(l)}}} \tag{3}$$

where $B^{(k)}$ is the Bayesian Information Criteria for model $M^{(k)}$, and the information criteria itself is estimated [4]

$$B^{(k)} \approx N \log{(1 - R^{2(k)})} + p^{(k)} \log N \tag{4}$$

Here $R^{2(k)}$ is the R^2 correlation value for model $M^{(k)}$, $p^{(k)}$ is the number of methods used by the model (not including the intercept), and N is the number of NIS values to be predicted. BMA's aggregation thus weights each model's expected parameter vector $b_j^{(k)}$ with the probability value that is based on that model's ability to balance trade-offs between model complexity (i.e., the number of methods used) and goodness of fit. Models that use a larger number of methods, or that do not fit the observations well, are penalized and can be eliminated from the final aggregation process (i.e., their posterior probabilities are effectively 0). In this context, BMA combines the *best* models to provide an accurate estimate for the *true* parameter terms, β_j.

The resulting parameter vector, β_j^{BMA}, obtained from (2) helps to address model uncertainty by accounting for all systems of linear equations that can model the relationship between the measured NIS values y_i and values x_{ij} forecasted by each method j. More importantly, β_j^{BMA} can be used to forecast NIS by combining new x_{ij} predictions.

B. Description of Data and the Forecasting Ensemble

We apply the BMA approach to an ensemble of prediction methods that were trained to forecast NIS. We used a real-world dataset from two different years to train and evaluate the BMA approach and each of the forecasting methods. In this context, our results are presented in Section IV as two separate experiments: experiment one is based on data from 2012 (January through October), and experiment two is based on data from 2013 (January to June). These experiments allow to test the BMA model for winter, summer and spring season; fall is very similar to spring with regards to energy demand patterns and NIS.

The data contains the dependent variable, total NIS submitted to PJM by the neighboring pools, and independent variables like load and shadow price in PJM, time and seasonal variables. It is well known that Time of Day (ToD), Day of Week (DoW) and season significantly affect power and energy demand, and consequently NIS. This intuition is confirmed in the plots of Section IV. The load and shadow price are always positive, while NIS can be positive or negative when a neighboring pool has respectively scheduled an import or export of power from or to PJM. There is also a number of derived variables used that are based on a trend analysis of some of the variables, as described in [27]. Both original and derived variables get time-aligned to form an input data matrix that serves as input for most of the forecasting methods described below.

The ensemble itself contains forecasting techniques from a variety of statistical methods including Sparse Regression, Support Vector Regression, Autoregression, Random Forests, and Artificial Neural Networks. We summarize these techniques below.

- ***Sparse Regression (SR)*** [28] The sparse regression implementation with Figueiredos Normal-Jeffreys prior is a Bayesian approach to sparse regression that uses a hierarchical Normal-Jeffreys prior instead of the Lasso L1 penalty. The advantage of the method is that it is scale-invariant and adapts automatically to the level of sparsity of the data without the need to tune parameters. In this work, our sparse regression implementation takes as input 12 hours of the input data matrix described above to specify a model that forecasts one value of NIS.
- ***Support Vector Regression (SVR)*** [29] SVR methods work by solving a constrained quadratic problem where the convex objective function for minimization is defined by a loss function and a regularization term that controls

the complexity of the hypothesis space. As SVRs minimize both training error and complexity, their forecasting performance is often significantly better than techniques that rely only on loss function minimization. In this work, our SVR implementation takes as input 12 hours of the input data matrix described above to specify a model that forecasts one value of NIS.

- ***Autoregression (AR)*** [30] AR methods forecast future values in a time series through a weighted average of previously observed values. The number of historical values used by the AR model indicates the model's *order*. Formally, AR(p) indicates that forecasts are based on a linear combination of the previous p values in the series. Solving for the weights in the linear combination is done through maximum likelihood analysis. This work uses an AR(12) model to forecast NIS.
- ***Random Forest (RF)*** [31] RFs extend general regression techniques with a learning strategy. The strategy begins by constructing an ensemble of regression trees during training, and then uses these trees to forecast the next value in the time series. The final RF forecast is based on the mean forecast of all regression trees. RFs are similar to BMA in that they combine an ensemble of forecasts to form an aggregated forecast. However, while RFs assign equal weights to each forecast in the ensemble, the BMA approach weights each forecast based on the posterior probability that the underlying statistical model is the correct model (2). This weighting helps to better account for statistical model uncertainty as models can be weighted according to information criteria. This work's RF model takes as input 12 hours of the input data matrix described above to specify a model that forecasts one value of NIS.
- ***Artificial Neural Network (ANN)*** ANNs model the relationship between a set of explanatory variables and a response variable (e.g., NIS). While most regression methods model such relationships through a linear combination of the explanatory variables, ANNs model this relationship through the use of arbitrary, nonlinear multi-parametric discriminant functions. Resultantly, ANNs can model more complex dependencies between y and X and so many domains employ ANNs for forecasting for their increased fidelity. The ANN used in this work is based on 12 hours of the input data matrix described above to specify a model that forecasts one value of NIS.

C. Training and Estimating with BMA

Each forecasting strategy discussed in Section III-B used a sliding window approach for training and testing. In this procedure, each method used 12 hours worth of data to forecast NIS for the next hour. For example, methods whose training set spans 9am to 8:59pm will forecast NIS for the interval that starts at 9pm and ends at 9:59pm. Once forecasts are made, the window slides ahead by a one hour interval. This process is illustrated in Fig. 1.

Similarly, the BMA model was trained with a sliding window strategy. To train BMA, we combined 12 hours of

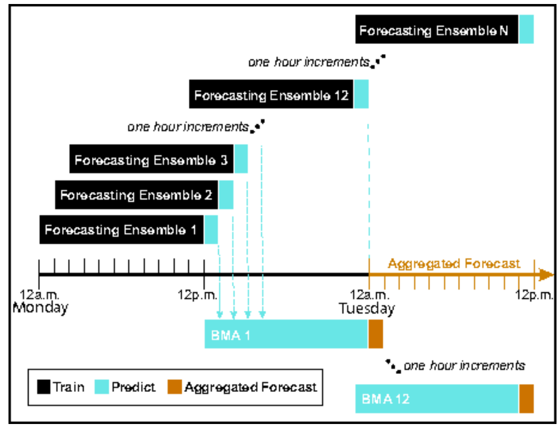

Fig. 1: This figure illustrates the sliding time window strategy used to train and then forecast NIS.

initial forecasts by the forecasting ensemble. The combination of these forecasts form the numerical ensemble estimate matrix ,$x_{i,j}$, in (1). We then solved for β_j^{BMA} as discussed in Section III-A. With this coefficient vector, we then combined the next hour's worth of forecasts made by the ensemble to form an aggregate forecast for the next hour. After the forecasts were made, the training window slides up by one hour.

We used the root mean squared error (RMSE) to assess the forecasting performance for each method. Thus we assess method j's forecasting accuracy

$$RMSE_j = \sqrt{\frac{\sum_{i=1}^{N}(y_i - x_{i,j})^2}{N}} \tag{5}$$

where y_i for $i = 1, \ldots, N$ is the series NIS observations, and x_{ij} is the estimate from method j for these observations. For each hour, there are four NIS observations. Accordingly, N = 4 and we compute RMSE for each hour for each model to asses forecasting performance.

IV. Results

We apply our aggregate forecasting approach to the NIS data described in Section III-B. To assess forecasting performance for the five ensemble constituents and BMA, we separate our results into performance calculated based on RMSE for 2012 and performance calculated in 2013. For each year we examine this error based on three attributes: 1) we report the overall annual forecasting performance for the ensemble of methods, as well as BMA's aggregate forecast (Section IV-A); 2) we analyze the mean forecasting error observed through out the year based on the day of the week (Section IV-B); and finally, 3) we analyze the mean forecasting error observed through out the year based on the time of day (Section IV-C).

A. Annual Forecasting Performance

One of the most important measures of the performance of a forecasting algorithm is the overall accuracy and performance improvement in comparison to other methodologies. As mentioned in Section II, accurate forecasting is critical to

TABLE I: Performance by Forecasting Method

Method class	2012	2013
BMA	378.5	280.8
ANN	707.9	517.0
AR	600.3	459.5
SR	872.7	648.1
RF	533.3	357.0
SVR	741.6	456.7

maintain reliable and efficient operation of a power grid. The NIS overall forecasting accuracy using the suggested BMA ensemble method is discussed, based on the results presented in Fig. 2 and Table I.

Fig. 2 shows the cumulative RMSE of each forecasting method in the ensemble, as well as BMA's aggregated forecast. This figure shows the results for NIS data between January and November of 2012 (left) and data observed between January and June of 2013 (right). The cumulative error in both plots is generated by summing the RMSE computed for each hour that NIS was observed.

In both years, the aggregated forecasts made by BMA significantly outperform each individual method in the ensemble. Table I summarizes the total benefits by listing the percent reduction in error BMA provides over each forecasting method for the complete data set. With respect to the forecasting ensemble, the random forest (RF) method provides the next best forecasting performance for both years. Note that the performance trends for the next best performing methods change between 2012 and 2013. More specifically, in 2012 the best performing forecasts after RF are AR, ANN, and SVR where as in 2013 the best performing forecasts after RF are SVR, AR, and ANN. In both years, sparse reduction (SR) method provides the least accurate forecasts. Note that despite the changing performance of the methods, the BMA-based forecast that aggregates these methods provides consistently better performance. Also note that the general performance of forecasts is better in 2013 than 2012.

B. *Performance Based on the Day of Week*

The power grid operation is driven by consumer and in general human behavior. As a result weekly effects on the power demand and supply are expected, whether for residential or industrial units. For example, during the working days of a week, Monday to Friday for the USA, there is more uncertainty on the power supply and demand cycle because all industrial units are operational. During the weekend a lot of the industrial units and office space units require much less power, because they are not operational or just perform basic maintenance functions. This phenomenon could be the driving reason behind some of the trends observed from our NIS forecasting analysis.

Fig. 3 shows the mean RMSE per day for the 2012 data set (left) and 2013 data set (right). These numbers are determined by calculating the mean NIS value for each day of the week based on the total days in the given year's data. The performance trends for best forecasting methods parallel the performance trends in Fig. 2: BMA provides the best forecasting performance. Note also the forecasting performance for all methods improves between 2012 and 2013. We observe slight Day of Week effects as NIS forecasting error (RMSE) increases during the middle of the week and then tapers off during the weekend. These effects are based on forecasting errors that increase due to the increase of the power supply and demand uncertainty during the working week.

C. *Performance Based on the Time of Day*

The human behavior effect on the power demand and NIS has another temporal component, the Time of Day. It is well known that during certain times in the day the power demand is higher than other times. There is usually a peak observed between 6:00am to 9:00am and another between 16:00pm to 19:00pm. The peaks are driven by standard human working schedules. When assessing the BMA forecasting accuracy, its performance based on ToD was analyzed and the results are summarized in Fig. 4.

Fig. 4 shows the mean RMSE per hour (for any given day based on 24 hours) for the 2012 data set (left) and 2013 data set (right). These numbers are determined by calculating the mean NIS value for each hour of the day based on the total hours in the given year's data. This figure illustrates that BMA-based forecasts provide the best NIS forecasts for any given hour. Note that forecasting trends in this figure are even more pronounced than the time of day effects for error shown in Fig. 3. More specifically, around 6:00am in both 2012 and 2013 data, forecasting errors rise sharply for all forecasting methods. This effect is based on forecasting errors that increase in magnitude as NIS becomes more uncertain with higher variability during the start of the work day. Note that there is also a similar spike in RMSE between 22:00pm and 23:00pm. Also note that these effects appear to be damped in 2013 in comparison to 2012.

The spikes in the forecasting error can be further reduced by including a forecasting model in the ensemble capable of capturing this periodic behavioral pattern more efficiently. This is part of the future work for this research and is discussed in Section V.

V. CONCLUSION

There are certain features of the power grid that are changing to accommodate for new technological developments, such as incorporation of renewable energy resources and increase in power and energy demand. The new features introduce a higher degree of uncertainty that causes currently applied forecasting methods to be challenged. There are efforts in progress for testing new statistical and machine learning methods to forecast power grid variables with the desired accuracy ensuring robustness of the power grid system. NIS is one of the important variables that influences the power grid system operation. In this research we have concentrated on applying an ensemble approach called BMA to forecast NIS. Ensemble approaches are strong candidates for performing forecasting and other statistical analysis because of their unique capability of combining diverse individual statistical methods and producing a single output that better models the variable or system under examination.

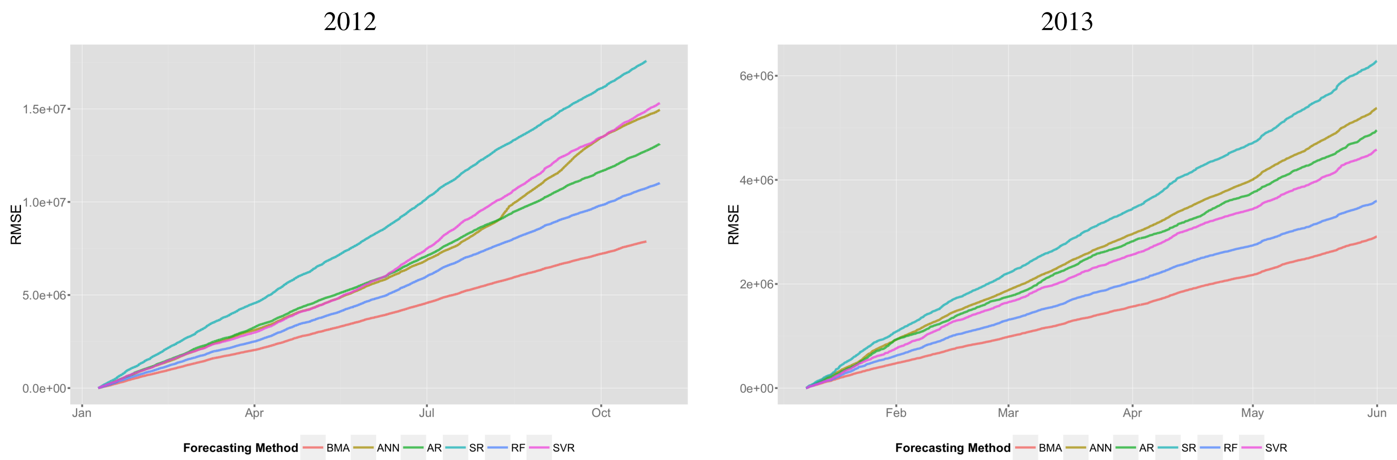

Fig. 2: This figure shows the cumulative forecasting error for two different years - 2012 (left) and 2013 (right) - based on root mean square error (RMSE). In this figure we contrast BMA's forecasting performance to the performance of methods in the forecasting ensemble. Here BMA is shown to provide significantly more accurate NIS forecasts throughout the year in comparison to any of the other forecasting methods.

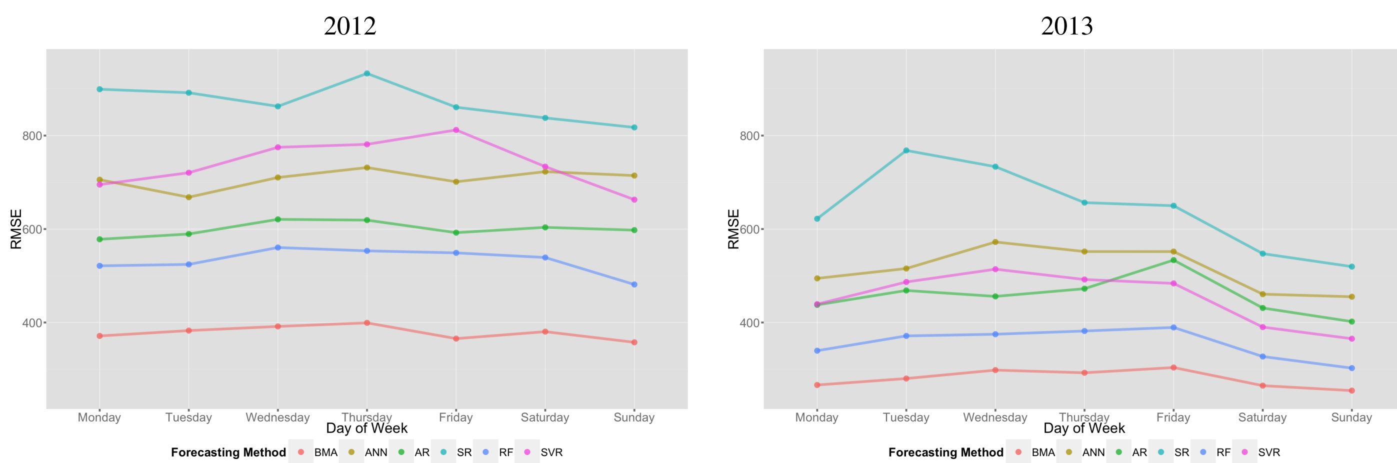

Fig. 3: This figure shows forecasting errors for given days of the week based on two different sets of data: 2012 (left) and 2013 (right). The error, which is based on root mean square error (RMSE), is calculated by taking the mean RMSE of the forecasts made for that day throughout the given year. This figure demonstrates that BMA provides consistently better performance for any given day of the week in comparison to the ensemble of forecasting methods.

The results of applying BMA for forecasting NIS are very promising and lay the foundation for applying this methodology on other power system variables. The results presented in Fig. 2 to Fig. 4 demonstrate that BMA's NIS forecasting accuracy is significantly better than the accuracy of each individual forecasting method. The performance of the BMA on NIS forecasting has been tested with respect to the entire dataset, the Time of Day and the Day of Week. In all cases, BMA's performance is superior to the performance of individual models.

As future work, there are a few areas where there is still room for improvement. First, the ToD analysis demonstrated that there are still patterns on the NIS forecasting error that could be removed by adding more forecasting models in the ensemble capable of capturing the temporal periodicity of human behavior and its effect on power demand and NIS. These model will probably be temporarily adaptive. Second, addition of more and more diverse forecasting models in the ensemble will render the BMA forecast more robust and accurate. We believe that the BMA superior performance in comparison to other ensemble approaches will be obvious. Finally, in this study the dependent variable to-be-forecasted is the aggregated NIS from all PJM's neighboring pools. Forecasting NIS of the neighboring pools individually might reveal more trends and information that gets averaged out when aggregated into total NIS to PJM.

ACKNOWLEDGEMENT

The authors would like to thank Francis K Tuffner with PNNL for his valuable suggestions and guidance in the area of power engineering. The authors would also like to thank Wayne Moodie with the PJM RTO for providing us with the datasets and for his technical insights. We would like to thank Aimee E Holmes with PNNL and Eleanor Wong with the

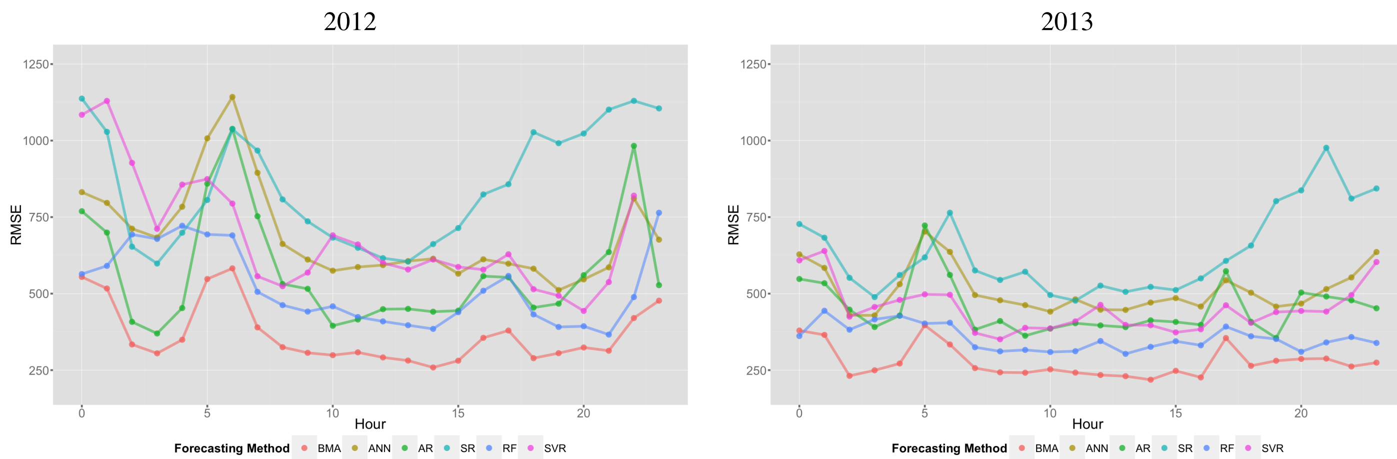

Fig. 4: This figure shows forecasting errors for given hours of the day based on two different sets of data: 2012 (left) and 2013 (right). The error, which is based on root mean square error (RMSE), is calculated by taking the mean RMSE of the forecasts made for that hour throughout the given year. This figure demonstrates that BMA provides consistently better performance for any given day of the week in comparison to the ensemble of forecasting methods. We address the performance spikes seen at 3:30am and 22:00pm in Section IV.

University of Utah for their comments and contribution related to the statistical methods used in this research. Finally, the authors would like to thank Tom Ferryman and David Haglin for laying the foundation for the development of this project.

References

[1] G. Seni and J. F. Elder, "Ensemble methods in data mining: Improving accuracy through combining predictions," *Synthesis Lectures on Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 1–126, 2010. [Online]. Available: http://dx.doi.org/10.2200/S00240ED1V01Y200912DMK002

[2] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: A tutorial," *Statistical Science*, vol. 14, no. 4, pp. 382–417, 1999.

[3] A. E. Raftery, D. Madigan, and J. A. Hoeting, "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association*, vol. 92, pp. 179–191, 1998. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.42.4400

[4] A. Raftery, "Bayesian Model Selection in Social Research," *Sociological Methodology*, vol. 25, pp. 111–163, 1995.

[5] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[6] E. Morales-Casique, S. P. Neuman, and V. V. Vesselinov, "Maximum likelihood Bayesian averaging of airflow models in unsaturated fractured tuff using Occam and variance windows," *Stochastic Environmental Research and Risk Assessment*, vol. 24, pp. 863–880, 2010.

[7] A. E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski, "Using Bayesian model averaging to calibrate forecast ensembles," *Monthly Weather Review*, vol. 133, no. 5, pp. 1155–1174, 2005. [Online]. Available: http://dx.doi.org/10.1175/MWR2906.1

[8] I. Davidson and W. Fan, "When efficient model averaging out-performs boosting and bagging." in *Principles and Practice of Knowledge Discovery in Databases (PKDD)*, ser. Lecture Notes in Computer Science, vol. 4213, 2006, pp. 478–486.

[9] M. Vlachopoulou, L. Gosink, T. C. Pulsipher, T. Ferryman, N. Zhou, and J. Tong, "An ensemble approach for forecasting net interchange schedule," in *Power and Energy Society General Meeting, 2013 IEEE*, 2013.

[10] C. T. Volinsky, D. Madigan, A. E. Raftery, and R. A. Kronmal, "Bayesian model averaging in proportional hazard models: Assessing the risk of a stroke," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 46, no. 4, pp. 433–448, 1997. [Online]. Available: http://dx.doi.org/10.1111/1467-9876.00082

[11] J. W. Taylor and R. Buizza, "Neural network load forecasting with weather ensemble predictions," *Power Systems, IEEE Transactions on*, vol. 17, no. 3, pp. 626–632, 2002.

[12] C. Guan, P. B. Luh, L. D. Michel, Y. Wang, and P. B. Friedland, "Very short-term load forecasting: Wavelet neural networks with data pre-filtering," 2013.

[13] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *Power Systems, IEEE Transactions on*, vol. 16, no. 1, pp. 44–55, 2001.

[14] B.-J. Chen, M.-W. Chang *et al.*, "Load forecasting using support vector machines: A study on eunite competition 2001," *Power Systems, IEEE Transactions on*, vol. 19, no. 4, pp. 1821–1830, 2004.

[15] L. Ghelardoni, A. Ghio, and D. Anguita, "Energy load forecasting using empirical mode decomposition and support vector regression," 2013.

[16] R. E. Abdel-Aal, "Short-term hourly load forecasting using abductive networks," *Power Systems, IEEE Transactions on*, vol. 19, no. 1, pp. 164–173, 2004.

[17] S. Fan and R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model," *Power Systems, IEEE Transactions on*, vol. 27, no. 1, pp. 134–141, 2012.

[18] L. C. M. de Andrade and I. N. da Silva, "Very short-term load forecasting based on arima model and intelligent systems," in *Intelligent System Applications to Power Systems, 2009. ISAP'09. 15th International Conference on*, 2009, pp. 1–6.

[19] M. Alexiadis, P. Dokopoulos, and H. Sahsamanoglou, "Wind speed and power forecasting based on spatial correlation models," *Energy Conversion, IEEE Transactions on*, vol. 14, no. 3, pp. 836–842, 1999.

[20] G. Sideratos and N. D. Hatziargyriou, "An advanced statistical method for wind power forecasting," *Power Systems, IEEE Transactions on*, vol. 22, no. 1, pp. 258–265, 2007.

[21] J. Shi, W.-J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *Industry Applications, IEEE Transactions on*, vol. 48, no. 3, pp. 1064–1069, 2012.

[22] P. Areekul, T. Senjyu, H. Toyama, and A. Yona, "A hybrid arima and neural network model for short-term price forecasting in deregulated market," *Power Systems, IEEE Transactions on*, vol. 25, no. 1, pp. 524–530, 2010.

[23] R. Bo and F. Li, "Probabilistic lmp forecasting considering load uncertainty," *Power Systems, IEEE Transactions on*, vol. 24, no. 3, pp. 1279–1289, 2009.

[24] Q. Zhou, L. Tesfatsion, and C.-C. Liu, "Short-term congestion forecasting in wholesale power markets," *Power Systems, IEEE Transactions on*, vol. 26, no. 4, pp. 2185–2196, 2011.

[25] P. T. Reiss, L. Huang, J. E. Cavanaugh, and A. K. Roy, "Resampling-based information criteria for best-subset regression," *Annals of the Institute of Statistical Mathematics*, vol. 64, no. 6, pp. 1161–1186, 2012. [Online]. Available: http://dx.doi.org/10.1007/s10463-012-0353-1

[26] C. L. Mallows, "Some comments on C_p," *Technometrics*, vol. 15, no. 4, pp. 661–675, 1973. [Online]. Available: http://www.jstor.org/stable/1267380

[27] M. Vlachopoulou, T. Ferryman, N. Zhou, and J. Tong, "A stepwise

regression method for forecasting net interchange schedule," in *Power and Energy Society General Meeting, 2013 IEEE*, 2013.

[28] M. A. Figueiredo, "Adaptive sparseness for supervised learning," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 9, pp. 1150–1159, 2003.

[29] Smola, Alex J. and Schölkopf, Bernhard, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, aug 2004.

[30] Harvill, Jane L. and Ray, Bonnie K., "Functional coefficient autoregressive models for vector time series," *Computational Statistics & Data Analysis*, vol. 50, no. 12, pp. 3547–3566, 2006.

[31] Breiman, Leo, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

Maria Vlachopoulou (M10) received her M.Sc. in Electrical and Computer Engineering and her M.Sc. in Industrial Engineering (Operations Research) from Purdue University in 2010. She received her B.Sc. in Electronic and Electrical Engineering from the University of Sheffield, U.K. in 2006. She currently works for the Pacific Northwest National Laboratory as a Research Engineer in the area of machine learning, data mining and artificial intelligence. Her research revolves around application of these methods to the US power grid. She is a member of the IEEE Power and Energy Society (PES).

Dr. Luke Gosink joined Pacific Northwest National Laboratory in January of 2010. His primary research interests and expertise are in combining statistical modeling methods (e.g., Bayesian methods) and machine learning techniques with visualization strategies in order to obtain insight from large, complex scientific. His current research addresses topics ranging from ensemble analysis and design to uncertainty quantification and signature discovery in high-throughput systems. He has applied his research across a variety of application spaces including the power grid, sub-surface flow and transport, protein modeling, and experimental evaluation of instrumentation. Dr. Gosink has degrees in both Chemistry (B.S.) and Computer Science (Ph.D.) from the University of California at Davis.

Trenton Pulsipher holds degrees in Statistics (B.S. and M.S.) from Brigham Young University, and currently applies his knowledge as a Statistician at Pacific Northwest National Laboratory. His research interests include Bayesian model aggregation, big data analytics using Hadoop, and general statistical modeling and visualization methods.

Dr. Ryan Hafen received his B.S. in Statistics from Utah State University in 2004, M.Stat. in Mathematics from University of Utah in 2006, and Ph.D. in Statistics from Purdue University in 2010. He currently works at Pacific Northwest National Laboratory as a statistical scientist. His research interests include statistical model building, visualization, massive data, time series, and nonparametric statistics.

Jeremiah Rounds Jeremiah Rounds received his B.S. in Mathematics from University of Arizona in 2007, M.S. in Mathematical Statistics from Utah State University in 2009, and is currently working toward a Ph.D. in Statistics from Purdue University. His research interests include statistical model building, methods for learning from massive data, and statistical inference utilizing digital gene expression data from next generation sequencing technologies in RNA-Seq experiments.

Dr. Ning Zhou (S01- M05- SM08) is currently with the Electrical and Computer Engineering Department at the Binghamton University as an assistant professor. In 2005, he received his Ph.D. in Electrical Engineering with a minor in statistics from the University of Wyoming. From 2005 to 2013, Dr. Zhou worked as a power system engineer at the Pacific Northwest National Laboratory. His research interests include power system dynamics and statistical signal processing. Dr. Zhou is a senior member of the IEEE Power and Energy Society (PES).

Dr. Jianzhong Tong senior strategist for the Applied Solutions Division of PJM, is responsible for manage, design and develop new applications, systems and tools for PJM real time system operations and market systems by using advanced technologies.