ADABOOST^{+}: an ensemble learning approach for estimating weather-related outages in distribution systems

Padmavathy Kankanala, Sanjoy Das, and Anil Pahwa

Published Version Information

Citation: Kankanala, P., Das, S., & Pahwa, A. (2014). ADABOOST+: An ensemble learning approach for estimating weather-related outages in distribution systems. IEEE Transactions on Power Systems, 29(1), 359-367.

Digital Object Identifier (DOI): doi:10.1109/TPWRS.2013.2281137

Publisher's Link: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6607244

ADABOOST$^+$: An Ensemble Learning Approach for Estimating Weather-Related Outages in Distribution Systems

Padmavathy Kankanala, *Student Member, IEEE*, Sanjoy Das, *Non-Member, IEEE*, and Anil Pahwa, *Fellow, IEEE*

***Abstract*— Environmental factors, such as weather, trees and animals are major causes of power outages in electric utility distribution systems. Of these factors, wind and lightning have the most significant impacts. The objective of this paper is to investigate models to estimate wind and lighting related outages. Such estimation models hold the potential for lowering operational costs and reducing customer downtime. This paper proposes an ensemble learning approach based on a boosting algorithm, ADABOOST$^+$, for estimation of weather caused power outages. Effectiveness of the model is evaluated using actual data, which comprised of weather data and recorded outages for four cities of different sizes in Kansas. The proposed ensemble model is compared with previously presented regression, neural network, and mixture of experts models. The results clearly show that ADABOOST$^+$ estimates outages with greater accuracy than the other models for all four data sets.**

***Index Terms*— Artificial intelligence, ensemble learning, environmental factors, power distribution systems, power system reliability.**

I. INTRODUCTION

IT is a well recognized fact that weather conditions, specifically wind and lightning, have a great effect on outages in power distribution systems [1]. Literature on outage analysis shows that especially overhead lines are highly susceptible to environmental factors such as weather, trees, and animals [2]. Proper design and maintenance can help in reducing weather related outages, but it is hard to prevent them completely. Although outages are more likely during severe weather, their occurrences are highly irregular rendering them very difficult to predict. A model which can accurately estimate outages based on weather data can help utilities in outage management, system design and upgrades.

Weather is typically categorized into normal weather, severe weather and extreme weather. The National Weather Service defines extreme weather as any dangerous meteorological phenomena with the potential to cause damage, serious social disruption, or loss of human life. Extreme weather conditions include hurricanes, tornadoes, severe thunderstorms, snowstorms, and ice storms. Severe weather conditions are characterized by lightning, high wind, extreme temperature, and heavy rainfall. Utilities usually separate outages caused by extreme weather conditions from those caused by severe weather conditions while evaluating the system performance. Since the outages occur randomly with higher probabilities during adverse conditions and outage recordings can have significant human errors, obtaining high correlation between estimated outages and observed outages is a very challenging task. Various models have been proposed in the literature to study effects of different weather phenomenon on outages with different levels of success.

An exponential model as a function of time for forecasting cumulative outages during different extreme weather events has been proposed in [3]. In this paper, the authors have classified storms by the intensity of temperature and wind speeds. Also, flash data has been considered for analysis of outages caused by storm with lightning activity. Similarly, statistical models to predict the number of outages due to hurricanes and ice storms have been developed in [4, 5]. In these papers, the authors have developed the hurricane and ice storm models as a function of explanatory variables such as number of protective devices, maximum wind gust and duration, ice thickness, hurricane rainfall, storm indicator covariate, land cover type, soil drainage level and soil depth. These methods have limitation such as evolving power system inventory with time and presence of huge matrix of spatial correlation makes it computationally challenging. Poisson regression and Bayesian hierarchical network for risk management of power outages caused by extreme weather conditions is investigated in [6]. In this study, surface wind speed, gust speed, gust frequency, daily rainfall, daily minimum pressure and daily maximum and minimum temperature have been considered, while other weather factors such as lightning are excluded. A Poisson regression model and a Bayesian network model to predict the yearly weather-related failure events on overhead lines are presented in [7]. Similarly, in [8] Poisson regression is used to study the significance of weather variables on outages using outage data from substations within 10 miles of National Weather Service sites under severe weather conditions.

Prior work of authors of this paper to study effects of wind and lightning includes investigation of linear, quadratic and exponential regression models [9, 10], multilayered neural

This work was supported by the National Science Foundation under Award ECCS-0926020.

Padmavathy Kankanala (e-mail: pk150@ksu.edu), Anil Pahwa (e-mail: pahwa@ksu.edu), and Sanjoy Das (e-mail: sdas@ksu.edu) are with Department of Electrical and Computer Engineering, Kansas State University, Manhattan, KS 66506 USA.

networks [11], and mixture of experts (ME) [12]. Although these prior methods show acceptable performance, there remains enough scope for further improvement using state-of-the-art machine learning algorithms.

The main focus of this paper is build upon the prior work of the authors and the existing literature to explore techniques using ensemble learning to estimate with greater accuracy outages in power distribution systems caused by weather. Specifically, demonstrate that ensemble-based methods can do a better job than a single learner approach. In addition to a standard ADABOOST, an approach based on a new boosting algorithm, ADABOOST+, which is a modification proposed by the authors of this paper, is presented. Ensemble learning is a technique that embodies one of the main directions of current machine learning research. Although in this paper the ensemble's constituent units are referred to as neural networks, an ensemble could equally well comprise of other learning models, such as support vector machines [13, 14], kernel-based models [15], radial basis function networks [16, 17], decision trees [14, 18], fuzzy logic [19] or ARMA models [21] in addition to neural networks [16, 17, 19-21]. Random forests [22] are another good method for classification and regression. However, during preliminary investigation for the problem, we found the random forests algorithm for classification but not for regression. We also tried some kernel based approaches before settling down with backprop-trained neural networks which seemed to perform best. Furthermore, since the motivation behind ADABOOST is to obtain a strong learner using an ensemble of weak learners, the specific choice of weak learner is not a significant issue.

The rest of the paper is organized as follows. Section II is an exposition to ensemble learning along with a survey of the applications in power and energy systems that have only recently begun to appear. In Section III, the specific ensemble algorithm used in this research is presented in greater detail. Section IV outlines how available historical weather observations and outage data are processed in this investigation. Section V provides results obtained from this study including comparisons with earlier modeling approaches. Finally, section VI concludes this research.

II. ENSEMBLE LEARNING

Aggregating models into ensembles is performed due to a variety of reasons. For example, when the input data is both extensive and spatially distributed, instead of transporting large volumes of data into a single location, it is preferable to train neural networks at different nodes, and communicate only their outputs into a central facility for collective decision making. Motivated by these considerations, an ensemble learning algorithm whose inputs range from historical records of grid failures to distributed real-time sensor measurements of the grid, is suggested. The task is to estimate the mean times between failures of various power equipment of the electricity grid of New York city [13].

Elsewhere, data heterogeneity has led to schemes where neural networks are trained to only process a subset of input fields, with a separate dedicated unit used for decision fusion. Multiple neural networks are trained with different inputs, and an ensemble is used to forecast load conditions [19].

Ensembles are also used for classification and regression tasks, to increase prediction accuracies beyond what can be accomplished with a single neural network. The expected error of a trained neural network for test data consists of three components: (*i*) random noise, (*ii*) *bias*, and (*iii*) *variance*. Random noise pertains to anomalies present in the test data that cannot be alleviated through computational means. The second component – bias, refers to the neural networks own topological inadequateness when modeling the data. It can be reduced by increasing the network's complexity – such as adding more hidden neurons. Unfortunately, increased network size also leads to higher variance, *i.e.* the sensitivity of the network's parameters to the training samples, which is the third source of error. In other words, increasing the neural network size to improve its performance with respect to the training samples has the undesirable effect of degrading the network's overall performance. This is the well-known *bias-variance dilemma* in machine learning theory; decreasing the bias increases the variance and vice versa.

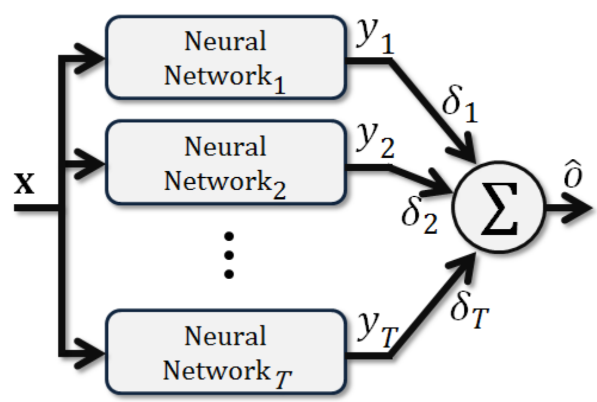

Fig.1. Schematic of an ensemble of neural networks.

Ensemble learning offers an alternative route to lower the variance without compromising the bias term [23-27]. This is done by aggregating the output over multiple, separately trained neural networks (Fig. 1). Although individual neural networks in the ensemble can exhibit high sensitivities to the training data, the variance of the collective output remains quite low. Even simple aggregation techniques, such as averaging the outputs of all neural networks, have shown great promise. A recent study reports that averaging the individual 1-9 day ahead weather predictions of several radial basis function networks is significantly more accurate than each network's output [20].

More advanced theoretical considerations have led to even better ensemble methods to lower the variance. In *bagging*, bootstrapped subsets of the training data are used to train each network in the ensemble [28]. This approach has been used for the short-term load forecasting using meteorological information [19].

On the other hand, in *boosting* weighted training of the neural networks is applied [26, 27]. ADABOOST is the most widely used algorithm for boosting ensemble outputs [25-27, 29-34]. The New York power grid study uses a realization of this basic method called RANKBOOST [13].

Each phase of ADABOOST involves the complete training of a separate neural network in the ensemble. In the initial phase, a neural network is trained with equal weights assigned to all samples in the test data. In each subsequent phase, a new neural network is trained with samples associated with high output errors from earlier networks receiving exponentially increasing weights. This scheme was originally used in classification tasks in the ADABOOST.M1 and ADABOOST.M2 algorithms [25, 31]. ADABOOST.R is an adaptation of ADABOOST.M2 for regression problems [33]. The algorithm that is considered for comparison in this paper is ADABOOST.RT, which is also meant for regression [27, 34].

III. PROPOSED APPROACH

Each of the T models used in the ensemble is a standard multilayered neural network with a single hidden layer. The output y of such a network is [11],

$$y = \sum_{j=1}^{M} \boldsymbol{W}_{jk}^{O} * \sigma\left(\sum_{i=1}^{D} \boldsymbol{W}_{ij}^{H}(x_i) + \boldsymbol{b}_i^{H}\right) + b^{O}. \quad (1)$$

In the above equation, M is the number of hidden neurons, and D, the size of the input $\mathbf{x}$ whose i^{th} component is x_i. The quantities $\boldsymbol{W}_{ij}^{H}$ and $\boldsymbol{W}_{jk}^{O}$ are the synaptic strengths of the neurons in the hidden and output layers, while $\mathbf{b}_i^{h}$ and b^{o} are their corresponding biases, with subscripts denoting the neurons they are associated with. The function $\sigma: \Re^{D} \to [0, 1]$ is the usual sigmoid function.

During training, the synaptic strengths and biases are updated iteratively using stochastic gradient descent. With $\{\boldsymbol{x}(n), o(n)\}$ representing the n^{th} input-output training sample out of a total of N (*i.e.* $1 \leq n \leq N$), the updating rule is,

$$\pi \leftarrow \pi + \frac{1}{2}\eta \times d(n) \times \nabla_{\pi}[y(n) - o(n)]^2. \quad (2)$$

In the above equation, η is the learning rate while π may be any of the network's parameters (*i.e.* $\boldsymbol{W}_{ij}^{H}$, $\boldsymbol{W}_{jk}^{O}$, $\mathbf{b}_i^{h}$ or b^{o}). When the training sample input is $\mathbf{x}(n)$, the corresponding output $y(n)$ of the neural network is determined in accordance with (1). The parameter $d(n)$ is the weight assigned to the n^{th} sample by ADABOOST.

As multiple neural networks are present in the ensemble, subscripts are applied to distinguish between the quantities pertaining to them. Thus $d_t(n)$ denotes the n^{th} sample's weight when training the t^{th} neural network ($1 \leq t \leq T$), while $y_t(n)$ is its corresponding output.

The sample weights begin with equal initial assignments,

$$d_1(n) = \frac{1}{N}, \quad (1 \leq n \leq N). \quad (3)$$

Hence the first neural network receives an equal amount of training for each sample.

In each subsequent network with index $t+1$, each sample weight $d_{t+1}(n)$ is determined based on the fraction error produced by the preceding (t^{th}) neural network with sample $\{\boldsymbol{x}(n), o(n)\}$. In order to do so, the algorithm maintains a threshold value θ. The neural network output for this sample is considered to be error-free when the absolute relative error lies within θ,

$$\frac{|y_t(n) - o(n)|}{o(n)} \leq \theta. \quad (4)$$

The new weights $d_{t+1}(n)$ are determined from the prior $d_t(n)$ in accordance with (5) below,

$$d_{t+1}(n) = \begin{cases} d_t(n)\varepsilon_t, & \dfrac{|y_t(n) - o(n)|}{o(n)} \leq \theta, \\ d_t(n), & \dfrac{|y_t(n) - o(n)|}{o(n)} > \theta. \end{cases} \quad (5)$$

The quantity ε_t in (5) is the error rate produced by the t^{th} network at the end of its training with $\{\boldsymbol{x}(n), o(n)\} \sim d_t(n)$. Using (4) as the criterion for a sample to be error-free, the set of erroneous samples is,

$$\mathcal{H}_t = \left\{ n \,\middle|\, \frac{|y_t(n) - o(n)|}{o(n)} > \theta \right\}. \quad (6)$$

Hence the network's error rate is given by,

$$\varepsilon_t = \sum_{n \in \mathcal{H}_t} d_t(n). \quad (7)$$

In order to ensure that the new weights constitute a probability distribution, they are normalized as follows,

$$d_{t+1}(n) = \frac{d_{t+1}(n)}{\sum_n d_{t+1}(n)}. \quad (8)$$

Following normalization, the weights add up to unity,

$$\sum_n d_{t+1}(n) = 1. \quad (9)$$

The overall training algorithm used by both ADABOOST.RT as well as ADABOOST$^+$ is outlined below.

1. Initialize d_1 using (3).
For each neural network $t = 1$ to T do
    2. Train network t using (1) and (2).
    3. Compute error rate ε_t using (6) and (7).
    4. Compute distribution d_{t+1} using (5).
    5. Normalize distribution d_{t+1} using (8).
    6. Add network t to ensemble.
End.

The algorithms ADABOOST.RT and ADABOOST$^+$ differ in how the ensemble output is determined. In ADABOOST.RT, the ensemble output $\hat{o}(n)$ is the weighted sum of all T neural networks, with the neural networks receiving weights proportional to the logarithm of their inverse error rates. Thus the weight δ_t applied the output of the t^{th} neural network ($1 \leq t \leq T$) is,

$$\delta_t = \frac{\log \varepsilon_t^{-1}}{\sum_t \log \varepsilon_t^{-1}}. \quad (10)$$

Accordingly, the ensemble output by ADABOOST.RT is,

$$\hat{o}(n) = \sum_t \delta_t y_t(n) \, . \quad (11)$$

However, in the proposed ADABOOST$^+$, the weights are determined to explicitly minimize the sum of the squared errors of all samples. Arranging the sample outputs and the ensemble outputs as $N \times 1$ column vectors $\mathbf{o}$ and $\hat{\mathbf{o}}$ respectively, the sum squared error can be expressed as,

$$E = \sum_n \left(o(n) - \hat{o}(n)\right)^2$$
$$= (\mathbf{o} - \hat{\mathbf{o}})^{\mathrm{T}}(\mathbf{o} - \hat{\mathbf{o}}). \quad (12)$$

Likewise, the outputs of each network can be organized as an $N \times 1$ vector $\mathbf{y}_t$. In a similar manner the network weights can be arranged as a $T \times 1$ vector $\boldsymbol{\delta}$. Defining the $N \times T$ output matrix $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_T]$, the output vector $\hat{\mathbf{o}}$ can be expressed as,

$$\hat{\mathbf{o}} = \mathbf{Y}\boldsymbol{\delta}. \quad (13)$$

Note that (13) is only a vector-matrix reformulation of (11).

It can be shown that the choice of $\boldsymbol{\delta}$ that minimizes the sum squared error E is given by,

$$\boldsymbol{\delta} = (\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{o}. \quad (14)$$

The above equation is the pseudo-inverse rule with the $T \times N$ matrix $\mathbf{Y}^{+} = (\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{-1}\mathbf{Y}^{\mathrm{T}}$ being the pseudo-inverse of $\mathbf{Y}$. Regularization can be incorporated for numerical stability of the matrix inversion, in which case $\boldsymbol{\delta}$ can be obtained as follows with σ being a small constant,

$$\boldsymbol{\delta} = (\mathbf{Y}^{\mathrm{T}}\mathbf{Y} + \sigma\mathbf{I})^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{o}. \quad (15)$$

In ADABOOST^{+}, (14) or (15) is applied to determine the network weights. The ensemble output is determined in accordance with (13).

IV. OUTAGE AND WEATHER DATA PREPARATION

Typical outage management systems in utilities record necessary information related to circuit outages including service area, circuit reference number, outage cause, outage weather, outage duration, number of customers affected, tripped equipment's, outage date and time, etc. The weather during outage is a set of weather conditions that utilities define based on their priorities and local weather characteristics. The most reliable weather information can always be obtained from the local weather stations, which record weather data including date, temperature, weather phenomenon, snow/ice, precipitation, pressure and wind on daily basis.

Existing literature [2] suggests that either gust or sustained wind can be used to study effects of outages with neither having any specific advantage over the other. Gust is recorded for days with high wind speeds and significant variation between peak and average speeds. In other words gust is an indicator of high wind speed as well as large fluctuations in wind speed or conditions which are likely to cause outages. In this paper, maximum daily wind gust measured on 5-second basis is used as the variable to study the wind effects because in our previous research we had found it to provide the best correlation to outages compared to other variables. However, for days with low wind speeds, which don't have gust recorded, 1-minute sustained speed is used. Additional investigation to identify other suitable wind related variables from the available data to include in the analysis will be pursued as part of future research.

Daily aggregate lightning stroke currents are calculated by summing the magnitudes of all the lightning strokes in kiloAmps (kA) including the first stroke and the flashes [35] within 500m around the feeders for each day of the study. Since our intent was to study combined effects of wind and lightning as well as that of wind alone, all the days including those that didn't have any recorded lightning were included. Also, the days of extreme weather conditions were excluded. Three such days for Lawrence, six days for Topeka, and eight days for Wichita were in this category, which were considered outliers and were removed from the data for analysis, which spanned a period of seven years from 2005 to 2011.

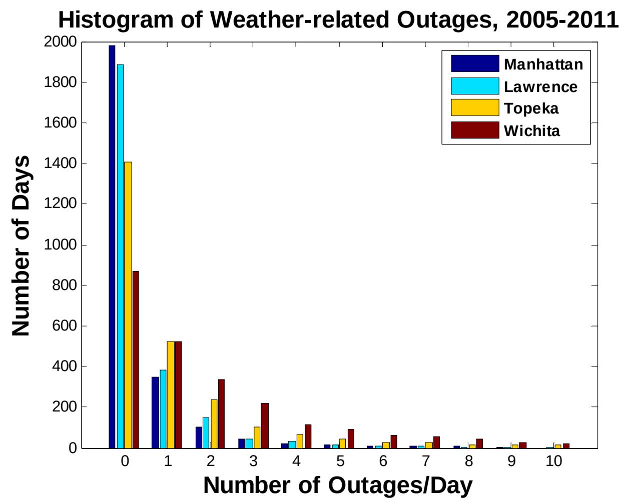

Fig.2. Outages caused by wind and lightning

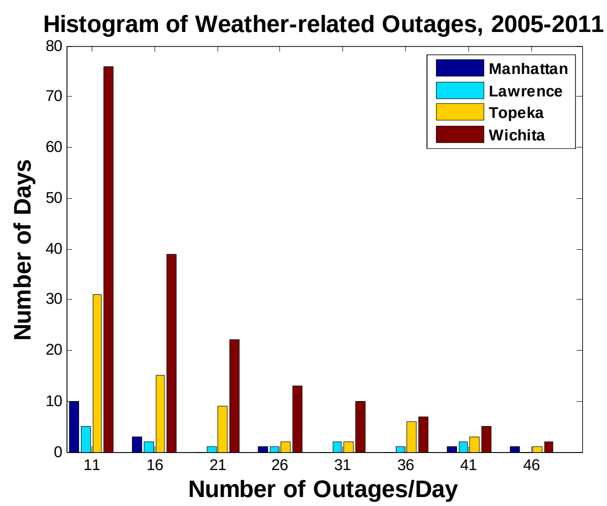

Fig.3. Outages in the higher range caused by wind and lightning (each bar graph represents outages with a bin size of five)

The daily maximum wind gust or 1-minute maximum sustained wind and aggregate lightning strokes were used as inputs for the model. In addition to these variables, trees around the feeders and vegetation management are important issues to consider because trees interact in a complex manner with wind to cause outages. However, since we aggregated all the feeders in the entire city in our analysis and each city was analyzed separately, tree density is not an important variable because it remains constant throughout the analysis for the specific city. Some spatial aggregation of feeders is necessary for smoothing of data to obtain meaningful statistical patterns [2]. If the tree density changes over time and utilities keep a good record of this change, this information could be included in the analysis. The utility that provided the data for analysis does vegetation management on a rolling basis over a period of four years. Specifically, trees in one-fourth of the city are trimmed each year, which allows completion of the entire city in four years. After that the cycle starts again. Therefore, if one looks at the entire city at any time, it remains roughly in the same state with respect to exposure to trees. If the feeders that had trees trimmed in a particular year were aggregated together to form four groups, it would be possible to include vegetation information as an input to the model. This approach might work for larger cities but might not work for smaller cities. This is an important issue, which requires additional data from the utility and further investigation.

The four cities included in this study are Manhattan (7 distribution substations with 176 miles of distribution feeders at 12.47 kV), Lawrence (7 distribution substations with 193 miles of distribution feeders at 12.47 kV), Topeka (22 distribution substations with 560 miles of distribution feeders mostly at 12.47 kV and a very small portion at 4 kV), and Wichita (42 distribution substations with 1165 miles of distribution feeders mostly at 12.47 kV and a very small portion at 4 kV).

Outages recorded in the database with lightning, trees, wind as cause, and equipment and unknown outages possibly caused by lightning and wind were included in the outage counts for the study. The weather at the time of all recorded lightning, equipment failure and unknown outages was manually examined to ensure that the lightning actually occurred on the feeder experiencing outage. Outages that were recorded as caused by lightning with no recorded lightning on the specific feeder were removed. On the other hand, equipment and unknown outages coinciding with recorded lightning on the specific feeders were included. In our previous studies [9-12] such detailed screening was not done. Therefore, the results shown in this paper based on these methods are different from those papers.

Fig. 2 and 3 shows histogram of outages (per day) in the study period for the four districts. Note that the scales of Fig 3 are different from that of Fig. 2. Also, in Fig. 3, each bar represents outages in a range covering five different values. For example, 11 on the x-axis represents outages from 11 to 15 and so on. The figure doesn't show a few additional days that had outages higher than 50. These figures show that there are a large number of days with zero or low number of recorded outages. Manhattan has the largest number of days with zero outages and Wichita has the smallest number of days with zero outages with Lawrence and Topeka falling in between in order. The trend reverses for one or higher number of outages. This is an outcome of the spatial aggregation of outages. Since Wichita has the largest service area, the probability of outages at each level greater than zero is higher for it than the cities with smaller service areas.

V. Experimental results

The data of the four cities were divided into training (2005 to 2009) and test (2010-2011) sets to evaluate performance of the AdaBoost models. The results obtained from these models were compared with prior research results [9-12]. In the previous research, linear, quadratic and exponential regression models have been considered [9, 10]. The model shown in (16) is considered for comparison because it showed the best performance out of them,

$$\hat{o} = \beta_1 Li + \beta_2 Wd + \beta_3 Wd \times Li + \beta_4 Wd^2 + \beta_5 Li^2 \quad (16)$$

Here, $\hat{o}$ is the estimated number of outages, Li is the accumulated lightning strokes in kA per day, and Wd is the maximum wind gust speed in miles per hour for the day.

Another model included for comparison is a neural network, which was applied to perform regression [11]. The output of this network is determined in accordance with (1), with $\hat{o} = y$. The training is performed as in (2) with $d(n) = 1$ for each sample n. In addition, a model based on mixture of experts (ME) [12] was considered for comparison.

To evaluate performance of the models, different criteria for comparison are used which are presented below:

(i) Mean Absolute Error (MAE) gives the average deviation of the estimated values from the observed values. This is given by

$$MAE = \frac{1}{N}\left(\sum_{i=1}^{N} |\hat{o}(i) - o(i)|\right) \quad (17)$$

(ii) Mean Square Error (MSE) between the observed and estimated outages defines the goodness of fit of the models. For N observations MSE is given by,

$$MSE = \frac{1}{N}\left(\sum_{i=1}^{N} \left(\hat{o}(i) - o(i)\right)^2\right) \quad (18)$$

(iii) Correlation Coefficient, R

$$R = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(o(i) - \bar{o})\left(\hat{o}(j) - \bar{\hat{o}}\right)}{\sqrt{\sum_{i=1}^{N}(o(i) - \bar{o})^2 \sum_{j=1}^{N}\left(\hat{o}(j) - \bar{\hat{o}}\right)^2}} \quad (19)$$

where, $\bar{o}$ is the average of observed outages and $\bar{\hat{o}}$ is the average of estimated outages.

Fig. 4 and 5 show the percentage MSE of AdaBoost.RT and AdaBoost$^+$ against the number of networks for the training data set of the four cities. The performance of AdaBoost$^+$ with Wichita data improved with regularization with σ = 0.01 in (15). In all other cases regularization was not used since it didn't change the results. The percentage MSE drops as the number of networks increase and it stabilizes after a certain number of networks. For example, for Wichita, the percentage MSE drops to 65% for AdaBoost.RT with four neural networks whereas for AdaBoost$^+$ the percentage MSE drops to 43% for the same number of neural networks; clearly illustrating the better performance of AdaBoost $^+$. Increasing the number of neural networks beyond that didn't change the results significantly. We believe that this is because even with only one neural network the results are reasonable and thus only some additional neural networks are required to improve the results and reach a stable point. This could also be dependent on the nature of the problem and initial selection of the neural network. Some problems could require a large number of neural networks to stabilize the error. For comparison results obtained with five neural networks are used.

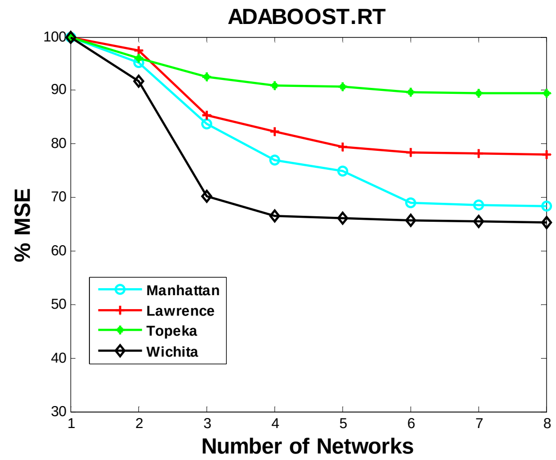

Fig.4. Performance of the ADABOOST.RT model

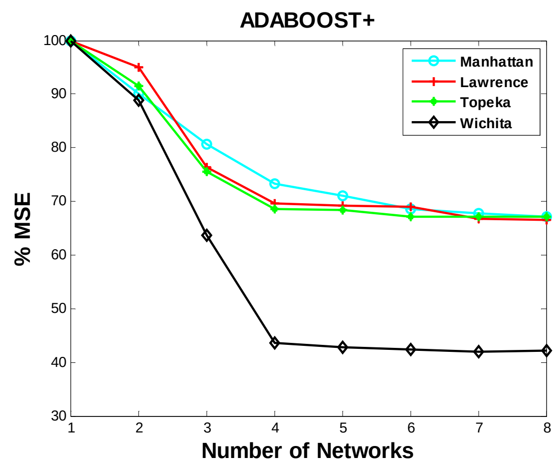

Fig.5. Performance of the ADABOOST$^+$ model

Performance measures of the models based on average absolute error (MAE) and mean squared error (MSE) are given in Tables from I to IV for the four cities. The R-square for regression between the estimated and the observed values of outages are not very large, but they are within a range similar to those presented previously in the literature for outage analysis. The nature of the data, which has significant natural randomness as well as errors introduced by people while collecting and recording observations make it very difficult to get very high correlation.

TABLE I
SUMMARY RESULTS FOR MANHATTAN

Manhattan						
	Training Data			**Test Data**		
	MAE	**MSE**	**R**	**MAE**	**MSE**	**R**
Regression	0.61	2.78	0.627	0.69	2.23	0.428
Neural Network	0.60	2.49	0.676	0.64	2.34	0.435
ME	0.60	2.38	0.676	0.64	2.32	0.434
ADABOOST.RT	0.38	1.92	0.781	0.56	2.07	0.622
ADABOOST$^+$	0.37	1.83	0.786	0.56	2.06	0.607

TABLE II
SUMMARY RESULTS FOR LAWRENCE

Lawrence						
	Training Data			**Test Data**		
	MAE	**MSE**	**R**	**MAE**	**MSE**	**R**
Regression	0.69	4.37	0.292	0.87	5.00	0.528
Neural Network	0.70	4.36	0.296	0.88	5.01	0.413
ME	0.69	4.20	0.282	0.83	4.34	0.531
ADABOOST.RT	0.39	3.59	0.515	0.53	3.66	0.666
ADABOOST$^+$	0.31	2.62	0.695	0.43	3.04	0.717

TABLE III
SUMMARY RESULTS FOR TOPEKA

Topeka						
	Training Data			**Test Data**		
	MAE	**MSE**	**R**	**MAE**	**MSE**	**R**
Regression	1.38	13.14	0.541	2.41	41.30	0.461
Neural Network	1.39	12.96	0.549	2.44	37.15	0.463
ME	1.37	11.85	0.549	2.43	36.94	0.469
ADABOOST.RT	0.91	11.79	0.643	1.92	32.07	0.703
ADABOOST$^+$	0.71	8.89	0.745	1.46	22.99	0.793

TABLE IV
SUMMARY RESULTS FOR WICHITA

Wichita						
	Training Data			**Test Data**		
	MAE	**MSE**	**R**	**MAE**	**MSE**	**R**
Regression	2.89	37.38	0.552	3.35	63.57	0.577
Neural Network	2.81	35.93	0.576	3.39	62.74	0.579
ME	2.79	34.68	0.576	3.33	62.47	0.589
ADABOOST.RT	1.67	24.39	0.769	2.44	48.02	0.796
ADABOOST$^+$	1.46	18.42	0.826	3.47	49.95	0.640

From the comparison of results from different methods for train and test data, it is found that ADABOOST+ performs relatively better over others followed by ADABOOST.RT. The only case where ADABOOST+ showed a slight reduction in performance compared to ADABOOST.RT based on these parameters is for the test data of Wichita. Although it appears to be a reduction is performance, it will be shown later that in fact ADABOOST+ performed better than ADABOOST.RT.

Fig. 6 and 7 show scatter plots with regression line of observed vs. estimated outages for training and test data of best regression model, neural network, ME, ADABOOST.RT, and ADABOOST+. In addition to the previously considered parameters, slope of the regression line between the observed and the estimated outages in an indicator of performance of the models. Higher slope would mean better performance with a slope of one giving the ideal performance. These graphs show clear improvement in performance of the ADABOOST+ model, which provides better slope than other models for all the training as well as test cases. ADABOOST+ performs distinctly better than the other models for outages in the lower range. However, all the models under predict outages in the higher range. This can be expected because the data in the higher range is sparse and thus the models are not able to fully learn the characteristics in the data in this range.

VI. CONCLUSIONS

In this paper, a new boosting algorithm ADABOOST+ is proposed to determine the effects of wind and lightning on outages in overhead distribution systems The models were trained and tested using the available historical data from 2005-2011 to verify their robustness. Comparison of the results show that the ADABOOST+ performs better than ADABOOST.RT and both the boosting models provide better estimates of the outages than the models based on standard regression, neural network, and mixture of experts.

The results are useful for utilities for system design and upgrades. Further research to improve ADABOOST+ will be focused on automating the choice of optimal value of threshold depending on the characteristics of the data set. Other machine learning models will be investigated to further improve the results, specifically for outages in the higher range. Other variables to represent wind in addition to gust speed and inclusion of vegetation related information into the models will be explored. The current research is suitable for end of the year evaluation based on past data. Ongoing research will focus on outage prediction in the future based on weather scenarios for the future.

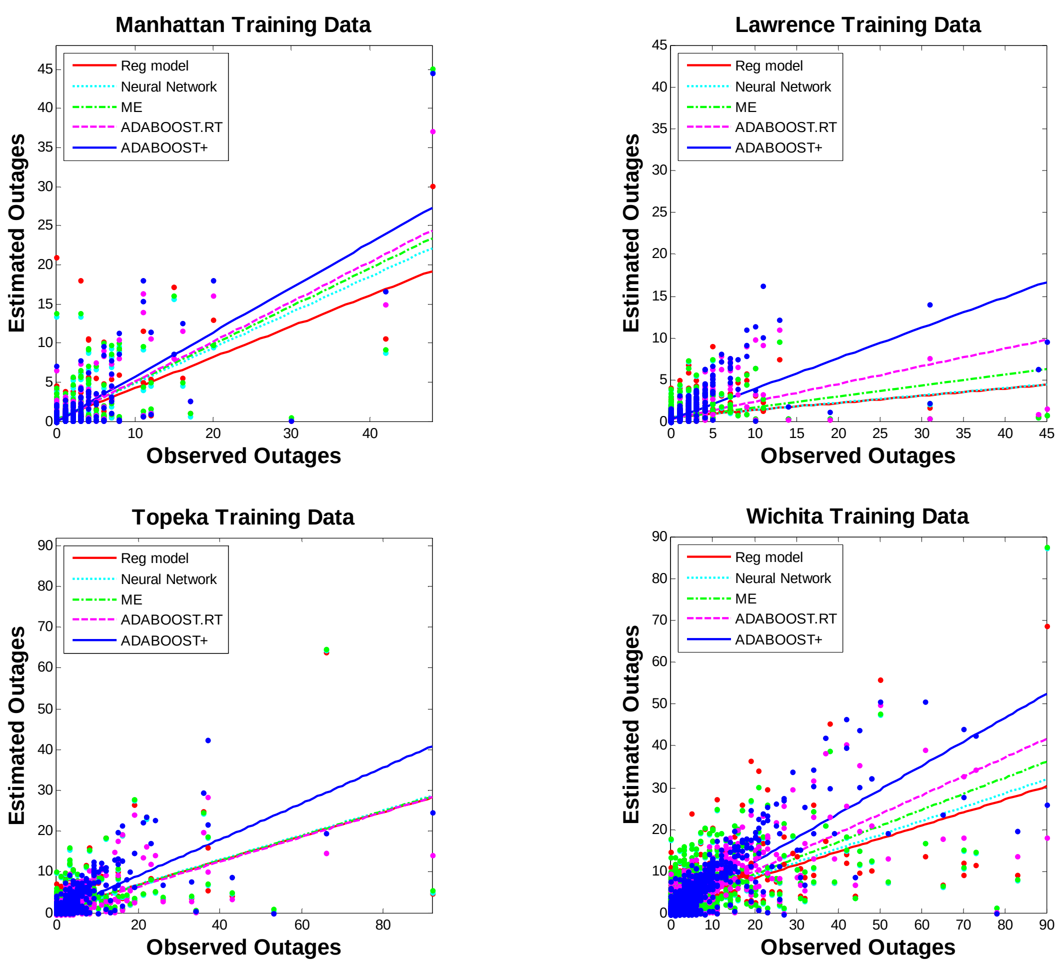

Fig.6. Scatter plot along with regression line of observed vs. estimated outages for 2005-2009 training data for different models (red-regression, cyan-neural network, green-mixture of experts, magenta-ADABOOST.RT and blue-ADABOOST+).

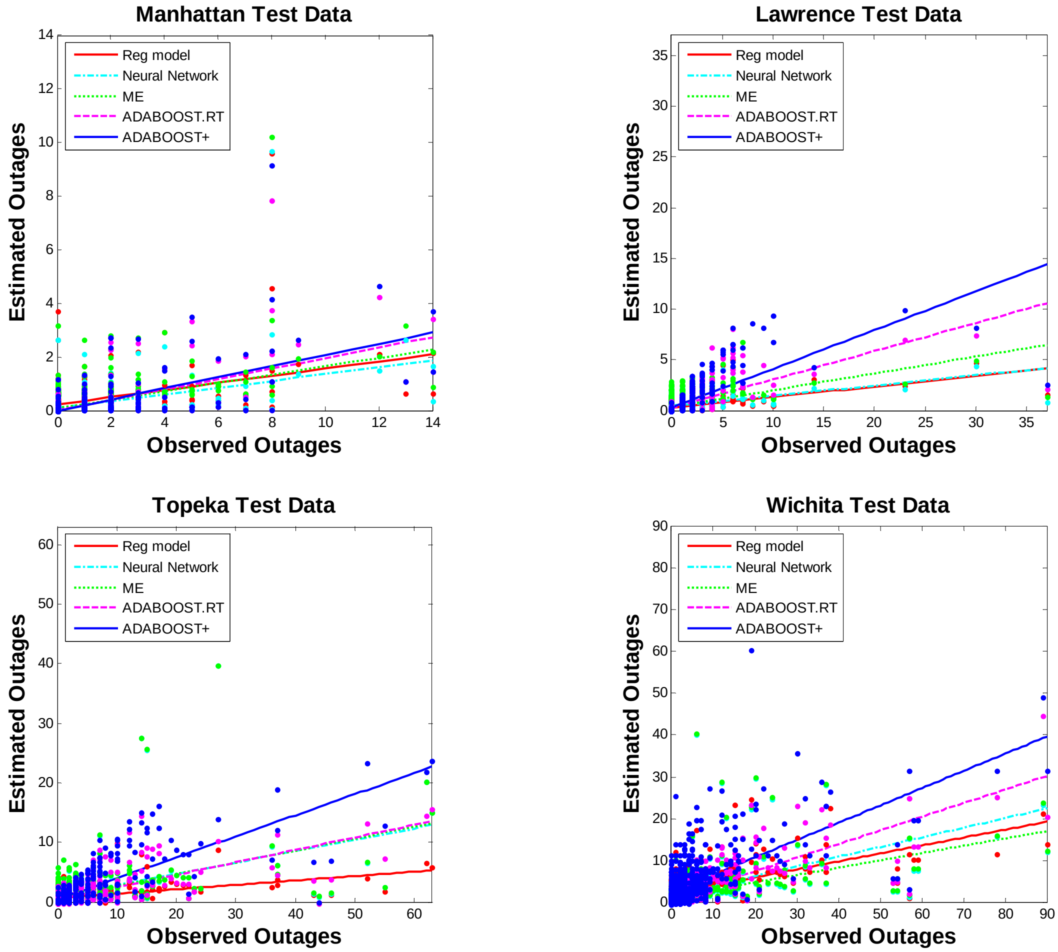

Fig.7. Scatter plot along with regression line of observed vs. estimated outages for 2010-2011 test data for different models (red-regression, cyan-neural network, green-mixture of experts, magenta-ADABOOST.RT and blue-ADABOOST^{+}).

ACKNOWLEDGMENT

We would like to thank David Rose and Rodney Robinson of Westar Energy, Topeka, KS for providing the outage and lightning data, Mary Knapp of State Climate Office at Kansas State University for providing the weather data. We would also like to thank NSF for providing financial support for the project. The views expressed in this paper are of the authors.

REFERENCES

[1] A. Pahwa, M. Hopkins, and, T. C. Gaunt, "Evaluation of Outages in Overhead Distribution Systems of South Africa and of Manhattan, Kansas, USA," Proceedings of International Conference on Power Systems Operation and Planning, Cape Town, South Africa, Jan. 2007.

[2] H. Caswell, V. Forte, J. Fraser, A. Pahwa, T. Short, M. Thatcher, and V. Verner, "Weather Normalization of Reliability Indices," IEEE Transaction on Power Delivery, vol. 26, 2011, pp. 1273-1279.

[3] D. Zhu, D. Cheng, R. P. Broadwater, and C. Scirbona, "Storm modeling for prediction of power distribution system outages," Electric power systems research, vol. 77, no. 8, June 2007, pp. 973-979.

[4] H. Liu, R. A. Davidson, D. V. Rosowsky, and J. R. Stedinger, "Negative binomial regression of electric power outages in hurricanes," Journal of infrastructure systems 11, no. 4, Dec, 2005, pp.258-267.

[5] H. Liu, R. A. Davidson, and T. V. Apanasovich., "Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms," *Reliability Engineering & System Safety*, vol. 93, no. 6, June 2008, pp. 897-912.

[6] H. Li, L. Treinish and J. R. M. Hosking., "A Statistical Model for Risk Management of Electric Outage Forecasts", *IBM Journal of Research and Development*, vol. 54, no. 3, May, 2010, pp. 8:1-8:11.

[7] Y. Zhou, A. Pahwa, and S. Yang, "Modeling weather-related failures of overhead distribution lines," *IEEE Transactions on Power Systems*, vol. 21, no. 4, Nov. 2006, pp. 1683-1690.

[8] A. Domijan Jr, R. K. Matavalam, A. Montenegro, W. S. Wilcox, Y. S. Joo, L. Delfron, J. R. Diaz, L. Davis, and J. D'agostini, "Effects of normal weather conditions on interruptions in distribution systems," *International journal of power & energy system*, vol. 25, no. 1, 2005, pp. 54-62.

[9] P. Kankanala, A. Pahwa, and S. Das, "Regression Models for Outages Due to Wind and Lightning on Overhead Distribution Feeders," *IEEE PES General Meeting* 2011, July 24-28, 2011, Detroit, Michigan, USA.

[10] P. Kankanala, A. Pahwa, and S. Das, "Exponential Regression Models for Wind and Lightning Caused Outages on Overhead Distribution Feeders," *North America Power Symposium* (NAPS) 2011, Aug 04-06, 2011, Boston, Massachusetts, USA.

[11] P. Kankanala, A. Pahwa, and S. Das, "Estimation of Overhead Distribution Outages Caused by Wind and Lightning Using an Artificial Neural Network," 9th *International Conference on Power System Operation and Planning*, Jan 16-19, 2012, Nairobi, Kenya.

[12] P. Kankanala, A. Pahwa, and S. Das, "Mean Field Annealing Based Committee Machines for Outage Estimation in Power Distribution System", the 2nd International Association of Science and Technology for Development (IASTED) International Conference on Power and Energy Systems and Applications (PESA), November 12 – 14, 2012, Las Vegas, USA.

[13] C. Rudin, D. Waltz, R. N. Anderson, A. Boulanger, A. Salleb-Aouissi, M. Chow, H. Dutta, P. Gross, B. Huang, S. Ierome, D. Isaac, A.

Kressner, R. J. Passonneau, A. Radeva, L. Wu, "Machine learning for the New York City power grid," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no.2, February 2012 pp. 328-345.
[14] I. Kamwa, S. R. Samantaray, and G. Joós, "On the Accuracy Versus Transparency Trade-Off of Data-Mining Models for Fast-Response PMU-Based Catastrophe Predictors," *IEEE Transaction on Smart Grid*, vol. 3, no. 1, 2012, pp. 152-161.
[15] M. Alamaniotis, A. Ikonomopoulos, and L. H. Tsoukalas, "Evolutionary Multiobjective Optimization of Kernel-Based Very-Short-Term Load Forecasting," *IEEE Transaction on Smart Grid*, vol. 27, no. 3, Aug 2012, pp.1477-1484.
[16] J. Guo, and P. B. Luh, "Improving Market Clearing Price Prediction by Using a Committee Machine of Neural Networks", *IEEE Transactions on Power System*, vol. 19, no. 4, November 2004, pp. 1867-1876.
[17] A. Chaouachi, R. M. Kamel, R. Andoulsi, K. Nagasaka, "Multiobjective Intelligent Energy Management for a Microgrid," *IEEE Transactions on Industrial Electronics*, vol. 60 , issue: 4, 2013, pp. 1688 – 1699.
[18] I. Kamwa, S. R. Samantaray, G. Joos, "Catastrophe Predictors From Ensemble Decision-Tree Learning of Wide-Area Severity Indices," *IEEE Transaction on Smart Grid*, vol. 1, no. 2, Sep 2010, pp. 144-158.
[19] S. Fan, L. Chen, W. Lee, "Short-Term Load Forecasting Using Comprehensive Combination Based on Multimeteorological Information," *IEEE Transactions on Industry Applications*, vol. 45, no. 4, 2009, pp: 1460-1466.
[20] J. W. Taylor and R. Buizza, "Neural Network Load Forecasting With Weather Ensemble Predictions," *IEEE Transaction on Smart Grid*, vol.17, no. 3, Aug 2002, pp. 626-632.
[21] M. De Felice, I. Enea, and X. Yao, "Short-Term Load Forecasting with Neural Network Ensembles: A Comparative Study," *IEEE Computational Intelligence Magazine*, vol. 6, issue 3, Aug 2011, pp. 47 – 56.
[22] D.S. Siroky, "Navigating Random Forests and related advances in algorithmic modeling," Statistical Reviews, Vol. 3, pp.147–163, 2009.
[23] M. Galar, A. Fern´andez, E. Barrenechea, H. Bustince, and F. Herrera, "A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches", *IEEE Transactions on Systems, Man, and Cybernetics*—Part C: Applications and Reviews, vol. 42, nO. 4, July 2012, pp. 463 – 484.
[24] T. M. Khoshgoftaar, J. V. Hulse, and A. Napolitano, "Comparing Boosting and Bagging Techniques with Noisy and Imbalanced Data", *IEEE Transactions on Systems, MAN, and Cybernetics*—Part A: Systems and Humans, vol. 41, no. 3, May 2011, pp. 552-568.
[25] Y. Freund, and R. Schapire, "Experiment with a new boosting algorithm", *Proc. of the 13th Int. Conf. on Machine Learning*, Bari, Italy, 1996, pp. 148-156.
[26] R. Schapire, "The strength of weak learnability", *Machine Learning*, vol. 5, no. 2, 1990, pp. 197-227.
[27] D. P. Solomatine, and D. L. Shrestha, "Experiments with AdaBoost.RT, an Improved Boosting Scheme for Regression", *Neural Computation*, vol. 18, no. 7, 2006, pp. 1678-1710.
[28] L. Breiman, "Bagging Predictors", *Machine Learning*, vol.24, 1996, pp.123–140.
[29] Nigel Duffy, David Helmbold, "Boosting Methods for Regression," *Machine Learning*, vol. 47, 2002, pp. 153–200.
[30] H. Drucker, "Boosting Using Neural Networks", in A. J. C. Sharkey (ed.), In Combining Artificial Neural Nets, London: Springer-Verlag, 1999, pp. 51-77.
[31] Y. Freund, and R. Schapire, "A decision-theoretic generalisation of on-line learning and an application of boosting", *Journal of Computer and System Sciences*, vol. 55, no. 1, 1997, pp. 119-139.
[32] V. Tresp, "Committee Machines", Published as a book chapter in: *Handbook for Neural Network Signal Processing*, YuHen Hu and Jenq-Neng Hwang (eds.), CRC Press, 2001.
[33] J. Friedman, T. Hastie, & R. Tibshirani, "Additive Logistic Regression: A Statictical View of Boosting", *The Annals of Statistics*, vol. 28, no. 2, 00, pp. 337-374.
[34] D. P. Solomatine, and D. L. Shrestha, "AdaBoost.RT: a Boosting Algorithm for Regression Problems", *Proc. of the International Joint Conference on Neural Networks*, IEEE, Budapest, Hungary, 2004, pp. 1163-1168.
[35] R. Orville, G. R. Huffines, W. R. Burrows, R. L. Holle and K. L. Cummins, "The North American Lightning Detection Network (NALDN)—First Results: 1998–2000," *Monthly Weather Review*, vol. 130, August 2002, pp. 2098 – 2109.

BIOGRAPHIES

Padmavathy Kankanala is pursuing the PhD degree in Electrical Engineering at Kansas State University. She received the M. S. in Electrical Engineering from Mississippi State University in 2009 and Bachelor's degree in Electrical and Electronics Engineering from Jawaharlal Nehru Technological University in 2006. Her research interests include power system real-time modeling and simulations, application of computational intelligence systems to power engineering, and multi-objective optimization.

Sanjoy Das received the Ph.D. in Electrical Engineering from Louisiana State University, Baton Rouge, in 1994. Between 1994 and 1997 he received postdoctoral training from the University of California, Berkeley and the Smith-Kettlewell Institute, San Francisco. Since 2001, he has been in the Electrical and Computer Engineering (ECE) Department at Kansas State University, where he is currently an associate professor. Das's research interests include multi-agent systems, machine learning, computational game theory, quantum computing, bioinformatics, and applications to power systems and smart grids.

Anil Pahwa received the B.E. (honors) degree in electrical engineering from BITS-Pilani, India, in 1975, the M.S. degree in electrical engineering from University of Maine, Orono, in 1979, and the Ph.D. degree in electrical engineering from Texas A&M University, College Station, in 1983. Since 1983, he has been with Kansas State University, Manhattan, where presently he is Logan-Fetterhoof Chair Professor in the ECE Department. His research interests include distribution automation, and intelligent computational methods for distribution system applications.