MACHINE LEARNING METHODS FOR THE ESTIMATION OF WEATHER AND ANIMAL-RELATED POWER OUTAGES ON OVERHEAD DISTRIBUTION FEEDERS

by

PADMAVATHY KANKANALA

B.Tech., Jawaharlal Nehru Technological University, 2006
M.S., Mississippi State University, 2009

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSPHY

Department of Electrical and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Abstract

Because a majority of day-to-day activities rely on electricity, it plays an important role in daily life. In this digital world, most of the people's life depends on electricity. Without electricity, the flip of a switch would no longer produce instant light, television or refrigerators would be nonexistent, and hundreds of conveniences often taken for granted would be impossible. Electricity has become a basic necessity, and so any interruption in service due to disturbances in power lines causes a great inconvenience to customers.

Customers and utility commissions expect a high level of reliability. Power distribution systems are geographically dispersed and exposure to environment makes them highly vulnerable part of power systems with respect to failures and interruption of service to customers. Following the restructuring and increased competition in the electric utility industry, distribution system reliability has acquired larger significance. Better understanding of causes and consequences of distribution interruptions is helpful in maintaining distribution systems, designing reliable systems, installing protection devices, and environmental issues. Various events, such as equipment failure, animal activity, tree fall, wind, and lightning, can negatively affect power distribution systems. Weather is one of the primary causes affecting distribution system reliability. Unfortunately, as weather-related outages are highly random, predicting their occurrence is an arduous task. To study the impact of weather on overhead distribution system several models, such as linear and exponential regression models, neural network model, and ensemble methods are presented in this dissertation. The models were extended to study the impact of animal activity on outages in overhead distribution system.

Outage, lightning, and weather data for four different cities in Kansas of various sizes from 2005 to 2011 were provided by Westar Energy, Topeka, and state climate office at Kansas State University weather services. Models developed are applied to estimate daily outages. Performance tests shows that regression and neural network models are able to estimate outages well but failed to estimate well in lower and upper range of observed values. The introduction of committee machines inspired by the 'divide & conquer" principle overcomes this problem. Simulation results shows that mixture of experts model is more effective followed by AdaBoost model in estimating daily outages. Similar results on performance of these models were found for animal-caused outages.

MACHINE LEARNING METHODS FOR THE ESTIMATION OF WEATHER AND ANIMAL-RELATED POWER OUTAGES ON OVERHEAD DISTRIBUTION FEEDERS

by

PADMAVATHY KANKANALA

B.Tech., Jawaharlal Nehru Technological University, 2006
M.S., Mississippi State University, 2009

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Approved by:

Co-Major Professor
Anil Pahwa

Approved by:

Co-Major Professor
Sanjoy Das

Abstract

Because a majority of day-to-day activities rely on electricity, it plays an important role in daily life. In this digital world, most of the people's life depends on electricity. Without electricity, the flip of a switch would no longer produce instant light, television or refrigerators would be nonexistent, and hundreds of conveniences often taken for granted would be impossible. Electricity has become a basic necessity, and so any interruption in service due to disturbances in power lines causes a great inconvenience to customers.

Customers and utility commissions expect a high level of reliability. Power distribution systems are geographically dispersed and exposure to environment makes them highly vulnerable part of power systems with respect to failures and interruption of service to customers. Following the restructuring and increased competition in the electric utility industry, distribution system reliability has acquired larger significance. Better understanding of causes and consequences of distribution interruptions is helpful in maintaining distribution systems, designing reliable systems, installing protection devices, and environmental issues. Various events, such as equipment failure, animal activity, tree fall, wind, and lightning, can negatively affect power distribution systems. Weather is one of the primary causes affecting distribution system reliability. Unfortunately, as weather-related outages are highly random, predicting their occurrence is an arduous task. To study the impact of weather on overhead distribution system several models, such as linear and exponential regression models, neural network model, and ensemble methods are presented in this dissertation. The models were extended to study the impact of animal activity on outages in overhead distribution system.

Outage, lightning, and weather data for four different cities in Kansas of various sizes from 2005 to 2011 were provided by Westar Energy, Topeka, and state climate office at Kansas State University weather services. Models developed are applied to estimate daily outages. Performance tests shows that regression and neural network models are able to estimate outages well but failed to estimate well in lower and upper range of observed values. The introduction of committee machines inspired by the 'divide & conquer" principle overcomes this problem. Simulation results shows that mixture of experts model is more effective followed by AdaBoost model in estimating daily outages. Similar results on performance of these models were found for animal-caused outages.

Table of Contents

List of Figures

List of Tables

List of Symbols

Li	Lightning Stroke Current in KA
Wd	Wind Gust Speed
O	Observed Outages
$\widehat{O}$	Estimated Outages
X	Input Variable
Y	Actual Output Variable
$\widehat{Y}$	Model Output Variable
$\hat{y}_k$	k^{th} Learner Output
β	Regression Coefficients
ε	Error due to Variability
T	Temperature
T_c	Critical Temperature
T_f	Final Temperature
d	Weight Distribution to Input Data for AdaBoost Model
W_{ji}	Weights j^{th} Hidden Neuron to i^{th} Input Neuron
W_{kj}	Weights k^{th} Output Neuron to j^{th} Input Neuron
e(.)	Error Function
w	Weight to Input Data for AME Model
σ	Sigmoidal Function
δ	Error Rate
E(.)	Sum Squared Error
γ	Temperature Cooling Parameter
K	Number of Base Learners
N	Number of data points

List of Abbreviations

NN - Neural Network

MLP - Multi-layer Perceptron

CM - Committee Machines

EL - Ensemble Learning

ME - Mixture of Experts

MFA - Mean Field Annealing

AME - Annealed Mixture of Experts

WNN - Wavelet-based NN Model

AIS - Artificial Immune System

SVM - Support Vector Machines

GP - Gaussian Process

GMM - Gaussian Mixture Model

HMM - Hidden Markov Model

EM - Expectation Maximization

DA - Deterministic Annealing

OMS - Outage Management System

MAE - Mean Absolute Error

MSE - Mean Square Error

RMSE - Root Mean Square Error

MAPE - Mean Absolute Percentage Error

R - Correlation Coefficient

S - Slope

Acknowledgements

I would like to give my thanks to all people who have helped me in the past four years while I pursued my Ph.D study and research. First of all, I would like to express my sincere thanks to my advisors, Dr. Anil Pahwa and Dr. Sanjoy Das, for giving me an opportunity to pursue my Ph.D degree at Kansas State University, for their inspiration, patience, moral support, and encouragement on my research work. It was their guidance and encouragement that helped me conquer the difficulties and led me to the success in my Ph.D study. I would like to thank Dr. Anil Pahwa, Dr. Sanjoy Das, Dr. Noel Schulz, and Dr. Caragea Doina, for serving on my committee and for the valuable guidance they have given to me. I would also like to thanks Dr. Richard Marston for agreeing to serve as outside chair for my defense at the last moment as Dr. Kevin Blake was on sabbatical.

In the past four years, Dr. Anil Pahwa and Dr. Sanjoy Das gave me much support in both my research and writing publications. I would like to give my appreciation to them for the inspiration they provided me and for the academic discussions and research group meetings we had that truly motivated my research. I also would like to thank Dr. Schulz for being a constant support through my Master's studies at Mississippi State University and Ph.D studies and for sharing her vast experience in research while encouraging me to explore research ideas.

Also, I would like to thank Westar Energy in Topeka, Kansas for providing the outage data and Mary Knapp of Kansas State University weather services for providing the weather information. I want to thank the National Science Foundation for providing financial support through grant no ECCS-0926020.

I would like express my gratitude to my parents, Mr. Shyam Sunder Reddy Kankanala and Mrs. Vijayalaxmi Kankanala, for bringing me this far through my education over the years. Without them this would not be possible. Finally, I thank my brother, Mr. Rajender Reddy Kankanala, my In-Laws, Mr. Ganapathi Reddy Kommera and Mrs. Padmaja Devi Kommera, and all my family and friends in USA and in India for their emotional support, understanding, encouragement, trust, and love.

Dedication

I dedicate this work to my husband, Venkat Reddy Kommera, for his love and support.

Chapter 1 - Introduction

Because a majority of day-to-day activities rely on electricity, it plays an important role in daily life. Without electricity, the flip of a switch would no longer produce instant light, television or refrigerators would be nonexistent, and hundreds of conveniences often taken for granted would be impossible. Households, businesses, and industry depend on electricity. In fact, electricity has become a basic necessity, and so any interruptions in service due to disturbances in power lines causes a great inconvenience to customers. Significant economic loss and business interruptions have been reported in the past.

Customers expect a high level of reliability, and electric utility companies have the responsibility to supply the interrupted electricity to the customers [1]. Ensuring electric power system reliability is a particularly challenging task for electric companies because maintaining a high level of reliability requires constant commitment. Utility companies must report system performance reliability annually to the utility regulatory commission [1]. The reliability assessment is concerned with system performance at the customer end, which would be considered as the system load points. Considerable interest has been shown in the development of reliability modeling and evaluating techniques for power distribution system [2,3].

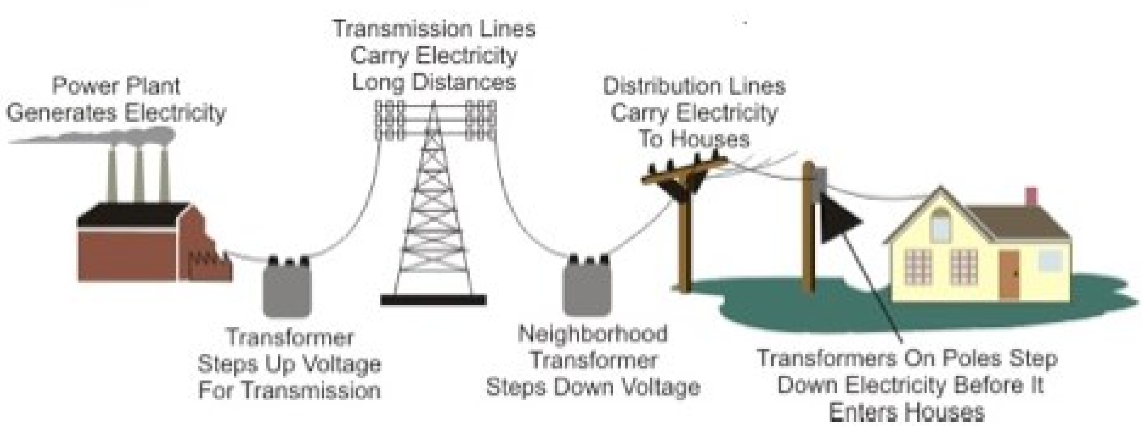


Figure 1.1 Components of the Electric Power System [1]

Electric power is generated at generation stations and transferred through high voltage transmission lines to substations which reduce voltage levels for distribution to end-use customers/demand point, as shown in Figure 1.1. The network enabling electric power to be sent to customers is either overhead or underground. Both the designs have advantages as well as disadvantages. Overhead feeders are less expensive and easy to install and maintain. In addition, individual faults on overhead circuits can be repaired more quickly than on underground circuits. The overhead feeders are vulnerable to severe weather events such as hurricanes, wind, rain, lightning, ice, freezing rain, and snow. Because of this vulnerability, underground circuit are preferred in some situations. However, underground systems are not immune to the effects of weather. Repairs for underground outages are typically more complex, more expensive, and result in longer restoration times. Since the distribution system in USA are predominantly overhead, reliability indices of distribution systems are more sensitive to failure rates of overhead feeders.

Power outages can result from seasonal storms which often combine strong winds, rain, snow, or ice. Extremely severe weather events typically cause greatest damage to electric power transmission and distribution infrastructure resulting in damage from trees or branches falling on electricity lines. While data on storm-related power outages exist, they are not generally considered to be complete or well characterized in relation to causes of outage events. Current data estimate that 90% of customer outage-minutes are due to events which affect local distribution systems. The remaining 10% outage-minutes stem from generation and transmission problems, which can cause wider-scale outages affecting more customers. [1]. These exposures create complications for the power distribution industry by causing interruptions in distribution systems. In order to accurately analyze component reliability data and predict system reliabilities, better models must be found and utilized to study outages caused by environmental factors on overhead distribution lines.

1.1 Overhead Distribution System Reliability Assessment

Because of restructuring and increased competition in the electric utility industry, distribution system reliability has acquired larger significance. There is an increase in demand of high reliability both from the digital age customers and the utility commissions. In many states, regulatory bodies require utility companies to annually disclose reliability related performance, and some states impose penalties and/or rewards based on performance [3]. Therefore, utility companies strive to maintain a high level of reliability.

In the last 20 years, significant research has been conducted for evaluating and improving distribution system reliability [4-20]. Reliability, in general, is defined as the probability that a component or system will perform a required function for a given period of time when used under stated operating conditions [6]. Reliability of electric power distribution systems is defined as the ability to deliver uninterrupted service to customers [9]. Customer interruptions are divided into two major categories: sustained interruptions and momentary interruptions based on the length of interruption. According to IEEE standard [14, 16], five minutes is the cut-off between momentary and sustained interruption. Utilities typically report annual performance of the distribution system using the most commonly used reliability indices: SAIFI (System Average Interruption Frequency Index), SAIDI (System Average Interruption Duration Index), and MAIFI (Momentary Average Interruption Frequency Index) [14-17]. In addition to these three indices, other indices, such as CAIFI, CAIDI, and ASAI, are also used [18]. Since distribution systems are radial in nature and located in population dense areas, perfect reliability is nearly impossible to provide [16]; however, proper design, maintenance, upgrades, and monitoring of the system contribute to a very high level of reliability [16]. For over more than 30 years, IEEE has periodically published a bibliography on power system reliability evaluation. For the reliability assessment of distribution systems, researchers primarily use three approaches: historical, predictive, and feature-based. Gui, Pahwa and Das provide a detailed review of these approaches and list relevant papers published over the years [21-23].

1.2 Causes of Outages in Distribution System

Various factors causing distribution system outages can be broadly classified into three categories: (*i*) Intrinsic factors, such as equipment age, manufacturing defects, conductor size; (*ii*) Environmental factors, such as trees, animals, wind, lightning; and (*iii*) Human error factors, such as vehicular accidents and accidents by utility crews [27-31]. Ten categories for general interruption causes are suggested for comparison in benchmark studies. These categories are intentionally broad and they make possible more precise benchmark comparisons between different distribution utilities. The ten categories as suggested by an IEEE Task Force are [32]:

- equipment;
- lightning;

- planned;
- power supply;
- public;
- vegetation;
- weather (other than lightning);
- wildlife;
- unknown;
- other.

The recommended categories do not prevent a utility from collecting additional detailed data, and that is indeed encouraged; however, the collected data should be classified into one of the recommended ten categories. Figure 1.2 shows a typical pie chart of outage causes in Manhattan, Kansas recorded by Westar Energy from 2005 to 2011. The chart illustrates that approximately 31% of outages were caused by environmental. Among these causes, weather was a primary cause on overhead distribution systems. Environmental factors influence system performance in a complex way and, without knowing this influence, correct evaluation of the distribution system reliability cannot be made.

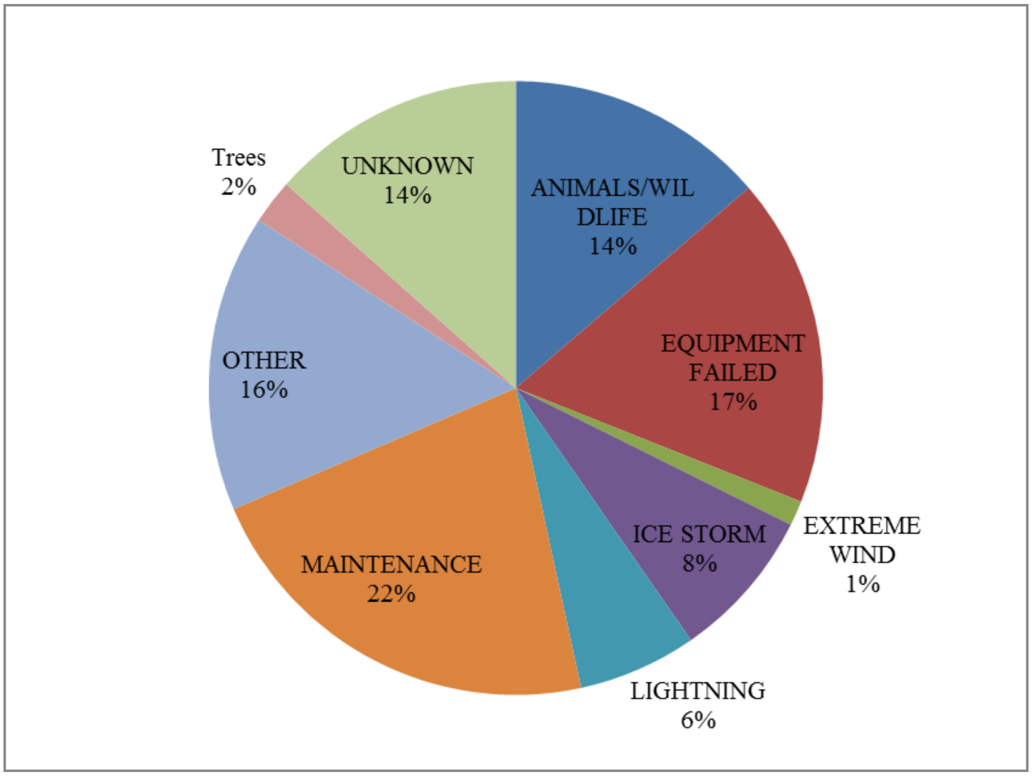


Figure 1.2 Percentage of Outages by Different Causes in Manhattan between 2005 and 2011

Figure 1.3 shows the impact of different weather factors on overhead transmission and distribution systems. 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. While evaluating the system performance, utilities usually separate outages caused by extreme weather conditions from those caused by severe weather conditions.

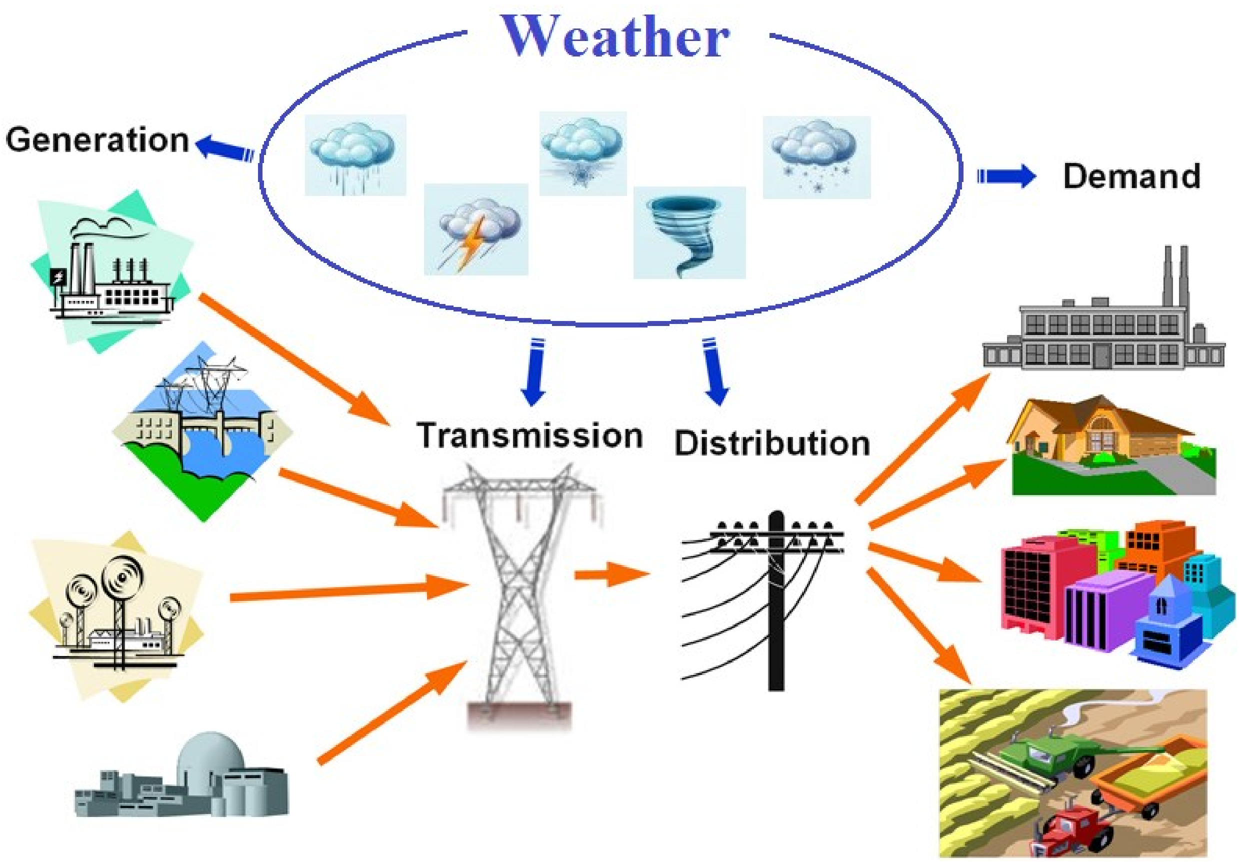


Figure 1.3 Weather Impacts on Power System

Wind induces conductor swinging, galloping, and aeolian vibration which are reliability concerns [18]. On the other hand, wind blowing over trees and poles causes branches to drop or poles to fall causing shorts or breaks in the overhead conductors, resulting in outages. In many areas of the United States, lightning is a major source of distribution feeder faults [34-36], and it affects distribution system reliability by direct or indirect strokes in which flashovers and high voltages are two major products. Direct strokes on overhead feeders bring big immediate damages on conductors and are not easy to protect, whereas indirect strokes cause short circuits or open circuits by affecting trees or poles surrounding the lines.

1.3 Related Work

Over the years, various models have been proposed to study effects of different environmental factors on outages with varying levels of success. As part of prior research support provided by the National Science Foundation (NSF), "Investigating the Influence of

Environmental Factors on Reliability of Distribution Systems," Pahwa and his team have developed models to study the impact of environmental factors on system reliability. Zhou, Pahwa and Yang researched weather impact on overhead distribution line failure rates [38]. A linear regression model, a Poisson model, and a Bayesian model were constructed for the prediction of the monthly average number of failures based on monthly weather conditions. The methods used wind gust speeds and lighting stroke currents as inputs for Manhattan from 1998 to 2003 and attempted to correlate each weather state and failure level. The lightning was regarded as a system-wide measurement and included only failures that resulted in outages to customers. Simulations with historical data showed that the Bayesian model provided a good way to model failure rates of overhead distribution lines. Sahai and Pahwa performed research on weather impact on animal-related outages in overhead distribution systems [39]. Examination of historical data showed that animal-related outages primarily occur on fair weather days. Also, behavioral patterns of animal activity in different months and their impact on animal-related outages were discovered [39]. A Bayesian model was constructed to predict animal-related outages in overhead distribution systems given two factors, month type and the number of fair days per week [39]. This Bayesian model was applied to data of five cities in Kansas from 1998 to 2002. Weekly and monthly estimations were obtained out and confidence intervals for the estimations were found. Gui, Pahwa and Das refined the Bayesian model and investigated other models to study the impact of animal activity on outages in distribution systems [23-25]. A Poison model, NN model, wavelet based NN model, and a refined Bayesian model were constructed for prediction of animal-related outages in overhead distribution systems. These methods considered the month type and the number of fair days per week as inputs for Manhattan, Lawrence, Topeka, and Wichita from 1998 to 2007. Weekly predictions were made, and experimental results showed that the WNN model performed better in predicting animal-related outages compared to other models.

Recently published papers related to this research are summarized next. An exponential model as a function of time for forecasting cumulative outages during different extreme weather events has been proposed in [41]. 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 predicting the number of outages due to hurricanes and ice storms have been developed [42, 43]. 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. However, these methods have limitations, such as evolving power system inventory with time and presence of huge matrix of spatial correlation, making it computationally challenging. Poisson regression and Bayesian hierarchical network for risk management of power outages caused by extreme weather conditions is investigated in [44]. 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. In [45], Poisson regression is used to study the significance of weather variables on outages using outage data from substations under severe weather conditions within 10 miles of National Weather Service sites.

1.4 Challenges and Motivation

In this work, outages on distribution feeders caused by severe weather conditions are studied. Among various weather factors, literature has shown that wind and lightning are primary causes of outages in the distribution system [28, 29]. They not only cause shorts or breaks directly on overhead lines, but they also disrupt trees which interrupt pathways delivering electricity. Previous studies show that the highest correlation between system interruptions and weather variables occur for wind, followed by ground flashes, with little or no correlation for temperature and rainfall.

To incorporate weather-caused outages in the reliability assessment, effective models are needed. To develop the models, understanding of how weather conditions effect power distribution interruptions must be gained. Weather factors influence system performance in a complex way and, without knowing this influence, correct evaluation of reliability performance of the distribution system is impossible. The complex interaction between different weather factors and their impact on the distribution system makes modeling of these processes very difficult and challenging. Modeling the effects of various weather factors on distribution system reliability helps utilities identify systems with high outages and provide a comparative analysis of actual performance with expected performance. Models are needed which are able to explain reliability trends due to weather conditions and aid in developing indicators to anticipate interruptions.

Unfortunately, because weather-related outages are highly random, predicting their occurrence is quite an arduous task. Additionally, various physical and data collection issues could impede performance of the models. Specifically, some possible reasons are

1. Inconsistencies and errors in recording outages by the utilities.
2. Although reliable weather observations exist from weather stations, they are an imperfect representation of weather conditions at the specific points.
3. Weather stations may not be in optimal locations. For example, nearby buildings and trees can act as shields, causing inaccurate wind measurement.
4. The distance between outage location and the airport where weather parameters are measured, can be large.

In initial investigation, linear, quadratic and exponential regression models, multilayered neural network model are considered to study the effects of wind and lightning on power outages on overhead distribution feeders [47-49]. Although these methods show acceptable performance, they are limited in their ability in estimating outages in lower and upper range of observed values. This can be due to unavailability of complete information from the historical data. The possible solution to overcome this problem is to utilize machine learning algorithms. Ensemble learning or committee machines is the process in which multiple models are strategically generated and individual solutions/outputs are combined to obtain a final solution. Committee machines are primarily used to improve the performance of a model.

The principle of combining predictions has been of interest to several fields over many years. Ensemble learning refers to procedures employed to train multiple learning machines and combine their outputs, treating them as a “committee” of decision makers. The principle is that the committee decision, with individual predictions combined appropriately, should have better overall accuracy, on average, than any individual committee member. Numerous empirical and theoretical studies have demonstrated that ensemble models often attain higher accuracy than single models. Members of the ensemble may predict real-valued numbers, class labels, posterior probabilities, rankings, clustering, or any other quantity. Therefore, their decisions can be combined by many methods, including averaging, voting, and probabilistic approaches. The majority of ensemble learning methods are generic and applicable for broad classes of model types and learning tasks.

The financial forecasting community has analyzed model combination in the context of stock portfolios for several decades. The contribution of the Machine Learning (ML) community emerged in the 1990s in automatic construction (from data) of the models and the method to combine them. While the majority of ML literature on this topic is from 1990 onward, the principle has been explored briefly for historical accounts by several independent authors since the 1960s [50].

The initial motivation for selecting the Committee Machine (CM) approach was to design a system in which individual learners are responsible for modeling different regions in input space. This modularity leads to greater modeling capability and a potentially meaningful and interpretable segmentation of the map.

Primary reasons for using CM are: (1) better performance, (2) statistical because the algorithm searches a space of hypothesis too large for the amount of available training data, (3) computational because the algorithm cannot guarantee finding the best hypothesis within the hypothesis space, (4) representational because the hypothesis space does not contain any hypotheses that are good approximations to true function, and (5) computational efficiency [50].

1.4 Organization of the Dissertation

The rest of this dissertation is organized as follows:

- Chapter 2 discusses causes of outages in the distribution feeders. Characteristics of outage data and weather data for four cities in Kansas from 2005 to 2011 are presented. This chapter also discusses data pre-processing.
- Chapter 3 presents modeling of six regression models and application of these models to the given data. Analysis of experimental results is discussed. Even though the models estimate the outages, outages are under-estimated in the higher range and over-estimated in the lower range. Specific discussions and results related can be found in our papers [47, 48].
- Chapter 4 presents the neural network modeling and its application to the data. The NN model can more accurately approximate high complexity equations. Simulation results

shows that NN models perform better than regression models but still have under-fitting and over-fitting issue [49].

- Chapter 5 presents the general theory of modeling based on ensemble learning. The modeling and parameter learning algorithms of two different AdaBoost models and a ME model are discussed [51, 52].
- Chapter 6 presents the application of ensemble models to the given data in order to estimate weather-related outages.
- Chapter 7 studies the effect of separating the data into lightning and non-lightning days. All the models are applied separately to these datasets. The results are compared with those obtained previously.
- Chapter 8 presents the modeling of simple NN model and ensemble models for the estimation of animal-related outages. This is an extension of the study presented in [23-25].
- Chapter 9 summarizes the entire dissertation. Concluding remarks and recommendations for future work are presented.

1.5 Performance Measure of the Models

To evaluate model performance, different criteria for comparison are used.

(i) Mean Absolute Error (MAE)

$$MAE = \frac{1}{N}\left(\sum_{i=1}^{N} |\hat{Y}(i) - Y(i)|\right) \tag{1.1}$$

(ii) Mean Square Error (MSE)

$$MSE = \frac{1}{N}\left(\sum_{i=1}^{N} \left(\hat{Y}(i) - Y(i)\right)^2\right) \tag{1.2}$$

(iii) Correlation Coefficient, R

$$R = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(Y(i) - \bar{Y})\left(\hat{Y}(j) - \bar{\hat{Y}}\right)}{\sqrt{\sum_{i=1}^{N}(Y(i) - \bar{Y})^2 \sum_{j=1}^{N}\left(\hat{Y}(j) - \bar{\hat{Y}}\right)^2}} \tag{1.3}$$

(iv) Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{N}\left(\sum_{i=1}^{N} |\hat{Y}(i) - Y(i)|\right) \times 100 \tag{1.4}$$

(v) Root Mean Square Error (RMSE)

$$E = w_i \sum_{i=1}^{N} RMSE = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{N}\left(\hat{Y}(i) - Y(i)\right)^2\right)} \tag{1.5}$$

(vi) Slope

$$S = {}^{Y'\hat{Y}}/_{Y'Y} \tag{1.6}$$

where, Y is the desired output, $\hat{Y}$ is the model output, $\bar{Y}$ is the average of desired output, and $\bar{\hat{Y}}$ is the average of model output.

Chapter 2 - Historical Data

The data period considered in this study ranged from January 1, 2005 to December 31, 2011. Historical lightning and outage data considered in this work were provided by Westar Energy, Topeka, Kansas, and wind data were provided by the State Climate Office at Kansas State University, Manhattan, Kansas. The four cities included in this study are Manhattan (seven distribution substations with 176 miles of distribution feeders at 12.47 kV), Lawrence (seven 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).

The utilities use geographical information system (GIS) to track their facilities in the distribution system, thus allowing for easier obtainment of information and maintenance of the database on system exposure to external factors. Currently, Westar Energy possesses GIS maps of the distribution system and maintains a log of outages caused by various factors for each district within their service territory. These system data include outages and causes in selected districts of Westar Energy, detailed layout of feeders with lengths, location and number of distribution transformers.

2.1 Characteristics of Weather Data

As mentioned in Chapter 1, wind and lightning are weather factors which strongly impact overhead feeders. Weather during an outage includes a set of weather conditions that utilities define based on priorities and local weather characteristics. The most reliable weather information can be obtained from local weather stations which record daily weather data including date, temperature, weather phenomenon, snow/ice, precipitation, pressure and wind on daily basis.

General weather data were provided by the state climate office, Kansas State University weather service. Daily weather data recorded on daily base include:

- Date
- Temperature
- Dew point
- Weather Phenomenon
- Snow/Ice
- Precipitation
- Pressure
- Maximum wind speed
- Wind Gust

Figure 2.2 shows the screenshot of weather recordings from the weather station for Manhattan, Kansas. Each column represents the elements recording in a day, as summarized in Table 2.1.

Table 2.1 Weather Elements Description

Code	**Description**
STN	Station number
WBAN	Weather bureau air force nave number
YearMODA	Year-month-day
TEMP	Mean temperature for the day in degrees Fahrenheit to tenths
COUNT	Number of observations used in calculating mean temperature
DEWP	Mean dew point for the day in degrees Fahrenheit to tenths
COUNT	Number of observations used in calculating mean dew point
SLP	Mean sea level pressure for the day in millibars to tenths
STP	Mean station pressure for the day in millibars to tenths

VISIB	Mean visibility for the day in miles to tenths
WDSP	Mean wind speed for the day in knots to tenths
MXSPD	Maximum sustained wind speed reported for the day in knots to tenths
GUST	Maximum wind gust reported for the day in knots to tenths
MAX	Maximum temperature reported during the day in Fahrenheit to tenths
MIN	Minimum temperature reported during the day in Fahrenheit to tenths
PRCP	Total precipitation (rain and/or melted snow) reported during the day in inches and hundredths

1 knot = 1.151 miles per hour approximately. Maximum wind speed and wind gust are converted to miles per hour.

Westar Energy provided lightning stroke data upon request, which details every stroke in the service territory from 2005 to 2011. Figure 2.2 shows the screenshot of lightning recordings from Westar Energy for Manhattan, Kansas. Each column represent the

- Date of the stroke event
- Time of the stroke
- Latitude (in decimal degrees)
- Longitude (in decimal degrees)
- Peak Current (in kiloAmps)
- Equipment name
- Length of asset (in kilometers)
- Radius (in kilometers)

Lightning data within 200m, 400m and 500m around the feeder were considered. In the initial analysis lightning strokes within 200m and 400m on either side of the distribution feeders were considered and simulations were performed. Results showed that the consideration of lightning data within a distance of 400m around the feeders slightly improved the performance in comparison to data within a distance of 200m around the feeders [23-25]. Although the reported value for median accuracy by the North American Lightning Detection Network (NLDN) is 500 m [36], 200m and 400m were utilized because the utility providing the data wanted to know whether results obtained with these distances had a significant difference. In subsequent analysis for this dissertation, lightning within 500m around the feeder were used to study lightning stroke influence on system interruptions. Figure 2.1 shows the lightning region within 500m around the feeder for Manhattan.

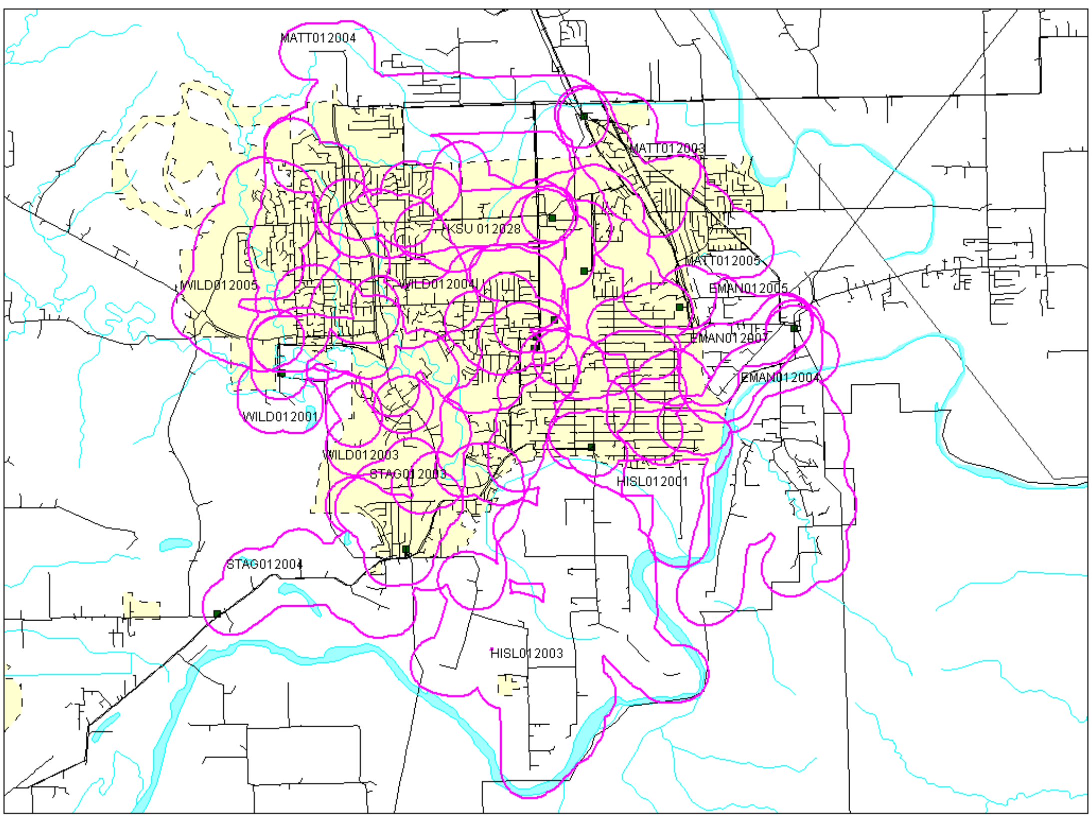


Figure 2.1 Area of Lightning for Manhattan

STN---	WBAN	YEARMODA	TEMP		DEWP		SLP		STP		VISIB		WDSP		MXSPD	GUST	MAX	MIN	PRCP
724555	3936	20090101	33.2	24	17.2	24	1014.4	24	974.8	24	10	24	9	24	14	22	45	24.1	0
724555	3936	20090102	33.9	24	24	24	1011.6	24	972.2	24	10	24	2.3	24	7		55.9	19	0
724555	3936	20090103	41.5	24	37.8	24	1004.1	14	966.6	11	5.6	24	5.5	24	15	24.1	50	26.1	0
724555	3936	20090104	19.9	24	7.2	24	1027.7	21	985.5	24	10	24	8.1	24	17.1	26	26.1	14	0
724555	3936	20090105	25.3	24	8.2	24	1024.4	24	984.1	24	10	24	3.4	24	11.1		37.9	14	0
724555	3936	20090106	23.8	24	13.1	24	1005.2	24	965.7	24	10	24	2.3	24	6		39.9	12.9	0
724555	3936	20090107	32.9	24	21.5	24	1003.9	24	964.8	24	10	24	4	24	8		45	21	0
724555	3936	20090108	30.8	24	20.4	24	1013	24	973.6	24	9.9	24	3.7	24	8.9		46.9	18	0
724555	3936	20090109	40.1	24	27.9	24	1008.9	24	969.9	24	10	24	6	24	19	28.9	59	28.9	0
724555	3936	20090110	24.6	24	13.6	24	1025.7	20	985.7	23	9.8	24	8.4	24	15	26	34	18	0
724555	3936	20090111	29.3	24	17.4	24	1024.7	24	984.5	24	10	24	2.8	24	7		50	15.1	0
724555	3936	20090112	36.2	24	22	24	1019	24	979.5	24	10	24	5.7	24	18.1	29.9	51.1	24.1	0
724555	3936	20090113	18.6	24	2.8	24	1028.3	24	987.8	24	10	24	7.2	24	17.1	25.1	37.9	6.1	0
724555	3936	20090114	29.7	24	11.1	24	1022.6	24	982.7	24	10	24	10.5	24	18.1	27	46	12.9	0
724555	3936	20090115	7.3	24	-8.7	24	1043.3	24	1001.5	24	10	24	6.7	24	12		16	1.9	0
724555	3936	20090116	17.1	24	4.7	24	1035.3	20	994	20	9.2	24	7.5	24	15	17.1	36	1.9	0.01
724555	3936	20090117	35.1	24	21.5	24	1019.4	24	979.6	24	10	24	5.4	24	13	22	55.9	17.1	0
724555	3936	20090118	40.2	24	23	24	1018	24	978.7	24	10	24	4.9	24	12	20	59	25	0
724555	3936	20090119	37.7	24	22.8	24	1017.2	23	978	24	10	24	8.2	24	14	24.1	51.1	27	0
724555	3936	20090120	33.6	24	22.3	24	1023.3	21	983.5	24	10	24	6.8	24	17.1	27	45	28	0
724555	3936	20090121	30.7	24	19.6	24	1017.6	24	977.7	24	9.6	24	2.1	24	5.1		54	16	0

Figure 2.2 Screenshot of Weather Recording in Manhattan

time	lat	lon	peak_ka	num_dfrs	asset_name	asset_len_km	radius
04/22/10 07:03:04.14	39.1775	-96.557	44	25	EMAN012004	21.0683	0.5
04/22/10 07:19:15.63	39.1606	-96.5406	-6	3	EMAN012004	21.0683	0.5
04/22/10 07:25:02.98	39.155	-96.5464	-9	4	EMAN012004	21.0683	0.5
04/22/10 07:39:00.41	39.1689	-96.5367	22	15	EMAN012004	21.0683	0.5
04/22/10 07:47:25.50	39.1727	-96.5519	28	15	EMAN012004	21.0683	0.5
04/22/10 23:03:02.50	39.1803	-96.5421	49	4	EMAN012004	21.0683	0.5
04/23/10 00:07:55.98	39.167	-96.5396	-16	8	EMAN012004	21.0683	0.5
04/23/10 00:11:33.01	39.1638	-96.5349	-101	10	EMAN012004	21.0683	0.5
04/23/10 00:13:35.18	39.1746	-96.5358	-21	9	EMAN012004	21.0683	0.5
04/23/10 00:15:06.47	39.1872	-96.5354	-36	16	EMAN012004	21.0683	0.5
04/23/10 00:15:06.51	39.187	-96.5381	-15	7	EMAN012004	21.0683	0.5
04/23/10 00:15:06.53	39.1867	-96.536	-14	5	EMAN012004	21.0683	0.5
04/23/10 00:15:06.70	39.1861	-96.5391	-23	10	EMAN012004	21.0683	0.5
04/23/10 00:15:06.76	39.188	-96.5387	-7	3	EMAN012004	21.0683	0.5
04/30/10 00:48:12.35	39.1621	-96.5546	-10	5	EMAN012004	21.0683	0.5
05/10/10 08:48:39.49	39.1814	-96.5573	-14	8	EMAN012004	21.0683	0.5
05/10/10 08:52:41.76	39.1718	-96.5588	-7	3	EMAN012004	21.0683	0.5
05/10/10 08:59:29.22	39.1847	-96.5398	-26	13	EMAN012004	21.0683	0.5
05/30/10 19:47:35.79	39.1903	-96.5434	-61	17	EMAN012004	21.0683	0.5
06/08/10 05:31:47.42	39.1835	-96.5699	-7	2	EMAN012004	21.0683	0.5
06/08/10 07:06:59.14	39.1554	-96.5414	-10	4	EMAN012004	21.0683	0.5
06/08/10 07:06:59.28	39.1563	-96.5437	-12	4	EMAN012004	21.0683	0.5
06/08/10 07:06:59.37	39.1576	-96.5371	-34	19	EMAN012004	21.0683	0.5
06/08/10 07:06:59.41	39.1567	-96.5396	-14	5	EMAN012004	21.0683	0.5
06/12/10 07:50:20.85	39.1874	-96.5446	-42	21	EMAN012004	21.0683	0.5
06/12/10 09:07:18.78	39.1904	-96.5423	-34	17	EMAN012004	21.0683	0.5

Figure 2.3 Screenshot of Lightning Recording in Manhattan

2.2 Characteristics of Outage Data

Westar Energy in Topeka, Kansas provided outage data extracted from their Outage Management System (OMS), which is a computer system used by operators of electric distribution systems to assist in restoration of power. Typical OMS 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, outage date, and outage time. Figures 2.4 and 2.5 show the screenshot of outages recorded by Westar Energy for Manhattan, Kansas. Each column represents the elements recorded as summarized in Table 2.2.

Table 2.2 Outage Recording Description

Code	Description
Cause CD	Outage Cause
Customer minutes	Customer without power in minutes
OFFC	Office name code
CUST	Number of customers on each device includes all customers downstream of that device
DUR	Duration of the outage
CAUSE	Outage cause code
COMMENT	Outage cause reason
DVC	Type of device failed, ex. Switch, circuit breaker, transformer
ISOEQ	Equipment isolated
PH	phase
DT OUT	Date of outage occurrence
TM OUT	Time of outage occurrence
DT RSTRD	Data of power is restored

TM RSTRD	Time of power is restored
NM CIRC	Circuit name
Failed	Number of devices failed
PLND	Planned outage in Yes/No
PLND Type	Planned outage type code
RSPSYS	Feeder design – overhead or underground
RSPKV	System voltage code

CAUSE CD	Customer Minutes	OFFC	CUST	DUR	CAUSE	COMMENT	DVC	ISOEQ	PH	DT OUT	TM OUT
CUSTOMER PROBLEM	52	24	1	52	12	CUST MAIN PANEL PROBLEM, LL CALLING ELECTRIAN. JIM GOFF 04573	PR	188302		1/1/2010	10:07
Trees (ALL)	180	24	3	60	10	20 amp fuse trees slapping primary	SW	M007F70855	C	1/3/2010	9:19
NON OUTAGE	213	24	1	213	25	WRONG TS / MISASSIGNED	TS	129139		1/4/2010	19:02
EQUIPMENT FAILED	73	24	1	73	3	secondary j boxs had the wire come out of the sec. block	PR	702129		1/4/2010	19:04
NON OUTAGE	-	24	1	0	25	CALL LANDLORD/MISASSIGNED TS	TS	129129		1/4/2010	19:18
NON OUTAGE	40	24	1	40	25	the customers breaker was off	PR	713148		1/5/2010	6:52
NON OUTAGE	18	24	1	18	25	CUST CALLED BACK	TS	205524		1/5/2010	8:17
ANIMALS/WILDLIFE	360	24	12	30	15	restored	SW	M007F70895	C	1/5/2010	10:19
SAFETY/HAZARD	41	24	1	41	27	TURNED OFF METER AT FIREMANS REQUEST, HOUSE FIRE 16249	PR	207922		1/5/2010	12:10
EQUIPMENT FAILED	49	24	1	49	3	REPLACED FAILED TRAN/05475	SW	M007F82693	B	1/5/2010	14:25
EQUIPMENT FAILED	142	24	1	142	3	replaced service line	PR	207924		1/5/2010	16:39
PROCEDURAL ERROR	17	24	1	17	24	was not turned on when said by service person	PR	218820		1/6/2010	15:35
EQUIPMENT FAILED	-	24	0	82	3	ARRESTERS FAILED @ M34 2601 & M34 2602	CR	JEFF342547	3P	1/7/2010	8:17
EQUIPMENT FAILED	-	24	0	131	3	ARRESTERS FAILED @ M34 2601 & M34 2602	SW	JEFF342547	3P	1/7/2010	8:17
EQUIPMENT FAILED	-	24	0	178	3	ARRESTERS FAILED @ M34 2601 & M34 2602	SW	JEFF342547	3P	1/7/2010	8:17
EQUIPMENT FAILED	-	24	0	499	3	ARRESTERS FAILED @ M34 2601 & M34 2602	SW	JEFF342547	3P	1/7/2010	8:17
CUSTOMER PROBLEM	19	24	1	19	12	CUST BREAKER	PR	205923		1/7/2010	9:43
EQUIPMENT FAILED	-	24	0	0	3	TRYING TO ISOLATE PROBLEM - CLOSED IN ON FAULT F28TAB	CR	JEFF342547	3P	1/7/2010	11:29
EQUIPMENT FAILED	-	24	0	0	3	TRYING TO ISOLATE PROBLEM - CLOSED IN ON FAULT F28TAB	CR	JEFF342547	3P	1/7/2010	12:53
OTHER	35	24	1	35	16	FUSE PULLED APART IN EXTREME COLD	TS	305638		1/7/2010	14:47
EQUIPMENT FAILED	-	24	0	8	3	TRYING TO ISOLATE PROBLEM - CLOSED IN ON FAULT F28TAB	CR	JEFF342547	3P	1/7/2010	15:32
NON OUTAGE	3	24	1	3	25	CUSTOMER CALLED CSR AND CANCELLED	TS	721712		1/7/2010	17:45
Trees (ALL)	102	24	1	102	10	broken net. on the service wire	PR	219522		1/7/2010	19:00
NON OUTAGE	8	24	1	8	25	CUST CALLED APT MAINT	TS	219610		1/9/2010	10:49
NON OUTAGE	50	24	1	50	25	Cust said all back on now??? Jim Goff	PR	218071		1/9/2010	13:18

Figure 2.4 Screenshot 1 of Outage Recording in Manhattan

TM RSTRD	NM CIRC	CIRC CUST	ISOL EQ	OLD NORM	IEEE 13	KCC 10PCT	FAILED	PLND	PLND TYPE	RSPSYS	RSPKY
10:59	LNVR012001	250	8	N	N	N		N		OH	10
10:19	CARN007000	21	12	N	N	N		N		OH	1
22:35	MATT012002	1249	8	N	N	N		N		OH	10
20:17	WILD012005	677	8	N	N	N	2	N		UG	12
19:18	MATT012002	1251	8	N	N	N		N		OH	10
7:32	WILD012009	1704	8	N	N	N		N		UG	1
8:35	HISL012003	1206	8	N	N	N		N		OH	10
10:49	MATT012004	1440	12	N	N	N		N		OH	1
12:51	HISL012002	1136	8	N	N	N		N		OH	10
15:14	MATT012002	1249	12	N	N	N	5	N		UG	1
19:01	HISL012002	1136	8	N	N	N	3	N		OH	11
15:52	WILD012001	1659	8	N	N	N		N		OH	10
9:39	JEFF342547	0	1	N	N	N	8	N		OH	3
10:28	JEFF342547	0	1	N	N	N	8	N		OH	3
11:15	JEFF342547	0	1	N	N	N	8	N		OH	3
16:36	JEFF342547	0	1	N	N	N	8	N		OH	3
10:02	KSU 012024	1003	8	N	N	N		N		OH	10
11:29	JEFF342547	0	1	N	N	N	20	N		OH	3
12:53	JEFF342547	0	1	N	N	N	20	N		OH	3
15:22	STGE012000	552	8	N	N	N		N		OH	1
15:40	JEFF342547	0	1	N	N	N	20	N		OH	3
17:48	WILD012009	1706	8	N	N	N		N		OH	10
20:42	WILD012001	1659	8	N	N	N		N		OH	11
10:57	WILD012004	1378	8	N	N	N		N		OH	10
14:08	RAND012000	528	8	N	N	N		N		OH	10
14:51	WILD012009	1706	8	N	N	N		N		OH	10
15:44	WILD012009	1705	12	N	N	N		Y	4	OH	1
16:48	EMAN012004	869	8	N	N	N		N		OH	10

Figure 2.5 Screenshot 2 of Outage Recording in Manhattan

Utility crews check outaged feeders and report presumed causes based on their experience and on the spot circumstances. For example, if a dead squirrel is found on a transformer or fallen branches are found near outage feeders, “squirrel on transformer” or “tree burned service down” is the reasonable explanation. However, when no clue is found, crews may guess by giving priority to the most plausible reason. Sometimes they record unknown is a reason if not found conclusively. Therefore, errors cannot be avoided in determining the outage cause.

Weather condition was also an important reference to analyze the true reason for outages in the OMS. Table 2.1 shows weather codes with descriptions.

Table 2.3 Weather Code & Description

Code	Description
1	THUNDERSTORM
2	LIGHTNING IN AREA
3	WET
5	ICE
6	ICE AND WIND
7	WIND
8	HEAT
9	COLD
10	FAIR
11	TORNADO
12	MICROBURST

Outage causes are also recorded in the Outage Management System for each outage incidence. The classification of outage causes is shown in Table 2.2.

Table 2.4 Cause Code & Description

Cause Code	Description
1	CUSTOMER REQUEST
3	EQUIPMENT FAILED
9	OVERLOAD
10	Trees (ALL)
11	PUBLIC DAMAGE
12	CUSTOMER PROBLEM
15	ANIMALS/WILDLIFE
16	OTHER
17	LIGHTNING
18	EXTREME WIND
19	ICE STORM
20	Trees (ALL)
21	DEBRIS NATURE/WTHR
22	UNKNOWN
23	COMPANY DAMAGED
24	PROCEDURAL ERROR
25	NON OUTAGE
26	LOAD TRANSFER
27	SAFETY/HAZARD
28	TURN-ON (VALID)
29	LOAD SHED
30	MAINTENANCE

Figure 2.5 shows the number of outages possibly caused directly or indirectly by severe weather conditions in Manhattan, Kansas as recorded by Westar Energy from 2005 to 2011. Not all the outages caused by failed equipment or unknown causes are due to weather, but several of

these are caused by weather. This figure provides a clear illustration of weather impacts on overhead feeders.

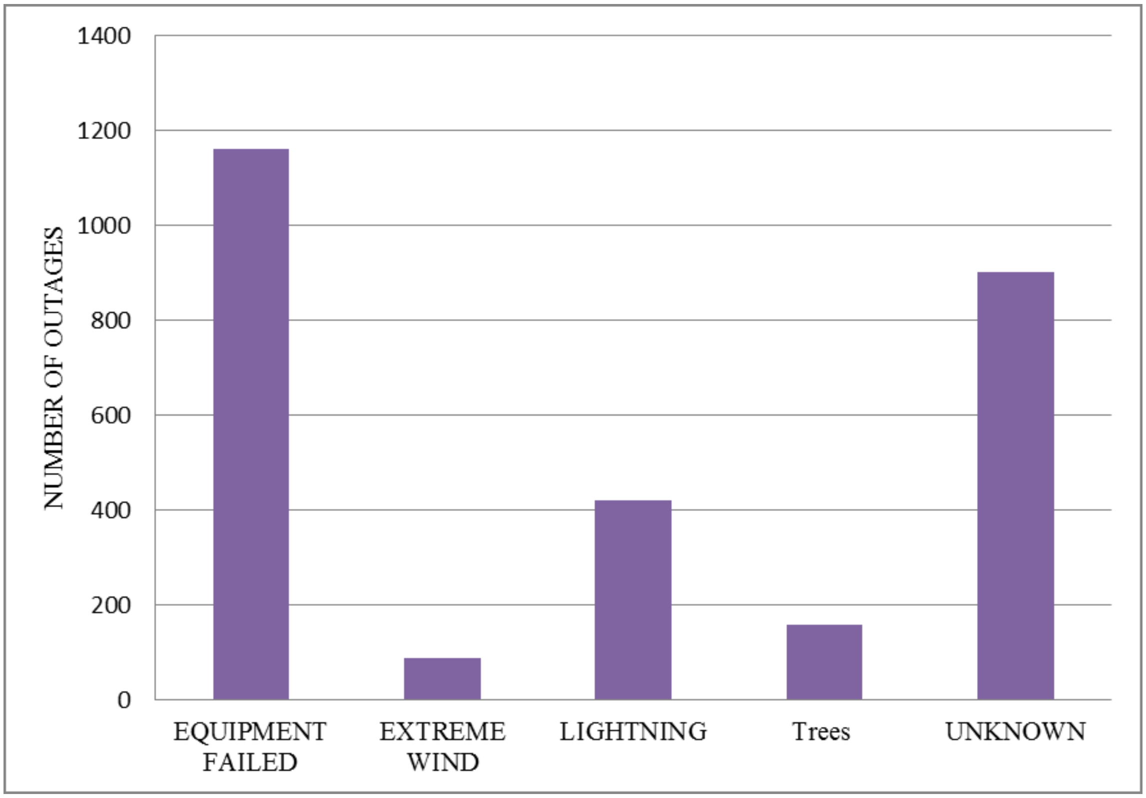


Figure 2.6 Total Number of Outages Caused by Different Weather Factors between 2005 and 2011 in Manhattan

Figures 2.6 and 2.7 show the histogram of outages per day in the study period for the four cities in Kansas. Figure 2.7 shows the histogram in higher range. These figures show that a large number of days occurred with zero or low number of recorded outages. Manhattan had the most days with zero outages and Wichita had the least number of days with zero outages, while Lawrence and Topeka were in the middle. The trend reversed for one or more outages. This is an outcome of the spatial aggregation of outages. Since Wichita is the largest service area, outage probabilities at each level greater than zero is higher than cities with smaller service areas.

2.3 Data processing

Data processing extracts useful information from historical outage data, weather data, and lightning stroke data and integrates them into a single data set. Existing literature [38] suggests that either gust or sustained wind can be used to study outage effects, with neither being more advantageous. 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 work, maximum daily wind gusts measured on a five-second basis was used as the variable to study wind effects because our previous research had found it to provide the best correlation of outages compared to other variables. However, for days with low wind speeds which did not have gusts recorded, a one-minute sustained speed was used. Investigations to identify additional suitable wind related variables from available data to include in the analysis will be pursued as future research.

Daily aggregate lightning stroke currents are calculated by totaling magnitudes of all lightning strokes in kiloAmps (kA), including the first stroke and flashes within 500m around the feeders for each day of the study [36]. Since the research goal was to study combined effects of wind and lightning as well as just wind, all days, including those that did not have any recorded lightning, were included. However, days of extreme weather conditions were excluded, including three such days for Lawrence, six days for Topeka, and eight days for Wichita. These days were considered outliers and were removed from the data for analysis, which spanned a period of seven years from 2005 to 2011.

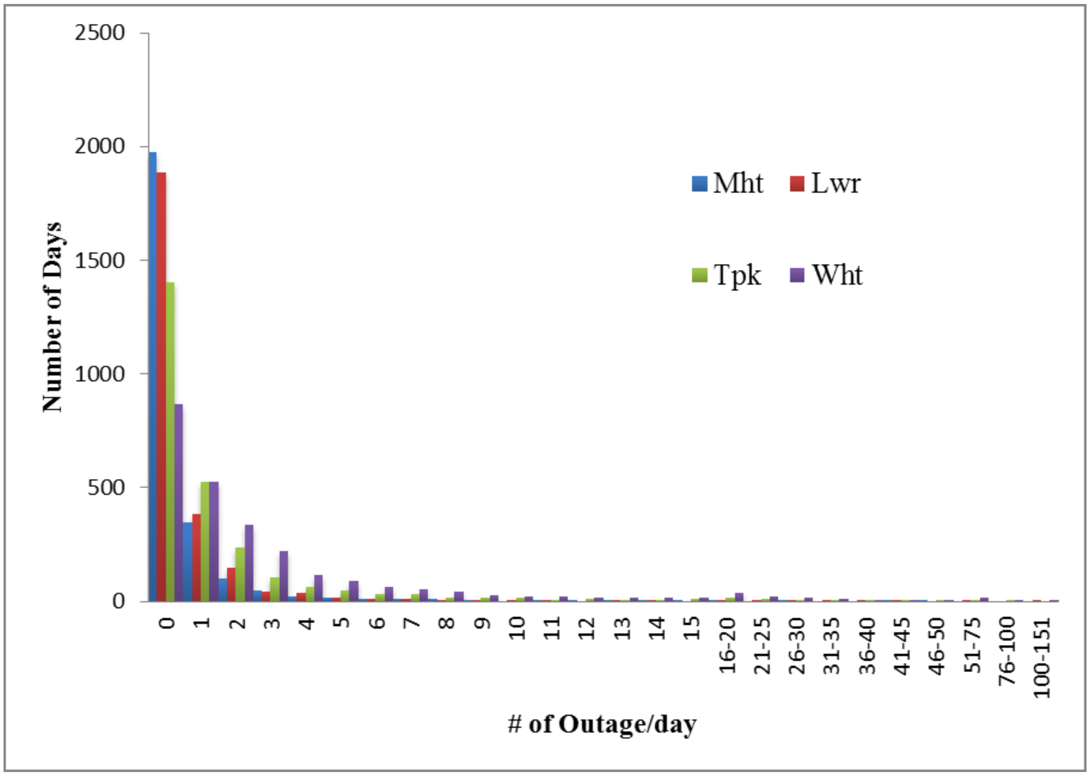


Figure 2.7 Histogram of Outages Caused by Wind and Lightning, 2005-2011

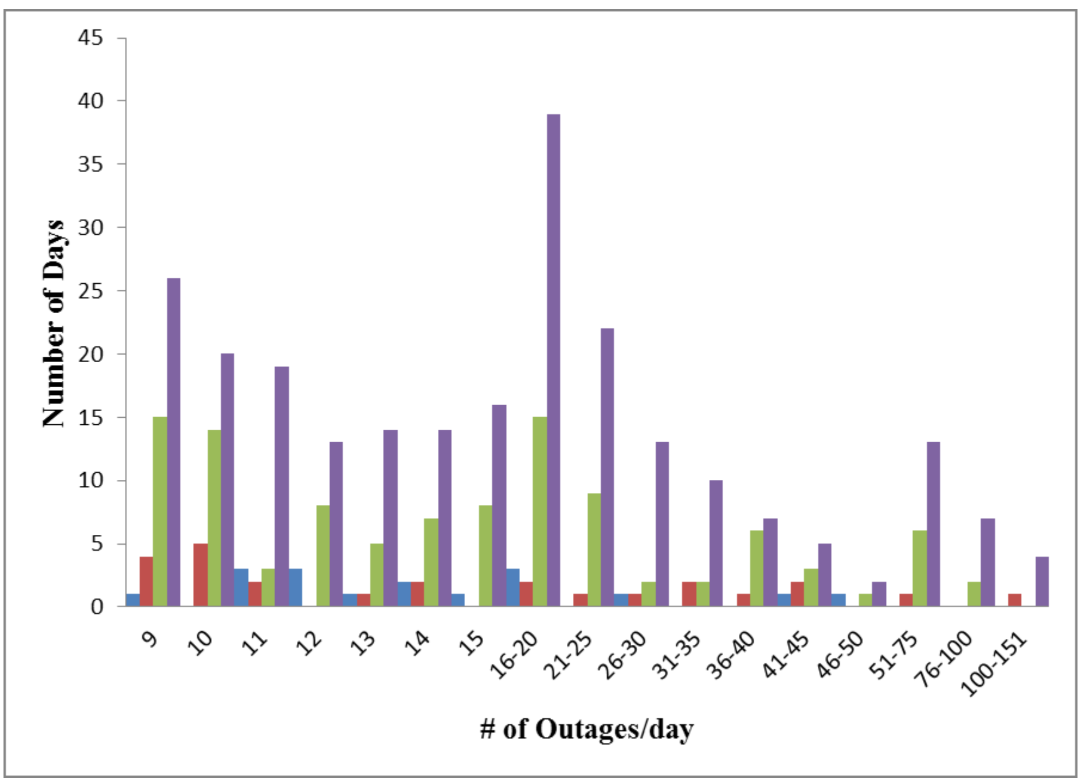


Figure 2.8 Histogram of Outages in the Higher range caused by Wind and Lightning, 2005-2011

Since this work focuses on outages caused by weather, lightning, trees, wind, equipment, and unknown factors, outages possibly caused by lightning and wind were included in outage counts for the study. Weather at the time of recorded lightning, equipment failure, and unknown outages were manually examined to ensure that the lightning actually occurred on the feeder experiencing outage. Outages 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. Two hour and four hour time window after the recorded lightning was considered during the day and night respectively for inclusion of these outages in the counts. Since a power outage during nighttime is typically reported late compared to outage report during daytime, a wider time window was considered.

Chapter 3 - Multiple Regression Models

Regression models are statistical methods that offer the combined advantages of model simplicity and low computation overheads. To find the relation between lightning and wind with outages occurring on overhead distribution feeders, multiple-regression models were developed. The idea of a regression model came from previous work of Zhou, Pahwa and Yang in which they used linear regression and poison regression to predict the weather-related outages (wind and lightning) on a distribution system from 1998 to 2003 [38]. In their study, the lightning was regarded as a system-wide measurement and included only failures that resulted in outages to customers. The proposed models were for estimating monthly weather-related outages, and the regression models utilize the least square criterion to estimate the regression coefficient.

3.1 Introduction to Regression Model

Regression models are employed to find the relationship between one or many dependent variables and one or many independent variables. The objective of regression models is to express the response variable as a function of the predictor variables. The model performance and conclusion drawn from model results depend on the data used. Hence, non-representative or incomplete data result in poor fits and conclusions. Thus, for effective use of regression analysis the data collection process must be checked, any limitations in data collected must be found, and conclusions must be restricted accordingly. Once the relationship between response and predictor variables is obtained, the regression analysis can be used to predict values of the response variable, model specification, and parameter estimation [54].

Differences exist between multiple-regression models and multi-variate regression models. In multiple-regression models, only one dependent variable and one or many independent variables are present, whereas multi-variate regression models contain one or many dependent variables and one or many independent variables [55].

3.1.1 Linear Regression Model

A regression model is the linear regression model which is a linear relationship between response variable, Y, and the predictor variable, $X_j = \left(X_{1j}, X_{2j}, \cdots, X_{Nj}\right)$, of the form

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \qquad (3.1)$$

where,

$X_1, X_2, \cdots, X_N$ are known variables

$\beta_0, \beta_1, \beta_2, \cdots, \beta_p$ are regression coefficients (unknown model parameters),

$\varepsilon_i, 1 \leq i \leq N$ is the error due to variability in the observed responses and assumed independent $N(0, \sigma^2)$, and

$i = 1,2,3 \ldots.., N$ where N is the number of data entries.

Matrix representation of the linear regression model (3.1)

$$Y_{N\times 1} = X_{N\times p} \times \beta_{p\times 1} + \varepsilon_{N\times 1} \tag{3.2}$$

where,

$Y_{N\times 1}$ is a vector of responses, $\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}$

$\beta_{p\times 1}$ is a vector of parameters, $\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$

$X_{N\times p}$ is a matrix of predictors, $\begin{bmatrix} 1 & X_{11} & \cdots & X_{1,p-1} \\ 1 & X_{21} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{N1} & \cdots & X_{N,p-1} \end{bmatrix}$

$\varepsilon_{N\times 1}$ is a vector of independent normal random variables, $\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$

3.1.2 Regression Coefficients

Regression coefficients can be obtained by any of the following criteria:

1. QR Factorization
2. Least Square
3. Weighted Least Square (WLS)
4. Gaussian
5. Pseduoinverse

6. Singular Value Decomposition (SVD)

The least square criterion has important statistical interpretations. The least criterion is frequently used for solving over-determined or inexactly specified systems of equations in an approximate sense. Instead of solving equations exactly, minimization of the sum of the residual squares is used. Zhou, Pahwa and Yang used the lease square criterion for estimating coefficients [38].

The most widely known type of matrix pseudoinverse is the Moore–Penrose pseudoinverse, which was independently described by E. H. Moore in 1920 [56], Arne Bjerhammar in 1951 [57], and Roger Penrose in 1955 [58]. The definition of the pseudoinverse makes use of the Frobenius norm of an *mxn* matrix *A* defined as the square root of the sum of the absolute squares of its elements,

$$||A||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2} \tag{3.3}$$

The sum of the square of the error (SSE) for (3.2) is:

$$SSE(\beta_0, \beta_1, \cdots, \beta_p) = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{3.4}$$

$$\hat{Y}_i = \left(\beta_0 + \sum_{j=1}^{p-1}\beta_j X_{ij} + \varepsilon_i\right) \tag{3.5}$$

To calculate vector β, which minimizes SSE

$$\frac{\partial SSE}{\partial \beta_j} = -2\left(Y - X\hat{\beta}\right)^T X_j = 0, \quad 0 \le j \le p-1 \tag{3.6}$$

Equivalent to

$$\left(Y - X\hat{\beta}\right)^T X_j = 0 \tag{3.7}$$

$$Y^T X = \hat{\beta}^T (X^T X) \tag{3.8}$$

$$\hat{\beta} = (X^T X)^{-1}(X^T Y) \tag{3.9}$$

A general exponential regression model:

$$Y_i = exp\left(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{i,p-1} + \varepsilon_i\right) \tag{3.10}$$

And in matrix form,

$$Y_{N\times 1} = \exp(X_{N\times p} \times \beta_{p\times 1} + \varepsilon_{N\times 1}) \tag{3.11}$$

Several methods are available for function approximation in order to predict values of model parameters. Gradient-descent methods are the most widely used of all function approximation methods [55]. The steepest descent method has been used to solve nonlinear equations in order to obtain model parameters while minimizing the sum of weighted least square error between the observed and estimated outages given below:

$$E = w_i \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i\right)^2 \tag{3.12}$$

$$\hat{Y}_i = exp\left(\beta_0 + \sum_{j=1}^{p-1} \beta_j X_{ij} + \varepsilon_i\right) \tag{3.13}$$

The partial derivative of weighted sum square error with respect to the coefficients is given by

$$\Delta_\beta E_i = w_i\left(Y_i - \hat{Y}_i\right) X_i \tag{3.14}$$

Computation begins with an initial guess of the values of ***β***. For every subsequent iteration, the coefficients are updated

$$\beta = \beta - \alpha * \Delta_\beta E_i \tag{3.15}$$

where, α is a constant in the range 0.001 to 0.1. For larger $\Delta_\beta E_i$, α equal to 0.001, and for the smaller $\Delta_\beta E_i$, α equal to 0.1, was used. In this research, all the weights, w_i, were considered to be equal to 1.

3.2 Model Construction

Zhou, Pahwa and Yang attempted several mathematical functions on response and explanatory variables, such as square root and natural log, and all possible regression procedures were conducted for the preceding 15 explanatory variables and three response variables. They found that taking the original form of the number of failures as the response variable and wind gust speed, lightning stroke current, and natural log of lightning stroke current as explanatory variables, yield comparatively small MSE [38].

From Zhou, Pahwa and Yang work and other published works [38], the following six regression functions were considered:

Model 1: $$\hat{O} = \beta_1 Li + \beta_2 Wd \tag{3.16}$$

Model 2: $$\hat{O} = \beta_1 Li + \beta_2 Wd + \beta_3 Wd^2 \tag{3.17}$$

Model 3: $$\hat{O} = \beta_1 Li + \beta_2 Wd + \beta_3 Wd^2 + \beta_4 Li^2 \tag{3.18}$$

Model 4: $$\hat{O} = \beta_1 Li + \beta_2 Wd + \beta_3 Wd \times Li + \beta_4 Wd^2 + \beta_5 Li^2 \tag{3.19}$$

Model 5: $$\hat{O} = \exp(\beta_0 + \beta_1 Li + \beta_2 Wd + \beta_3 Wd \times Li) \tag{3.20}$$

Model 6: $$\hat{O} = \exp(\beta_0 + \beta_1 \ln(Li) + \beta_2 Wd + \beta_3 Wd \times \ln(Li)) \tag{3.21}$$

where,

Li are the daily accumulated lightning stroke current in kiloAmps

$\ln(Li)$ are the natural log of lightning stroke current

wd are daily maximum wind gust speed in miles per hour

$\hat{O}$ observed outages, and β's are the regression coefficients.

Model 1 is a linear model, Model 2 considers linear relationship for lightning and quadratic relationship for wind, Model 3 considers quadratic relationship both for wind and lightning and Model 4 considers quadratic relationship both for wind and lightning and combined effect of lightning and wind. Interactions of wind and lightning are represented by multiplication of any wind and lightning variables. Model 5 and 6 are the exponential models, both have same structure except in model 6 logarithmic relation for lightning is considered.

Given the wind, lightning, and observed outages, model parameters for the first four models were estimated using the pseudo inverse method. For model 5 and 6, parameters were estimated using the steepest descent approach while minimizing the least square error between the observed and estimated outages.

3.3 Simulations and Model Performance

Table 3.1 tabulates the MAE, MSE, and correlation coefficient values for the six different regression models for Manhattan service area for training and test data. The table demonstrates that, among all six models, model 4 has the lowest MAE, MSE, and the highest correlation coefficient, R, for both the training and test. Similarly, for Lawrence, Topeka, and Wichita, MAE, MSE, and correlation are tabulated in Tables 3.2, 3.3, and 3.4. Performance measure (MAE, MSE) values for the training and test data decreases from model 1 to 4 for all four cities in Kansas. For all the cities, correlation coefficients are all positive, which indicates there is a positive linear relationship between the estimated outages and observed outages. Additionally, slope of the regression line between the observed and estimated outages, S, is shown in these tables. This slope is an indicator of performance of the models. A higher slope indicates better performance, with a slope of one giving the ideal performance.

Table 3.1 Results of Regression Models for Manhattan

Manhattan								
	Training Data				Test Data			
	MSE	MAE	R	S	MSE	MAE	R	S
Model 1	3.3181	0.6702	0.5258	0.2712	2.0779	0.7180	0.4824	0.1556
Model 2	3.2239	0.6055	0.5449	0.3012	2.1021	0.6622	0.4929	0.1826
Model 3	3.1218	0.5924	0.5650	0.3233	2.0185	0.6509	0.5077	0.2362
Model 4	2.7824	0.6056	0.6268	0.3946	2.2247	0.6885	0.4283	0.1328
Model 5	2.8980	0.6334	0.6077	0.3934	2.1622	0.6992	0.4477	0.1467
Model 6	4.1716	0.7795	0.4669	0.3773	2.4705	0.6919	0.4608	0.0589

Table 3.2 Results of Regression Models for Lawrence

Lawrence								
	Training Data				Test Data			
	MSE	MAE	R	S	MSE	MAE	R	S
Model 1	4.4220	0.7421	0.2755	0.0699	4.6686	0.8653	0.5185	0.1420
Model 2	4.3814	0.6879	0.2890	0.0849	4.7475	0.8099	0.5219	0.1431
Model 3	4.3805	0.6860	0.2893	0.0851	4.7243	0.8028	0.5543	0.1396
Model 4	4.3731	0.6872	0.2920	0.0864	4.9956	0.8261	0.5316	0.1051
Model 5	4.4043	0.7114	0.2816	0.0722	4.8554	0.8471	0.5449	0.1168
Model 6	4.4999	0.6502	0.2468	0.0554	5.7250	0.8713	0.4655	0.0371

Table 3.3 Results of Regression Models for Topeka

Topeka								
	Training Data				Test Data			
	MSE	MAE	R	S	MSE	MAE	R	S
Model 1	13.7319	1.4523	0.5114	0.2515	40.3696	2.4318	0.4443	0.0724
Model 2	13.3099	1.3690	0.5325	0.2892	40.3150	2.3851	0.4756	0.0905
Model 3	13.1783	1.3753	0.5391	0.2964	40.9059	2.4054	0.4613	0.0810
Model 4	13.1423	1.3754	0.5409	0.2979	41.2994	2.4105	0.4604	0.0736
Model 5	13.7337	1.4077	0.5114	0.2433	41.1660	2.4329	0.4721	0.0636
Model 6	15.5890	1.3537	0.4189	0.1298	45.0651	2.5160	0.4713	0.0284

Table 3.4 Results of Regression Models for Wichita

WICHITA								
	Training Data				**Test Data**			
	MSE	**MAE**	**R**	**S**	**MSE**	**MAE**	**R**	**S**
Model 1	39.7079	3.0767	0.5151	0.2413	71.8361	3.5966	0.4775	0.1050
Model 2	37.5084	2.8773	0.5501	0.3107	63.3287	3.3308	0.5811	0.1919
Model 3	37.4824	2.8841	0.5506	0.3111	63.5145	3.3392	0.5809	0.1892
Model 4	37.3769	2.8891	0.5524	0.3135	63.5677	3.3468	0.5771	0.1902
Model 5	39.7459	3.0183	0.5114	0.2541	71.0337	3.5369	0.4970	0.1128
Model 6	42.0023	2.9238	0.4678	0.2144	77.2927	3.4298	0.4814	0.0837

Comparing the model performance measures, it is observed that model 4 has better performance with low MSE and MAE values and high correlation compared to models 1, 2 and 3 for all four cities for the training data. The reason behind this very possibly is the simplicity of linear relations in model 1 to 3. The MSE, MAE and R values for model 4 and 5 differ by a very small value. For the test data, it is observed that model 5 has lower MSE and high correlation but high MAE values compared to model 4 for service areas in Kansas except Wichita. For Wichita, model 4 has better performance values compared to model 5. This might be size of service area because the bigger cities have a wider range of outages.

Figure 3.1 to 3.6 show the scatter plots between the observed and estimated outages with a regression line for the training and test data for models 1 to 6 for four service areas in Kansas. From the scatter plot, it is observed that the regression models underestimate the outages when in the higher range and overestimate outages when in the lower range. It can be observed from the scatter plots, the models performance for Lawrence compared to other cities is not good and this might be because of some anomalies is the data.

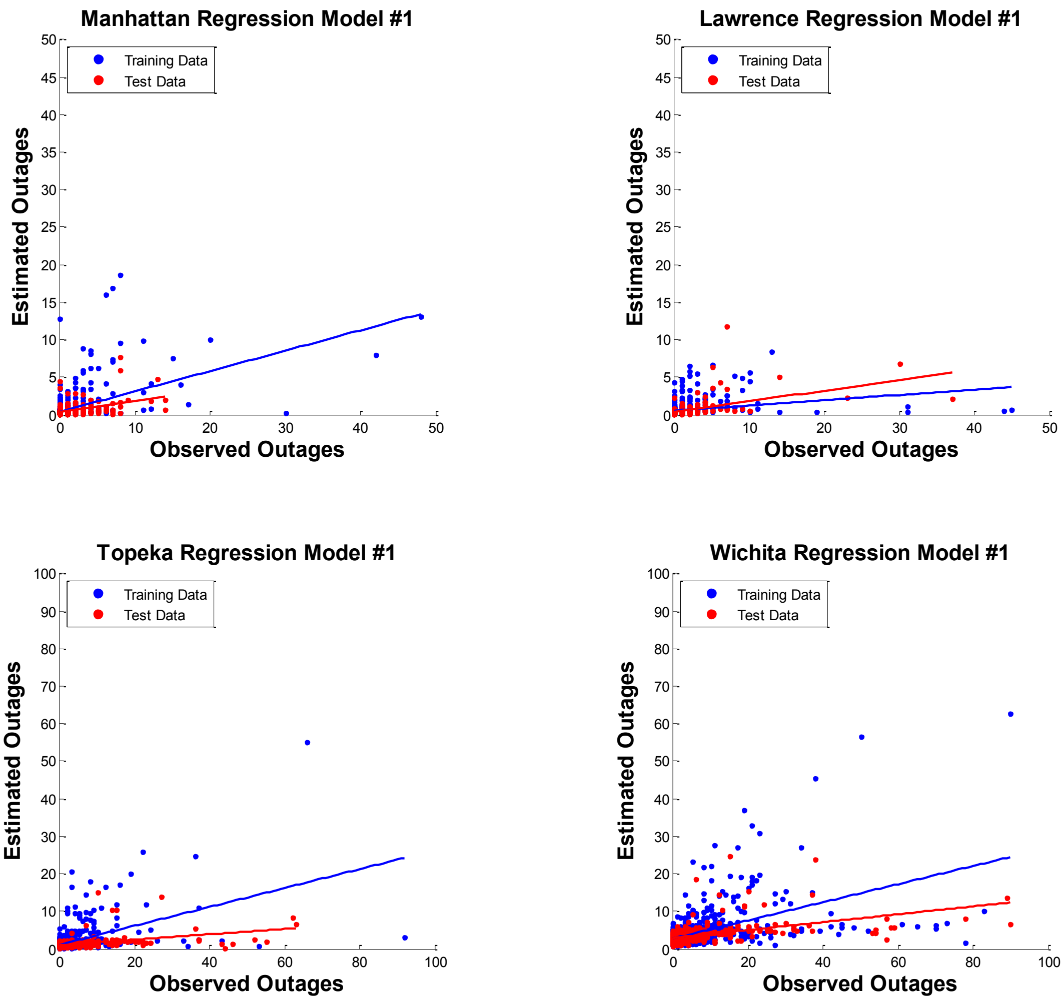


Figure 3.1 Plot of Observed and Estimated Weather-related Outages Obtained with Regression Model 1 in Overhead Distribution Systems for Four Cities from 2005 to 2011

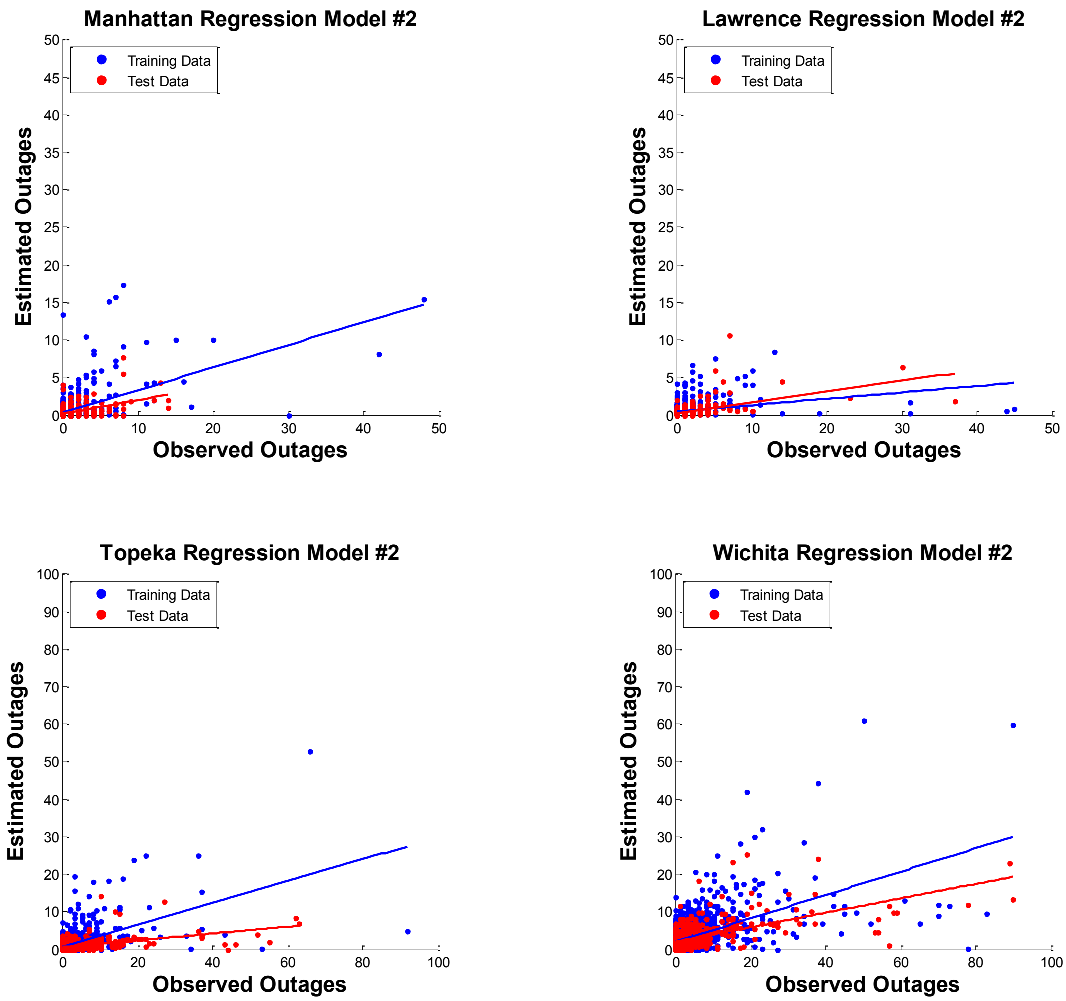


Figure 3.2 Plot of Observed and Estimated Weather-related Outages Obtained with Regression Model 2 in Overhead Distribution Systems for Four Cities from 2005 to 2011

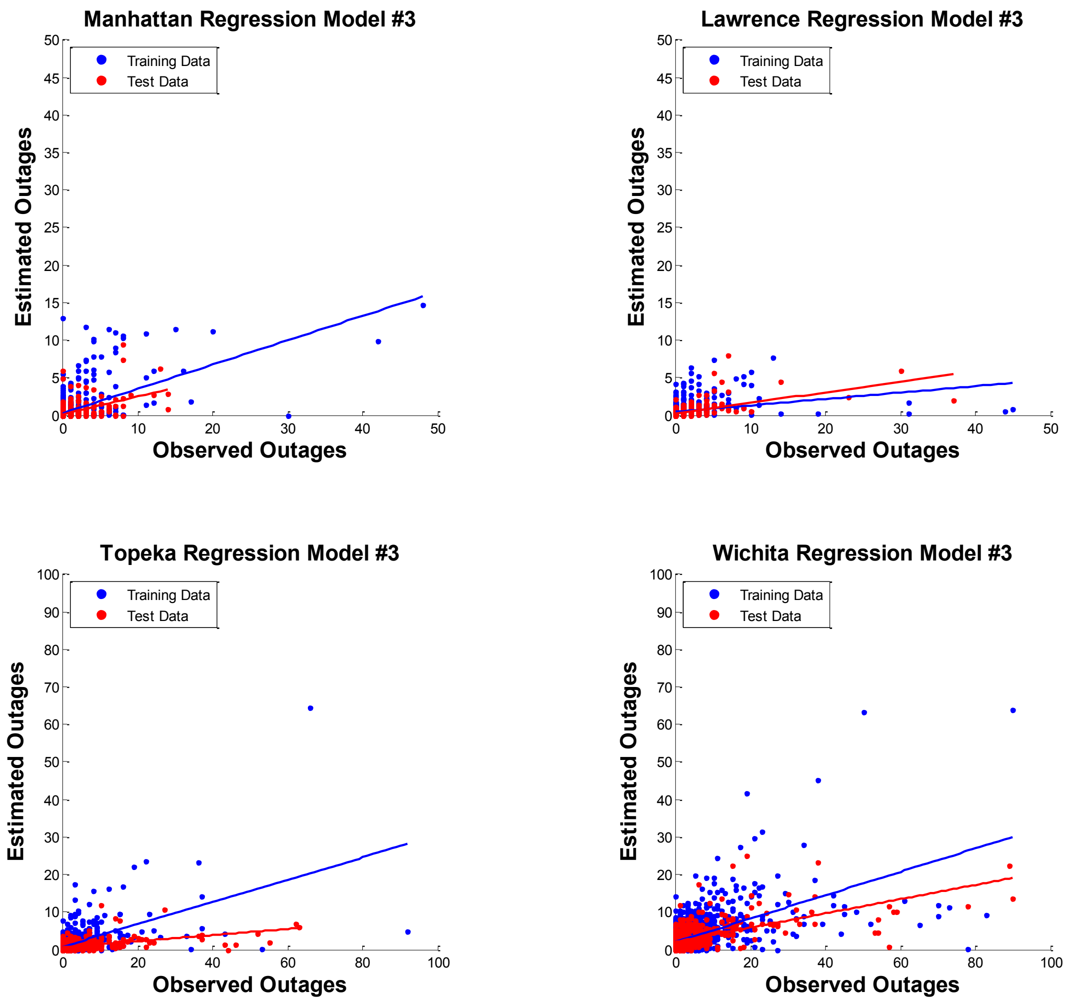


Figure 3.3 Plot of Observed and Estimated Weather-related Outages Obtained with Regression Model 3 in Overhead Distribution Systems for Four Cities from 2005 to 2011

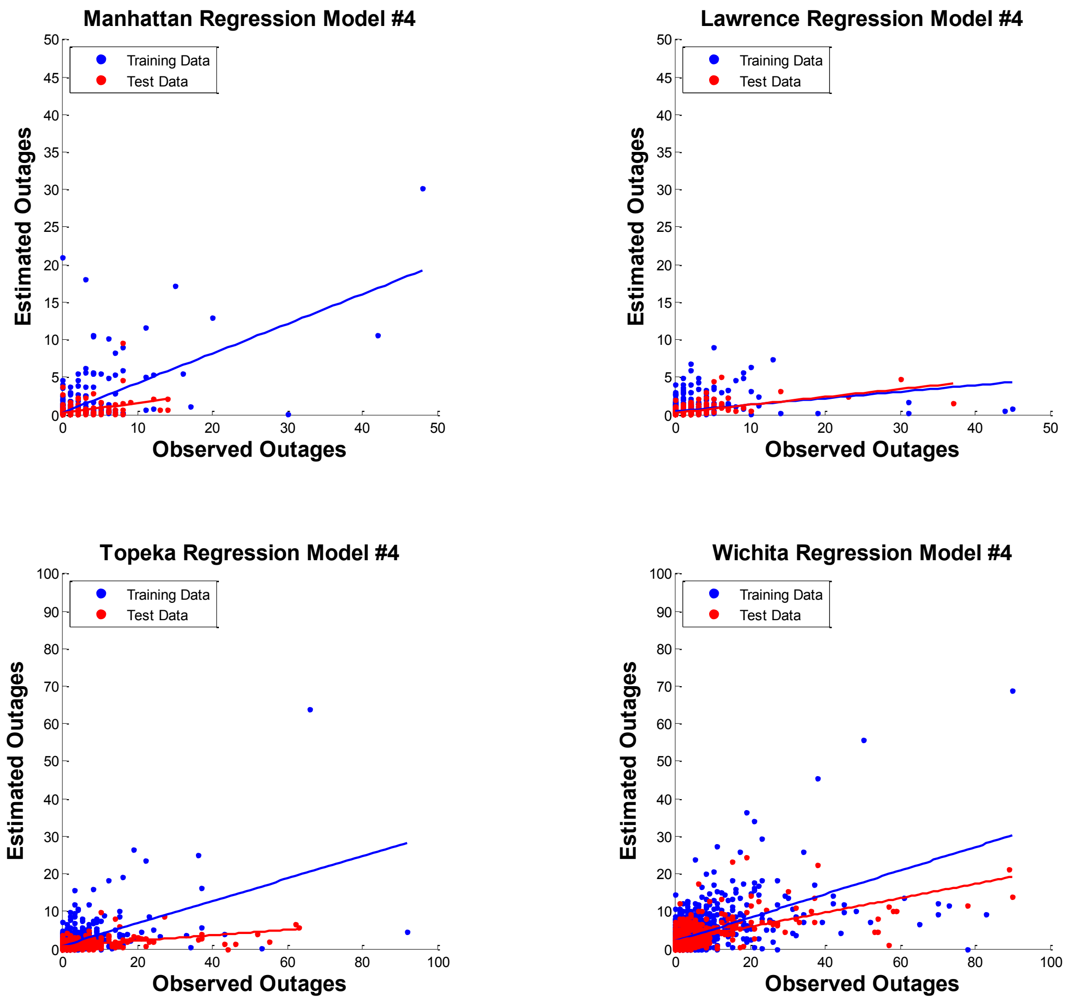


Figure 3.4 Plot of Observed and Estimated Weather-related Outages Obtained with Regression Model 4 in Overhead Distribution Systems for Four Cities from 2005 to 2011

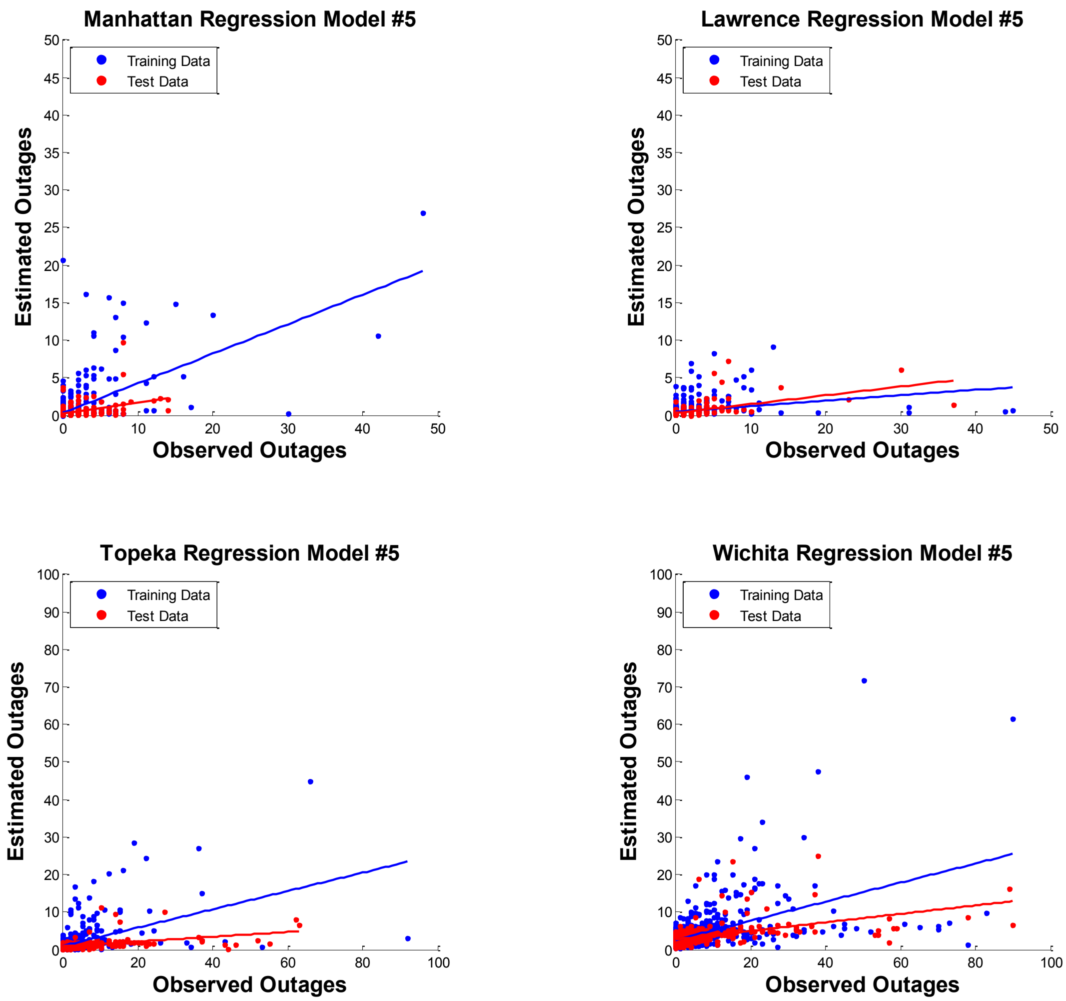


Figure 3.5 Plot of Observed and Estimated Weather-related Outages Obtained with Regression Model 5 in Overhead Distribution Systems for Four Cities from 2005 to 2011

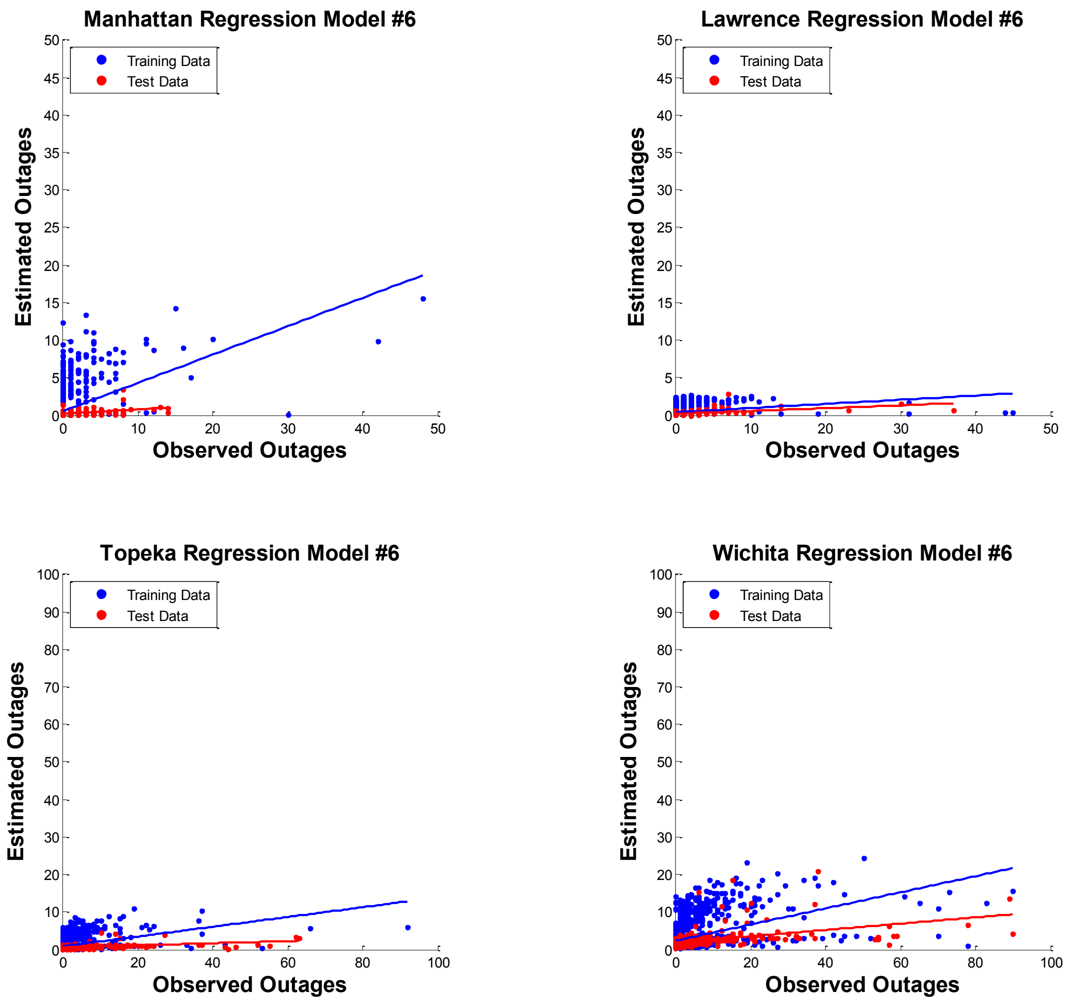


Figure 3.6 Plot of Observed and Estimated Weather-related Outages Obtained with Regression Model 6 in Overhead Distribution Systems for Four Cities from 2005 to 2011

The linear and exponential regression models presented here were also applied for the estimation of outages caused by weather considering wind and lightning stroke within 200m and 400m around the feeder as input to study impact of distance in recording the lightning stroke in reliability assessment [47, 48]. Simulation results shows that considering 400m lighting stroke either side of the overhead distribution feeders provides better estimation [47].

3.4 Summary

Linear, quadratic and exponential regression models are investigated. From performance measure tables for all cities it is observed that model 4 does better estimation compared to other models. Although the multiple regression models are able to approximate complex relations between the input and output, the models underestimate when observed outages are in the higher range and slightly overestimate when observed outages are in the lower range. Better learning models are needed to accurately correlate the input and output.

Chapter 4 - Neural Network Model

Because of the random nature of weather-related outages, conventional methods are limited in capturing nonlinearities in the time series of weather-related outages. In Chapter 3, results based on a multiple regression model showed that regression models overestimate when observed outages are in the lower range, thus prompting demand for a nonlinear model for the estimation for this research. Artificial neural networks-based methods have gained wide attention in engineering and are widely used for forecast because of their ease of use and ability to approximate high complexity functions.

4.1 Neural Network

Neural network (NN) is a general mathematical computing paradigm that models operations of biological neural systems. The non-linear nature of the neural networks allows them to learn from the environment in supervised and unsupervised ways, and the universal approximation property of neural networks makes them highly suited for solving complex problems. The study of artificial neural networks (ANNs) is loosely motivated by biological learning systems which are built of densely interconnected neurons. By mimicking the brain, artificial neural networks acquire knowledge by learning from data and storing it within the connections between neurons. The most widely used neural model was devised in 1943 on McCulloch, a neurobiologist, and Pitts, a statistician [59, 60]. A simple mathematical model of the neuron is illustrated in Figure 4.1. Neural networks have advantages over traditional linear models because they are able to represent linear and non-linear relationships and they can learn these relationships directly from the data being modeled.

4.1.1 McCulloch and Pitts' Neuron Model **[59]**

Network consists of units arranged in layers with only forward connections to units in subsequent layers. The connections have weights associated with them. Each signal traveling along the link is multiplied by the connection weight. The first layer is the input layer, and the input units distribute inputs to units in subsequent layers. In the following layers, each unit sums its inputs, adds a bias or threshold term to the sum, and nonlinearly transforms the sum to produce an output. This nonlinear transformation is called the activation function of the unit.

Output layer units often have linear activations. The layer located between the input layer and output layer are called hidden layers, and units in hidden layers are called hidden units.

Neural networks are composed of nodes or units connected by directed links. A link from unit i to unit j serve to propagate the activation a_i from i to j. Each link is also associated with a numeric weight W_{ij}, which determines the strength and sign of the connection. Each unit j first computes a weighted sum of its inputs, defined as the net function,

$$net_j = \sum_{i=1}^{N} w_{ji}x_i + b_i \tag{4.1}$$

The bias weight, b_j, included in (4.1), is used to model the threshold.

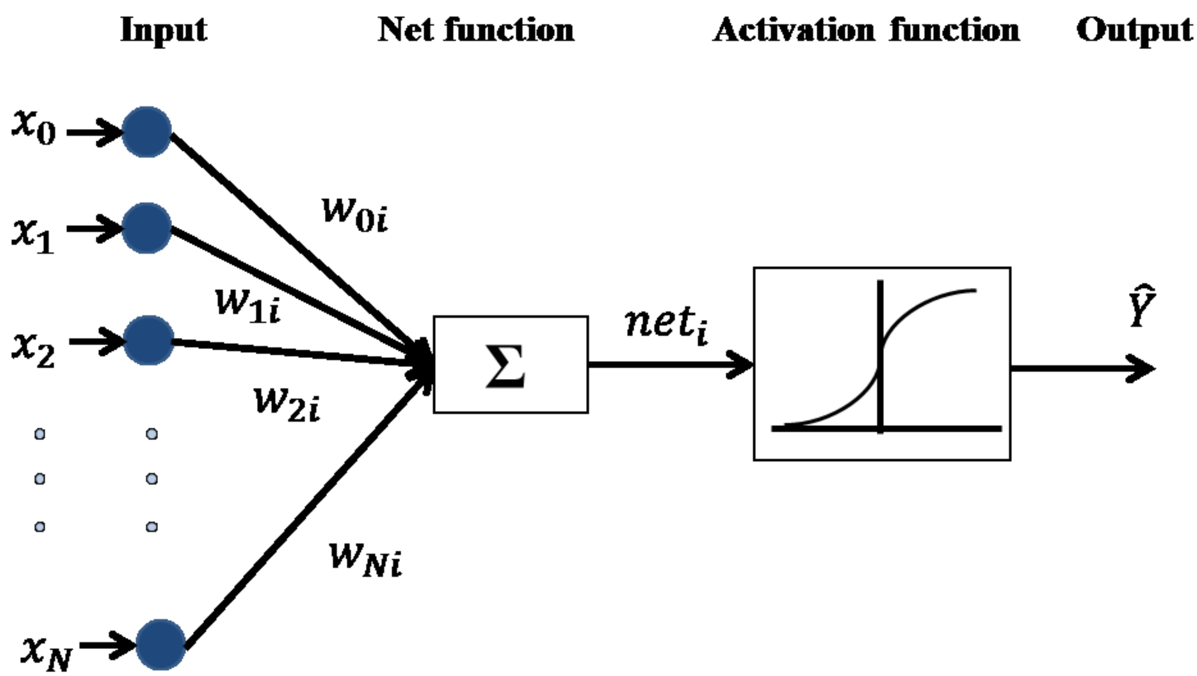


Figure 4.1 Simple Mathematical Model for a Neuron

The output of the neuron, Y_j, is related to the network input, net_i, via a linear or nonlinear transformation called the activation function,

$$Y_i = f(net) \tag{4.2}$$

Additional commonly used activations functions are summarized in Table 4.1. The derivative of the activation function is also provided.

Table 4.1 Neuron Activation Functions

Activation Function	$\boldsymbol{f(u)}$	**Derivative** $\frac{\partial f(u)}{\partial u}$
Sigmoid	$f(u) = \frac{1}{1+e^{-u}}$	$f(u)[1 - f(u)]$
Hyperbolic Tangent	$f(u) = \tanh(u)$	$1 - (f(u))^2$
Threshold	$f(u) = \begin{cases} 1 & u > 0 \\ -1 & u < 0 \end{cases}$	Derivates do not exist at $u = 0$
linear	$f(u) = yu + b$	y

4.1.2 Neural Network Topology

In a neural network, multiple neurons are interconnected to form a network that facilitates distributed computing. Configuration of the interconnections can be described efficiently with a directed graph. A directed graph consists of nodes and directed arcs. The topology of the graph can be categorized as either acyclic or cyclic. A neural network with acyclic topology represents a function of its current input; thus, it has no internal state other than the weights. No feedback connection is present from units in one layer to those in a previous layer or the same layer. Such an acyclic neural network is called a feed-forward network, which is often used to approximate a nonlinear mapping between its input and output. A neural network with cyclic topology contains at least one cycle formed by directed arcs. Such a neural network is known as a recurrent network. A recurrent network feeds its outputs back into its own inputs.

The most popular neural network used in the application of engineering problems is a multi-layer feed-forward network. Although a neural network can have any number of layers, the universal approximation theorem proves that only one layer of hidden units with non-linear activation functions is enough to approximate any function with finitely many discontinuities of an arbitrary degree of precision [59]. Hence, in most applications, a three-layer feed-forward network is used, consisting of an input layer, a hidden layer, and an output layer.

4.2 Multilayer Feed-Forward Network

A network with all inputs connected directly to the outputs is called a single-layer neural network, or a perceptron network. Multilayer perceptron (MLP) neural networks model consists of a feed-forward, layered network of McCulloch and Pitts' neuron. Each neuron in an MLP has a nonlinear activation function that is often continuously differentiable. Some of the most frequently used activation functions for MLP include the sigmoid function and the hyperbolic tangent function. The advantage of adding hidden layers is that it enlarges the space of hypotheses that the network can represent. A typical MLP configuration is depicted in Figure 4.2.

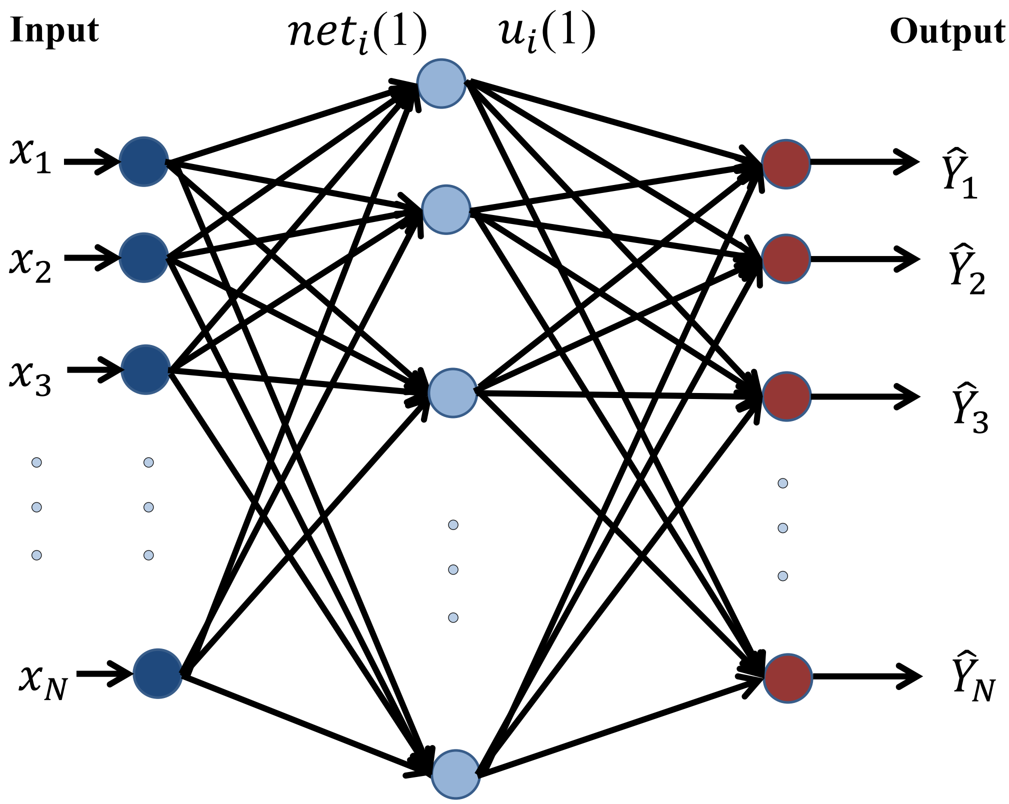


Figure 4.2 A Three-layer MLP Configuration

The performance of a neural network is highly dependent on the training algorithm. A well-trained neural network has minimal error in training data and, thus, is able to accurately approximate the targets. An adequate learning method is needed to obtain such a network. Learning rules can be grouped into two distinct types: supervised learning and unsupervised learning [66]. In supervised learning, the inputs and corresponding outputs are provided to the network from outside. Supervised learning allows the network to adjust weights based on differences between network outputs and provided outputs. Several popular supervised learning methods exist, among which back-propagation is the most common. Unsupervised training is also called self-organized learning in which no matching outputs are provided and output units must independently make sense of the inputs. Unsupervised learning varies from supervised learning because no existing groups are present into which the inputs are classified, and the network discovers the feature within the inputs [66]. Unsupervised learning is commonly used in data mining in which large sets of data and features of data are not known.

4.2.1 Back-Propagation Training Algorithm

One supervised learning method, back-propagation, is routinely used in many neural network applications in which the connection weights are adjusted to minimize the error between the output of each output unit and a target output [66]. This process requires the computation of error derivative of the weights, starting from the output layer and moving from layer to layer in a direction opposite to the propagation of data through the network. The term “back-propagation” originates from the fact that the error is propagated back to modify the incoming weights. For a MLP model, the mathematical steps of back-propagation algorithm are given in [59].

The training data set consists of N training patterns $\{(x_i, y_i)\}$, where i is the pattern number. The input vector, x_i, and desired output vector, y_i, have dimensions M. $\widehat{y_\iota}$ is the network output vector for the i^{th} pattern. The thresholds are handled by augmenting the input vector with an element x_0 and setting it equal to one.

For the j^{th} hidden unit, the net input $net_i(j)$ and output activation $u_i(j)$ for the i^{th} training pattern are

$$net_i(j) = \sum_{i=0}^{N} w_{ji} x_i \tag{4.3}$$

where w_{ji} denotes the weight connecting the i^{th} input unit to the j^{th} hidden unit.

$$u_i(j) = f(net_i(j)) \tag{4.4}$$

For MLP networks, a typical activation function f is the sigmoid

$$f(net_i(j)) = \frac{1}{1 + e^{-net_i(j)}} \tag{4.5}$$

The k^{th} ouput for the i^{th} training pattern in $\widehat{y_{\iota k}}$ and is given by

$$\hat{Y}_{ik} = \sum_{i=0}^{N} w_{ji} x_i + \sum_{j=1}^{M} w_{kj} u_j \tag{4.6}$$

where w_{kj} denotes the weight connecting the j^{th} hidden unit to the k^{th} output unit.

The mapping error or sum-squared loss function for the i^{th} pattern is

$$E = \frac{1}{2}\sum_{i=1}^{N}\left(Y_i - \widehat{Y_i}\right)^2 \tag{4.7}$$

In order to update the weights by the gradient descent method, the gradient of the loss function with respect to each weight w_{ji} of the network needs to be computed. According to the chain rule, the gradient can be represented as:

$$\Delta w_{ji} = \frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial net_j}\frac{\partial net_j}{\partial w_{ji}} \tag{4.8}$$

The second factor is actually the output of unit j:

$$\frac{\partial net_j}{\partial w_{ji}} = \frac{\partial}{\partial w_{ji}}\sum w_{ji}x_i = x_j \tag{4.9}$$

In Equation (4.8), the first factor can be considered as two cases: the case where unit j is an output unit for the network, and the case where j is an internal unit.

Case 1: Training rule for output unit weights

The first factor of Equation (4.8), by chain rule can be written as

$$\frac{\partial E}{\partial net_j} = \frac{\partial E}{\partial \widehat{Y_j}}\frac{\partial \widehat{y_j}}{\partial net_j} \tag{4.10}$$

The first term in Equation (4.10)

$$\frac{\partial E}{\partial \widehat{Y_j}} = \frac{\partial}{\partial \widehat{Y_j}}\frac{1}{2}\sum\left(Y_i - \widehat{Y_i}\right)^2 \tag{4.11}$$

The derivates of Equation (4.11) will be zero for all outputs units except i=j.

$$\frac{\partial E}{\partial \widehat{Y}_j} = \frac{\partial}{\partial \widehat{Y}_j}\frac{1}{2}\left(Y_j - \widehat{Y}_j\right)^2$$

$$= \frac{1}{2}\ 2\left(Y_j - \widehat{y}_j\right)\frac{\partial\left(Y_j - \widehat{Y}_j\right)}{\partial Y_j}$$

$$= -\left(Y_j - \widehat{Y}_j\right) \tag{4.12}$$

Now consider the second term in Equation (4.10).

$$\frac{\partial \widehat{Y}_j}{\partial net_j} = \frac{\partial f(net_j)}{\partial net_j} = \widehat{Y}_j\left(1 - \widehat{Y}_j\right) \tag{4.13}$$

Substituting expressions (4.12) and (4.13) into (4.10),

$$\frac{\partial E}{\partial net_j} = -\left(Y_j - \widehat{Y}_j\right)\widehat{Y}_j\left(1 - \widehat{Y}_j\right) \tag{4.14}$$

Combining (4.9) and (4.14) into (4.8),

$$\Delta w_{ji} = \frac{\partial E}{\partial w_{ji}} = -\left(Y_j - \widehat{Y}_j\right)\widehat{Y}_j\left(1 - \widehat{Y}_j\right)x_j \tag{4.15}$$

Case 2: Training rule for hidden unit weights

In the case where j is a hidden unit in the network, derivation of the training rule for w_{ji} must account for the indirect ways in which w_{ji} can influence the network outputs and, hence, error E. All units immediately downstream of unit j in the network are considered.

$$\frac{\partial E}{\partial net_j} = \sum_k \frac{\partial E}{\partial net_k} \frac{\partial net_k}{\partial net_j}$$

$$\frac{\partial E}{\partial net_j} = \sum_k -\delta_k \frac{\partial net_k}{\partial net_j}$$

$$\frac{\partial E}{\partial net_j} = \sum_k -\delta_k \frac{\partial net_k}{\partial \hat{Y}_j} \frac{\partial \hat{Y}_j}{\partial net_j}$$

$$\frac{\partial E}{\partial net_j} = \sum_k -\delta_k \, w_{kj} \frac{\partial \hat{Y}_j}{\partial net_j}$$

$$\frac{\partial E}{\partial net_j} = \sum_k -\delta_k \, w_{kj} \hat{Y}_j \left(1 - \hat{Y}_j\right) \tag{4.16}$$

Combining (4.9) and (4.16) into (4.8),

$$\Delta w_{ji} = \frac{\partial E}{\partial w_{ji}} = -\delta_j x_j \tag{4.17}$$

where,

$$\delta_j = \hat{Y}_j \left(1 - \hat{Y}_j\right) \sum_k \delta_k \, w_{kj}$$

At last, weights are updated by the gradient which has been computed,

$$w_{ji} = w_{ji} + \Delta w_{ji} \tag{4.18}$$

A well-trained network can converge to a stable solution, but, unfortunately, divergence can occur during the learning procedure. To prevent divergence, a learning rate ε is introduced to the weight-update scheme. Learning rate is a constant value used to multiply the gradient [59]. By choosing a different value for the learning rate, the increment amount for weight-update at each step can be controlled. Divergence occurs when ε is too large, and the algorithm misses the optimal solution and oscillate. If the learning rate is too small, though, the algorithm is less efficient since it will take a long time for the algorithm to converge. The most suitable value for a learning rate is the largest one without causing oscillation.

$$w_{ji} = w_{ji} + \varepsilon\, \Delta w_{ji} \tag{4.19}$$

The aim of training a neural network with training data is to learn the complex relation between input and output data and generalize this knowledge to new data. The training algorithm must be performed many times in order to completely acquire the knowledge and back-propagation algorithm to converge. However, overtraining happens if the number of iteration is too large because the network memorizes every detail of the training set or error and noise and, thus, is not able to generalize new data. In practical cases, it results in a network that has very small error on the training set, but large error on new test data.

4.3 Model Construction

The most commonly used three-layer, feed forward neural network topology, which is able to adequately approximate nonlinear functions with sufficient accuracy is considered here [59]. The network has a single hidden layer with sigmoid activation functions and is trained in the batch mode according to the error back-propagation algorithm with gradient decent. Lightning and wind speed are two feature-related inputs to the neural network model.

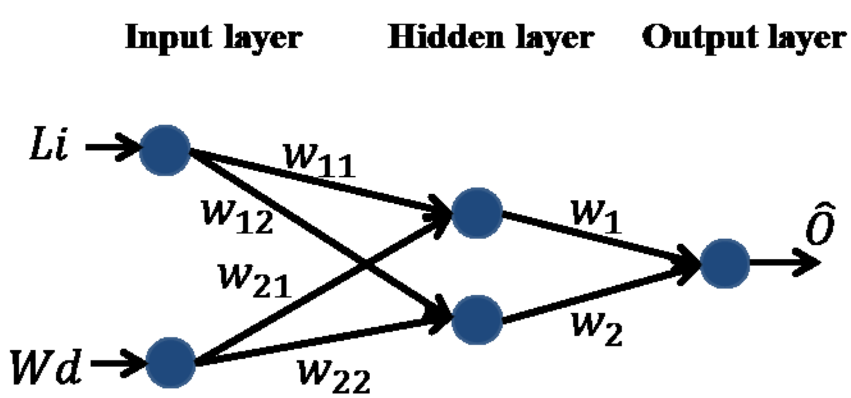


Figure 4.3 Three-layer Feed-forward NN Model for Weather-related Outages

No theoretical guidance exists for choosing the number of neurons in the hidden layer. From all applications of NN, experience shows that preference goes to the structure in which fewer neurons are present in the hidden layer rather than neurons in the input layer. The number of hidden neurons typically is half the total number of input neurons and output neurons. When applying neural networks to time series estimation, the number of output neurons is very important [66]. As suggested in [66], the number of required targets was minimized in the weather-related outage estimation of this research. The NN model structure for the estimation is shown in Figure 4.3. Input and target data are normalized between 0.1-0.9, but the desired output must never be set to 0 or 1 because whatever the inputs, the node outputs in the hidden layer must remain between 0 and 1 (these values are asymptotes of the function). Approaching these values requires enormous weights and/or input values, and most importantly, they cannot be exceeded. By contrast, setting a desired output of 0.9, for example, allows the network to approach and ultimately reach this value from either side. Experiments show that the learning rate is 0.5 and optimum training times are 3000.

4.4 Simulations and Model Performance

The model was trained with historical data for the four cities from 2005 to 2009 and tested for 2010 to 2011. Performances of the model were measured using the average absolute error (MAE) and mean square error (MSE) given in Table 4.2. Results showed increasing MSE and MAE with increasing city size, which does not necessarily mean the model works better for smaller cities because the outages have a greater range in bigger cities.

Table 4.2 Results of Neural Network Model for Four Cities

Neural Network								
	Training Data				Test Data			
	MSE	MAE	R	S	MSE	MAE	R	S
Manhattan	2.4879	0.6009	0.6761	0.4555	2.3370	0.6433	0.4335	0.1254
Lawrence	4.3621	0.6973	0.2958	0.0871	5.0176	0.8778	0.4130	0.1012
Topeka	12.9613	1.3913	0.5494	0.3016	37.1506	2.4418	0.4231	0.1909
Wichita	35.9343	2.8051	0.5756	0.3312	76.7436	3.2873	0.4003	0.2314

For all cities, however, correlation coefficients are positive, indicating a positive linear relationship between estimated outages and observed outages. Because the correlation coefficients are much smaller than 1, the estimated outages cannot accurately follow observed outages for Wichita, Topeka, Lawrence and Manhattan. Scatter plots of daily observed outages and estimated outages for four cities for training and testing duration are shown in Figure 4.4. The NN model presented here are applied for the estimation of outages caused by weather considering wind and lightning stroke within 200m and 400m around the feeder as weather variables [49].

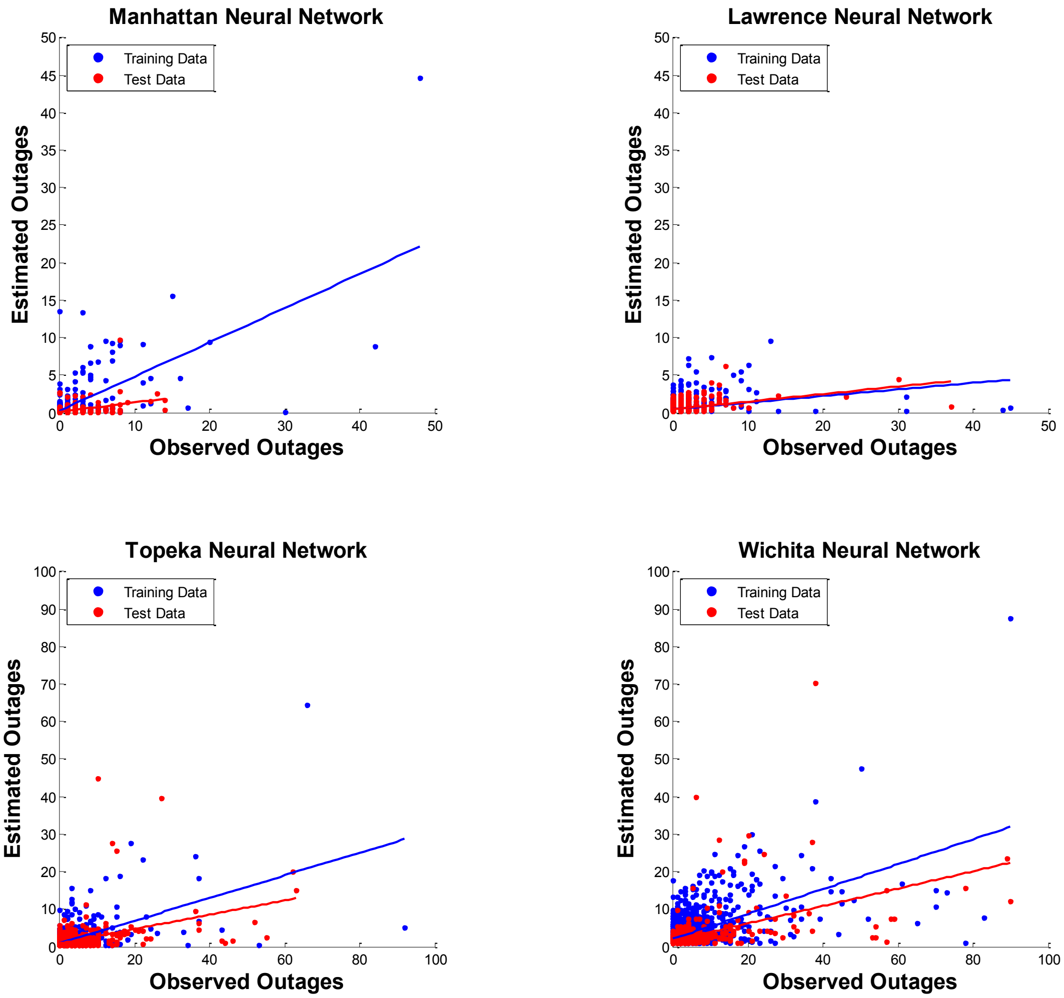


Figure 4.4 Plot of Observed and Estimated Weather-related Outages Obtained with Neural Network Model in Overhead Distribution Systems for Four Cities from 2005 to 2011

4.5 Summary

The neural network model is able to approximate complex relations between the input and output, causing it to outperform the traditional regression models. Although, the performance measure values are slightly better than the regression model 4, the NN model also overestimates and underestimates the outages for the lower and higher range of observed values. The performance of the model can be improved with the use of machine learning algorithms based on ensemble systems. In an ensemble system, multiple networks are trained using identical data but each network is initialized with different random weights. The network output can be weighted average of individual network output or best performing individual network's output.

Chapter 5 - Committee Machines

Two key problems associated with time series modeling of weather-related outages are noise and non-stationarity in the data which could lead to overestimation and/or underestimation. A potential solution to the above problems is to utilize Committee Machine (CM) architecture [50] inspired by the "divide-and-conquer" principle often used to attack a complex problem by dividing it into simpler problems whose solutions are combined to yield a solution to the complex problem. The motivation of the CM is that individual expert networks can focus on specific regions and attack them well.

5.1 Concept of a Committee

Noisy characteristics of the data refer to the unavailability of complete information from past behavior of the time series in order to fully capture dependency between the future and the past. The noise in the data could lead to over-fitting or under-fitting so that the obtained model has a poor level of performance when applied to new data patterns. Non-stationarity implies that the time series switches dynamically between regions, leading to gradual changes in dependency between input and output variables. In general, it is hard for a single model to capture such a dynamic input–output relationship inherent in the data and causes over-fitting or under-fitting. Apart from random noise, the expected error of a trained model for test data consists *bias*, and *variance* components. The component, bias, refers to topological inadequateness when modeling the data. It can be reduced by increasing network complexity, such as adding additional parameters. For example, as more hidden neurons are added to a neural network, complexity of the resulting model complexity will be greater. Unfortunately, increased parameter also leads to higher variance, or sensitivity of network parameters to the training samples, thus creating the third source of error. In other words, increasing model complexity to improve training sample performance 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.

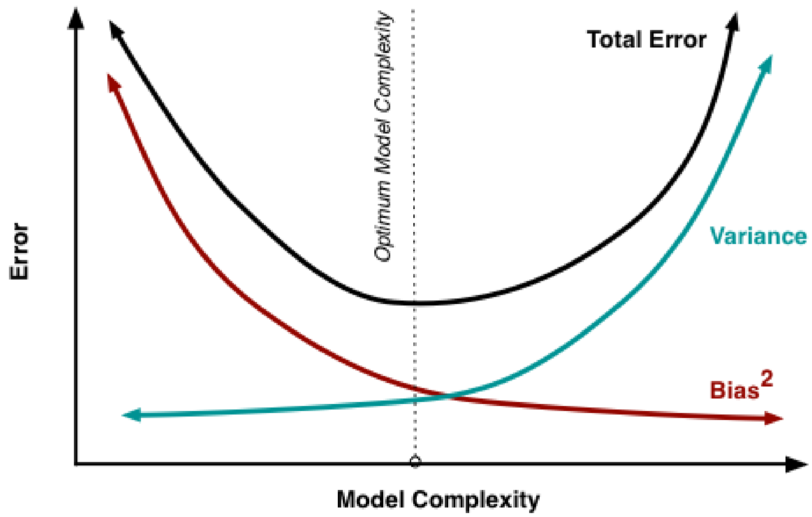


Figure 5.1 Bias-Variance Trade-off as a Function of Model Complexity [50]

Understanding bias and variance is critical for understanding the behavior of estimation models. However, in general, overall error is the greater concern, not specific decomposition. The optimal point for any model is the level of complexity at which the increase in bias is equivalent to the reduction in variance. If the model complexity exceeds this optimal point, the model is over-fitting; while if complexity falls short of the optimal point, the model is under-fitting. Ensemble methods are proved to be very effective technique in reducing bias and variance. In the literature, a plethora of terms other than ensembles has been used, such as fusion, combination, aggregation, committee, to indicate sets of learning machines that work together to solve a machine learning problem, but here the term committee or ensemble is used. Likewise, each learning algorithm has plethora of terms such as base models, elementary units, experts, weak learner, and base learner.

A complex computational task is solved by dividing the task into a number of computationally simple tasks and then combining the solutions to those tasks. In supervised learning, computational simplicity is achieved by distributing the learning task among a number of experts, who divide the input space into a set of subspaces. The combination of experts is said to constitute a committee machine because it fuses knowledge acquired by experts to arrive at an overall decision that is supposedly superior to that attainable by any one of them acting alone. The idea of a committee machine originated with Nilsson (1965); the network structure considered therein consisted of a layer of elementary units followed by a vote-taking unit in the second layer.

Potential benefits of the CM include:

1. Better overall performance.
2. Reuse of existing pattern classification expertise.
3. Heterogeneity
    - Expert classifiers need not be of the same type.
    - Different features can be used for different classifiers.
4. Anonymity: black box, proprietary expert classifiers can be used.

5.1.1 Types of CM

CM is comprised of multiple experts which are strategically generated and combined to solve a particular computational problem. Committee machines are classified into two major categories:

- **Static structures.** In this class of committee machines, the responses of several predictors (experts) are combined by means of a mechanism that does not involve the input signal, hence the designation "static." Examples include Averaging, Boosting, voting.
- **Dynamic structures.** In this second class of committee machines, the input signal is directly involved in actuating the mechanism that integrates the outputs of individual experts into overall outputs, hence designation "dynamic." [50]. Examples include mixture of experts, hierarchical mixture of experts.

5.1.2 Base-Learner Selection and Combining Outputs

Three basic questions that arise while generating ensemble model:

1. How to choose a base-learners among many competing models?
2. Given a particular learning algorithm, which realization of this algorithm should be chosen?
3. How to combine the outputs of base-learners for maximum accuracy?

Each learning algorithm dictates a certain model that comes with a set of assumptions. This inductive bias leads to error if the assumptions do not hold for the data. Since learning is an ill-posed problem and with finite data, each algorithm converges to a different solution and fails under different circumstances. Learner performance may be fine-tuned to achieve the highest possible accuracy on a validation set, but this fine tuning is a complex task and instances still occur in which even the best learner is not accurate enough. Some of the choices are:

- Match the assumptions for particular model to what is known about the problem, or
- Try several model and choose the one that performs the best, or
- Use several models and allow each sub-result to contribute to the final result.

Several base learners, such as multilayer perceptron [80, 83], support vector machine (SVM) [84], Gaussian Process (GP) [85], Gaussian Mixture Models (GMM) [86], Hidden Markov Models (HMM) [87], kernel-based models [88], radial basis function networks [89], decision trees [90], fuzzy logic [91], ARMA models [101] and Bayesian models [102, 103], exist.

The No Free Lunch Theorem [72] states that there is no single learning algorithm in any domain that always induces the most accurate learner. From literature, the usual approach is to try many and choose the one that performs best on a separate validation set. Various learning algorithms can be used to train different base-learners, since algorithms make different assumptions about the data and lead to different base learners, or the same learning algorithm can be used with different hyper-parameters. Some of possible choices are:

- Use same base-learner with different input representation, such as sensor fusion and random subspace [72].
- Use different base-learners by different subsets of the training set by drawing random training sets from the given sample; this is called bagging.
- Train the base-learner serially so that when preceding base-learners were not accurate, they are given more emphasis in training later base-learners, such as boosting and cascading.
- Use base-learners on each region by partition the input space, such as a mixture of experts.

The final output can be obtained in many different ways from the outputs of multiple base-learners. In multi-expert combination method, base-learners work in parallel and the final output can be generated in two ways:

- In the global approach, all base-learners are given an input, they generate an output and all these outputs are used, such as voting and averaging.
- In the local approach, a gating model is present which looks at the input and selects one (or very few) learners as responsible for generating the output.

In multistage combination method, a serial approach is used in which the next base-learner is trained with training samples where previous base-learners are not accurate enough. In this dissertation, boosting and mixture of experts are considered.

5.2 Boosting

Boosting is a powerful technique for combining multiple base learners in order to produce a form of committee whose performance can be significantly better than any base learner. The original boosting algorithm was developed by Schapire [104]. Structural illustration is given in Figure 5.1. In boosting, learners are trained sequentially and the training of a particular learner is dependent on the training and performance of previously trained learners. Each data point is associated with weighting coefficient which is dependent on past learner performance. In particular, points that are misclassified by one of the learners are given more weight when used to train the next learner in the sequence. Once all learners have been trained, their outputs are combined through a combining rule. Since more emphasis is put on data points misclassified by previously trained learners, boosting reduces both variance and bias [72]. In this dissertation, the most widely used boosting algorithm, AdaBoost, is used. Freund and Schapire proposed AdaBoost algorithm short for adaptive boosting in 1995 [104].

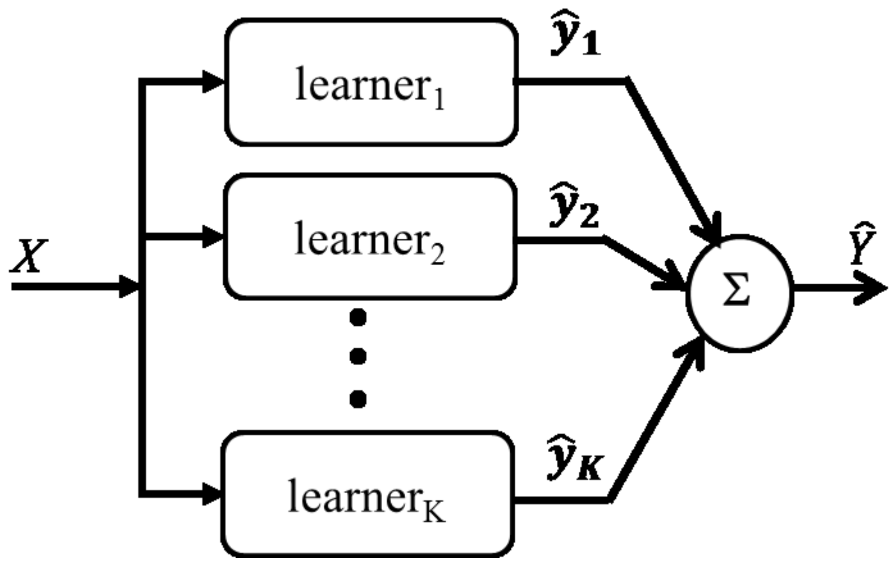


Figure 5.2 Structural Representation of Boosting

5.2.1 AdaBoost Model

The training data comprises input vector $X = \{x_1, x_2, \dots \dots .., x_N\}$ and desired output vector $Y = \{y_1, y_2, \dots \dots .., y_N\}$. Each data point is given an associated weighting parameter, $d_k; k = 1,2, \dots, K$, which is initially set $1/_N$ for all data points. Each learner is trained using associated weights to give output, $\widehat{\boldsymbol{y}_k}$ by minimizing the error function given as

$$E_k = \frac{1}{2}\sum_{i=1}^{N} d_k(i)(Y(i) - \widehat{\boldsymbol{y}_k}(i))^2 \tag{5.1}$$

At each stage of the algorithm, a new learner is trained using a data set in which the weighting coefficients are adjusted according to performance of the previously trained learner so as to give more weight to misclassified data points. The learner output for this sample is considered to be error-free when the absolute relative error lies within threshold, θ

$$\frac{|\widehat{\boldsymbol{y}_k}(i) - Y(i)|}{Y(i)} \leq \theta \tag{5.2}$$

The new weights $d_{k+1}(i)$ are determined from the prior $d_k(i)$ in accordance

$$d_{k+1}(i) = \begin{cases} d_k(i)\delta_k, & \frac{|\widehat{\boldsymbol{y}_k}(i) - Y(i)|}{Y(i)} \le \theta, \\ d_k(i), & \frac{|\widehat{\boldsymbol{y}_k}(i) - Y(i)|}{Y(i)} > \theta. \end{cases} \tag{5.3}$$

From (5.3), results show that in subsequent iterations, weighting coefficients $d_k(i)$ are increased for data points that have larger error than threshold, θ. Therefore, successive learners are forced to emphasize points that have larger estimation error by previous learners, and data points that continue to have higher error by successive learners receive increasing weight.

The quantity δ_k in (5.3) is the error rate produced by the k^{th} network at the end of its training with $\{\boldsymbol{x}(i), y(i)\} \sim d_k(i)$. Using (5.2) as the criterion for a sample to be error-free, the set of erroneous samples is,

$$h_k = \left\{ i \middle| \frac{|\widehat{\boldsymbol{y}_k}(i) - Y(i)|}{Y(i)} > \theta \right\} \tag{5.4}$$

Hence, the network error rate is given by,

$$\delta_k = \sum_{i \in h_k} d_k(i) \tag{5.5}$$

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

$$d_{k+1}(n) = \frac{d_{k+1}(i)}{\sum_i d_{k+1}(i)} \tag{5.6}$$

Following normalization, the weights add up to unity,

$$\sum_i d_{k+1}(i) = 1 \tag{5.7}$$

Finally, when the desired number of base learners has been trained, they are combined to form a committee using coefficients that give different weight to different base learners.

The algorithms, AdaBoost.RT and AdaBoost$^+$, differ in how the ensemble output is determined. In AdaBoost.RT, the ensemble output $\hat{Y}(i)$ is the weighted sum of all K learners, with the learners receiving weights proportional to the logarithm of their inverse error rates. Accordingly, the ensemble output by AdaBoost.RT is,

$$\hat{Y}(i) = \sum_{k} \left(\frac{\log \frac{1}{\delta_k}}{\sum_k \log \frac{1}{\delta_k}} \right) \times \hat{y}_k(i) \tag{5.8}$$

In AdaBoost+, the weights are determined to explicitly minimize the sum of the squared errors of all samples. The sum squared error can be expressed as,

$$E = \sum_{i} \left(Y(i) - \hat{Y}(i) \right)^2 = \left(\mathrm{Y} - \hat{\mathrm{Y}} \right)^{\mathrm{T}} \left(\mathrm{Y} - \hat{\mathrm{Y}} \right) \tag{5.9}$$

where, Y and $\hat{\mathrm{Y}}$ are the desired output and the ensemble output of size $N \times 1$.

Outputs of each network, $\hat{\mathbf{y}}_k$ can be organized as an $N \times 1$ vector. Defining the $N \times K$ output matrix $\hat{\mathbf{y}} = [\hat{\boldsymbol{y}}_1 \ldots \hat{\boldsymbol{y}}_{\boldsymbol{K}}]$, the output vector $\hat{\mathbf{Y}}$ can be expressed as,

$$\hat{\mathrm{Y}} = \hat{\mathbf{y}}(\hat{\mathbf{y}}^{\mathrm{T}}\hat{\mathbf{y}})^{-1}\hat{\mathbf{y}}^{\mathrm{T}}\mathrm{Y} \tag{5.10}$$

Regularization can be incorporated for numerical stability of the matrix inversion, in which case (5.10) can be obtained as follows, with a being a small constant,

$$\hat{\mathrm{Y}} = \hat{\mathbf{y}}(\hat{\mathbf{y}}^{\mathrm{T}}\hat{\mathbf{y}} + a\mathbf{I})^{-1}\hat{\mathbf{y}}^{\mathrm{T}}\mathrm{Y} \tag{5.11}$$

Actual boosting performance on a particular problem is clearly dependent on the data and the base-learner. Enough training data should be available and the base-learner should be weak but not too weak. Boosting is especially susceptible to noise and outliers.

5.3 Mixture of Experts

The original ME model, introduced by Jacobs et al. [111] in 1991, can be viewed as a tree-structured architecture based on the principle of "divide and conquer", having three main components: several experts that are either regression functions or classifiers; a gate that makes soft partitions of the input space and defines regions where individual expert opinions are trustworthy; and a probabilistic model to combine the experts and the gate. Structural representation of ME is shown in Figure 5.2. The model is a weighted sum of experts, where the weights are the input-dependent gates. In this simplified form, the original ME model has three important properties: 1) it allows individual experts to specialize in smaller parts of a larger problem; 2) it uses soft partitions of the data; and 3) it allows splits to be formed along hyperplanes at arbitrary orientations in the input space [112].

Some of models, such as support vector machines [114], Gaussian process (GP) [115], hidden markov models [116, 117], and Bayesian [118-121] are used as base-learners in the literature. Different inference techniques, such as fuzzy c-means [91], EM-based methods like IRLS [122], generalized EM [127], single loop EM [128], Newton-Raphson [130], deterministic annealing (DA) [133], and Bayesian inference [136], are used to train experts for faster convergence rates and learn the arameters of these models. Gaussian mixture model GMM [138], softmax of GPs [133], Dirichlet distribution [134], Dirichlet process (DP) [135], max/min networks [139], and neural networks (NNs) [140] are expert models for the gating.

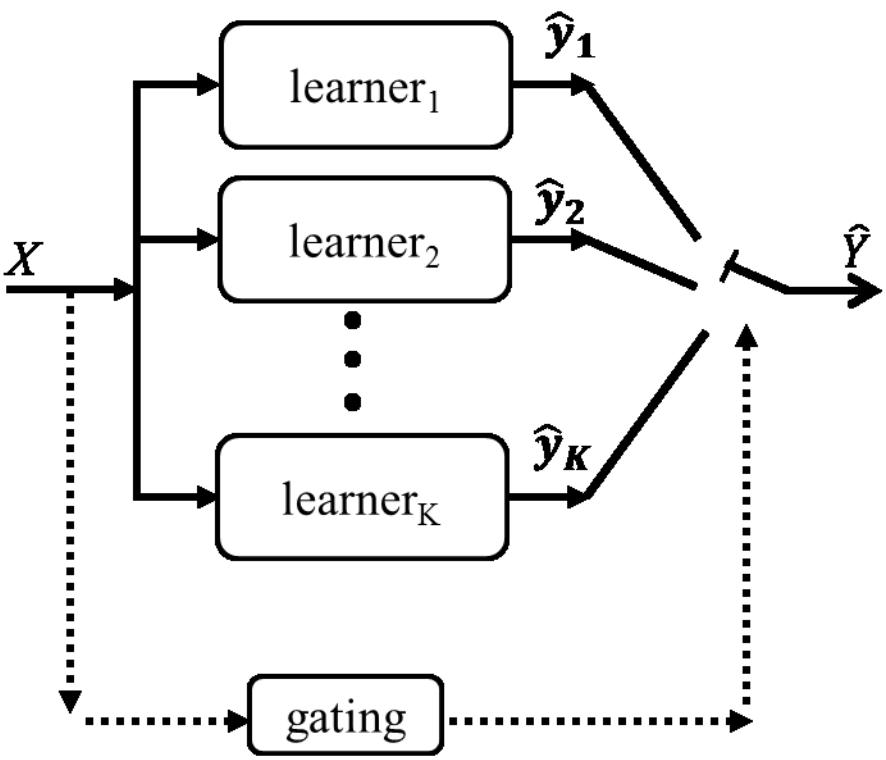


Figure 5.3 Mixture of Experts Architecture

5.3.1 Mean Field Annealing

Mean field annealing is an optimization approach developed for use in statistical mechanics. Statistical mechanics is an area of physics which describes the slow Ising Hamiltonian process of thermal cooling for spin articles with many degrees of freedom until reaching equilibrium states [141]. Annealing particles in solids provide a framework for optimization of the properties of very large and complex systems. This idea is now incorporated into algorithms for solving several prototype combinatorial optimization problems. The annealing models degrees of freedom of collection of atoms slowly being cooled into a ground state corresponding to the optimal solution to the problem with the temperature *T* as the controlling parameter. The energy surface, defined as *E(s)* for a particle state, *s*, is a Boltzmann distribution function that allows changes in s to increase *E,* thus providing the network with a mechanism to escape from entrapment in a local minimum.

The relaxation is done according to the Boltzmann distribution

$$P(S) = \frac{e^{-E(S)}/T}{Z} \tag{5.12}$$

where,

S is any one possible configuration specified by the corresponding expert set

E(S) is the energy of the corresponding configuration;

T is the temperature;

Z is the partition function given by

$$Z = \sum_{S} e^{-E(S)}/T \tag{5.13}$$

and the summation covers all possible expert configurations.

The efficacy of method depends on a proper choice of temperature T. The goal is to spend most relaxation time around the critical temperature T_c, where global minima begin to be noticeable (i.e., when escaping from deeper minima begins to be significantly more difficult). For T >> T_c, the system evolves randomly, whereas for T << T_c, the system 'freezes' in a local minimum. The usual procedure is to start at a sufficiently high initial temperature T_0 and decrease slowly until some final temperature T_f .

5.3.2 AME Model

The training data comprises input vector $X = \{x_1, x_2, \dots \dots .., x_N\}$ and desired output vector $Y = \{y_1, y_2, \dots \dots .., y_N\}$. The learning algorithm used by the AME localizes the base-learners such that each becomes an expert in a different part of the input space and has its weight, $w_k(i)$ close to 1 in its region of expertise.

$$w_k(i) = \frac{1}{K} \; ; k = 1,2, \dots K; i = 1,2, \dots N \tag{5.14}$$

Suitable initial values for the thermostatic temperature T_0 and cooling parameter γ are also chosen. In first iteration, all experts receive an equal amount of training for each sample. Each learner is trained using associated weights to give output, $\widehat{y_k}$. The committee is trained to find optimal combination weights to minimize the mean squared error between the desired and the expert output with respect to the training data distribution. The combination weights is a linear combination of the estimators based on the empirical MSE and defined as

$$E_k = \frac{1}{2} \sum_{i=1}^{N} w_k(i) * \left(e_k(i)\right)^2 \tag{5.15}$$

where the w_k satisfies the constraint that $\sum w_k = 1$. Choose w_k*'s* so as to minimize the MSE with respect to the desired output, Y. The error $e_k(i)$of sample i is defined as

$$e_k(i) = \left(\frac{\widehat{\boldsymbol{y}}_k(i) - Y_k(i)}{1 + Y_k(i)} \right) \tag{5.16}$$

where $Y_k(i)$ denotes the desired output for the input *i* and $\hat{Y}_k(i)$ the output of kth member of the AME.

The new weights w_k are determined as

$$w_k(i) = 1 - \left[\frac{e^{-\frac{e_k(i)}{T}}}{Z} \right] \tag{5.17}$$

where,

$$Z = \sum_{k=1}^{K} \left(e^{-\frac{e_k(i)}{T}} \right) \tag{5.18}$$

In our earlier work, the initial ME model, the weight update rule is just the Boltzmann distribution and details can be found in [52].

The weighted MSE is computed w.r.t the new weights. At every cycle, the weighted MSE change is observed closely as it determines whether the selected neuron updates its values or not. The next step is to repeat annealing. The temperature T is reduced according to $T \leftarrow \gamma * T$ for O< γ <l, and repeated until the system is stabilized. As the temperature is lowered, a phase transition is passed at $T = T_c$ and as $T \to 0$ fixed points w_k^* emerge, representing a specific decision made as to the solution. The fixed points are characterized by

$$\sum_{k=1}^{K} w_k^* = 1 \tag{5.19}$$

When no significant change occurs in the weighted MSE, the neurons are said to have stabilized at the current temperature. This entire process is repeated for several iterations until a stopping criteria is reached. The temperature first exhibiting this observation is called the *critical temperature* T_c, and MFA is said to have reached its equilibrium.

The position of T_c depends on T_0, γ, and w. Setting initial parameters for annealing techniques has always been troublesome. Beginning at too high a temperature is wasteful since no progress is made toward a solution until the critical temperature is reached. Beginning at a low temperature, however, can quench the system and quickly force it into a poor solution.

At high temperatures, the clustering/partitioning is maximally disordered. As the temperature lowers, a critical temperature T_c is reached where each node begins to move predominantly into one or another of the clusters. At sufficient low temperatures, the MSE saturates, completing the clustering process.

The output of the AME architecture is determined by the gating network, given by

$$\hat{Y}(i) = \hat{\boldsymbol{y}}_k(i) | \; k = \max\left(w_k(i)\right) \; or \; min\left(e_k(i)\right) \tag{5.20}$$

The model output is the output of the individual expert which ever gives the minimum error or have the maximum weight.

The initial investigation of the ME model to study the weather impact on power outages are presented and published [52].

5.4 Summary

In this chapter, an overview and the practical reasons of using committee machines were discussed. Also presents, the different approaches for selection of base-learner and types of combining outputs of multiple learners to generate the final output. Two distinct static and dynamic structures, boosting and mixture of experts were discussed. In boosting, each case the outputs from the base-learner constituent networks are combined with no reference to the inputs. The construction of AdaBoost.RT and AdaBoost$^+$ models were discussed. In mixtures of experts, the outputs of the base-learners are combined by gates which learn appropriate dependencies on the inputs. The mean field theory is used to learn the learning algorithm parameters in the AME. Application of the three proposed methods, AdaBoost.RT, AdaBoost$^+$ and AME are presented in the following chapter.

Chapter 6 - Application of CM Models for Estimation of Weather-related Outages

In this chapter, the AdaBoost and AME models presented in Chapter 5 are applied for the estimation of weather-related outages. The model construction of AdaBoost and AME models are also presented and the model performance is investigated by simulating the model with available data.

6.1 AdaBoost Model

In the proposed algorithm, four base-learners are considered. The multilayer neural network discussed in Chapter 4 and represented in Figure 4.3 is considered as the base-learner. All networks have an identical number of layers and neurons. AdaBoost structural representation is shown in Figure 6.1.

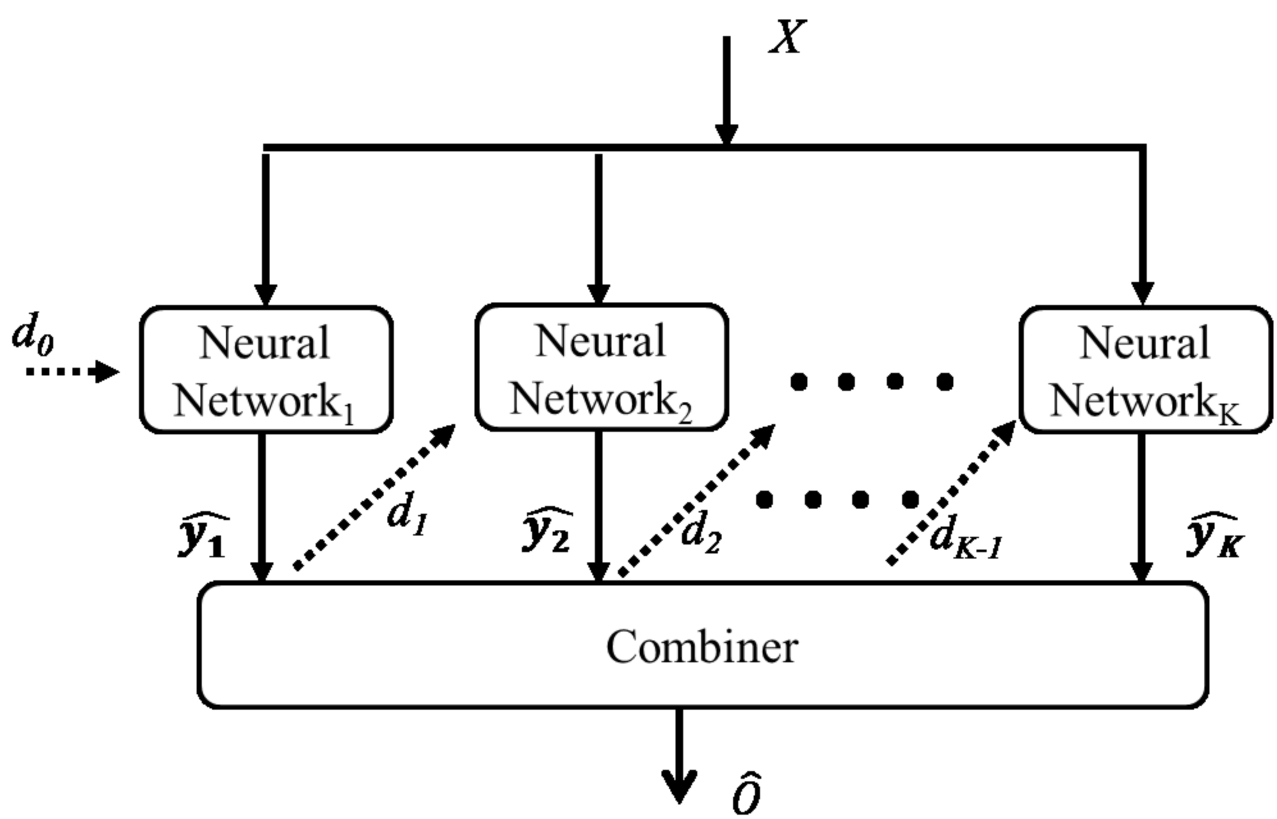


Figure 6.1 Structure of AdaBoost Model for Weather-related Outage Estimation

Input

- Training data $X = \{li, wd\}$ and $Y = \{O\}$.
- K is the total number of base-learners.

Initialize

Assign initial distribution d_k^0 to data points

AdaBoost Learning

For each neural network k=1 to K do

- Train network k
- Compute error rate δ_k using (5.5).
- Compute distribution d_{k+1} using (5.3).
- Normalize distribution d_{k+1} using (5.6).
- Add network k to ensemble.

End

Output

AdaBoost.RT: compute ensemble output using (5.8).

AdaBoost$^+$: compute ensemble output using (5.11).

The models were trained with historical data for the four cities, Manhattan, Lawrence, Topeka and Wichita from 2005 to 2009 and tested for 2010 and 2011. The learning algorithm for the proposed architecture is outlined above. The lightning stroke, wind gust and observed outages are given as input to the model. All data points are associated with distribution, d. Initially, the distribution, d for all training samples are given equal value, 1. The number of base-learners, K considered is 4. At each stage of the algorithm, a base-learner is trained using training data with associated distribution. Based on the performance of the previous learner, the distribution is updated using equation (5.3) and (5.5). To have a probability distribution, the distribution is normalized using equation (5.6). When K number of base learners has been trained, the final outputs are computed using equation (5.8) and (5.11) for AdaBoost.RT and AdaBoost$^+$ respectively. The performance of AdaBoost$^+$ for Wichita data improved with regularization using a = 0.01 in equation (5.11). For all other cases, regularization was not used since it did not change the results.

Figure 6.2 shows the percentage MSE of AdaBoost.RT and AdaBoost$^+$ compared to the number of networks for the training data set of the four cities. The percentage MSE dropped as the number of networks increased, and it stabilized after a certain number of networks. In Wichita, for example, the percentage MSE dropped to 65% for AdaBoost.RT with four neural networks, whereas for AdaBoost$^+$, the percentage MSE dropped to 43% for the same number of neural networks, clearly illustrating the better performance of AdaBoost$^+$. Since increasing the number of neural networks beyond that did not significantly change the results, further comparison of the models uses results with five networks.

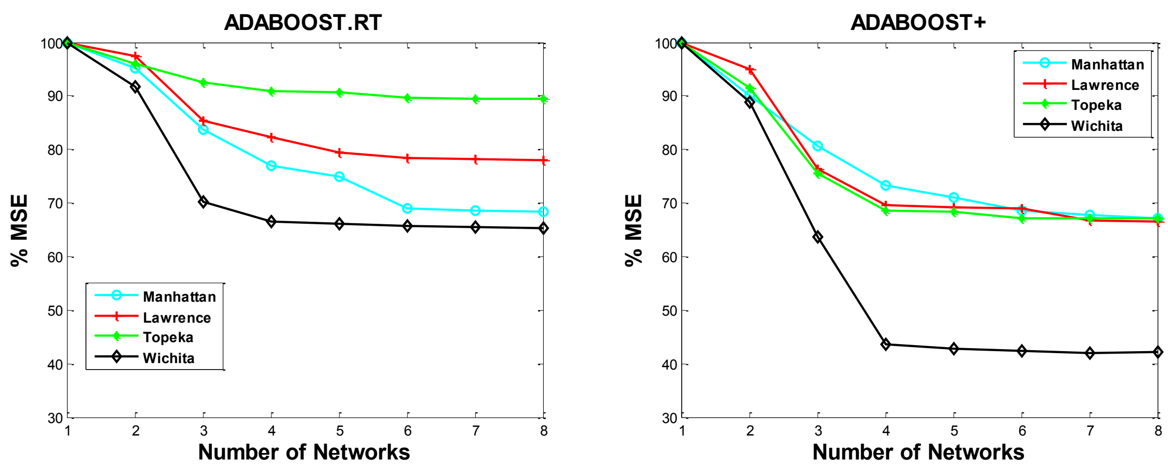


Figure 6.2 The % Mean Square Error as a Function of Number of Learners

The performance of the proposed AdaBoost.RT and AdaBoost$^+$ models are measured using MAE, MSE, R and S, given in Tables 6.1 and 6.2. Comparison of MSE and MAE in Tables 6.1 and 6.2, shows the better performance of AdaBoost$^+$ compared to AdaBoost.RT. Scatter plots of daily observed outages and estimated outages for AdaBoost.RT and AdaBoost$^+$ for four cities for training and testing duration are shown in Figures 6.1 and 6.2. The scatter plots clearly show that the models precisely estimated outages in the lower range, but underestimated outages in the upper range. This can be expected because data in the higher range is sparse and thus the models are not able to fully learn data characteristics in this range.

Table 6.1 Results of AdaBoost.RT Model for Four Cities

AdaBoost.RT								
	Training Data				Test Data			
	MSE	MAE	R	S	MSE	MAE	R	S
Manhattan	1.9163	0.3789	0.7813	0.5082	2.0737	0.5578	0.6216	0.1952
Lawrence	0.3884	3.5923	0.5150	0.2112	3.6619	0.5316	0.6662	0.2792
Topeka	11.7940	0.9053	0.6434	0.3042	32.0706	1.9056	0.7026	0.2095
Wichita	4.9390	1.6712	0.7692	0.4539	49.9527	3.1672	0.4145	0.6401

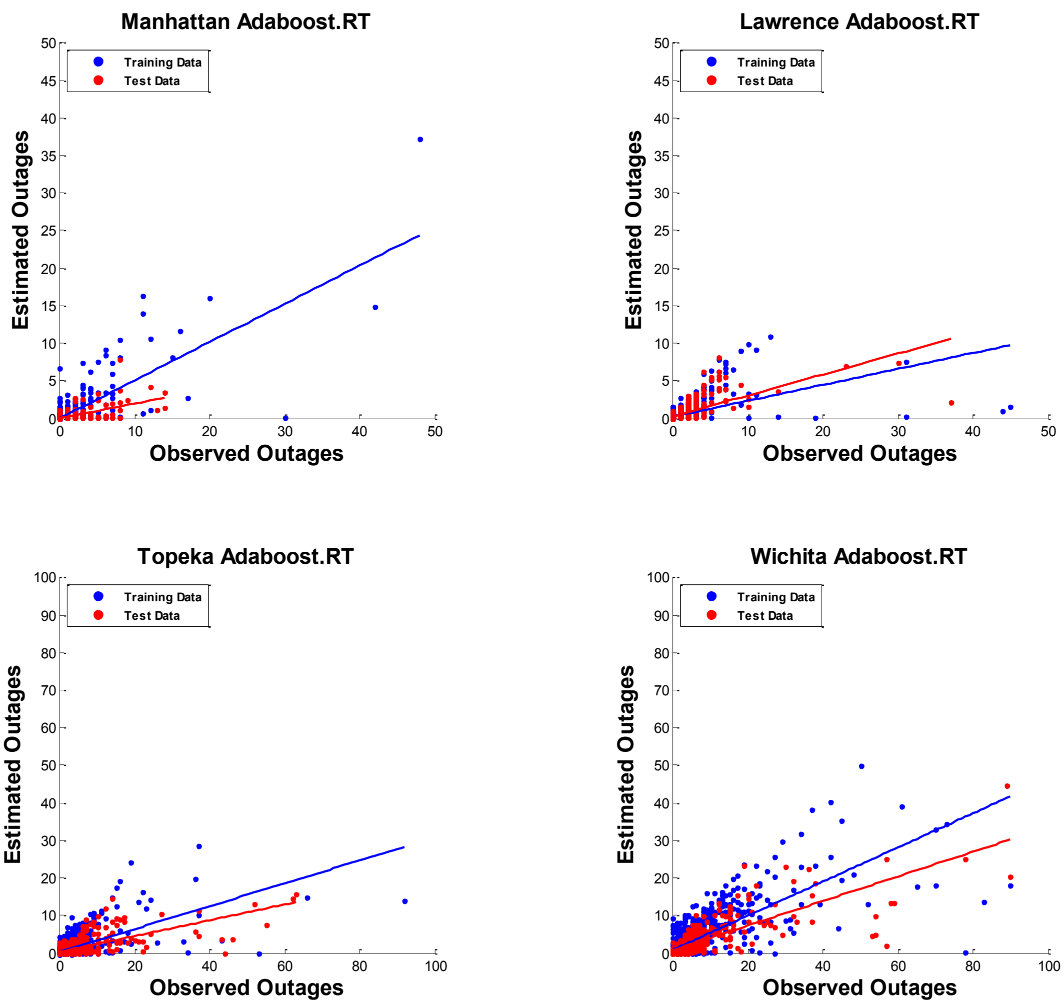


Figure 6.3 Plot of Observed and Estimated Weather-related Outages Obtained with AdaBoost.RT Model in Overhead Distribution Systems for Four Cities from 2005 to 2011

Table 6.2 Results of AdaBoost^{+} Model for Four Cities

AdaBoost^{+}								
	Training Data				Test Data			
	MSE	MAE	R	S	MSE	MAE	R	S
Manhattan	1.8251	0.3691	0.7860	0.5694	2.0558	0.5578	0.6079	0.2083
Lawrence	0.3095	2.6210	0.6947	0.3662	3.0395	0.4340	0.7173	0.3825
Topeka	8.8922	0.7070	0.7448	0.4409	22.9910	1.4621	0.7928	0.3530
Wichita	4.2925	1.4615	0.8263	0.5755	48.0201	2.4409	0.7963	0.3271

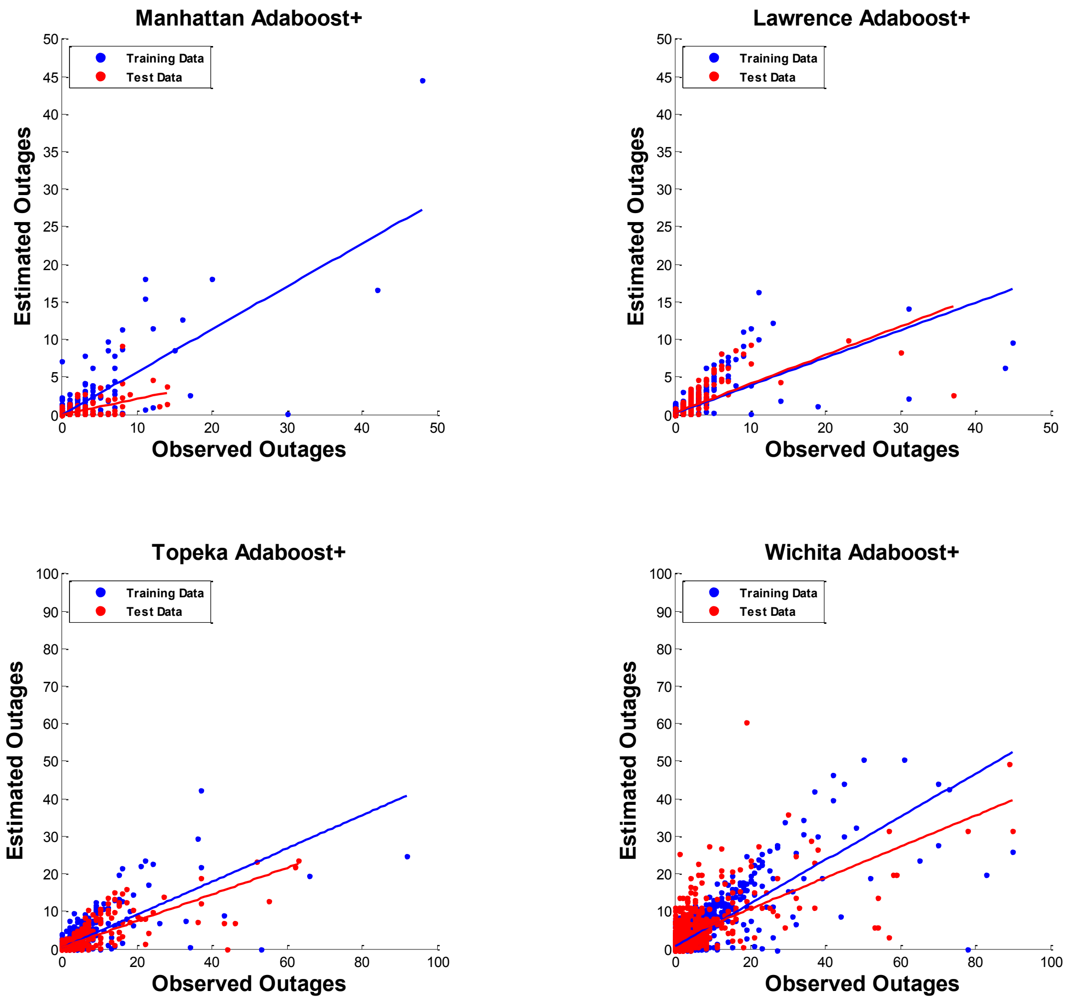


Figure 6.4 Plot of Observed and Estimated Weather-related Outages Obtained with AdaBoost^{+} Model in Overhead Distribution Systems for Four Cities from 2005 to 2011

6.2 Annealed Mixture of Experts (AME) Model

Multilayer neural networks used as base-learners are structurally identical to the neural network model discussed in Chapter 4 and represented in Figure 4.3. All networks have identical number of layers and neurons. Structural representation of AME is shown in Figure 6.5.

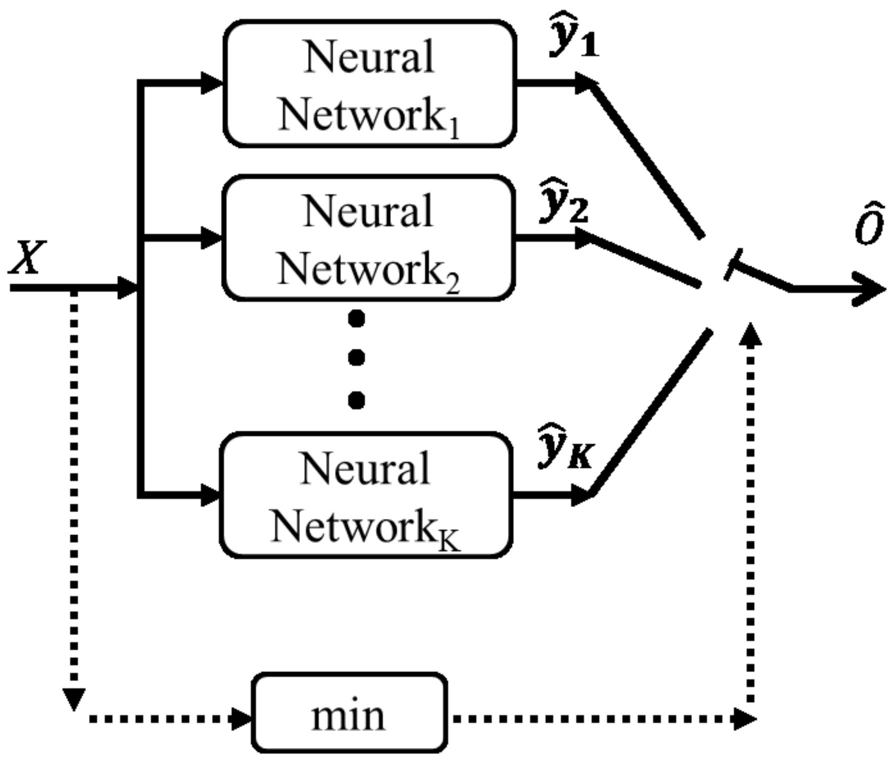


Figure 6.5 Structure of AME for Weather-related Outage Estimation

The models were trained with historical data for the four cities, Manhattan, Lawrence, Topeka and Wichita from 2005 to 2009 and tested for 2010 and 2011. The learning algorithm for the proposed architecture is outlined below. The lightning stroke, wind gust and observed outages are given as input to the model. The number of base-learners, K was set to 4, temperature cooling parameter, γ was set to 0.98, initial temperature T_0 was 1 and maximum number of iteration were 5000.

Input

- Training data $X = \{li, wd\}$ and $Y = \{O\}$.
- K is the total number of base-learners.

Initialize

Assign initial weights w_k^0 to data points

Set initial temperature T_0, cooling parameter γ, max_iteration

Adaboost Learning

For j=1 to max_iteration

For each neural network k=1 to K do

- Train network k
- Compute error rate, e_k using (5.16).
- Compute weights, w_{k+1} using (5.17) and (5.18).
- Add network k to ensemble.

End

Cool temperature using $T \leftarrow \gamma \times T$

End

Output

Compute model output using (5.20).

Each base-learner is trained using training data with associated weights. All training samples are associated with weights, w. Initial weights, w for all training samples are given equal value, 1. In next iteration, a base learner that performs relatively well with any sample input in the past iteration receives increased training with similar samples only through a weight adjustment using equation (5.17). As a result, each learner in the ensemble is trained to "specialize" in only one region of the entire input space. The parameters in the learning algorithm are trained using MFA method. After each iteration the temperature is reduced by the cooling parameter. At high temperatures, the clustering is maximally disordered. As the temperature lowers, a critical temperature T_c is reached where each node begins to move predominantly into one or another of the clusters. At sufficiently low temperatures, the MSE saturates, completing the clustering process. Figure 6.6 shows the drop of MSE as a function of negative logarithm of temperature. The temperature behavior of MFA during estimation is approximately analyzed and shown to possess a critical/curie temperature lying between 1 and 2. It can be observed from Figure 6.6, the MSE saturates once the critical temperature has reached. By experiments, it was found that optimal maximum number of iterations is 5000. Once the maximum number of iterations has reached, the training is stopped and the final output is computed using equation (5.20).

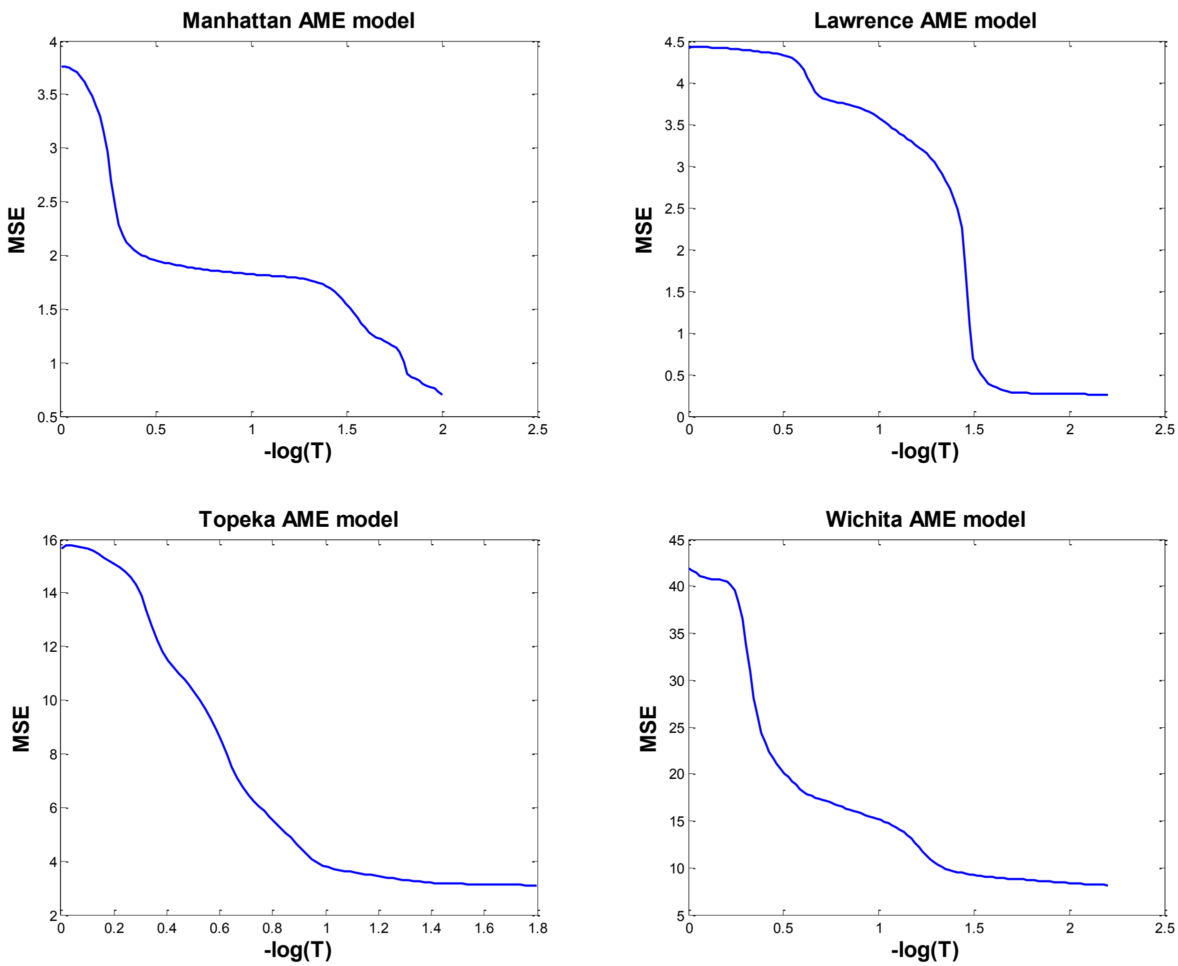


Figure 6.6 Temperature vs. MSE plots Obtained with AME Model in Overhead Distribution Systems for Four Cities

Cluster plots of training data and testing data for the four cities are shown in Figures 6.7 to 6.15. The clustering is shown using the input variables, wind and lightning stroke. Since the aggregate lightning stroke recorded in a day ranges from 0 to tens of thousands of kiloamps, it is hard to show them in a scatter plot along with wind speed, which ranges from 0 to 75 miles per hour. Therefore, natural log of lightning is considered. Since many days have zero recorded lightning, one is added to the lightning values to avoid singularities in the data.

The color code of the input data shows regions in which base learners in the AME model are specialized. The color red, blue, green and black are for cluster 1, cluster 2, cluster 3 and cluster 4. One base-learner performs well on cluster 1 input data points, other on cluster 2 data points and so on. The clustering is done based on the observed and estimated outages. The cluster plots shown here are on input variables wind and lightning. The two or more of the data points will have the same values on the input variables for different output variables, when this happens, the points are plotted on top of each other, and it is hard from the cluster plot to state how many data points each symbol on the plot represents and hard to find the cluster boundaries. From the scatter plots it can observed that, all four cities have same similar pattern for the training and test data and it is hard to categorize the data points in each clusters.

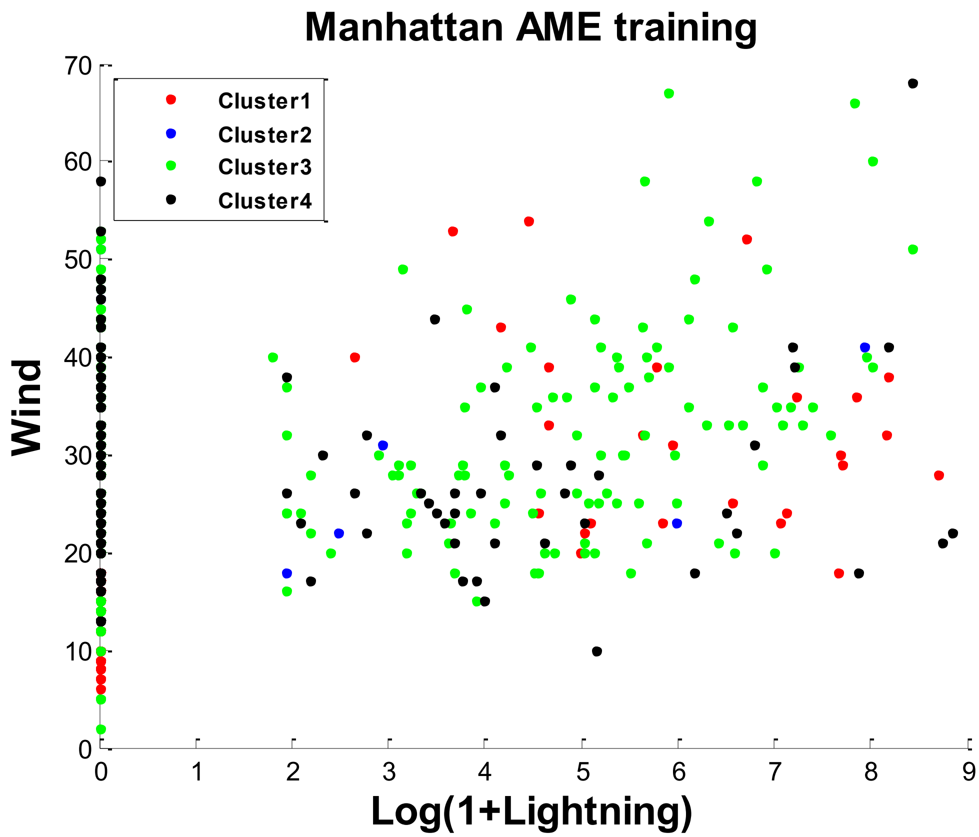


Figure 6.7 Cluster Plots of Input Data for Weather-related Outages Obtained with AME Model in Overhead Distribution Systems for Manhattan from 2005 to 2009

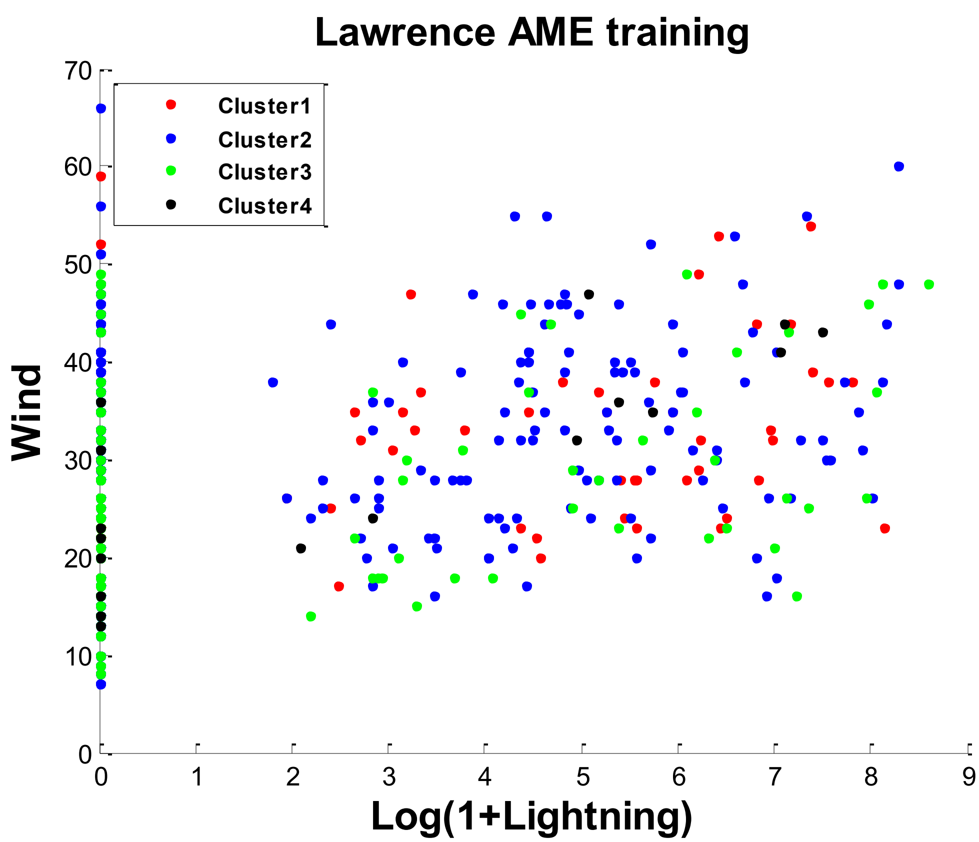


Figure 6.8 Cluster Plots of Input Data for Weather-related Outages Obtained with AME Model in Overhead Distribution Systems for Lawrence from 2005 to 2009

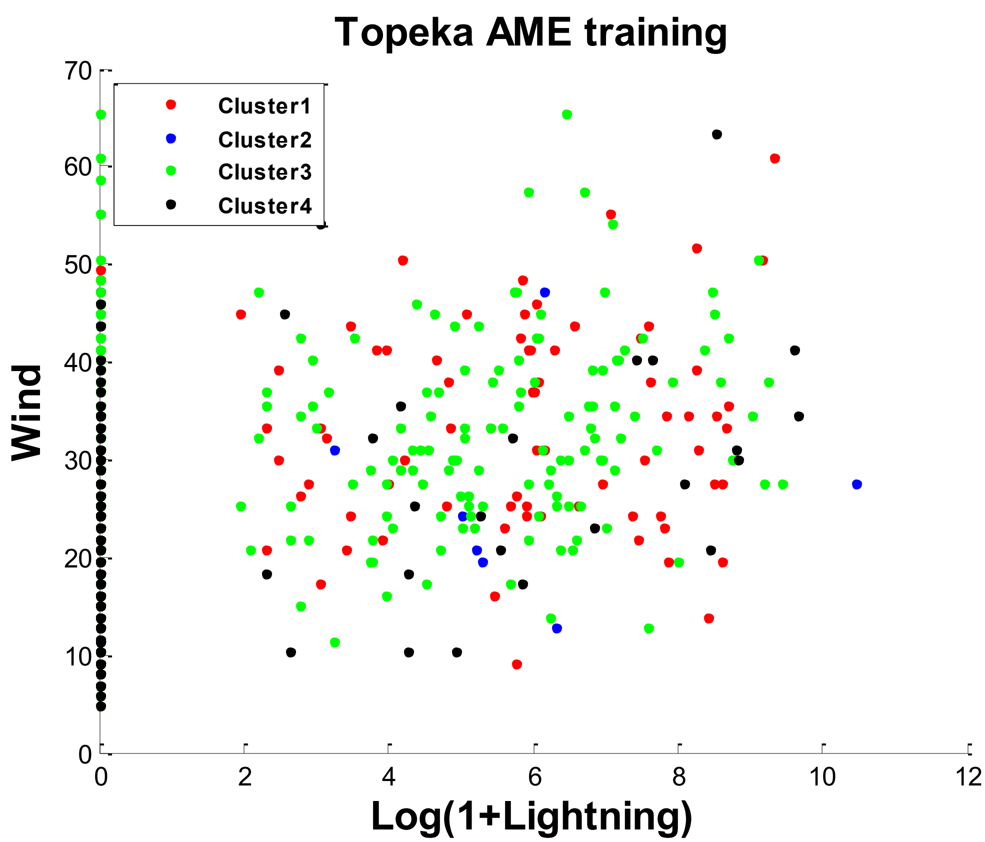


Figure 6.9 Cluster Plots of Input Data for Weather-related Outages Obtained with AME Model in Overhead Distribution Systems for Topeka from 2005 to 2009

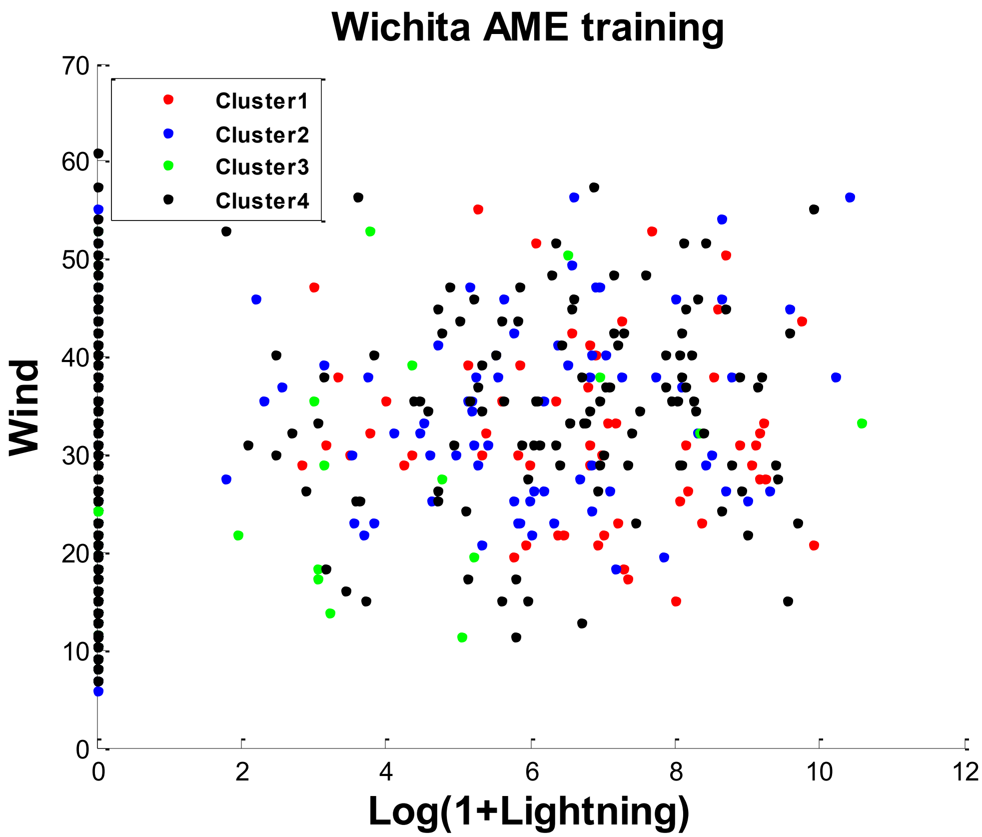


Figure 6.10 Cluster Plots of Input Data for Weather-related Outages Obtained with AME Model in Overhead Distribution Systems for Wichita from 2005 to 2009

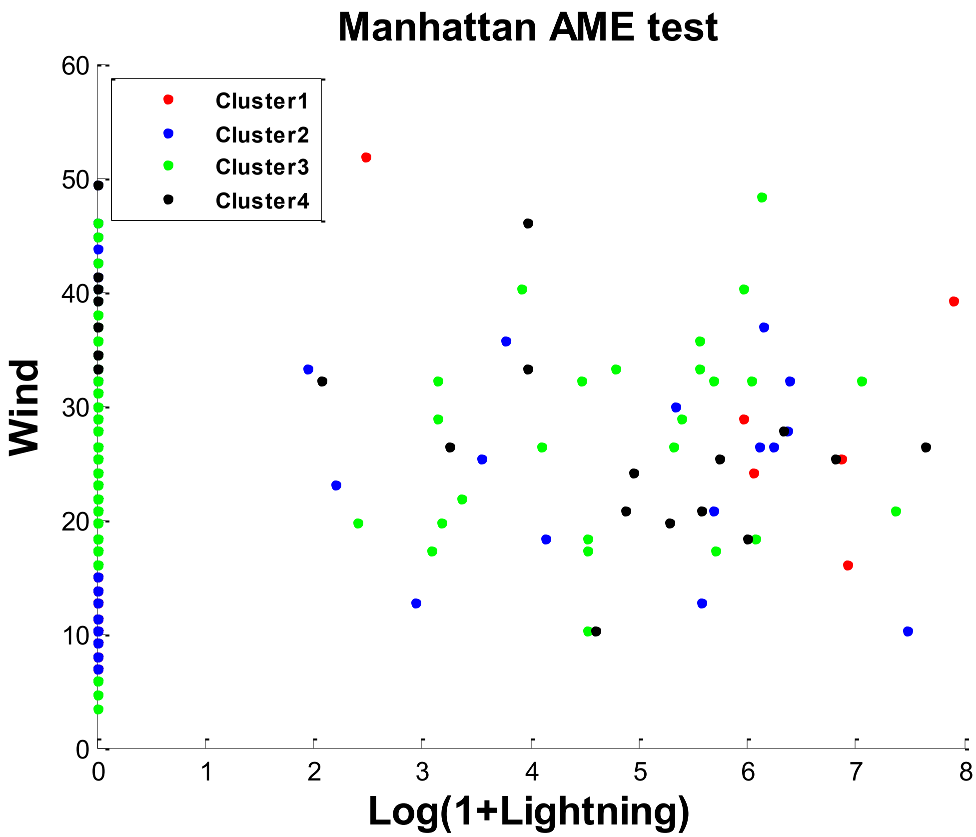


Figure 6.11 Cluster Plots of Input Data for Weather-related Outages Obtained with AME Model in Overhead Distribution Systems for Manhattan from 2010 to 2011

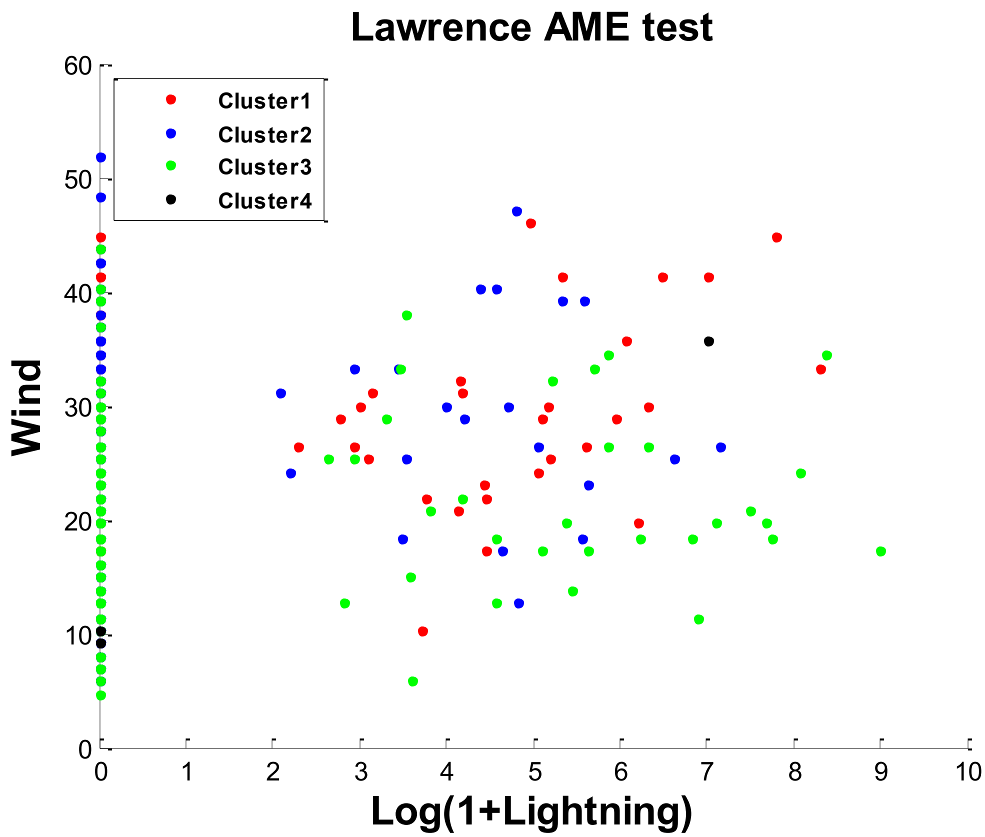


Figure 6.12 Cluster Plots of Input Data for Weather-related Outages Obtained with AME Model in Overhead Distribution Systems for Lawrence from 2010 to 2011

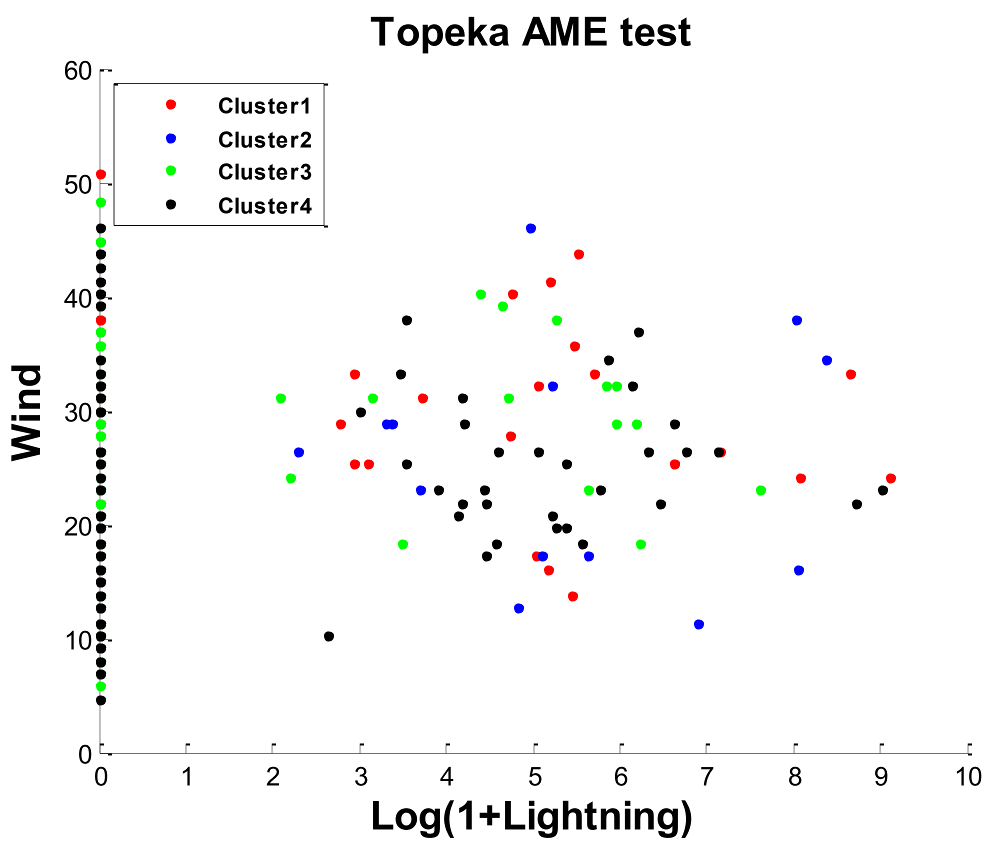


Figure 6.13 Cluster Plots of Input Data for Weather-related Outages Obtained with AME Model in Overhead Distribution Systems for Topeka from 2010 to 2011

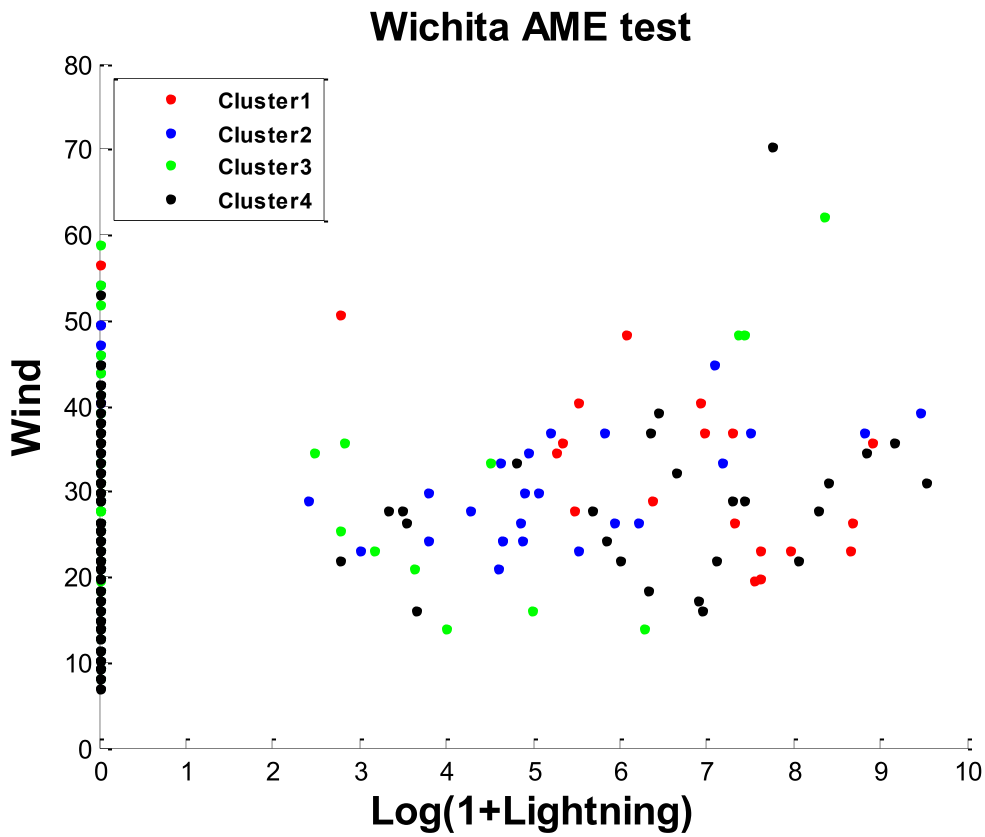


Figure 6.14 Cluster Plots of Input Data for Weather-related Outages Obtained with AME Model in Overhead Distribution Systems for Wichita from 2010 to 2011

Performance of the proposed AME model is measured using MAE, MSE, R and S, given in Table 6.3. The correlation for the four cities for the training data is close to one with a very high value of slope, indicating better performance of the model. The results for test data are slightly inferior, but they are better than those obtained with other models. It can be observed that for Lawrence, the results for training data have improved but not as good on the test data. The scatter plot of daily observed outages and estimated outages for AME for four cities for training and test duration are shown in Figure 6.7. The scatter plot clearly shows that the model estimated outages very well in the lower range as well as in the upper range.

Table 6.3 Results of AME Model for Four Cities

AME								
	Training Data				Test Data			
	MSE	MAE	R	S	MSE	MAE	R	S
Manhattan	0.7044	0.2611	0.9225	0.8406	0.6679	0.2802	0.8715	0.6400
Lawrence	0.1815	0.2603	0.9724	0.9414	2.9584	0.3856	0.7382	0.3743
Topeka	3.0957	0.4493	0.9138	0.8276	9.6359	1.0227	0.8871	0.8172
Wichita	2.8548	0.9677	0.9225	0.8289	18.6595	1.4358	0.8883	0.7524

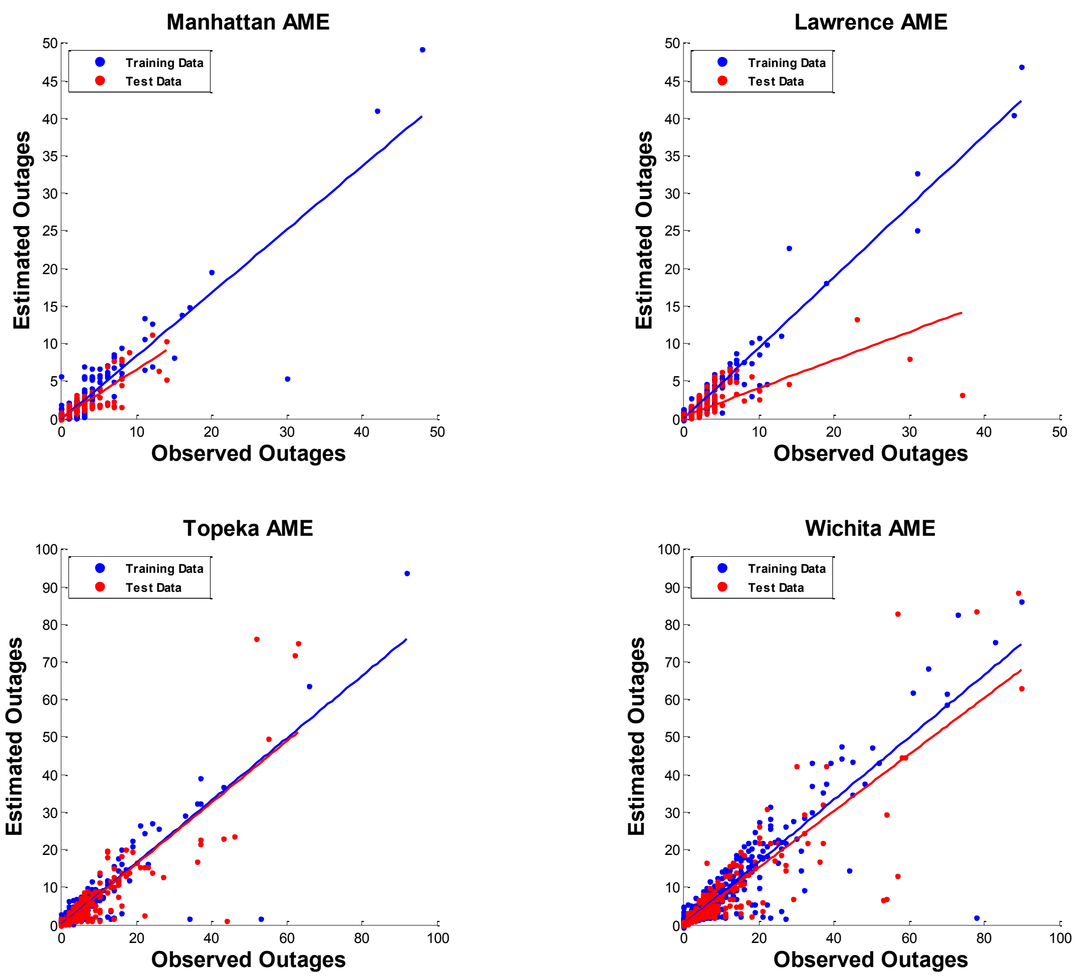


Figure 6.15 Plot of Observed and Estimated Weather-related Outages Obtained with AME Model in Overhead Distribution Systems for Four Cities from 2005 to 2011

6.3 Summary

In this chapter, ensemble learning models for the estimation of weather caused outages were investigated. Training the network on a specialized region on the input space and computing the ensemble output using combining rules performs better than the individual network. AdaBoost.RT, AdaBoost^{+} and AME models are applied to estimate weather-related outages. AdaBoost models were able to approximate complex relations between the input and output, and they were able to estimate outages in the lower range but under-estimated in the upper range. The AME model proposed was able to accurately estimate outages in the lower as well as upper ranges, compared to the single network in which the multilayered neural network over-estimated in the lower range and under-estimated in the upper range.

Chapter 7 - Comparison of Models

Different models, including linear and exponential regression, neural network, AdaBoost, and a mixture of experts were presented to estimate outages caused by weather in overhead distribution systems. Model performance was evaluated by computing MAE, MSE, RMSE, MAPE, slope, and correlation coefficient for four cities in Kansas, Wichita, Topeka, Lawrence, and Manhattan, which represent the two large cities and two smaller cities in Kansas. Available historical data was divided into training data from 2005 to 2009 and test data from 2010 to 2011. Tables 7.1 to 7.8 tabulate performance measure values for training and testing data for four cities. Because of the random nature and noise in the data, traditional regression models faced a big challenge in estimating outages both in the lower range and the higher range. Therefore, the NN model, which has the ability to approximate high complexity functions, was introduced and it outperformed regression models. However, the NN model still could not give accurate estimations in extreme cases which demanded methods based on committee machines. Generalized performance for various committee machines was certainly impressive. A particular feature of boosting, one of the method used, is that it reduces bias and variance. Training the network on a specialized region on the input space and computing the ensemble output using combining rules gives better results than the individual network. Since AdaBoost models are able to approximate complex relations between the input and output, they were able to estimate outages in the lower range. However, they underestimate in the upper range of outages. The AME model proposed was able to accurately estimate outages both in lower and upper ranges.

Tables 7.1 - 7.8 shows the performance measure values for training and test data for all models for Manhattan, Lawrence, Topeka and Wichita. It can be observed that all the performance measures, MAE, MSE, RMSE, and MAPE, dropped from traditional models to ensemble models. The slopes of the best-fit lines between the estimated and the observed outages are also higher than those for traditional models, indicating better performance. For ensemble models for all the cities, the correlation coefficients for the training data and the testing data are closer to one too which shows a high degree of relationship between the estimated and the observed values. AME performed distinctly better than other models, followed by AdaBoost$^+$ for all the four cities demonstrating the reasons for using the ensemble learning methods.

Table 7.1 Performance Measure for Manhattan Training Data by Different Models for Weather-related Outages

Manhattan - Training Data						
	MAE	**MSE**	**RMSE**	**MAPE**	**S**	**R**
Regression Model 1	0.6702	3.3181	1.8216	39.7715	0.2712	0.5258
Regression Model 2	0.6055	3.2239	1.7955	33.1950	0.3012	0.5449
Regression Model 3	0.5924	3.1218	1.7669	31.7251	0.3233	0.5650
Regression Model 4	0.6056	2.7824	1.6681	34.4878	0.3946	0.6268
Regression Model 5	0.6334	2.8980	1.7024	36.7020	0.3934	0.6077
Regression Model 6	0.7795	4.1716	2.0424	48.4178	0.3773	0.4669
Neural Network	0.6009	2.4879	1.5773	35.1208	0.4555	0.6761
AdaBoost.RT	0.3789	1.9163	1.3843	14.4025	0.5082	0.7813
AdaBoost^{+}	0.3691	1.8251	1.3510	13.6323	0.5694	0.7860
AME	0.2611	0.7044	0.8393	10.9287	0.8406	0.9225

Table 7.2 Performance Measure for Manhattan Test Data by Different Models for Weather-related Outages

Manhattan - Test Data						
	MAE	**MSE**	**RMSE**	**MAPE**	**S**	**R**
Regression Model 1	0.7180	2.0779	1.4415	35.7854	0.1556	0.4824
Regression Model 2	0.6622	2.1021	1.4499	29.5598	0.1826	0.4929
Regression Model 3	0.6509	2.0185	1.4207	28.8090	0.2362	0.5077
Regression Model 4	0.6885	2.2247	1.4915	30.8150	0.1328	0.4283
Regression Model 5	0.6992	2.1622	1.4704	32.5445	0.1467	0.4477
Regression Model 6	0.6919	2.4705	1.5718	27.8129	0.0589	0.4608
Neural Network	0.6433	2.3370	1.5287	23.5203	0.1254	0.4335
AdaBoost.RT	0.5578	2.0737	1.4400	16.6458	0.1952	0.6216
AdaBoost^{+}	0.5578	2.0558	1.4338	16.5486	0.2083	0.6079
AME	0.2802	0.6679	0.8172	8.7964	0.6400	0.8715

Table 7.3 Performance Measure for Lawrence Training Data by Different Models for Weather-related Outages

Lawrence - Training Data						
	MAE	**MSE**	**RMSE**	**MAPE**	**S**	**R**
Regression Model 1	0.7421	4.4220	2.1028	45.4708	0.0699	0.2755
Regression Model 2	0.6879	4.3814	2.0932	39.9166	0.0849	0.2890
Regression Model 3	0.6860	4.3805	2.0930	39.7650	0.0851	0.2893
Regression Model 4	0.6872	4.3731	2.0912	39.9118	0.0864	0.2920
Regression Model 5	0.7114	4.4043	2.0987	41.8992	0.0722	0.2816
Regression Model 6	0.6502	4.4999	2.1213	35.1645	0.0554	0.2468
Neural Network	0.6973	4.3621	2.0886	40.8074	0.0871	0.2958
AdaBoost.RT	0.3884	3.5923	1.8953	17.3300	0.2112	0.5150
AdaBoost^{+}	0.3095	2.6210	1.6189	12.9578	0.3662	0.6947
AME	0.1815	0.2603	0.5102	8.6236	0.9414	0.9724

Table 7.4 Performance Measure for Lawrence Test Data by Different Models for Weather-related Outages

Lawrence - Test Data						
	MAE	**MSE**	**RMSE**	**MAPE**	**S**	**R**
Regression Model 1	0.8653	4.6686	2.1607	38.6730	0.1420	0.5185
Regression Model 2	0.8099	4.7475	2.1789	31.9075	0.1431	0.5219
Regression Model 3	0.8028	4.7243	2.1735	31.6856	0.1396	0.5543
Regression Model 4	0.8261	4.9956	2.2351	32.4379	0.1051	0.5316
Regression Model 5	0.8471	4.8554	2.2035	35.4599	0.1168	0.5449
Regression Model 6	0.8713	5.7250	2.3927	30.8729	0.0371	0.4655
Neural Network	0.8778	5.0176	2.2400	40.7603	0.1012	0.4130
AdaBoost.RT	0.5316	3.6619	1.9136	18.4797	0.2792	0.6662
AdaBoost$^+$	0.4340	3.0395	1.7434	14.6647	0.3825	0.7173
AME	0.3856	2.9584	1.7200	10.9470	0.3743	0.7382

Table 7.5 Performance Measure for Topeka Training Data by Different Models for Weather-related Outages

Topeka - Training Data						
	MAE	**MSE**	**RMSE**	**MAPE**	**S**	**R**
Regression Model 1	1.4523	13.7319	3.7057	78.1175	0.2515	0.5114
Regression Model 2	1.3690	13.3099	3.6483	64.8792	0.2892	0.5325
Regression Model 3	1.3753	13.1783	3.6302	66.4634	0.2964	0.5391
Regression Model 4	1.3754	13.1423	3.6252	66.4296	0.2979	0.5409
Regression Model 5	1.4077	13.7337	3.7059	72.0787	0.2433	0.5114
Regression Model 6	1.3537	15.5890	3.9483	60.3896	0.1298	0.4189
Neural Network	1.3913	12.9613	3.6002	69.5512	0.3016	0.5494
AdaBoost.RT	0.9053	11.7940	3.4342	26.8615	0.3042	0.6434
AdaBoost^{+}	0.7070	8.8922	2.9820	19.6678	0.4409	0.7448
AME	0.4493	3.0957	1.7594	12.0277	0.8276	0.9138

Table 7.6 Performance Measure for Topeka Test Data by Different Models for Weather-related Outages

Topeka – Test Data						
	MAE	**MSE**	**RMSE**	**MAPE**	**S**	**R**
Regression Model 1	2.4318	40.3696	6.3537	62.7603	0.0724	0.4443
Regression Model 2	2.3851	40.3150	6.3494	50.0993	0.0905	0.4756
Regression Model 3	2.4054	40.9059	6.3958	51.1664	0.0810	0.4613
Regression Model 4	2.4105	41.2994	6.4265	51.2496	0.0736	0.4604
Regression Model 5	2.4329	41.1660	6.4161	58.9479	0.0636	0.4721
Regression Model 6	2.5160	45.0651	6.7131	49.8850	0.0284	0.4713
Neural Network	2.4418	37.1506	6.0951	62.6446	0.1909	0.4231
AdaBoost.RT	1.9056	32.0706	5.6631	29.5779	0.2095	0.7026
AdaBoost$^+$	1.4621	22.9910	4.7949	21.3953	0.3530	0.7928
AME	1.0227	9.6359	3.1042	14.7260	0.8172	0.8871

Table 7.7 Performance Measure for Wichita Training Data by Different Models for Weather-related Outages

Wichita - Training Data						
	MAE	**MSE**	**RMSE**	**MAPE**	**S**	**R**
Regression Model 1	3.0767	39.7079	6.3014	135.4532	0.2413	0.5151
Regression Model 2	2.8773	37.5084	6.1244	107.3498	0.3107	0.5501
Regression Model 3	2.8841	37.4824	6.1223	108.0321	0.3111	0.5506
Regression Model 4	2.8891	37.3769	6.1137	108.1678	0.3135	0.5524
Regression Model 5	3.0183	39.7459	6.3044	129.0661	0.2541	0.5114
Regression Model 6	2.9238	42.0023	6.4809	105.4414	0.2144	0.4678
Neural Network	2.8051	35.9343	5.9945	107.8690	0.3312	0.5756
AdaBoost.RT	1.6712	24.3932	4.9390	39.7066	0.4539	0.7692
AdaBoost$^+$	1.4615	18.4254	4.2925	37.9276	0.5755	0.8263
AME	0.9677	8.1498	2.8548	22.4145	0.8289	0.9225

Table 7.8 Performance Measure for Wichita Test Data by Different Models for Weather-related Outages

Wichita - Test Data						
	MAE	**MSE**	**RMSE**	**MAPE**	**S**	**R**
Regression Model 1	3.5966	71.8361	8.4756	105.5216	0.1050	0.4775
Regression Model 2	3.3308	63.3287	7.9579	80.8640	0.1919	0.5811
Regression Model 3	3.3392	63.5145	7.9696	81.4191	0.1892	0.5809
Regression Model 4	3.3468	63.5677	7.9729	81.5540	0.1902	0.5771
Regression Model 5	3.5369	71.0337	8.4281	100.3206	0.1128	0.4970
Regression Model 6	3.4298	77.2927	8.7916	73.6672	0.0837	0.4814
Neural Network	3.2873	76.7436	8.7603	67.2931	0.2314	0.4003
AdaBoost.RT	3.1672	49.9527	7.0677	38.2060	0.4145	0.6401
AdaBoost$^+$	2.4409	48.0201	6.9297	32.4919	0.3271	0.7963
AME	1.4358	18.6595	4.3197	17.6520	0.7524	0.8883

Figures 7.1 and 7.2 show scatter plots with regression lines of observed and estimated outages for training and test data of best regression model, neural network, AdaBoost.RT, ADABOOST$^+$, and AME. These graphs show clear improvement in AME model performance, which provides better slope than other models for all training and test cases.

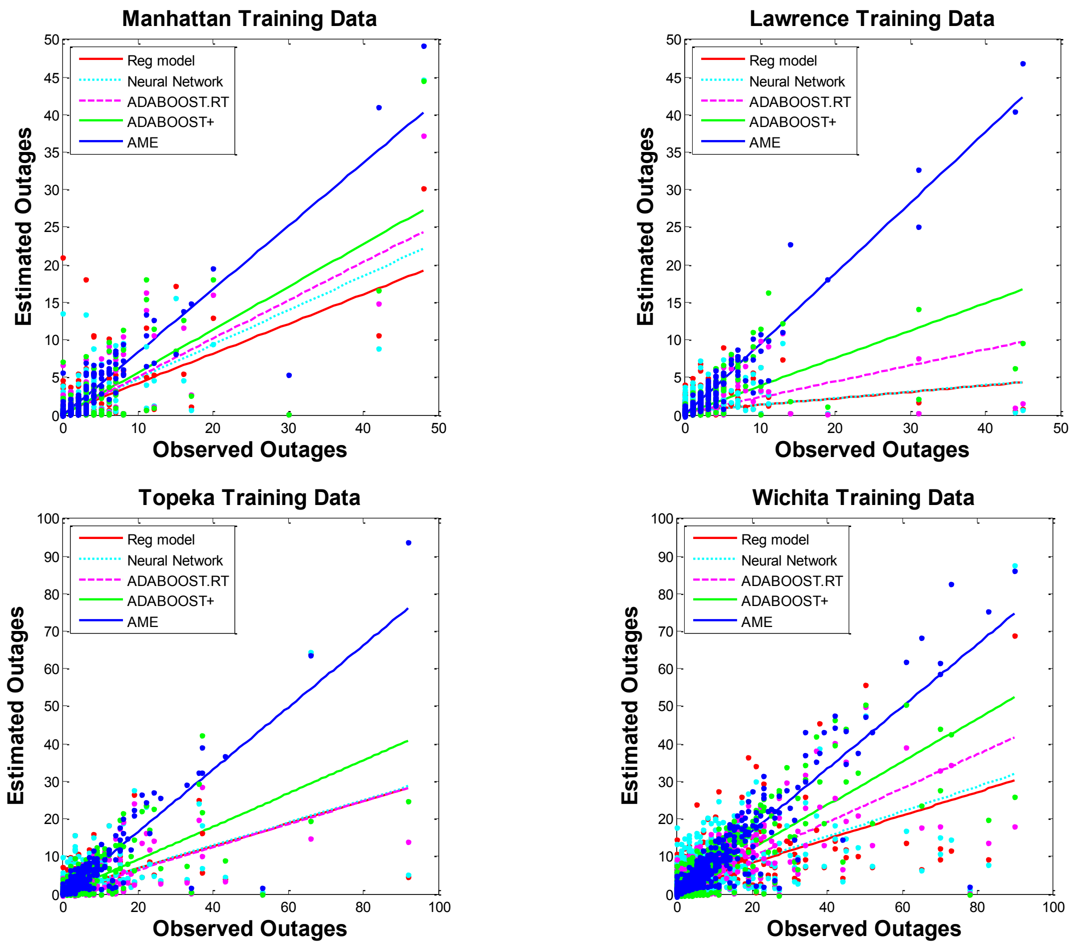


Figure 7.1 Scatter Plot with Regression Line of Observed and Estimated Outages for 2005-2009 Training Data for Different Models

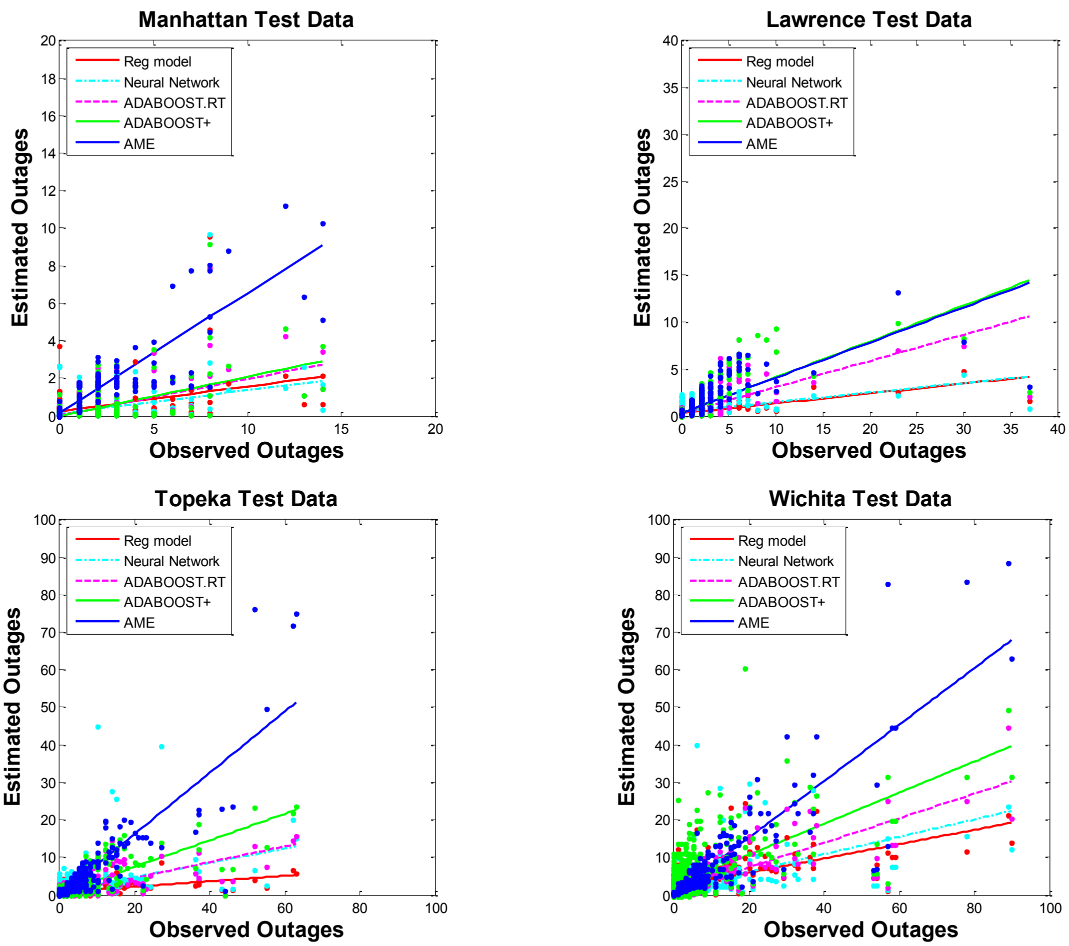


Figure 7.2 Scatter plot with Regression Line of Observed vs. Estimated Outages for 2010-2011 Testing Data for Different Models

7.1 Summary

In this chapter, the models discussed in Chapters 3 through 6 for the estimation of weather-related outages in the overhead distribution system were compared. The analysis of performance metrics of six regression models, NN model, ensemble models were presented. Experimental results concluded that AME model outperformed the other models, followed by AdaBoost^{+} model. Also, results indicated that ensemble of networks accurately estimates outage, compared to the single network. Overall, the ensemble learning methods gave significantly better performance compared to traditional linear, quadratic and exponential regression models and NN model.

Chapter 8 - Application of Models for Estimation of Outages for Only Lightning Days

In order to evaluate performance of models discussed in Chapters 3 through 7, all days in the study period were considered. In this chapter, results with non-lightning days were excluded from the study period are presented. The motivation is to determine whether the models will provide better results for days with lightning. All the models discussed previously were used to find correlation between lightning, wind, and outages.

8.1 Data Overview

In the previous study 2555 days from January 1, 2005 to December 31, 2011 were considered. Table 8.1 shows the number of days on which lightning occurred during the study period in the four cities. The histogram of outages caused by lightning and wind is shown in Figure 8.1. Since outages are spatially aggregated over the service area, smaller cities such as Manhattan and Lawrence have the most number of days with zero outages, whereas Topeka and Wichita have the least number of days with zero outages.

Table 8.1 Number of Lightning Days from 2005 to 2011

Cities	Number of Days
Manhattan	239
Lawrence	288
Topeka	314
Wichita	329

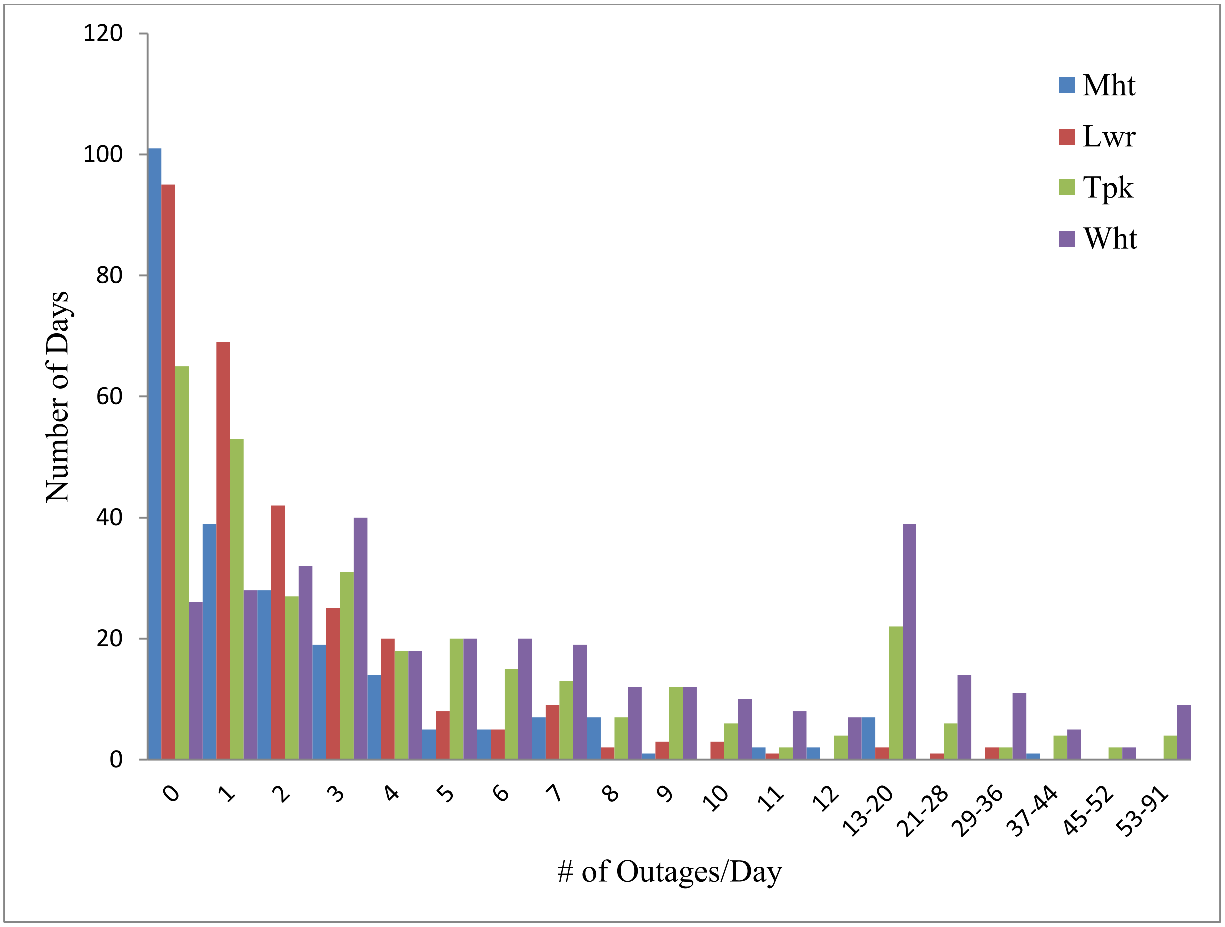


Figure 8.1 Histogram of Weather-related Outages Excluding Non-lightning Days, 2005-2011

8.2 Models & Results

Performances of the models were evaluated using the same metrics as done previously, that is by computing MAE, MSE, R and S values. The division of training and test data is also similar with 2005-2009 as training data and 2010-2011 as test data.

8.2.1 Regression Models

Tables 8.2 to 8.5 tabulate the MAE, MSE, R and S for six different regression models of the four service areas for training and testing data. The tables show that, among all six models, model 4 has the lowest MAE and MSE and the highest correlation coefficient, R, for the training data; whereas for the test data, model 1 has lower MAE and MSE for Manhattan, Lawrence, and Topeka, and model 2 has lower MAE and MSE for Wichita. Figures 8.2 to 8.5 show scatter plots between observed and estimated outages with a regression line for the six regression models for the four cities. The scatter plots indicate that the regression models underestimate outages when in the higher range and overestimate in the lower range.

Table 8.2 Results of Regression Models for Manhattan

Manhattan								
	Training Data				Test Data			
	MSE	MAE	R	S	MSE	MAE	R	S
Model 1	22.2554	2.4144	0.5411	0.2674	12.2207	2.3069	0.4113	0.1382
Model 2	20.9057	2.2984	0.5762	0.3386	13.8272	2.3392	0.3930	0.1557
Model 3	20.2859	2.2961	0.5932	0.3590	13.2797	2.3449	0.4198	0.2072
Model 4	17.4771	2.1212	0.6643	0.4453	14.8278	2.3655	0.2804	0.0983
Model 5	18.5958	2.2975	0.6402	0.4263	13.1955	2.3204	0.3540	0.1323
Model 6	31.8471	4.1990	0.4952	0.2279	16.9278	2.5182	0.3638	0.0487

Table 8.3 Results of Regression Models for Lawrence

Lawrence								
	Training Data				Test Data			
	MSE	MAE	R	S	MSE	MAE	R	S
Model 1	7.9714	1.6617	0.4109	0.1424	33.0010	2.5279	0.4502	0.4502
Model 2	7.7517	1.6337	0.4319	0.1891	34.6061	2.6660	0.4379	0.4379
Model 3	7.7414	1.6381	0.4331	0.1901	34.7633	2.7106	0.3833	0.3833
Model 4	7.7413	1.6389	0.4331	0.1902	34.7534	2.7152	0.3789	0.3789
Model 5	7.8939	1.6212	0.4161	0.1564	34.2368	2.5952	0.4782	0.4782
Model 6	8.6825	1.6526	0.3786	0.0563	42.2372	3.2598	0.4436	0.4436

Table 8.4 Results of Regression Models for Topeka

Topeka								
	Training Data				Test Data			
	MSE	MAE	R	S	MSE	MAE	R	S
Model 1	57.4702	3.6942	0.5658	0.2939	234.7566	8.4734	0.4108	0.0641
Model 2	54.7092	3.4509	0.5908	0.3524	248.4255	9.1163	0.4379	0.0806
Model 3	52.4079	3.4189	0.6135	0.3802	251.7194	9.1459	0.4432	0.0640
Model 4	52.4079	3.4176	0.6135	0.3803	251.7084	9.1458	0.4432	0.0640
Model 5	58.1699	3.5117	0.5587	0.2756	244.5595	8.7217	0.4180	0.0561
Model 6	74.4721	3.8945	0.4061	0.0766	288.7147	10.0273	0.4066	0.0242

Table 8.5 Results of Regression Models for Wichita

WICHITA								
	Training Data				Test Data			
	MSE	MAE	R	S	MSE	MAE	R	S
Model 1	120.6749	6.2875	0.5678	0.2849	228.3920	7.7286	0.4953	0.1312
Model 2	113.9089	6.0441	0.5954	0.3592	204.6407	7.4305	0.6088	0.2423
Model 3	112.7194	6.1154	0.6011	0.3659	207.8384	7.4830	0.6103	0.2238
Model 4	109.8211	6.0993	0.6146	0.3825	207.7634	7.5412	0.6036	0.2393
Model 5	122.5367	6.0590	0.5533	0.2927	231.7494	7.7373	0.5097	0.1330
Model 6	143.6969	7.2987	0.4671	0.1425	292.3997	8.8891	0.4370	0.1110

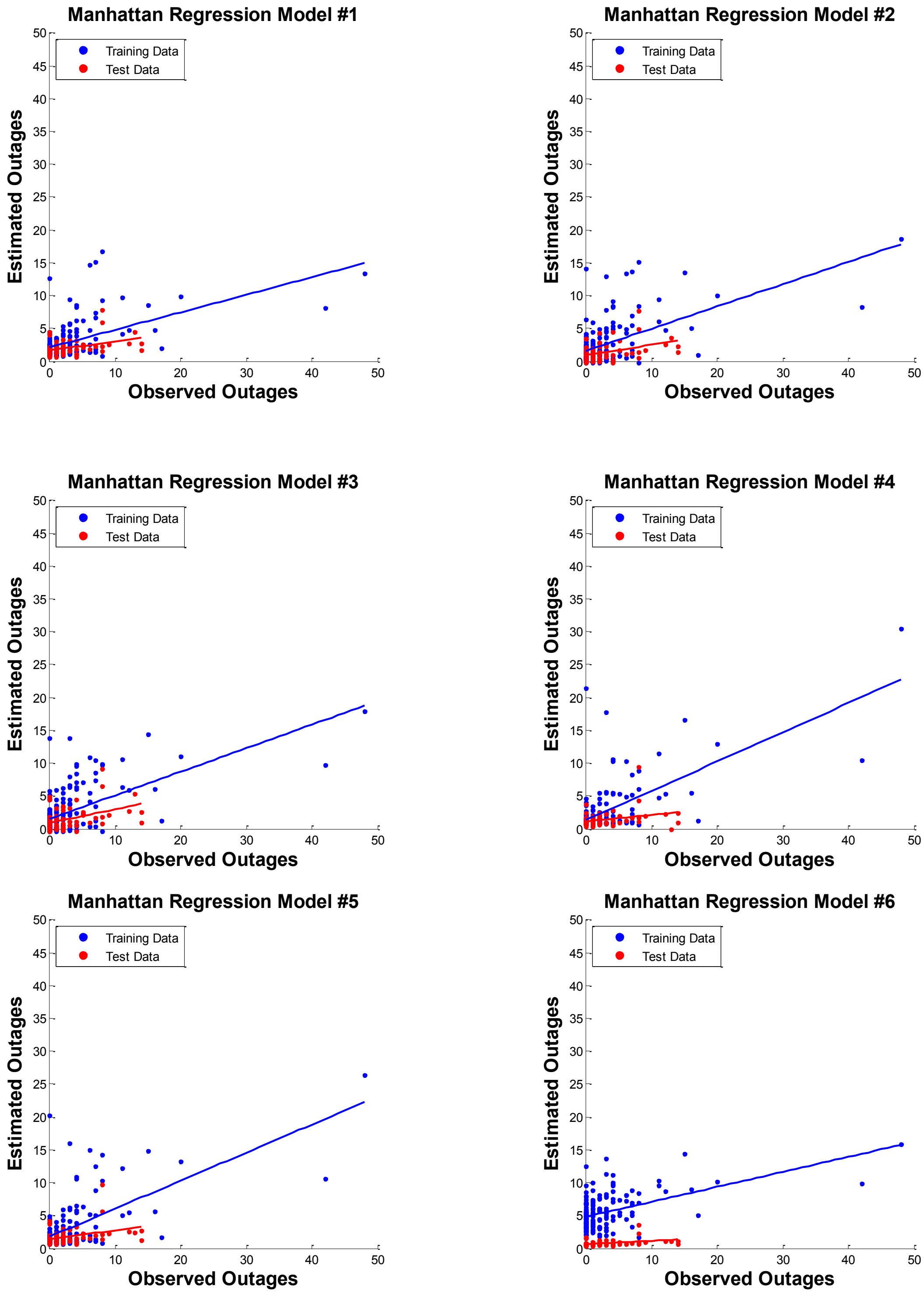


Figure 8.2 Plot of Observed and Estimated Weather-related Outages, Excluding Non-lightning days, Obtained with Six Regression Models in Overhead Distribution Systems for Manhattan from 2005 to 2011

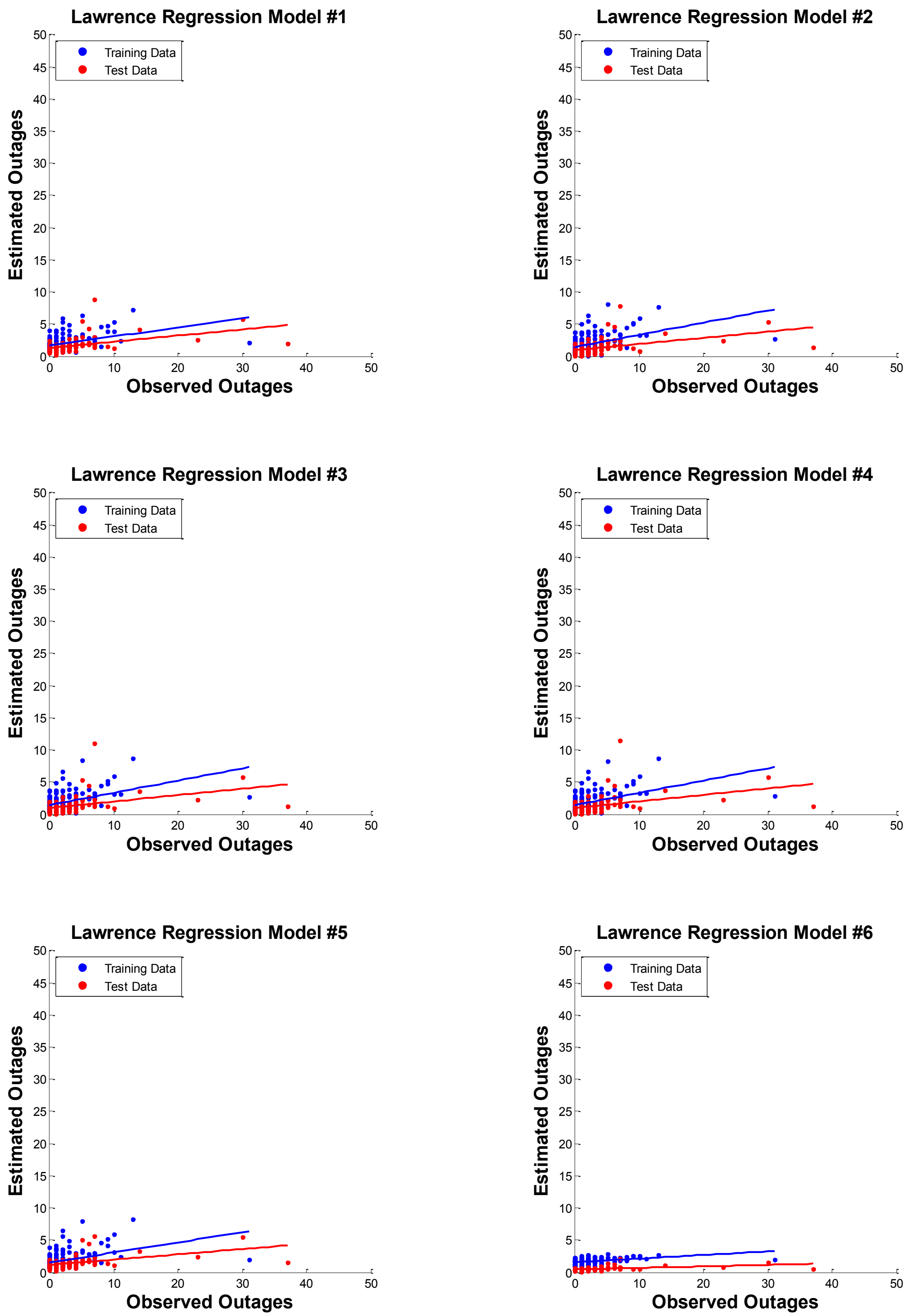


Figure 8.3 Plot of Observed and Estimated Weather-related Outages, Excluding Non-lightning days, Obtained with Six Regression Models in Overhead Distribution Systems for Lawrence from 2005 to 2011

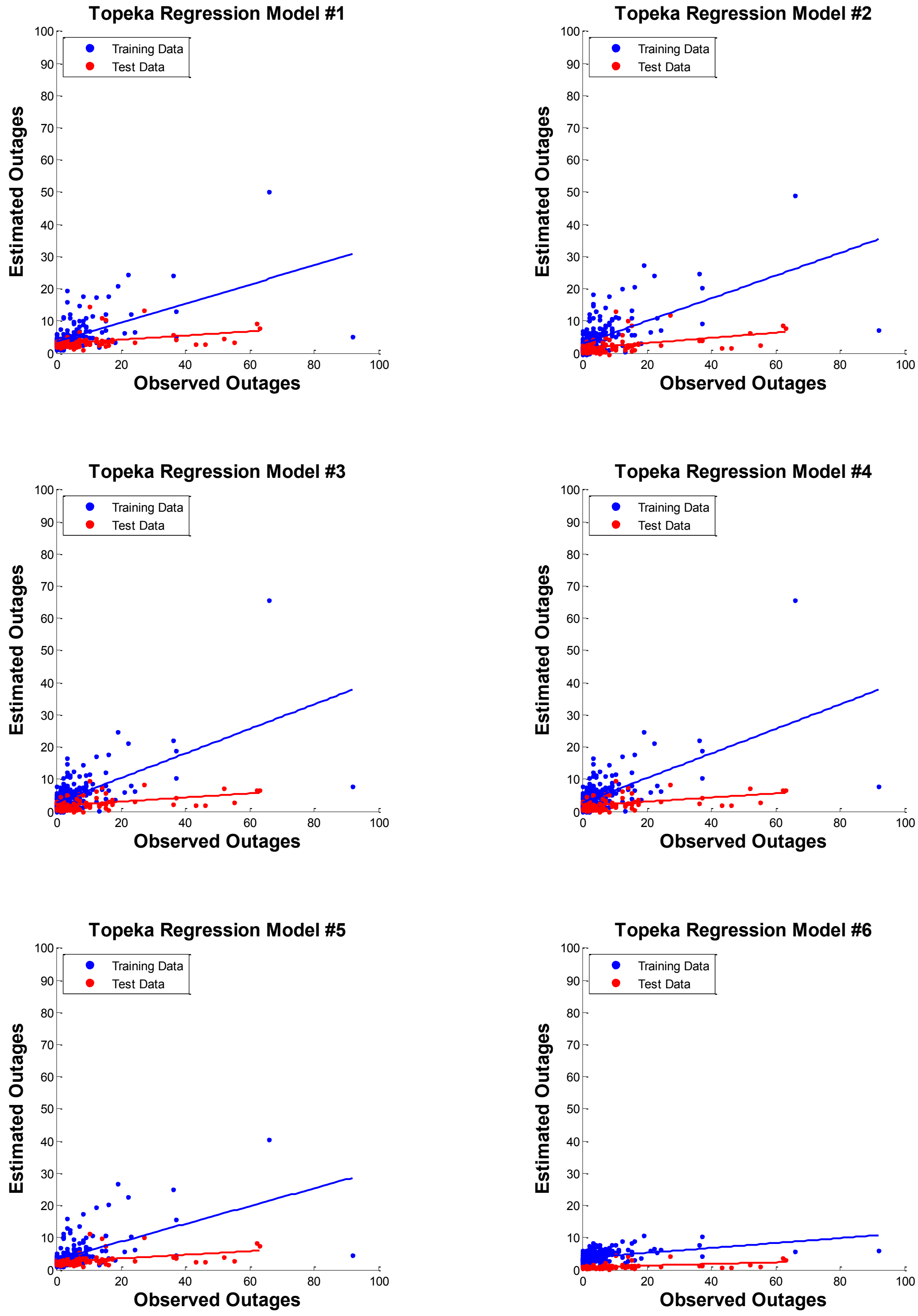


Figure 8.4 Plot of Observed and Estimated Weather-related Outages, Excluding Non-lightning days, Obtained with Six Regression Models in Overhead Distribution Systems for Topeka from 2005 to 2011

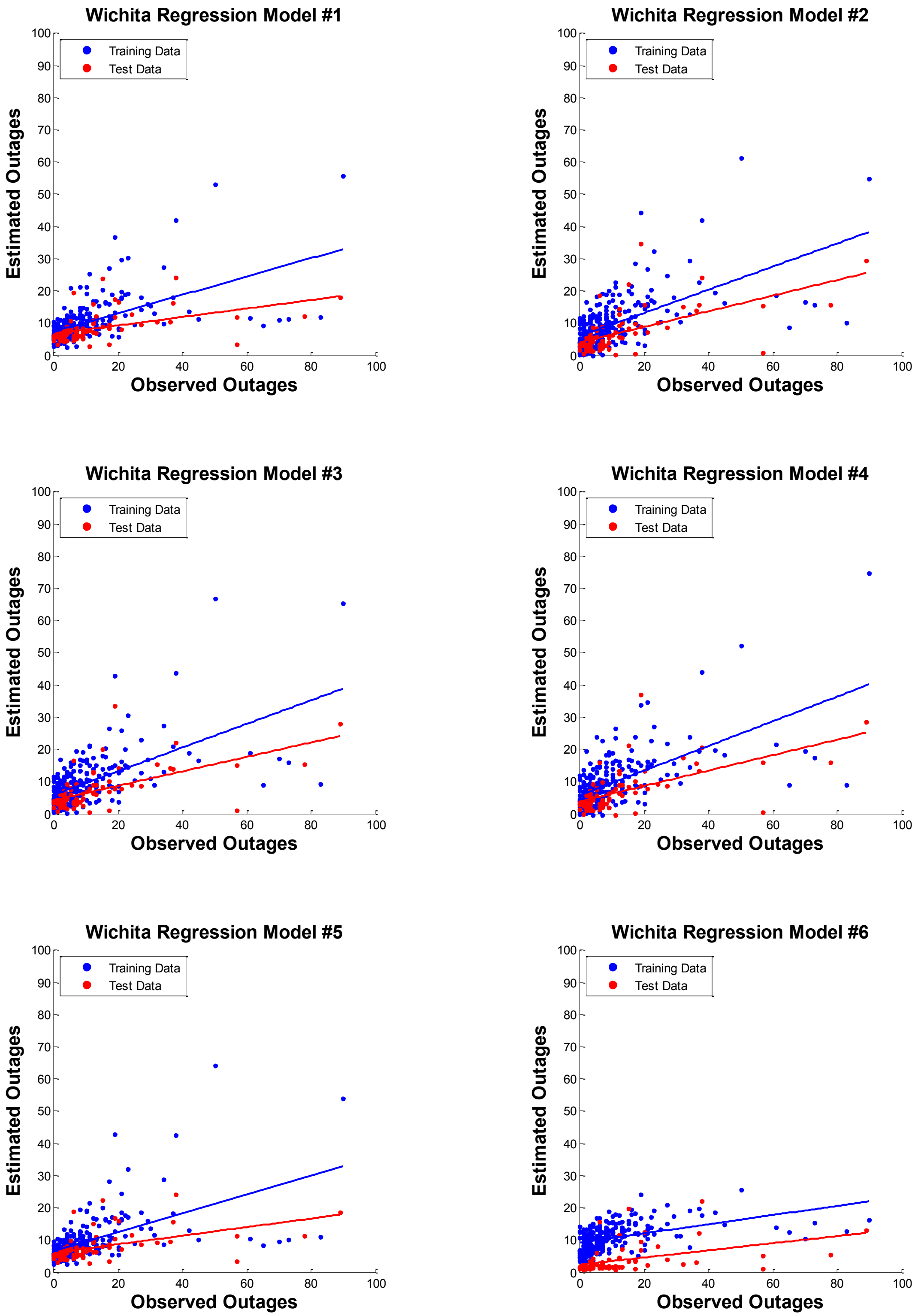


Figure 8.5 Plot of Observed and Estimated Weather-related Outages, Excluding Non-lightning Days, Obtained with Six Regression Models in Overhead Distribution Systems for Wichita from 2005 to 2011

8.2.2 Neural Network

The NN model, as discussed in Chapter 4, which is a 3x2x1 multilayered feed forward network, shown in Figure 4.3, is considered in this section. The MSE, MAE, R and S for the four cities are tabulated in Table 8.6.

By comparing the MSE and MAE values obtained from NN model with the six regression models presented in the previous section, the NN model has performed better. Figure 8.6 shows a scatter plot of observed and estimated outages, with a regression line for the four cities.

Table 8.6 Results of Neural Network Model for Four Cities

Neural Network								
	Training Data				**Test Data**			
	MSE	**MAE**	**R**	**S**	**MSE**	**MAE**	**R**	**S**
Manhattan	17.0073	2.0839	0.6672	0.4504	11.1545	2.2044	0.4519	0.2819
Lawrence	7.5003	1.6066	0.4614	0.2124	31.7455	2.4905	0.4575	0.0991
Topeka	52.1686	3.3817	0.6158	0.3795	217.2222	8.3975	0.4832	0.1815
Wichita	107.7232	5.9835	0.6241	0.3886	203.9876	6.9484	0.6123	0.2947

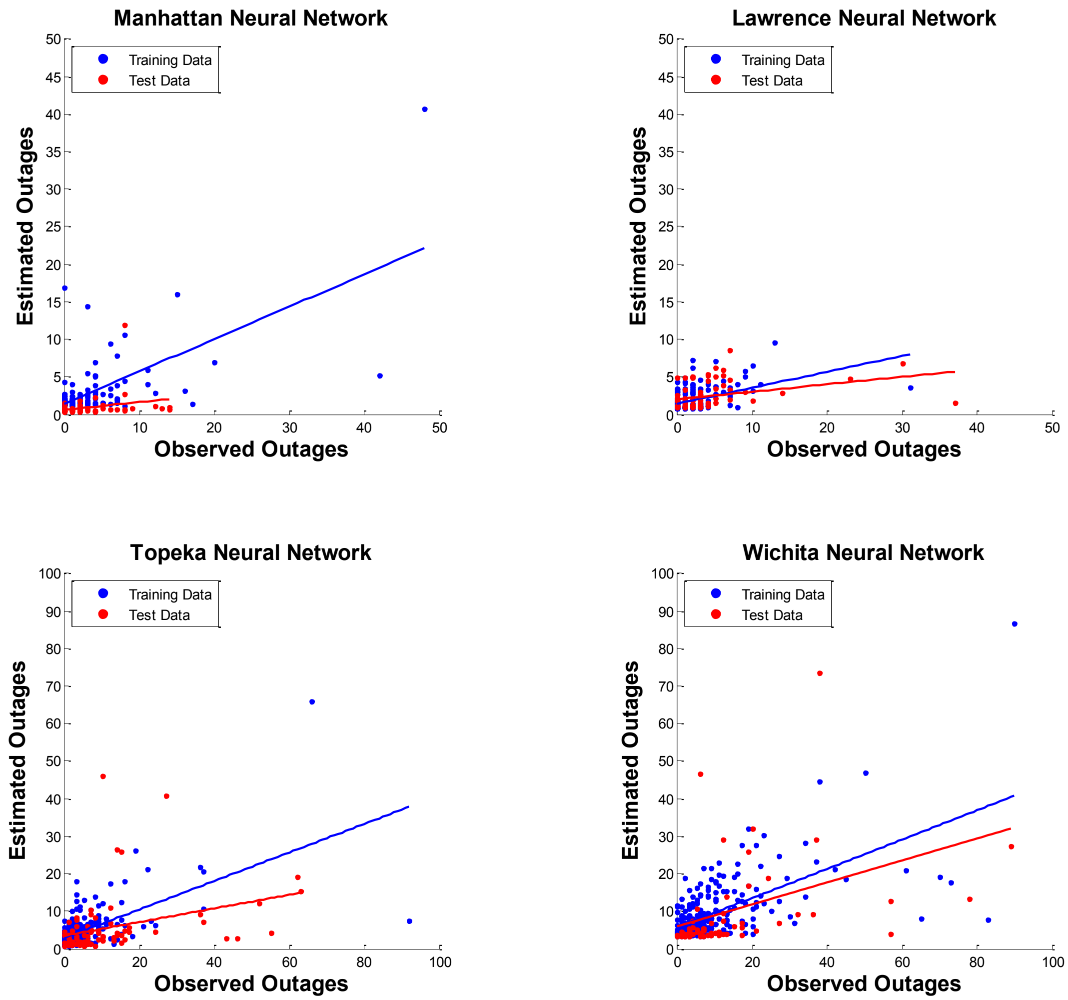


Figure 8.6 Plot of Observed and Estimated Weather-related Outages, Excluding Non-lightning Days, Obtained with NN Model in Overhead Distribution Systems for Four Cities from 2005 to 2011

8.2.3 AdaBoost Model

The MSE, MAE, R and S for the four cities for the AdaBoost.RT and AdaBoost^{+} are tabulated in Tables 8.7 and 8.8. Figures 8.7 and 8.8 show the scatter plot between observed and estimated outages, with a regression line for the AdaBoost.RT and AdaBoost^{+} models for the four cities.

Table 8.7 Results of AdaBoost.RT Model for Four Cities

AdaBoost.RT								
	Training Data				Test Data			
	MSE	**MAE**	**R**	**S**	**MSE**	**MAE**	**R**	**S**
Manhattan	7.0555	1.1504	0.8905	0.6729	11.0939	1.7922	0.6791	0.2542
Lawrence	4.1884	0.7555	0.7772	0.4506	22.8446	1.6073	0.6961	0.2462
Topeka	37.0523	1.8908	0.7781	0.4478	143.3025	5.4715	0.7425	0.6693
Wichita	64.9567	3.5652	0.8125	0.5238	132.4641	5.1816	0.8205	0.8317

Table 8.8 Results of AdaBoost$^+$ Model for Four Cities

AdaBoost$^+$								
	Training Data				Test Data			
	MSE	**MAE**	**R**	**S**	**MSE**	**MAE**	**R**	**S**
Manhattan	6.3902	1.1065	0.8945	0.7457	10.8610	1.7671	0.6576	0.2751
Lawrence	1.8573	0.6011	0.9056	0.7107	17.7038	1.7692	0.7352	0.3908
Topeka	3.9552	1.2753	0.9762	0.9431	42.8308	3.4174	0.9013	0.9452
Wichita	13.4385	2.5290	0.9646	0.8785	49.7171	3.4863	0.9036	0.8448

From tables 8.6 - 8.8, it can be observed that the AdaBoost$^+$ model estimates the outages more accurately compared to the NN model. For bigger cities, Topeka and Wichita, the drop in MAE and MSE is very significant compared to smaller cities, Manhattan and Lawrence. The correlation and slope for all cities for the AdaBoost models are high indicating better performance of the models.

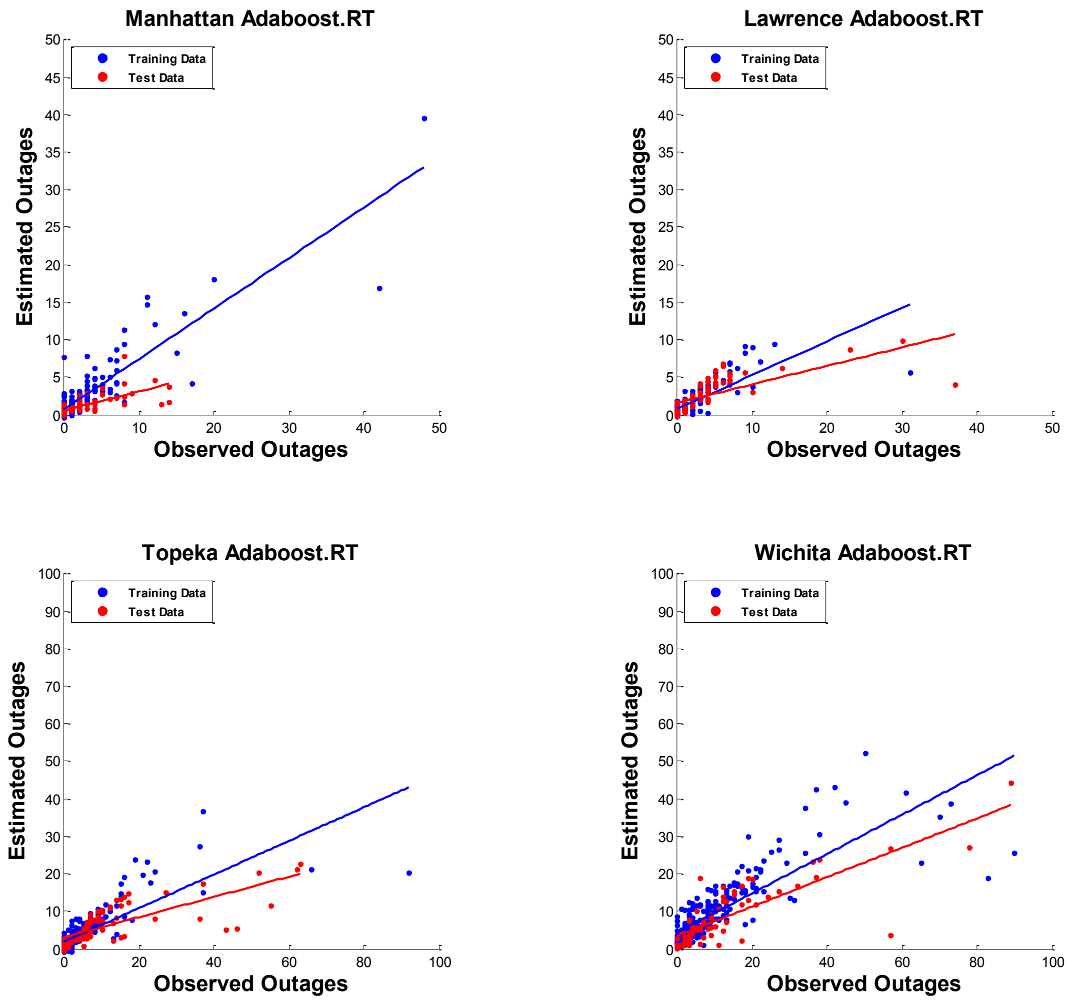


Figure 8.7 Plot of Observed and Estimated Weather-related Outages, Excluding Non-lightning days, Obtained with AdaBoost.RT Model in Overhead Distribution Systems for Four Cities from 2005 to 2011

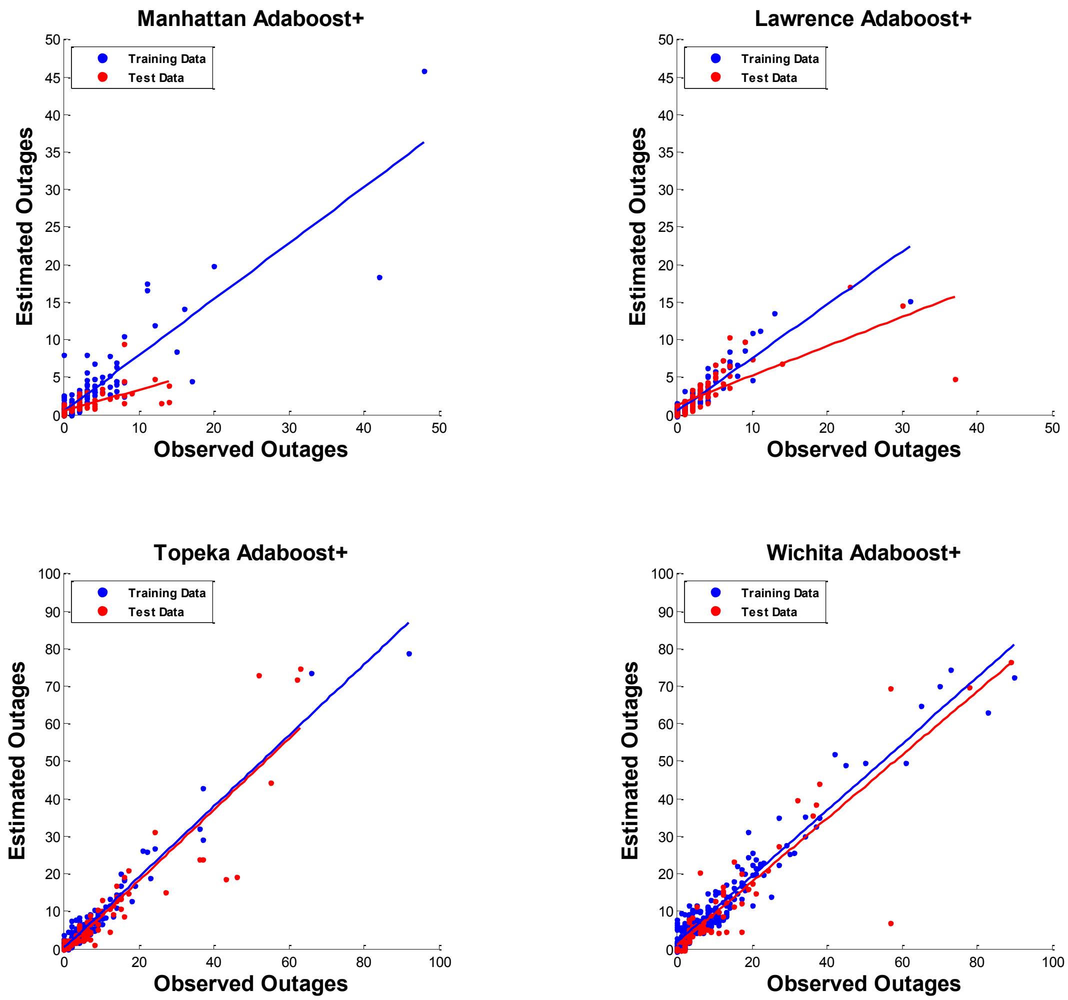


Figure 8.8 Plot of Observed and Estimated Weather-related Outages, Excluding Non-lightning Days, Obtained with AdaBoost^{+} Model in Overhead Distribution Systems for Four Cities from 2005 to 2011

8.2.4 Annealed Mixture of Experts (AME) Model

The MSE, MAE, R and S for the four cities for AME are tabulated in Table 8.9. Figure 8.9 shows the scatter plot between observed and estimated outages, with a regression line for the AME model for the four cities. The scatter plot indicates that the model accurately estimated outages both in the lower range and the higher range. The high correlation also indicates the better performance of AME model in estimation of weather-caused outages. However, results for the test data for Lawrence are still inferior as compared to other cities.

Table 8.9 Results of AME Model for Four Cities

AME								
	Training Data				**Test Data**			
	MSE	**MAE**	**R**	**S**	**MSE**	**MAE**	**R**	**S**
Manhattan	1.2064	0.6169	0.9806	0.9547	3.7862	0.9601	0.8864	0.6259
Lawrence	0.3535	0.4144	0.9814	0.9598	16.7268	1.4352	0.7352	0.4577
Topeka	2.5126	0.1115	0.9879	0.9706	37.0199	3.3236	0.9150	0.9045
Wichita	8.9861	2.2542	0.9754	0.9503	29.2457	2.9082	0.9479	0.8634

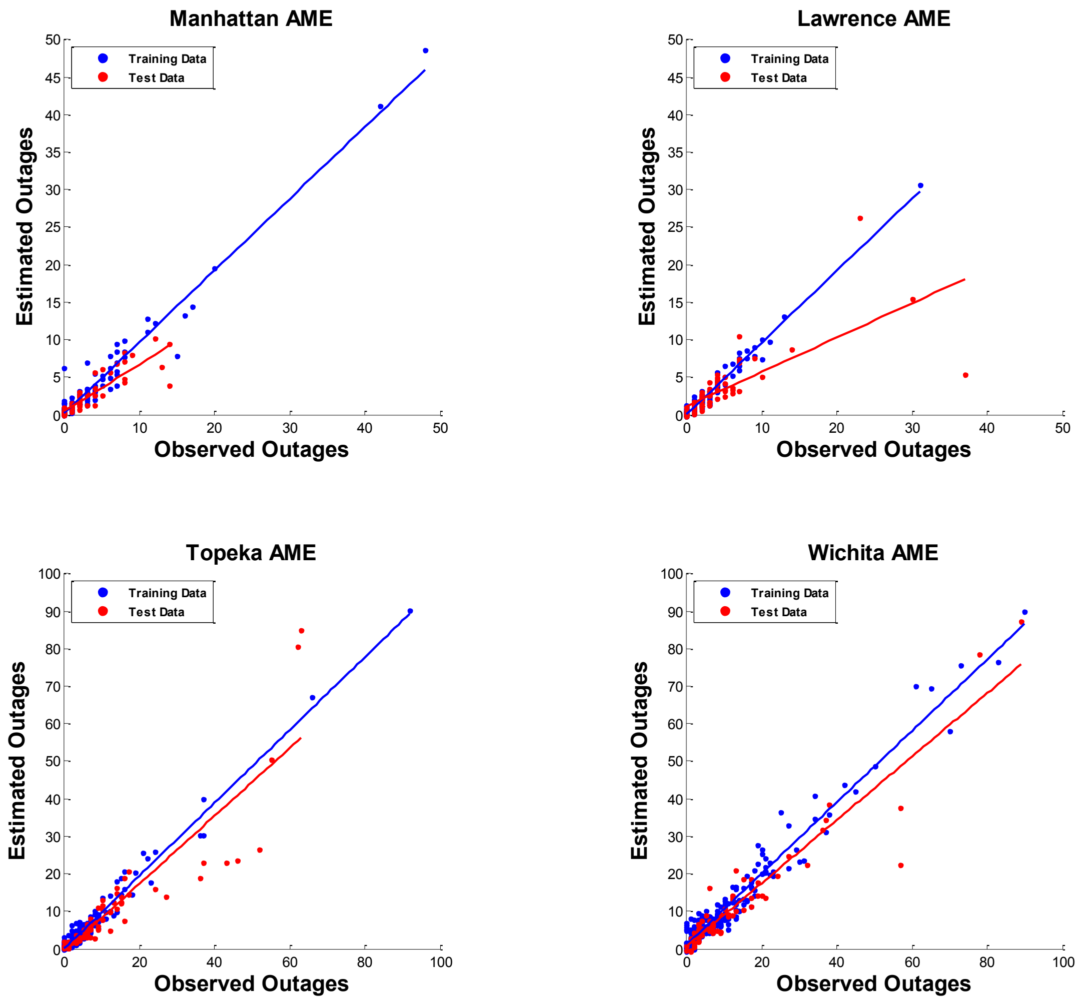


Figure 8.9 Plot of Observed and Estimated Weather-related Outages, Excluding Non-lightning Days, Obtained with AME Model in Overhead Distribution Systems for Four Cities from 2005 to 2011

8.3 Analysis of Models Result

The available weather and outage data from 2005 to 2011 excluding non-lightning days were used. The models discussed in Chapters 3 through 6, that is six regression models, NN model, and ensemble models were investigated to estimate outages and find correlation. Similar metrics as previously used that is by computing the MAE, MSE, R and S values are used to evaluate the performance of the models. The performance measure values are summarized in Table 8.10 for training data and testing data for Manhattan, Lawrence, Topeka and Wichita. Experimental results concluded that AME model outperformed the other models, followed by AdaBoost models. For the tested cities, all models have positive correlation coefficient, indicating a positive relationship between observed and estimated outages. The drop is MAE and MSE values from NN model to AME model is very significant for training and test data for all four cities. Also, high correlation of ensemble models which are close to one indicates that ensemble of networks more accurately estimates outages, compared to the traditional regression models and single NN model.

To observe the impact of only lightning on the distribution outages, the results were obtained by considering only days on which lightning happened as input and they are compared with the model results with all days as input. To be simple, we define dataset 1 and dataset 2 as,

- Dataset 1 – Only the days that have lightning in the study period
- Dataset 2 – All days included in the study period

With dataset 2 as input the models are trained and the outages are estimated. Although all the days were used for training, the performance measure values are computed for the days in which lightning happened and are summarized in Table 8.11 for the training data and test data for all four cities. Comparison of Tables 8.10 and 8.11, shows that dataset 1 has better performance measure values. Figures 8.10 to 8.17 shows the observed and estimated outages for AdaBoost$^+$ and AME model for dataset 1 and 2 for training and test data for Manhattan and Wichita. The plots shows that both AdaBoost$^+$ and AME model are able to estimate outages very well for dataset 1 and 2. From Figures 8.12, 8.13, 8.16 and 8.17, it can be clearly observed that for dataset 1 and 2, the AME model is able to estimates outages accurately, compared to AdaBoost$^+$ model.

Table 8.10 Summary of Model Results for four Cities for Dataset 1

		Training Data				Test Data			
		MAE	**MSE**	**R**	**S**	**MAE**	**MSE**	**R**	**S**
Mht	**NN**	2.0839	17.0073	0.6672	0.4504	2.2044	11.1545	0.4519	0.2819
	AdaBoot.RT	1.1504	7.0555	0.8905	0.6729	1.7922	11.0939	0.6791	0.2542
	AdaBoost^{+}	1.1065	6.3902	0.8945	0.7457	1.7671	10.8610	0.6576	0.2751
	AME	0.6169	1.2064	0.9806	0.9547	0.9601	3.7862	0.8864	0.6259
Lwr	**NN**	1.6066	7.5003	0.4614	0.2124	2.4905	31.7455	0.4575	0.0991
	AdaBoot.RT	0.7555	4.1884	0.7772	0.4506	1.6073	22.8446	0.6961	0.2462
	AdaBoost^{+}	0.6011	1.8573	0.9056	0.7107	1.4692	17.7038	0.7397	0.3908
	AME	0.4144	0.3535	0.9814	0.9598	1.4352	16.7268	0.7352	0.4577
Tpk	**NN**	3.3817	52.1686	0.6158	0.3795	8.3975	217.2222	0.4832	0.1815
	AdaBoot.RT	1.8908	37.0523	0.7781	0.4478	5.4715	143.3025	0.7425	0.6693
	AdaBoost^{+}	1.2753	3.9552	0.9762	0.9431	3.3236	37.0199	0.9150	0.9452
	AME	0.1115	2.5126	0.9879	0.9706	3.4174	42.8308	0.9013	0.9045
Wht	**NN**	5.9835	107.7232	0.6241	0.3886	6.9484	203.9876	0.6123	0.2947
	AdaBoot.RT	3.5652	64.9567	0.8125	0.5238	5.1816	132.4641	0.8205	0.8317
	AdaBoost^{+}	2.5290	13.4385	0.9646	0.8785	3.4863	49.7171	0.9036	0.8448
	AME	1.8249	7.4726	0.9791	0.9503	2.9082	29.2457	0.9479	0.8634

Table 8.11 Summary of Model Results for four Cities for Dataset 2

		Training Data				Test Data			
		MAE	**MSE**	**R**	**S**	**MAE**	**MSE**	**R**	**S**
Mht	**NN**	1.8530	14.6236	0.7347	0.5594	2.4163	15.8752	0.3548	0.1318
	AdaBoot.RT	1.3730	5.8565	0.8636	0.6074	1.9593	12.2358	0.6219	0.2241
	AdaBoost⁺	1.2889	7.8769	0.8695	0.7006	1.9410	11.9084	0.6127	0.2505
	AME	0.7346	1.4625	0.9765	0.9588	0.9167	3.2285	0.9061	0.6844
Lwr	**NN**	1.5256	7.9216	0.4373	0.2053	2.8599	37.0763	0.3912	0.0715
	AdaBoot.RT	0.7694	3.6690	0.8007	0.5171	1.9683	28.1243	0.5417	0.1743
	AdaBoost⁺	0.6788	2.2221	0.8893	0.6556	1.7538	25.05	0.6042	0.2318
	AME	0.5471	0.6764	0.9644	0.9603	1.6626	23.7575	0.6460	0.2504
Tpk	**NN**	3.1090	54.1858	0.6075	0.3774	9.1558	234.7022	0.3206	0.1754
	AdaBoot.RT	2.2712	48.0067	0.7018	0.3921	6.6575	185.5079	0.6532	0.1929
	AdaBoost⁺	1.8595	33.5809	0.7990	0.4849	5.1769	132.5051	0.7671	0.2936
	AME	1.1337	2.9294	0.9825	0.9738	3.5183	38.7901	0.9126	0.9500
Wht	**NN**	5.4880	110.5012	0.6246	0.4037	9.1286	319.9919	0.3587	0.3187
	AdaBoot.RT	3.8783	80.6723	0.7696	0.4342	5.6632	145.9637	0.8286	0.3693
	AdaBoost⁺	2.9238	55.2778	0.8463	0.5763	7.5024	155.2768	0.6444	0.4078
	AME	2.2542	8.9861	0.9754	0.9579	3.7394	53.3616	0.9061	0.9236

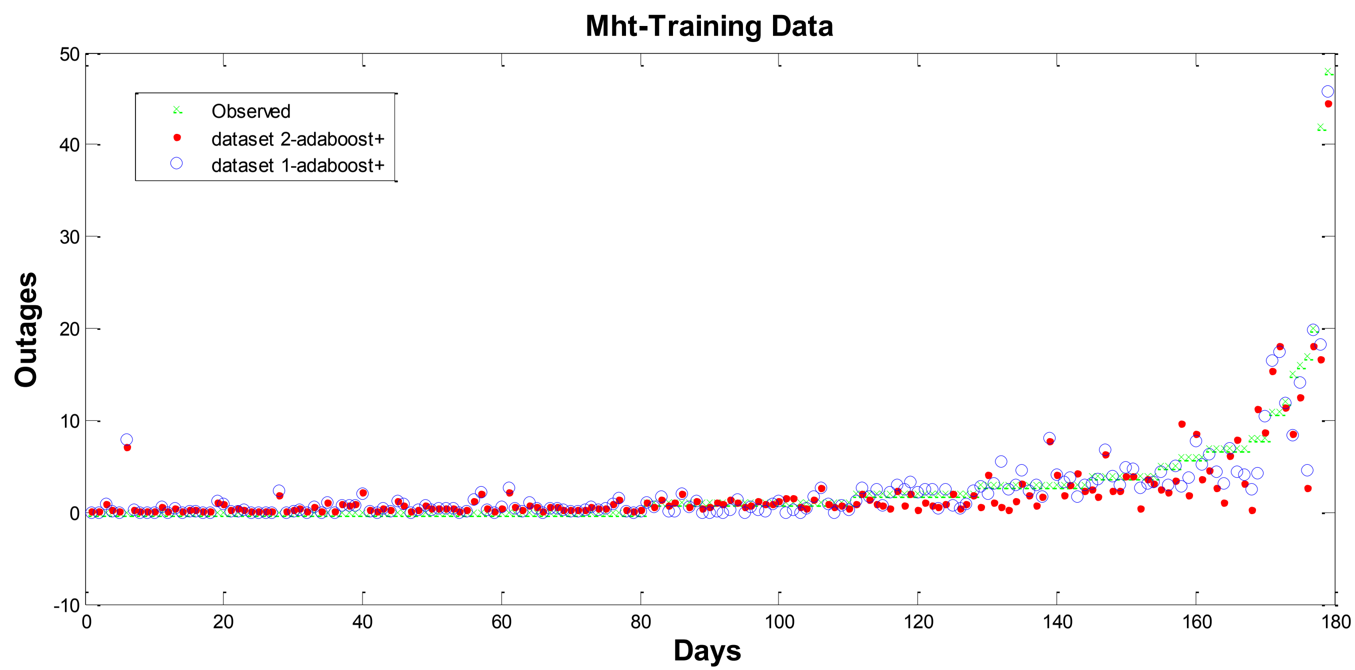


Figure 8.10 Observed and Estimated Outages for Dataset 1 & 2 for Adaboost^{+} model for Manhattan Training Data

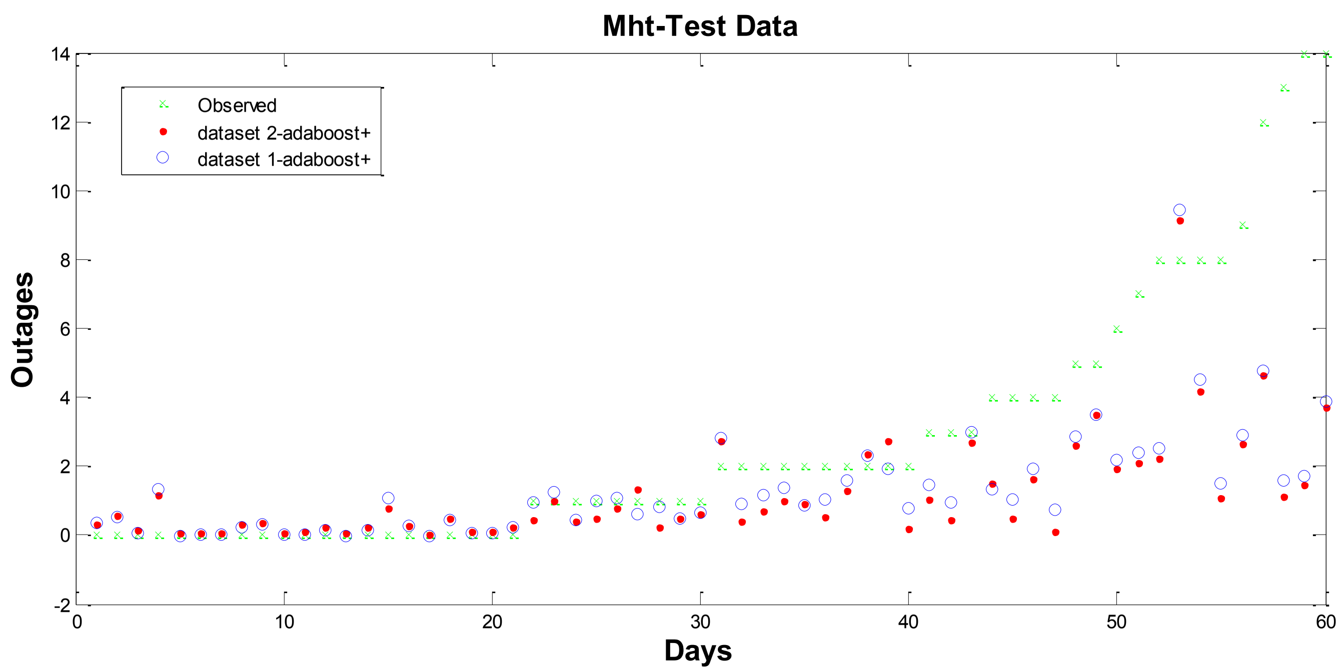


Figure 8.11 Observed and Estimated Outages for Dataset 1 & 2 for Adaboost$^+$ model for Manhattan Test Data

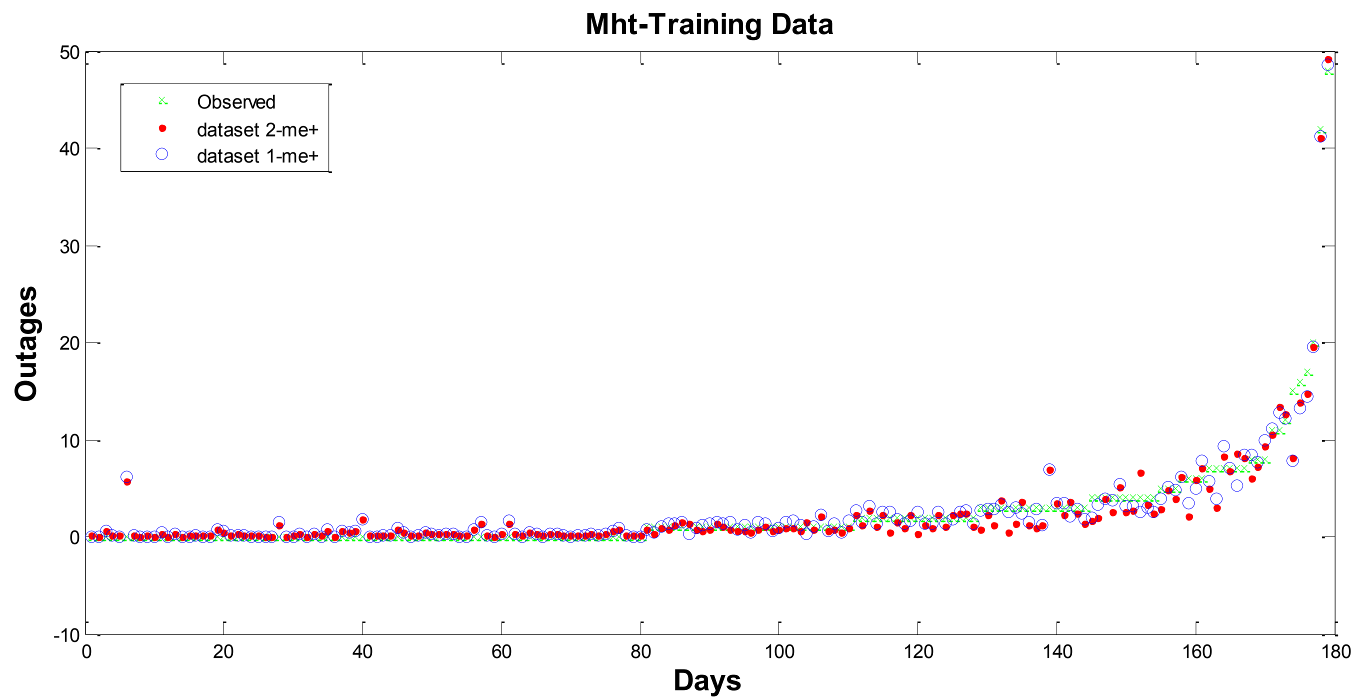


Figure 8.12 Observed and Estimated Outages for Dataset 1 & 2 for AME model for Manhattan Training Data

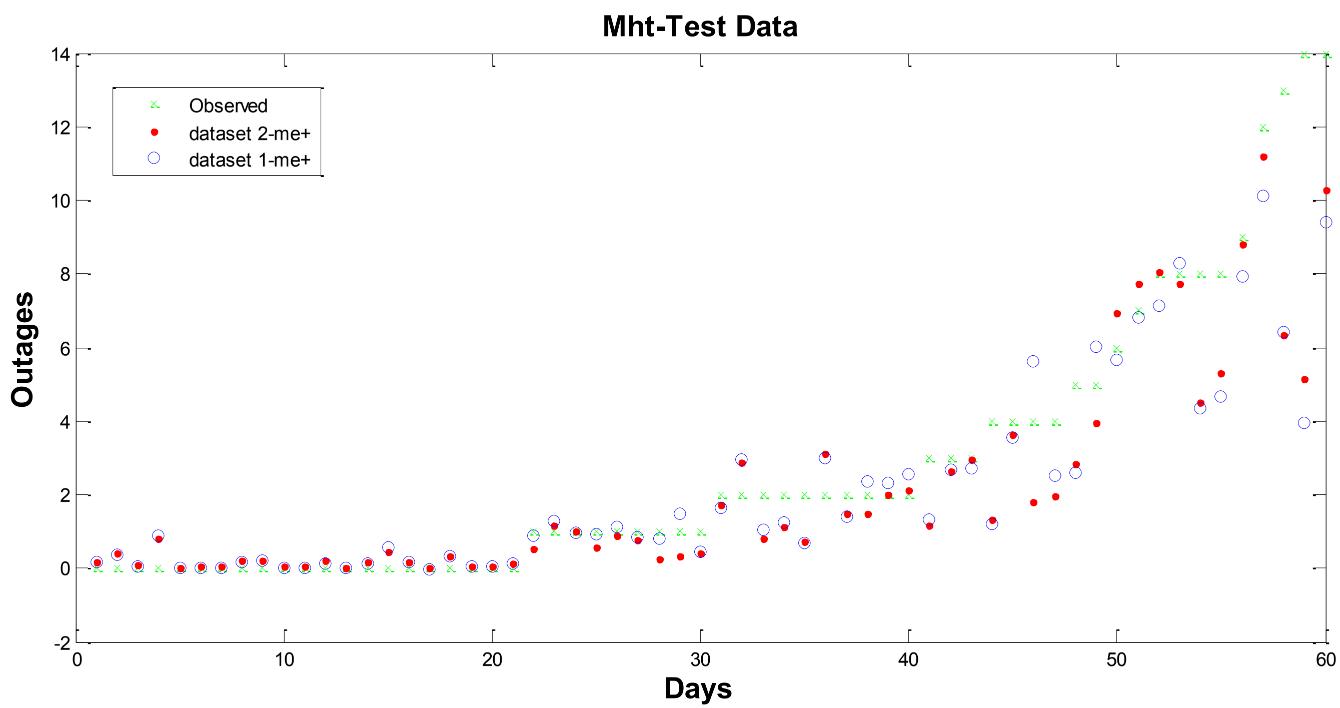


Figure 8.13 Observed and Estimated Outages for Dataset 1 & 2 for AME model for Manhattan Test Data

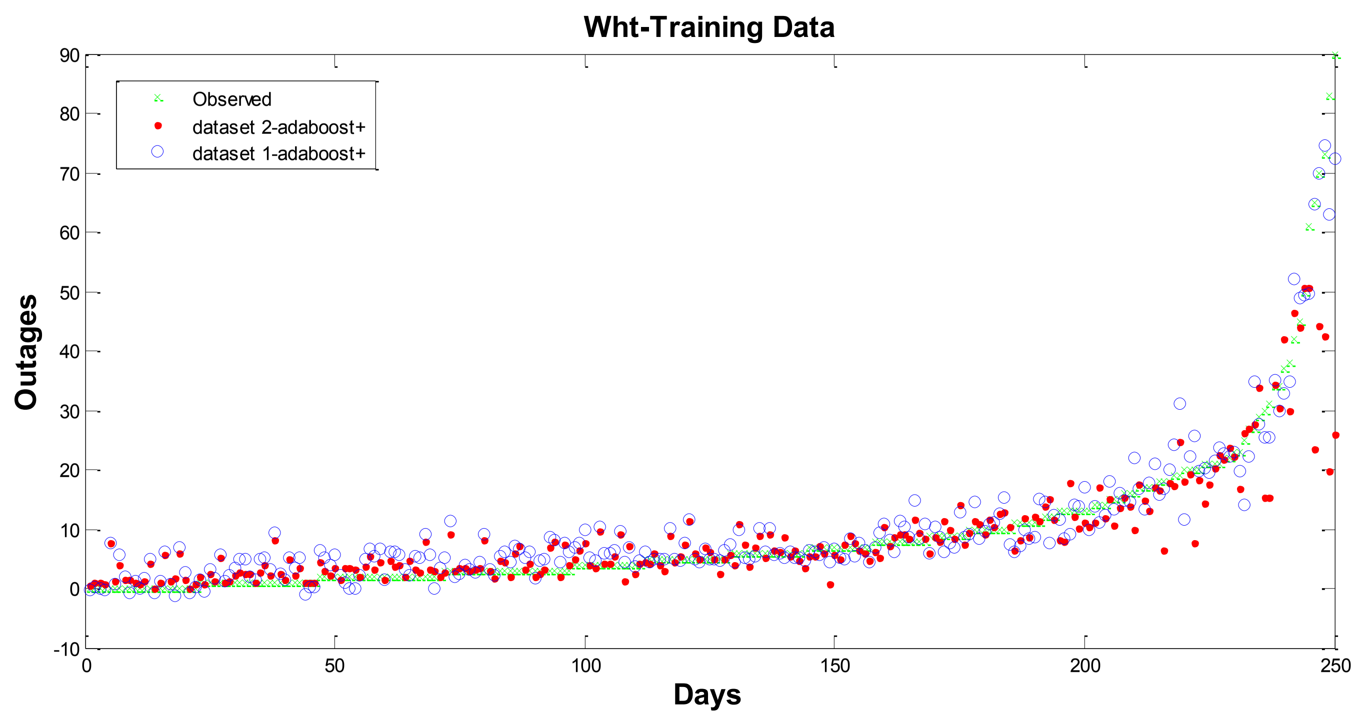


Figure 8.14 Observed and Estimated Outages for Dataset 1 & 2 for Adaboost^{+} model for Wichita Training Data

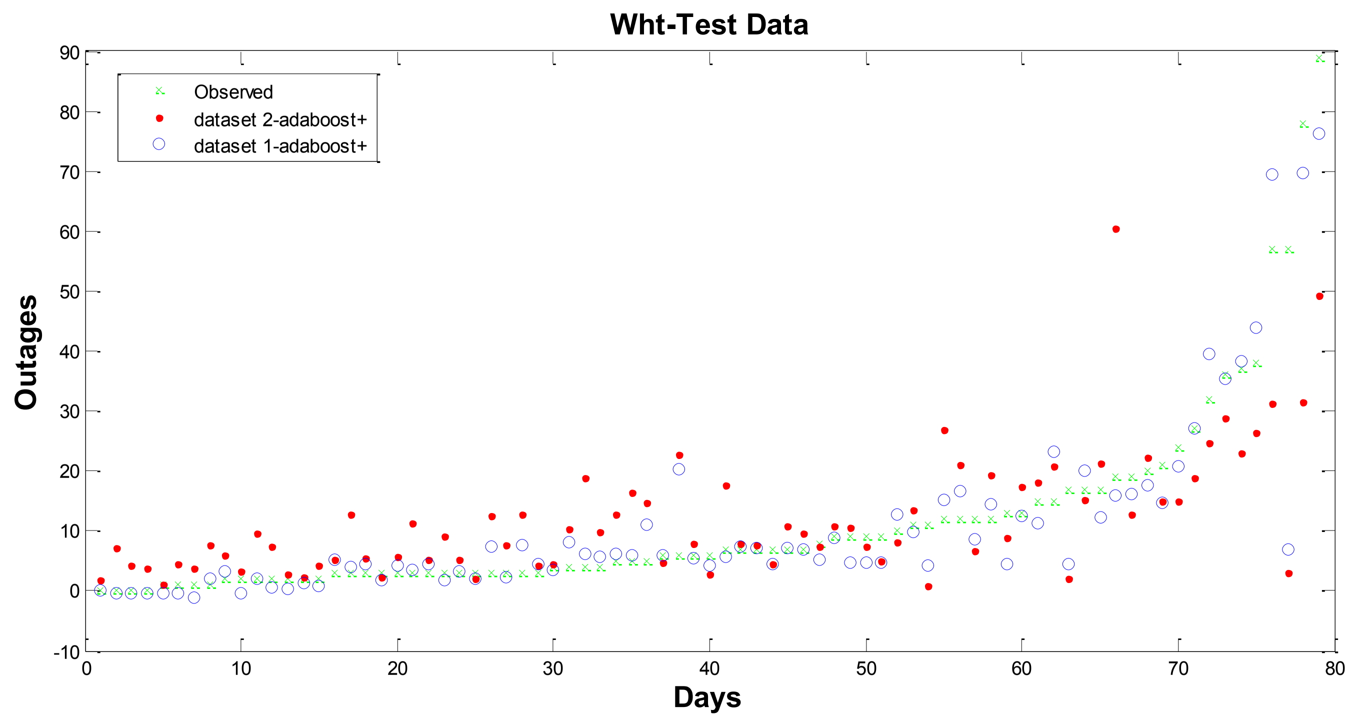


Figure 8.15 Observed and Estimated Outages for Dataset 1 & 2 for Adaboost+ model for Wichita Test Data

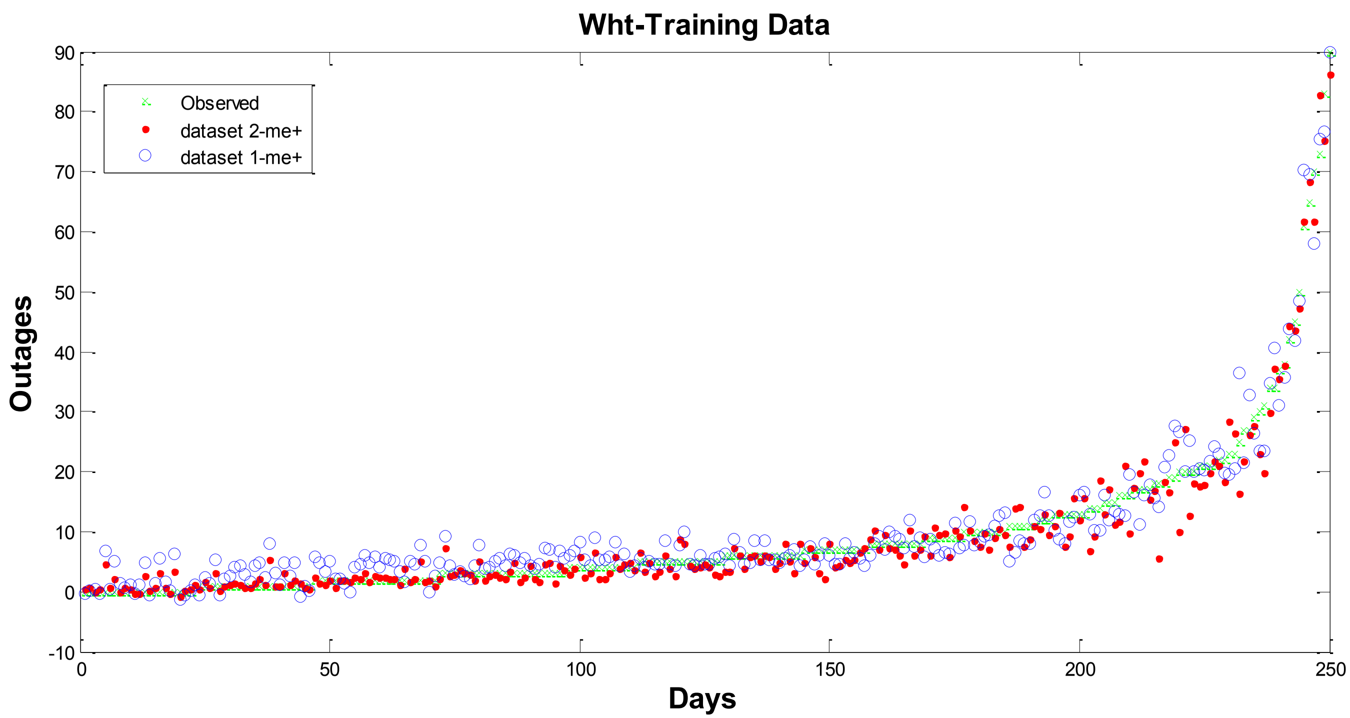


Figure 8.16 Observed and Estimated Outages for Dataset 1 & 2 for AME model for Wichita Training Data

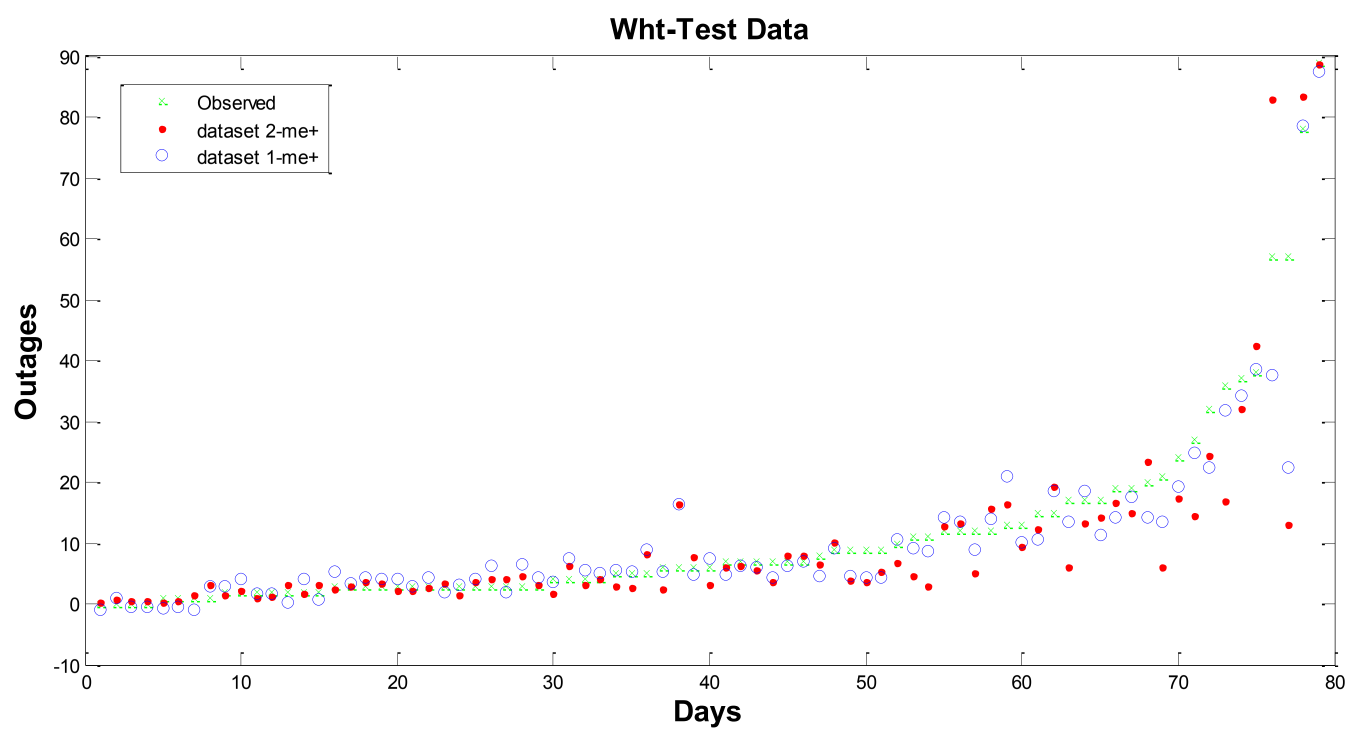


Figure 8.17 Observed and Estimated Outages for Dataset 1 & 2 for AME model for Wichita Test Data

8.4 Summary

In this chapter, only days that have lightning were considered to study the effect of lightning on distribution feeders outages. The six regression models, NN model, and ensemble models were investigated and performance of the models were evaluated by computing performance measures. The results for the days that had lightning were obtained with two models, one with all days included and the other with lightning days only. Results show that the models with only days with lightning as input have better performance and were able to capture the time series of daily observed outages more accurately.

Chapter 9 - Application of Models for Estimation of Outages for Only Non-Lightning Days

The better performance of model with only days that have lightning as input compared to the model with all days included as input, prompts for the investigation of models with days in which no lightning happened. A NN model and ensemble models were tested with available historical data from 2005 to 2011 with days that had no recorded lighting.

9.1 Data Overview

Table 9.1 shows the number of days on which no lightning occurred during the study period in the four cities.

Table 9.1 Number of non-Lightning Days from 2005 to 2011

Cities	Number of Days
Manhattan	2316
Lawrence	2267
Topeka	2241
Wichita	2226

9.2 Model Results

Since it was found that ensemble methods have better performance compared to traditional regression models, in this chapter, regression models were not investigated. A NN model, AdaBoost$^+$ and AME models were tested. The division of training and test data is also similar with 2005-2009 as training data and 2010-2011 as test data. The model construction, training, testing and performances evaluation are the same as done previously.

The MSE, MAE, R and S values for the training data and testing data for the four cities for the NN model, AdaBoost$^+$ and AME model are tabulated in Table 9.2.

Table 9.2 Summary of Model Results for four Cities for Non-Lightning Days

		Training Data				Test Data			
		MAE	**MSE**	**R**	**S**	**MAE**	**MSE**	**R**	**S**
Mht	**NN**	0.4404	1.1721	0.1667	0.1073	0.4955	1.1582	0.1747	0.0128
	AdaBoost^{+}	0.2379	1.2383	0.3726	0.1076	0.4654	1.2298	0.5472	0.0165
	AME	0.1634	0.5774	0.7394	0.4642	0.1882	0.4527	0.7772	0.4970
Lwr	**NN**	0.5791	4.0713	0.0947	0.5092	0.5204	1.1586	0.2990	0.0198
	AdaBoost^{+}	0.3124	3.5770	0.4060	0.6923	0.4193	0.9440	0.5982	0.1404
	AME	0.1414	0.3792	0.9528	0.9076	0.3270	0.5346	0.7429	0.5733
Tpk	**NN**	1.0559	6.5898	0.2626	0.2689	1.5973	12.7715	0.3315	0.1895
	AdaBoost^{+}	0.3943	3.1630	0.7496	0.5156	0.7692	5.7346	0.7759	0.5656
	AME	0.3504	2.8433	0.7765	0.5654	0.6874	5.2513	0.7940	0.6025
Wht	**NN**	2.2790	24.1046	0.3297	0.1110	2.6812	41.5480	0.5751	0.1749
	AdaBoost^{+}	1.0159	10.4977	0.7893	0.5867	1.2411	10.8465	0.9024	0.7635
	AME	0.8085	7.4343	0.8541	0.6881	0.9783	6.8607	0.9392	0.8174

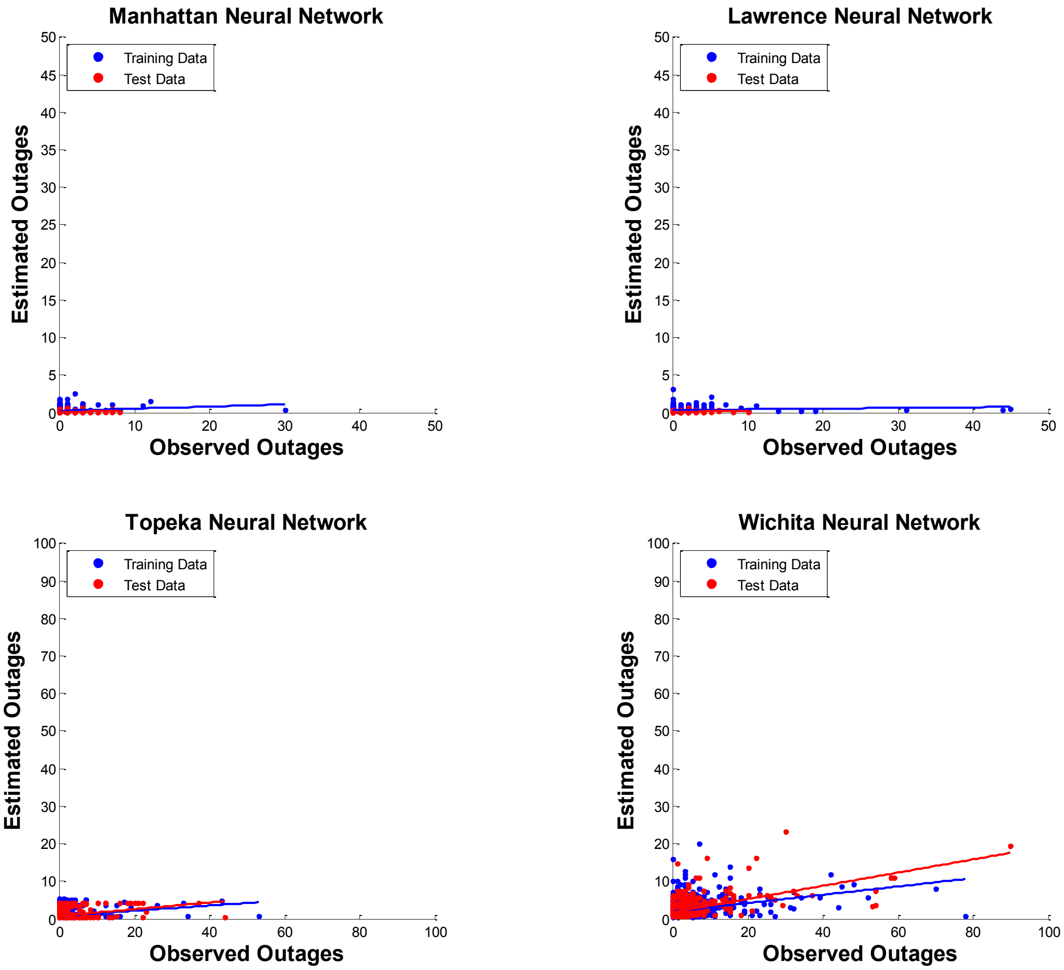


Figure 9.1 Plot of Observed and Estimated Weather-related Outages, for Non-lightning Days, Obtained with NN Model in Overhead Distribution Systems for Four Cities from 2005 to 2011

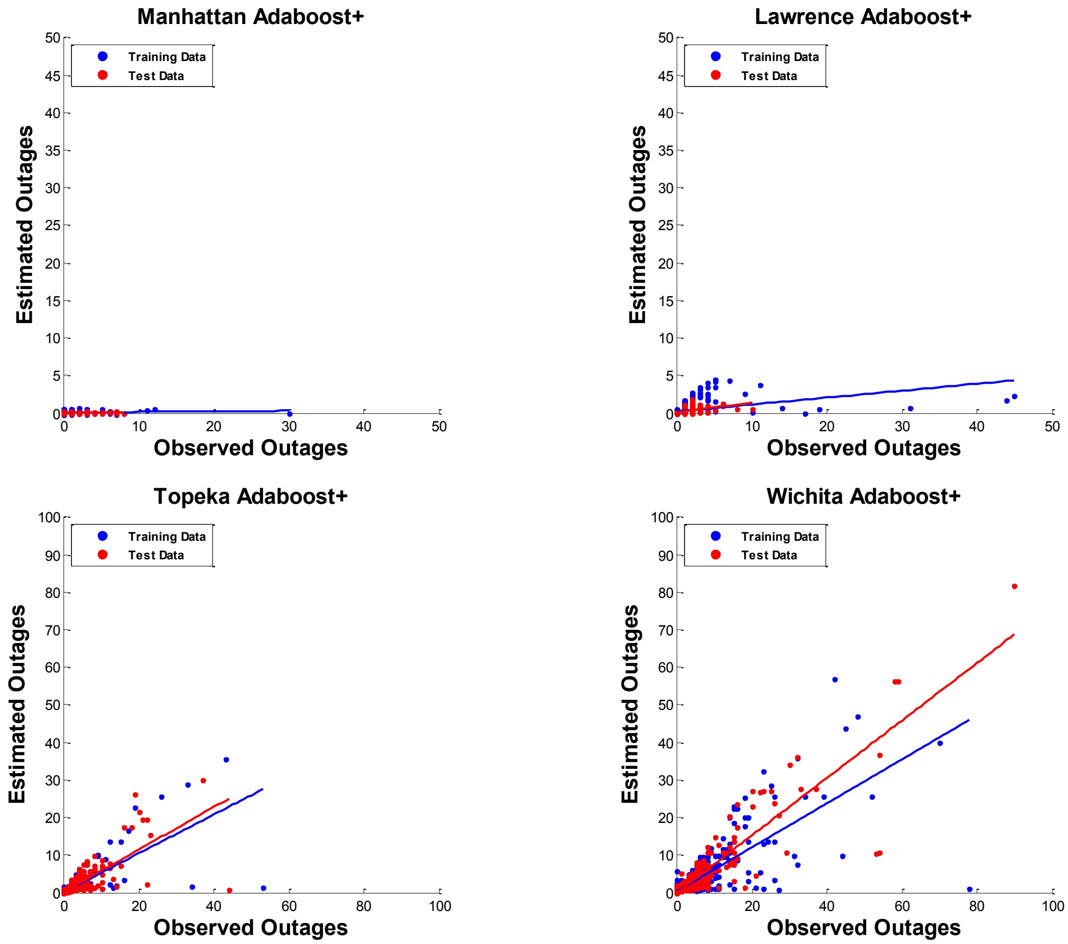


Figure 9.2 Plot of Observed and Estimated Weather-related Outages, for Non-lightning Days, Obtained with AdaBoost$^+$ Model in Overhead Distribution Systems for Four Cities from 2005 to 2011

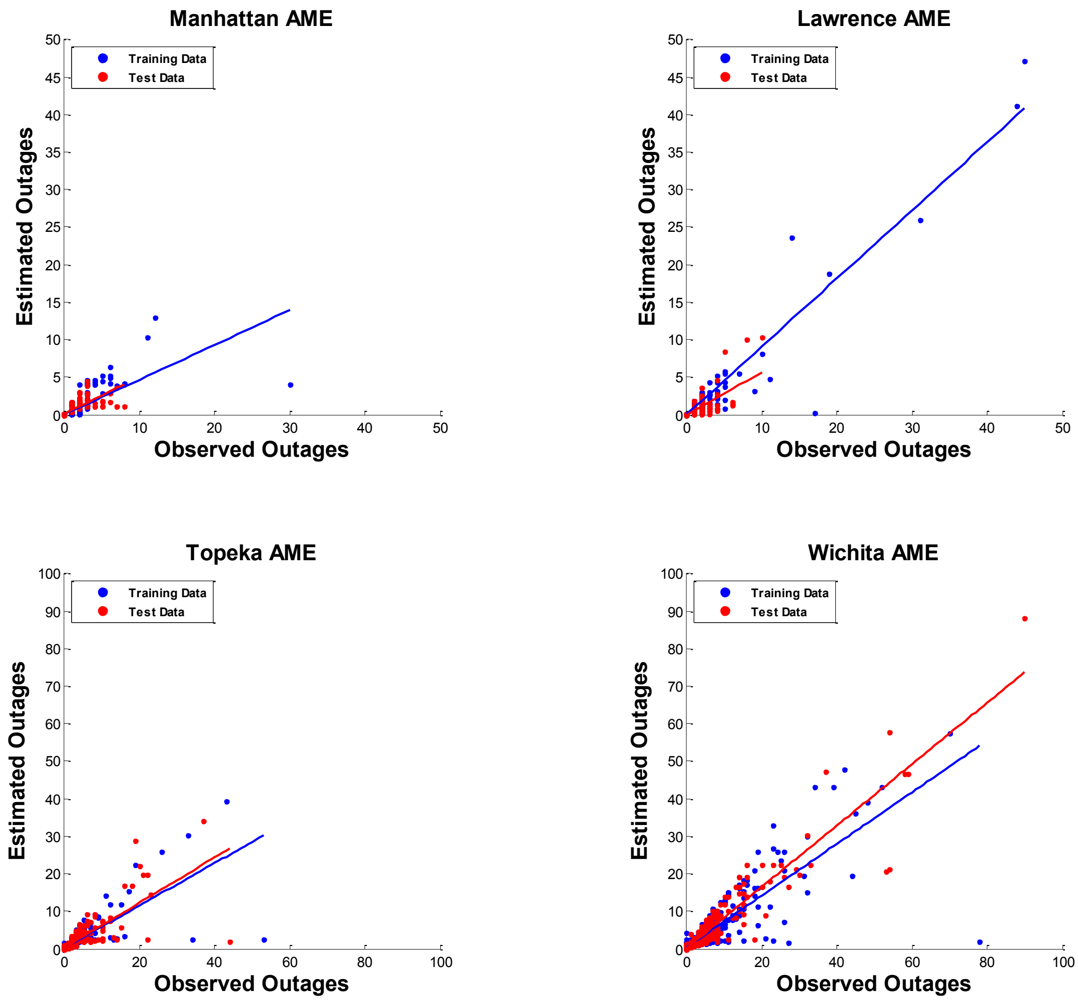


Figure 9.3 Plot of Observed and Estimated Weather-related Outages, for Non-lightning Days, Obtained with AME Model in Overhead Distribution Systems for Four Cities from 2005 to 2011

Figures 9.1 - 9.3 show the scatter plot between observed and estimated outages, with a regression line for the Neural Network, AdaBoost$^+$ and AME models for the four cities. The scatter plots for Manhattan and Lawrence for NN and AdaBoost$^+$ models are poor. One reason might be skewed data distribution and the model assigning too much weight onto a few hard-to-learn data points. As seen in Chapter 2, the Manhattan and Lawrence had a large number of days with zero observed outages compared to bigger cities, Topeka and Wichita.

A significant difference in MAE and MSE values between NN model and AME model can be observed for all four cities for training and testing data. Higher correlation and higher slope between observed and estimated outages indicates better performance of AME model.

For comparison, the results for days with no lightning were computed from the estimates for these obtained with models considering all days as input. The MAE, MSE, R and S values for all four cities for training and testing data obtained from these models are summarized in Table 9.3.

Table 9.3 Summary of Model Results for four Cities for All Days Included

		Training Data				Test Data			
		MAE	**MSE**	**R**	**S**	**MAE**	**MSE**	**R**	**S**
Mht	**NN**	0.4645	1.1658	0.1135	0.1182	0.4848	1.1246	0.2823	0.0274
	AdaBoost$^+$	0.2689	1.1458	0.3642	0.1292	0.4340	1.1734	0.5022	0.0331
	AME	0.2095	0.6218	0.7098	0.4220	0.2232	0.4386	0.7697	0.5441
Lwr	**NN**	0.5918	3.9089	0.1060	0.0168	0.6266	0.9545	0.2972	0.1333
	AdaBoost$^+$	0.2625	2.6718	0.6131	0.2381	0.2667	0.2500	0.8765	0.7711
	AME	0.1350	0.2073	0.9735	0.9301	0.2237	0.3224	0.8935	0.8368
Tpk	**NN**	1.1396	6.9203	0.2069	0.1689	1.5962	12.3802	0.3575	0.1060
	AdaBoost$^+$	0.5381	5.2743	0.5790	0.2059	0.9963	9.2594	0.6759	0.2757
	AME	0.3490	3.1200	0.7543	0.5214	0.7098	5.9804	0.7701	0.5161
Wht	**NN**	2.3784	24.0757	0.3460	0.1490	2.6863	47.0425	0.5773	0.1069
	AdaBoost$^+$	1.2290	12.5646	0.7604	0.4595	2.9745	37.0924	0.5668	0.3271
	AME	0.8314	8.2575	0.8378	0.6542	0.8787	14.4222	0.8787	0.6293

A comparison of the performance measures shows that the results for days without lightning are better if lightning days are removed from the data.

9.3 Comparative Analysis

Three different input datasets are defined as follows:

- Input dataset 1 – All days included
- Input dataset 2 – Only days with lightning

- Input dataset 3 – Only days with no lightning occurrence

The input dataset 1 contains all the days in the study period from 2005 to 2011.The input dataset 2 contains only the days with lightning recording in the study period. The input dataset 3 contains only the days with zero for lightning values. All the input datasets are divided into training data and test data with 2005 - 2009 as the training duration and 2010 – 2011as the test duration.

The six different output datasets which are considered are defined as follows:

- Output dataset 1 – estimated outages for all days with models trained using input dataset 1
- Output dataset 2 – estimated outages for days with lighting with models trained using input dataset 2
- Output dataset 3 – estimated outages for days without lightning with models trained using input dataset 3
- Output dataset 4 – estimated outages for days with lightning with models trained using input dataset 1
- Output dataset 5 - estimated outages for days that have no lightning with models trained using input dataset 1
- Output dataset 6 – estimated outages for all days with models trained separately using input datasets 2 and 3

Table 9.4 summarizes detailed categorization of different output datasets.

Table 9.4 Summary of Output Dataset

Output dataset	Input dataset for training (2005 -2009)	Input dataset for testing (2010-2011)	Model Output	Outputs considered for model performance evaluation
1	1	1	All days	All days
2	2	2	Only days with lightning recorded	Only days with lightning recorded
3	3	3	Only days with no lightning recorded	Only days with no lightning recorded
4	1	1	All days	Only days with lightning
5	1	1	All days	Only days with no lightning recorded
6	2	2	Only days with lightning recorded	Combined to get all Days
	3	3	Only days with wind	

The MSE, MSE, R and S values are computed between observed outages and six output datasets of a NN model, AdaBoost.RT, AdaBoost$^+$ and AME models. The performance measure values for output datasets 1, 2, 3, 4, 5 and 6 are tabulated in Table 9.5, 8.10, 9.2, 8.11, 9.3 and 9.6. Table 9.5 and 9.6 summarizes the performance measure values for output dataset 1 and 6 for all four cities for NN model, AdaBoost.RT, AdaBoost$^+$ and AME models. Results shows that model performs better for output dataset 6 compared to 1 that is the models perform better when trained separately with days with only lightning and only no lightning.

Comparing the tables 8.10, 8.11, 9.2, 9.3, 9.5, 9.6, the improvement in the model performance can be observed. One of the reason of using ensemble models stated earlier in the dissertation is the statistical reasons. The ensemble models allows the individual learner to specialize local on the input data space and output is obtained by combining the results of individual learners. Our results demonstrates significantly better performance of the ensemble learning methods compared to traditional regression models and NN model.

Table 9.5 Summary of Model Results for four Cities for Output dataset 1

		Training Data				Test Data			
		MAE	**MSE**	**R**	**S**	**MAE**	**MSE**	**R**	**S**
Mht	**NN**	0.6009	2.4879	0.6761	0.4555	0.6433	2.3370	0.4335	0.1254
	AdaBoost.RT	0.3789	1.9163	0.7813	0.5082	0.5578	2.0737	0.6216	0.1952
	AdaBoost⁺	0.3691	1.8251	0.7860	0.5694	0.5578	2.0558	0.6079	0.2083
	AME	0.2611	0.7044	0.9225	0.8406	0.2802	0.6679	0.8715	0.6400
Lwr	**NN**	0.6973	4.3621	0.2958	0.0871	0.8778	5.0176	0.4130	0.1012
	AdaBoost.RT	0.3884	3.5923	0.5150	0.2112	0.5316	3.6619	0.6662	0.2792
	AdaBoost⁺	0.3095	2.6210	0.6947	0.3662	0.4340	3.0395	0.7173	0.3825
	AME	0.1815	0.2603	0.9724	0.9414	0.3856	2.9584	0.7382	0.3743
Tpk	**NN**	1.3913	12.9613	0.5494	0.3016	2.4418	37.1506	0.4231	0.1909
	AdaBoost.RT	0.9053	11.7940	0.6434	0.3042	1.9056	32.0706	0.7026	0.2095
	AdaBoost⁺	0.7070	8.8922	0.7448	0.4409	1.4621	22.9910	0.7928	0.3530
	AME	0.4493	3.0957	0.9138	0.8276	1.0227	9.6359	0.8871	0.8172
Wht	**NN**	2.8051	35.9343	0.5756	0.3312	3.3873	76.7436	0.4003	0.2314
	AdaBoost.RT	1.6712	24.3932	0.7692	0.4539	2.4409	48.0201	0.7963	0.4145
	AdaBoost⁺	1.4615	18.4254	0.8263	0.5755	3.4672	49.9527	0.6401	0.3271
	AME	0.9677	8.1498	0.9225	0.8289	1.4358	18.6595	0.8883	0.7524

Table 9.6 Summary of Model Results for four Cities for Output dataset 6

		Training Data				Test Data			
		MAE	**MSE**	**R**	**S**	**MAE**	**MSE**	**R**	**S**
Mht	**NN**	0.6013	2.4831	0.6784	0.3905	0.6606	2.4647	0.3496	0.1072
	AdaBoost.RT	0.3126	1.7949	0.7903	0.5499	0.5711	2.0419	0.6194	0.2043
	AdaBoost$^+$	0.3097	1.7261	0.7955	0.5970	0.5694	2.0107	0.6083	0.2195
	AME	0.2078	0.6390	0.9294	0.8504	0.2516	0.7266	0.8674	0.6046
Lwr	**NN**	0.6951	4.3583	0.3082	0.0953	0.7420	4.5990	0.5073	0.1968
	AdaBoost.RT	0.5396	4.0956	0.4224	0.1436	0.6379	3.6254	0.6963	0.3007
	AdaBoost$^+$	0.3450	3.3817	0.5720	0.2669	0.5373	2.8289	0.7624	0.4289
	AME	0.1722	0.3762	0.9610	0.9217	0.4516	2.3559	0.7867	0.5390
Tpk	**NN**	1.3527	12.4057	0.5751	0.3311	2.3883	35.5507	0.4619	0.2173
	AdaBoost.RT	1.0009	10.6172	0.6769	0.3556	1.9180	27.7728	0.7279	0.2767
	AdaBoost$^+$	0.5066	3.2639	0.9084	0.8094	1.0538	9.2203	0.8918	0.8302
	AME	0.4475	2.8011	0.9217	0.8402	0.9915	9.4382	0.8882	0.8080
Wht	**NN**	2.7867	35.5654	0.5810	0.3386	3.3631	72.2817	0.4386	0.2697
	AdaBoost.RT	1.4610	19.086	0.8135	0.5728	1.7961	26.5714	0.8646	0.5648
	AdaBoost$^+$	1.2217	10.8976	0.8937	0.7851	1.4794	15.0616	0.9101	0.8080
	AME	0.0944	7.6422	0.9262	0.8556	1.1823	9.2845	0.9473	0.8330

9.4 Summary

In this chapter, the models were trained with days that have no lightning and performance of the models were discussed. Three different input datasets and six different output datasets were defined. Comparison of different models performance with these input and output datasets are discussed. Overall, the models trained with separate data, days with lightning and no lightning, perform better when compared to models trained with all day in the data.

Chapter 10 - Application of CM Models for Estimation of Animal-Related Outages

Gui, Pahwa and Das found the correlation between animal activity and animal-caused outages by analyzing historical outage information for different months of the year under different weather conditions and relating it to behavioral patterns of animals [23-25]. A Poison model, NN model, wavelet-based NN model, and a Bayesian model combined with Monte Carlo simulation are presented in [23-25]. The models were trained with historical data for the four cities in Kansas, Manhattan, Lawrence, Topeka, and Wichita, from 1998 to 2006 and tested for 2007. The data was aggregated on a weekly basis since a larger sample size evened out some of the randomness in the daily data. In this chapter, the ensemble learning models are applied for estimation of animal-related outages on overhead distribution systems to test their performance for animal-related outages.

10.1 Data Overview

The same historical data used by Gui, Pahwa and Das was considered to evaluate performance of the ensemble models [23-25]. Figure 10.1 illustrates that, under fair weather conditions, animal activity is the most significant cause of outages as compared to other factors. Fair weather days have temperature within 40 and 85 degrees F with no other weather activity. From historical data it is observed that animals are most active in fair weather [23]. When there is strong wind, ice, thunder storm and other unfavourable weather conditions, they stay in their nests [23]. Additionally, animals have different behavioral patterns in different months of the year and thus months have considerable impacts on animal-related outages in overhead distribution system. The months are grouped based on low, medium and high level of animal activity and classified as Month type 1, 2, and 3 as follows:

- Month Type 1: January, February, March;
- Month Type 2: April, July, August, December ;
- Month Type 3: May, June, September, October, and November.

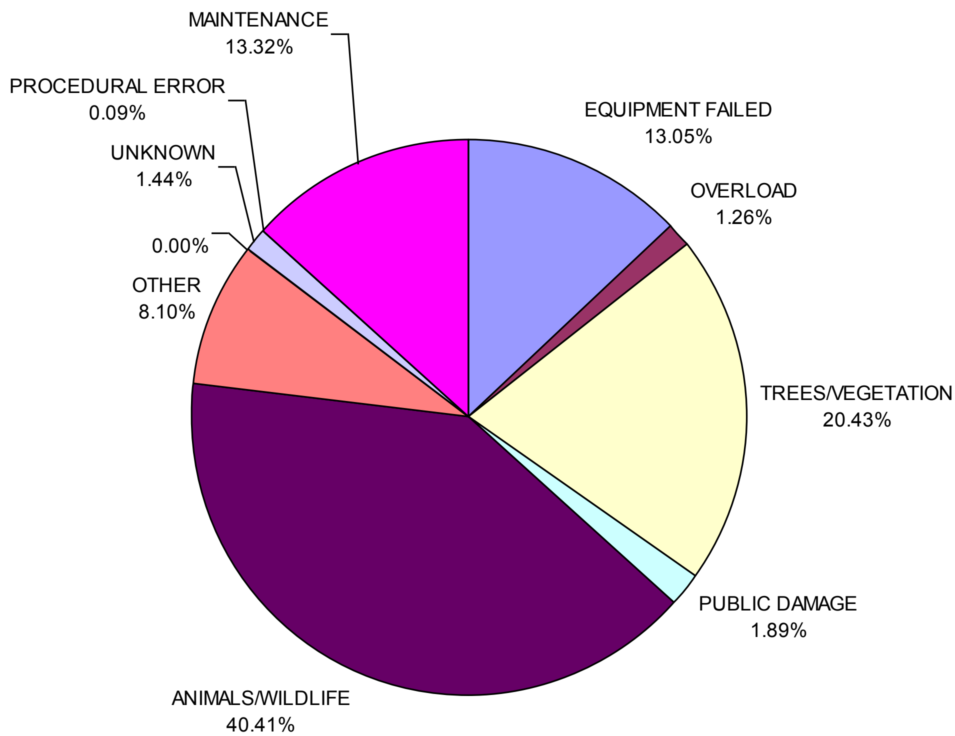


Figure 10.1 Percentage of Outages Caused under Fair Weather Conditions between 2003 and 2004 [23]

Since a month can have 28, 29, 30, or 31 days, even allocation of the weeks in a particular month is difficult. In order to ensure that every week belongs to only one month, some weeks have eight days [23]. For months that have 31 days, the first week has seven days and the remaining three weeks each have eight days. For months that have 30 days, the first two weeks each have seven days and the other two weeks each have eight days. In February, which typically has 28 days, each week has seven days. If it is a leap year, the last week of February has eight days. Classification of week as mentioned above does not impact results because both the input state and output have the same classifications for weeks [23]. Based on this classifications, the number of fair weather days per week can vary from zero to seven or eight [23]. Since the number of fair days in a week impacts outage occurrences in that week, the number of fair days per week is used as an input factor in the models for weekly animal-related outages. Also, the month type of the month in which that week lies is taken as the second input factor for weekly animal-related outages. Detailed information, such as animal characteristics, outages caused by animals, and data processing can be found in [23-25].

10.2 Previous Neural Network Model

Structure of the NN model presented by Gui, Pahwa and Das for outage estimations is shown in Figure 10.2 [23]. The model has the number of fair days per week, month type, and outages from the previous four weeks as inputs.

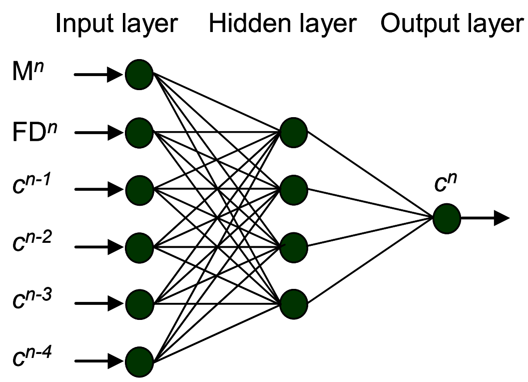


Figure 10.2 Structure of NN Model [23]

With a single output node, this model gives one step ahead estimation. Input and target data are normalized between 0.1-0.9. The learning rate is considered as 0.5, momentum as 0.2, and optimum training times as 3000. Simulations were carried out for the four cities, and performances of the model were measured using the mean absolute error (MAE), given in Table 10.1. The correlation, R and slope, S between the observed and estimated outages tabulated are the overall R and S for training and test data. The results show that the model performs better with increase in the size of the cities.

Table 10.1 Previous NN Model Results for Four Cities [23]

NN Model	Training MAE	Testing MAE	R	S
Manhattan	1.93	2.09	0.29	0.76
Lawrence	2.74	3.38	0.36	0.86
Topeka	6.29	6.94	0.76	0.90
Wichita	7.67	6.38	0.69	0.90

10.3 Model Construction & Simulation Results

In this section, the construction of various models for estimation of outages caused by animals is discussed and experimental results are presented. The NN model, AdaBoost models, and AME model were considered in this study.

10.3.1 Neural Network

The most common three-layered feed forward neural network topology was used, which is able to adequately approximate nonlinear functions with sufficient accuracy. The network has a single hidden layer with sigmoid activation functions and is trained in the batch mode according to the error back-propagation algorithm with gradient decent. The number of fair days per week and month type are the two feature-related inputs to the neural network. Furthermore, outages of previous weeks are taken as additional inputs since similar patterns are observed in historical data. Thus the NN can learn the patterns and predict future outages based on learned patterns. By experimentation, Gui found the four previous week outages as a suitable number for additional inputs [23]. To make the model computationally simple, only one previous week outage is considered as additional input in this dissertation. A 3x2x1 multilayered feed forward NN model structure considered in this work is shown in Figure 10.3, whereas Gui considered 6x4x1 NN model, as shown in Figure 10.2.

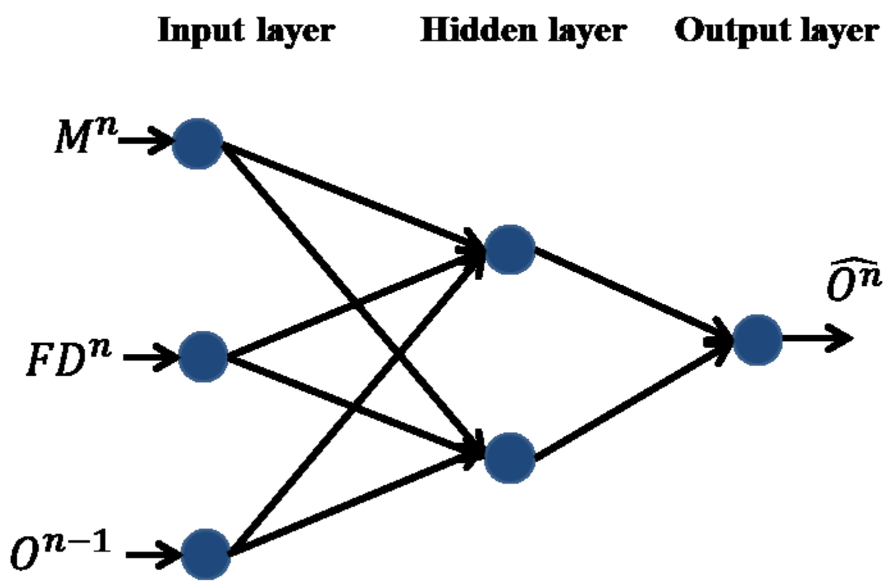


Figure 10.3 Three-layer Feed-Forward NN Model

where,

M^n: the month type index of the forecasting week n;

FD^n: the number of fair days during the forecasting week n;

O^{n-1}: values of the time series O for *one* week before the forecasting week n;

O^n: estimated output of the time series O of week n.

Gui, Pahwa and Das also considered the sigmoidal function in the hidden and output layer. Learning rate, momentum and the optimum number of training times were 0.5, 0.2 and 3000, respectively. In this work, sigmoidal function in the hidden layer, linear function in the output layer, the learning rate of 0.01, momentum of 0 and 3000 as the number of training times were considered.

Table 10.2 Results of Neural Network Model for Four Cities

Neural Network								
	Training Data				**Test Data**			
	MSE	**MAE**	**R**	**S**	**MSE**	**MAE**	**R**	**S**
Manhattan	9.8982	2.1843	0.5593	0.3021	5.1144	1.8683	0.3885	0.2867
Lawrence	15.8521	3.1250	0.6857	0.4680	19.7885	3.7601	0.4263	0.4075
Topeka	84.4520	6.8084	0.7729	0.5896	54.3965	5.7810	0.7552	0.7059
Wichita	154.8318	8.6310	0.7800	0.5912	68.8767	6.2469	0.7399	0.9819

The NN model was trained with historical data for the four cities from 1998 to 2006 and tested for 2007. Model performances were measured using the MAE, MSE, correlation, R, and slope, S given in Table 10.2. Figure 8.4 shows the scatter plot of observed and estimated outages with a regression line for the four cities. MAE for the training and test data obtained by Gui are shown in Table 10.1. It can be observed that the 6x4x1 model has slightly lower MAE values compared to 3x2x1 model. However, to use the NN model as the base learner in ensemble learning methods, it is kept computationally simple by using the 3x2x1 model.

Figures 10.5-10.8 show the plot of weekly observed outages and estimated outages for Manhattan, Lawrence, Topeka, and Wichita. The figures indicate that the model can reproduce basic patterns of the time series quite well in training and testing durations. However, it still underestimates outages in the higher range and overestimates in the lower range, shown at the high peaks and base in the time series in Figures 10.5-10.8.

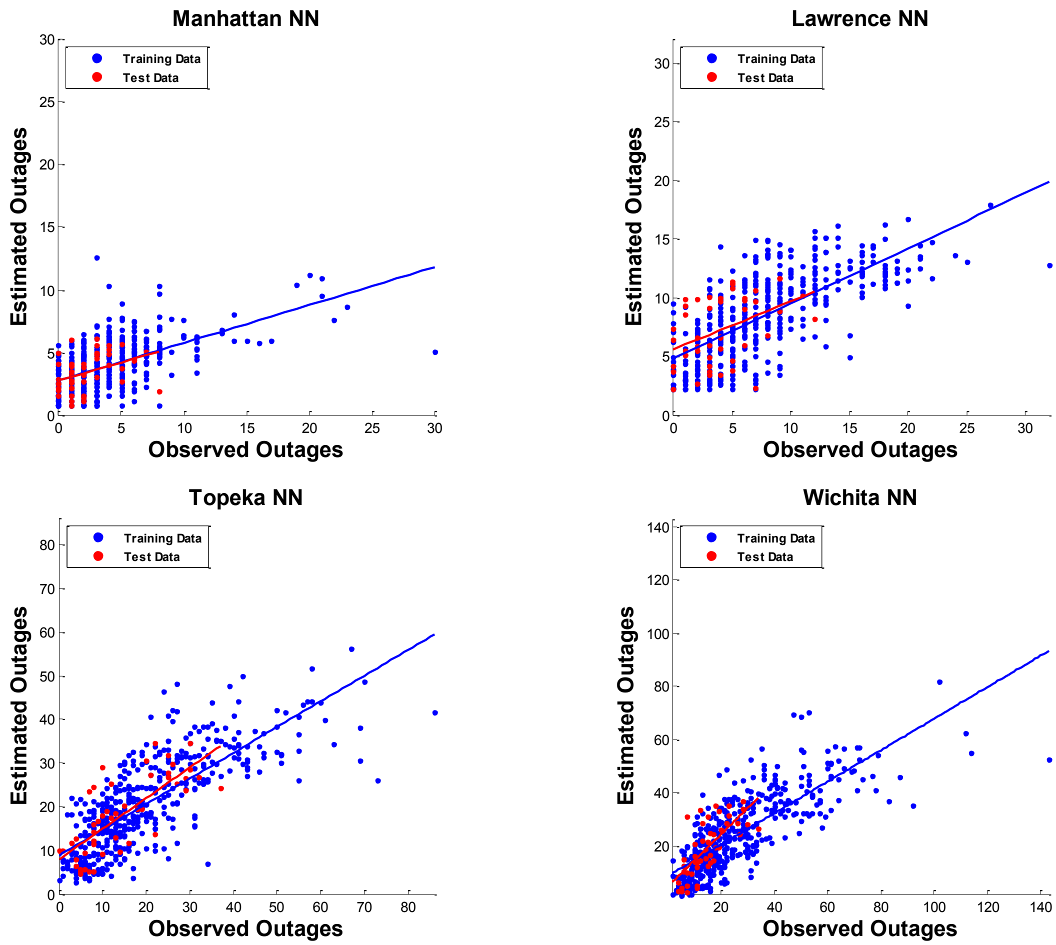


Figure 10.4 Plot of Observed and Estimated Animal-related Outages Obtained with Neural Network Model in Overhead Distribution Systems for Four Cities from 1998 to 2007

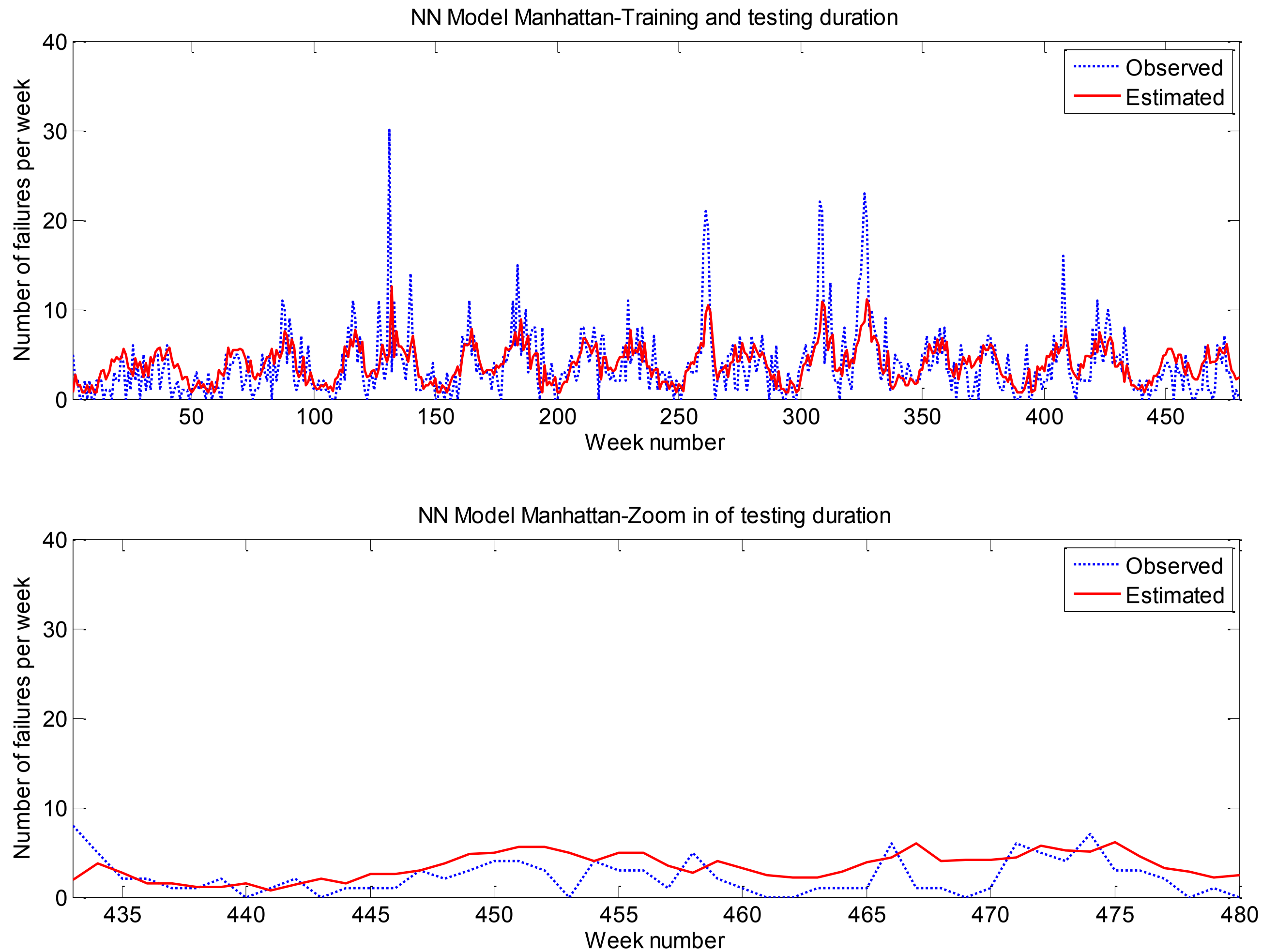


Figure 10.5 Animal-related Outages Observed and Estimated by NN Model in Manhattan from 1998 to 2007

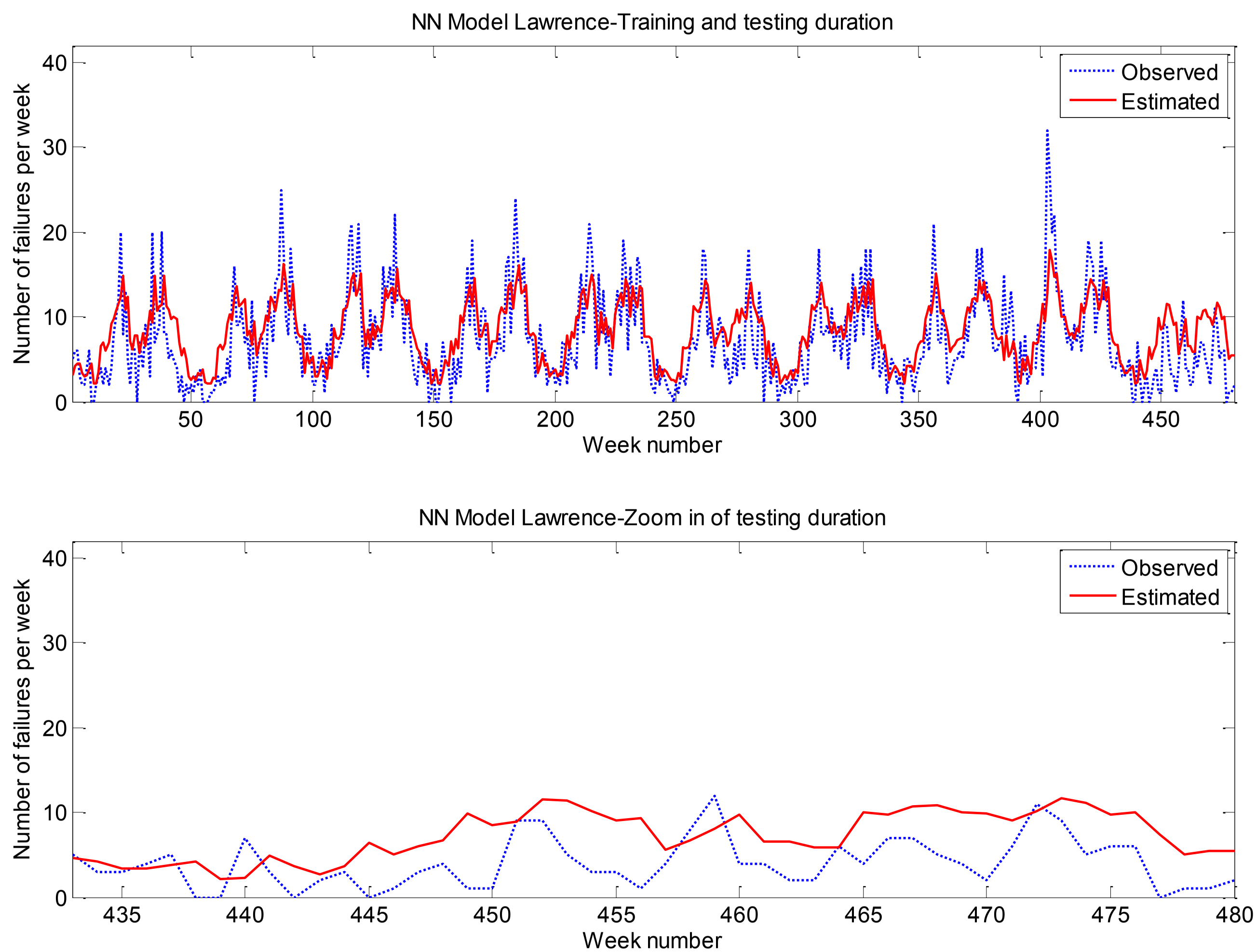


Figure 10.6 Animal-related Outages Observed and Estimated by NN Model in Lawrence from 1998 to 2007

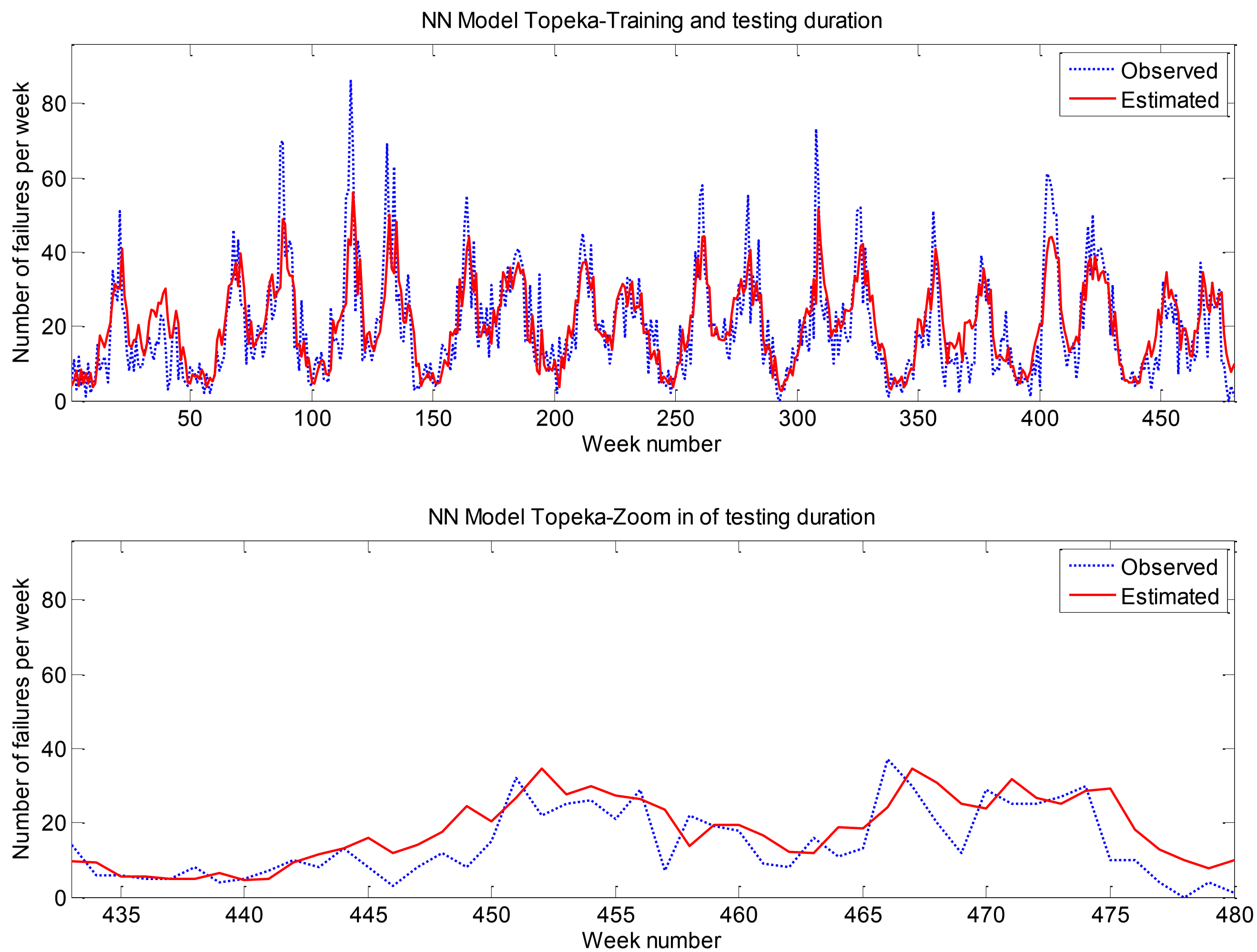


Figure 10.7 Animal-related Outages Observed and Estimated by NN Model in Topeka from 1998 to 2007

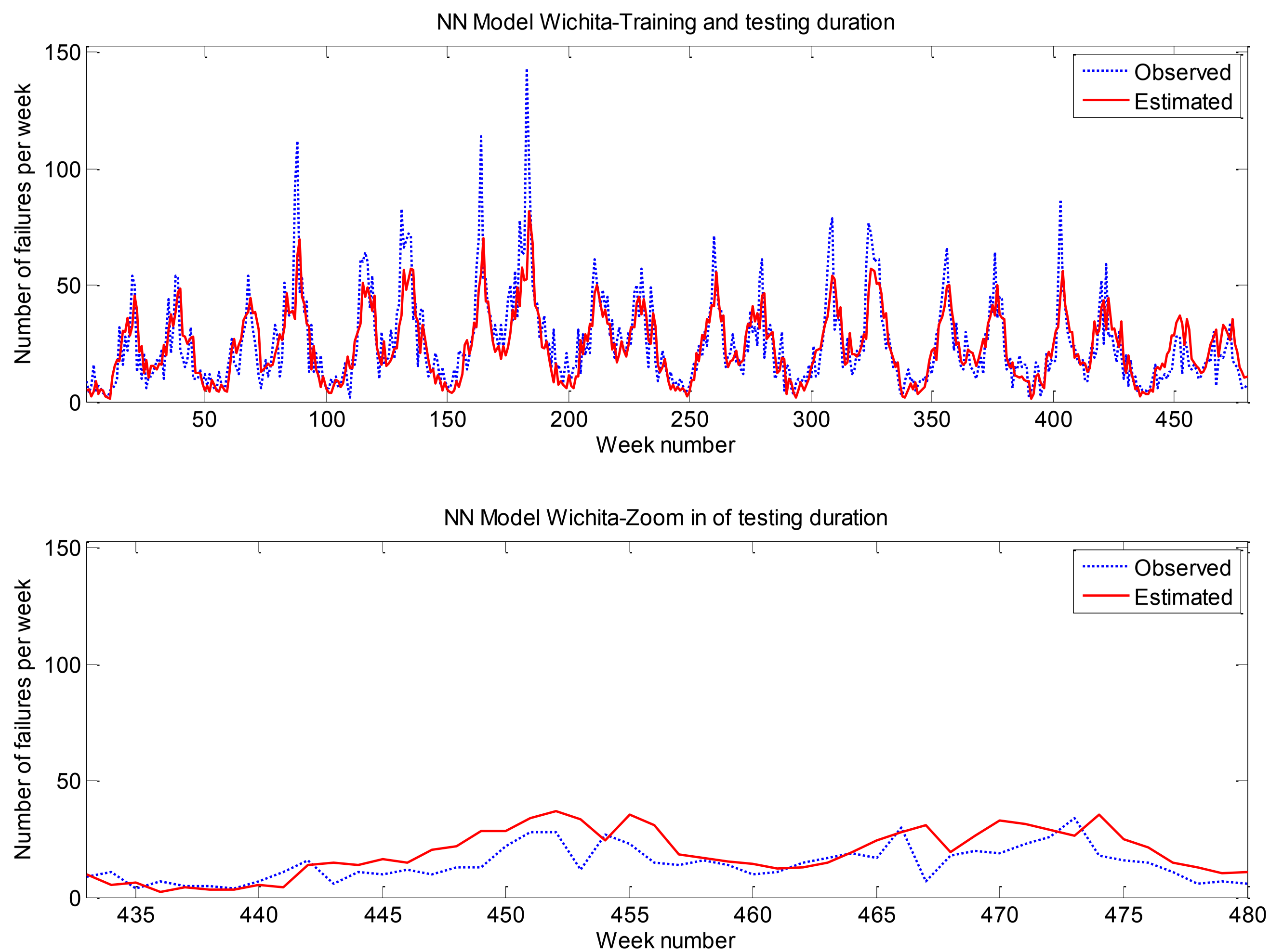


Figure 10.8 Animal-related Outages Observed and Estimated by NN Model in Wichita from 1998 to 2007

10.3.2 AdaBoost Model

The weekly animal-related outages time series is fluctuant and seasonal. In addition, noise could have deteriorated performance of the NN model. Ensemble learning methods may be used to approximate characteristics of the time series where each learner can be trained in one specific region. The AdaBoost structure, shown in Figure 6.1, is used here. In this work, four base-learners are considered. The base learner is the NN model presented in Section 10.3.1 with the structure shown in Figure 10.3. The training method is the same as that presented in Chapter 6. In this proposed model, each NN model is trained in one specific input region and outputs of the NN models are combined using the combining rule specified in Chapter 5 to find the final estimation for the original series.

The results are tabulated in Tables 10.3 and 10.4. The tables indicate that AdaBoost$^+$ has better performance, followed by AdaBoost.RT compared to an NN model with smaller MSE and MAE values and high correlation coefficient in training and test durations.

Figures 10.9 and 10.14 show the scatter plot of observed and estimated outages with regression line for the four cities for AdaBoost.RT and AdaBoost$^+$ models. Figures 10.10-10.13 and Figures 10.15-10.18 show the plot of weekly observed outages and estimated outages for Manhattan, Lawrence, Topeka, and Wichita for the AdaBoost.RT and AdaBoost$^+$ model. The AdaBoost.RT and AdaBoost$^+$ models are able to reproduce fluctuating patterns of the time series quite well in training and testing durations; however, they still underestimate outages in the higher range.

Table 10.3 Results of Adaboost.RT Model for Four Cities

Adaboost.RT								
	Training Data				**Test Data**			
	MSE	**MAE**	**R**	**S**	**MSE**	**MAE**	**R**	**S**
Manhattan	3.3883	0.9927	0.8836	0.6686	2.1028	0.8251	0.7125	0.4467
Lawrence	7.7892	1.8940	0.8723	0.6501	4.5417	1.4592	0.7298	0.3919
Topeka	50.0755	3.3548	0.8893	0.6751	41.7213	4.3797	0.8809	0.5057
Wichita	55.2318	4.0011	0.9292	0.7777	21.6314	3.1175	0.8479	0.5773

Table 10.4 Results of Adaboost^{+} Model for Four Cities

Adaboost^{+}								
	Training Data				Test Data			
	MSE	MAE	R	S	MSE	MAE	R	S
Manhattan	1.6513	0.8592	0.9518	0.8219	1.3111	0.6566	0.8325	0.5847
Lawrence	3.5305	1.3293	0.9407	0.8170	2.4946	0.9759	0.8594	0.6102
Topeka	22.0193	2.8002	0.9485	0.8337	20.9659	2.6823	0.9589	0.7065
Wichita	33.8051	3.2699	0.9569	0.8637	12.5119	2.2847	0.9082	0.7104

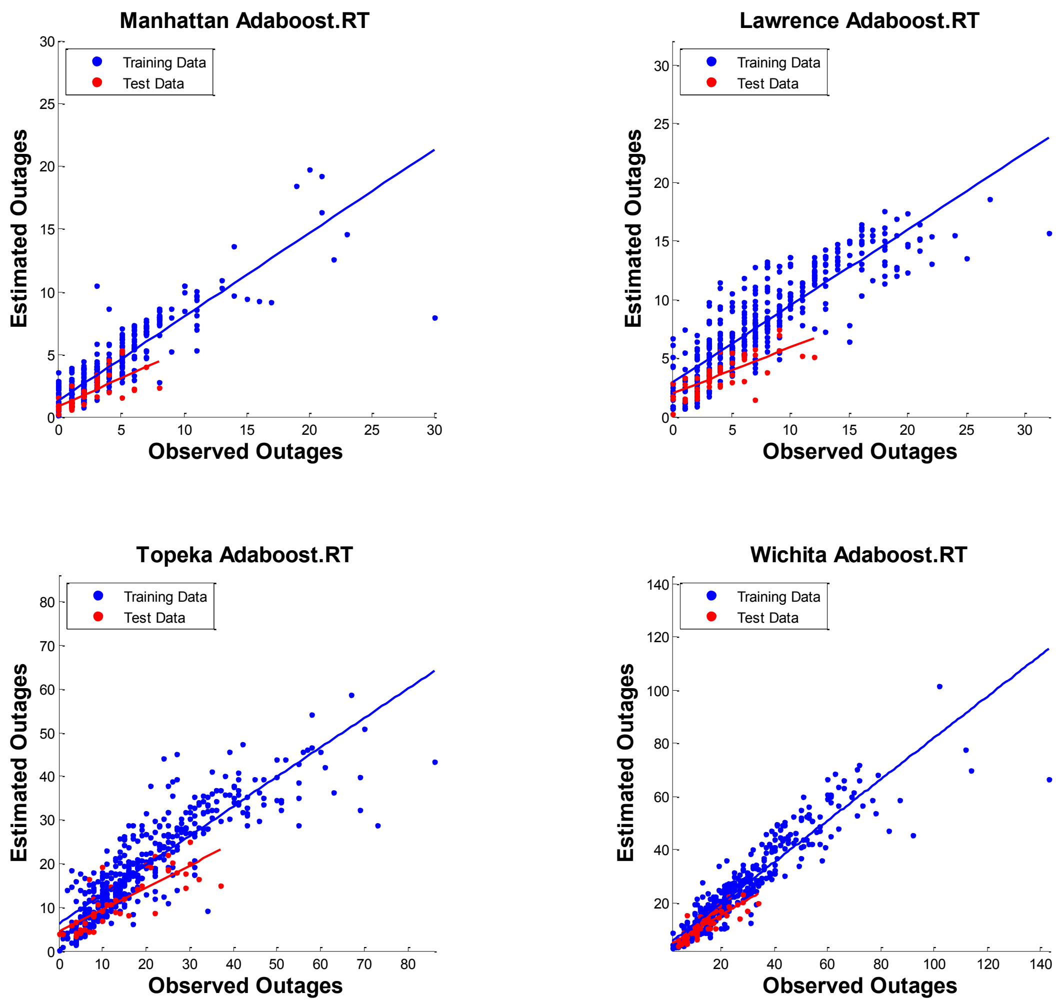


Figure 10.9 Plot of Observed and Estimated Animal-related Outages Obtained with Adaboost.RT Model in Overhead Distribution Systems for Four Cities from 1998 to 2007

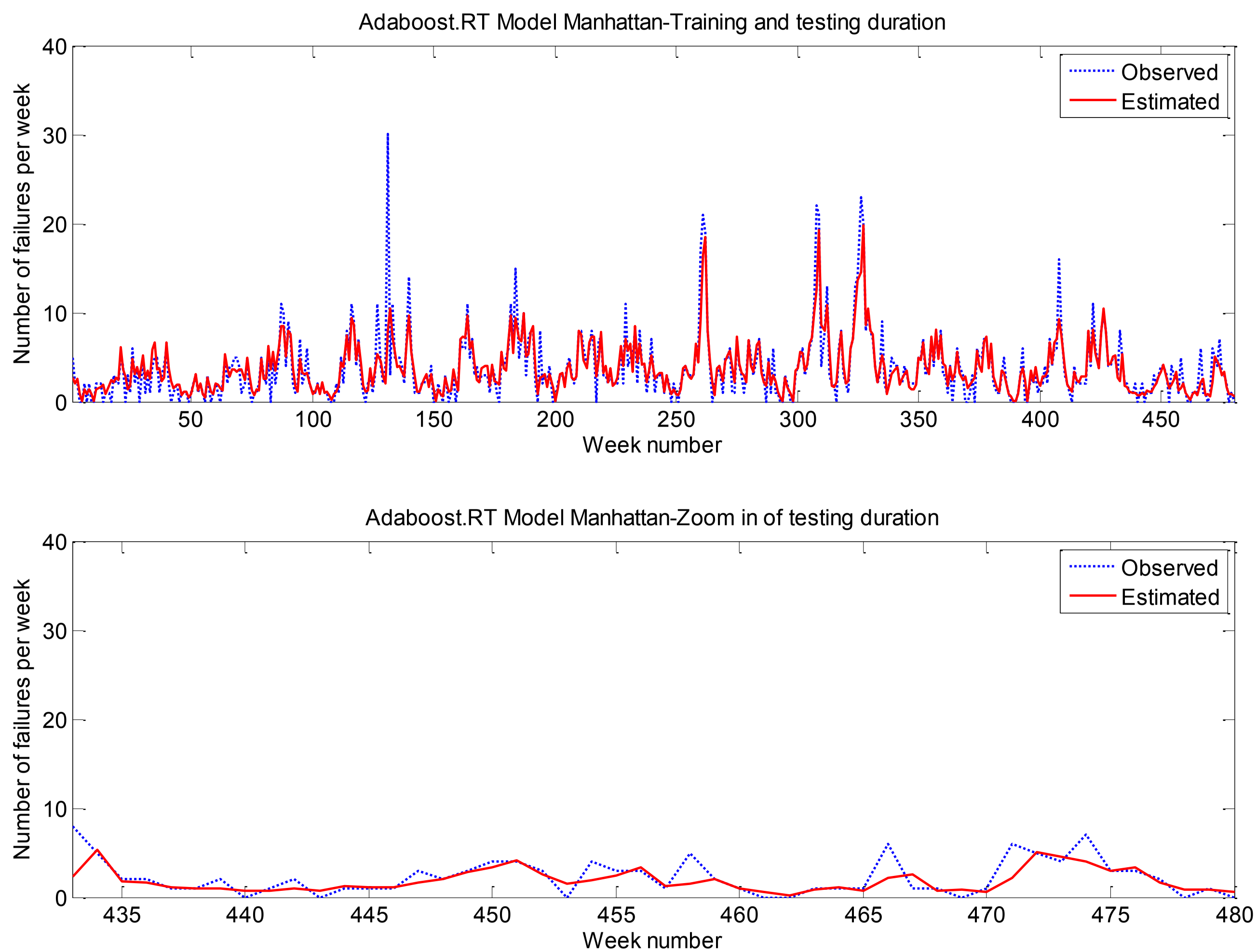


Figure 10.10 Animal-related Outages Observed and Estimated by AdaBoost.RT Model in Manhattan from 1998 to 2007

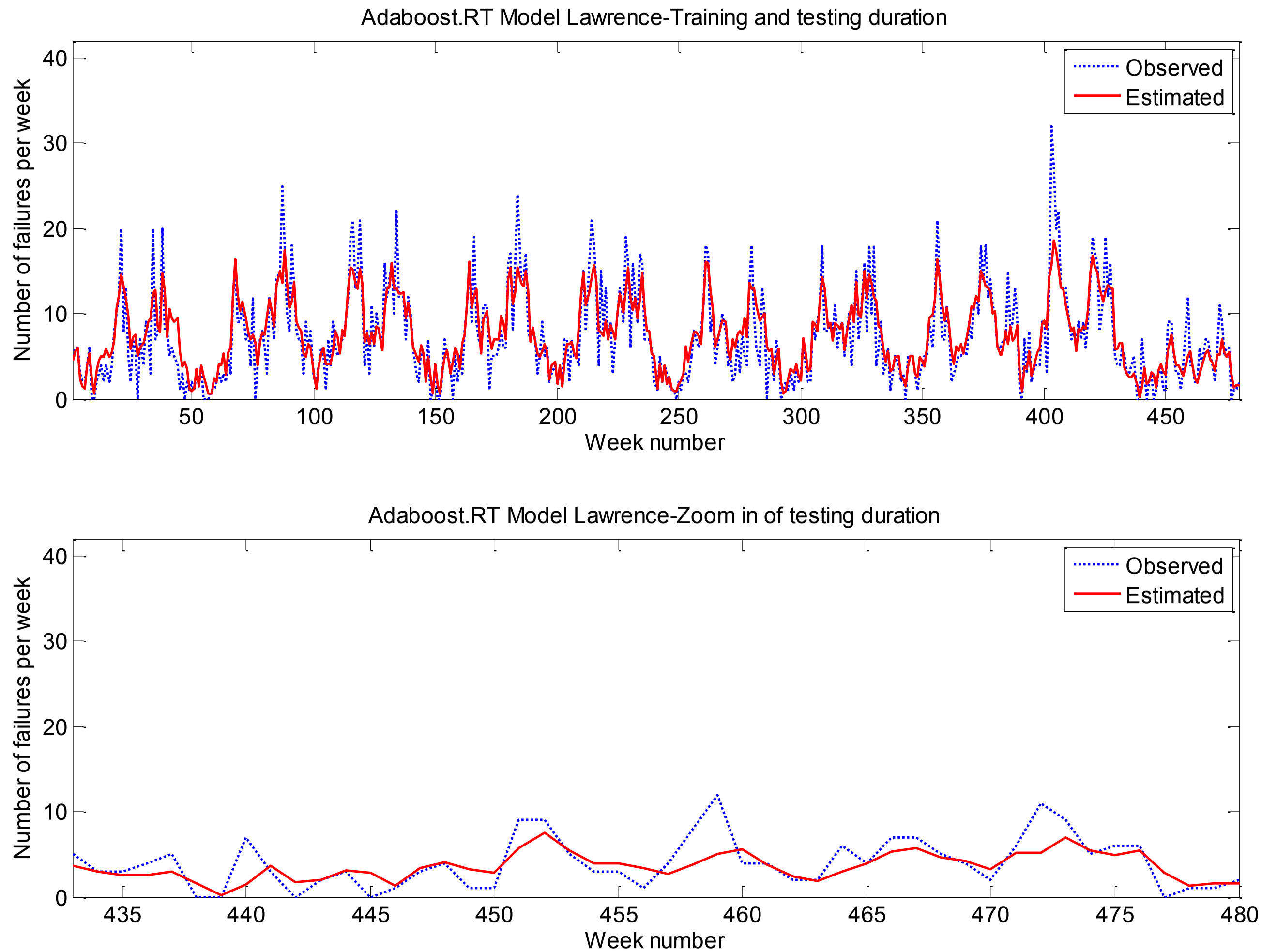


Figure 10.11 Animal-related Outages Observed and Estimated by AdaBoost.RT Model in Lawrence from 1998 to 2007

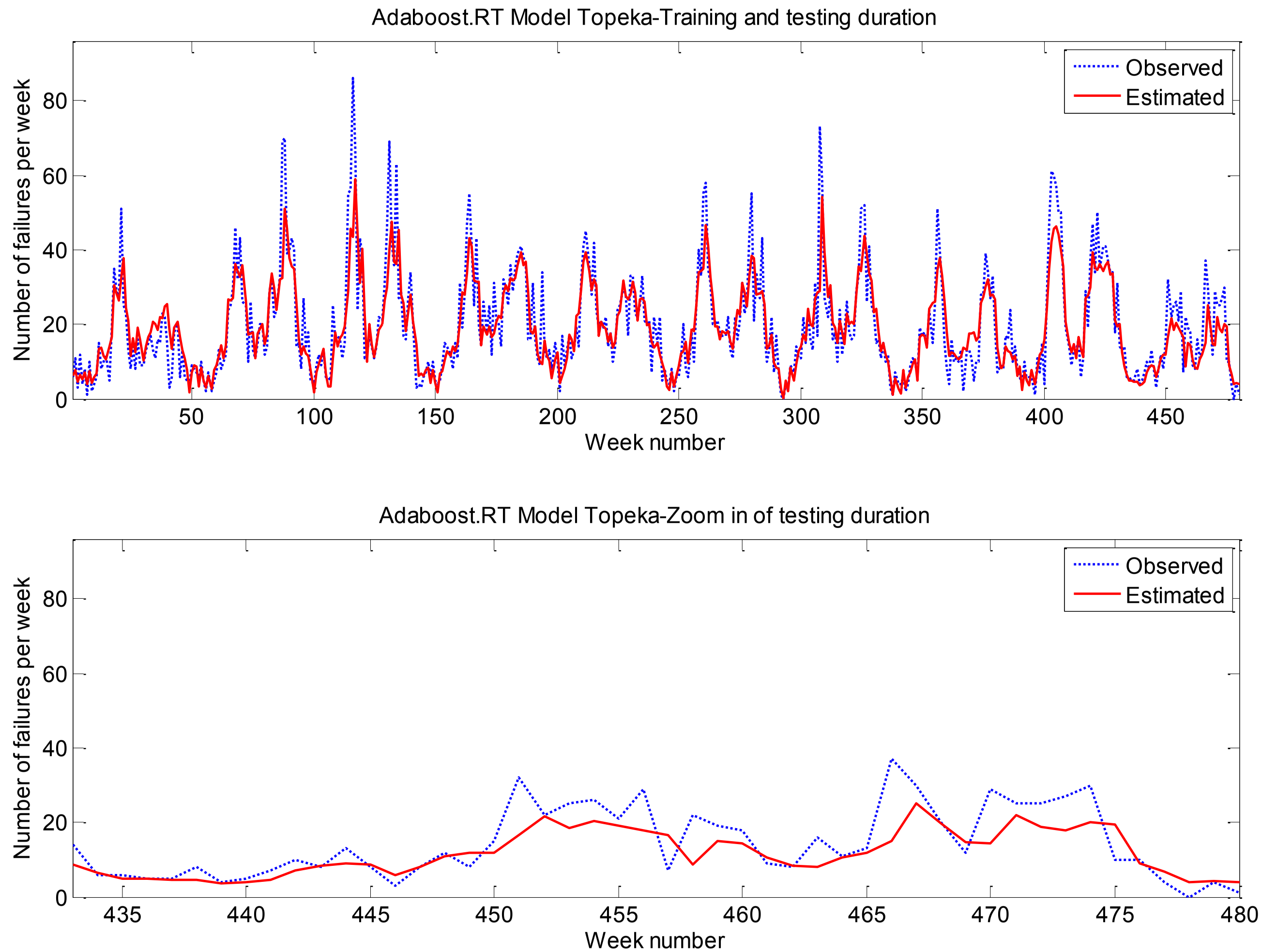


Figure 10.12 Animal-related Outages Observed and Estimated by AdaBoost.RT Model in Topeka from 1998 to 2007

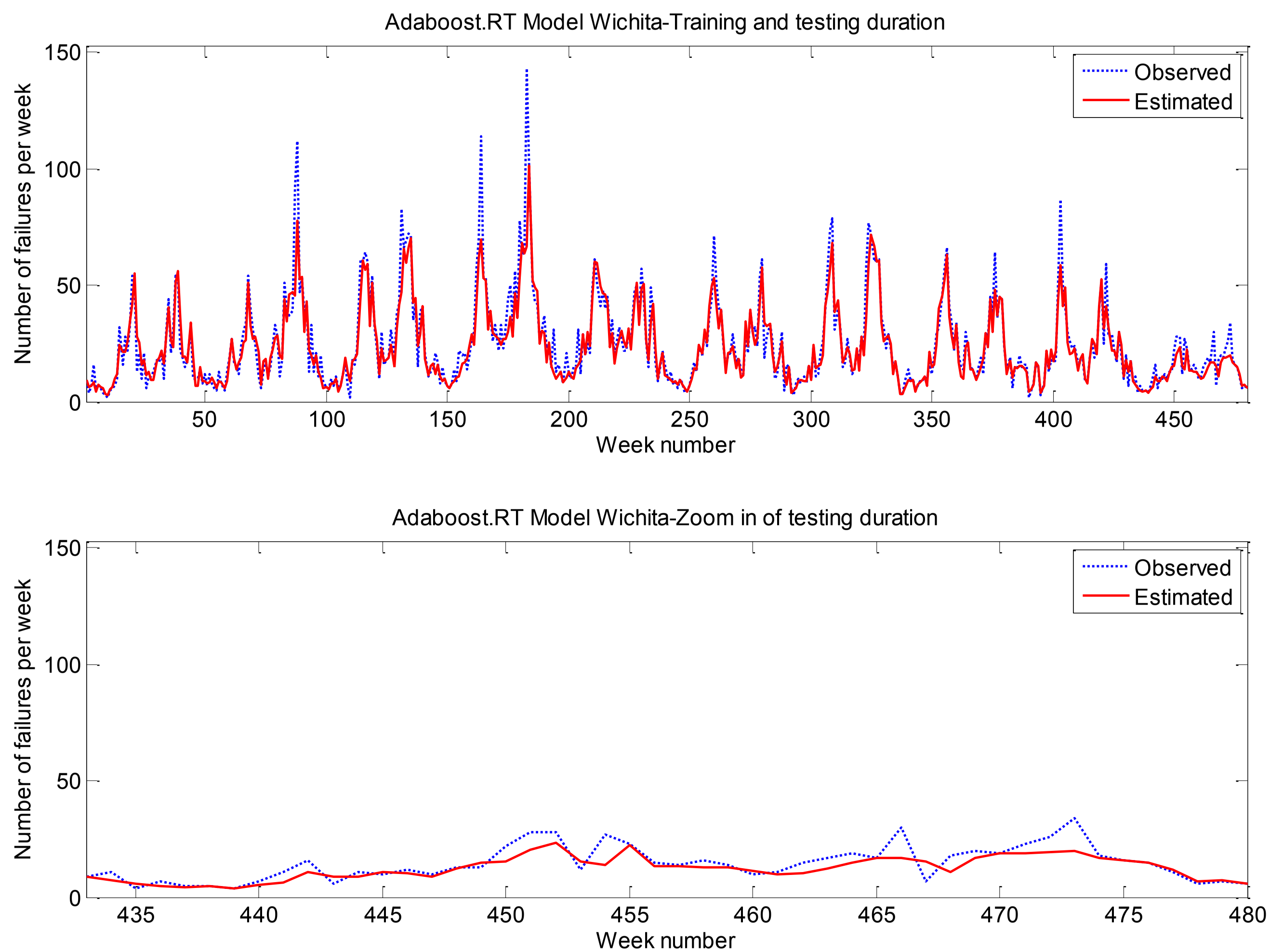


Figure 10.13 Animal-related Outages Observed and Estimated by AdaBoost.RT Model in Wichita from 1998 to 2007

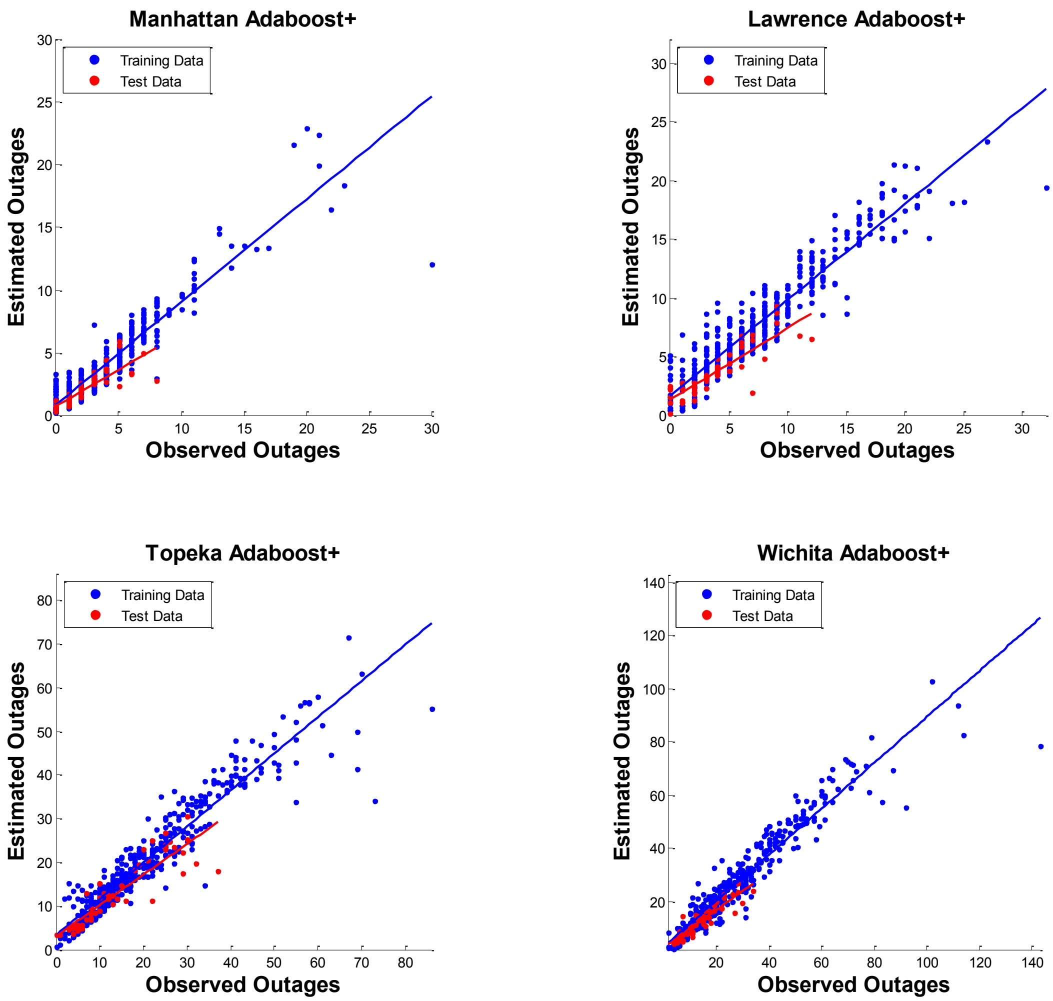


Figure 10.14 Plot of Observed and Estimated Animal-related Outages Obtained with Adaboost+ Model in Overhead Distribution Systems for Four Cities from 1998 to 2007

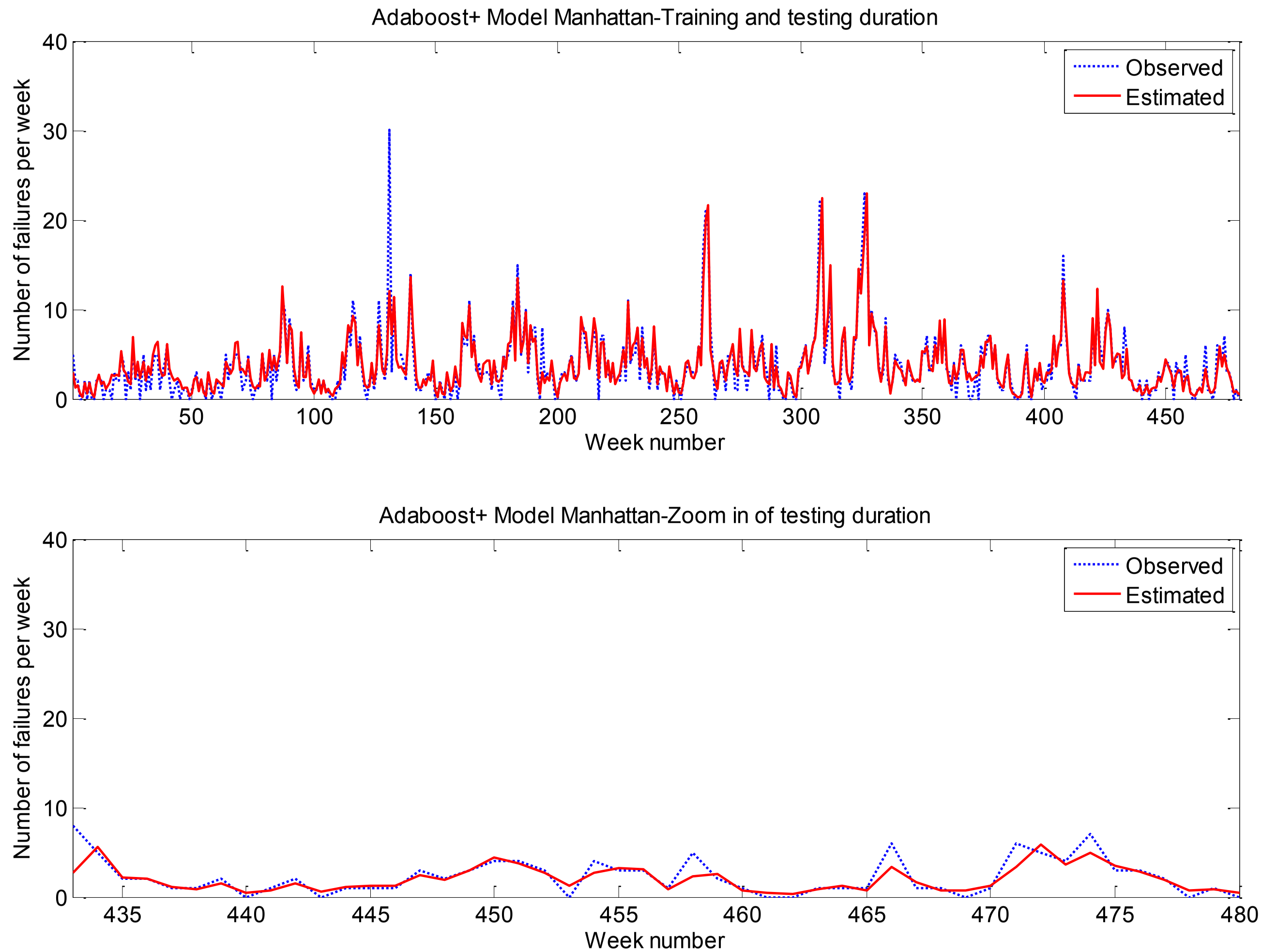


Figure 10.15 Animal-related Outages Observed and Estimated by AdaBoost$^+$ Model in Manhattan from 1998 to 2007

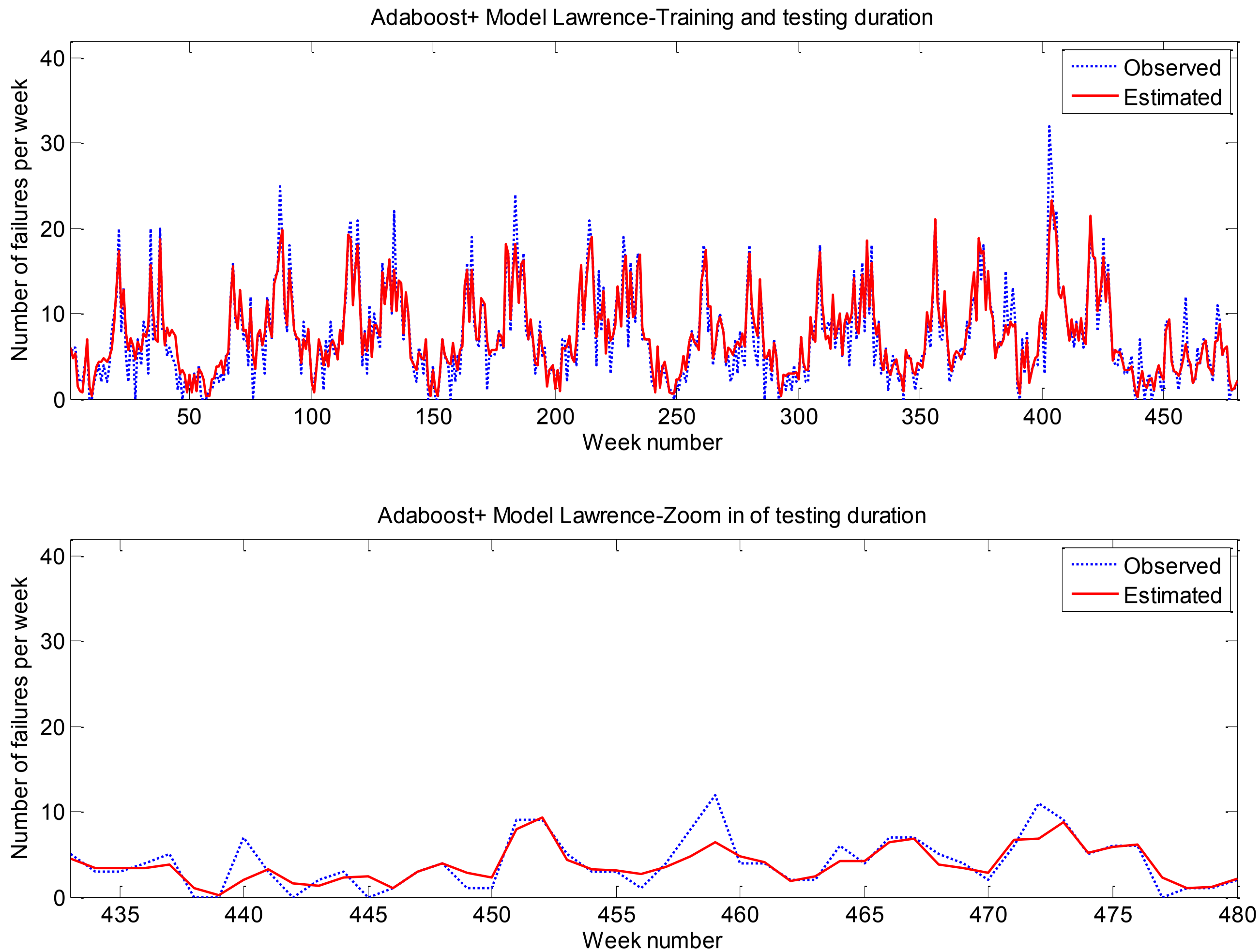


Figure 10.16 Animal-related Outages Observed and Estimated by AdaBoost^{+} Model in Lawrence from 1998 to 2007

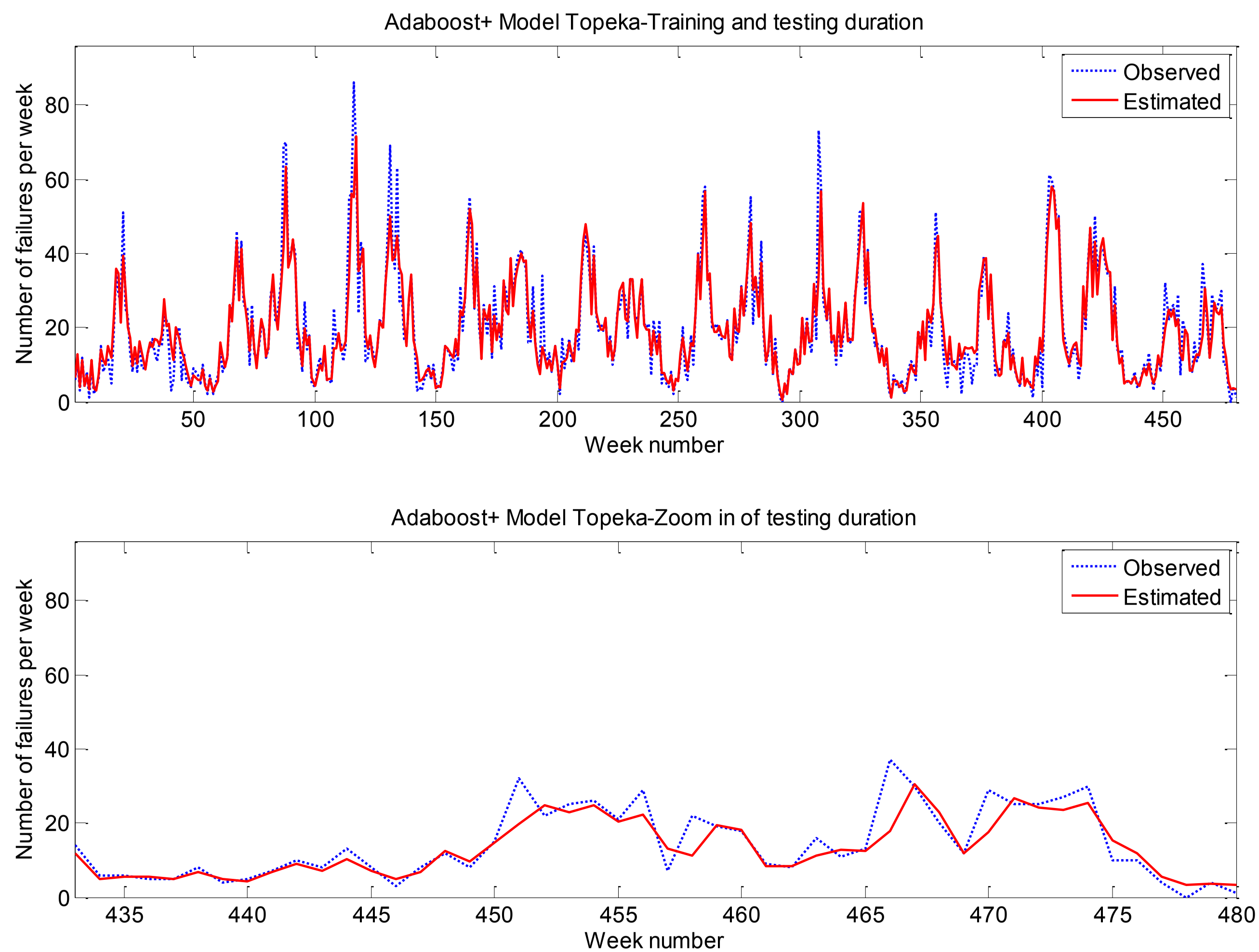


Figure 10.17 Animal-related Outages Observed and Estimated by AdaBoost$^+$ Model in Topeka from 1998 to 2007

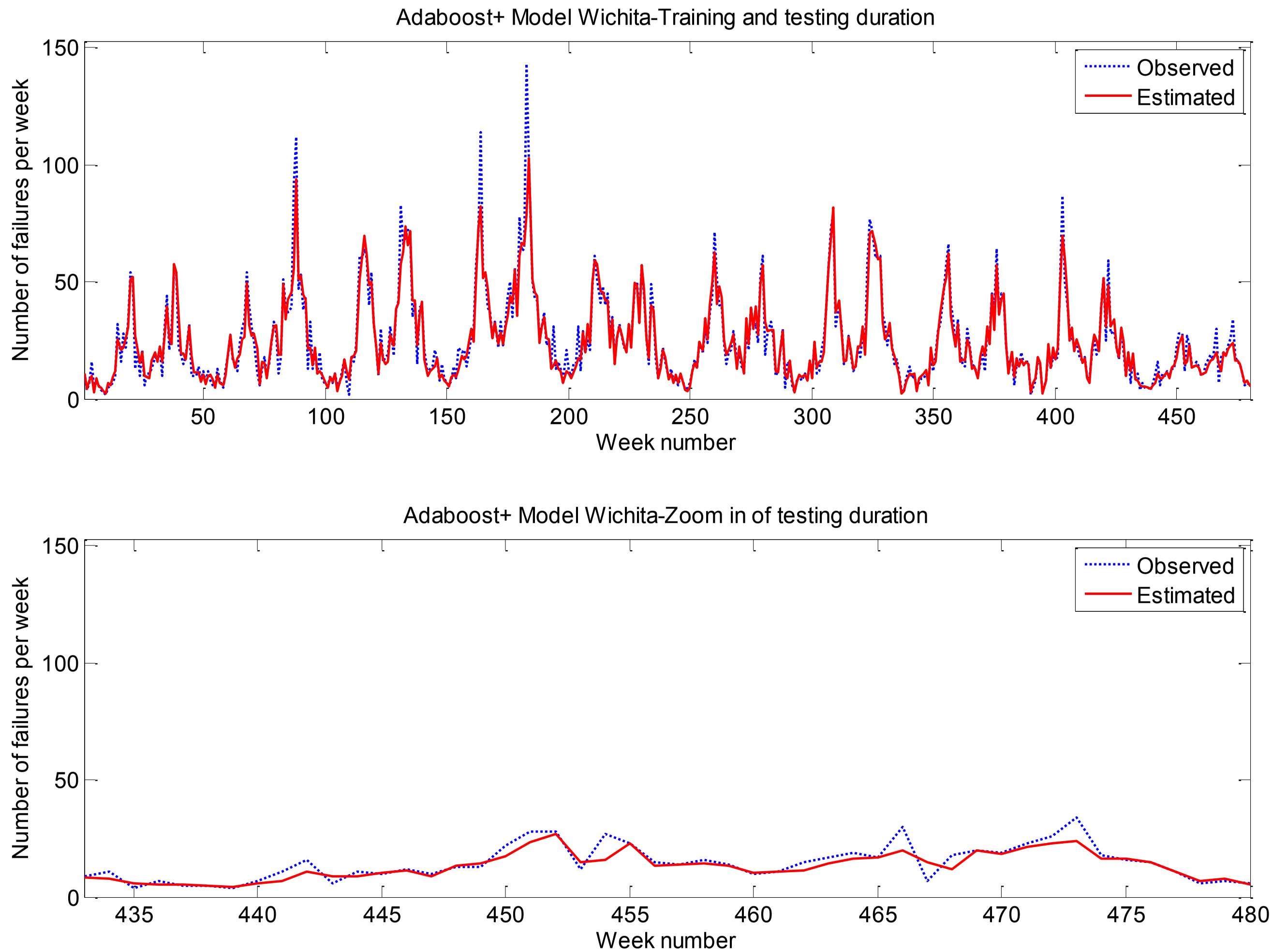


Figure 10.18 Animal-related Outages Observed and Estimated by AdaBoost^{+} Model in Wichita from 1998 to 2007

10.3.3 Annealed Mixture of Experts (AME) Model

Performances of the proposed models were measured by MAE, MSE, and correlation, R, which are tabulated in Table 10.5. The table shows that AME has better performance compared to other models with smaller MSE and MAE values and high correlation coefficient in training and test durations. For all the cities, the AME model has correlation coefficient near 0.9, indicating the positive relationship between observed and estimated outages. Figures 10.19 - 10.22 show clusters of data for observed and estimated outages for training and test data for the four cities. Table 10.6 summarizes the number of data points each cluster receives in the training and test data for the four cities. Figure 10.23 shows the scatter plot of observed and estimated outages with regression line for the four cities for AME model. Most of the points are around the unity slope line. Figures 10.24-10.25 show the plot of weekly observed outages and estimated outages for Manhattan, Lawrence, Topeka, and Wichita for the AME model. The model accurately estimates outages in the lower and higher range for all the cities.

Table 10.5 Results of AME Model for Four Cities

AME								
	Training Data				**Test Data**			
	MSE	**MAE**	**R**	**S**	**MSE**	**MAE**	**R**	**S**
Manhattan	1.5039	0.8246	0.9467	0.8576	0.9390	0.5576	0.8859	0.6624
Lawrence	3.1367	1.2829	0.9468	0.8423	2.2260	0.9125	0.8739	0.6530
Topeka	11.0242	2.2725	0.9736	0.9344	11.4240	2.1174	0.9425	0.8028
Wichita	19.7962	2.8531	0.9741	0.9320	7.5639	1.8747	0.9333	0.8686

Table 10.6 Number of Data Points in each Cluster for Four Cities

	Training Data					**Test Data**				
	Total	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Total**	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Manhattan	432	18	205	118	91	48	11	14	15	8
Lawrence	432	107	57	193	75	48	11	14	16	7
Topeka	432	124	70	175	63	48	8	19	11	10
Wichita	432	126	151	97	58	48	15	7	9	17

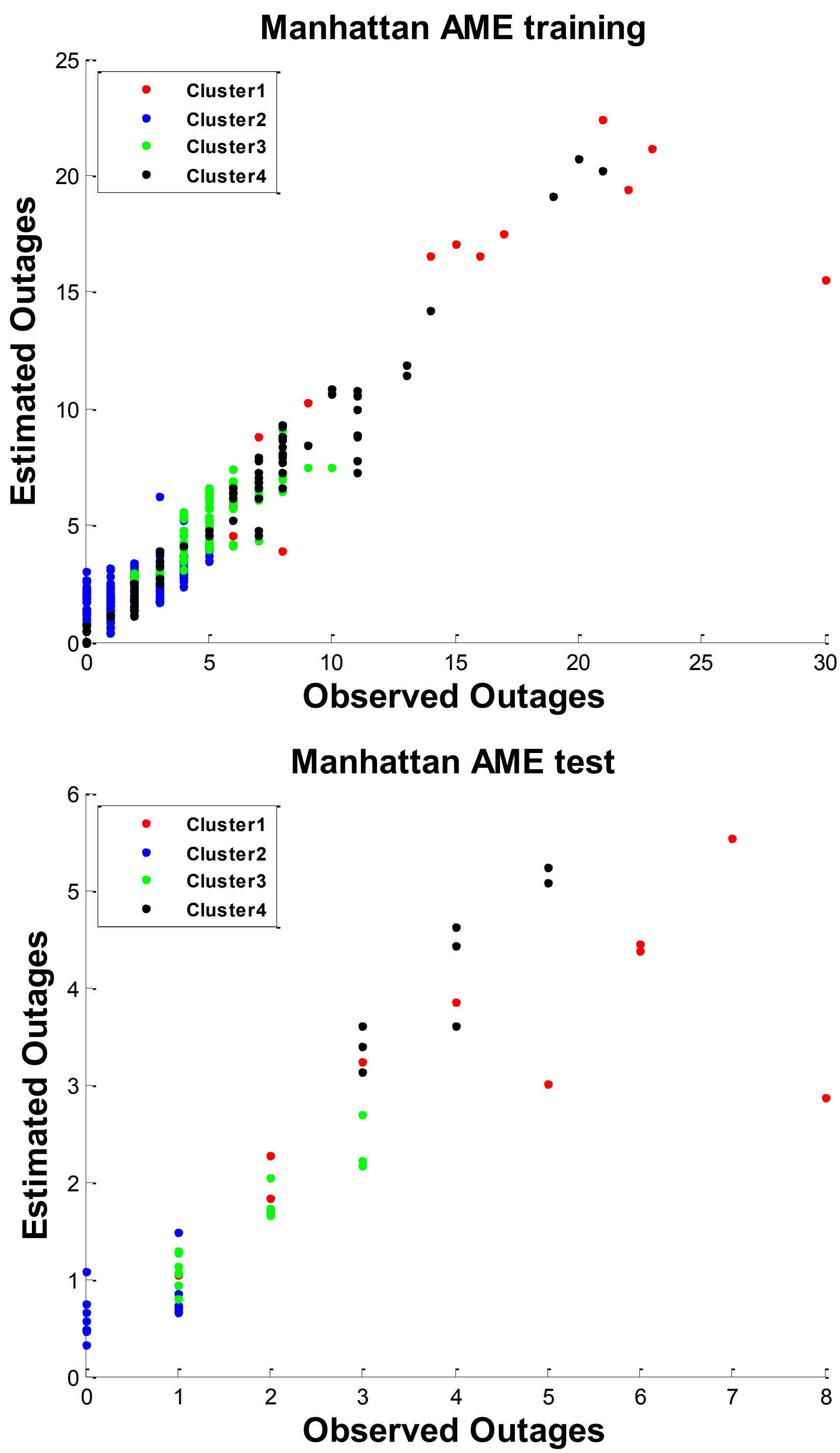
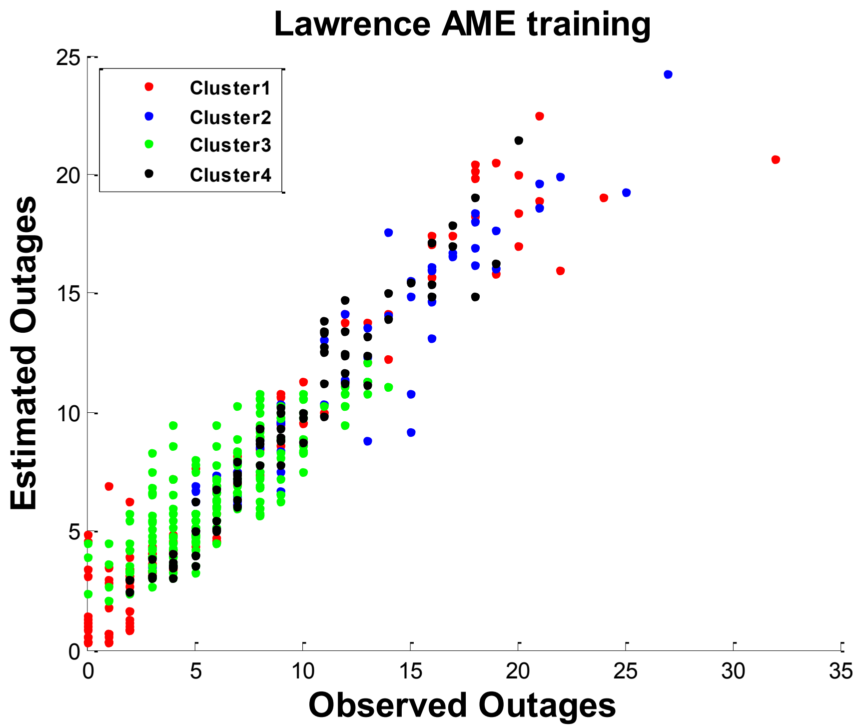


Figure 10.19 Clustering of Estimated Animal-related Outages Obtained with AME Model in Overhead Distribution Systems for Manhattan for 1998-2006 and 2007



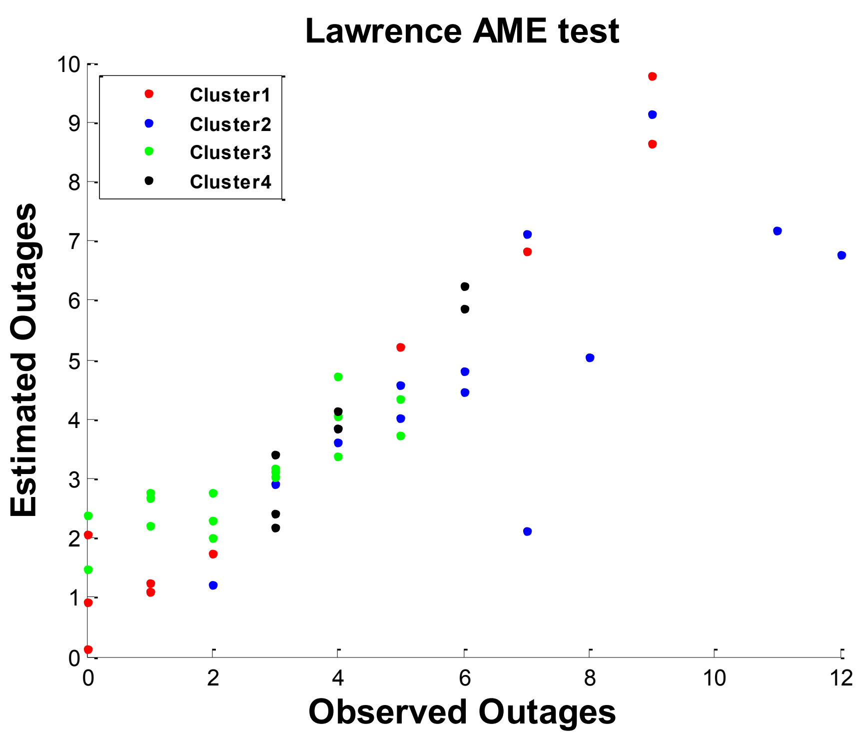


Figure 10.20 Clustering of Estimated Animal-related Outages Obtained with AME Model in Overhead Distribution Systems for Lawrence for 1998-2006 and 2007

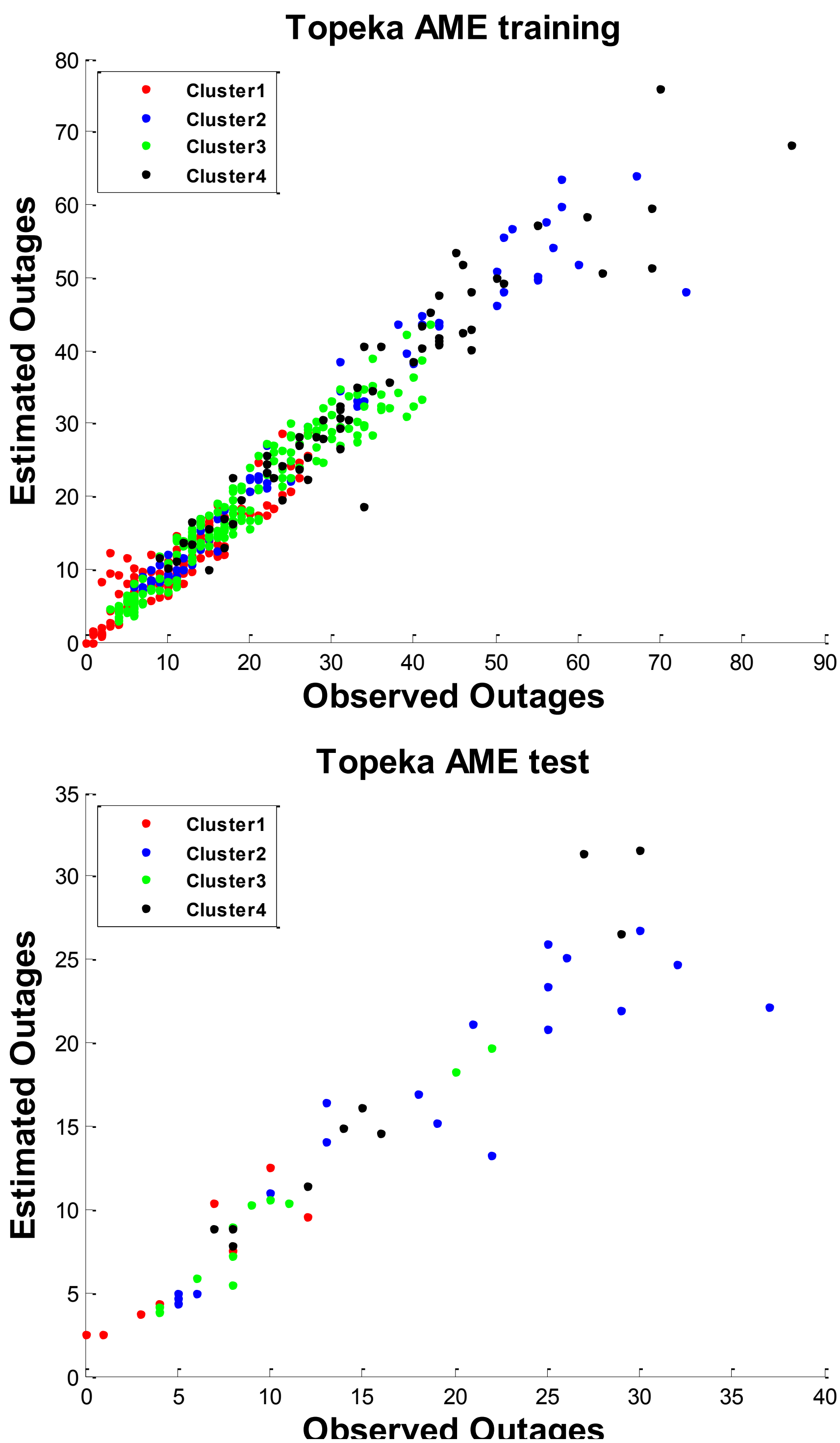


Figure 10.21 Clustering of Estimated Animal-related Outages Obtained with AME Model in Overhead Distribution Systems for Topeka for 1998-2006 and 2007

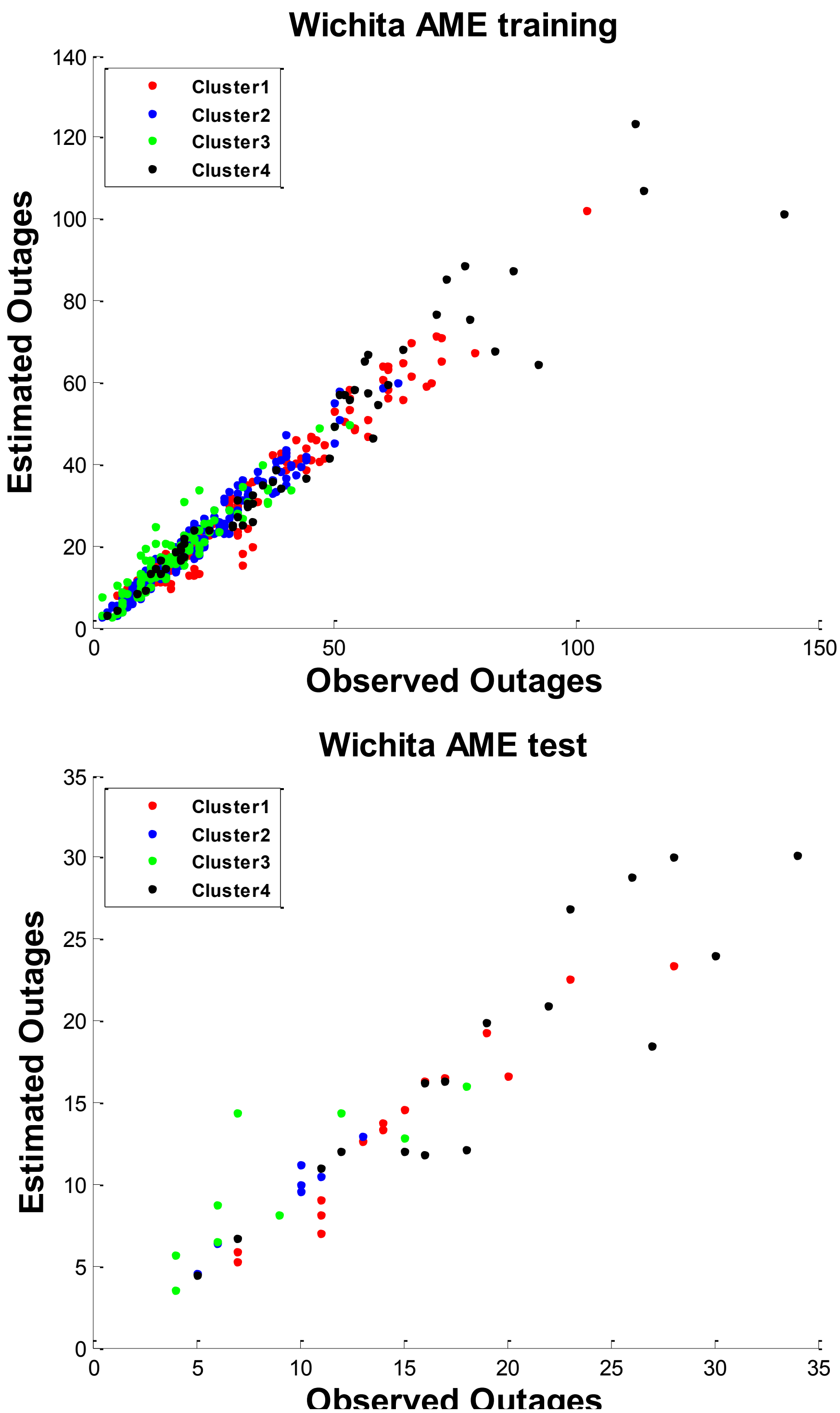


Figure 10.22 Clustering of Estimated Animal-related Outages Obtained with AME Model in Overhead Distribution Systems for Wichita for 1998-2006 and 2007

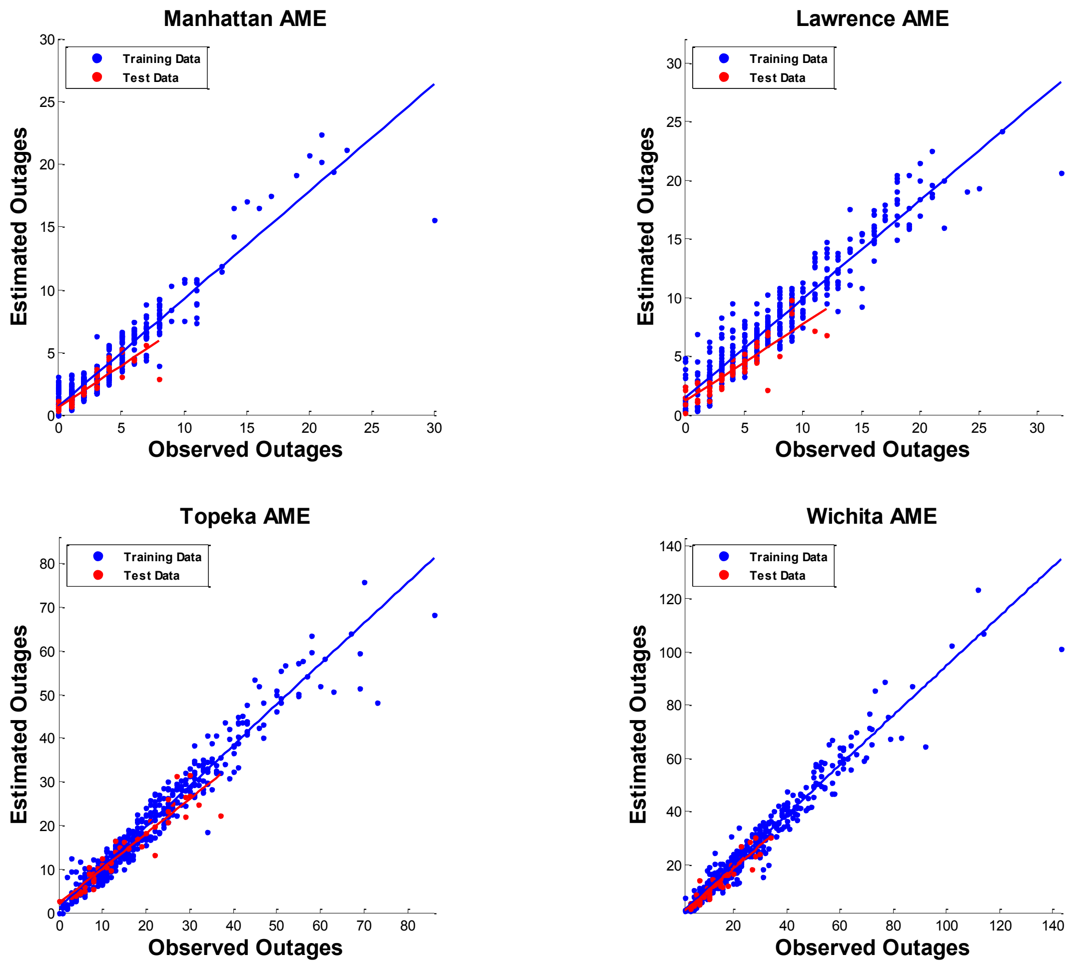


Figure 10.23 Plot of Observed and Estimated Animal-related Outages Obtained with AME Model in Overhead Distribution Systems for Four Cities from 1998 to 2007

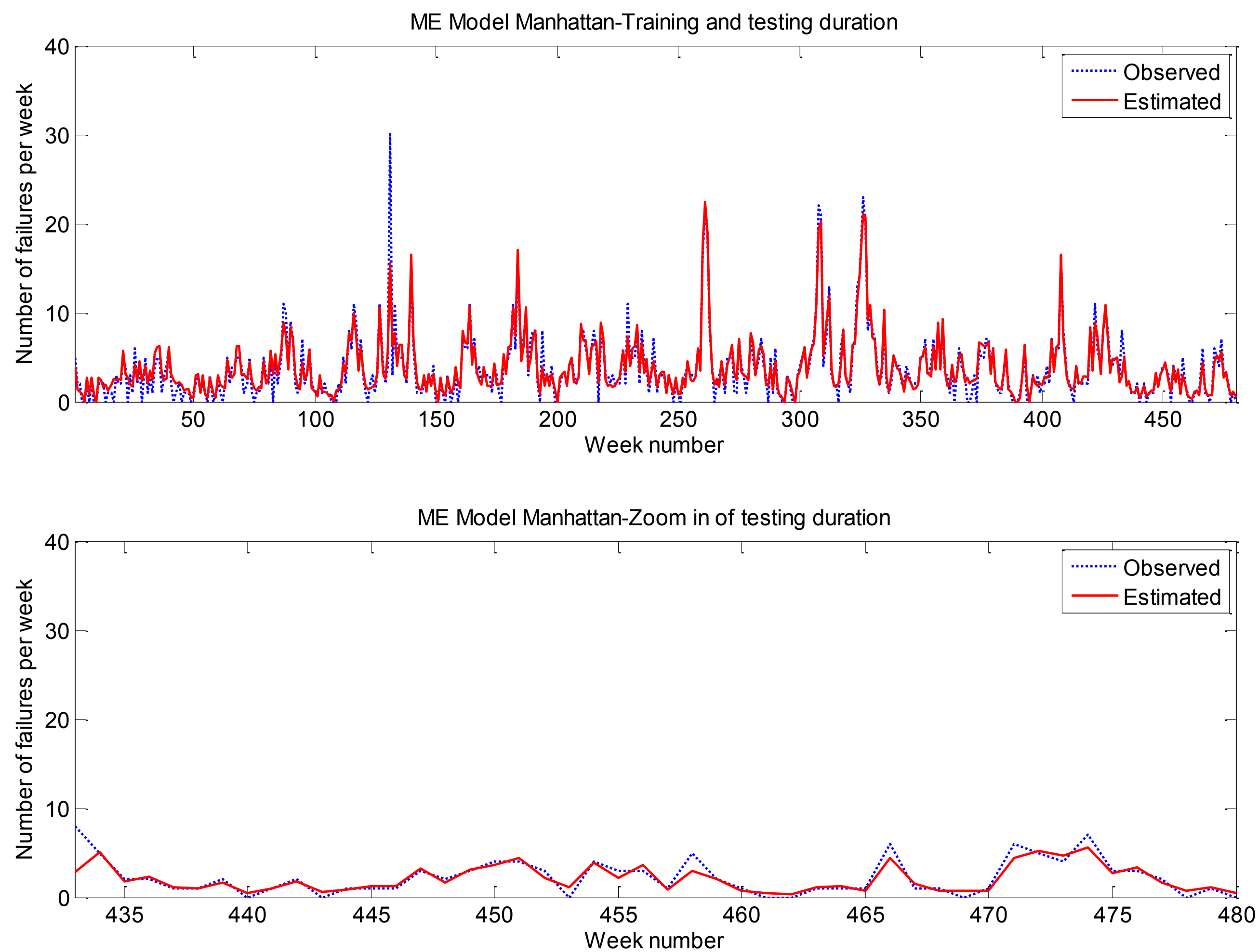


Figure 10.24 Animal-related Outages Observed and Estimated by AME Model in Manhattan from 1998 to 2007

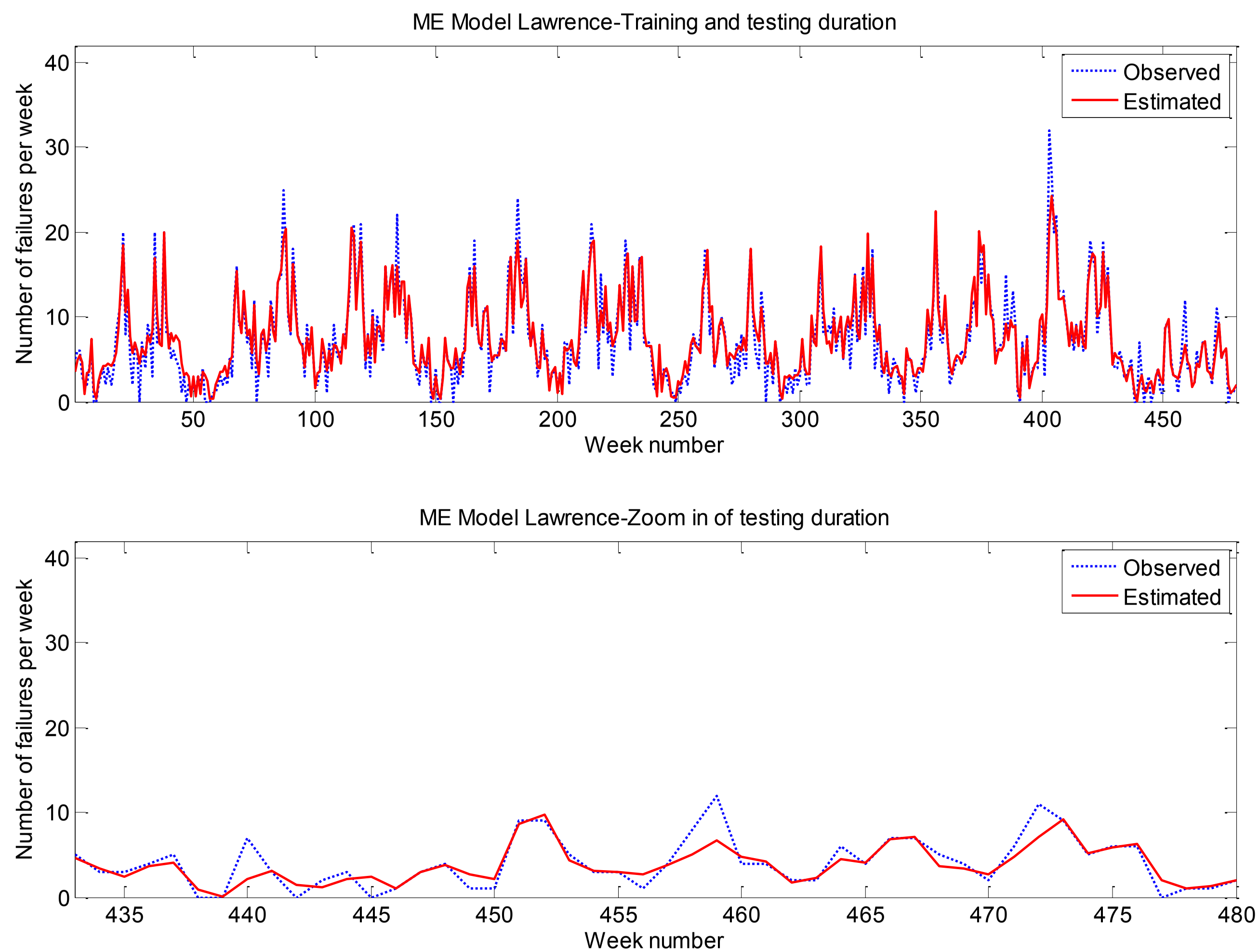


Figure 10.25 Animal-related Outages Observed and Estimated by AME Model in Lawrence from 1998 to 2007

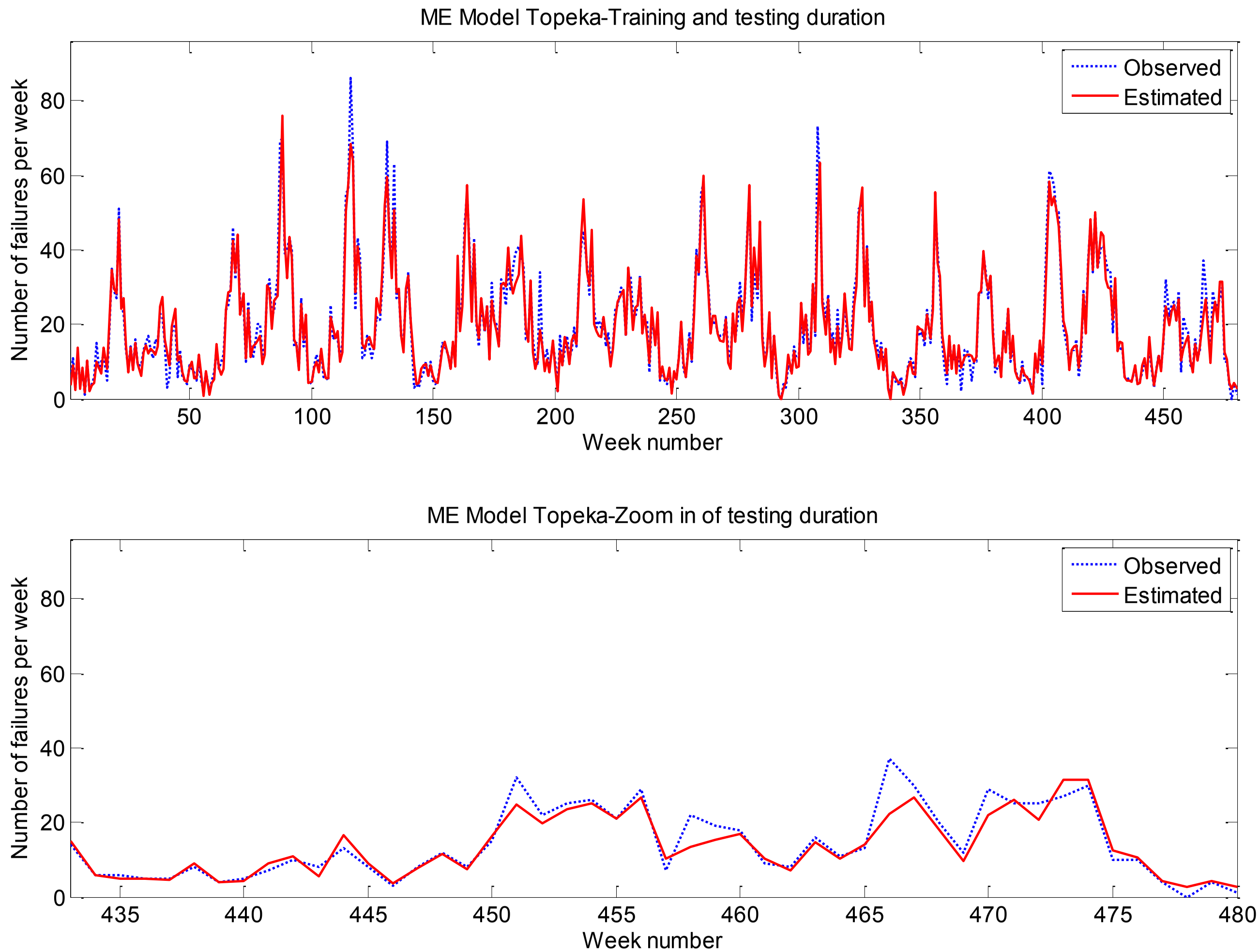


Figure 10.26 Animal-related Outages Observed and Estimated by AME Model in Topeka from 1998 to 2007

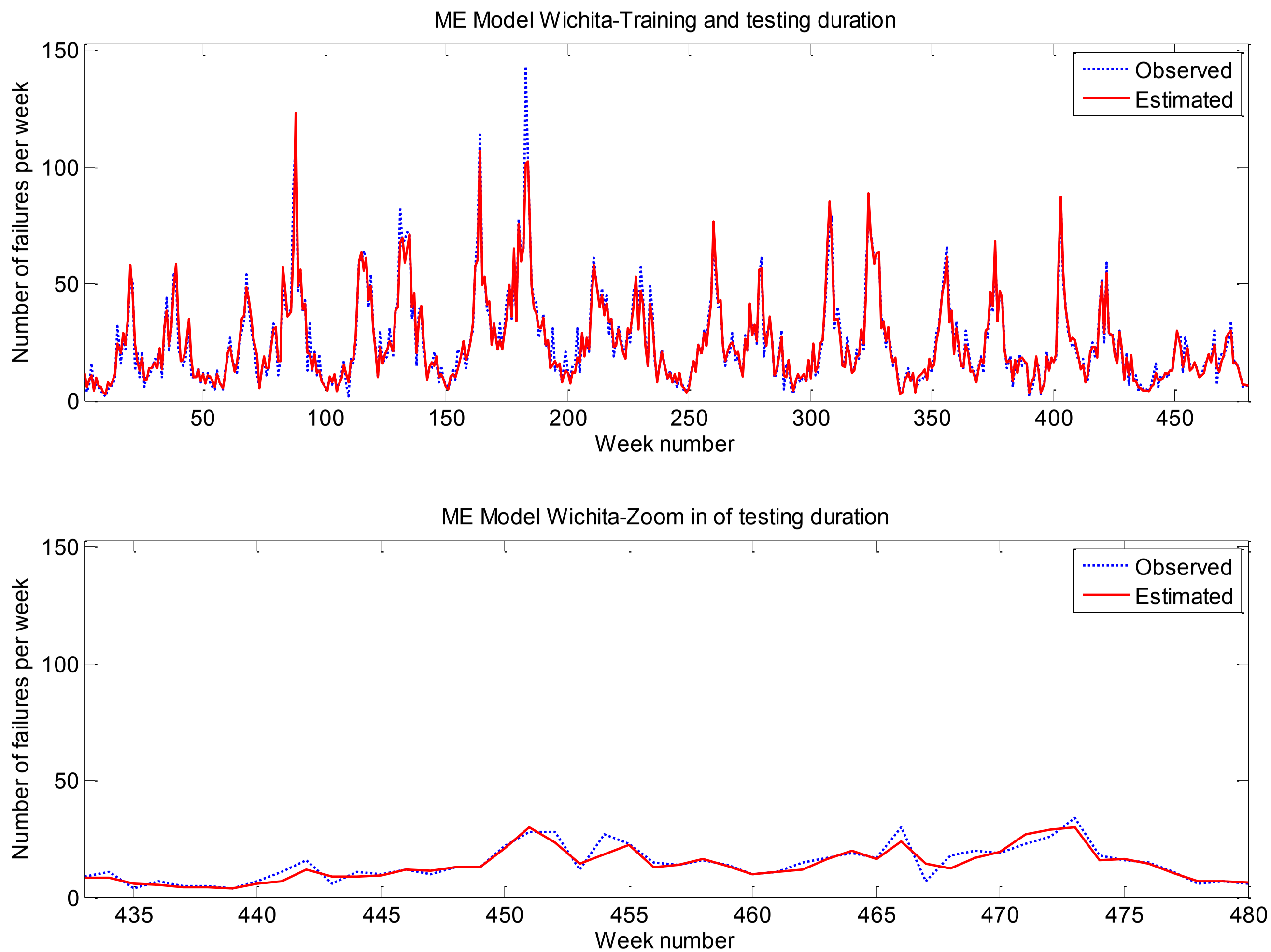


Figure 10.27 Animal-related Outages Observed and Estimated by AME Model in Wichita from 1998 to 2007

10.4 Comparison of Models

Gui, Pahwa and Das proposed the poison model, NN model, wavelet-based NN (WNN) model to find the correlation between animal activity and animal-caused outages [23]. The NN model with 6x4x1 structure with month type, number of fair days and previous four week outages as inputs was constructed. Wavelet-based NN model with three-stage DWT decomposition (WNN), with one stage wavelet decomposition (MWNN) was constructed. The hybrid model presented in [23] represents MWNN, one stage wavelet decomposition and artificial immune system (AIS). Historical data from 1998 to 2007 was considered and carried out the simulations. As mentioned in [23] the hybrid model outperformed other models followed by MWNN model for Manhattan, Lawrence, Topeka, and Wichita for the test data. An NN model was developed and because of randomness in the data, it could not catch the high peak in the time-series of weekly animal caused outages. A wavelet based NN models which decompose the data into approximate and detail subseries, used NN models for the subseries estimation and summed up the outputs from the NN models to get the final estimation for the original data. Three stage and one state wavelet decomposition was considered in [23]. To overcome the overtraining problem in the application of NNs, the Artificial Immune System was applied in one stage DWT decomposition for hyper mutation and retraining of the networks during the testing stage. From the simulations results, the hybrid model has better performance followed by one stage wavelet decomposition NN model [23].

In this dissertation, to make model computationally simple, a 3x2x1 structure NN model with month type, number of fair days and previous week outage as inputs was constructed. An ensemble models, AdaBoost.RT, AdaBoost$^+$ and AME, with 3x2x1 NN model as base learner was constructed. The historical data as considered in [23] was used to evaluate the performance of the models. The construction of the models and simulation results are discussed in section 8.3.

Table 10.7 The MAE of Models for Four Cities

Models	Wichita	Topeka	Lawrence	Manhattan
Hybrid [23]	4.67	3.70	1.50	1.22
MWNN [23]	4.71	3.77	1.53	1.25
WNN [23]	6.75	5.62	2.39	1.51
AdaBoost$^+$	2.7773	2.7413	1.1526	0.7579
AME	2.3639	2.1950	1.0977	0.6911

Table 10.7 shows the mean absolute error (MAE) with different models for the four cities. Note that the MAE for ensemble models are overall MAE for training data and testing data combined, which are different from the ones in Table 10.4 and 10.5. Comparing the MAE values, it can be observed that ensemble models have outperformed the hybrid model. The MAE drop for hybrid model and AME is very significant: for Wichita it dropped from 4.67 to 2.36, for Topeka it dropped from 3.7 to 2.1, for Lawrence it dropped from 1.5 to 1.09 and for Manhattan it dropped from 1.22 to 0.69.

10.5 Summary

In this chapter, ensemble learning methods were applied for estimation of animal-related outages in the distribution system. The construction of simple NN model and ensemble models were presented. Comparison of models developed by Gui and ensemble models were also discussed. The simulation results shows that ensemble methods gave significantly higher performance compared to wavelet-based NN models. For ensemble models for all the cities, the correlation coefficients are close to 1 too which indicates the high degree of relation between estimated and observed values. Overall, the ensemble model outperformed wavelet based NN model and gave significantly good performance.

Chapter 11 - Conclusion and Future Work

In this dissertation, different feature-based methods are proposed for estimation of weather and animal-related outages in the overhead distribution system. The proposed methods are applied to available historical weather and outage data to demonstrate their effectiveness. In this chapter, the primary results and conclusion are summarized, followed by suggestions for possible research directions for future work.

11.1 Conclusion

Considering discussions of previous research on the impact of various environmental factors on the distribution system, assessment of distribution reliability, and outage estimations, this dissertation aims to extend and implement algorithms in order to more accurately estimate outages caused by weather and animal activity. The utilities can use the approaches presented in this dissertation to find distribution reliability at the end of the year. By comparing the reliability of a specific year with past, the utilities can identify critical areas and plan remedial action.

The weather and outage data from 2005 to 2011 for Manhattan, Lawrence, Topeka and Wichita in Kansas are considered in this dissertation to evaluate the performance of the proposed models. Separate analysis for outages caused by wind, lightning and animals are considered. In the initial analysis, lightning strokes recorded within 200m and 400m around the feeder were used to know whether the distance of the recorded lightning will have any impact on system performances. Experimental results showed that the considering the lightning data within a distance of 400m around the feeder slightly improved the performance. In this dissertation, lightning stroke recorded within 500m around the feeder, as per the standard industry practice, were used to study the impact of lightning occurrence on the distribution system outages. For this study, the outages caused by wind, lightning, trees, equipment and unknown factors, outages possibly caused by lightning and wind are included in the outage count. All the proposed models in this dissertation were successfully implemented in MATLAB and model performances were evaluated using available historical data by computing MAE, MSE and R.

Linear, quadratic and exponential regression models were used to compare the performance. Simulation were carried out with lightning and wind as independent variables and results show that among six regression models developed, Model 4 (eqn 3.19), which considers quadratic relationship both for wind and lightning and combined effect of lightning and wind, have better performance than other models. However, due to random nature of weather effects and presence of noise in the data, regression models fail to estimate outages in the lower and upper range of observed outages.

The NN model, which has the ability to approximate nonlinear functions, is used to find the complex relationship between the weather and weather caused outages. With lightning, wind and observed outages as input to the model, the network structure and parameters were optimized by experiments. The NN model outperforms the regression models, but fails to estimate outages accurately. The performance of the NN model can be improved with the use of machine learning algorithms, specifically ensemble learning methods. In ensemble learning methods, multiple networks are trained using identical data with assigned weights. Based on performance of network in the previous training, the weight associated with data are updated in the current training. The final output can be weighted average of individual network output or weighted sum of individual network with weights being input dependent. This way with use of many networks, individual networks are specialized to perform in specific region. The number of learners to be used in the ensemble is found by experiments.

The AdaBoost.RT, AdaBoost$^+$, and AME based ensemble learning methods were proposed to estimate outages caused by weather in overhead distribution feeders. In AdaBoost models, the learners are trained sequentially based on the performance of past learner. The AdaBoost.RT and AdaBoost$^+$ model differ in the way the individual network outputs are combined to obtain the final output. In AdaBoost.RT, the ensemble output is the weighted sum of all learners, with the learners receiving weights proportional to the logarithm of their inverse error rates. In AdaBoost$^+$, the ensemble output is the pseudo inverse of all learners. To stabilize the matrix inversion in pseudo inversion, regularization has been incorporated. In AME method, the learners are trained in parallel and algorithm parameters are learned using MFA. When the temperature reaches a critical temperature, the MSE of the learners stabilizes. The optimum number for the maximum iterations is obtained by experiments. The critical temperature for all four cities is found to lie between 1.5 and 2. The ensemble output is the output of the individual learner whichever gives the minimum error. Simulation results shows that ensemble of networks perform significantly better compared to the other models. Among proposed ensemble methods, it is found that AME has better performance followed by AdaBoost$^+$. Also, results show that the ensemble learning methods are resistant to over-fitting due to noisy data and that they perform significantly better than a single network. We can see a significant improvement from regression model to ensemble models in estimating outages. And also the positive correlation close to one indicates the high correlation between observed and estimated outages.

To further test the models, the data were divided into two sets, the days with lightning and the days with no lightning. The proposed models are trained separately with two data sets and the MSE and MAE values were computed. The models were trained with all days included and separately for days with lightning and with no lightning. Also, the estimated outages from the model trained separately were combined to find the MAE and MSE between the estimated outages and observed outages. These values were compared with the MAE and MSE values obtained from models trained with all days included. It was found that the prior approach has better performance values indicating better performance of the models trained separately.

Since the results of ensemble learning models for weather-related outages were very promising, the models were extended to study their application for estimation of outages caused by animals. The historical data from 1998 to 2005 were used to evaluate the performance of the models. The model performance were compared with models presented in [23] and the simulation results shows that the ensemble models outperformed the hybrid and wavelet-based neural network model.

11.2 Future Work

Further research to improve AdaBoost models should be focused on automating the choice of optimal value of threshold depending on characteristics of the data set. Other learning models as a base learner in the ensemble learning method can be investigated in the future.

Other variables to represent wind, such as wind gust duration and gust speed, can be investigated. Furthermore, in this dissertation, cutoff for the wind speed is not considered. When the wind gust is less than 15-25 mph and with no lightning occurrence, the probability of outage happening is zero. This calls for investigation of the models with a cutoff value for wind speed.

Tree density information along the overhead feeders will greatly improve the estimations of animal-related outages since tree attracts animals and animals can cause outages indirectly through trees. And the proposed models can be applied for available 2005 -2011 data.

Finally, the proposed models are suitable for end-of-the-year reliability evaluation based on past data. Based on the weather scenarios, future outage prediction can be researched. Further, the proposed models provide estimated outages without a confidence bound associated with them. These models can be extended to include determination of confidence bound both for outages caused by weather and animals. Such statistical analysis can be used to benchmark the performance of the system. If the number of outages observed for a year fall within the confidence bounds, the utilities can justify the results to the regulatory commissions. However, if the observed outages are outside the bounds, the system performed either better or worse than expected. Specifically, if higher than upper bound, the utility will have to do further investigation to determine specific reasons for higher outages and fix them to prevent large number of outages in the future.

References

[1] L. Silverstein, Transmission 101, National Association of Regulatory Utility Commissioners, April 20, 2011, pp. 30. http://www.naruc.org/grants/Documents/Silverstein%20NCEP%20T-101%200420111.pdf.

[2] C. A. Warren, "Distribution Reliability: What is it?," *IEEE Industry Application Magazine*, Jul./Aug. 1996, pp. 32 – 37.

[3] R. Billington and J. E. Billington, "Distribution System Reliability Indices," *IEEE Transactions on Power Delivery*, Jan. 1989, pp. 561 – 568.

[4] A. M. Rei and M. T. Schilling, "Reliability Assessment of the Brazilian Power System Using Enumeration and Monte Carlo," *IEEE Transactions on Power Systems,* Vol. 23, Issue 3, Aug. 2008 , p1480 – 1487

[5] H. L. Willis, *Power Distribution Planning Reference Book*, Marcel Dekker, Inc., New York, 1997.

[6] C. A. Warren and M.J. Adams, "Reliability of the Regulatory Horizon," *IEEE Transmission and Distribution Conference and Exposition*, Atlanta, GA, Oct. 2001.

[7] M. Amin, "Modeling and Control of Complex Interactive Networks [Guest Editorial]", *IEEE Control Systems Magazine*, Feb. 2002, pp. 22-37.

[8] Edison Electric Institute Transmission and Distribution Committee Report, *Guide for Reliability Measurement and Data Collection*, Oct. 1971.

[9] U. S. Department of Energy Report, *The National Electric Reliability Study: Executive Summary*, DOE/EP-0003 Dist. Category UC-97C, Apr. 1981.

[10] Electric Power Research Institute Report, *Development of Distribution System Reliability and Risk Analysis Models*, EL-2018, Aug. 1981.

[11] Canadian Electric Association Report, *Distribution System Reliability Statistics*, 148 D255, Oct. 1984.

[12] T. Gonen, *Electric Power Distribution System Engineering*, McGraw-Hill, Inc., New York, 1986.

[13] R. Billington and W. Li, *Reliability Evaluation of Power Syst*ems, 2nd Edition, Plenum Press, New York, 1995.

[14] IEEE Std 1366, *IEEE Guide for Electric Power Distribution Reliability Indices*, IEEE, Piscataway, NJ, 2001.

[15] IEEE Std. 1366-2003, *IEEE Guide for Electric Power Distribution Reliability Indices*, IEEE, New York, NY, 2003.

[16] R. E. Brown and J. R. Ochoa, “Distribution System Reliability: Default Data and Model Validation,” *IEEE Transactions on Power Systems*, Vol. 13, No. 2, May 1998, pp. 704 – 709.

[17] R. E. Brown, A. P. Hanson, H. L. Willis, F. A. Luedtke and M. F. Born, “Assessing the Reliability of Distribution Systems,” *IEEE Computer Applications in Power*, Jan. 2001, pp. 44 – 49.

[18] R. E. Brown, *Electric Power Distribution Reliability*, Marcel Dekker, Inc., New York, 2002.

[19] G. Kjolle, L. Rolfseng and E. Dahl, “The Economic Aspect of Reliability in Distribution System Planning,” *IEEE Transactions on Power Delivery,* Apr. 1990, pp. 1153 – 1157.

[20] L. Goel and R. Billinton, “Determination of Reliability Worth for Distribution System Planning,” *IEEE Transactions on Power Delivery*, July 1994, pp. 1577 – 1583.

[21] M. Gui, A. Pahwa, and S. Das, “Analysis of animal-related outages in overhead distribution systems with wavelet decomposition and immune systems based neural networks,” *IEEE Transactions on Power Systems*, vol. 24, pp. 1765–1771, 2009.

[22] M. Gui, A. Pahwa, and S. Das, “Classification of input and output variables for a Bayesian model to analyze animal-related outages in overhead distribution systems,” *Proceedings International Conference Probabilistic Methods Applied to Power Systems*, Singapore, Jun. 2010, pp. 469–474.

[23] M. Gui, A. Pahwa, S. Das, “Bayesian Network Model With Monte Carlo Simulations for Analysis of Animal Overhead Distribution Systems”, *IEEE Transactions on Power Systems*, Vol. 26, No. 3, Aug 2011.

[24] W. F. Horton, S. Goldberg and C. A. Volkmann, “The Failure Rates of Overhead Distribution System Components,” *Proceedings of the 1991 IEEE Power Engineering Society Transmission and Distribution Conference*, 1991, pp. 713 –717.

[25] R. H. Stillman, “Modeling Failure Data of Overhead Distribution Systems,” *IEEE Transactions on Power Delivery,* Oct. 2000, pp. 1238 – 1242.

[26] T. E. Nigon, “A Method of Reporting Distribution Field Inspection Data,” *41st Annual Rural Electric Power Conference*, Minneapolis, USA, 1997, pp. 1 – 6.

[27] J. G. Tine and D. A. Walder, “Improved Distribution System Reliability at Northern States Utilities: a Case History,” *IEEE Transactions on Power Delivery,* Apr. 1993, pp. 689 – 696.

[28] 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.

[29] 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.

[30] W. H. Kersting, W. H. Phillips and R. C. Doyle, “ Distribution Feeder Reliability Studies,” *IEEE Transactions on Industry Applications*, Mar./Apr. 1999, pp. 319 –323.

[31] J. McDaniel, C. Williams, and A. Vestal, “Lightning and Distribution Reliability – A Comparison of Three Utilities,” *Proceedings of IEEE PES T&D Conference and Expo*, Dallas, TX, Sep. 2003.

[32] C. A. Warren, “Collecting and Categorizing Information Related to Electric Power Distribution Interruption Events: Data Consistency and Categorization for Benchmarking Surveys,” *IEEE Transactions on Power Delivery*, Vol. 21, No. 1, Jan. 2006.

[33] M. Darveniza, C. Dunn, and B. Holcombe, "A Comparison of Faults and Outages on an Electrical Distribution System Caused by Lightning and Wind Gusts," *29th International Conference on Lightning Protection*, Uppsala, Sweden, 2008.

[34] IEEE Std 1410-2004 (Revision of IEEE Std 1410-1997), *IEEE Guide for Improving the Lightning Performance of Electric Power Overhead Distribution Lines*, IEEE, New York, NY, 2004.

[35] P. P. Barker and R. T. Mancao, “Lightning Research Advances with Digital Surge Recordings,” *IEEE Computer Applications in Power*, April 1992, pp. 11-16.

[36] 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.

[37] N. Balijepalli, Advances in Distribution System Reliability Assessment, Ph.D. dissertation, Iowa State University, Ames, IA, 2002.

[38] Y. Zhou, A. Pahwa, and S. Yang, "Modeling Weather-Related Failures of Overhead Distribution Lines," *IEEE Transactions on Power Systems,* vol. 21, 2006, pp. 1683-1690.

[39] S. Sahai and A. Pahwa, “A probabilistic approach for animal-caused outages in overhead distribution systems,” *Proceedings International Conference Probabilistic Methods Applied to Power Systems*, Stockholm, Sweden, Jun. 2006, pp. 1–7.

[40] N. Balijepalli, S. S. Venkata, and R.D. Christie, “Predicting Distribution System Performance Against Regulatory Reliability Standards,” *IEEE Transactions on Power Delivery*, Jan. 2004, pp. 350 – 356.

[41] 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.

[42] 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*, vol. 11, no. 4, Dec, 2005, pp.258-267.

[43] 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.

[44] 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.

[45] A. Domijan Jr, R. K. Matavalam, A. Montenegro, W. S. Wilcox, Y. S. Joo, L. Delforn, 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] P. Kankanala, A. Pahwa, and S. Das, “ADABOOST+: An Ensemble Learning Approach for Estimating Weather-related Outages in Distribution Systems,” IEEE Transaction on Power Systems, vol. 1, issue. 99, Sep 2013, pp. 1-9.

[52] 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.

[53] P. McCullagh, and J. A. Nelder, *Generalized linear models*, *2 Ed*, Chapman and Hall, New York, 1989.

[54] R. H. Myers, D. C. Montgomery and G. G. Vining, *Generalized Linear Models with Applications in Engineering and the Science*, John Wiley & Sons, Inc., New York, 2002.

[55] J. Neter, M. H. Kutner, C. J. Nachtsheim and W. Wasserman, *Applied Linear Statistical Models*, McGraw-Hill, New York, 1996.

[56] E. H. Moore, "On the reciprocal of the general algebraic matrix", bulletin of the American Mathematical Society, 26 (9): pp: 394–395, 1920.

[57] A. Bjerhammar, "Application of calculus of matrices to method of least squares; with special references to geodetic calculations", *Transaction of the Royal Institute of Technology*, Stockholm, vol. 49, 1951.

[58] R. Penrose, "A generalized inverse for matrices," *Proceedings of the Cambridge Philosophical Society*, vol. 51, pp. 406–413, 1955.

[59] C. M. Bishop, "*Neural networks for pattern recognition*," Oxford Univ. Press, Oxford, U.K, 1995.

[60] S. Haykin, "*Neural Networks: A Comprehensive Foundation*", 2nd edition, Prentice-Hall, 1999.

[61] X. Yao and Y. Liu, "Making use of population information in evolutionary artificial neural networks," *IEEE Transaction on Systems, Man, Cybernetics, Part B*, vol. 28, pp. 417–425, June 1998.

[62] P. Sollich, and A. Krogh, "Learning with ensembles: How over-fitting can be helpful," in *Advances in Neural Information Processing Systems,* MIT Press, pp.190- 196, 1996.

[63] E. J. Hartman, J. D. Keeler, and J. M. Kawalski, "Layered neural networks with gaussian hidden units as universal approximators," *Neural Computation*, vol. 35, no. 2, pp. 210–215, 1990.

[64] E. J. Hartman, J. D. Keeler, and J. M. Kawalski, "Layered neural networks with gaussian hidden units as universal approximators," *Neural Computation*, vol. 35, no. 2, pp. 210–215, 1990.

[65] K. Hornik, M. Stinchcombe, and H. White, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks," *Neural Computation*, pp. 551–560, 1990.

[66] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Englewood Cliffs, NJ: Prentice-Hall, 1985.

[67] I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley Press, 2004.

[68] B. Krose and P. V. D. Smagt, *A Introduction To Neural Networks*, 8th Edition, Nov. 1996.

[69] T. Poggio and F. Girosi, "Networks for approximation and learning", *Proceedings of IEEE*, Vol. 78, 1990, pp. 1481 – 1497.

[70] B. T. Zhang and J. G. Joung, "Time series prediction using committee machines of evolutionary neural trees," in *Proceedings of the Congress on Evolutionary Computation (CEC'99)*, vol. 1, July 1999, pp. 281–286.

[71] D. Opitz, R. Maclin, "Popular ensemble methods: An empirical study," *Journal of Artificial Intelligence Research*, vol. 11, pp.169–198, 1999.

[72] E. Alpaydin. *Introduction to Machine Learning*, MIT, 2004.

[73] V.N. Vapnik, "An overview of statistical learning theory," *IEEE Transaction on Neural Networks*, vol. 10, issue. 5, pp. 988–999, 1999.

[74] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 2000.

[75] E. R. Schapire, "The strength of weak learnability," *Machine Learning,* 5, 197, 1990.

[76] A. Barrat, M. Barthelemy, and A. Vespignani, Dynamical Processes on Complex Networks: Cambridge University.

[77] M. P. Perrone and L. N. Cooper, "When networks disagree: Ensemble methods for hybrid neural networks," Artificial Neural Networks for Speech and Vision, Chapman & Hall, pp. 126- 142, 1994.

[78] Z. Zhou, J. Wu, W. Tang, "Ensembling neural networks: Many could be better than all," Artificial Intelligence, vol. 137, no. 1–2, pp. 239–263, May 2002.

[79] B. T. Zhang and J. G. Joung, "Time series prediction using committee machines of evolutionary neural trees," in *Proceedings of the Congress on Evolutionary Computation (CEC'99)*, vol. 1, July 1999, pp. 281–286.

[80] A.S. Weigend, M. Manageas, "Analysis and prediction of multi-stationary time series, in: Neural Networks in Financial Engineering," *Proceedings of the third International Conference on Neural Networks in the Capital Markets*, 1995, London, UK, pp. 597–611.

[81] A.S. Weigend, M. Manageas, A.N. Srivastava, "Nonlinear gated experts for time series: discovering regimes and avoiding over-3ting," *International Journal on Neural Systems*, vol. 6, issue 4, pp. 373–399, (1995).

[82] 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.

[83] 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.

[84] L. Cao, "Support vector machines experts for time series forecasting," *Neurocomputing*, vol. 51, pp. 321–339, Apr. 2003.

[85] E. Meeds and S. Osindero, "An alternative infinite mixture of Gaussian process experts," in *Advances in Neural Information Processing Systems*, Cambridge, MA: MIT Press, 2006, pp. 883–890.

[86] C. Yuan and C. Neubauer, "Variational mixture of Gaussian process experts," in *Advances in Neural Information Processing Systems 21*, Cambridge, MA: MIT Press, 2009, pp. 1897–1904.

[87] Y. Li, "Hidden Markov models with states depending on observations," *Pattern Recognition Letters*, vol. 26, no. 7, pp. 977–984, 2005.

[88] 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.

[89] 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.

[90] 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.

[91] H.-J. Xing and B.-G. Hu, "An adaptive fuzzy c-means clustering-based mixtures of experts model for unlabeled data classification," *Neurocomputing*, vol. 71, nos. 4–6, pp. 1008–1021, Jan. 2008.

[92] L. Breiman, "Bagging Predictors", *Machine Learning*, vol.24, 1996, pp.123–140.

[93] N. Duffy, D. Helmbold, "Boosting Methods for Regression," *Machine Learning*, vol. 47, 2002, pp. 153–200.

[94] H. Drucker, "Boosting Using Neural Networks", Combining Artificial Neural Nets, London: Springer-Verlag, 1999, pp. 51-77.

[95] 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.

[96] J. Friedman, T. Hastie, & R. Tibshirani, "Additive Logistic Regression: A Statictical View of Boosting", *Annals of Statistics*, vol. 28, no. 2, 2000, pp. 337-374.

[97] 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.

[98] 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.

[99] 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.

[100] 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.

[101] 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.

[102] 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.

[103] 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.

[104] Y. Freund, and R. Schapire, "Experiment with a new boosting algorithm", *Proceedings of the 13th International Conference on Machine Learning*, Bari, Italy, 1996, pp. 148-156.

[105] R. Schapire, "The strength of weak learnability", *Machine Learning*, vol. 5, no. 2, 1990, pp. 197-227.

[106] H. Miihlenbein, and D. Schlierkamp-Voosen, "The science of breeding and its application to the breeder genetic algorithm," *Evolutionary Computation*, vol. l, no. 4, pp. 335-360, 1994.

[107] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *Annals of Statistics,* vol. 28, no. 2, pp. 337-407, 2000.

[108] 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.

[109] P. Solomatine, and D. L. Shrestha, "AdaBoost.RT: a Boosting Algorithm for Regression Problems", *Proceedings of the International Joint Conference on Neural Networks*, IEEE, Budapest, Hungary, 2004, pp. 1163-1168.

[110] R.A. Jacobs, M.A. Jordan, S.J. Nowlan, "G.E. Hinton, Adaptive mixtures of local experts," *Journal on Neural Computation*, vol. 3, pp. 79–87, 1991.

[111] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, vol. 3, no. 1, pp. 79–87, 1991.

[112] R. Avnimelech and N. Intrator, "Boosted mixture of experts: An ensemble learning scheme," *Neural Computation*, vol. 11, no. 2, pp. 483–497, 1999.

[113] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Network*s, Cambridge University.

[114] C. A. M. Lima, A. L. V. Coelho, and F. J. Von Zuben, "Hybridizing mixtures of experts with support vector machines: Investigation into nonlinear dynamic systems identification," *Information Sciences*, vol. 177, no. 10, pp. 2049–2074, 2007.

[115] V. Tresp, "Mixtures of Gaussian processes," *Advances in Neural Information Processing Systems 13*, Cambridge, MA: MIT Press, 2001, pp. 654–660.

[116] M.I. Jordan, R.A. Jacobs, "Hierarchical mixtures of experts and the EM algorithm", *Neural Computation*, vol. 6, pp. 181–214, 1994.

[117] X. Wang, P. Whigham, D. Deng, and M. Purvis, "Time-line hidden Markov experts for time series prediction," *Neural Information Processing, Lett. Rev.*, vol. 3, no. 2, pp. 39–48, 2004.

[118] C. M. Bishop and M. Svensen, "Bayesian hierarchical mixtures of experts," *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence*, 2003, pp.57–64.

[119] C. Sminchisescu, A. Kanaujia, Z. Li, and D. Metaxas, "Learning to reconstruct 3D human motion from Bayesian mixture of experts, a probabilistic discriminative approach," Department of Computer Science, University of Toronto, Toronto, ON, Canada, Tech. Rep. CSRG-502, Oct. 2004.

[120] S. Mossavat, O. Amft, B. de Vries, P. Petkov, and W. Kleijn, "A Bayesian hierarchical mixture of experts approach to estimate speech quality," in *Proceedings of the 2nd International Workshop Qual. Multimedia Exper.*, 2010, pp. 200–205.

[121] S. Mossavat, O. Amft, B. de Vries, P. Petkov, and W. Kleijn, "A Bayesian hierarchical mixture of experts approach to estimate speech quality," in *Proc. 2nd Int. Workshop Quality of Multimedia Experience*, 2010, pp. 200–205.

[122] M. I. Jordan and L. Xu, "Convergence results for the EM approach to mixtures of experts architectures," *Neural computation*, vol. 8, no. 9, pp.1409–1431, 1995.

[123] Y. Yang and J. Ma, "Asymptotic convergence properties of the EM algorithm for mixture of experts," *Neural Computation*, vol. 23, no. 8, pp. 2140–2168, 2011.

[124] L. Xu, M. I. Jordan, and G. E. Hinton, "An alternative model for mixtures of experts," in *Advances in Neural Information Processing Systems 7*, Cambridge, MA: MIT Press, 1995, pp. 633–640.

[125] K. Chen, L. Xu, and H. Chi, "Improved learning algorithms for mixture of experts in multiclass classification," *Neural Computation.*, vol. 12, no. 9, pp. 1229–1252, 1999.

[126] S. Ng and G. McLachlan, "Extension of mixture-of-experts networks for binary classification of hierarchical data," *Artificial Intelligence in Medicine*, vol. 41, no. 1, pp. 57–67, 2007.

[127] Y. Yang and J. Ma, "An efficient EM approach to parameter learning of the mixture of Gaussian processes," *Proceedings of the 8th International Conference on Advances in Neural Networks*, 2011, pp. 165–174.

[128] Y. Yang and J. Ma, "A single loop EM algorithm for the mixture of experts architecture," *Proceedings of the 6th Internatinal Symposium on Neural Networks*, 2009, pp. 959–968.

[129] D. M. Titterington, *The EM Algorithm, Variational Approximations and Expectation Propagation for Mixtures*, New York: Wiley, 2011, ch. 1, pp. 1–29.

[130] S.-K. Ng and G. McLachlan, "Using the EM algorithm to train neural networks: Misconceptions and a new algorithm for multiclass classification," *IEEE Transactions on the Neural Networks*, vol. 15, no. 3, pp. 738–749, May 2004.

[131] M. I. Jordan, "Hierarchical mixtures of experts and the EM algorithm," *Neural Computation*, vol. 6, no. 2, pp. 181–214, 1994.

[132] A.V. Rao, D.J. Miller, K. Rose, A. Gersho, "Mixture of experts regression modeling by deterministic annealing," *IEEE Transactions on Signal Processing*, vol. 45, issue. 11, 1997, pp. 2811 – 2820.

[133] A.V. Rao, D.J. Miller, K. Rose, A. Gersho, "A deterministic annealing approach for parsimonious design of piecewise regression models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, issue. 2, 1999, pp. 159 – 173.

[134] C.C. Chuang, S.F. Su, and C.C. Hsiao, "The annealing robust backpropagation (ARBP) learning algorithm," *IEEE Transactions on the Neural Networks*, vol. 11, 2000, pp.1067–1077.

[135] D. E. Van der Bout and T. K. Miller, "Graph partitioning using annealed neural networks," *IEEE Transactions on the Neural Networks*, vol. 1, pp. 192–203, March 1990.

[136] F. Peng, R. A. Jacobs, and M. A. Tanner, "Bayesian inference in mixtures-of-experts and hierarchical mixtures-of-experts models with an application to speech recognition," *Journal on American. Statistical Assocation*, vol. 91, no. 435, pp. 953–960, 1996.

[137] J. Shi, R. M. Smith, and D. M. Titterington, "Bayesian regression and classification using mixtures of multiple Gaussian processes," *International Journal of Adaptive Control and Signal Processing*, vol. 17, no. 2, pp. 149–161, 2003.

[138] C. E. Rasmussen and Z. Ghahramani, "Infinite mixtures of Gaussian process experts," *Advances in Neural Information Processing Systems 14*, Cambridge, MA: MIT Press, 2002, pp. 881–888.

[139] P. Estevez and R. Nakano, "Hierarchical mixture of experts and max-min propagation neural networks," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 1. Nov.–Dec. 1995, pp. 651–656.

[140] C.-M. Carlos, M. K. Sen, and P. L. Stoffa. "Hopfield neural networks, and mean field annealing for seismic deconvolution and multiple attenuation." *Geophysics*, vol. 62, issue. 3, 1997, pp. 992-1002.

[141] N. Ansari, E. Hou, "Mean Field Annealing," *Computational Intelligence for Optimization*, 1997, pp 71-81.

Appendix A - Additional Simulation Results

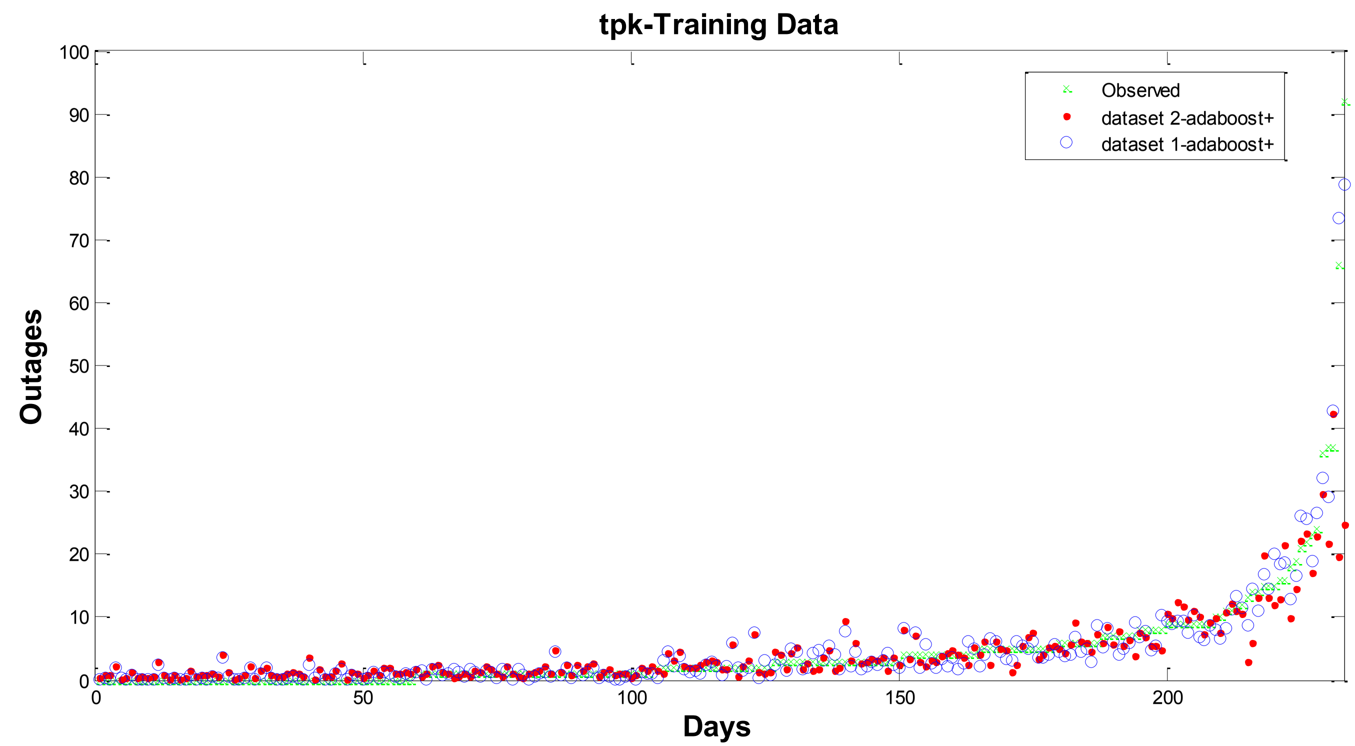


Figure A.1 Observed and Estimated Outages for Dataset 1 & 2 for Adaboost^{+} model for Topeka Training Data

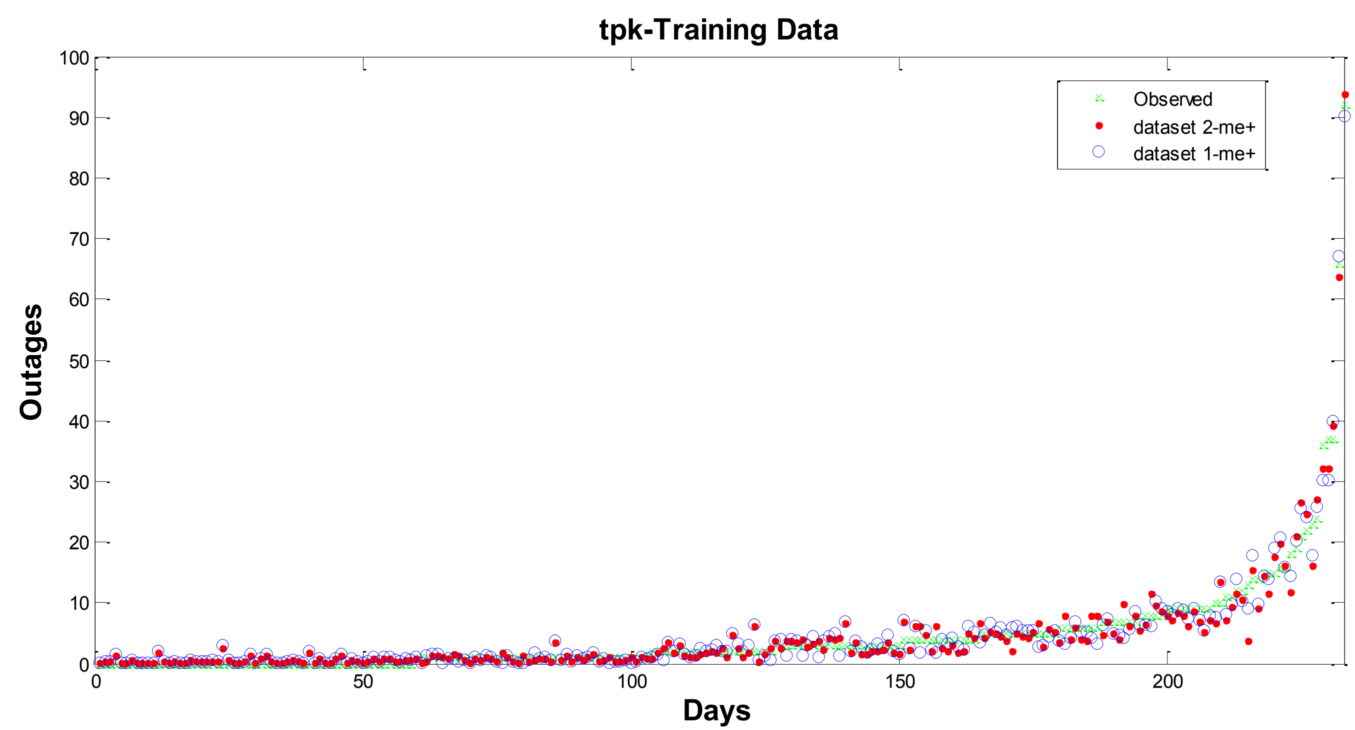


Figure A.2 Observed and Estimated Outages for Dataset 1 & 2 for AME model for Topeka Training Data

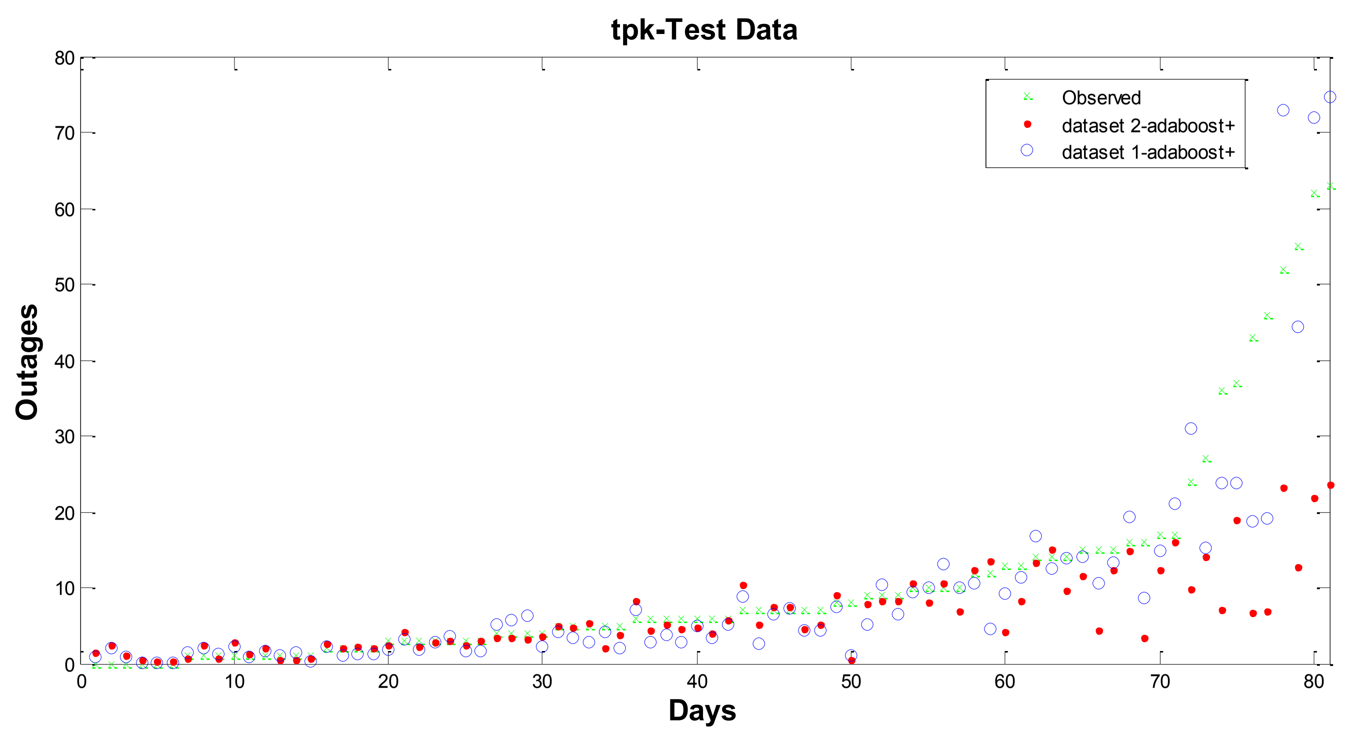


Figure A.3 Observed and Estimated Outages for Dataset 1 & 2 for Adaboost^{+} model for TopekaTest Data

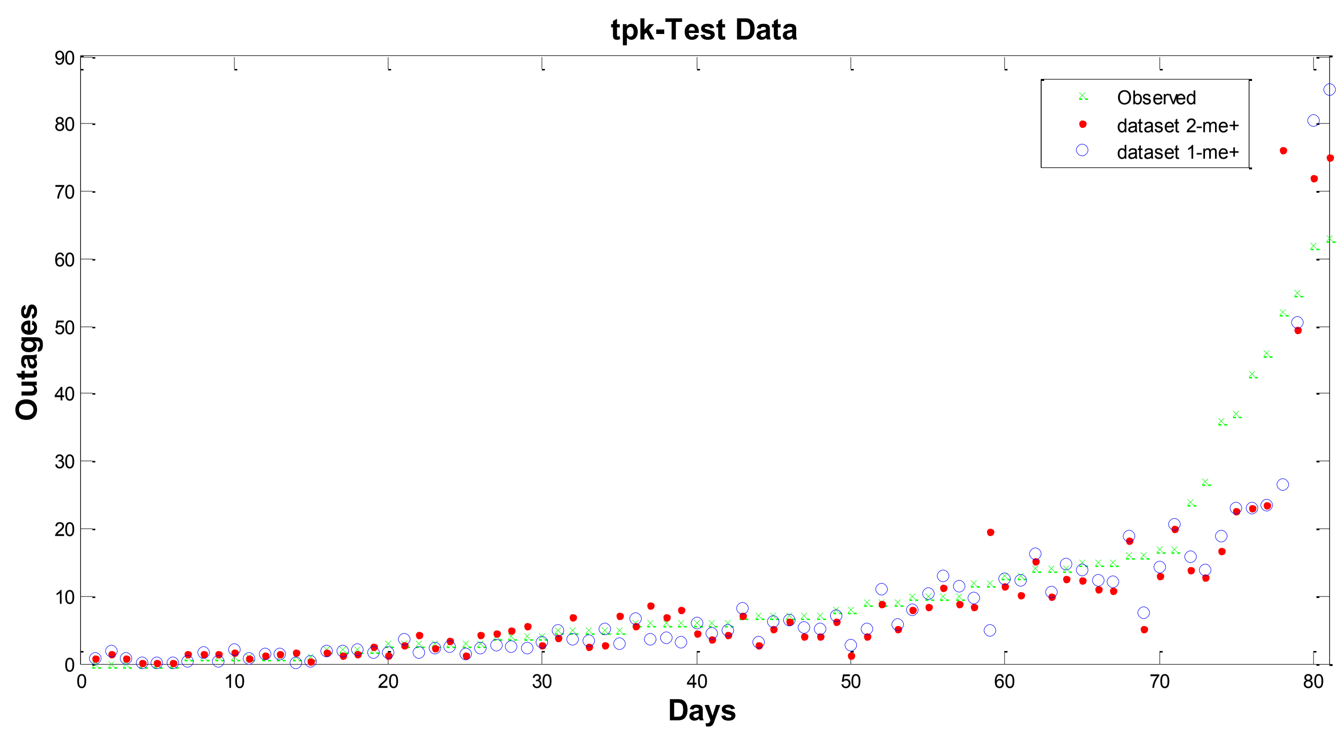


Figure A.4 Observed and Estimated Outages for Dataset 1 & 2 for AME model for Topeka Test Data

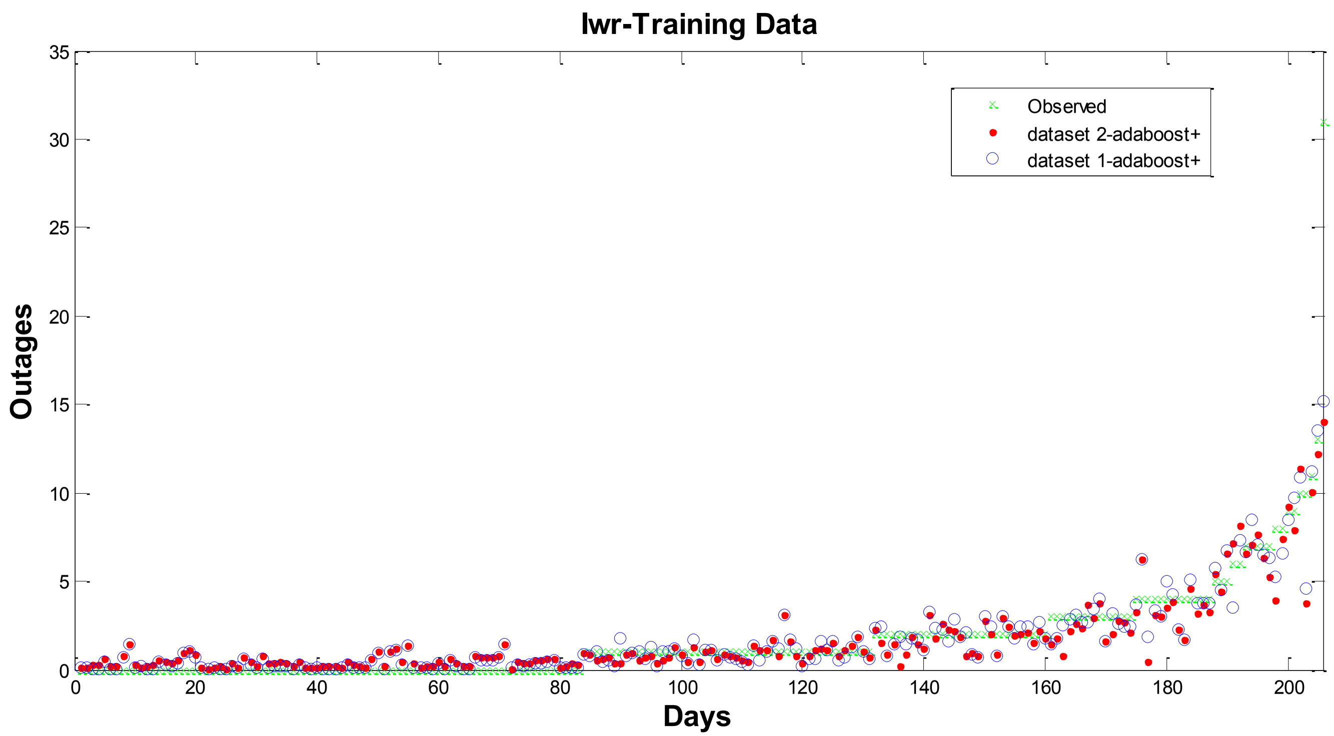


Figure A.5 Observed and Estimated Outages for Dataset 1 & 2 for Adaboost^{+} model for Lawrence Training Data

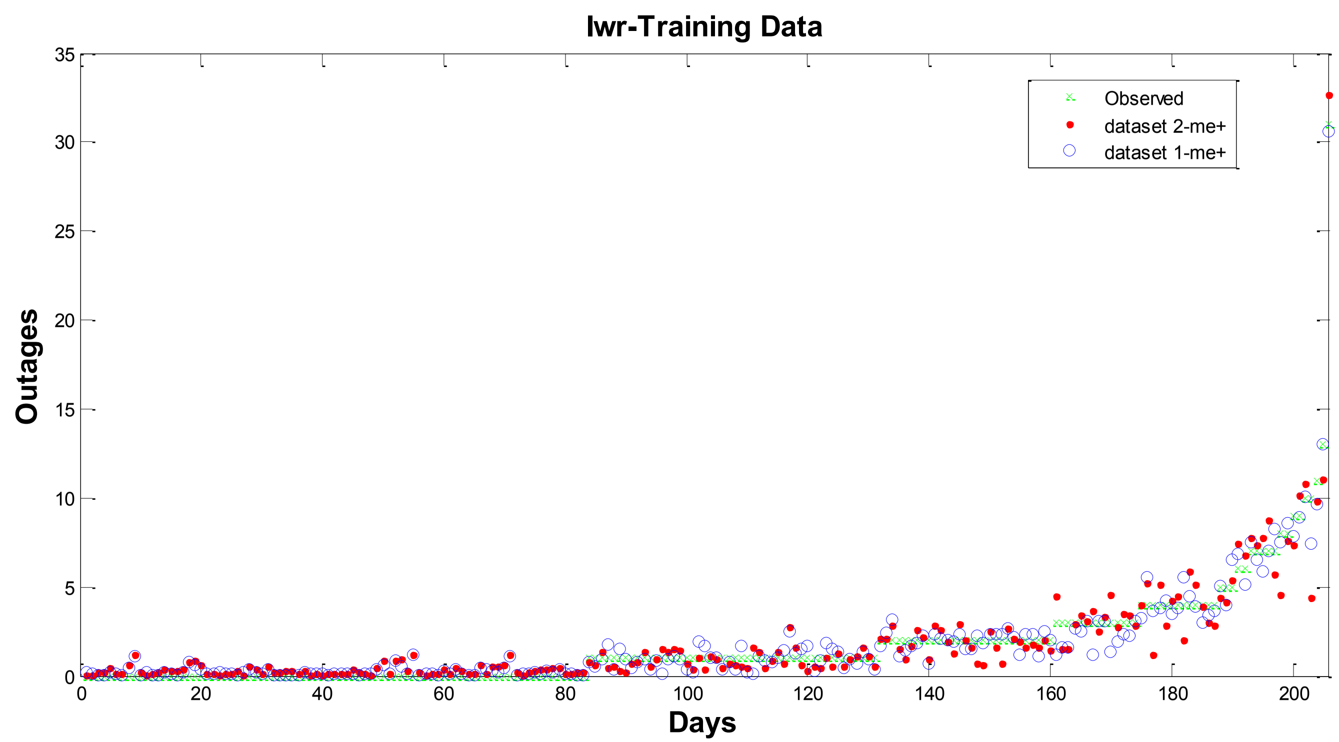


Figure A.6 Observed and Estimated Outages for Dataset 1 & 2 for AME model for Lawrence Training Data

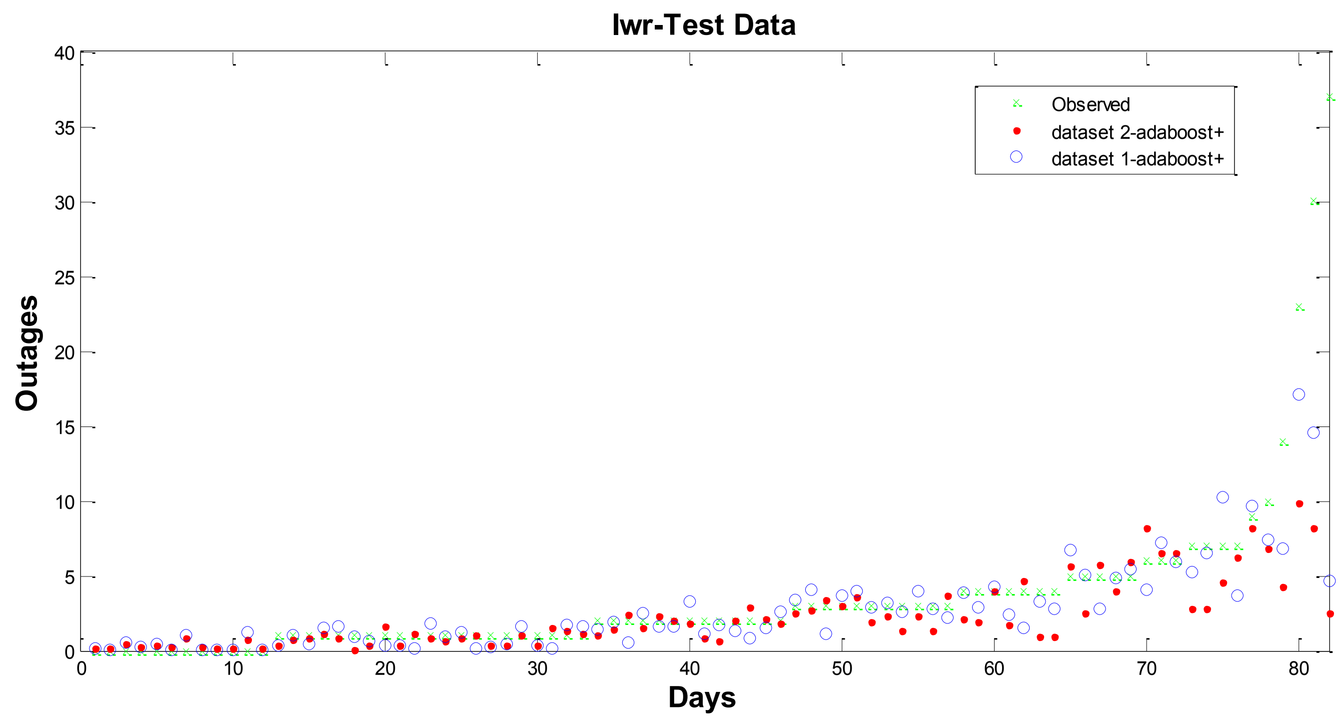


Figure A.7 Observed and Estimated Outages for Dataset 1 & 2 for Adaboost^{+} model for Lawrence Test Data

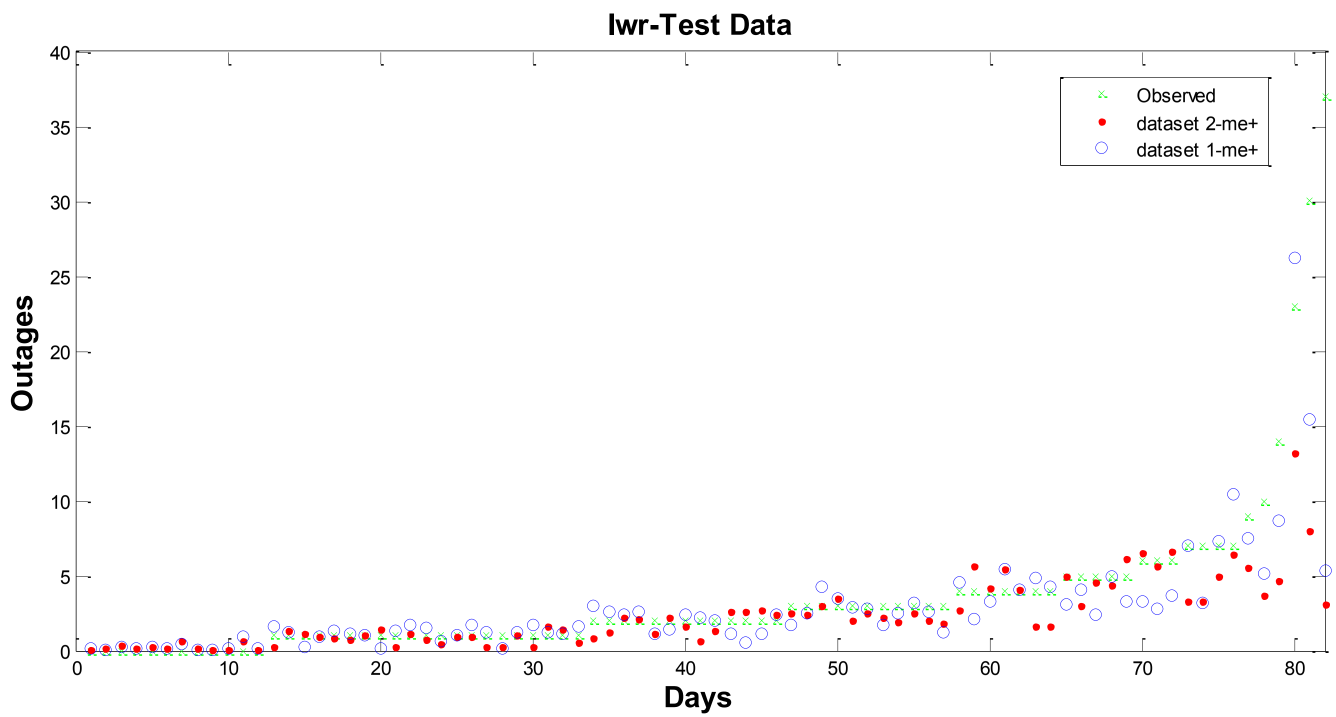


Figure A.8 Observed and Estimated Outages for Dataset 1 & 2 for AME model for Lawrence Test Data