ADVANCED METHODS FOR PREDICTION OF ANIMAL-RELATED OUTAGES IN OVERHEAD DISTRIBUTION SYSTEMS

by

MIN GUI

B.S., Central South University, 2000
M.S.,Central South University, 2003

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Electrical Engineering and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

Abstract

Occurrence of outages in overhead distribution systems is a significant factor in determining distribution system reliability. Analysis of animal-related outages has practical value since animals cause a large number of outages in overhead distribution systems. This dissertation presents several different methods to investigate the impact of weather and time of the year on the animal-related outage rate. The animal-related outages from year 1998 to year 2007 for different cities in Kansas are provided by Westar Energy. From examinations of the historical data, two factors which influence the animal-related outages, the month type and the number of fair weather days are taken as inputs along with historical outage data for prediction models. Poisson regression model, neural network model, wavelet based neural network model and Bayesian model combined with Monte Carlo simulations are applied to the weekly data of different cites. Even though Poisson regression models, Bayesian models and neural network models are able to recognize the changing pattern of outage rates under different weather conditions, they are limited in their ability to follow the high peaks in the time series of weekly animal-related outages. The introduction of wavelet transform techniques overcomes this problem. Simulation results indicate that the wavelet based neural network models are able to capture the pattern of fast fluctuations in the weekly outages of different cities in Kansas of various sizes. A hyperpermutation method inspired by artificial immune system algorithm is used to solve the overtraining problem in the application of neural networks. Finally, Monte Carlo simulations based on conditional probability tables from Bayesian models are used to find out the confidence intervals of the predictions. We aggregate the weekly data and carry out the analysis on a monthly and yearly basis too. Simulation results indicate that the models are able to capture the pattern as at least 90% of the observed values are within the upper limits of 95% confidence in the predictions for weekly, monthly and yearly animal-related outages of different cities in Kansas. The results obtained from Monte Carlo simulations are compared with the wavelet based neural network model to indentify years with more than expected level of outages.

ADVANCED METHODS FOR PREDICTION OF ANIMAL-RELATED OUTAGES IN OVERHEAD DISTRIBUTION SYSTEMS

by

MIN GUI

B.S., Central South University, 2000
M.S.,Central South University, 2003

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Electrical Engineering and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

Approved by:

Co-Major Professor
Anil Pahwa

Approved by:

Co-Major Professor
Sanjoy Das

Abstract

Occurrence of outages in overhead distribution systems is a significant factor in determining distribution system reliability. Analysis of animal-related outages has practical value since animals cause a large number of outages in overhead distribution systems. This dissertation presents several different methods to investigate the impact of weather and time of the year on the animal-related outage rate. The animal-related outages from year 1998 to year 2007 for different cities in Kansas are provided by Westar Energy. From examinations of the historical data, two factors which influence the animal-related outages, the month type and the number of fair weather days are taken as inputs along with historical outage data for prediction models. Poisson regression model, neural network model, wavelet based neural network model and Bayesian model combined with Monte Carlo simulations are applied to the weekly data of different cites. Even though Poisson regression models, Bayesian models and neural network models are able to recognize the changing pattern of outage rates under different weather conditions, they are limited in their ability to follow the high peaks in the time series of weekly animal-related outages. The introduction of wavelet transform techniques overcomes this problem. Simulation results indicate that the wavelet based neural network models are able to capture the pattern of fast fluctuations in the weekly outages of different cities in Kansas of various sizes. A hyperpermutation method inspired by artificial immune system algorithm is used to solve the overtraining problem in the application of neural networks. Finally, Monte Carlo simulations based on conditional probability tables from Bayesian models are used to find out the confidence intervals of the predictions. We aggregate the weekly data and carry out the analysis on a monthly and yearly basis too. Simulation results indicate that the models are able to capture the pattern as at least 90% of the observed values are within the upper limits of 95% confidence in the predictions for weekly, monthly and yearly animal-related outages of different cities in Kansas. The results obtained from Monte Carlo simulations are compared with the wavelet based neural network model to indentify years with more than expected level of outages.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First, I would like to thank my advisors, Dr. Anil Pahwa and Dr. Sanjoy Das for their guidance, encouragement and moral support during my graduate studies here at Kansas State University. Next, I would like to thank the professors and the staff of Electrical and Computer Engineering Department for their kindness and availability anytime I needed help. My graduate studies in EECE Department of Kansas State University have been the most enriching and exciting period of my education. Also, I would like to thank Westar Energy, Topeka, Kansas for providing the outage data and Mary Knapp of Kansas State University Weather Services for providing the weather information. I would like to specially thank Dr. Dwight Day and Dr. S.S Yang for their valuable advices in my research work at Kansas State University. Thanks to Dr. Todd Cochrane for agreeing to serve on my committee at the last moment due to Dr. Yang's illness and to Dr. Bharat Ratra for serving as outside chair for my defense. I want to thank the National Science Foundation for providing financial support through grant no ECS-0501288. Finally, I would like to thank my parents for encouraging me all the time, and also my boyfriend Zachery Moore for his support.

CHAPTER 1 - Introduction

Customers in the digital age have increased the demand for high reliability in electricity service since due to the wide usage of computers, clocks, and other electrical devices even momentary power interruptions cause major inconveniences and loss of productivity [1]. Also there is an increased demand for high reliability from regulators. In many states the utilities are required by the utility commissions to report reliability related system performance on an annual basis [2]. The utility companies strive to maintain a certain level of reliability in the generation, transmission and distribution part of the power system.

1.1 Significance of Overhead Distribution System Reliability

In relation to reliability issues, distribution system reliability is drawing more and more attention. Power distribution systems receive electricity from transmission systems and deliver it to customers. The reliability of distribution systems correlates directly with their ability to deliver power to the customers without outages. A very large part of the distribution systems in the US consists of overhead feeders with radial configuration, which saves capital cost but is not always reliable. Locating in highly populated areas, unique configurations and relatively simple protection mechanisms in distribution systems makes distribution systems more responsible for most of the interruptions experienced by customers than generation and transmission systems [3]. It has been reported that 80% of the interruptions experienced by customers are due to outages in distribution systems [4] and on average, an outage of a segment on a feeder will interrupt service to about half of the customers it serves [5]. Although historically utilities have maintained a very high level of reliability, pressure on them to continue to maintain this has gradually increased over the past several years. This is because some state utility commissions are imposing or proposing penalties on utilities for not providing certain expected levels of reliability. The situation is further compounded by the fact that customers of the digital age expect a higher level of reliability, while the utilities operate under a tighter budget. Thus, distribution system reliability is becoming a very significant part of the utility business.

In terms of reliability, the overhead distribution feeders distinguish themselves from the other components. Power distribution systems are made up of a network of both overhead and underground wires. Overhead feeders make up the highest percentage of the total number of equipment units in distribution systems. Compared to its alternative, underground cables, overhead feeders are less expensive, and easy to install and maintain [3]. However, because of the entire exposure to outside environments and vicinity to trees and houses, they are highly susceptible to wind, ice storm, vegetation, and animals. These exposures create a variety of problems for the power distribution industry by causing interruptions in the distribution systems. Finding good models to study outages caused by these factors on overhead distribution lines is an important step for analyzing the component reliability data and for predicting system reliabilities. Among other factors affecting outages, weather has a great deal of influence. It not only directly causes shorts or breaks on overhead lines but also interacts with trees in damaging the electricity delivery path. Moreover, outages caused by weather and other environment factors are random, with a higher probability under adverse conditions and are difficult to eliminate completely.

1.2 Distribution System Reliability Assessment

Power system reliability has been a subject of interest to researchers for over 30 years. IEEE has periodically published a bibliography on power system reliability evaluation. The latest were published in 1994 [6] and in 1988 [7]. A study of these bibliographies reveals that a majority of the literature on reliability deals with generation or transmission systems. Interest in distribution systems started in 1971 with an Edison Electric Institute report [8]. However, it is in the 1980 when interest in this subject grew as a result of studies conducted by the US Department of Energy [9], Electric Power Research Institute [10], and the Canadian Electric Association [11]. This was followed by analytical and simulation-based methods [12 – 14] for evaluation of distribution system reliability and their application to various reliability improvement studies. Reliability of electric power distribution systems is defined as the ability to deliver uninterrupted service to customers [15]. Based on the duration of the interruptions, customer interruptions are divided into two categories, sustained interruptions and momentary interruptions. According to the IEEE standard [16], five minutes is the cut-off between momentary and sustained interruptions. The most commonly used reliability indices are SAIFI (System Average Interruption Frequency Index), SAIDI (System Average Interruption Duration

Index) and MAIFI (Momentary Average Interruption Frequency Index) [17]. Besides these three indices, other indices such as CAIFI, CAIDI and ASAI are also used [18].

Utilities seek high reliability to serve customer needs, while trying to minimize the cost spent on reliability improvement. This is called the reliability worth/cost trade-off. It has stimulated significant interest in value-based distribution reliability assessment in the last fifteen years. The value-based reliability planning approach aims at finding the minimum cost solution to meet the target level of reliability. The cost is identified as the sum of capital cost, operation and maintenance cost, and cost of outage to the customers. Some examples of this approach are available in [19 – 26]. The cost of outage to the customers needed for value-based planning is determined from customer surveys [27 – 29].

Some of the literature on distribution system reliability also deals with data collection, data modeling, and practical projects for reliability improvement. In [30] and [31] the authors have described statistical models for the analysis of outage data of overhead lines. A procedure for reporting distribution field inspection data is given in [32], and a database management system for distribution system reliability evaluation is described in [33]. Results of projects implemented at different utilities are discussed in [34 – 36]. Culmination of research and experience over the years has resulted in three approaches for reliability assessment of distribution systems, which are historical, predictive and feature-based. These three methods are reviewed next.

1.2.1 Historical Approach

Analyzing historical data or looking back is the most common approach used by utilities to assess reliability of distribution systems. The utilities record all the sustained outages and a cause code is assigned to that outage based on the available evidence. The outage database also contains information on the number of customers affected and the duration for each outage. These data are then used to compute reliability indices, such as SAIFI and SAIDI for the whole system as well as for individual feeders. Distribution system feeders with high values of these indices are targeted for improvement. Though useful, this is a reactive approach and it can only be used to identify the less reliable parts based on the past performance. This method cannot reliably forecast future problems.

1.2.2 Predictive Approach

Nowadays, the utilities are looking towards more proactive methods for finding distribution system reliability, so that they can distinguish the less reliable parts of the system and act proactively rather than waiting for the outage to take place before taking corrective measures [17,37]. This way the utilities can improve the system reliability indices [37]. In addition, a knowledge of the expected number of outages that can take place in the system given a particular combination of certain causal factors, might help the utilities in justifying the bad performance of their system in a particular year, thereby saving them from paying performance-based penalties to the regulatory bodies. Various methods for predictive reliability assessment of distribution systems have appeared in the literature. Some of these methods have been discussed in [14, 15, 38-43].

In comparison to historical reliability assessments, predictive reliability assessments are more significant since they provide more information in guiding utilities to identify future reliability-involved activities. Good predictive reliability assessments need good predictive data. Predictive or *looking ahead* methods for reliability assessment of distribution systems rely either on analytical means or on simulation-based approaches. Analytical methods use probabilistic outage models of different components and then combine them with network models or state-space models to predict overall reliability of the system [40, 44, 45]. Since analytical models are mathematically quite involved, simplifying assumptions such as constant failure rates for systems components and pre-defined probability distribution functions for outages are used to compute expected values and variances of reliability indices. Some probability distribution functions make it difficult to implement analytical solutions. In those cases simulation-based methods such as Monte Carlo approaches are used [17, 41, 46 – 53]. These methods require extensive computing, but they have become more popular lately due to increased power of computers. Failure rates of different system components and their probability distributions are the most crucial data for these analyses. Usually, most of the existing methods use average values of failure rates published in the literature. However, using average failure rates in reliability analysis could be misleading [54, 55]. Both analytical and Monte Carlo methods are useful, but they suffer from a serious drawback, which is that the user has to pre-define the probability distribution function for failures and also select an average value for failure rate of components or parts of a system. The user has an option of adjusting the failure rates for components based on empirical judgment or calibrating them based on past performance [55]. In real-life, some components within a system may have different failure rates due to exposure to diverse factors [56, 57], which are very difficult to estimate. Thus, the computed reliability indices based on analytical as well as Monte Carlo methods may not represent a real-life scenario accurately. Some authors have used different functions to account for change in failure rates over time due to different factors based on empirical judgment [58]. This approach makes implementation of the Monte Carlo analysis more complicated and computationally more intensive.

1.2.2 Feature-Based Approach

Although a large amount of literature is available on distribution system reliability, very little of it addresses the modeling of the effects of environmental factors on reliability. Gilligan [39] first proposed a method based on circuit configuration and its features to obtain a numerical index to estimate reliabilities of segments of a feeder as well as for the entire feeder. This method combined the effects of length, conductor size, sectionalizing devices, and tree density with the number of customers affected to obtain a risk index for the feeders. Lang and Pahwa [15] did similar work based on a fuzzy knowledge based approach. They developed a fuzzy assessment model to calculate a risk index for every section of a circuit as well as for the whole circuit. Chow, Yee, and Taylor [59] used an approach based on artificial neural networks to identify animal caused outages. Radmer and others [60] studied several models for predicting vegetation-related outage rates of overhead lines. Their models included linear-regression, exponential, multivariable linear-regression, and neural networks with temperature, precipitation, and date last trimmed as the inputs. The authors suggest inclusion of other environmental variables as inputs to the model. Kuntz, Christie, and Venkata [61] followed this work, in which they determine optimal schedule for vegetation management in the vicinity of distribution lines. In this Monte Carlo based simulation study, the authors used a multi-step function to account for increase in outage rates over time due to vegetation. Similarly, Monte Carlo simulation has been used with time variable outage rates to evaluate effects of wind storms [62] and to evaluate effects of lightning [63, 64]. Williams [65] and McDaniel, Williams, and Vestal [66] recently presented results based on simple linear regression models for determining effects of wind and lightning on distribution system reliability indices. These models provide a very rough fit to the data. Another recent research discusses the categorization of tree-related outages [67]. The available literature clearly suggests growing interest in the industry on studying the influence of various environmental factors on the reliability of distribution systems. However, current methodologies have shortcomings, such as:

- weather conditions are categorized into only two or three states;
- only duration of a specific weather state is considered while the strength of the weather state is omitted;
- contributions of various factors to outage rates are modeled based on empirical findings without much emphasis on physical cause-effect relationship;
- modeling is based on one or two years of data;
- effects of only one or two factors are addressed.

Thus, these models do not provide a comprehensive methodology to address a large set of environmental factors. The proposed research builds upon the existing knowledge and introduces novel modeling methodologies to capture the effects of various environmental factors on distribution system reliability with an emphasis on causal relationships.

1.3 Causes of Outages in Distribution Systems

Utilities typically report annual performance of the distribution systems using commonly accepted indices, which are SAIFI, SAIDI and MAIFI [68 – 70]. Annual indices, however, are only a snapshot of the system for a given year and do not provide a true representation of the system performance [71, 72]. Various environmental and other factors influence the system performance in complex ways, and without knowing this influence it is not possible to correctly evaluate the reliability performance of distribution systems. Although the utilities exclude outages caused by extreme weather conditions [73], the system performance can have a large variance even for days considered non-extreme based on weather. Quite often the system performance over a given year deteriorates even after implementing several system improvements. The utility engineers intuitively know that this happened due to worse than average weather conditions or some other causes. Therefore, it is important to look into the causes and have a better insight of the system performance.

There are various factors that cause outages in distribution systems. Ten general categories for interruption cause are suggested for comparison in benchmarking studies. These are intentionally broad categories that will make possible more precise benchmark comparisons between different distribution utilities. There are numerous categories that could be chosen, but with the goal of uniformity for comparison the purpose of the IEEE task force arrived at the following ten categories [74]:

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

The recommended categories do not prevent utilities from collecting more detailed data, and that is indeed encouraged. However, the data collected should be able to be lumped into one of the ten categories recommended.

Of these causes, animals are a main cause of outages on overhead distribution systems. Animal/wildlife includes mammals, birds, reptiles, and insects or any other members of the animal kingdom. Squirrels and snakes cause outage in distribution systems by climbing up the distribution poles or transformers and creating short circuits between phase wires and ground [37]. Birds usually perch on the lines and spread their wings, which results in short circuits [37]. Wildlife can cause interruptions directly through contact, as with snakes, mice, ants, raccoons, squirrels, or birds, or indirectly as with nests and bird excrement. In Figure 1.1 [75] a squirrel climbed up the distribution pole and very possibly would have created a short circuit between the phase wire and ground. In Figure 1.2 [75], an owl perched on the lines, spread its wings and caused a short circuit fault.

Figure 1.1 An Owl Caused Outage in the Distribution System [75] (With Permission of Rick Harness)

Figure 1.2 A Squirrel Perched on a Power Line [75] (With Permission of Rick Harness)

In Figure 1.3 the percentage of outages by different causes in the overhead distribution system in the Manhattan area in the years 2003 and 2004 is shown. Note that the categories in Manhattan are different from the recommended categories by having two additional causes which are extreme winds and ice storms. However extreme winds and ice storms can be included within weather. We can see that animals caused 24% of the outages, which is the second most significant cause, next to vegetation. This translates into millions of dollars lost due to reduced power use, man-hours paid for repair, and the cost of replacing damaged equipment**.** Thus, a good method to evaluate impacts of animal activities on overhead distribution lines is to track the random process of animal-related outage events. This will thus allow utilities to better understand animal impacts on distribution reliability and to choose better operation and maintenance plans.

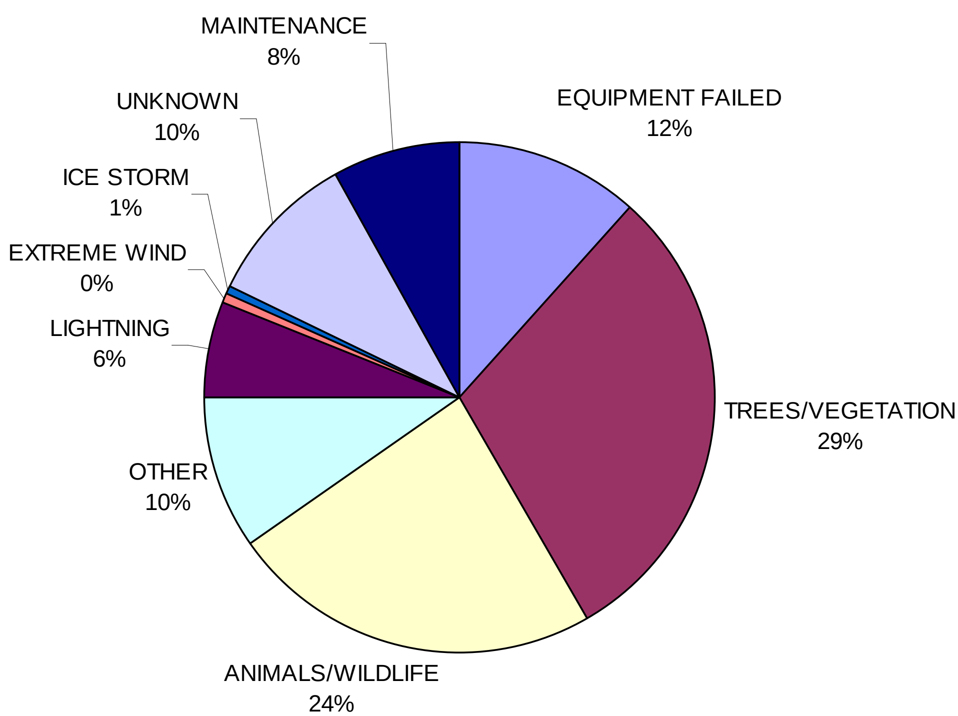

Figure 1.3 Percentage of Outages by Different Causes in Manhattan in the years 2003 and 2004

1.3 Previous Work

Zhou and Pahwa have done research on the weather's impact on overhead distribution lines' failure rates [3]. A Poisson model and a Bayesian model are presented in [3]. The methods took wind gust speeds and lighting stroke currents as inputs and tried to capture the probabilistic relationship between each weather state and failure level. The simulations with historical data showed that both the Poisson regression model and the Bayesian model provided good ways to model the failure rates of overhead distribution lines.

Sahai and Pahwa have done research on the weather's impact on animal-related outages in overhead distribution systems [76]. Examination of historical data showed that the animal-related outages mainly take place on fair weather days. Also, the behavioral patterns of animals in different months and their impact on animal-related outages were discovered [76]. Finally, a Bayesian model is constructed for prediction of animal-related outages in overhead distribution systems given the two factors, the month type and the number of fair days per week [76]. This Bayesian model was applied to data of five cities in Kansas. The weekly and the monthly predictions were done and the confidence intervals for the predictions were found.

1.4 Scope of This Dissertation

This dissertation is focused on the study of outage data and weather data for four different cities in Kansas from year 1998 to year 2007. Several analysis models are presented. In Chapter 3, Poisson regression model is applied to the given data. Even though the model can catch the changing pattern of animal-related outages caused by changing weather and time of year, it has a problem in precisely catching the high peaks in the time series of animal-related outages due to the random nature and noise in the outage data. In order to provide accurate prediction for extreme cases, neural network (NN) model, which has better ability to approximate high complexity equations, is applied in Chapter 4. In Chapter 5, we use discrete wavelet transform (DWT) to decompose the time series into more smooth components, which are an approximate coefficient series (low frequency information) and three detail coefficient series (high frequency information). We construct different feed-forward NNs for each coefficient series according to its characteristic and thereby improve the forecast accuracy. To overcome overtraining of neural networks, we have applied an artificial immune system (AIS) approach to fine-tune the trained neural networks during the testing stage. The output series are reconstructed to get the final output for the time series. The simulations show that the wavelet decomposition technique overcomes the under estimation for extreme cases. In Chapter 6 a method based on Bayesian models, which is an extension of the method presented in [76], is presented. Based on the conditional probability table and the smoothed histogram found in Bayesian model, we have used Monte Carlo simulations to find the 95% confidence limits of the weekly prediction. By aggregating Monte Carlo simulations for weekly outages, we obtained the confidence limits for monthly and yearly data. To justify the outliers in yearly prediction which lie outside of the confidence limits, we have taken the outliers as testing data and applied wavelet based neural network again. The confidence limit of the testing year is obtained from the training errors. In the conclusion section, we point out that the methods presented in this paper are useful to utilities for end of the year analysis of past year's reliability performance of the distribution systems. Performance of a specific year can be compared with the past performance to identify deviations. Significant increase in outages would require the utility to do further analysis and take remedial actions.

CHAPTER 2 - Characteristics of Animal-related Outages

In order to develop a mathematical model for prediction of animal-related outages, we need to find the inputs for the model which influence their occurrences. In previous research, an effort has been made to find the correlation between the population density of animals in Kansas and the total number of animal-related outages but no direct relation was discovered because of a lack of information [76]. Therefore, we have tried to find a correlation between animal activities and animal-caused outages by analyzing historical outage information under different weather conditions and in different months of the year and relating it to the behavioral patterns of animals.

2.1 Animal-related Outages and Weather

While recording outages, utilities also record the weather conditions at the time of outages. The weather is described by a code which specifies one of the following conditions:

1) Thunderstorm;
2) Lightening;
3) Wet;
5) Ice;
6) Ice and wind;
7) Wind;
8) Heat;
9) Cold;
10) Fair;
11) Tornado;
12) Microburst.

As we have already seen, animals are the second most significant cause, next to tree/vegetation, for outages in Manhattan in the years 2003 and 2004. However, further examination of outages by animal under different weather conditions shows that animals become the most significant cause under fair weather conditions as shown in Figure 2.1. This indicates that fair weather affects animal-related outages more than other causes.

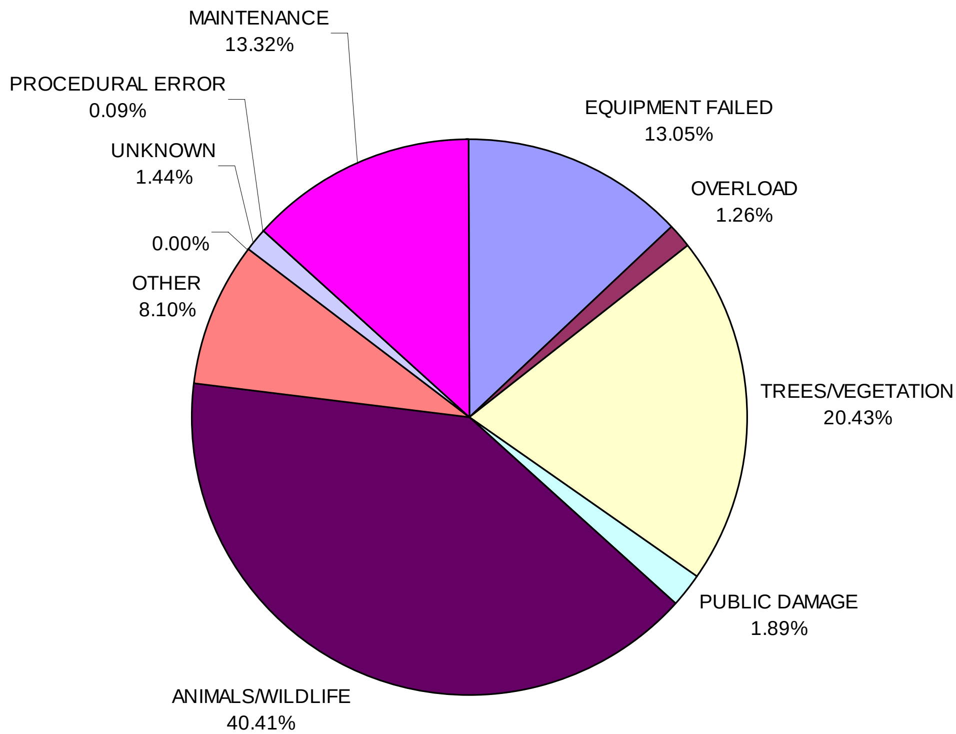

Figure 2.1 Percentage of Outages Causes under Fair Weather Conditions

Table 2.1 [76] shows the occurrences of animal-related outages under different weather conditions from the year 1998 to the year 2002. It can be clearly seen that almost all the outages due to animals took place in fair weather. Fair weather days have temperatures within 40 and 85 degrees Fahrenheit with no other weather activity as shown in Table 2.1. Under other weather conditions, there are not many occurrences of animal-related outages [76]. The reason behind this observation is the fact that animals are most active in fair weather [37]. When there are strong winds, ice, thunder storms or other unfavourable weather conditions, they stay in their nests [37, 77]. Since fair weather influences the occurrences of animal-related outages, the number of fair weather days in a given period of time is taken as one of the causal inputs for the prediction models for animal-related outages.

Table 2.1 Numbers of Outages under Different Weather Conditions in Manhattan from 1998 to 2002 [76]

Weather Codes	Description	Total Number of Animal Outages
1	THUNDERSTORM	3
3	WET	20
4	WIND >50 MPH	2
5	ICE	0
6	ICE AND WIND	0
7	WIND <50 MPH	0
8	HEAT	17
9	COLD	31
10	FAIR	763

2.2 Month Type

In the previous work of Sahai and Pahwa, animal-related outages in different months were compared with each other to find out the pattern of animal activities in different months and their impacts on the occurrences of outages [76]. We have done the same investigation with more data. From more than ten cities in Kansas, we have chosen two bigger sized cities, Wichita and Topeka, and two cities of smaller size, Lawrence and Manhattan, to do the data examination, in order to get reliable observations for the whole state. The historical animal-related outages data of Wichita, Topeka, Lawrence and Manhattan are aggregated on a monthly basis to discover the impact on outages of different behavioral patterns of animals in different months. The pattern during 12 months in one year is consistent from year to year. To obtain a comprehensive view of the data from the year 1998 to the year 2007, we take the average monthly animal-related outages over the 10 years.

These observations are consistent with those reported by Sahai and Pahwa [76]. In Figure 2.2 the average outages in every month from 1998 to 2007 for Wichita, Topeka, Lawrence and Manhattan are presented. This graph shows that the highest number of animal-related outages took place in the month of October, followed by May, June, September and November. These are the months with most active animal activities. It is observed that the majority of animals which cause outages in Kansas are squirrels and the squirrels are noticeably more active in September, October and November. During these three months, they are busy collecting foods such as nuts for the winter [77]. The least number of outages due to animals happens in the months of January, February and March, which are usually cold months with less animal activities. Thus it is reasonable to expect lesser outages due to animals in these months. In April, the weather becomes warmer and more animal activities are expected than during the winter. However, for some cities we have observed only slight increases in number of outages in April due to life cycle of squirrels. There are mainly two kinds of squirrels in Kansas: grey squirrels and fox squirrels. Every year grey squirrels mate twice while fox squirrels only mate once. The mating of grey squirrels starts in January and July. Fox squirrels, which are more common in Kansas [77], usually mate in January with gestation periods of about 44 to 45 days [77]. The fox squirrels give birth to baby squirrels in March and the babies do not venture out of the nest for 8 weeks. This coincides with the month of April. This explains why there is no noticeable increase in squirrel activities in April even when the weather is good. In May and June when weather is fair, the babies come out of the nest with their mothers and stay outside a lot. As a result we have observed a dramatic increase of the outages in May and June. July and August are very hot and during these months less animal activity can be expected. In December, the weather starts becoming cold and animals begin to hide. However, in some years the weather can stay pleasant in this month which results in more animal activities [76].

As we observed, animals have different behavioral patterns in different months of the year and thus months have considerable impacts on animal-related outages in overhead distribution systems. For convenience of model construction in the later chapters, we have done month type classifications based on the outages in each month. The months that have higher animal-related outages and indicate higher animal activities, which are May, June, September, October and November, are grouped together and classified as month type 3. The months in the winter with very low outage level, which are January, February and March, are grouped together and classified as month type 1. The animal activities in the very hot months of July and August and the temperature transition months of April and December are moderate. Hence, these months have been grouped together and classified as month type 2. This gives us the following classification for the month type:

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

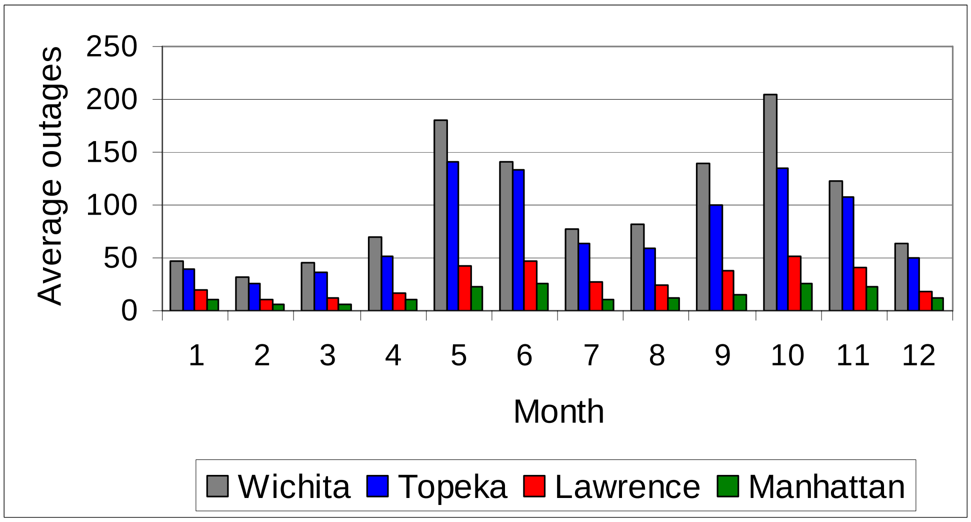

Figure 2.2 Average Monthly Animal-related Outages from 1998 to 2007 for Wichita, Topeka, Lawrence and Manhattan

Note that in the previous work [76], the month type classification was:

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

The difference between this dissertation and the previous work is that this dissertation has classified April as month type 2 instead of 1 and classified January as month type 1 instead of 2. In the previous work, the month type classification is only based on observation of historical data of one city, Manhattan, instead of four cities as in this dissertation. Also, the data in the previous work is only from the year 1998 to the year 2004, which is 3 years less than the data in this dissertation. With more data, this dissertation has better observations and leads to more reliable classification of the months.

Graphs on animal related outages in each month for each of the ten years are shown in Figure 2.3 to 2.12. These graphs appear to be consistent with one another.

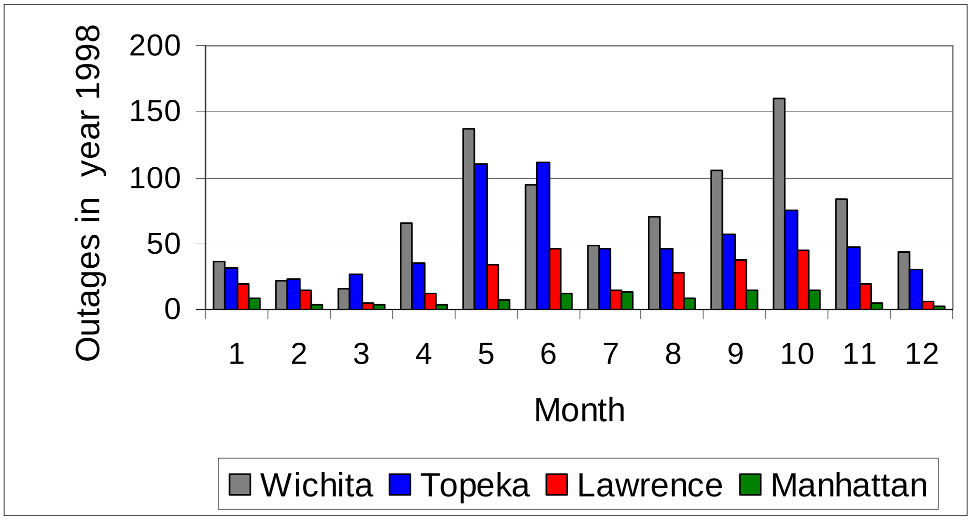

Figure 2.3 Average Monthly Animal-related Outages in the Year 1998 for Wichita, Topeka, Lawrence and Manhattan

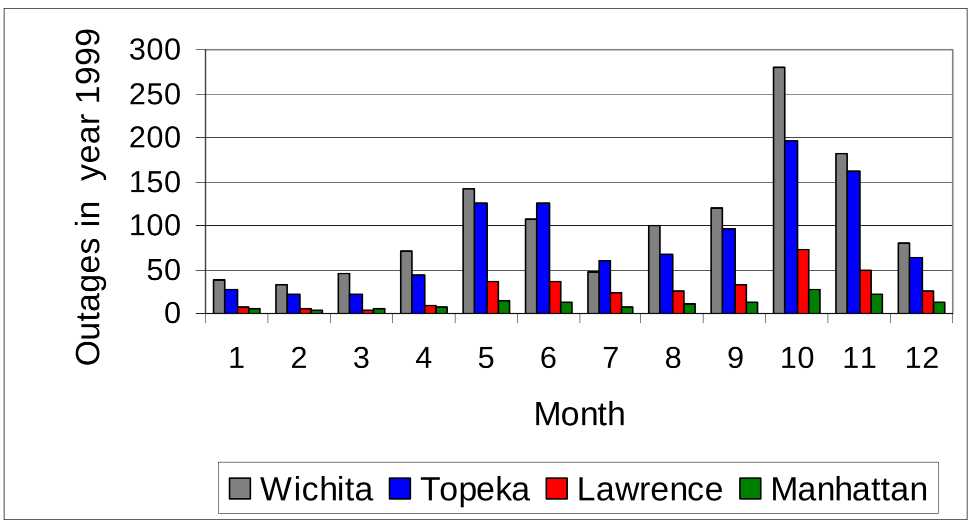

Figure 2.4 Average Monthly Animal-related Outages in the Year 1999 for Wichita, Topeka, Lawrence and Manhattan

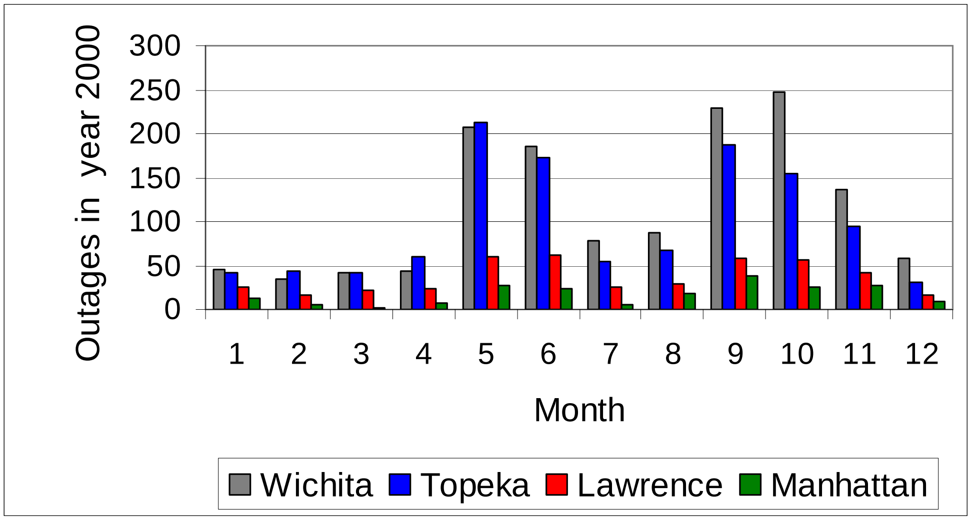

Figure 2.5 Average Monthly Animal-related Outages in the Year 2000 for Wichita, Topeka, Lawrence and Manhattan

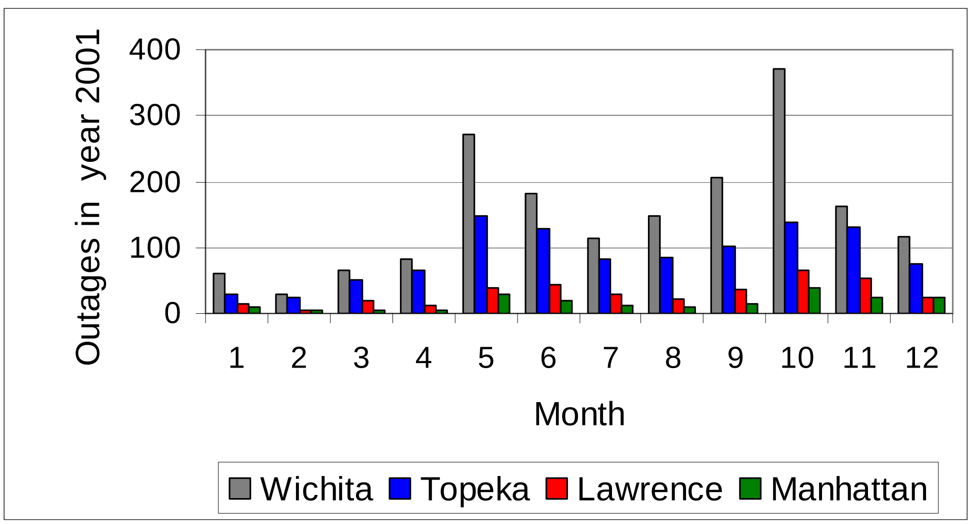

Figure 2.6 Average Monthly Animal-related Outages in the Year 2001 for Wichita, Topeka, Lawrence and Manhattan

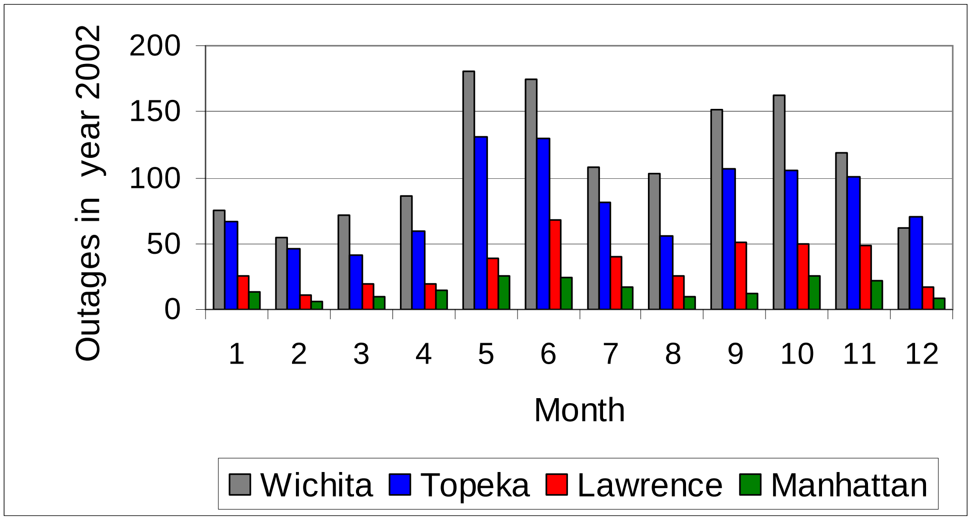

Figure 2.7 Average Monthly Animal-related Outages in the Year 2002 for Wichita, Topeka, Lawrence and Manhattan

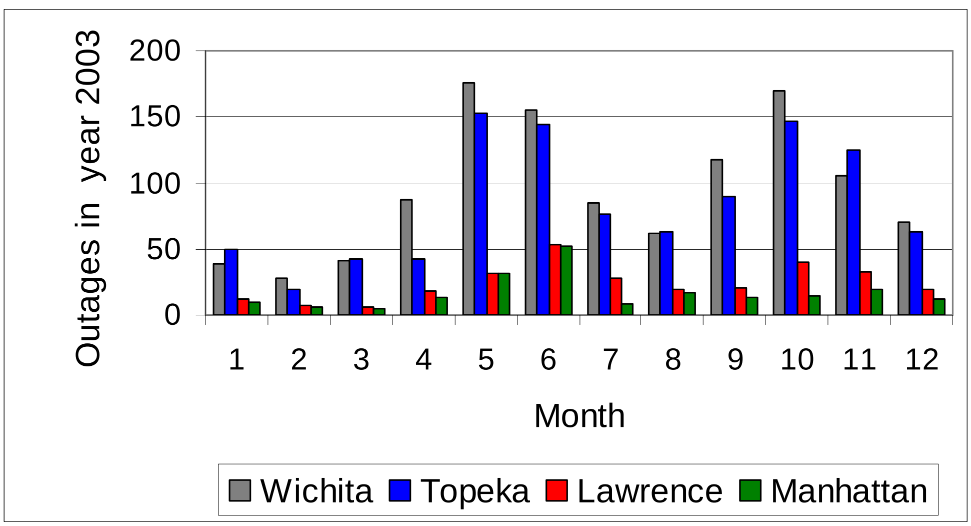

Figure 2.8 Average Monthly Animal-related Outages in the Year 2003 for Wichita, Topeka, Lawrence and Manhattan

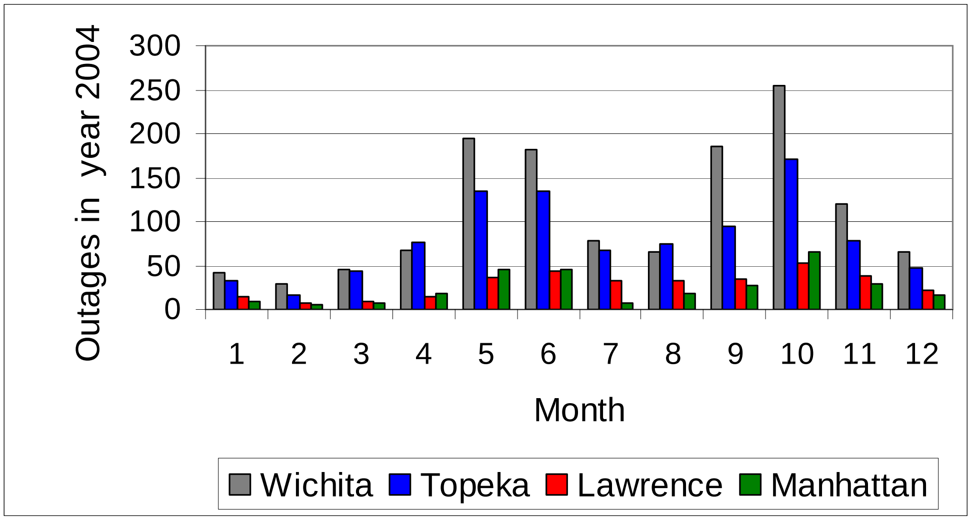

Figure 2.9 Average Monthly Animal-related Outages in the Year 2004 for Wichita, Topeka, Lawrence and Manhattan

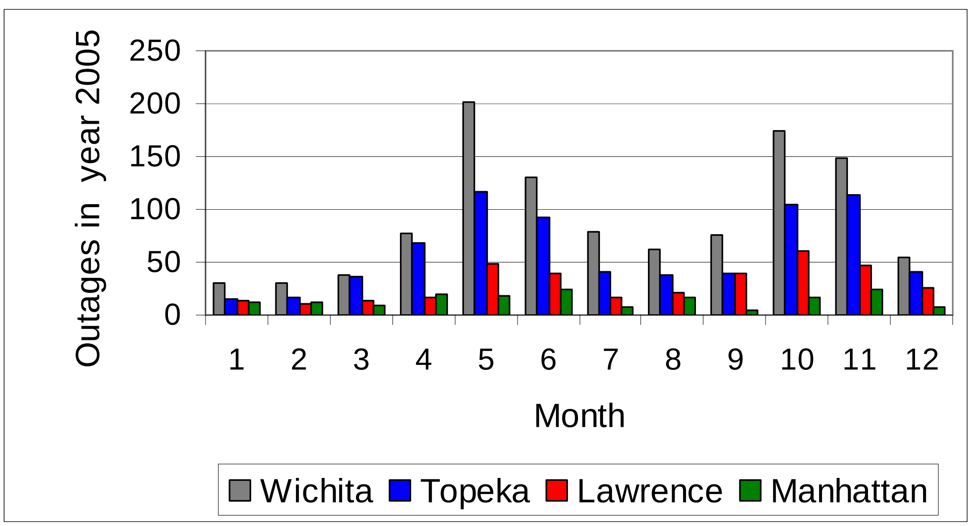

Figure 2.10 Average Monthly Animal-related Outages in the Year 2005 for Wichita, Topeka, Lawrence and Manhattan

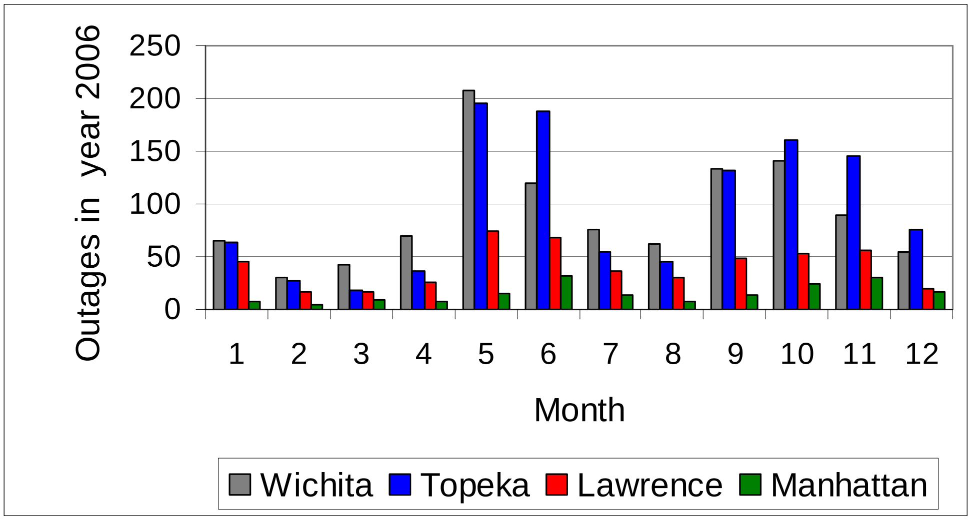

Figure 2.11 Average Monthly Animal-related Outages in the Year 2006 for Wichita, Topeka, Lawrence and Manhattan

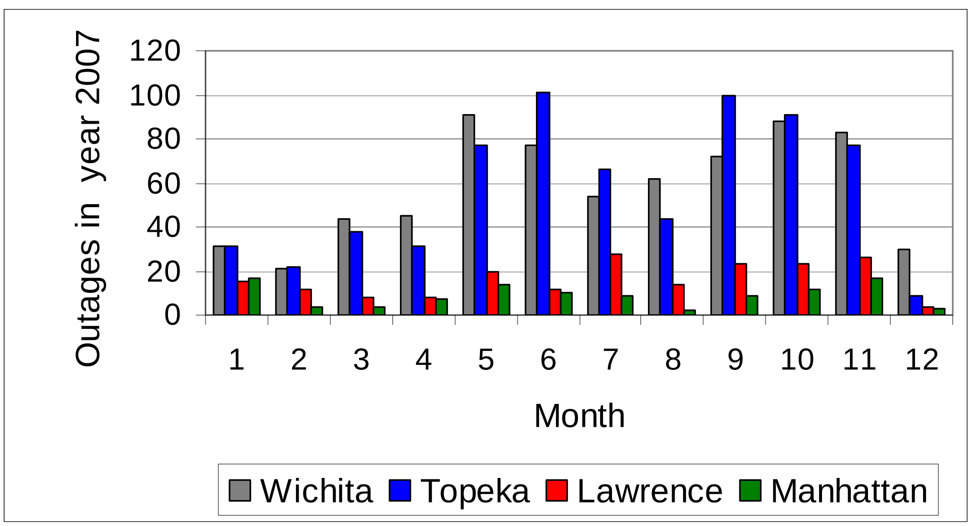

Figure 2.12 Average Monthly Animal-related Outages in the Year 2007 for Wichita, Topeka, Lawrence and Manhattan

2.3 Weekly Animal-related Outages

Because of randomness, it is not practical to accurately predict the animal-related outages for a specific day or for a certain overhead line. However, predictions for a relatively long period or large number of overhead lines can produce satisfactory results since randomness will be averaged due to a large sample size. Therefore, predictions of weekly animal-related outage in a whole distribution network for a district, in a certain period of time have a significant practical value. As done in the previous work [76], a weekly basis is found to be the proper duration for predictions. For uniformity and ease of classifications of data, each month is considered to be composed of exactly four weeks [76]. Since a month can have 28, 29, 30 or 31 days, it is difficult to allocate the weeks evenly in a particular month. To ensure that every week belong to only one month, some weeks have eight days [76]. For the months which have 31 days, the first week has 7 days and the rest three weeks each have 8 days. For the months which have 30 days, the first two weeks each have 7 days and the rest two weeks each have 8 days. However for February which usually have 28 days, each week has 7 days. If it is a leap year, the last week of February has 8 days. This classification does not impact the results because both the input and the output have the same classifications for weeks [76].

Based on the classifications, the number of fair weather days per week can vary from 0 to seven or eight [76]. Since the number of fair days in a week has an impact on the occurrences of outages in that week, it 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.

We plot the time series of weekly animal-related outages from the year 1998 to 2007 as shown in Figure 2.13. In total there are 480 weeks for ten years' data from 1998 to 2007. For all four cities, the year 2007 has less animal-related outages than the other years, which could possibly be due to Westar Energy implementing some protection plan to lower the animal-related outages in 2007. For the high peaks in the time series, additional and more detailed information such as fault locations is needed to explain these unusual phenomena.

Even though the outage levels vary from year to year for all four cities, the pattern in one year is similar to the others for all the cities. In every year there are two high peaks around the 20th and 40th weeks which lies in the months classified as month type 3. Periodicity of the historical time series make it possible to predict future weekly animal-related outages by the historical outages [78]. However, the occurrences of animal-related outages are random in nature, which means even under the exact same weather condition and the same week the outage count is not fixed. The randomness results in the noisy data which can cause big challenges for prediction models to precisely follow the trends, especially the high peaks.

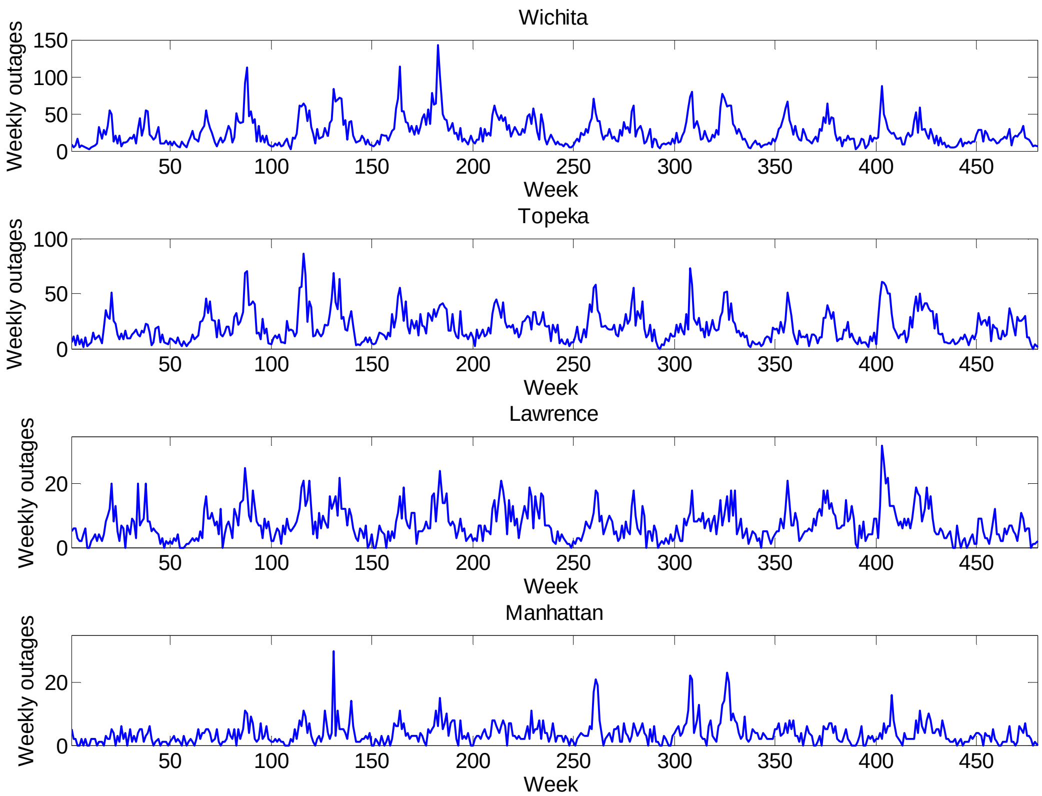

Figure 2.13 Historical Animal-related Outages of Four Cities from the Year 1998 to 2007

CHAPTER 3 - Poisson Regression Model

The idea of a Poisson regression model came from previous work of Zhou and Pahwa in which they used Poisson regression to analyze outages caused by winds and lightning [3]. Intuitively, the number of outages caused by animals that take place in the overhead distribution systems is a counting process in which usually there are less than two outages in a day in a certain distribution system. In probability theory and statistics, the Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time, provided these events occur with a known average rate and independently of the time since the last event [79]. A Poisson distribution is best suited for the cases in which there are large numbers of occurrences of values at the lower levels. Therefore, it is reasonable to assume that the number of animal-caused outages in a day follows a Poisson distribution.

3.1 Introduction

3.1.1 Poisson Distribution

Poisson distributions are suitable for random events that take place during a time-interval of given length. If the expected number of occurrences in this interval is λ, then the probability that there are exactly *k* occurrences (*k* being a non-negative integer) is equal to:

$$f(k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad (3.1)$$

where,

- $k = 0, 1, 2\ldots$,which is the number of occurrences of an event
- $f(k)$ denotes the probability that the outcome is k
- λ is a positive real number and the expected number of occurrences that happen during the given interval.
- $k!$ is the factorial of k
- e is the natural logarithmarithm base

The mean and variance of the Poisson probability distribution are both λ .

3.1.2 Additive Property

According to the additive property of Poisson distributions, if $X_i \sim Poisson(\lambda_i)$ follows a Poisson distribution with mean λ_i and the X_is are independent, then $Y = \sum_{i=1}^{N} X_i \sim Poisson(\sum_{i=1}^{N} \lambda_i)$ also follows a Poisson distribution with mean $\sum_{i=1}^{N} \lambda_i$ [80].

In our case, the number of animal-related outages in a certain overhead distribution system in a day is X_i, which follows a Poisson distribution with mean of λ_i. Furthermore the occurrence of animal-related outages in one day makes it neither more nor less probable than the animal-related outages occurring on the other days. In other words, X_i is independent. The summation of the number of animal-related outages in one week, Y, follow a Poisson distribution also because of the additive property of independent Poisson distributions. And the expected number of weekly animal-related outages equals the summation of the expected number of daily animal-related outages in that week.

3.1.3 Poisson Regression Model

In statistics, Poisson regression is a form of regression analysis used to model counting data. Poisson regression assumes the response variable Y has a Poisson distribution, and assumes its expected value λ is determined by a function of a linear combination of explanatory variables, $X_1, X_2, \ldots X_{p-1}$. This function is called link function. In other words, the link function provides the relationships between the mean response of a Poisson distribution and the explanatory variables $X_1, X_2, \ldots X_{p-1}$. It can be expressed by the following equation:

$$\lambda = \lambda(\mathbf{X}_i, \boldsymbol{\beta}) \tag{3.2}$$

where, $\boldsymbol{\beta}$ is the linear term of the explanatory variables.

There are many commonly used link functions such as the following [81]:

$$\lambda = \lambda(\mathbf{X}_i, \boldsymbol{\beta}) = \mathbf{X}_i^T \boldsymbol{\beta} \tag{3.3}$$

$$\lambda = \lambda(\mathbf{X}_i, \boldsymbol{\beta}) = e^{\mathbf{X}_i^T \boldsymbol{\beta}} \tag{3.4}$$

$$\lambda = \lambda(\mathbf{X}_i, \boldsymbol{\beta}) = \ln(\mathbf{X}_i^T \boldsymbol{\beta}) \tag{3.5}$$

In fact, any computation form can be set as the link function although the natural logarithm is the most popular one. The link function that fits a data set the best may vary according to the specific data set. It requires experiments to find out the best fit link function for different data sets. Once the link function is determined, the linear term $\boldsymbol{\beta}$ can be estimated by maximum likelihood, which ensures to gain the best model performance for a certain data set.

3.2 Model Construction

The counting nature of the response variable, occurrences of animal-related outages, leads to the choice of the Poisson regression model. It assumes that the numbers of outages follow independent Poisson distribution with outage rates, $E(Y_i) = \lambda_i$. The two factors, which are the number of fair days per week (*fd*) and the month type (*mt*), influence outage rates λ_i via the model:

$$\lambda_i = \lambda(fd_i, mt_i, \boldsymbol{\beta}) \tag{3.7}$$

The natural logarithm is found to be the best fit in our study. Experiments were carried out with outages data for Wichita from 1998 to 2007. Wichita was chosen because it is the biggest city in Kansas and gives more values of the observations of the occurrences of animal-related outages. Note that any zero outage will cause infinite natural logarithm and thus requires data examination before the model construction. If there are any zero outages in a data set, we need a link function other than natural logarithm. The size of the distribution system in Wichita is big enough not to have any zero weekly animal-related outages. Therefore the function (3.6), $\lambda_i = \lambda(\mathbf{X}_i, \boldsymbol{\beta}) = \log(\mathbf{X}_i^T \boldsymbol{\beta})$, is chosen to be the form of the Poisson regression model in our study.

Explanatory variables in the linear term $\mathbf{X}_i^T \boldsymbol{\beta}$ are selected by all-possible regression procedures. The pool of candidate variables consists of:

- Number of fair days per week
- Square root of the number of fair days per week
- Natural logarithm of the number of fair days per week

- Month type
- Square root of the month type
- Natural logarithm of the month type
- Interactions between the number of fair days and the month type, represented by multiplication of any two variables listed above.

Considering that for some cases the number of fair days per week is zero, which causes infinite natural logarithm of the number of fair days, the natural logarithm of the number of fair days per week is dropped. For the month type, there is no concern of this kind. Again, the linear term which gives the best performance is obtained by experimental results, which consists of:

- Square root of the number of fair days per week
- Month type
- Number of fair days per week × Natural logarithm of the month type

The final Poisson regression model for Wichita is obtained by least-squares regression method:

$$\lambda_i = \exp(1.55 + 0.061 \times \mathrm{X}_{1\mathrm{i}} + 0.596 \times \mathrm{X}_{2\mathrm{i}} + 0.02 \times X_{3i}) \tag{3.8}$$

where,

- $i = 1, 2, ..., n$
- X_{1i} = square root of the number of fair days per week
- X_{2i} = natural logarithm of the month type
- X_{3i} = square root of the number of fair days per week× natural logarithm of the month type

Similarly, individual models for Topeka, Lawrence and Manhattan are obtained experimentally too. The link function is $\lambda_i = \lambda(\mathbf{X}_i, \boldsymbol{\beta}) = \mathbf{X}_i^T \boldsymbol{\beta}$, which is linear function instead of natural logarithm, since for some weeks there are zero outages in these three cities. Explanatory variables in the linear term $\mathbf{X}_i^T \boldsymbol{\beta}$ for these three cities are the same as the ones for Wichita, except the coefficients vary according to different cities.

The final Poisson regression model obtained for Topeka is:

$$\lambda_i = 6.82 + 2.02 \times X_{1i} + 0.853 \times X_{2i} + 0.877 \times X_{3i}) \quad (3.9)$$

The final Poisson regression model obtained for Lawrence is:

$$\lambda_i = -1.35 + 0.285 \times X_{1i} + 3.90 \times X_{2i} - 0.0875 \times X_{3i}) \quad (3.10)$$

The final Poisson regression model obtained for Manhattan is:

$$\lambda_i = -0.09 - 0.287 \times X_{1i} + 1.687 \times X_{2i} + 0.24 \times X_{3i}) \quad (3.11)$$

3.3 Simulations and Model Performance

Time sequences of the weekly observations and predictions for Wichita, Topeka, Lawrence and Manhattan by the Poisson models are shown in Figure 3.1-3.4. The Poisson models are able to track fluctuations in the weekly number of outages from 1998 to 2007. However, from the plots we can clearly see that the predictions have a big challenge of catching the high peaks in the time sequences for all four cities. The reason behind this very possibly is the simplicity of linear relations in function (3.8), (3.9), (3.10) and (3.11). We need a better function with higher complexity to approximate the relations between input factors and output outages, which leads us to neural networks.

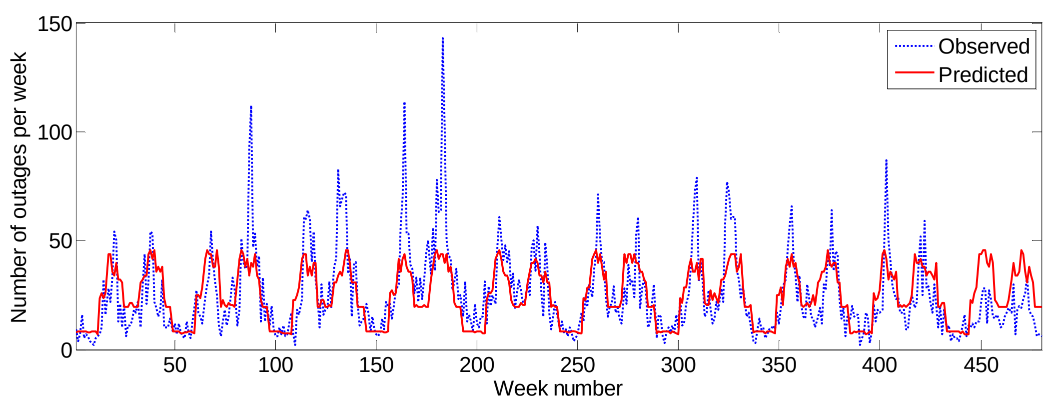

Figure 3.1 Time Sequences of Observed and Predicted Weekly Animal-related Outages in Wichita from the Year 1998 to 2007

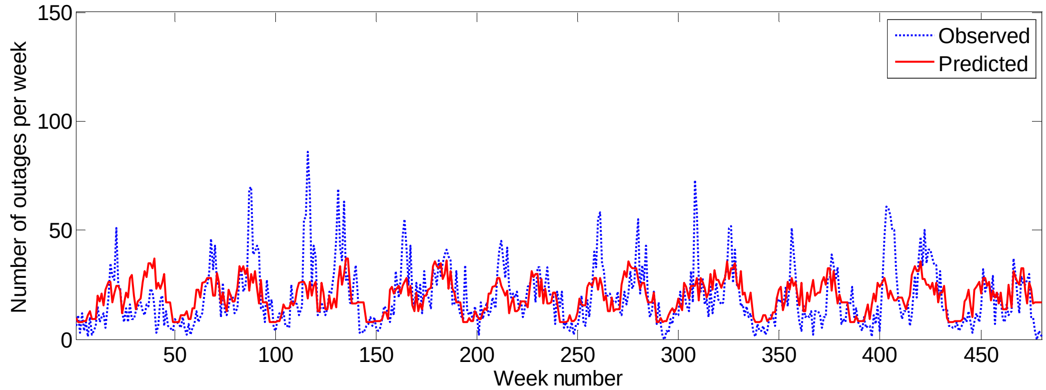

Figure 3.2 Time Sequences of Observed and Predicted Weekly Animal-related Outages in Topeka from the Year 1998 to 2007

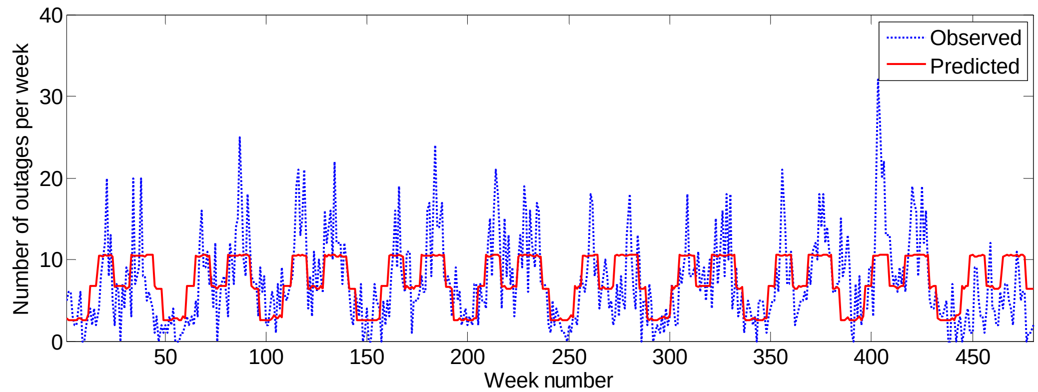

Figure 3.3 Time Sequences of Observed and Predicted Weekly Animal-related Outages in Lawrence from the Year 1998 to 2007

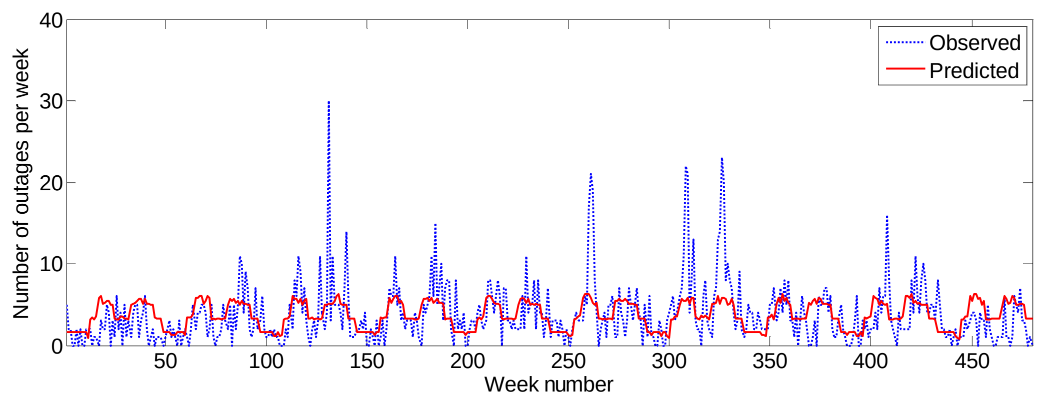

Figure 3.4 Time Sequences of Observed and Predicted Weekly Animal-related Outages in Manhattan from the Year 1998 to 2007

CHAPTER 4 - Neural Network Model

Because of the randomness and fluctuations in animal-related outages from one week to the next, conventional methods have limited abilities to capture the nonlinearities in the time series of animal-related outages. In Chapter 3 results based on a Poisson regression model showed that predicted values can track the basic patterns in the time series of observed values but are not able to follow the sharp peaks in the time series, which prompts the demand for nonlinear models for our predictions. Artificial neural networks based methods have gained wide attention in engineering and are widely used for time series forecast because of their ease of use and their ability to approximate high complexity functions.

4.1 Neural Network

An artificial neural network is a powerful approach which is able to capture complex nonlinear relationships between inputs and outputs. Artificial neural networks are modeled loosely after neural processes in the human brains. The human brain has about 10^{10} neurons which communicate through a network of axons and synapses [84]. It is a complex system which is capable of receiving information, computing, learning, remembering and reasoning. By mimicking the brains, artificial neural networks acquire knowledge by learning from data and storing it within the connections between neurons. The strengths of connections are called weights. Neural networks have advantage over the traditional linear models because they are able to represent both linear and non-linear relationships and they can learn these relationships directly from the data being modeled.

4.1.1 Mathematical Model

A neural network (NN) is a set of processing units and connections with weights which can be adjusted during the learning process [84]. Usually a NN has a multi-layer structure with one input layer, one or more hidden layers and one output layer. Even though the learning process of human brains is complicated, the working principal for each unit is relatively simple [84]: receive input from feeding units or external sources, compute an output signal and send it to other units. During this process, the weights of the connections are adjusted. Since many units can process simultaneously, the neural network is a parallel computing system with multi-layer units. Even though every unit in each layer operates in the same way, there are three different kinds of units: input units, hidden units and output units. Input units only receive data from external sources and feed their outputs to hidden units. The hidden units obtain data from input units or other hidden units, compute the output and propagate the data forward to the units on the next layer. Output units receive data from hidden units or input units when there are only two layers in the neural network. Output units give the final output of the neural network [84].

While constructing a mathematical model of a biological neuron, usually each weight is defined as w_{ji}, which reflects the connection strength unit i has on unit j. w_{ji} can have negative or positive values. Negative values imply inhibitory connections between units i and j, while positive values are associated with excitatory connections. Within each unit a summer is present to add all the inputs, multiplied by their respective weights. This sum is denoted as s_j. Furthermore, there is an external bias b_j which is equivalent to a weight applied to a constant input with value of 1. This unit also includes a nonlinear activation function $F_j(\,.\,)$. Given an input s_j, the output is defined as:

$$s_j = \sum_{i=1}^{p} w_{ji} y_i + b_j \tag{4.1}$$

$$y_j = F_j(s_j) \tag{4.2}$$

Figure 4.1 [84] shows a schematic for such a unit.

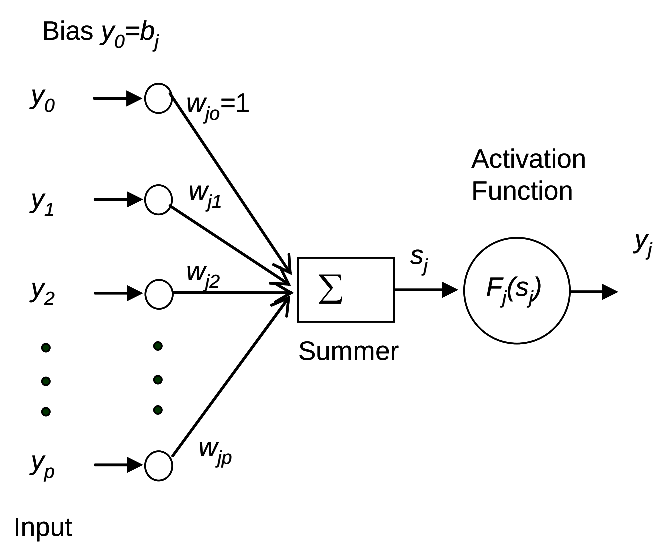

Figure 4.1 The Process of Computation in a Neuron

The output of the neuron, y_j, would therefore be the outcome of the activation function F_j on the value of s_j. Some functions are chosen as activation functions such that the output of a neuron in a neural network has certain range (usually 0 and 1, or -1 and 1). Three types of activation functions are commonly used. First, there is the discrete threshold function with a certain threshold. When the summed input s_j is below the threshold this function has value of 0 and when the input is equal or greater than the threshold this function has a value of 1. The second type is the Piecewise-Linear function which consists of several linear pieces. It has the values of 0 or 1, but can also take on values between 0 and 1 when the input falls into the linear pieces [84]. Thirdly, there is the sigmoid function which has an output range between 0 and 1:

$$F_j(s_j) = 1/(1 + e^{-s_j}) \tag{4.3}$$

4.1.3 Network Topologies

In this section we focus on the topology of the network and the propagation of data. Based on their topology, neural networks can be classified into two classes: feed-forward neural networks and recurrent neural networks. In a feed-forward neural network, neurons are arranged as layers, input data is applied to the input layer, which is then forwarded to the output units in a strictly feed-forward manner. There is no feedback connection from units in one layer to those in a previous layer or the same layer. On the other hand, when such feedback connections are present, the neural network is called recurrent neural network.

Multi-layer feed-forward neural networks are very popular in the applications for engineering problems. 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 [84]. Hence, in most applications, a three-layer feed-forward network is used, which consists of an input layer, a hidden layer and an output layer.

4.1.4 Training of Artificial Neural Networks

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 approximate the targets accurately. An adequate learning method is needed to obtain such a network.

Learning rules can be grouped into two distinct types [84]: supervised learning and unsupervised learning. In supervised learning the inputs and the corresponding outputs are provided to the network from outside. The supervised learning allows the network to adjust the weights based on the difference between network outputs and provided outputs. There are several popular supervised learning methods among which back-propagation is most commonly used. Unsupervised training is also called self-organized learning in which no matching outputs are provided and the output units have to make sense of the inputs on their own. Unsupervised learning is different from the supervised learning since there are no existing groups into which the inputs are to be classified and the network discovers the feature within the inputs [84]. Unsupervised learning is commonly used in data mining in which there are big sets of data and the features of data are not known.

4.1.4.1 The Back-propagation Algorithm

One supervised learning method, back-propagation, is routinely used in many applications of neural networks. Here, the weights of the connections are adjusted to minimize the error between the output of each output unit and a target output [84]. This process requires the computation of error derivative of the weights, which starts from the output layer and moves from layer to layer in a direction opposite to the propagation of data through the network. The name back-propagation comes from the fact that the error is propagated back to modify the incoming weights.

For a multilayer neural network, the mathematical steps of back-propagation algorithm are given in the following [84].

1. Definitions:

   The output of unit j: y_j

   The sum-squared loss function: $E = \frac{1}{2}\sum_o (t_o - y_o)^2$

   The weight of connection from units i to j: w_{ji}

   The summed input for unit j: $s_j = \sum w_{ji} y_i + b_j$

   In the above equation, the summation is carried out over all units in its previous layer.

The distributed error for unit *j*: $\delta_j = -\partial E / \partial s_j$

The gradient for weight w_{ji}: $\Delta w_{ji} = -\partial E / \partial w_{ji}$

The error for output unit o: $\delta_o = t_o - y_o$

2. Computation of gradient for weight-update:

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

$$\Delta w_{ij} = -\partial E / \partial w_{ij} = -\frac{\partial E}{\partial s_i} \frac{\partial s_i}{\partial w_{ij}} \qquad (4.4)$$

According the definitions, the first factor is the error of unit *i*. The second factor is actually the output of unit *j*:

$$\frac{\partial s_i}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} \sum w_{ik} y_k = y_j \qquad (4.5)$$

The summation is carried out over all units in the previous layer of unit *i*. Therefore the gradient of loss function with respect to weight w_{ij} is simply:

$$\Delta w_{ij} = \delta_i y_j \qquad (4.6)$$

From the above equation, we have to compute the output and the error for all units to obtain the increments of the weights.

3. Computation of output of each unit:

For units on the input layer, the output is simply the input received from outside sources. For the hidden layers and output layer, the output of unit *i* is determined by the summed input s_i and the activation function F_i:

$$y_i = F_i(s_i) = F_i(\sum w_{ij} y_j) \qquad (4.7)$$

In the above equation, the summation is carried out over all units in the previous layer of unit *i*. Note that the output of unit *i* cannot be computed unless the outputs of all the units in A_i have been calculated. This condition can be met in feed-forward networks since there are no cycles in the connections and thus the units can be numbered in an order from input layer to output.

4. Calculation of errors propagated to hidden units:

As we have mentioned, the error of output units is just the difference between outputs and the anticipated targets. For the hidden units there is no direct error without targets. However, the back-propagation algorithm allows us to propagate the error of output units back to hidden units. With the chain rule, the error of a hidden unit can be expressed as:

$$\delta_j = -\partial E / \partial s_j = -\sum \frac{\partial E}{\partial s_i} \frac{\partial s_i}{\partial y_j} \frac{\partial y_j}{\partial s_j} \qquad (4.8)$$

The summation is carried out over all units in the next layer of unit *j*. There are three factors in the formula. According to the definitions, the first factor is just the error of node *i.* The second is

$$\frac{\partial s_i}{\partial y_j} = \frac{\partial}{\partial y_j} \sum w_{ik} y_k = w_{ij} \qquad (4.9)$$

The summation is carried out over all units in the previous layer of unit *i*. And the third is the derivative of unit *j*'s activation function:

$$\frac{\partial y_j}{\partial s_j} = \frac{\partial f_j(s_j)}{\partial s_j} = f_j^{'}(s_j) \qquad (4.10)$$

Therefore, the error of unit *j* can be computed by this equation:

$$\delta_j = f_j'(s_j)\sum \delta_i w_{ij} \qquad (4.11)$$

In the above equation, the summation is carried out over all units in the next layer of unit *j*. In order to calculate the error for unit *j*, we must first know the error of all the units which get data from unit *j*. The property of non-cycle connections for feed-forward networks ensures this condition is fulfilled. Starting from known error of output units, the error for all the hidden units can be computed step by step. Note that error is not applicable to input units since there is no incoming weight for them.

5. Weights update:

At the end, the weights are updated using the gradient which has already been computed,

$$w_{ij} = w_{ij} + \Delta w_{ij} \qquad (4.12)$$

A well trained network can converge to a stable solution. Unfortunately divergence can occur during the learning procedure. To prevent divergence, a learning rate μ is introduced to the weight-update scheme. Learning rate is a constant value that is used to multiply the gradient [84]. By choosing different value for the learning rate, we can control the amount of increment for weight-update at each step. Divergence happens when μ is too big, and the algorithm will miss the optimal solution and oscillate. But if the learning rate is too small, the algorithm will be 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. Besides using learning rate, another way to avoid oscillation is to add a portion of past gradient $\Delta w_{ij}(t-1)$ in addition to the current gradient $\Delta w_{ij}(t)$ when updating weights:

$$\Delta w_{ij}(t) = \mu \delta_i y_j + \alpha \Delta w_{ij}(t-1) \qquad (4.13)$$

In this equation *t* represents the current training iteration and *t-1* represents the previous training iteration. μ is the learning rate and α is called momentum factor which determines the effect of the previous gradient.

The aim of training a neural network with training data is to learn the underlined relationships between input and output data and generalize this knowledge to new data. To let the network fully acquire the knowledge and let the back-propagation algorithm converge, we have to execute the training process with many iterations. However, overtraining will happen if the number of iteration is too large because the network will memorize every detail of the training set or even the error and noise and thus is not able to generalize to new data. In practical cases, it results in a network that has very small error on the training set but has large error on new test data. This problem can be overcome by using a permuted training method, which is introduced in Chapter 5.

4.2 Model Construction and Performance

4.2.1 Neural Network for Prediction

We have used the most common three-layer, feed forward neural network topology, which is able to adequately approximate nonlinear functions with sufficient accuracy [85]. 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 the month type are the two feature-related inputs to the neural network. Furthermore, the outages of previous weeks are taken as additional inputs since there are similar patterns observed in historical data and thus the NN can learn the patterns and predict future outages based on the learned patterns.

The procedure for designing neural network structures mainly involves selecting the number of neurons in the input, hidden and output layers. Preference is given to simple models in applications of neural networks. According to Ockham's razor principle, it's more possible that a simpler computing model has better generalization abilities [86]. With less weights in the network, there is greater probability that overtraining will not happen due to noises [86]. It is desirable to reduce the number of input nodes to an absolute minimum of essential nodes [87-89]. Following this line of thought, the optimum number of inputs for outage in previous weeks is obtained by experimentation, which indicates that outages in four previous weeks are sufficient to give enough historical information.

There is no theoretical guidance for choosing the number of neurons in the hidden layer. By experience from all the applications of NN, the preference goes to the structure in which there are fewer neurons in the hidden layer than neurons in the input layer. In most cases the number of hidden neurons is half of the total number of input neurons and output neurons. When applying neural networks to time series predictions, the number of output neurons is very important [90]. As suggested in [90], we minimize the number of targets required in our animal-related outage predictions. A single output neuron is the ideal case because the network is focused on one task and conflicting outputs on the output layer will not happen [86]. Once the number of units on the three layers are determined, the number of weights are determined too. The structure of the NN model for our predictions is shown in Figure 4.2 where,

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

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

c^{n-k}: the values of the time series c for k weeks before the forecasting week n, where k=1, 2, 3, 4;

c^n: the predicted output of the time series c of week n.

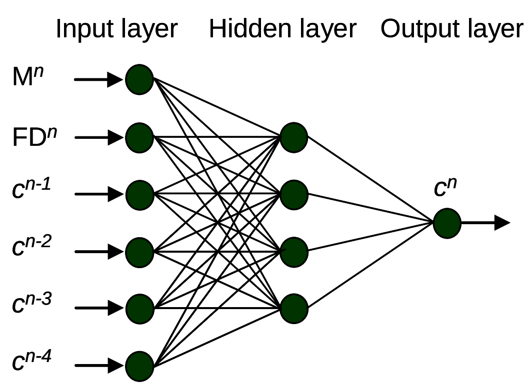

Figure 4.2 Three-layer Feed-forward NN Model

With a single output node, this model gives one step ahead prediction. The input and target data are normalized between 0.1-0.9. Note that the desired output must never be set to 0 or 1! The reason is simple: whatever the inputs are, the outputs of the nodes in the hidden layer are restricted to between 0 and 1 (these values are the asymptotes of the function). To approach these values would require enormous weights and/or input values, and most importantly, they cannot be exceeded. By contrast, setting a desired output of 0.9 allows the network to approach and ultimately reach this value from either side. Found by experimentation, the learning rate is 0.5, momentum is 0.2 and the optimum training times is 3000.

4.2.2 Simulations And Performance

The models were trained with historical data for the four cities from the year 1998 to 2006 and tested for 2007. The performances of the models are measured using the average absolute error (AAE), which are given in Table 4.1. Figure 4.3-4.6 show plots of weekly observed outages and outages predicted for Wichita, Topeka, Lawrence and Manhattan. It can be seen that the models can reproduce the basic patterns of the time series quite well both in training and testing durations. However, they still have the deficiency in catching the fast fluctuations in the time series, which are shown at the high peaks in the time series.

Plots of weekly observed outages and predicted outages for four cities for the testing year 2007 are shown in Figure 4.7-4.10. The slopes (S) of the best-fit line and correlation coefficients (R) between observed and predicted otuages for the year 2007 are also shown on these graphs. Values of 1 for both of S and R would indicate best fit around the ideal (y=x) line. We have observed increasing AAE with increasing city size, which doesn't necessarily mean the model works better for smaller cities. This is because the outages have a greater range in bigger cities. In fact, the correlation coefficient increases with the increase in the size of cities, which shows that the predictions can catch the patterns in the observed outages better for the bigger cities than the smaller cities. For all the cities, correlation coefficients are positive, which indicates that there is a positive linear relationship between the predicted outages and observed outages. For Wichita, Topeka and Lawrence, the slopes of best-fit line are slightly greater than 1, which implies overestimations in the predictions of 2007 for these three cities. Since the slope is close to 1 for Manhattan, there is no obvious overestimation or underestimations in the prediction of 2007 for Manhattan.

Table 4.1 Results of NN Models for Four Cities

NN Model	Training Error	Testing Error	Slope of Best-fit Line	Correlation Coefficient
Wichita	7.67	6.38	1.24	0.69
Topeka	6.29	6.94	1.21	0.76
Lawrence	2.74	3.38	1.19	0.36
Manhattan	1.93	2.09	0.97	0.29

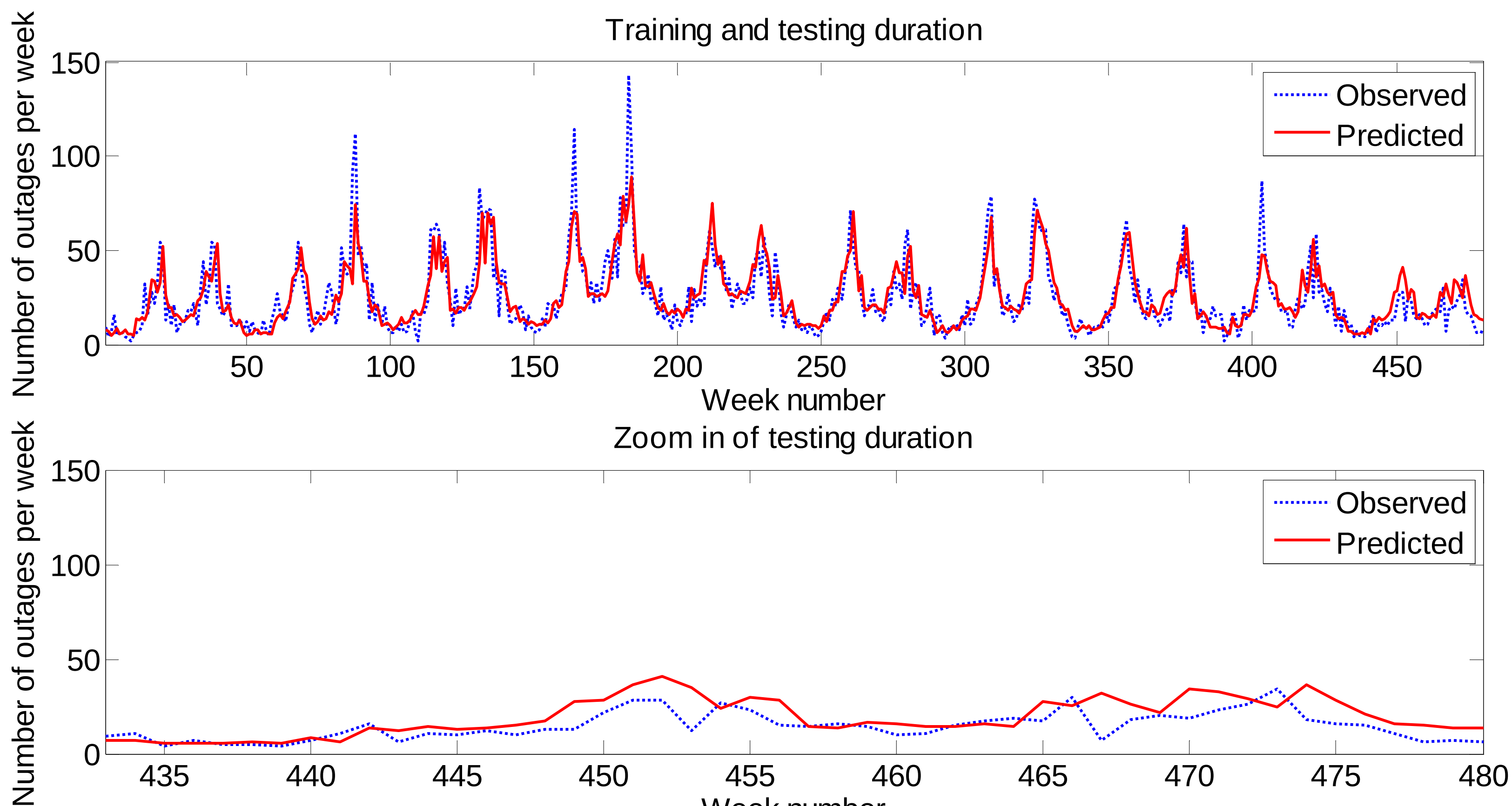

Figure 4.3 Outages Observed and Predicted by NN Model for Wichita from 1998 to 2007

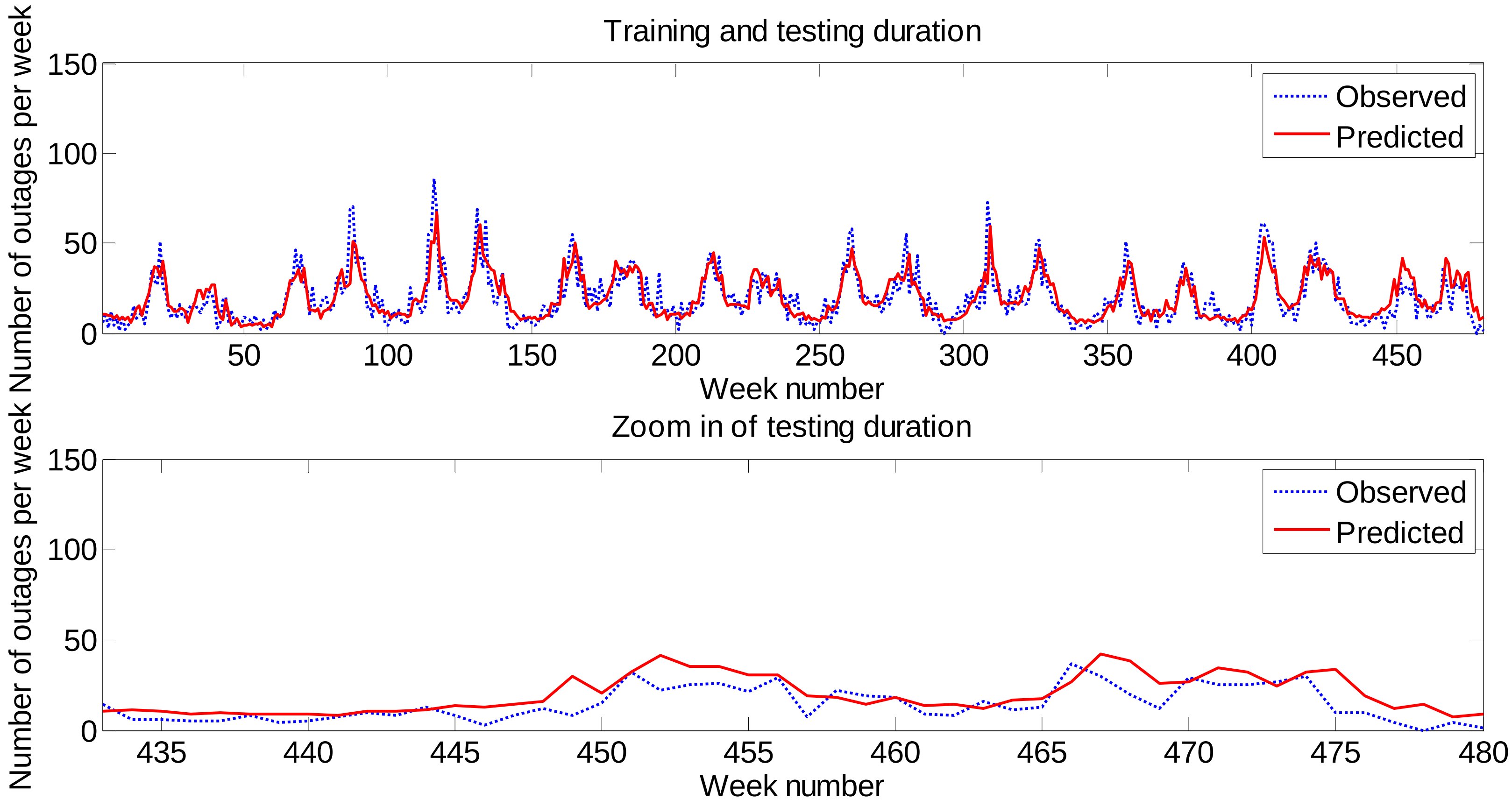

Figure 4.4 Outages Observed and Predicted by NN Model for Topeka from 1998 to 2007

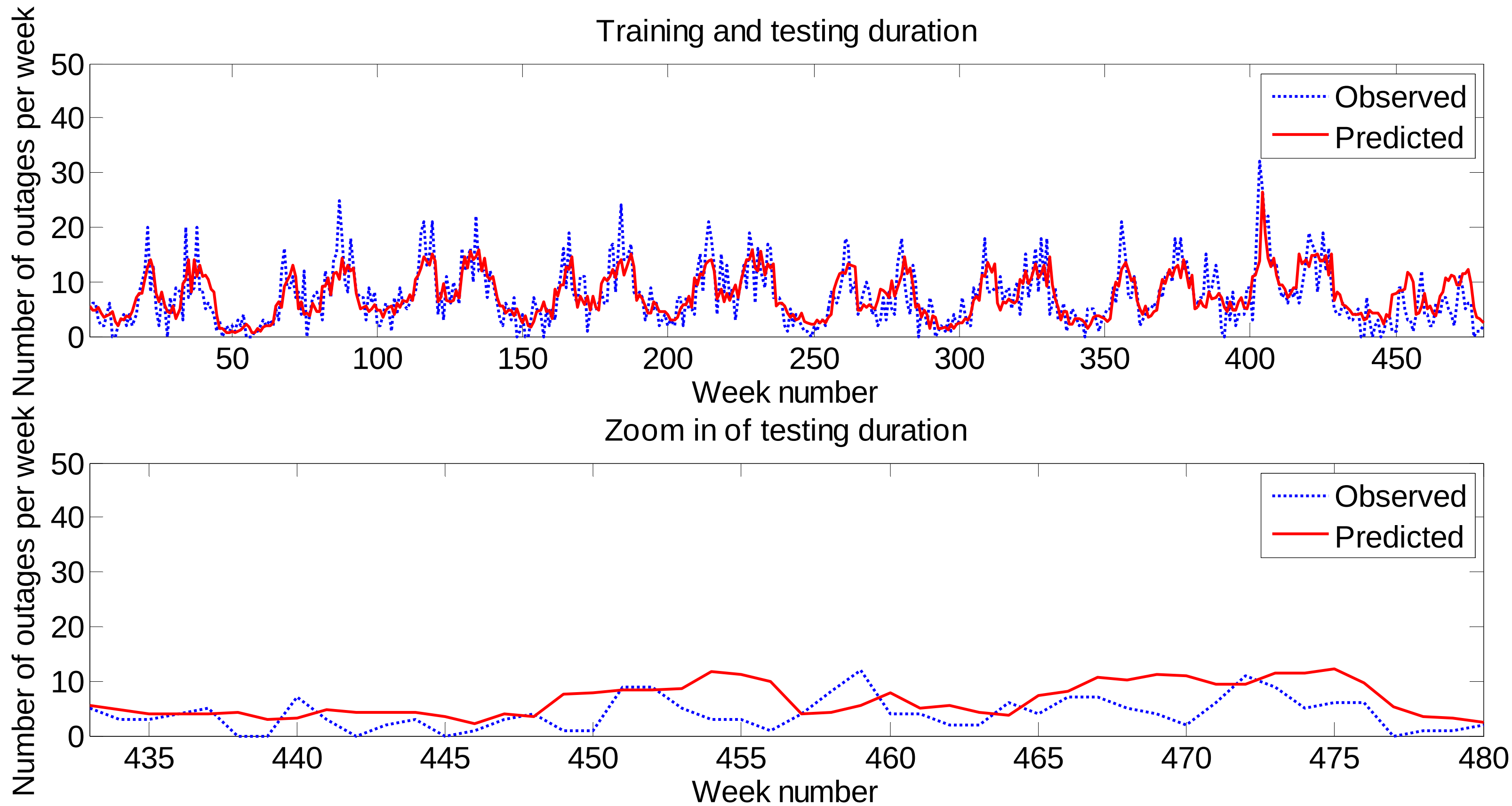

Figure 4.5 Outages Observed and Predicted by NN Model for Lawrence from 1998 to 2007

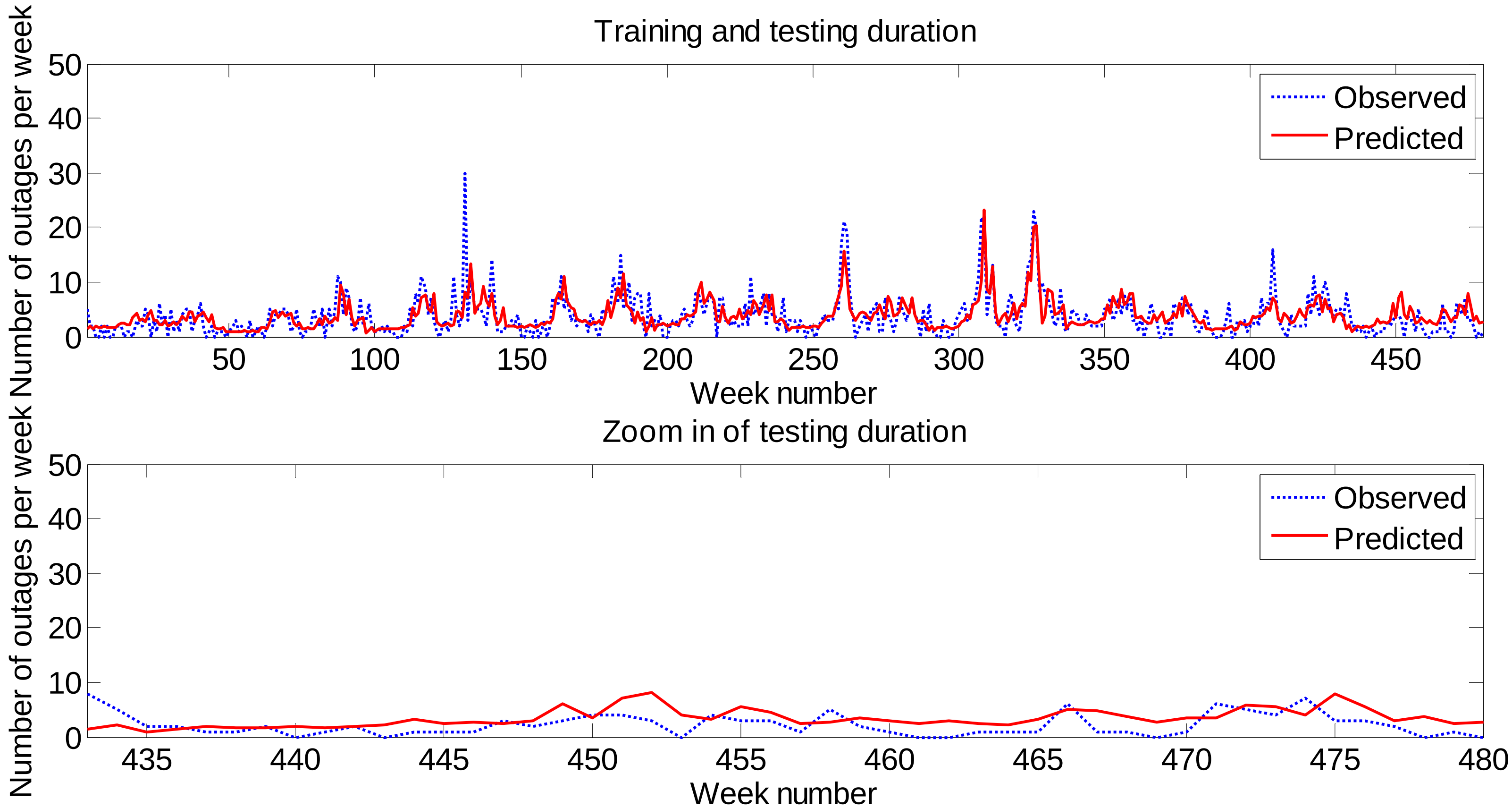

Figure 4.6 Outages Observed and Predicted by NN Model for Manhattan from 1998 to 2007

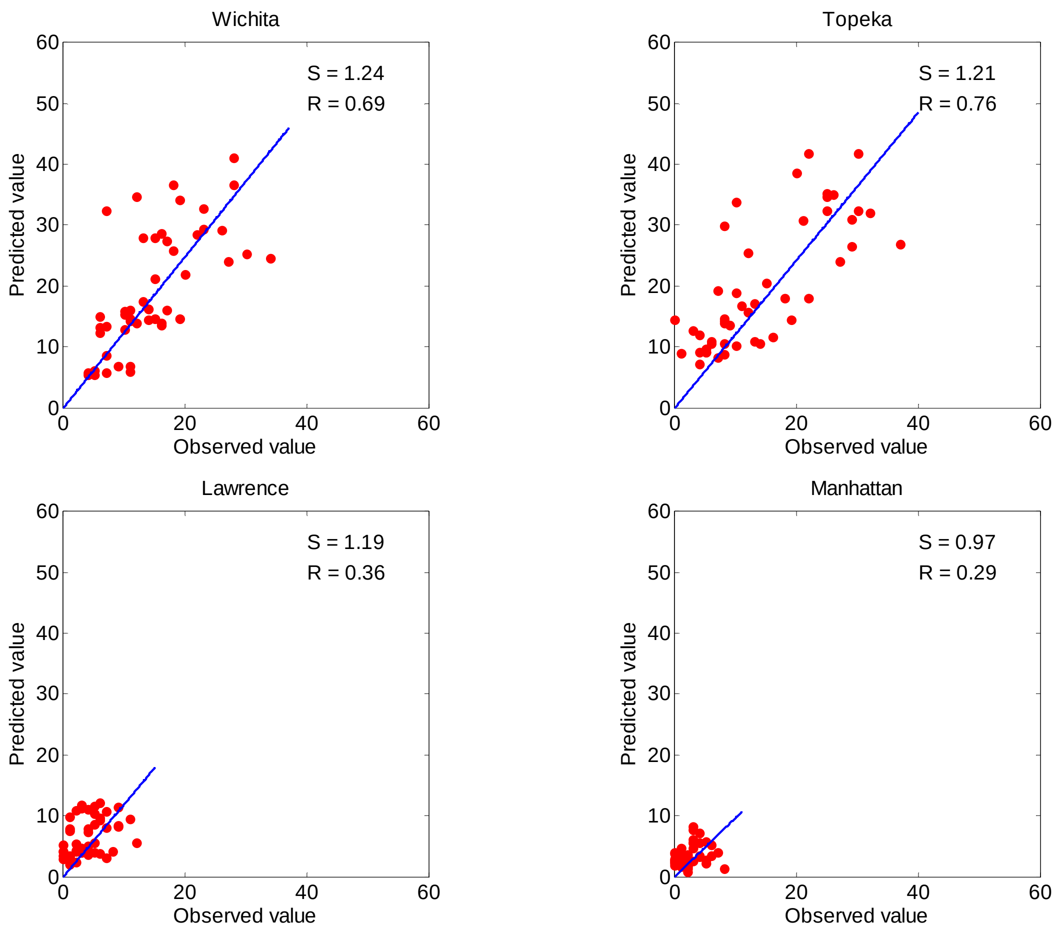

Figure 4.7 Scatter Plot of Observed and Predicted Outages Obtained with NN Model for Four Cities in the Year 2007

4.2 Conclusions for Neural Network Model

The NN model is able to approximate the complex relations between the inputs and outputs, which makes it outperform the traditional Poisson regression. However, the animal-related outages data are noisy and may deteriorate the out-of-sample performance of the neural network. Generally, an appropriate data preprocessing stage could lead to better models for representing the true features of the underlying systems.

CHAPTER 5 - Wavelet Based Neural Network Model

The seasonal and fluctuating time series of weekly animal-related outages do not have smooth characteristics that could be apparent in a larger sample. The data may also contain noise, which obscures the characteristics. The idea of wavelet transforms comes from [86] and [91] where wavelet based neural networks were constructed for load forecasting. The wavelet transforms are one technique which can remove (or partially remove) noise and can approximate the characteristics of the time series. “The wavelet technique can decompose time series into low frequency and high frequency sub-series” [92]. The low frequency sub-series (approximation coefficient series) are denoised and thus can approximate the smooth characteristics of original time series. The high frequency sub-series (detail coefficient series) contain the detailed information. In other words, based on wavelet decomposition, we can carry out multi-resolution analysis on the data. With different scales, wavelet technique can split the original data into different frequency bands. A smaller scale will allows us to analyze the gross feature of the data and a greater scale gives us the insight into the small details in the data. The features of the original data series are better presented in the decomposed series, which makes neural network prediction more accurate on each decomposed level rather than on the original series alone. Hence, decomposing the time series into sub-series and constructing different neural network models for each sub-series according to its characteristic is a method to overcome the problems associated with directly applying neural network models to the original time series.

5.1 Wavelet transform Technique

5.1.1 Wavelet Transform

Wavelets theory and its applications are popular in signal analysis and several other engineering fields. As a powerful signal analysis tool, the wavelet transform is usually compared to the Fourier transform, the traditional method for signal processing. A Fourier transform translates a time domain signal into frequency domain signal by expanding it as sine and cosine coefficients at different frequencies. Thus, the Fourier transform provides the analysis in frequency domain for signal in time domain. However, when the signal has sharp spikes, the Fourier transform has poor ability to approximate the spikes since the sine and cosine functions are not restricted to finite domains but instead are expanded along the whole time domain. To investigate the spikes in a short time duration, a more suitable technique such as wavelet transform is introduced [93]. Wavelet transform uses a set of functions to represent signals. These functions are called wavelet and scaling basis which are orthogonal [92]. The objective of wavelet transform is to develop representations of a signal *f(t)* in terms of wavelet and scaling functions. The approximation coefficient series can be computed by taking the inner products of the function *f(t)* with the scaling basis:

$$c_{a,b} = \langle f(t), \phi_{a,b}(t) \rangle = \int_{-\infty}^{\infty} f(t)\overline{\phi}_{a,b}(t)dt \qquad (5.1)$$

The wavelet coefficients (details) can be computed by taking the inner products of the function *f(t)* with the wavelet basis:

$$w_{a,b} = \langle f(t), \psi_{a,b}(t) \rangle = \int_{-\infty}^{\infty} f(t)\overline{\psi}_{a,b}(t)dt \qquad (5.2)$$

where scale function $\varphi_{a,b}(t)$ and wavelet function $\psi_{a,b}(t)$ are obtained by dilating and scaling particular mother functions with scaling parameter *a* and translation parameter *b*.

$$\begin{aligned} \psi_{a,b}(t) &= \frac{1}{\sqrt{a}} \times \psi\left(\frac{t-b}{a}\right) \\ \phi_{a,b}(t) &= \frac{1}{\sqrt{a}} \times \phi\left(\frac{t-b}{a}\right) \end{aligned} \qquad (5.3)$$

where $1/\sqrt{a}$ is an energy normalization factor that keeps the energy of the scaled functions the same as the mother function. To invert transform and gain full information of the original signal, *ψ(t)* much satisfy the admissibility condition which is expressed as follow:

$$C_{\psi} = \int_{-\infty}^{+\infty} \frac{|\psi(\omega)|^2}{|\omega|} d\omega < \infty \qquad (5.4)$$

In practical cases, equation 5.4 means that the average value of the wavelet in the time domain must be 0:

$$\psi(0) = \int_{-\infty}^{+\infty} \psi(t)dt = 0 \qquad (5.5)$$

where *ψ(ω)* is the Fourier Transform of *ψ(t)*. From equation 5.5 we can see that *ψ(t)* has to oscillate in the time domain like a wave.

5.1.2 Discrete Wavelet Transform

The continuous wavelet transform becomes discrete wavelet transform (DWT) when choosing a and b based on powers of two. The DWT has the same accuracy and gains more efficiency than choosing other values for a and b [86]. In this case, the scaling and translation parameters a and b are integers: $a = a_0^{-j}$ and $b = kb_0 a_0^{-j}$. With $a_0 = 2$ and $b_0 = 1$, which are commonly used values, the formula (5.3) becomes:

$$\begin{aligned} \phi_{j,k}(t) &= 2^{j/2}\phi(2^j t - k) \\ \psi_{j,k}(t) &= 2^{j/2}\psi(2^j t - k) \end{aligned} \qquad j,k \in Z \quad (5.6)$$

where $\psi_{j,k}$ is called the discrete binary wavelet function and j is the resolution level [94].

The most famous algorithm for implementing DWT is the fast pyramid algorithm proposed by Mallat[95]. In this algorithm, the signal is fed into filter pairs with down-sampling by a factor of 2 after each filter pairs, as shown in Figure 5.1. The length of the signal has to be $N=2^k$ for the down-sampling and totally N wavelet and approximate coefficients are computed. The filter coefficients are restrictively constructed based on the orthogonal wavelet $\psi_{a,b}(t)$ and scaling basis $\varphi_{a,b}(t)$.

The wavelet transform can be applied in matrix form:

$$w = M \times c_0 \qquad (5.7)$$

where M is the $N \times N$ matrix consists of the wavelet and scaling basis at all resolution levels. Given J as the highest resolution level, resolution level j = 1,2,..J, we can have a $\frac{N}{2^j} \times N$ matrix M_j which is comprised of the wavelet basis at resolution level j. Similarly the scaling basis at resolution level j forms a $\frac{N}{2^j} \times N$ matrix V_j. Matrix M is:

$$M = [\ M_1 \quad M_2 \quad \ldots \quad M_J \quad V_J\ ]^T \qquad (5.8)$$

As a result the wavelet coefficients w with length of N can be written as J+1 vectors:

$$w = [\ w_1 \quad w_2 \quad \ldots \quad w_J \quad c_J\ ]^T \qquad (5.9)$$

where w_j is the wavelet coefficients at resolution level j and c_J is the scaling (also called approximate) coefficients at the highest resolution level J. Since M is orthogonal, the inverse exists and the original signal can be obtained by

$$c_0 = M^{-1} \times w \tag{5.10}$$

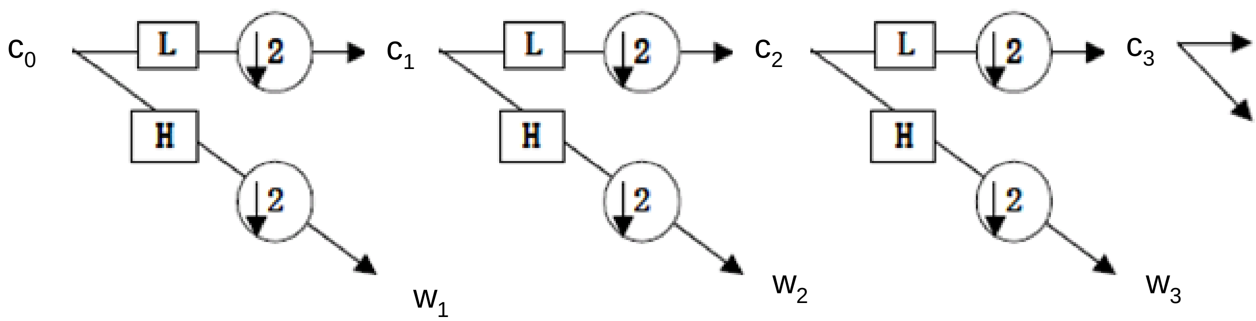

Figure 5.1 The Filtering Process for Fast Pyramid Algorithm [95]

This pyramid algorithm is called decimated DWT because the down-sampling discards many wavelet coefficients. It's also called non-redundant DWT since it only keeps the least wavelet coefficients needed to reconstruct the original signal without loss of information. This technique has remarkably contributed to data compression area. However, when DWT is applied to multi-resolution time series analysis, it encounters certain problems. First, the original signal at a certain time cannot be exactly located in the wavelet coefficients when the wavelet coefficients are shorter than the original signal. Second, DWT is not time-invariant and lacks stability during the transform [96]. Since the reconstruction is highly dependent on every wavelet coefficient, a small error in the prediction of wavelet coefficients can lead to a large error in the reconstruction. Third, the length of signal has to be restricted to powers of 2. These problems can be solved by employing a non-decimated or redundant wavelet transform. A non-decimated wavelet transform is better suited in predictions of time series since it can represent the original signal at a certain time in coefficients precisely and uniquely. Also, it can handle data of arbitrary length without causing any non-stability in the transform. The à trous algorithm [97] can be used to achieve such a redundant transform.

5.1.3 The à trous Algorithm

The basic idea behind this algorithm is similar to the classical non-redundant DWT except there is no down-sampling step. Down-sampling works as a data collecting tool to save storage space without loss of information for reconstruction. This is not applicable to predictions of time series in which the accuracy is the first priority and more information is needed to ensure stable performance. The only cost for discarding the down-sampling is a larger storage space. Using the non-redundant DWT, N wavelet coefficients can be obtained from N original signal. Unlike the classical DWT, in which the coefficients are shortened by half every time when resolution level increases by one, the redundant DWT has N coefficients at each resolution level. Hence, there are totally $(J+1)\times N$ coefficients when the redundant transformation is performed on J resolution levels. Therefore, information at each resolution level is directly related at each time point.

In order to apply the à trous algorithm, the data series $c_0(k)$ is passed through a low pass filter h_l and the output is the approximate coefficients $c_1(k)$ at the first resolution level. Subsequently, $c_n(k)$ is obtained when the time series goes through the filter n times. The decomposition process is given in Figure 5.2 and it is achieved using the following equation [97]:

$$c_j(k) = \sum_{l=0}^{L-1} h_l c_{j-1}(k + 2^{j-1} l) \quad (5.11)$$

Instead of using a high pass filter, the algorithm takes the difference between $c_j(k)$ and $c_{j-1}(k)$ as the wavelet coefficients, or the detail signal at level j, which can be expressed as:

$$w_j(k) = c_{j-1}(k) - c_j(k) \quad (5.12)$$

The reconstruction is simply to add up all the coefficients [97]:

$$c_0(k) = c_n + \sum_{j=1}^{n} w_j(k) \quad (5.13)$$

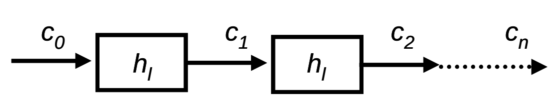

Figure 5.2 The Filtering Process of Obtaining Approximate Coefficients in à trous Algorithm

Our choice for low pass filter is the *Daubechie* 4 wavelet with length, *L,* of 4 . The coefficients are:

$$h_0 = \frac{1-\sqrt{3}}{4\sqrt{2}}, h_1 = \frac{-3+\sqrt{3}}{4\sqrt{2}}, h_2 = \frac{3+\sqrt{3}}{4\sqrt{2}}, h_3 = \frac{-1-\sqrt{3}}{4\sqrt{2}}$$

From equation (5.11), we can see there is a boundary problem for non-decimated DWT. When calculating $c_0(t)$ for time series up to time N, $c_0(N+1)$ and other farther samples are needed. There are several methods to handle this problem. We use the repetition treatment, where we add a copy of $c_0(t)$ at the end of the original $c_0(t)$.

5.2 Model Construction of WNN

In this proposed model, the input time series of historical data is decomposed into various sub-series and different neural network models for these sub-series are constructed. The outputs of the NN models are summed to find the final prediction for the original time series.

Wavelet decomposition serves as a data pre- and post-processing tool for the NNs to capture the patterns in the processed input data with different resolutions. The most suitable resolution level with the best performance in prediction is a matter for experimentation. Intuitively, with higher resolution level, the approximation coefficients are smoother and represent the underlying distribution of the original time series better. However, it doesn't ensure the overall model gives better predictions for the original time series since every further decomposition introduces one more detail coefficients into the whole model and thus introduces one more prediction error to the final output. To find a suitable resolution level, we carry out experiments according to two different schemes:

- Scheme 1: decompose the original time series into sub-series till it gives sufficiently smooth approximation coefficients;
- Scheme 2: only decompose the original time series with limited resolution level.

In each scheme, different resolution levels are tested and the AAE are computed to find out the optimum solution. In scheme 1, resolution level three provides smooth approximation coefficients c_3, third detail coefficients w_3, second detail coefficients w_2 and first detail coefficients w_1. Experiments showed that the model could show better performance if we further decompose the first and second detail coefficients, w_1 and w_2. Further decomposition of w_1 and w_2 resulted in ww_1, aw_1, ww_2 and aw_2. Totally we obtain six final sub-series c_3, w_3, aw_2, ww_2, aw_1 *and* ww_1 for the original time series of outages in Wichita as shown in Figure 5.3. For scheme 2, the simplest model with resolution level 1, sub-series c_1 and w_1 obtained are shown in Figure 5.4. The model with resolution level three is called WNN model, and the one with resolution level 1 is called MWNN model.

After preprocessing with wavelet decomposition, the sub-series of the historical outages, the month type and the number of fair days per week are fed into different NNs. The NNs have the same structure as in Figure 4.2 in Chapter 4. In the post processing stage, all the outputs from all the separate NNs are summed up to produce the final prediction for the original time series. The structure of the proposed wavelet based NN model with different resolution levels are shown in Figure 5.5 and 5.7. Note that there is a sub model for w_1 and w_2 in Figure 5.5, which is shown in Figure 5.6.

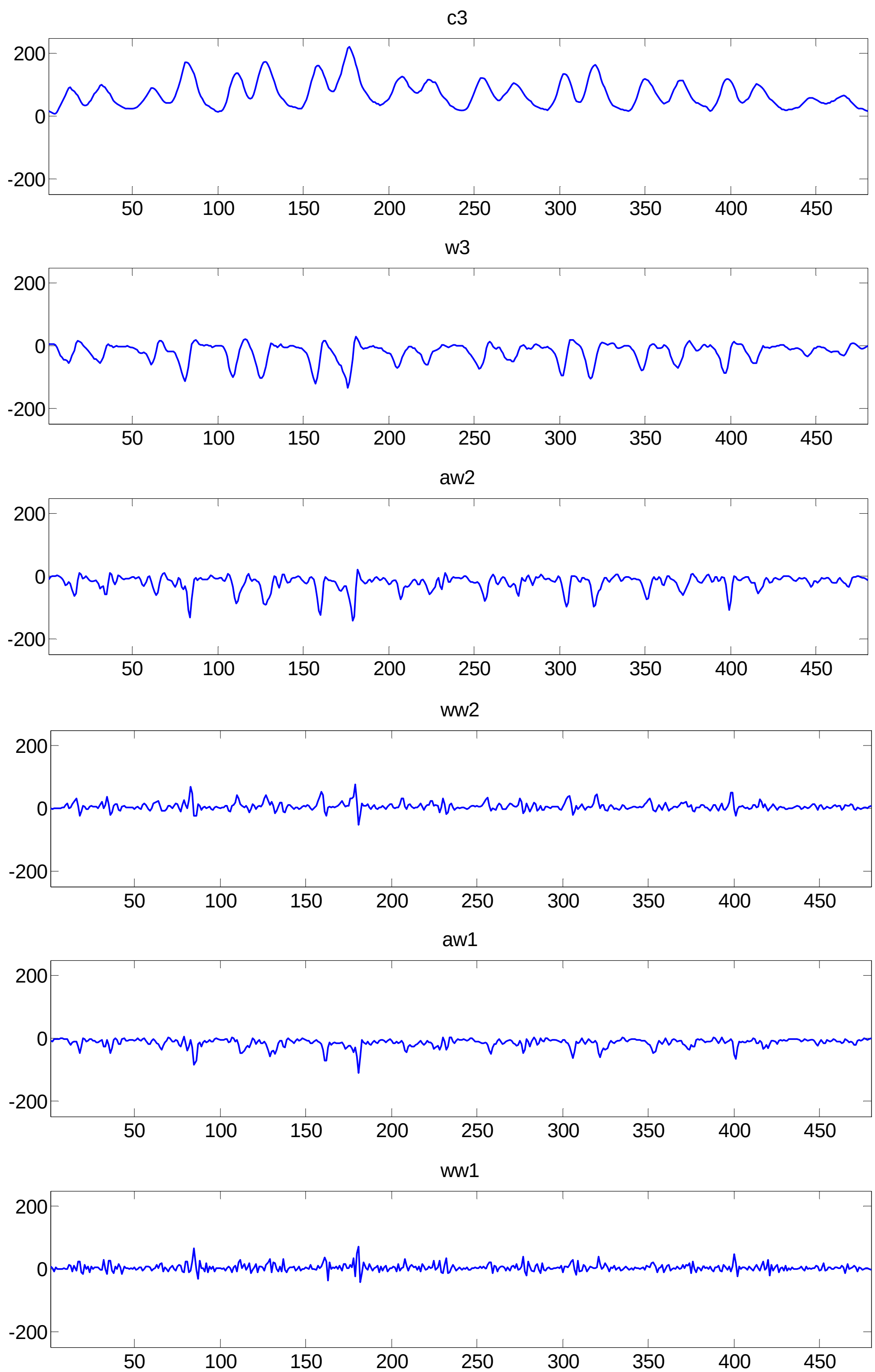

Figure 5.3 The Six Sub-series Obtained with Resolution Level 3 Wavelet Decomposition for Weekly Animal-related Outages in Wichita from the Year 1998 to 2007

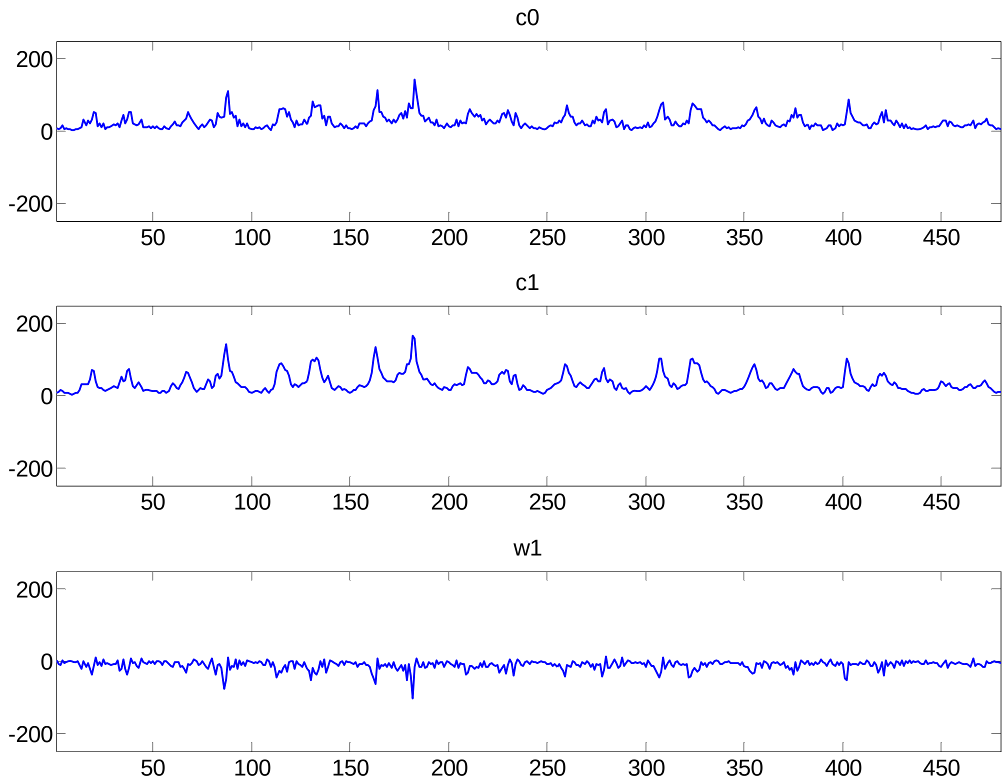

Figure 5.4 The Original Time Series and Sub-series with Resolution Level 1Wavelet Decomposition for Weekly Animal-caused Outages in Wichita from the Year 1998 to 2007

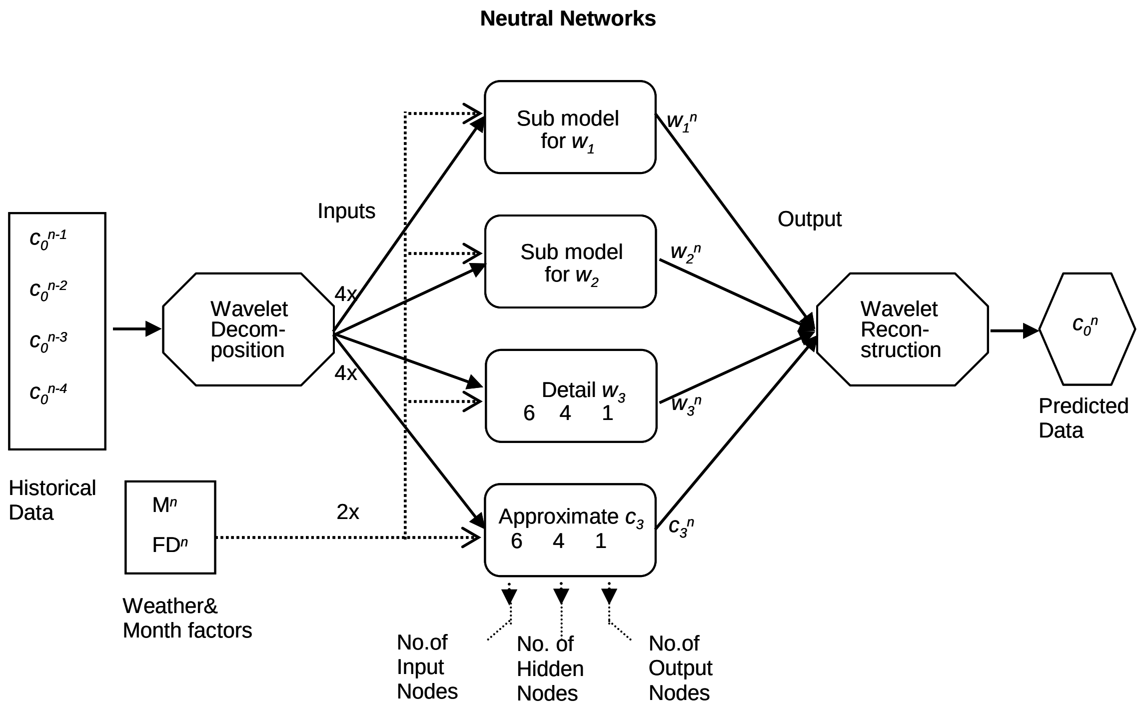

Figure 5.5 The Structure of Wavelet Based NN Model with Resolution Level 3

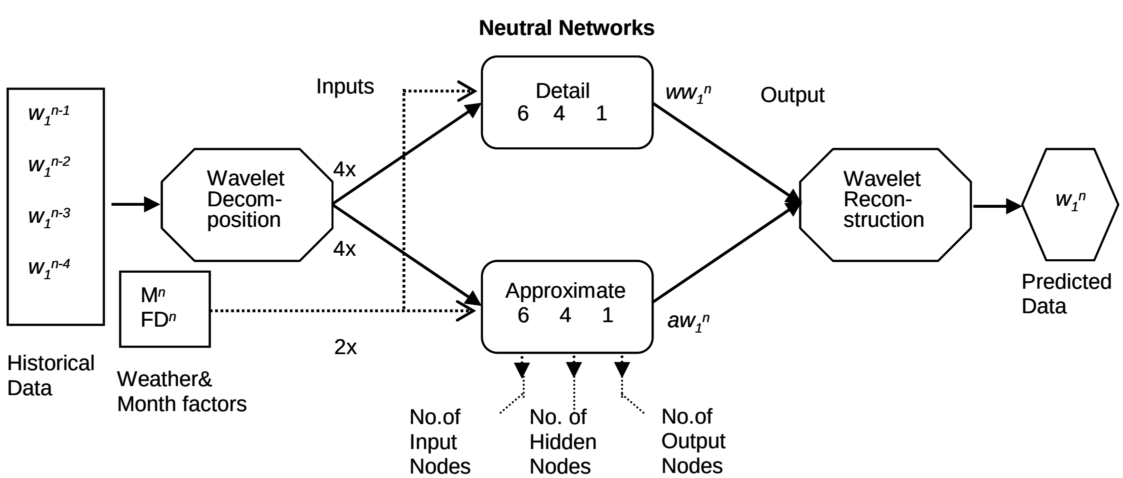

Figure 5.6 Sub Model for ww_1

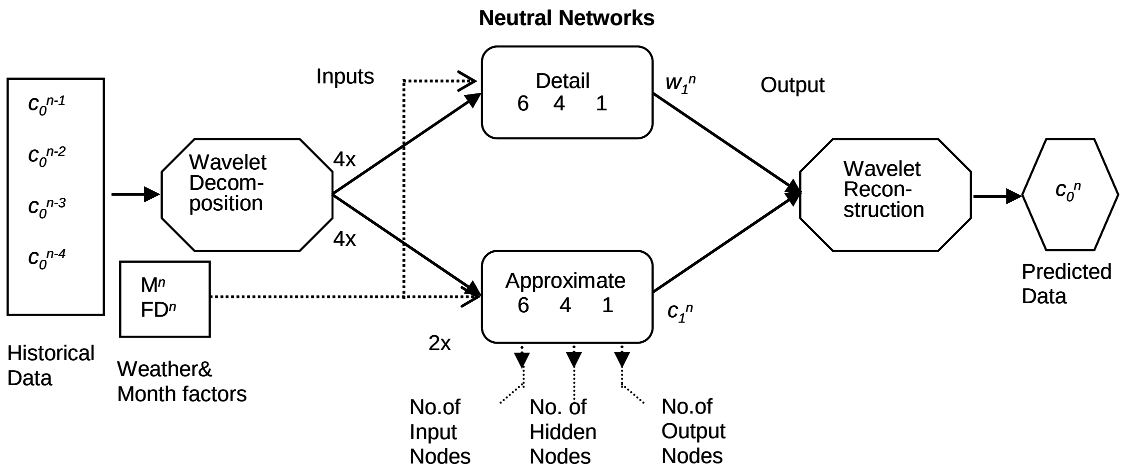

Figure 5.7 The Structure of Wavelet Based NN Model with Resolution Level 1

5.3 AIS Hybrid Model

As mentioned in Chapter 4, overtraining is a common problem for neural network training. To overcome overtraining of neural networks, we have applied an artificial immune system approach to fine-tune the trained neural networks during the testing stage. AIS is a field of computational intelligence, which has emerged in the past decade as an important tool that can be applied to many engineering problems [98]. In recent years, computational models of various aspects of the vertebrate immune system, such as clonal selection, have been proposed. In clonal selection, a class of immune system related cells called B cells play a key role. B cells produces antibodies that recognize antigens (*i.e.* foreign bodies) and bind to them, thereby marking these antigens for eventual elimination by the immune system. The ability of a B cell to bind to an antigen is referred to as its affinity. The immune system maintains a repertoire of B cells that are constantly improved. Those B cells with the highest affinities are allowed to proliferate through a process of cloning. In this step, a number of clones of each B cell, which are identical copies of itself, are produced. In order to introduce variation, and possible further improvement in the affinities of the antibodies, the cloned B cells next undergo a process called hyper-mutation, where small random perturbations are introduced in them. Those new B cells are merged with the original ones, and those that have higher affinities are then selected to enter the repertoire [99]. A method to apply the clonal selection principle to optimization was proposed recently

[100], which has found many applications in engineering. This approach has been applied for retraining of neural networks used in this research where clonal selection approach is used in conjunction with back propagation for improved performance.

In the proposed hybrid algorithm, a repertoire of 10 trained neural networks is maintained. All the networks have identical number of layers, and neurons, but differ in the weights. Each neural network is analogous to a B cell in an immune system. Better networks with lower errors are considered to have higher affinities while those with higher errors have lower affinities. The hybrid algorithm proceeds iteratively. In each time step during the prediction stage, the neural networks in the repertoire are ranked based on their affinities. A number of clones of each network is then obtained based on the latter's rank. The neural network with the highest affinity produces 4 clones, while those ranked 2, 3 and 4, produce 3 clones each. The 5^{th} network produces 2 clone, while the remaining ones are cloned only once each. The weights of each clone are identical to the original networks.

The clones are then subject to hyper-mutation. During this step, the weights are changed by the addition of small random perturbations. If the set of weights of any clone is represented as a matrix **W**, then hyper-mutation is implemented as,

$$\mathbf{W} = \mathbf{W} + \delta \cdot \mathbf{U}(-1,1). \qquad (5.14)$$

In the above equation, the quantity δ is a constant associated with hypermutation, and $\mathbf{U}(-1,1)$ is a matrix of the same dimensions as **W**, of uniformly distributed random numbers in the interval [-1, 1]. The standard back-propagation algorithm was iteratively applied 10 times to each clone with $\delta = 0.05$ and a learning rate of 0.4, which were selected based on trial and error.

Following the hypermutation and back-propagation steps, the affinities of the cloned neural networks are evaluated and they are inserted into the repertoire. Only the best 10 networks are retained for the next time step while the rest are discarded.

5.4 Simulations

The models were trained with historical data for the four cities from the year 1998 to 2006 and tested for 2007. The performance of the proposed MWNN hybrid model with resolution level 1 wavelet decomposition is compared to a model without AIS-based hyper-mutation and to a resolution level 3 wavelet decomposition model (WNN). The performances of the models are measured using the AAE, which are given in Table 5.1. MWNN+AIS represents the hybrid model. The hybrid model outperformed the other two models in the testing stage with the smallest AAE for all the cities and the MWNN model gave better results than the WNN for all the cities. Also, the MWNN gave better results than the WNN in the training stage for all the cities. Note that the hybrid model and the MWNN model have the same results for the training stage because they are the same model in this stage. The difference is during the testing phase, where the hybrid model improves the neural networks at every instant with newly gained data. Figure 5.8-5.19 show plots of weekly observed and predicted outages for four cities by the three models. It can be seen that the models can reproduce the fluctuating patterns of the time series quite well both in training and testing durations.

Plots of weekly observed and predicted outages by the three models for four cities for 2007 are shown in Figure 5.20-5.22. Slopes (S) of the best-fit lines and their correlation coefficients (R) are also shown on these graphs. Values of one for both these variables would indicate best fit around the ideal (y=x) line. The results clearly show that the hybrid model gives the best results and the level 3 DWT model gives the worst performance. The results of Topeka are very comparable with that of Wichita. Both Lawrence and Manhattan (the smaller cities) have slopes slightly different from the ideal value of 1 with Lawrence having a value lower than 1 and Manhattan having a value higher than 1. This could be attributed to the size of the cities. The correlation values are very similar for Topeka and Wichita. We hypothesize that spatial aggregation increases the accuracy of the models based on limited observations. However, further research is needed to verify this hypothesis.

Table 5.1 Results of Different Models for Four Cities

City	Model	Training Error	Testing Error	S	R
Wichita	MWNN+AIS	4.87	2.90	1.04	0.88
	MWNN	4.87	3.31	1.03	0.83
	WNN	6.79	6.40	1.33	0.79
Topeka	MWNN+AIS	3.81	2.75	1.00	0.93
	MWNN	3.81	3.47	1.01	0.90
	WNN	5.75	4.45	1.05	0.90
Lawrence	MWNN+AIS	1.56	1.00	0.91	0.91
	MWNN	1.56	1.24	0.90	0.87
	WNN	2.32	1.99	1.27	0.82
Manhattan	MWNN+AIS	1.28	0.68	1.05	0.90
	MWNN	1.28	0.99	1.13	0.85
	WNN	1.52	1.44	1.34	0.83

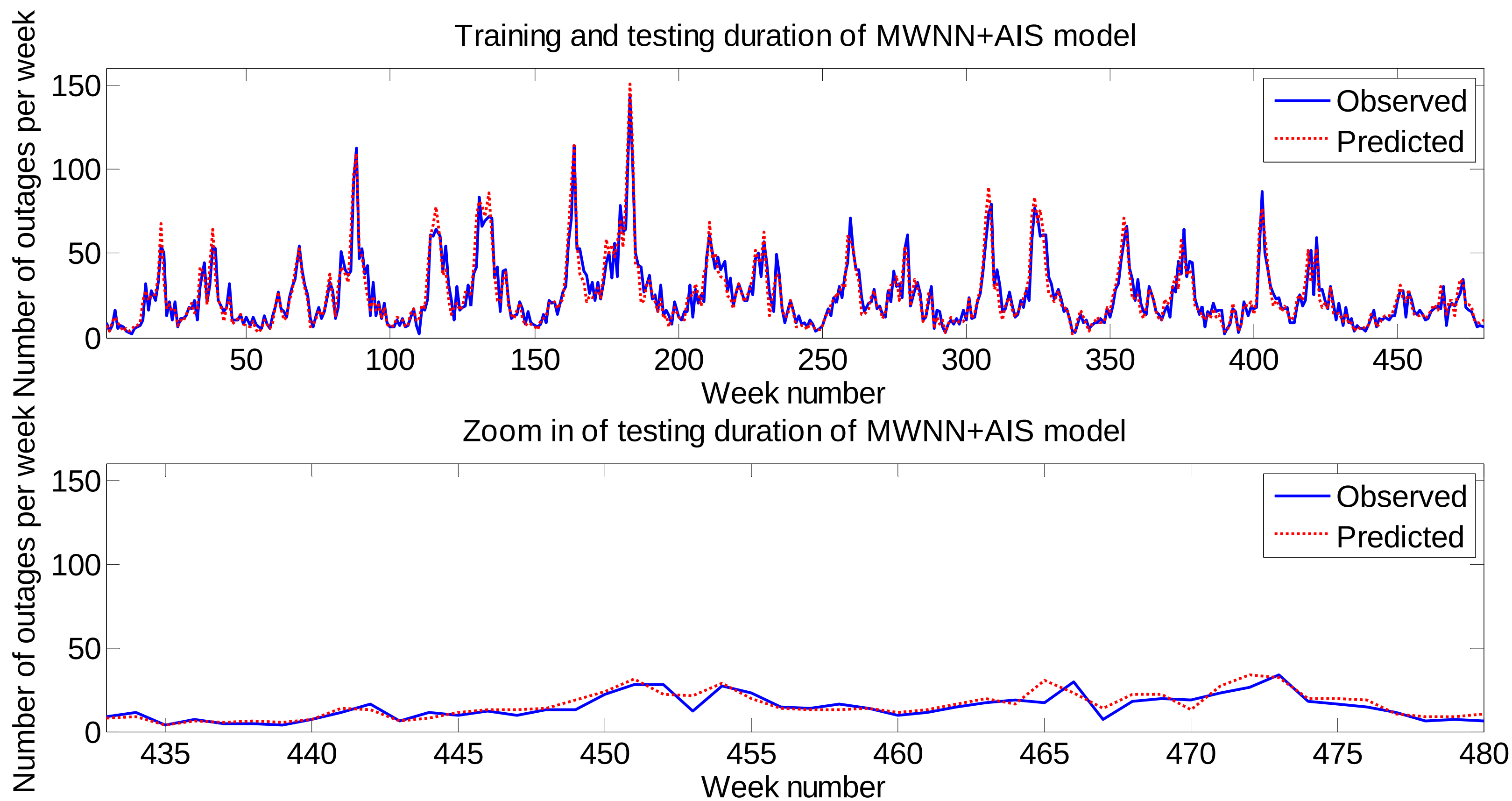

Figure 5.8 Outages Observed and Predicted by MWNN+AIS Model in Wichita from Year 1998 to 2007

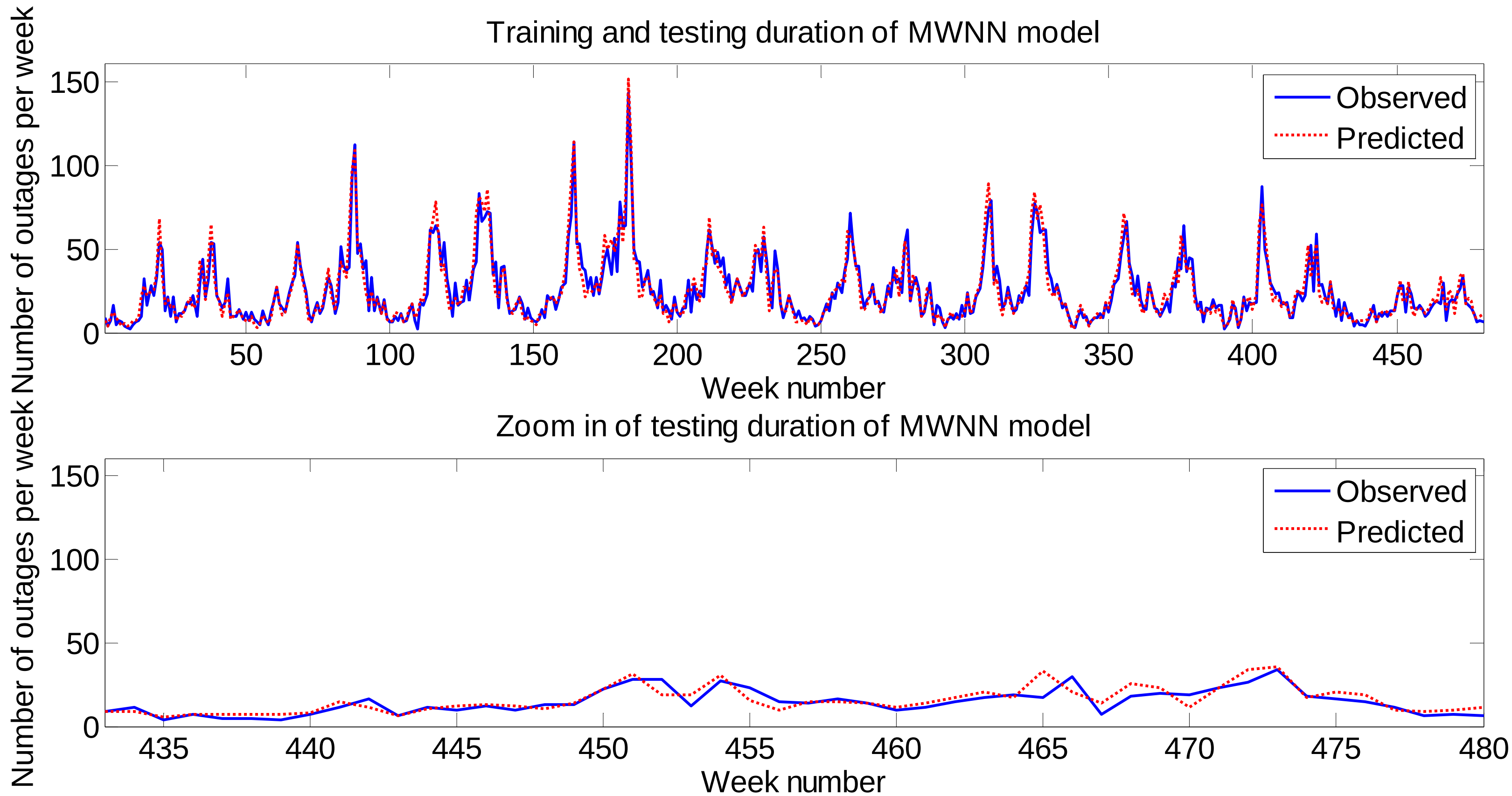

Figure 5.9 Outages Observed and Predicted by MWNN Model in Wichita from Year 1998 to 2007

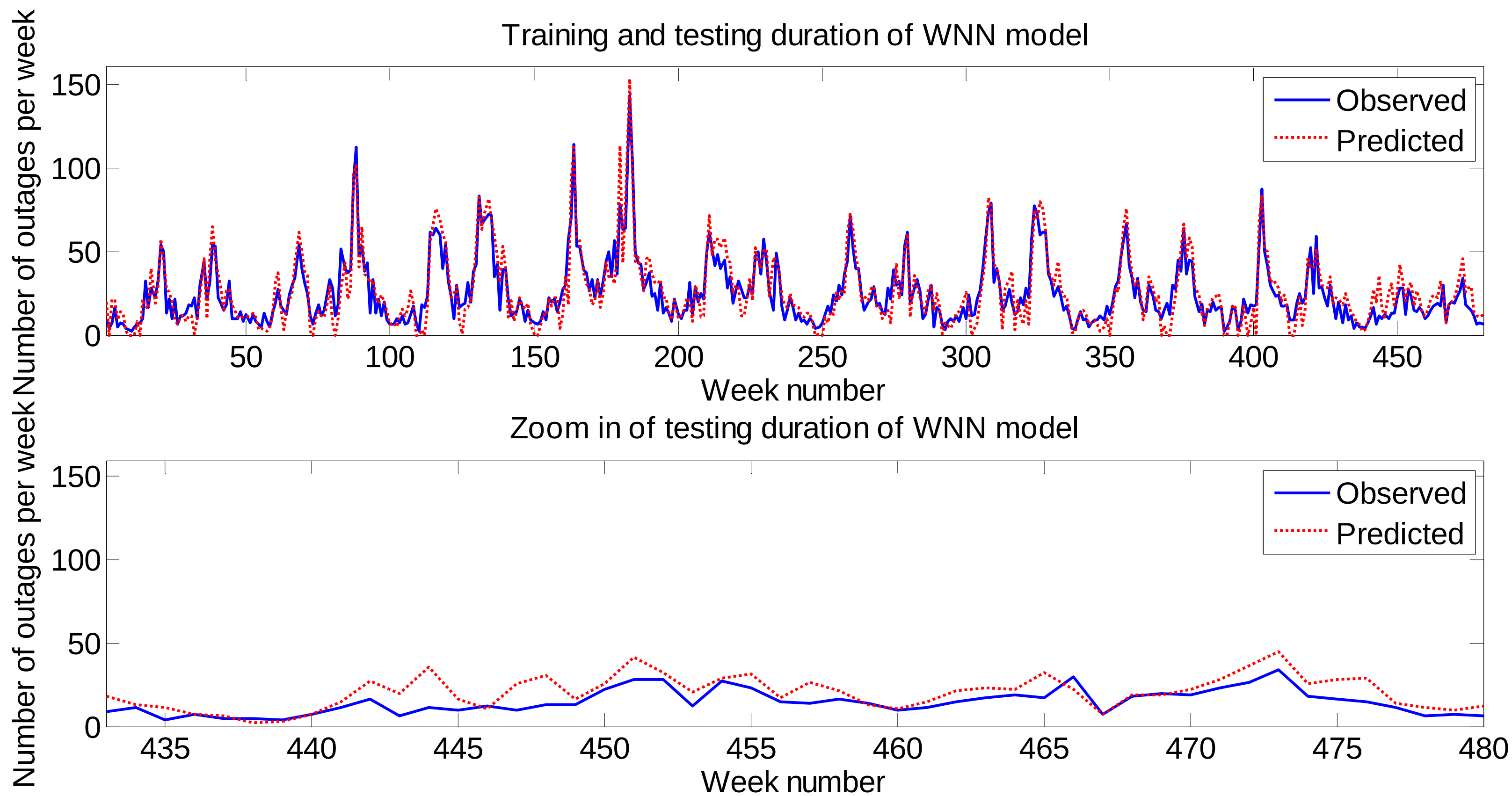

Figure 5.10 Outages Observed and Predicted by WNN Model in Wichita from Year 1998 to 2007

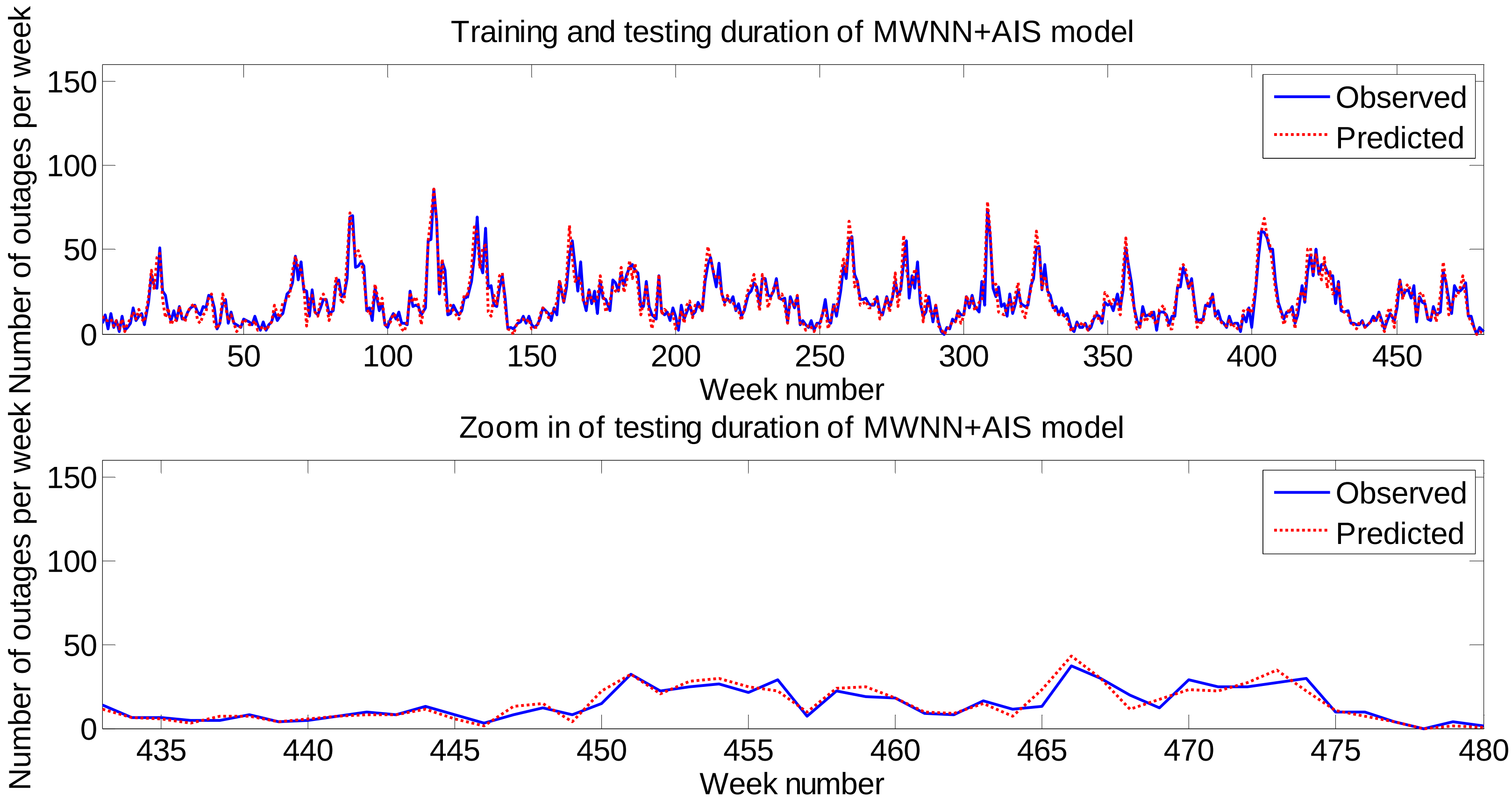

Figure 5.11 Outages Observed and Predicted by MWNN+AIS Model in Topeka from Year 1998 to 2007

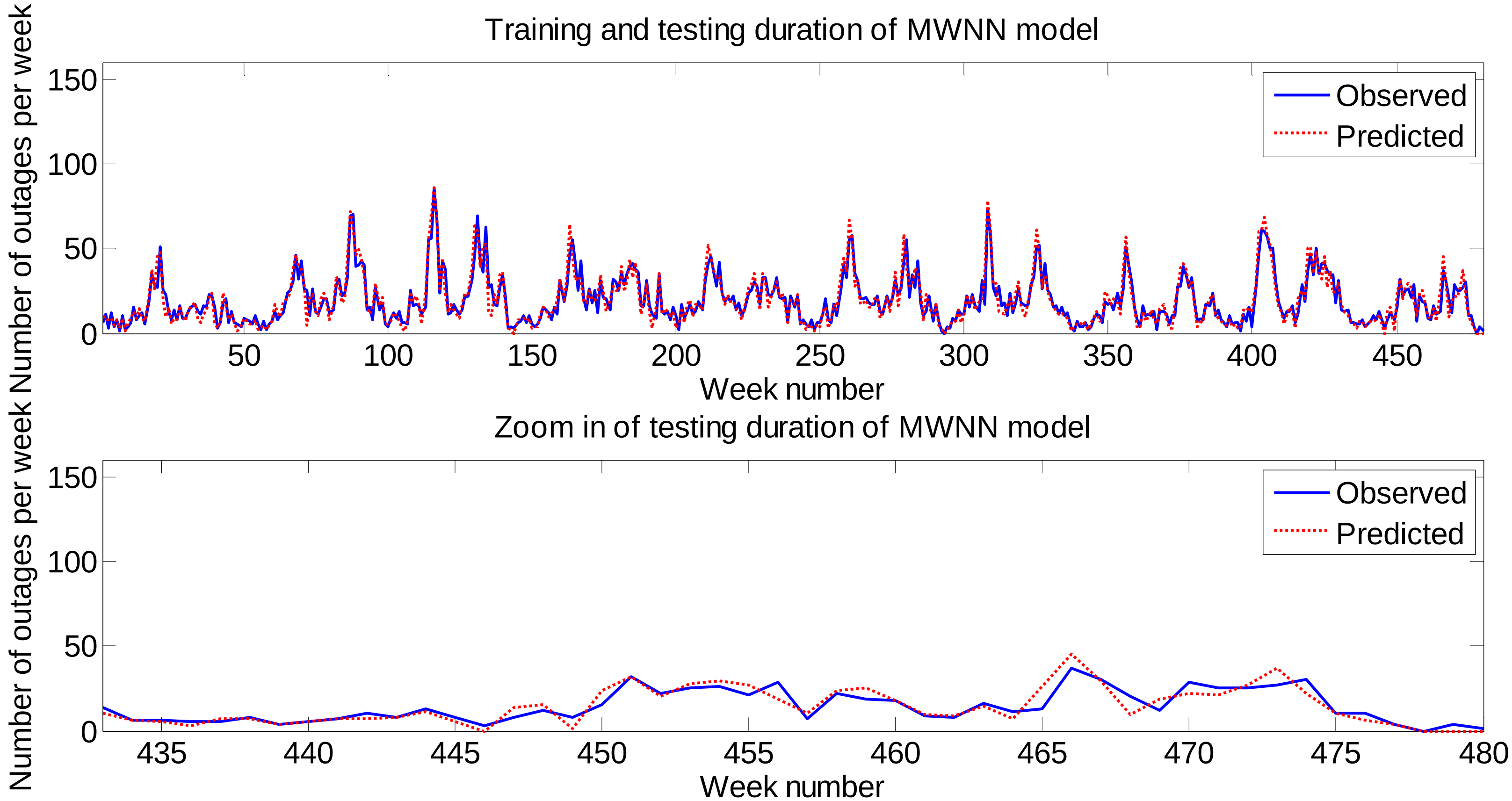

Figure 5.12 Outages Observed and Predicted by MWNN Model in Topeka from Year 1998 to 2007

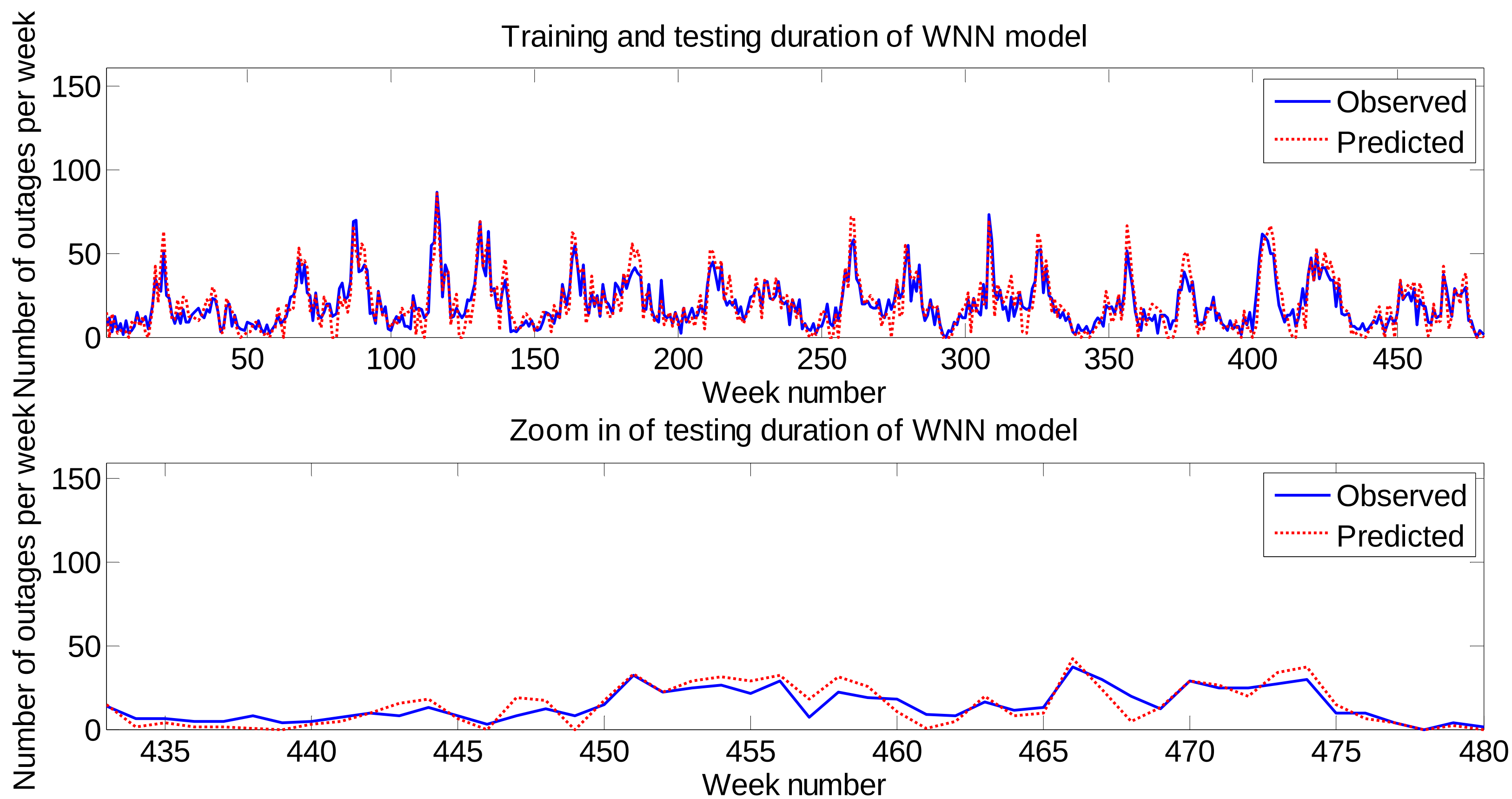

Figure 5.13 Outages Observed and Predicted by WNN Model in Topeka from Year 1998 to 2007

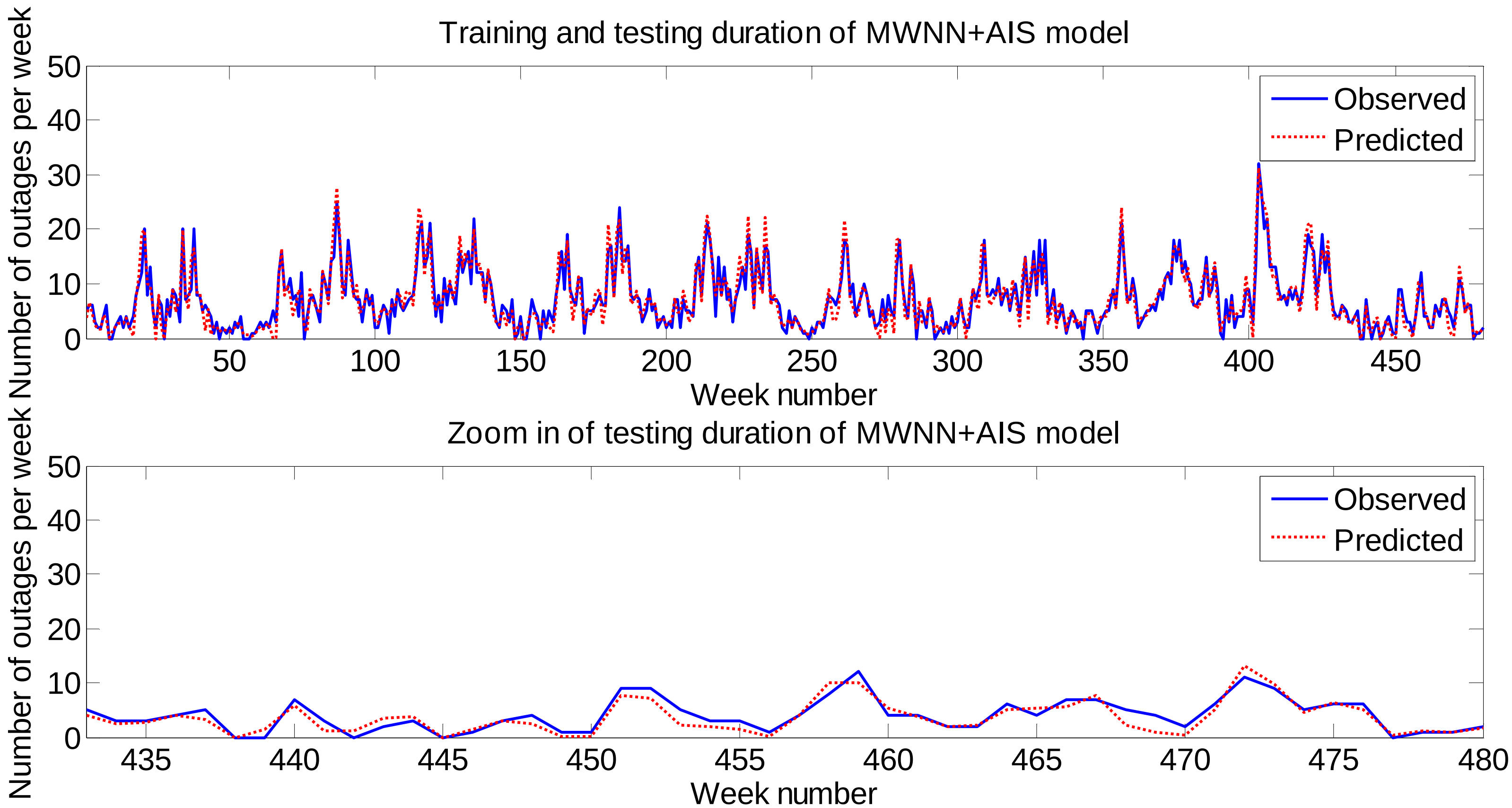

Figure 5.14 Outages Observed and Predicted by MWNN+AIS Model in Lawrence from Year 1998 to 2007

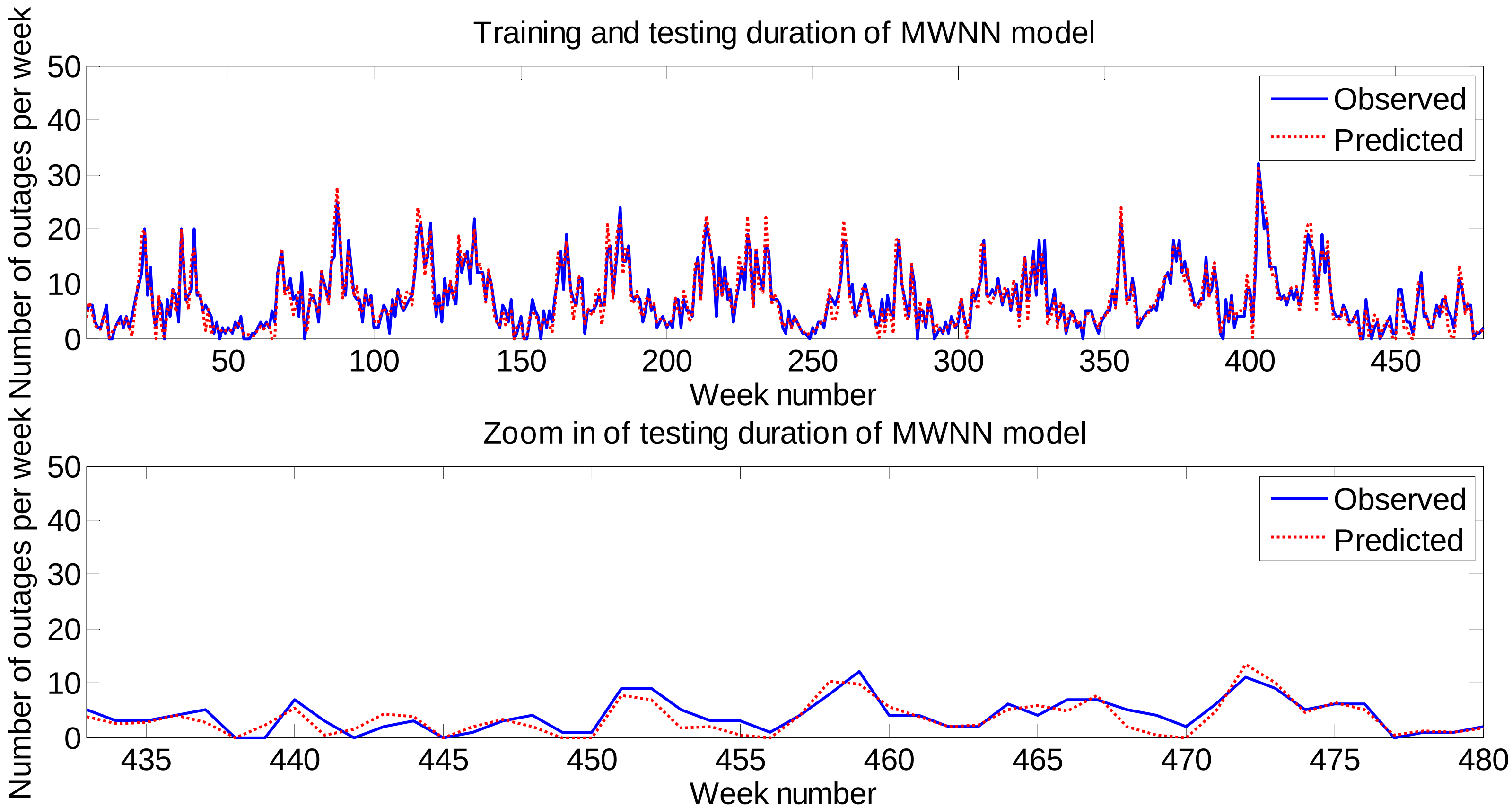

Figure 5.15 Outages Observed and Predicted by MWNN Model in Lawrence from Year 1998 to 2007

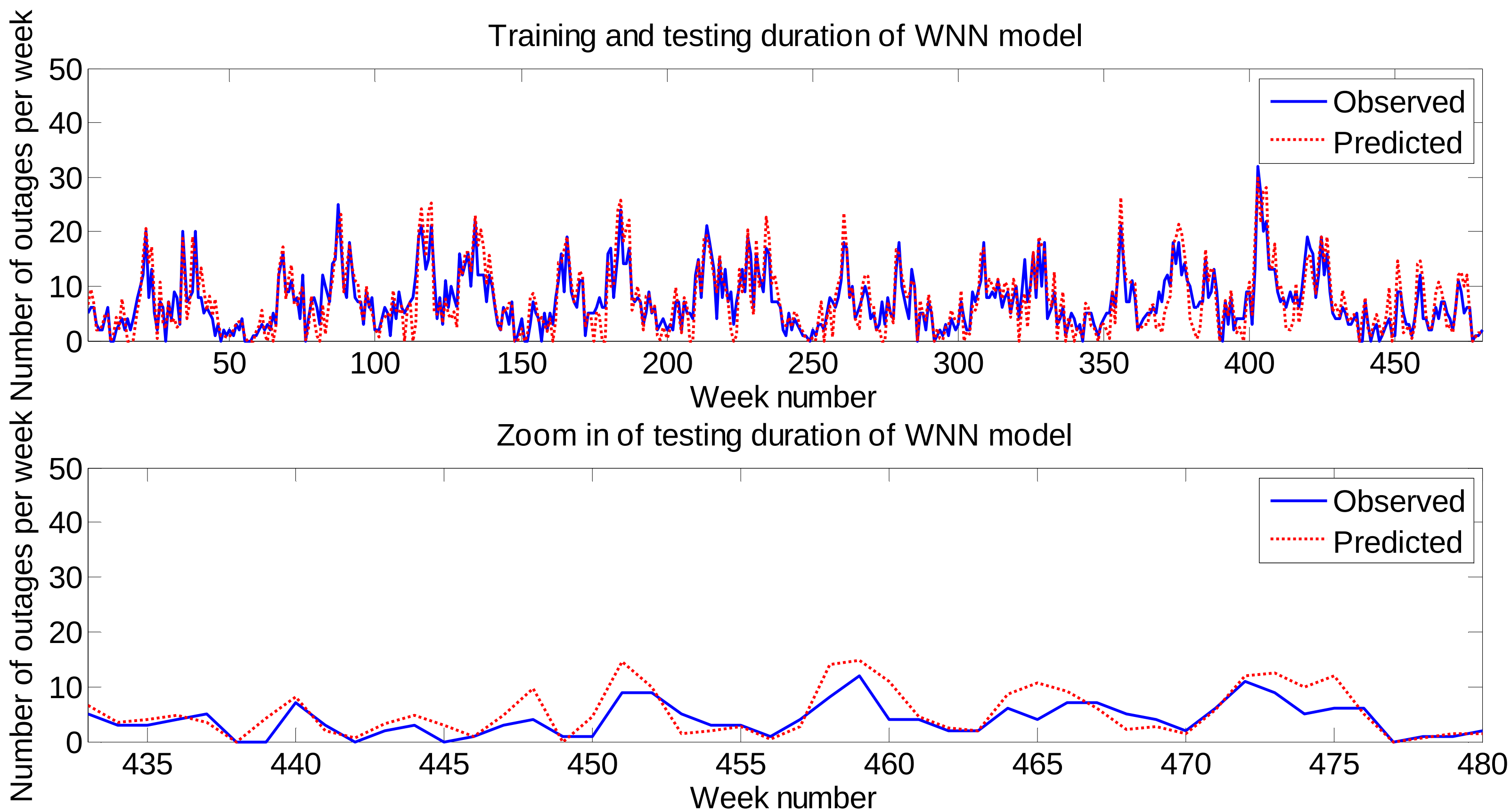

Figure 5.16 Outages Observed and Predicted by WNN Model in Lawrence from Year 1998 to 2007

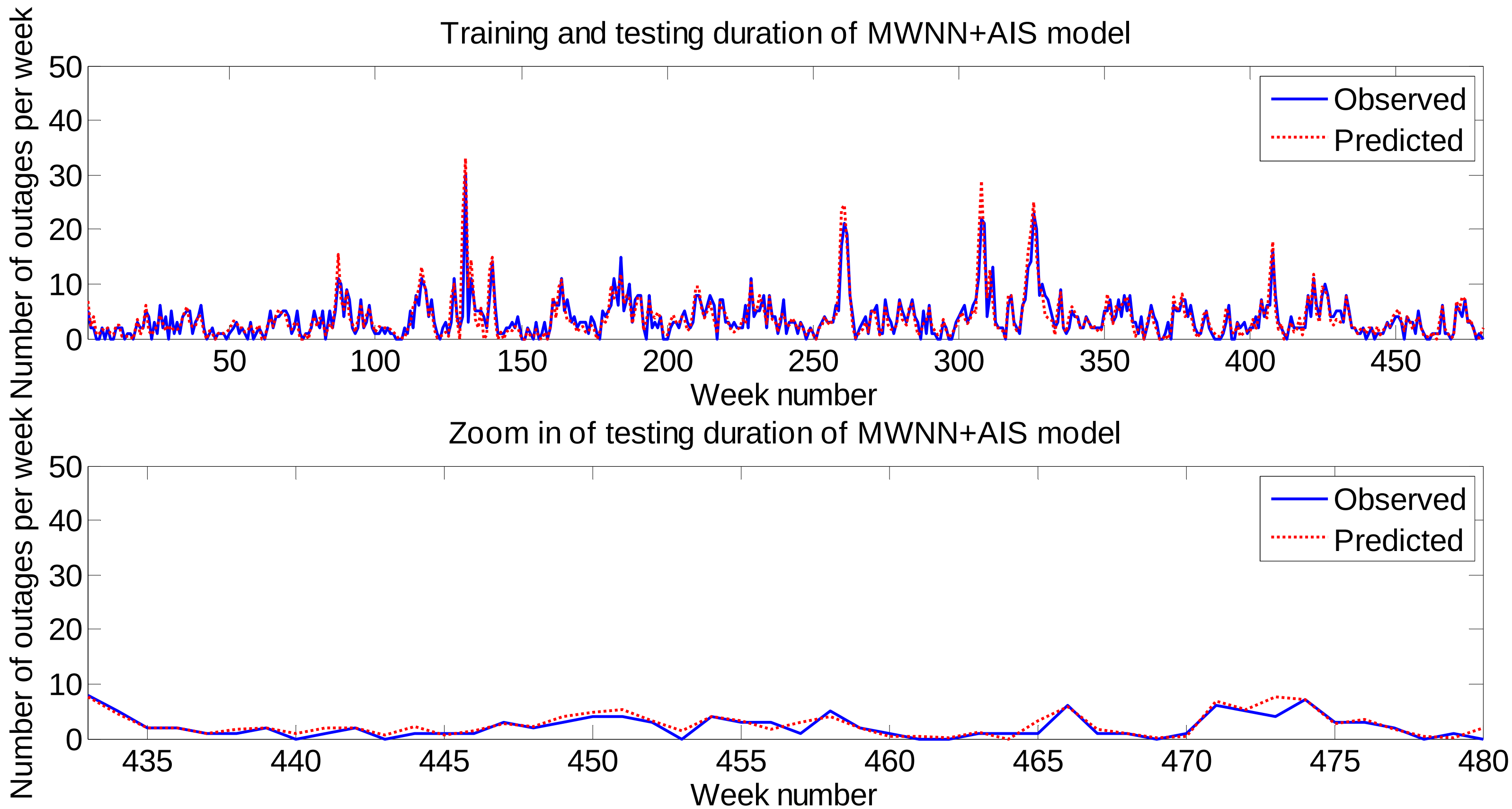

Figure 5.17 Outages Observed and Predicted by MWNN+AIS Model in Manhattan from Year 1998 to 2007

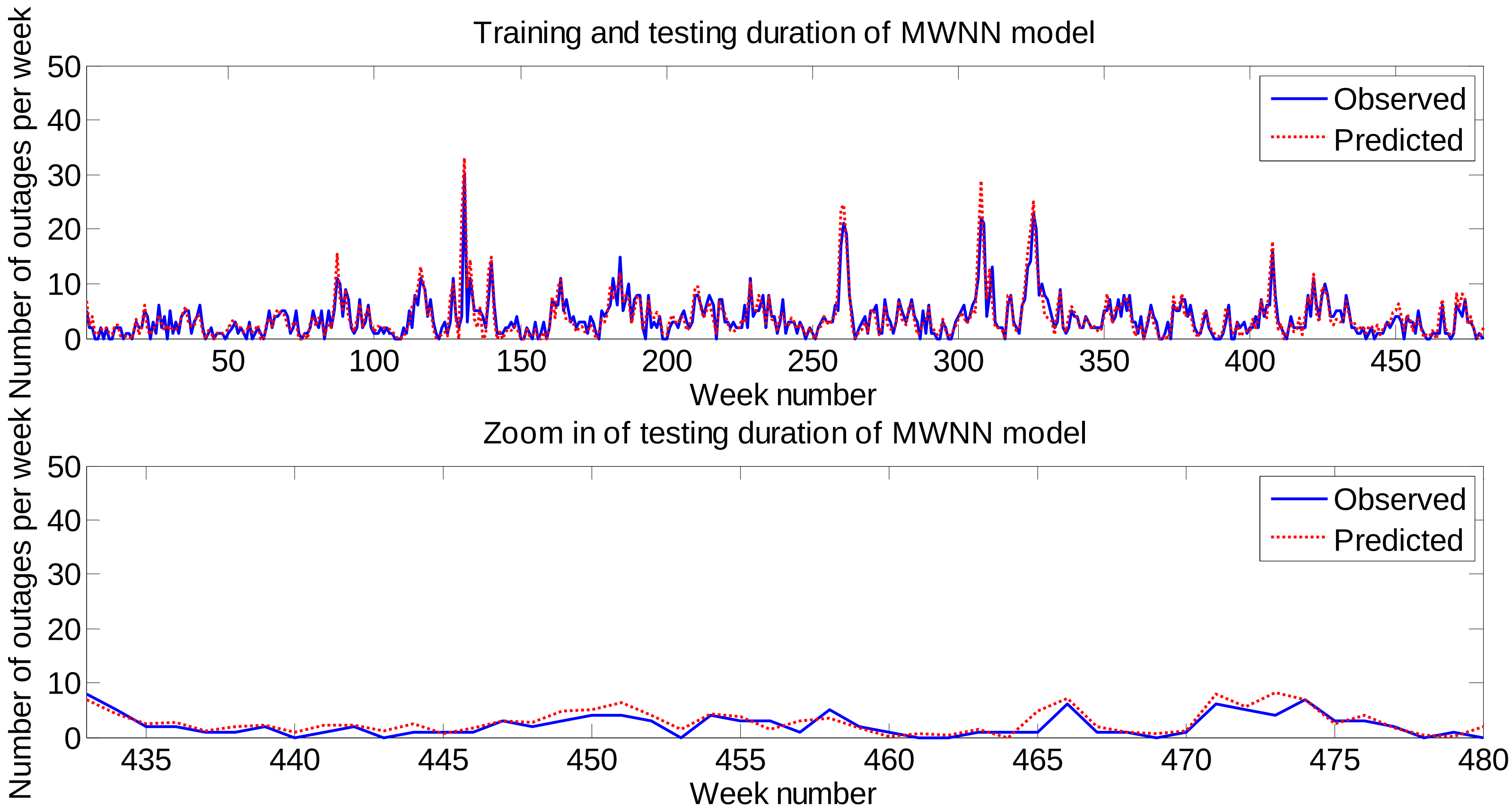

Figure 5.18 Outages Observed and Predicted by MWNN Model in Manhattan from Year 1998 to 2007

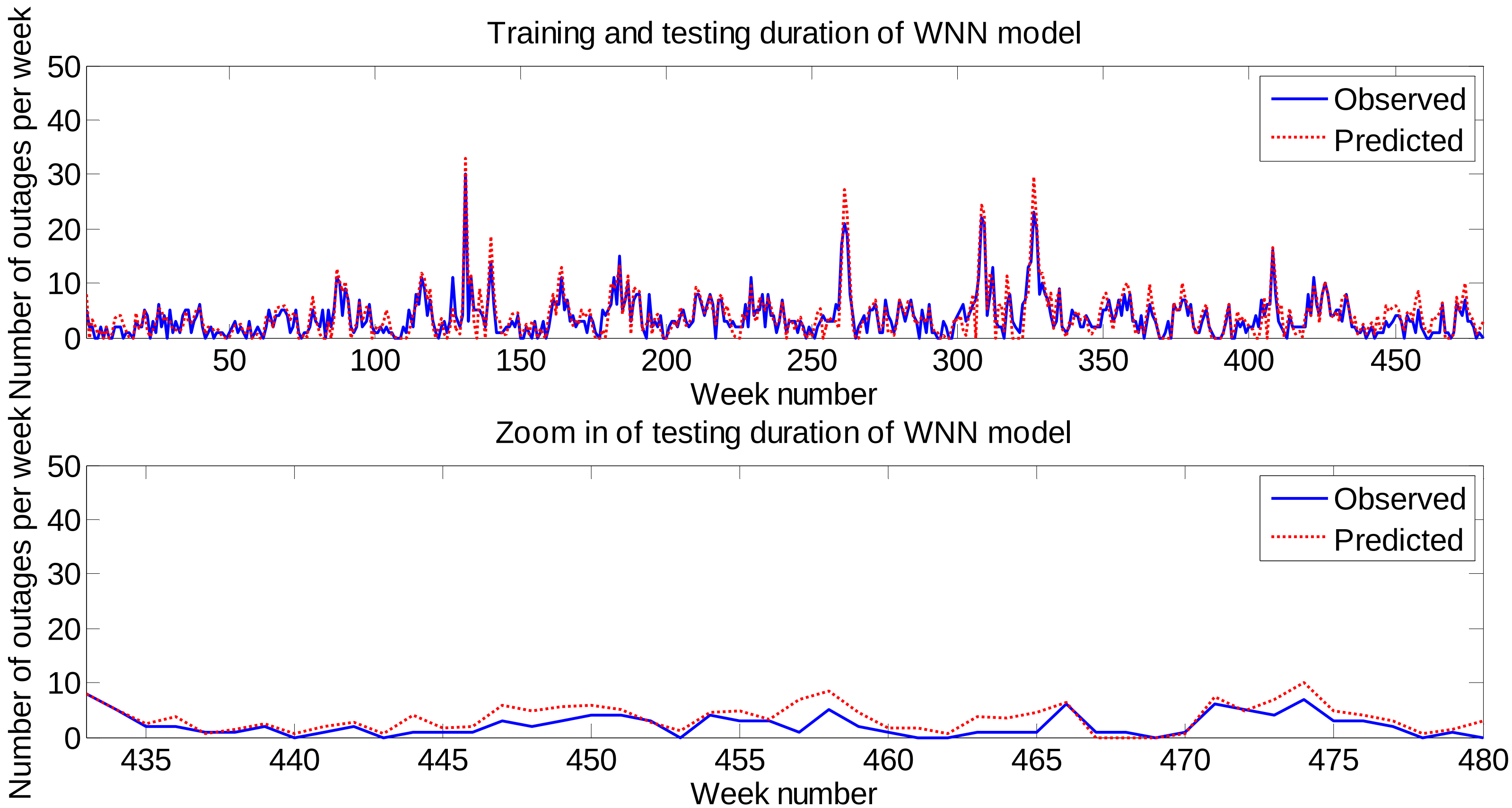

Figure 5.19 Outages Observed and Predicted by WNN Model in Manhattan from Year 1998 to

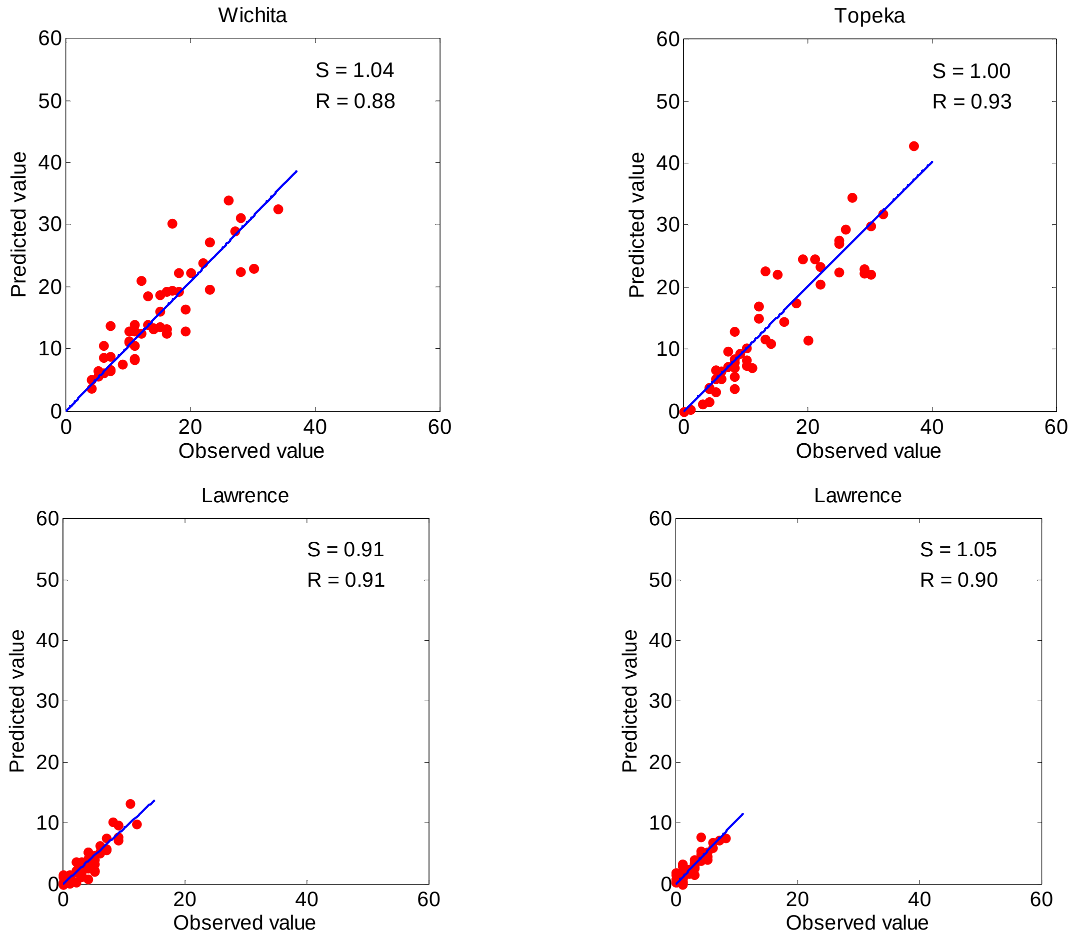

Figure 5.20 Outages Observed and Computed by MWNN+AIS models for Four Cities in 2007

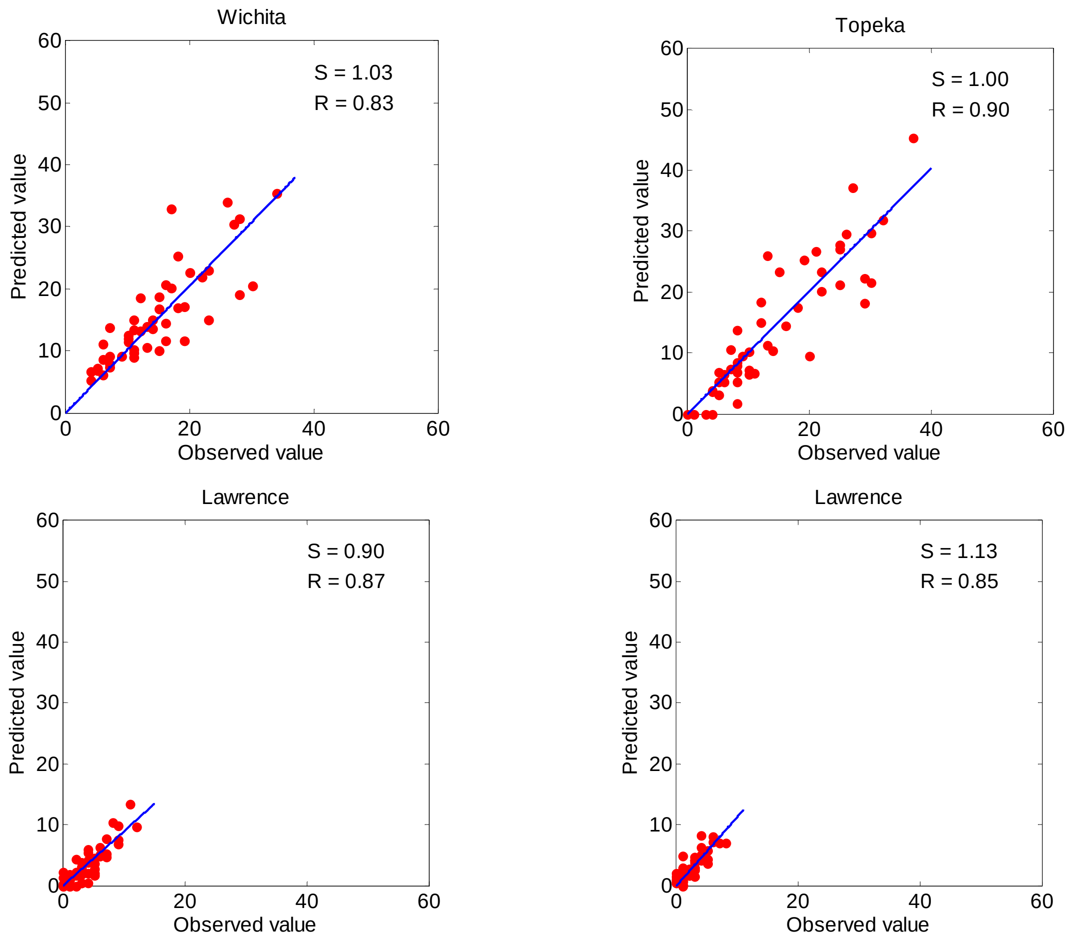

Figure 5.21 Outages Observed and Computed by MWNN Models for Four Cities in 2007

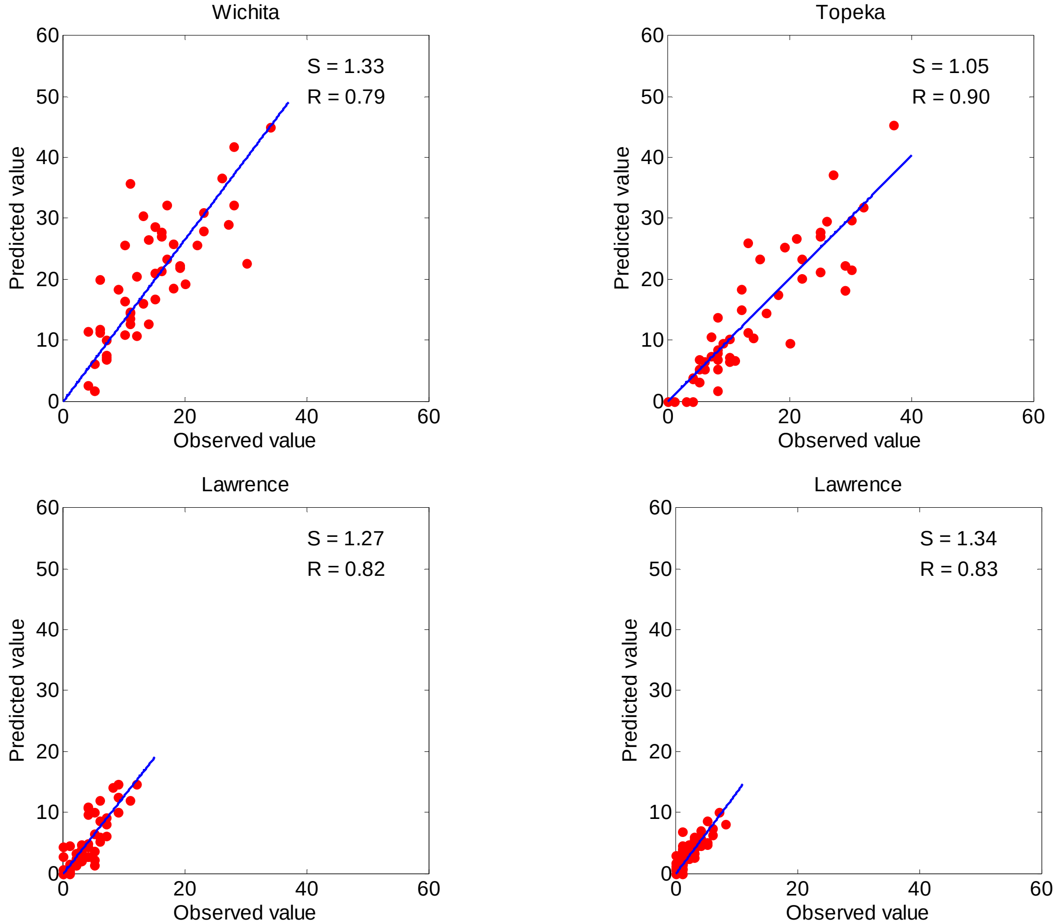

Figure 5.22 Outages Observed and Computed by WNN Models for Four Cities in 2007

CHAPTER 6 - Bayesian Models and Monte Carlo Simulations

Since the occurrences of animal outages in a distribution system are random events, they can be successfully modeled using probabilistic methods [76]. Sahai and Pahwa [76] have previously proposed using Bayesian networks for predictions of animal related outages on distribution feeders. Additional work using these models has been done and is presented in this chapter. Compared to the previous work, this dissertation has constructed Bayesian models with more inputs and applied them to more data. Besides the weekly basis, more predictions have been done on a monthly and yearly basis. Also, this dissertation uses Monte Carlo simulations to find the confidence intervals for the predictions.

6.1 Introduction to Bayesian Model

6.1.1 Bayes' Theorem

Bayes' theorem presents the relationships of the conditional probabilities and marginal probabilities of two random events. Usually it is used to update the conditional probability of a event A taking account of new observations of occurrences of event B. Mathematically, Bayes' theorem is formulated by the following equation [101]:

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)} \tag{6.1}$$

- P(A) is the prior probability or marginal probability of A. It is "prior" because no information about B is considered.
- P(A|B) is the conditional probability of A, given B. It is also called the posterior probability because it is computed after the event B has been observed.
- P(B|A) is the conditional probability of B given A.
- P(B) is the prior or marginal probability of B.

Note that B has to have a non-zero prior probability in equation 6.1.

6.1.2 Bayesian Network

A Bayesian network is comprised of a set of variables $\{x_1, x_2, \ldots, x_n\}$, a graphical structure and a set of conditional probability tables. A Bayesian network is a directed acyclic graph which means a graph with no loops [102-104]. Each variable is represented by a node in the graph. And there are connection arcs between nodes. An arc leads a parent (casual) node to a child (influenced) node and denotes the conditional dependence between the child and parent nodes. On the other hand, if there is no connection arc between two nodes, it indicates conditional independence. There is a conditional probability table for each child node, which can be computed by the prior probabilities of the parent nodes.

6.1.3 Learning and Prediction by Bayesian Model

Not only the conditional probability tables but also the casual relationships can be learned from the data [105]. However, it is much easier to learn the conditional probability tables as compared to graph topology learning [105]. Also, it is easier to learn with fully observed data, as compared to partially observed data where some nodes are hidden or some data is missing [105]. With fully observed data and known structure, the Maximum Likelihood Estimation (MLE) algorithm is effective [105]. For unknown graph structure, algorithms that search through model space are used [105]. MLE is a method of estimating parameters of a population such that the selected values maximize the likelihood of a sample [105]. The goal of learning in this case is to find the values of the parameters of each cumulative probability distribution, which maximizes the likelihood of the training data [105].

A Bayesian network can be used to learn causal relationships between parents and child nodes which are captured in the conditional probability tables [102]. After the graph structure and the conditional probability tables are learned, a Bayesian model can be used for predictions. Given the values of parent nodes and the learned conditional probability tables, the values of the child nodes can be estimated [105]. To predict the child nodes given the status of the parent nodes, top-down reasoning is used where we can compute the probability of an effect given the cause [105].

6.2 Model Construction

From Chapter 2, we know that weekly animal caused outages in distribution systems are influenced by the number of fair weather days per week and the month type. With this prior knowledge it is much easier for us to construct the Bayesian model for our study because we do not have to learn the structure from the data. A one-layer discrete Bayesian network with three nodes representing the three variables, the month type, the fair days level and the weekly animal-related outage level, is shown in Figure 6.1. This network has the same structure as the one used by Sahai and Pahwa [76]. The variables, the months, the number of fair weather days per week and the number of animal caused outages, are all classified into discrete levels. This is done primarily because with discrete variables, the conditional probability tables are simple to be computed and easy to use. Also we do not have enough training data to find conditional probability distributions associated with a continuous Bayesian model.

However, the classification of the input data to discrete levels is at the expense of the performance of the model in predictions because there is loss of information during the classifications and all the data points in each level are treated in the same manner. To make the model as accurate as possible, the data needs to be examined carefully to get the best classification. The classification for the parent nodes should be done in a way such that all the data points which have similar influences on the child nodes are grouped into the same level. Conversely, the data points which have different impacts on the child nodes should be grouped into different levels [76]. Also, there should be sufficient data entries for each combination of inputs because a reliable conditional probability distribution needs enough observations in the data. For the child node, the classification should be done so as to have as many levels as possible, with relevant number of data entries in each level [76]. The more levels we have for the child node, the more information we have about the effects of parent nodes on the child node and thus a better prediction of the outages will be obtained.

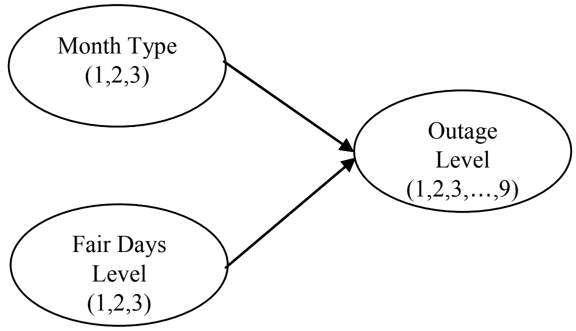

Figure 6.1 One-layer Bayesian Model for Predictions of Animal-related outages

6.2.1 Classification of Weather Conditions

The input variable month signifies the behavioral patterns of squirrels at different times of the year. As mentioned in Chapter 2, the squirrel activity in the months of January, February and March is the minimum. Hence, these months have been grouped together and classified as the month type 1. The squirrel activity in the very hot months of July and August and the temperature changing months of December and January is moderate. Hence, these months have been grouped together and classified as month type 2. Squirrels are most active in the months of May and June, and September, October and November. Therefore, these months have been grouped together and classified as month type 3. This gives us the following classification for month type:

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

Note that this classification is different from that used in previous work [76], where January is placed in type 2 and April is in type 1 category. The difference arises from observations with more data from different cities in this dissertation.

As defined in previous work, fair weather days has temperatures between 40 and 85 degrees Fahrenheit with no other weather activity [76]. We have counted the number of fair days per week for different cities based on the weather data obtained from Kansas State University Weather Services. The number of fair days per week can vary from 0 to seven or eight based on division of every month into exactly 4 weeks as discussed in Chapter 2. Given that we already have three input states for month type, dividing the number of fair days per week into three different levels gives nine input states. Similar to previous work [76], weeks with zero number of fair weather days are classified as level 1 of fair weather days per week assuming that they will have least impacts on animal caused outages. Weeks with one to three fair weather days have been classified as the mid-level or level 2 of fair weather days per week assuming that they will have moderate impacts on animal caused outages. The weeks in which nearly half or more than half of the days are fair weather days, will have the most impacts on animal caused outages and have been classified as level 3 of fair weather days per week. That is, the weeks that have four to seven (or eight) fair weather days are grouped under level 3. Finally, we get the fair weather days per week classification as follows:

- Fair Weather Days per Week Level 1: 0 Fair Weather Days
- Fair Weather Days per Week Level 2: 1 ~ 3 Fair Weather Days
- Fair Weather Days per Week Level 3: 4 ~ 7 (or 8) Fair Weather Days

This classification for the number of fair days per week is the same as that in previous work [76].

6.2.1 Classification of Weekly Animal-related Outages

The overhead distribution feeder outage information for different areas in Kansas from 1998 to 2007 was obtained from Westar Energy. Wichita has the biggest area in Kansas within the Westar Energy service territory. We take Wichita as an example to show the modeling since it has the most data. We computed the animal caused outages per week in the Wichita area for the past seven years, which means a total of 480 weeks. Figure 6.2 shows the histogram of animal caused outages per week in Wichita area, in which the n^{th} bar represents the observation of n-1 outages per week. The proper classifications of outages should improve the model performance. In previous work [76], the classification of the child node has been done so as to have as many levels as possible, with relevant number of data entries in each level. For example, Manhattan has outage range from 0 to 30. Due to simplicity, animal caused outages per week have been classified into 12 levels instead of the possible 31 levels. The lower outage levels have a bin size of one. This is done because the lower numbers of animal caused outages occur more frequently in the system, for example there are a lot more zero and one animal caused outage per week as compared to twenty outages per week [76]. In this dissertation, we have tried to find the proper numbers of outage levels for all four cities by experiments. The AAE of the predictions by Bayesian models with 9 input states with different outage levels is shown in Table 6.1. We have found that classifications with 9 outage levels give the best results for almost all of the cities. To classify outages into more levels than 9 doesn't significantly improve the model performance. Therefore, for uniformity and simplicity, 9 levels of outages are used for all the cities.

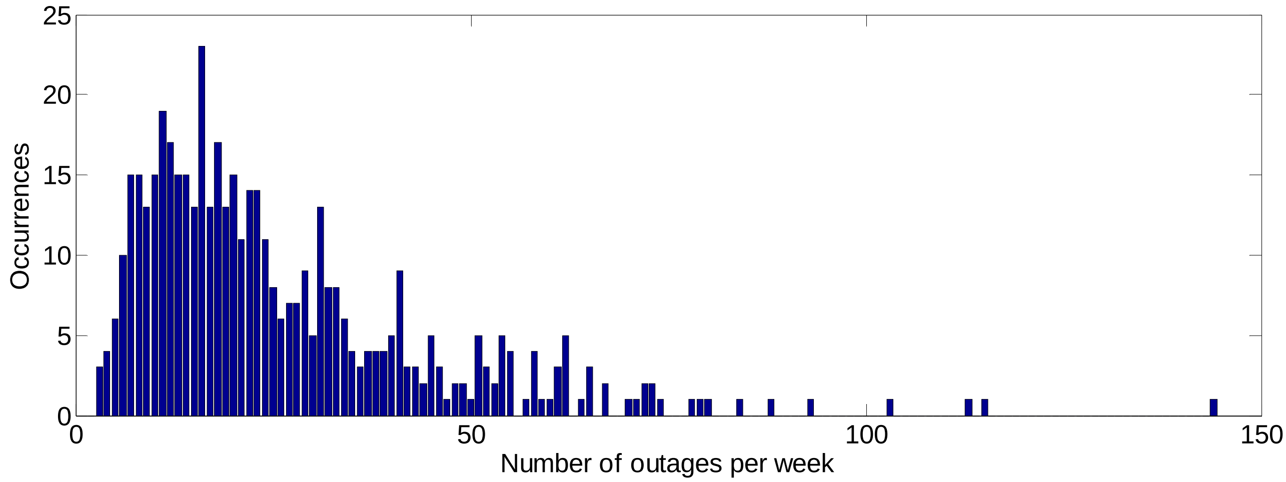

Figure 6.2 Histogram of Weekly Animal-related Outages from the Year 1998 to 2007 in Wichita

Table 6.1 The Performance of Predictions with Different Numbers of Outage Levels

Number of Outage Levels	AAE For Wichita	AAE For Topeka	AAE For Lawrence	AAE For Manhattan
1	14.216	10.960	4.080	2.540
2	10.734	9.186	3.450	2.409
3	9.892	7.357	3.187	2.184
4	9.609	7.195	3.174	2.181
5	9.605	7.175	3.172	2.182
6	9.542	7.171	3.173	2.177
7	9.540	7.171	3.177	2.177
8	9.536	7.171	3.175	2.178
9	9.534	7.172	3.174	2.176
10	9.531	7.174	3.174	2.178
11	9.535	7.178	3.174	2.179
12	9.536	7.176	3.174	2.178
13	9.536	7.176	3.174	2.178

14	9.534	7.177	3.175	2.179
15	9.530	7.179	3.171	2.179
16	9.530	7.179	3.171	2.179
17	9.532	7.178	3.170	2.181
18	9.533	7.177	3.171	2.182
19	9.534	7.177	3.172	2.182
20	9.534	7.177	3.172	2.182

Note that, with the same number of bins, the classifications for outages are different for each city since the histogram of weekly animal-related outages are different for each city. But we have tried to follow one general rule for the classifications which is to strive to make every bin have roughly the same count of occurrences as much as possible. For Wichita, in total there are 480 occurrences of weekly outages and the outages vary from 0 to 143. For every bin there should be about 53 (480 divided by 9) occurrences. In Figure 6.2, there are 144 bars in the histogram each of which records the occurrence of the corresponding weekly outages. The biggest bar is the 16th bar and it has 23 occurrences, and the two bars next to it have 13 occurrences each. We group these three bars together and get 49 occurrences which is the closest we can get to the average occurrences of 53. Then, we aggregate the bars in the order from the first one to the last one and every time when we find bars with total occurrences equal to or bigger than 49 we group these bars together as a bin. Following this method, we have obtained the outage levels for Wichita as:

- Outage Level 1: 0~9 Animal Caused Outages per Week;
- Outage Level 2: 10~13 Animal Caused Outages per Week;
- Outage Level 3: 14~17 Animal Caused Outages per Week;
- Outage Level 4: 18~21 Animal Caused Outages per Week;
- Outage Level 5: 22~25 Animal Caused Outages per Week;
- Outage Level 6: 26~32 Animal Caused Outages per Week;
- Outage Level 7: 33~42 Animal Caused Outages per Week;
- Outage Level 8: 43~62 Animal Caused Outages per Week ;
- Outage Level 9: 63~143 Animal Caused Outages per Week.

6.2.3 Conditional Probability Table

The conditional probability table (CPT) gives the probability of occurrence of each outage level given a month type and a level of fair weather days per week, that is,

P (Outage Level = i | Month Type = j, Fair Weather Days per Week Level = k)

where i = 1,…,9, j =1,2,3 and k = 1,2,3.

Since the graph structure is fully known, we have used Maximum Likelihood Estimation to learn the values in the CPT with fully observed historical data. The input states are tabulated in Table 6.2 and the learned conditional probabilities are listed in Table 6.3. There are sufficient training cases for each input state except input state 7, as shown in Table 6.2. As for input state 7, the months are January, February and March, which are too cold to have many fair weather days. The equation we use to compute the conditional probabilities for input state m is:

P (Outage level = i | Input state = m) =

Number of occurrences in outage level i / Total number of occurrences in input state m

Table 6.2 All Possible States and Number of Observations for Wichita

Input State	1	2	3	4	5	6	7	8	9
Month Type	1	2	3	1	2	3	1	2	3
Fair Day Level	1	1	1	2	2	2	3	3	3
Number of Occurrences	74	71	20	34	52	64	12	37	116

Table 6.3 The Conditional Probability Table with 9 Input States for Wichita

Outage Level	1	2	3	4	5	6	7	8	9
Input State 1	0.61	0.19	0.15	0.04	0.00	0.01	0.00	0.00	0.00
Input State 2	0.11	0.20	0.21	0.2	0.14	0.07	0.04	0.03	0.00
Input State 3	0.00	0.10	0.25	0.00	0.05	0.15	0.35	0.1	0.00
Input State 4	0.56	0.15	0.18	0.06	0.03	0.03	0.00	0.00	0.00
Input State 5	0.08	0.25	0.21	0.13	0.08	0.19	0.06	0.00	0.00
Input State 6	0.02	0.05	0.05	0.13	0.09	0.2	0.19	0.19	0.09

Input State 7	0.25	0.33	0.33	0.08	0.00	0.00	0.00	0.00	0.00
Input State 8	0.03	0.22	0.24	0.27	0.11	0.14	0.00	0.00	0.00
Input State 9	0.00	0.03	0.02	0.07	0.11	0.16	0.17	0.29	0.15

The conditional probability table represents the influence of month and number of fair weather days per week on the number of animal caused outages per week [76]. In the previous work, the conditional probability table was sparse matrix which is not the case for this dissertation. The reason is that in this dissertation we have 3 more years' data and we have different classifications for the month type and the outage levels. But there are still some zero entries in the conditional probability table we have obtained as shown in Table 6.3. The zero occurrence in high outage levels in the first row results from the unfavorable weather conditions. This indicates that if the month is of type 1 and there are no fair weather days in a week, then we should only expect a very low number of animal-caused outages. When the number of fair days per week increases to 1~3 and the input state becomes 4, we expect more occurrences of high outages. As a result, the numbers of zeros in the fourth row decreases and there are animal-related outages at five and six levels. Comparing row 1 to 4, which have the same month type, there is an increasing trend in outages when the fair days level increases from 1 to 2. However, this observation is not true when the fair days level changes from 2 to 3, which is indicated by the increase of zeros in row 7. This can be explained by the lack of data in input state 7. In rows 2, 5 and 8 which have the month type of 2, the occurrences mainly range from outage level 1 to 7 with very few random events at the higher levels. When the fair days level increases from 1 to 3, the majority of the probabilities shift from lower outage levels to higher outage levels. There are only a few zero entries in rows 3, 6 and 9 which are at the month type 3 with majority of the probabilities at outage levels higher than 5.

Comparing the input states with the same month type, we have observed an increasing trend in probabilities at higher outage levels with increase in the fair days level. On the other hand, by comparing the input states with the same fair days level, we can find the effects of the month type on the outages. In rows 1, 2 and 3 with the fair days level of 1 and the month type increasing from 1 to 3, the majority of probabilities shifts from lower outage levels to higher outage levels. In rows 4, 5 and 6, the number zero entries at high outage levels decrease which can be observed in rows 7, 8 and 9. These observations indicate that even under the same weather conditions, there are more outages when the month type increases from 1 to 3. We can also see that the month type has a bigger impact on outages than the fair days level. In row 3, when there are zero fair days per week and the month type is 3, there are still lots of occurrences of outages at high outage levels 6, 7 and 8. But in row 7 with month type 1, when more than half of the days in the week have fair weather, there are no outages at outage levels higher than 4. This is similar to that observed by Sawti and Pahwa [76].

6.2.4 Prediction for Wichita

In order to get the expected values of the outages, conditional probabilities obtained from the Bayesian model are multiplied by the average value or median of each corresponding output level. In the previous work, the median values are used [76]. We have chosen the average values for characterization of each input state because they give better representation of the historical outage data. The median values are only determined by the classifications of outage levels. On the contrary, the average values take account the distribution of outages in the same outage level and thus can characterize the outage levels better. The average values for the outage levels in the data for Wichita are tabulated in Table 6.4.

Table 6.4 Average Values for Each Outage Level for Wichita

Outage Level	1	2	3	4	5	6	7	8	9
Average Value	6.49	11.39	15.52	19.49	23.15	29.16	37.49	52.00	80.52

Expected number of animal-caused outages in each input state can thus be computed by the following equation[76]:

$$E\,(\text{Number of outages} \mid \text{Input state} = j) = \sum_{k=1}^{9} P\,(\text{outage level} = k \mid \text{input state} = j) \times \text{Average (Outage level} = k) \qquad (6.4)$$

where,

- E (Number of animal-caused outages | Input state = j) is the expected number of animal-caused outages in input state j, j = 1,…, 9.

- P (Outage level = k | Input state = j) is the conditional probability of the occurrence of outage level k, given input state j, which can be looked up in Table 6.3.
- Average (Outage level = k) is the average value of the outage level k, k=1,…, 9, which can be looked up in Table 6.4.

Expected values of animal-caused outages in each input state for Bayesian models with 9 input states are shown in Table 6.5. For better observation of the trends in the expected values, we plot the expected values in Figure 6.3. From this figure, we can observe apparent increasing trend in the expected values of animal-related outages when the month type increase from 1 to 3. When the fair days level increases from 1 to 3, we observe the similar but not as obvious increasing trend in expected values of outages. However, for month type of 2, when the fair days level increases from 2 to 3, there is a slight decrease from 18.80 (at input state 5) to 18.12 (at input state 8) in the expected values of outages. This contradiction might be the result of saturation effect. That is, 18.80 outages per week are already high for April and December considering that there is not many animal activities in these months. More fair days cannot cause higher outages.

Table 6.5 Expected Values of Animal-related Outages for Wichita by Bayesian Model with 9 Input States

Outage Level	Month Type	Fair Day Level	Expected Number
Input State 1	1	1	9.60
Input State 2	2	1	18.46
Input State 3	3	1	28.87
Input State 4	1	2	10.73
Input State 5	2	2	18.80
Input State 6	3	2	36.22
Input State 7	1	3	12.22

Input State 8	2	3	18.12
Input State 9	3	3	42.78

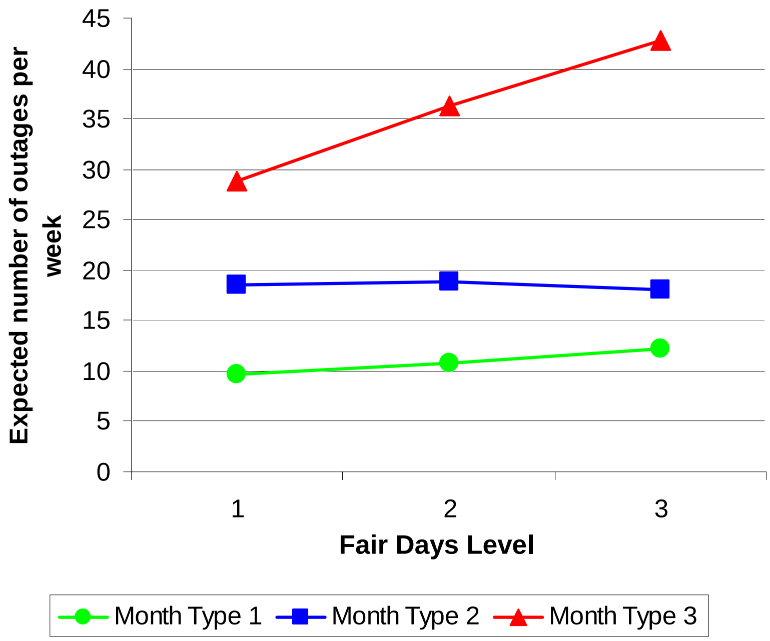

Figure 6.3 The Trends in Expected Values of Animal-related Outages for Wichita

We consider the expected value in any input state as the predicted value for the weeks with the same input state. Figure 6.4 shows the time series of predictions by Bayesian model with 9 input states for Wichita. As shown in Figure 6.4, this Bayesian model with 9 input states cannot capture the high peaks in the time series. In other word, it tends to underestimate in the months when the numbers of animal-related outages have been found to be high. Other than the underestimations the predicted values follow the observed values closely. A reason for such underestimations comes from the loss of information during classifications of outages. We are using the average values to represent an outage level during predictions and thus the higher observed values of the outages in one outage levels are ignored during predictions. To show performance of the Bayesian model without information loss of this kind, we can keep the outage levels as outputs instead of the numbers of outages. We just simply check the outage levels to which the expected values of outages in Table 6.5 belong and keep them as the expected outage levels which are shown in Table 6.6. The time series of predictions of outage levels for Wichita is shown in Figure 6.5. Comparing Figure 6.5 to Figure 6.4, we have observed better performance of predictions when the predictions are represented as outage levels instead of number of outages. We have observed overestimations in the year 2006 and 2007 both in Figure 6.4 and 6.5 which are because of the lower outages in these two years compared to the other years. A third input based on outages in previous week can possibly provide some corrections in the model.

Table 6.4 Expected Outage Levels for Wichita by Bayesian Model with 9 Input States

Input State	Month Type	Fair Day Level	Expected Outage Level
Input State 1	1	1	1
Input State 2	2	1	4
Input State 3	3	1	6
Input State 4	1	2	2
Input State 5	2	2	4
Input State 6	3	2	7
Input State 7	1	3	2
Input State 8	2	3	4
Input State 9	3	3	7

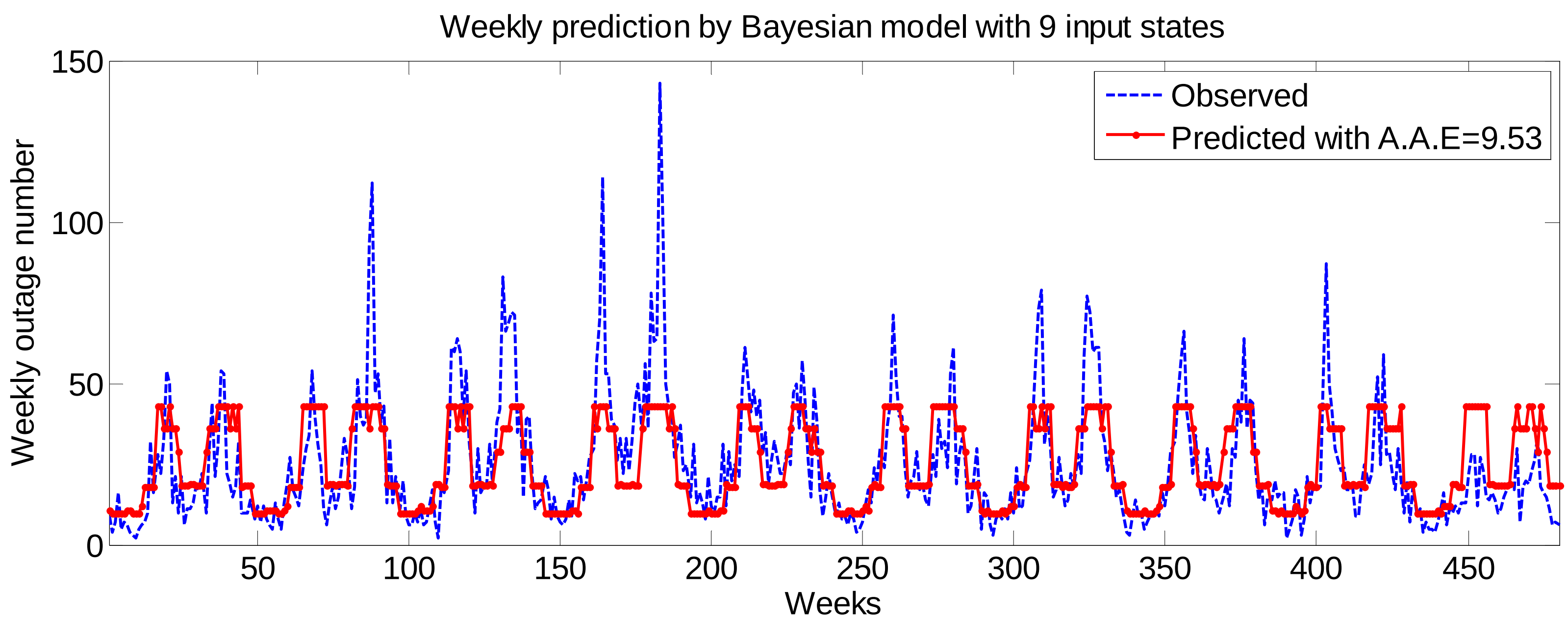

Figure 6.4 Outages Predicted and Observed by the Bayesian Model with 9 Input States for Wichita

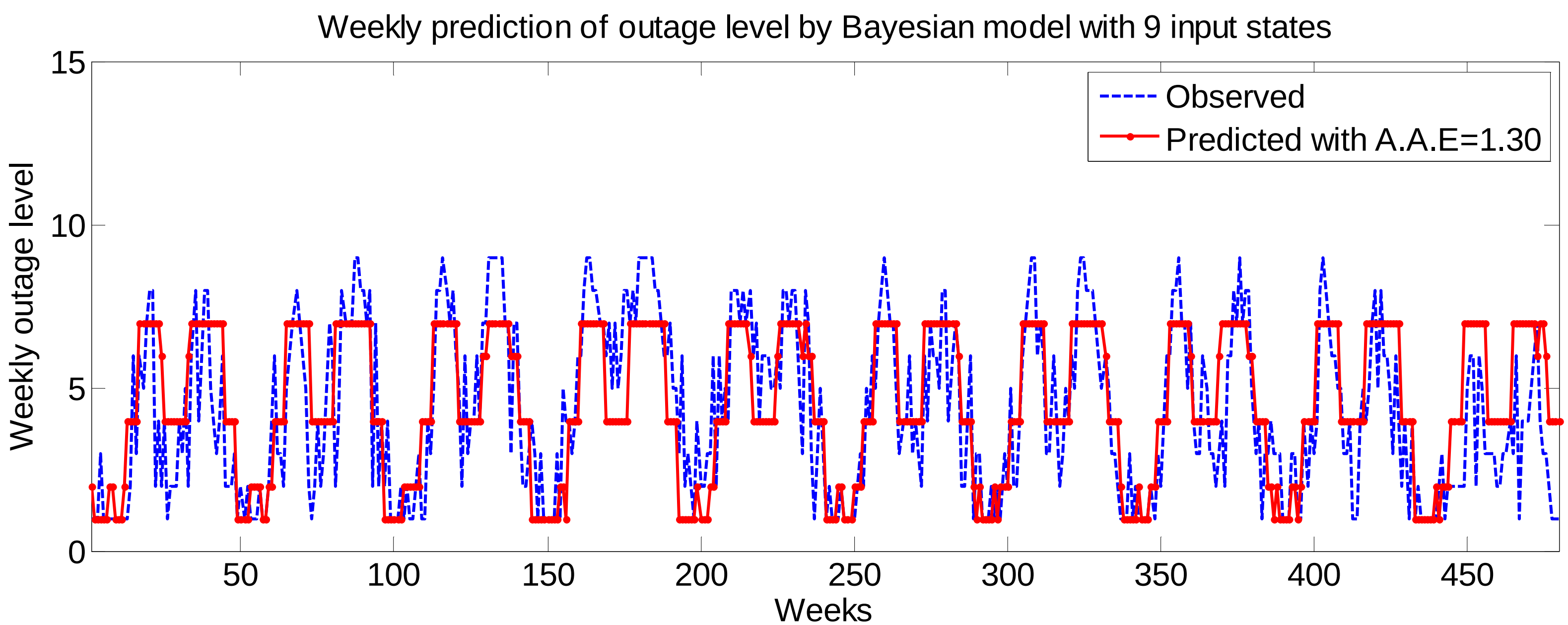

Figure 6.5 Outage Levels Predicted and Observed by the Bayesian Model with 9 Input States for

6.2.4 Models and Performance for More Cities

The Bayesian model has been built using Wichita as an example. In order to validate the model and compare the performance of the models for different cities, we have used the same approach for Topeka, Lawrence and Manhattan. The two inputs, the month type and the fair days level are the same as in the model for Wichita. The number of outage levels is 9 for all the cities. But the outage classification for each individual city differs from city to city, based on the histogram of historical outage data.

6.2.4.1 Model and Performance for Topeka

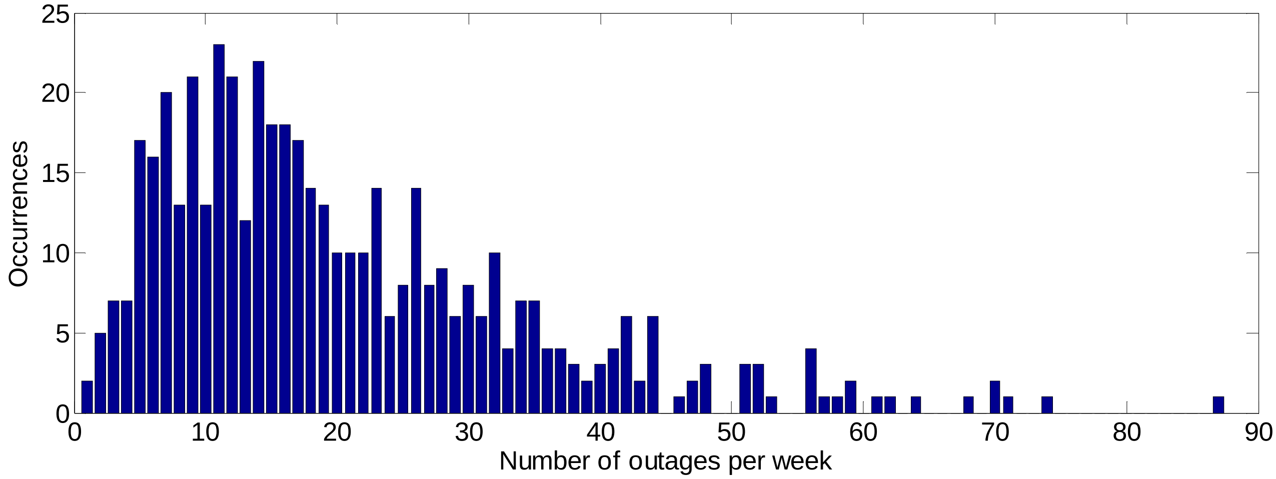

Figure 6.6 Histogram of Weekly Animal-related Outages from the Year 1998 to 2007 in Topeka

As mentioned in section 6.2.2, for uniformity and simplicity, 9 outage levels are used for models of all four cities. The outages range from 0 to 86. By following the same method introduced in section 6.2.2, we have obtained the outage levels for Topeka as:

- Outage Level 1: 0~7 Animal Caused Outages per Week;
- Outage Level 2: 8~10 Animal Caused Outages per Week;
- Outage Level 3: 11~13 Animal Caused Outages per Week;
- Outage Level 4: 14~16 Animal Caused Outages per Week;
- Outage Level 5: 17~19 Animal Caused Outages per Week;
- Outage Level 6: 20~24 Animal Caused Outages per Week;
- Outage Level 7: 25~29 Animal Caused Outages per Week;

- Outage Level 8: 30~36 Animal Caused Outages per Week ;
- Outage Level 9: 37~86 Animal Caused Outages per Week.

The numbers of observations for each input state are shown in Table 6.6 and the conditional probability table for Topeka is shown in Table 6.7. Similar observations have been found in the CPT for Topeka as in the one for Wichita. The average values for characterizing each input states are shown in Table 6.8.

Table 6.6 All Possible States and Number of Observations for Topeka

Input State	1	2	3	4	5	6	7	8	9
Month Type	1	2	3	1	2	3	1	2	3
Fair Day Level	1	1	1	2	2	2	3	3	3
Number of Occurrences	73	55	15	39	64	62	8	41	123

Table 6.7 The Conditional Probability Table with 9 Input States for Topeka

Outage Level	1	2	3	4	5	6	7	8	9
Input State 1	0.56	0.22	0.10	0.04	0.04	0.03	0.00	0.01	0.00
Input State 2	0.20	0.15	0.20	0.15	0.13	0.15	0.02	0.02	0.00
Input State 3	0.00	0.07	0.13	0.07	0.00	0.27	0.13	0.20	0.13
Input State 4	0.56	0.18	0.08	0.15	0.00	0.03	0.00	0.00	0.00
Input State 5	0.08	0.17	0.17	0.20	0.14	0.16	0.05	0.03	0.00
Input State 6	0.03	0.05	0.06	0.10	0.06	0.13	0.16	0.18	0.23
Input State 7	0.00	0.25	0.50	0.13	0.00	0.00	0.13	0.00	0.00
Input State 8	0.12	0.17	0.27	0.15	0.20	0.05	0.02	0.02	0.00
Input State 9	0.01	0.02	0.02	0.07	0.05	0.11	0.22	0.19	0.33

Table 6.8 Average Values for Each Outage Level for Topeka

Outage Level	1	2	3	4	5	6	7	8	9
Average Value	4.59	9.04	12.02	14.98	17.89	21.83	26.69	32.64	49.14

By following the same method we have used to compute the expected values of outages for Wichita, we have obtained the expected values of animal-caused outages in each input state for Topeka which are shown in Table 6.9. For better observations of the trends in the expected values, we plot the expected values in Figure 6.7. In this figure, we can observe increasing trends in animal-related outages when the month type and the number of fair days per week increase, which is similar to the results for Wichita. Like Wichita, there is a slight decrease from input state 5 to input state 8 in the expected values of outages.

Table 6.9 Expected Values and Levels of Animal-related Outages for Topeka by the Bayesian Model with 9 Input States

Outage Level	Month Type	Fair Day Level	Expected Value	Expected Outage Level
Input State 1	1	1	8.10	2
Input State 2	2	1	13.35	3
Input State 3	3	1	25.67	7
Input State 4	1	2	8.00	1
Input State 5	2	2	15.22	4
Input State 6	3	2	27.97	7
Input State 7	1	3	13.48	3
Input State 8	2	3	13.52	3
Input State 9	3	3	32.60	8

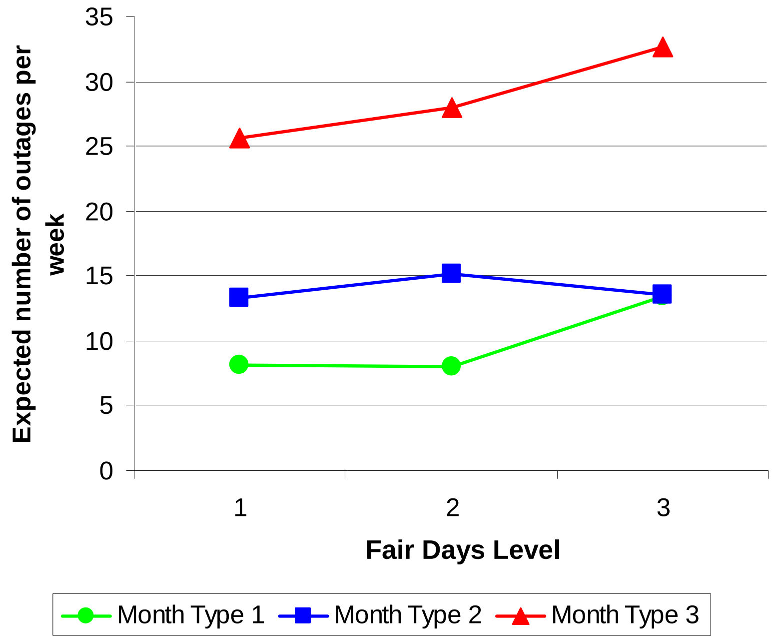

Figure 6.7 The Trends in Expected Values of Animal-related Outages for Topeka

Similar to what we have done for Wichita, we take the expected value in any input state as the predicted value for the weeks with the same input state. Figure 6.8 shows the time series of predictions and Figure 6.9 shows the predictions of outage levels for Topeka. The underestimations in months with high outages exist in the predictions for Topeka just like Wichita. Again, we have observed better performance of predictions when the predictions are represented as the outage levels instead of the number of outages. Unlike Wichita, we have not observed overestimations in the year 2007 both in Figure 6.8 and 6.9. It is because the variance in the outages for Topeka is not as great as the one for Wichita.

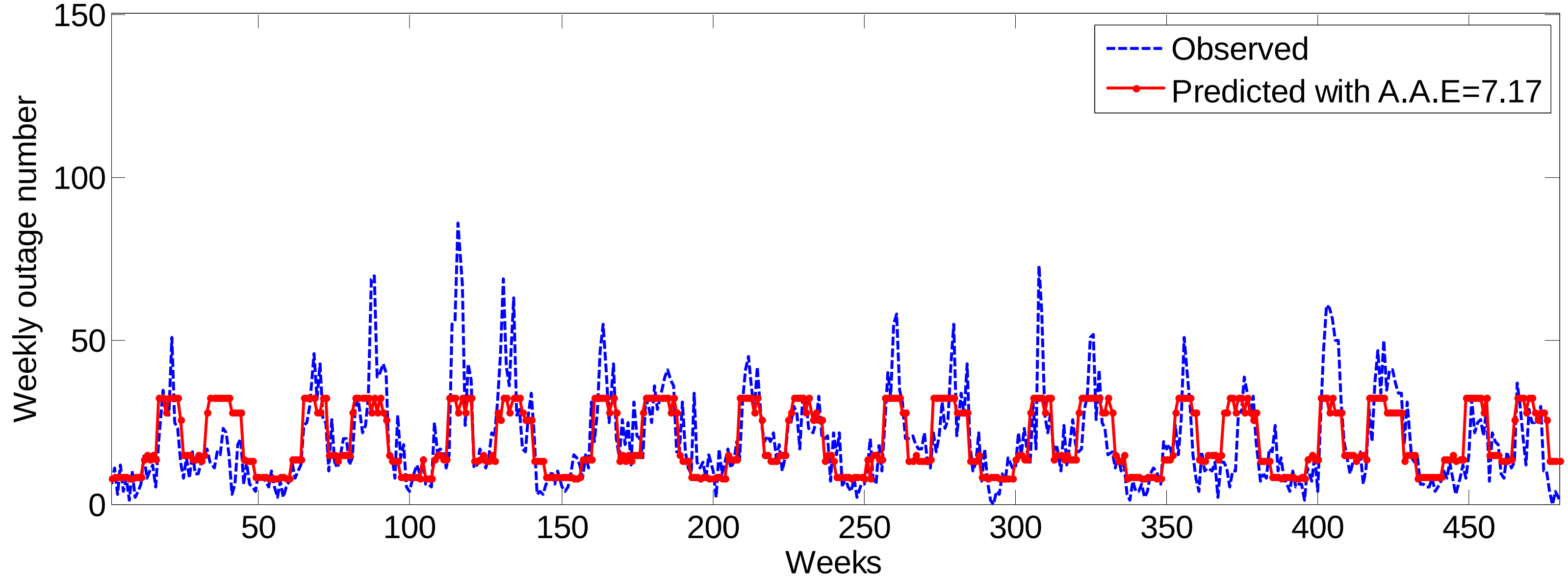

Figure 6.8 Outages Predicted and Observed by the Bayesian Model with 9 Input States for Topeka

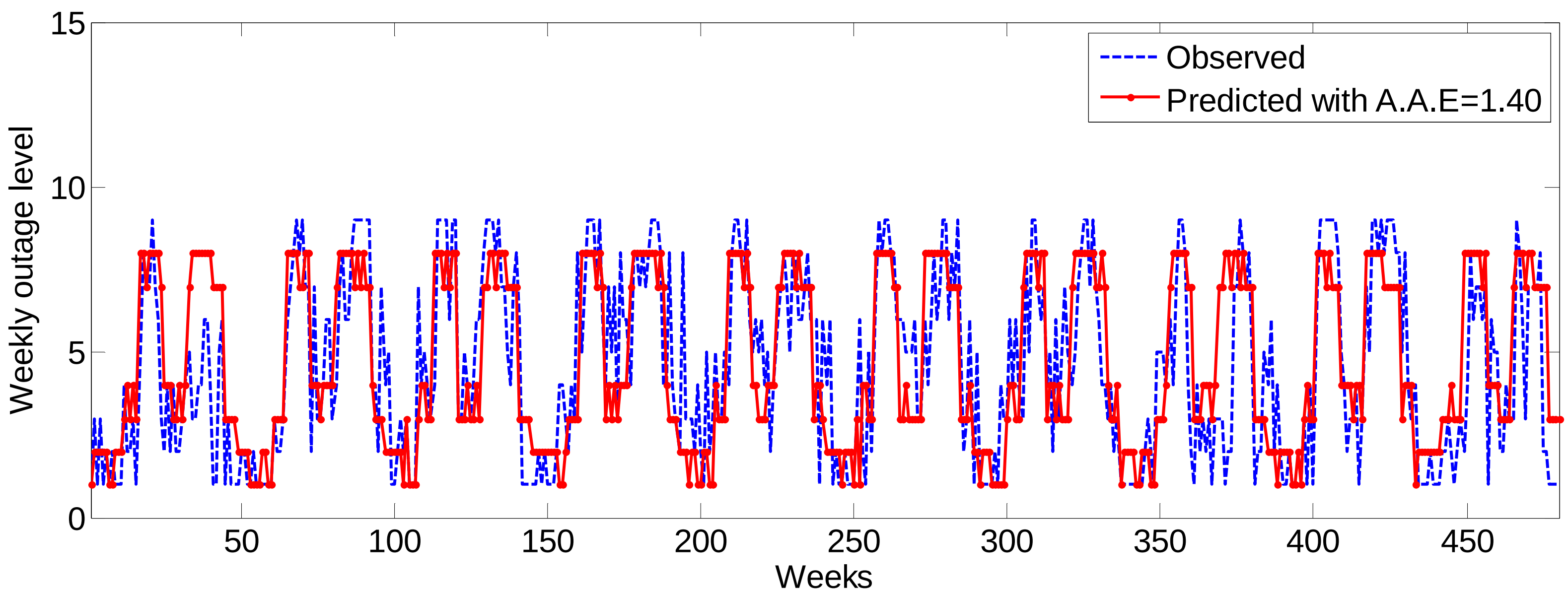

Figure 6.9 Outage Levels Predicted and Observed by the Bayesian Model with 9 Input States for Topeka

6.2.4.2 Model and Performance for Lawrence

The histogram of weekly animal-related outages in Lawrence from the year 1998 to 2007 is shown in Figure 6.10.

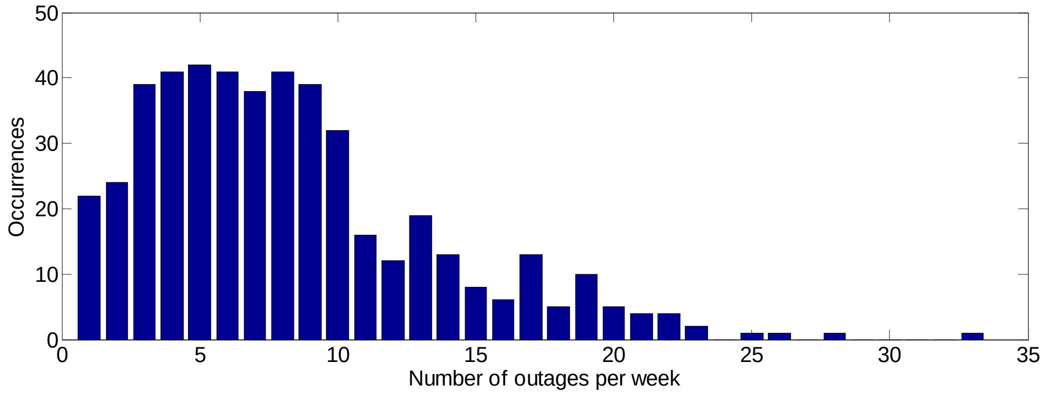

Figure 6.10 Histogram of Weekly Animal-related Outages from Year 1998 to 2007 in Lawrence

The outages range from 0 to 32. Since the size of Lawrence is smaller than the one of Topeka and Wichita, there are more occurrences of lower outages. By following the same method introduced in section 6.2.2, we have obtained the outage levels for Lawrence as:

- Outage Level 1: 0~3 Animal Caused Outages per Week;
- Outage Level 2: 4~5 Animal Caused Outages per Week;
- Outage Level 3: 6 Animal Caused Outages per Week;
- Outage Level 4: 7~8 Animal Caused Outages per Week;
- Outage Level 5: 9~10 Animal Caused Outages per Week;
- Outage Level 6: 11~12 Animal Caused Outages per Week;
- Outage Level 7: 13~15 Animal Caused Outages per Week;
- Outage Level 8: 16~20 Animal Caused Outages per Week ;
- Outage Level 9: 21~33 Animal Caused Outages per Week.

The numbers of observations for every input state are shown in Table 6.10 and the conditional probability table for Lawrence is shown in Table 6.11. Similar observations have been found in the CPT for Lawrence as in the ones for Wichita and Topeka. The average values for characterizing each input states are shown in Table 6.12.

Table 6.10 All Possible States and Number of Observations for Lawrence

Input State	1	2	3	4	5	6	7	8	9
Month Type	1	2	3	1	2	3	1	2	3
Fair Day Level	1	1	1	2	2	2	3	3	3
Number of Occurrences	74	56	18	38	64	67	8	40	115

Table 6.11 The Conditional Probability Table with 9 Input States for Lawrence

Outage Level	1	2	3	4	5	6	7	8	9
Input State 1	0.61	0.20	0.04	0.08	0.05	0.00	0.01	0.00	0.00
Input State 2	0.38	0.18	0.18	0.16	0.07	0.02	0.02	0.00	0.00
Input State 3	0.00	0.11	0.06	0.17	0.28	0.17	0.11	0.11	0.00
Input State 4	0.50	0.24	0.08	0.13	0.03	0.00	0.03	0.00	0.00
Input State 5	0.14	0.25	0.13	0.30	0.09	0.06	0.03	0.00	0.00
Input State 6	0.04	0.09	0.06	0.12	0.21	0.10	0.13	0.16	0.07
Input State 7	0.75	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 8	0.33	0.35	0.05	0.18	0.08	0.03	0.00	0.00	0.00
Input State 9	0.09	0.08	0.06	0.20	0.10	0.13	0.10	0.21	0.04

Table 6.12 Average Values for Each Outage Level for Lawrence

Outage Level	1	2	3	4	5	6	7	8	9
Average Value	1.79	4.49	6.00	7.49	9.33	11.61	13.74	17.51	23.60

By following the same method we have used previously to compute the expected values of outages, we have obtained the expected values of animal-caused outages in each input state for Lawrence which are shown in Table 6.13. For better observations of the trends in the expected values, we plot them in Figure 6.11. We can observe increasing trends in animal-related outages when the month type increases from 1 to 3 as in Topeka and Wichita. Also we can observe increasing trends in animal-related outages when the fair days level increases from 1 to 2. However, the expected values of outages decrease when the fair days level increases from 2 to 3. A reason might lie in the outages data for Lawrence. For some years between 1998 and 2007, there are higher outages in January than in April. Errors in weather information data could also cause this problem.

Table 6.13 Expected Values and Levels of Animal-related Outages for Lawrence by Bayesian Model with 9 Input States

Outage Level	Month Type	Fair Day Level	Expected Value	Expected Outage Level
Input State 1	1	1	3.54	1
Input State 2	2	1	4.87	2
Input State 3	3	1	10.08	5
Input State 4	1	2	4.02	2
Input State 5	2	2	6.38	3
Input State 6	3	2	11.38	6
Input State 7	1	3	2.46	1
Input State 8	2	3	4.75	2
Input State 9	3	3	10.77	5

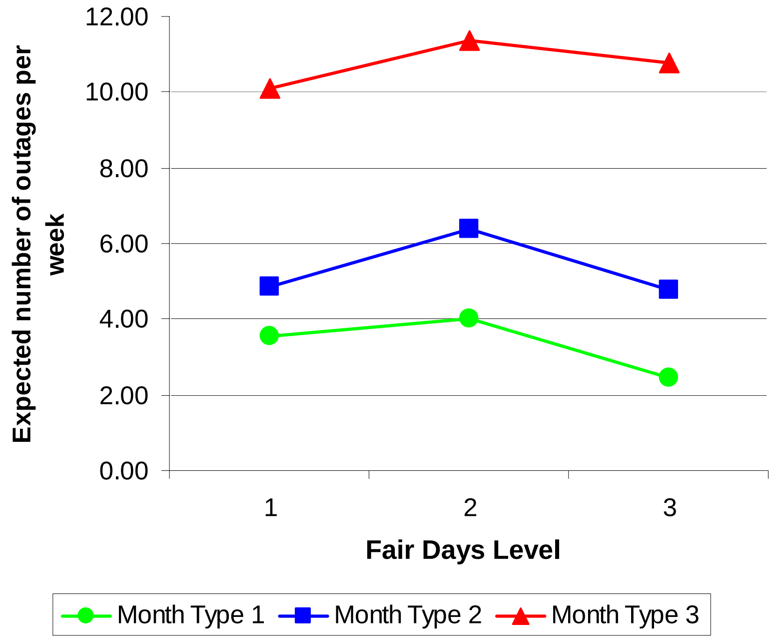

Figure 6.11 The Trends in Expected Values of Animal-related Outages for Lawrence

Again, we take the expected value in any input state as the predicted value for the weeks with the same input state. Figure 6.12 shows the time series of predictions and Figure 6.13 shows the predictions of outage levels for Lawrence. The underestimations in months with high outages exist in the prediction for Lawrence similar to that for Topeka and Wichita. Again, we have observed better performance of predictions when the predictions are represented as the outage levels instead of the number of outages.

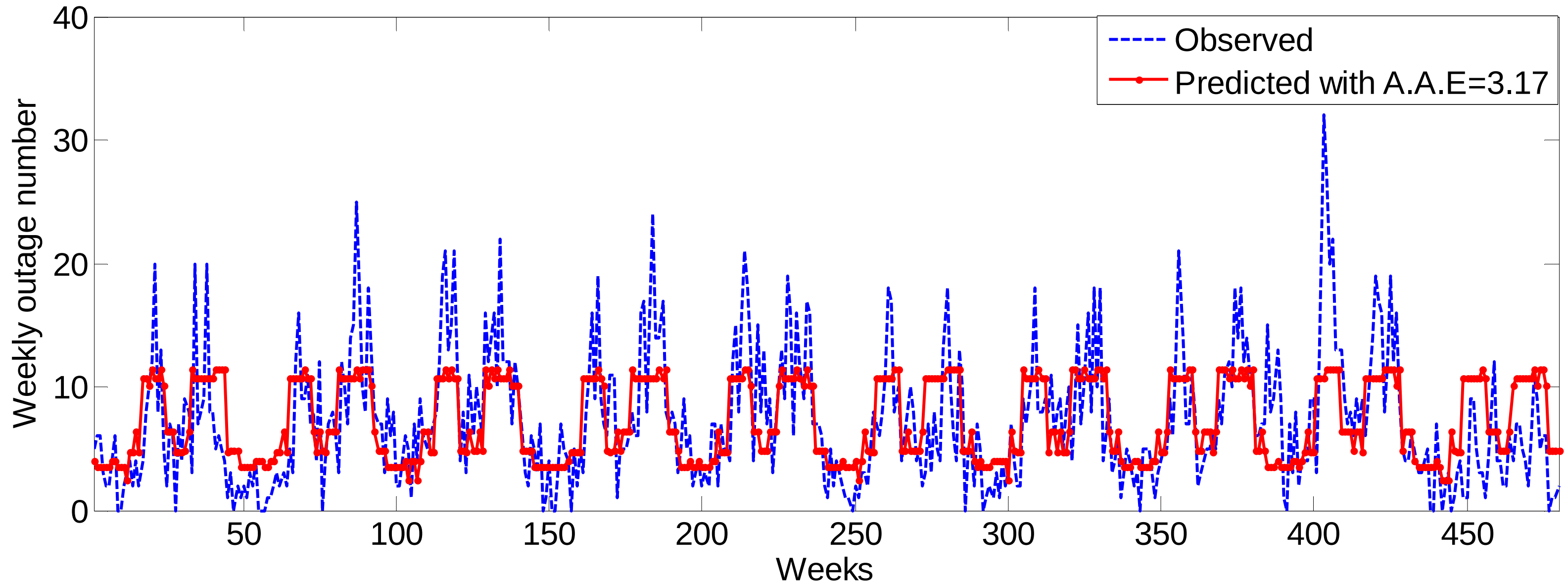

Figure 6.12 Outages Predicted and Observed by the Bayesian Model with 9 Input States for Lawrence

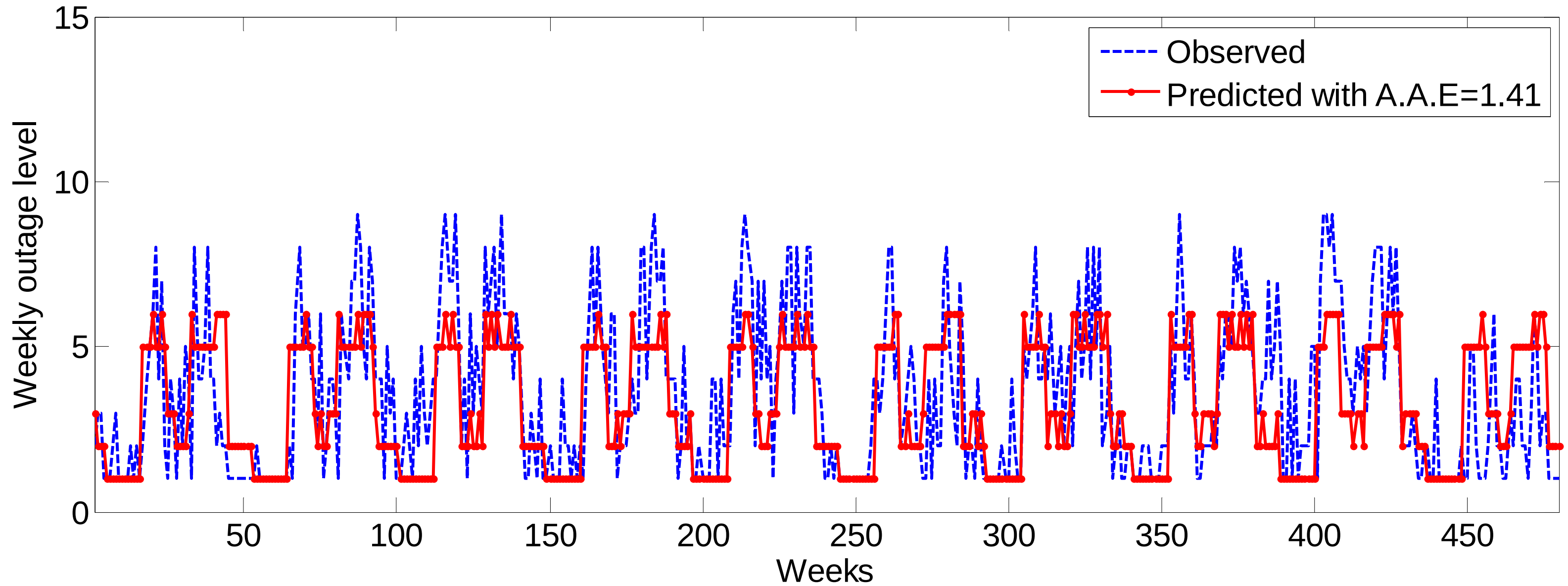

Figure 6.13 Outage Levels Predicted and Observed by the Bayesian Model with 9 Input States for Lawrence

6.2.4.3 Model And Performance For Manhattan

The histogram of weekly animal-related outages in Manhattan from the year 1998 to 2007 is shown in Figure 6.14.

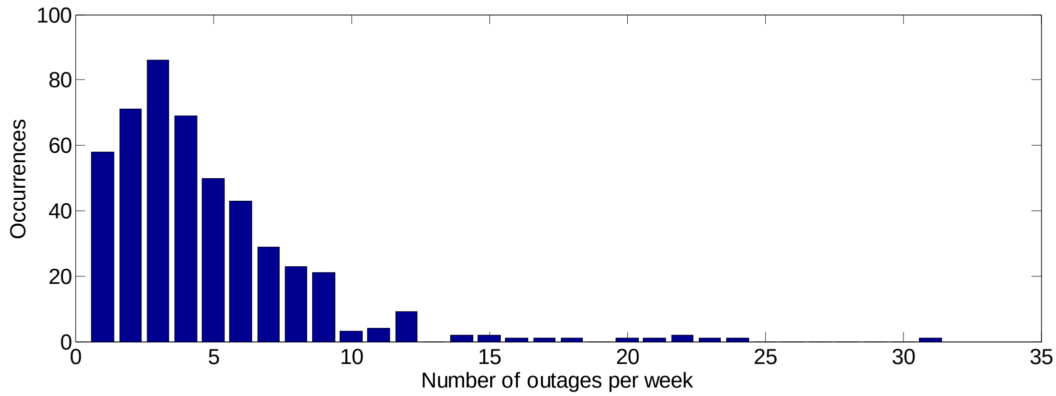

Figure 6.14 Histogram of Weekly Animal-related Outages from the Year 1998 to 2007 in Manhattan

The outages range from 0 to 30. Since most outages are low in Manhattan and the biggest bar has 86 occurrences, we cannot manage to classify the outages into 9 bins with equal number of occurrences in each bin. Therefore we use a different method for Manhattan from the ones we have used for other cities. We have kept the bin size as 1 for the first 4 bins. We have obtained the outage levels for Manhattan as:

- Outage Level 1: 0 Animal Caused Outages per Week;
- Outage Level 2: 1 Animal Caused Outages per Week;
- Outage Level 3: 2 Animal Caused Outages per Week;
- Outage Level 4: 3 Animal Caused Outages per Week;
- Outage Level 5: 4~6 Animal Caused Outages per Week;
- Outage Level 6: 7~8 Animal Caused Outages per Week;
- Outage Level 7: 9~13 Animal Caused Outages per Week;
- Outage Level 8: 14~24 Animal Caused Outages per Week ;
- Outage Level 9: 25~30 Animal Caused Outages per Week.

The number of observations for each input state are shown in Table 6.14 and the conditional probability table for Manhattan is shown in Table 6.15. Similar observations have been found in the CPT for Manhattan as in the ones for other three cities. The average values for characterizing each input states are shown in Table 6.16.

Table 6.14 All Possible States and Number of Observations for Manhattan

Input State	1	2	3	4	5	6	7	8	9
Month Type	1	2	3	1	2	3	1	2	3
Fair Day Level	1	1	1	2	2	2	3	3	3
Number of Occurrences	92	69	21	24	57	79	4	34	100

Table 6.15 The Conditional Probability Table with 9 Input States for Manhattan

Outage Level	1	2	3	4	5	6	7	8	9
Input State 1	0.24	0.18	0.27	0.16	0.12	0.02	0.00	0.00	0.00
Input State 2	0.13	0.28	0.16	0.16	0.19	0.07	0.01	0.00	0.00
Input State 3	0.00	0.00	0.19	0.10	0.38	0.29	0.00	0.05	0.00
Input State 4	0.29	0.21	0.38	0.08	0.04	0.00	0.00	0.00	0.00
Input State 5	0.05	0.21	0.18	0.25	0.23	0.07	0.02	0.00	0.00
Input State 6	0.09	0.06	0.09	0.09	0.34	0.18	0.10	0.04	0.01
Input State 7	0.25	0.25	0.25	0.00	0.25	0.00	0.00	0.00	0.00
Input State 8	0.18	0.12	0.26	0.18	0.24	0.03	0.00	0.00	0.00
Input State 9	0.03	0.08	0.10	0.12	0.40	0.12	0.08	0.07	0.00

Table 6.16 Average Values for Each Outage Level for Manhattan

Outage Level	1	2	3	4	5	6	7	8	9
Average Value	0.00	1.00	2.00	3.00	4.83	7.48	10.67	18.36	30.00

The expected values of animal-caused outages in each input state for Manhattan are shown in Table 6.17. The expected values are shown in Figure 6.15. We can again observe increasing trends in animal-related outages when the month type increases from 1 to 3 as in other three cities. However, there is no obvious increasing trend in the expected outages when the fair days level increases. At month type of 3, the expected values of outages are almost the same when the fair days level increases from 1 to 3. At month type of 2 the expected number of outage decreases from 3.11 to 2.53 when the fair days level increases from 2 to 3. At month type of 1, expected number of outages decreases from 1.96 to 1.41 when the fair days level increases from 1 to 2. The same problems exist in the outages and weather data for Manhattan as that for Lawrence. Topeka and Wichita are the two biggest cities in the service area of Westar Energy and both have sufficient observations of animal-related outages. Manhattan and Lawrence are the smaller cities and thus there are not as many observations of animal-related outages, which could cause problems in predictions by the Bayesian models.

Table 6.17 Expected Values and Levels of Animal-related Outages for Manhattan by Bayesian Model with 9 Input States

Outage Level	Month Type	Fair Day Level	Expected Number	Expected Outage Level
Input State 1	1	1	1.96	2
Input State 2	2	1	2.68	3
Input State 3	3	1	5.52	5
Input State 4	1	2	1.41	2
Input State 5	2	2	3.11	4
Input State 6	3	2	5.64	5
Input State 7	1	3	1.96	2
Input State 8	2	3	2.53	3
Input State 9	3	3	5.61	5

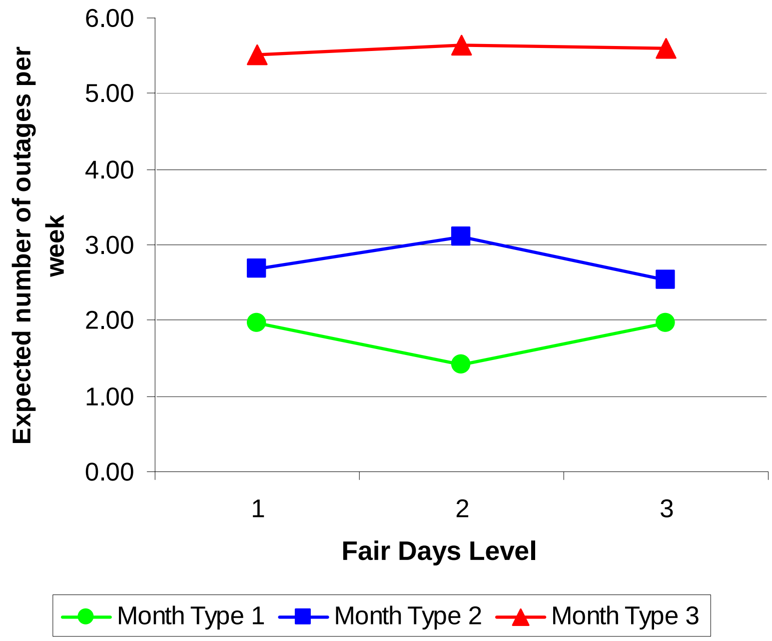

Figure 6.15 The Trends in Expected values of Animal-related Outages for Manhattan

Again, we take the expected value in any input state as the predicted value for the weeks with the same input state. Figure 6.16 shows the time series of predictions and Figure 6.17 shows the predictions of outage levels for Manhattan. The underestimations in months with high outages exist in the predictions for Manhattan as well as other three cities. Again, we have observed better performance of predictions when the predictions are represented as the outage level instead of the number of outages.

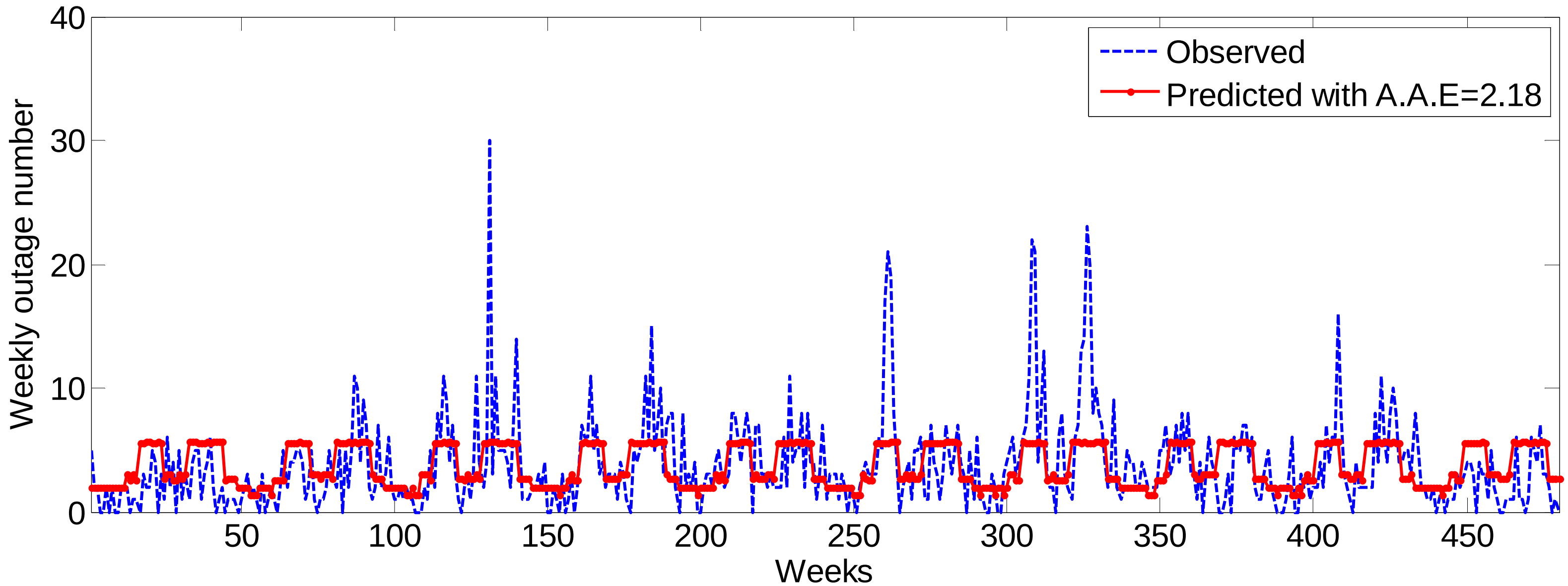

Figure 6.16 Outages Predicted and Observed by the Bayesian Model with 9 Input States for Manhattan

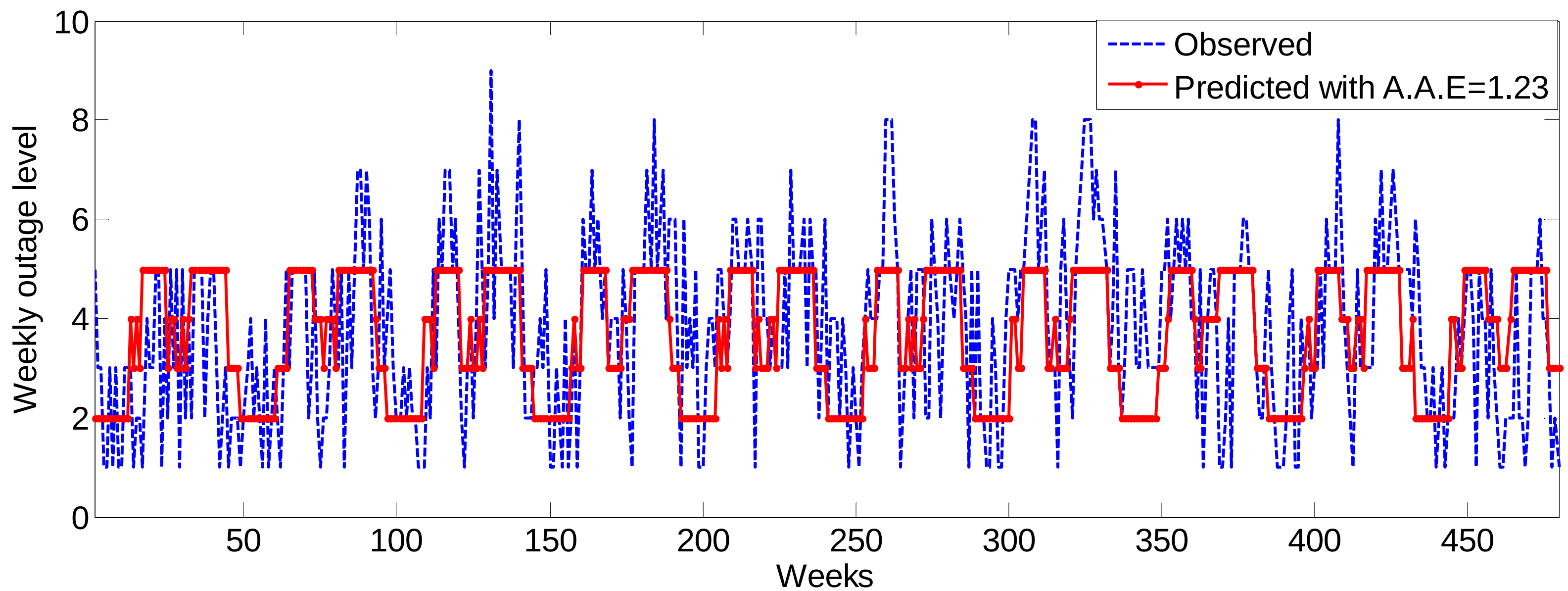

Figure 6.17 Outage Levels Predicted and Observed by the Bayesian Model with 9 Input States for Manhattan

6.3 Adding One More Input to the Model

In the same input state, the outages vary in a certain range because in different years the outages have different ranges even though the patterns are similar for each year. It is the reason that has caused underestimations for all four cities by the Bayesian models with two inputs. As we can see in Figure 6.18, the outages for Wichita in the last year are significantly lower than in the other years and the outages in the year 2001 are significantly higher than in other years. Therefore, it's a good idea to take the previous week's outages as the third input besides the month type and the fair days level.

To keep the conditional probability table from being too sparse, it is better to classify the outages in previous week into only two levels which consist of "High" and "Low", or three levels which consist of "High" , "Medium" and "Low". Recall that the outages of previous week are taken as inputs for the neural network models too, which yielded better results. The third input variable increases the computation load but at the same time it increases the performance of the model. The structure of the Bayesian model with three inputs is shown in Figure 6.19. With this structure, we have a model with 18 input states when the third input has two discrete values, and a model with 27 input states when the third input takes on three discrete values.

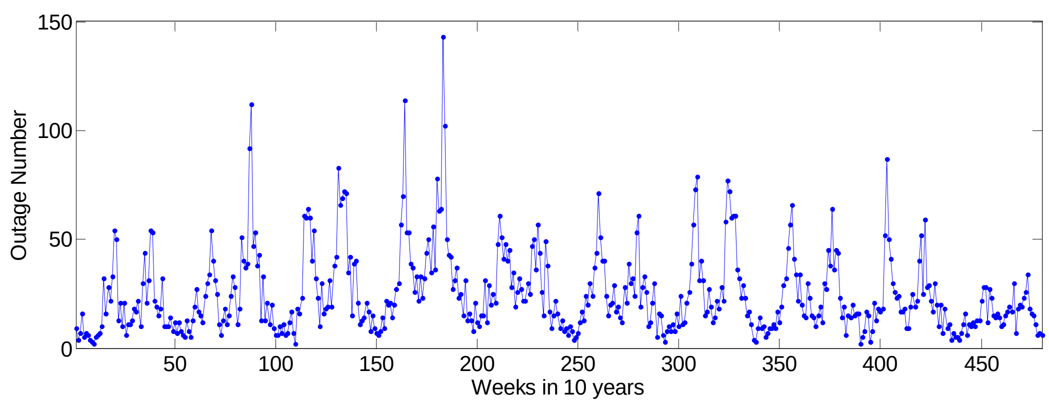

Figure 6.18 Weekly Animal-related Outages in Wichita from Year 1998 to 2007

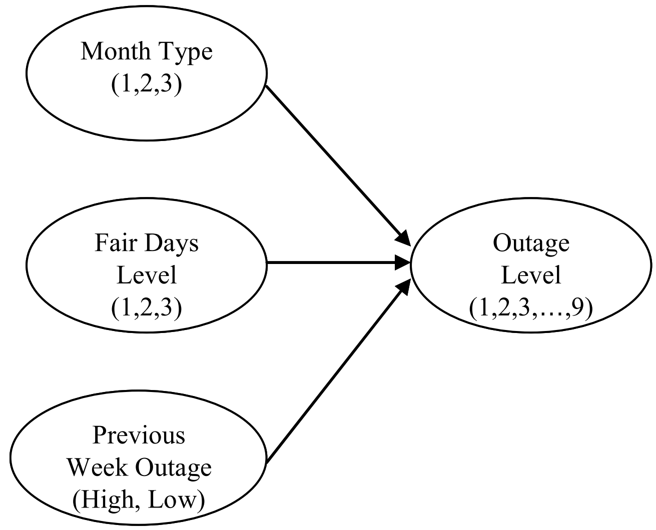

Figure 6.19 One-layer Bayesian Network with Three Inputs for Prediction of Animal-related Outages

6.3.1 Model with 18 Inputs

For the model with 18 input states, classifications of the outage levels, the month type and the fair days levels are the same as the model with 9 inputs and the model structure is the same except there is one more input node, which is previous week outage level. Wichita is taken as an example for modeling. Experiments were done to find the optimal values for the thresholds to be used for the outage levels in the previous week. AAE for predictions of animal-related outages with different values of threshold are shown in Table 6.18. The method used for predictions is the same as the one used for the model with 9 input states. According to the results in Table 6.18, the threshold to define "High" outage level in the previous week is set at the optimal value of 70th percentile, which means number of outages higher than 70% of the weeks in 480 weeks are defined as "High".

The training cases for each input state are given in Table 6.19 from which we can see that for 18 input states there are two input states with all zero entries. The zero entries happen because there are no sample data or the combination of inputs for these states never happens. For

Table 6.18 The Performance of the Model with Different Thresholds in Model with 18 Input States for Wichita

Threshold	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
AAE	9.53	9.37	9.37	9.11	9.1	8.84	8.81	8.31	8.57	8.87

example, input states 13 and 16 have month type 1 and high outage level in previous week which are rare. It is because month type of 1 is too cold and thus outages are low. These input states with all zero entries result in the all zero rows in the conditional probability table as shown in Table 6.20. The input states with all zero entries have potential deficiency during prediction stage because if they occur during the prediction stage we will have no idea what the corresponding outage level should be. Also, there are some cases with only 1 or 2 observations such as input states 10 and 17, which result in possibly unreliable predictions since the conditional probability table is computed based only on 1 or 2 observations. Totally, there are 4 input states have zero or only a few observations. Comparing Table 6.20 to Table 6.3, we can see the conditional probability table for the model with 18 input states is sparser than the one for the model with 9 input states.

Table 6.19 All 18 Possible Input States and Number of Observations in Each State for Wichita

	Month Type	Fair Day Level	Previous Week Outage	Number of Observations
Input State 1	1	1	Low	73
Input State 2	2	1	Low	58
Input State 3	3	1	Low	11
Input State 4	1	2	Low	34
Input State 5	2	2	Low	46
Input State 6	3	2	Low	30
Input State 7	1	3	Low	12

Input State 8	2	3	Low	35
Input State 9	3	3	Low	47
Input State 10	1	1	High	1
Input State 11	2	1	High	13
Input State 12	3	1	High	9
Input State 13	1	2	High	0
Input State 14	2	2	High	6
Input State 15	3	2	High	34
Input State 16	1	3	High	0
Input State 17	2	3	High	2
Input State 18	3	3	High	69

Table 6.20 The Conditional Probability Table with 18 Input States for Wichita

Outage Level	1	2	3	4	5	6	7	8	9
Input State 1	0.62	0.18	0.15	0.04	0.00	0.01	0.00	0.00	0.00
Input State 2	0.14	0.22	0.26	0.19	0.10	0.05	0.03	0.00	0.00
Input State 3	0.00	0.18	0.27	0.00	0.00	0.18	0.36	0.00	0.00
Input State 4	0.56	0.15	0.18	0.06	0.03	0.03	0.00	0.00	0.00
Input State 5	0.09	0.24	0.22	0.13	0.09	0.17	0.07	0.00	0.00
Input State 6	0.03	0.07	0.10	0.20	0.13	0.27	0.07	0.13	0.00
Input State 7	0.25	0.33	0.33	0.08	0.00	0.00	0.00	0.00	0.00
Input State 8	0.03	0.20	0.23	0.29	0.11	0.14	0.00	0.00	0.00
Input State 9	0.00	0.06	0.04	0.15	0.17	0.30	0.11	0.17	0.00
Input State 10	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 11	0.00	0.08	0.00	0.23	0.31	0.15	0.08	0.15	0.00
Input State 12	0.00	0.00	0.22	0.00	0.11	0.11	0.33	0.22	0.00
Input State 13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 14	0.00	0.33	0.17	0.17	0.00	0.33	0.00	0.00	0.00
Input State 15	0.00	0.03	0.00	0.06	0.06	0.15	0.29	0.24	0.18
Input State 16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Input State 17	0.00	0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00
Input State 18	0.00	0.00	0.00	0.01	0.07	0.07	0.22	0.38	0.25

Table 6.21 All 18 Possible Input States and Number of Observations in Each State for Wichita

	Month Type	Fair Day Level	Previous Week Outage	Expected Number of Outages	Expected Outage Level
Input State 1	1	1	Low	9.57	1
Input State 2	2	1	Low	16.35	3
Input State 3	3	1	Low	25.24	5
Input State 4	1	2	Low	10.73	2
Input State 5	2	2	Low	18.73	4
Input State 6	3	2	Low	26.72	6
Input State 7	1	3	Low	12.22	2
Input State 8	2	3	Low	18.39	4
Input State 9	3	3	Low	29.76	6
Input State 10	1	1	High	11.39	2
Input State 11	2	1	High	27.87	6
Input State 12	3	1	High	33.31	7
Input State 13	1	2	High	0.00	1
Input State 14	2	2	High	19.35	4
Input State 15	3	2	High	44.60	8
Input State 16	1	3	High	0.00	1
Input State 17	2	3	High	13.45	2
Input State 18	3	3	High	51.66	8

To compute the expected number of outages in each input states, we use the same method as the one used for the model with 9 input states. Since the classifications of outages are the same as the model with 9 input states, the average values of outages used in each level to multiply the CPT are the same as used for the model with 9 input states which are tabulated in Table 6.9 for Wichita. The expected number and levels of outages for each input state are computed and listed in Table 6.21. The trends in expected outages in each state are shown in Figure 6.20 and 6.21. In Figure 6.20, the third input, the outage level of previous week, is "Low" and thus the input states range from 1 to 9. In Figure 6.21, the third input, the outage level of previous week, is "High" and thus the input states range from 10 to 18.

In Figure 6.20, it is apparent that the expected values of animal-related outages show an increasing trend with increase of the month type from 1 to 3. Though the expected values of animal-related outages also increase with increase in the levels of fair days per week, there is a slight inconsistency when the month type is 2. In this case, the month type is the same, but even though the level of the fair weather days per week is increasing from 2 to 3, the expected number of animal-caused outages is almost the same. This might be because of the effect of saturation. In Figure 6.21, we have observed an increasing trend in the expected outages at the month type of 3. But it is not the case at the month type of 1 and 2. It is mainly because of lack of training data. The zero expected outage at month type 1 results from the zero number of training cases. The decreasing trend at the month type of 2 results from the fact that there are only 13, 6 and 2 training cases for input states 11, 14 and 17.

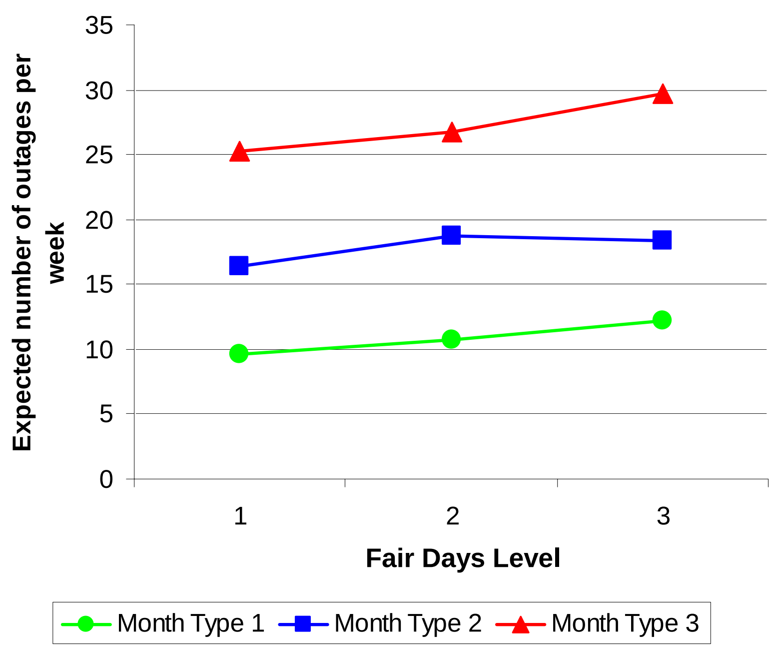

Figure 6.20 The Trends in Expected Values of Outages with Previous Week Outage as Low for Wichita

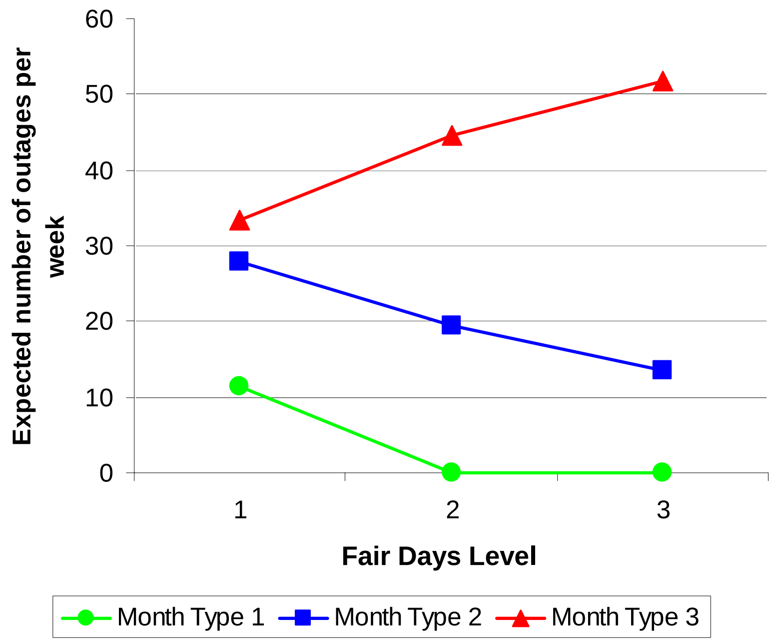

Figure 6.21 The Trends in Expected Values of Outages with Previous Week Outage as High for Wichita

Again, we take the expected value in each input state as the prediction for the weeks with the same input state. Figure 6.22 shows the time series of predicted numbers of outages and Figure 6.23 shows the predictions of outage levels for Wichita from the year 1998 to 2007. The model with 18 input states outperforms the one with 9 input states. The addition of the third input brings down the AAE from 9.54 to 8.31. In Figure 6.22, there are not as many underestimations in months with high outages as in Figure 6.4, which shows the predictions by model with 9 input states. Also there is no overestimation for the year 2007. Once again, we have observed better performance of predictions when the predictions are represented as the outage levels instead of the numbers of outages.

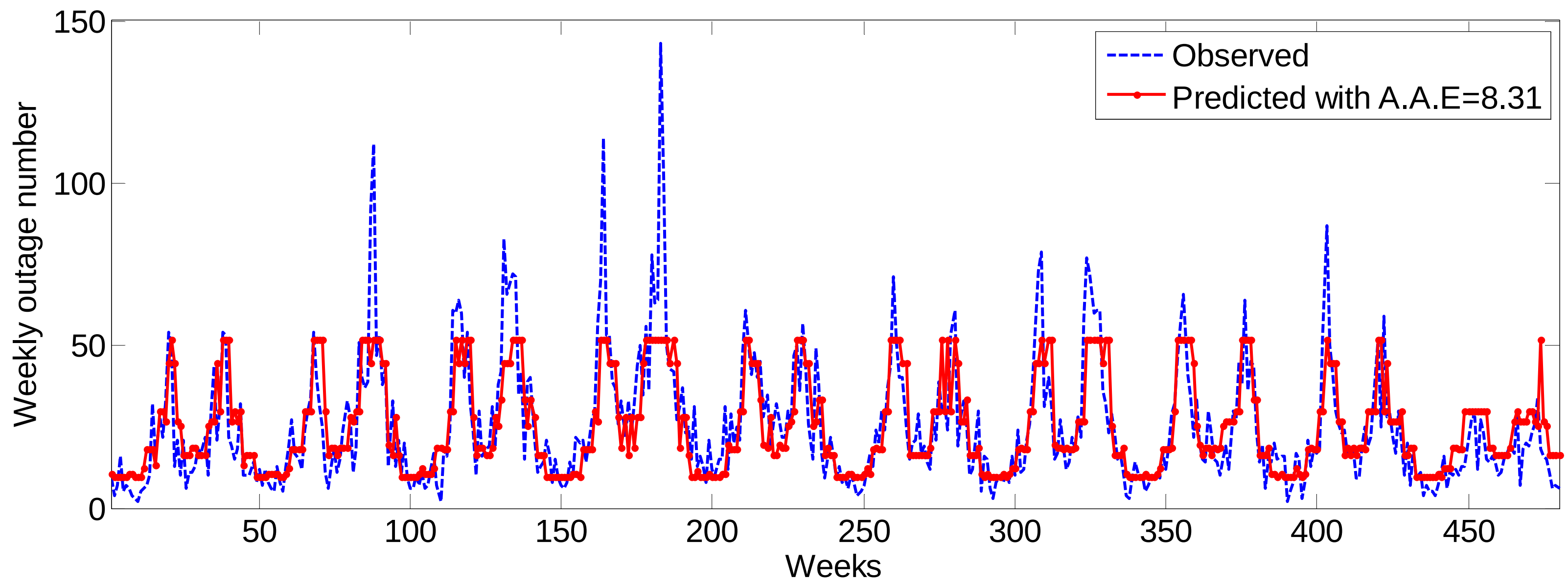

Figure 6.22 Outages Predicted and Observed by the Bayesian Model with 18 Input States for Wichita

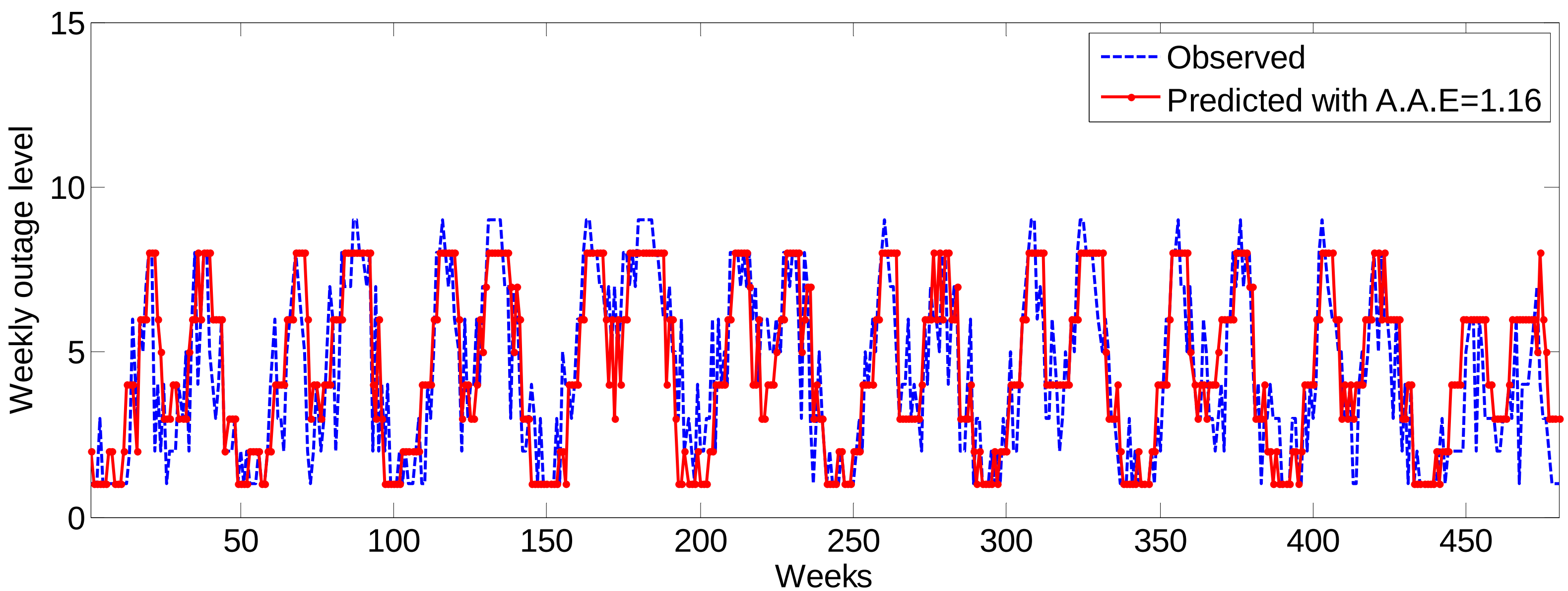

Figure 6.23 Outage Levels Predicted and Observed by the Bayesian Model with 18 Input States for Wichita

6.3.2 Model With 27 Input States

For the model with 27 input states, the classifications of outages, the month type and the number of fair days per week are the same as the models with 9 and 18 input states and the model structure is the same with the model with 18 input states. The only difference is that the third input is classified into three levels instead of two in the model with 18 input states. The same experiments are done to find the most suitable thresholds used to determine the outage level of the previous week as an input. The AAE of the predictions with different combination of values for two thresholds are shown in Table 6.22. The values in the first row are the possible values for the first threshold used to determine the "Low" outages in previous week. The values in the first column are the possible values for the second threshold used to determine the "High" outages in previous week. Intuitively, at most half of the outages in the total 480 weeks can be classified as "Low" and at most half of the outages in the total 480 weeks can be classified as "High". Therefore, six values from 0.1 to 0.5 for the first threshold and six values from 0.5 to 0.9 for the second threshold are tested. According to the results in Table 6.21, the thresholds are set at the optimal values of 70th and 30th percentile, which means outages in a week higher than 70% of all the weeks in 480 weeks are defined as "High", outages lower than 30% of all the weeks are defined as "Low" and the rest are defined as "Medium".

Table 6.22 The AAE of the Predictions with Different Values for the Threshold in Model with 27 Input States

Thresholds	0.10	0.20	0.30	0.33	0.40	0.50
0.50	8.743	8.738	8.701	8.700	8.786	8.844
0.60	8.687	8.696	8.633	8.645	8.720	8.733
0.66	8.644	8.655	8.568	8.576	8.636	8.649
0.70	8.177	8.195	8.109	8.111	8.180	8.190
0.80	8.433	8.443	8.320	8.321	8.371	8.319
0.90	8.544	8.552	8.372	8.371	8.400	8.255

The training cases for each input state are given in Table 6.23 from which we can see that for 27 input states there are four input states with all zero entries. These input states with all zero entries result in the rows with all zeros in the conditional probability table as shown in Table

6.24. Also there are four input states with only 1 or 2 observations, such as input state 3, 16, 19 and 26. We can also see the conditional probability table for the model with 27 input states is sparser than the one for the model with 9 input states and 18 input states. The model with 18 input states has four input states with only zero or a few observations while the model with 27 input states has totally eight input states with only zero or a few observations. These input states with insufficient training cases would cause difficulties during prediction stage.

To compute the expected number of outages in each input states, we use the same method as the one used for the model with 9 and 18 input states. The expected number and levels of outages for each input states are computed and listed in Table 6.25. The trends in expected values with different levels of outages in previous week are shown in Figure 6.24, 6.25 and 6.26. In Figure 6.24, the third input, outage level of previous week, is “Low” and thus the input states range from 1 to 9. In Figure 6.25, the third input, outage level of previous week, is “Medium” and thus the input states range from 10 to 18. In Figure 6.26, the third input, outage level of previous week, is “High” and thus the input states range from 19 to 27.

Since the thresholds we have used to determine “High” outage in previous week are both 0.7 in the models with 18 and 27 inputs, the input states with “High” outages in previous week are the same in these two models and thus have the same expected values of outages. Therefore, Figure 6. 26 is the same as Figure 6.21. Each input state with “Low” outage in previous week in the model with 18 inputs is divided into two input states in the model of 27 input states. Even though more details in the data can be captured in the model with 27 input states, less training cases are assigned to each input state which causes less reliable predictions. As a result, we have observed decreasing trends in Figure 6.24 and 6.25 when the month type increases. Again, we take the expected value in each input state as the predictions for the weeks with the same input state. Figure 6.27 shows the time series of predicted number of outages and Figure 6.28 shows the predictions of outage levels for Wichita from the year 1998 to 2007. The model with 27 input states slightly outperforms the one with 18 input states. The AAE decreases slightly from 8.31 to 8.11. Considering the fact that this little improvement costs more computation, more sparseness in CPT and more discontinuity in the trends in expected outages, we have decided to abandon the model with 27 input states and focus on the one with 18 inputs in the rest of our study.

Table 6.23 All 27 Input States and Number of Observations in Each State for Wichita

	Month Type	Fair ay Level	Previous Week Outage	Number of Observations
Input State 1	1	1	Low	52
Input State 2	2	1	Low	19
Input State 3	3	1	Low	1
Input State 4	1	2	Low	25
Input State 5	2	2	Low	15
Input State 6	3	2	Low	4
Input State 7	1	3	Low	10
Input State 8	2	3	Low	15
Input State 9	3	3	Low	6
Input State 10	1	1	Medium	21
Input State 11	2	1	Medium	39
Input State 12	3	1	Medium	10
Input State 13	1	2	Medium	9
Input State 14	2	2	Medium	31
Input State 15	3	2	Medium	26
Input State 16	1	3	Medium	2
Input State 17	2	3	Medium	20
Input State 18	3	3	Medium	41
Input State 19	1	1	High	1
Input State 20	2	1	High	13
Input State 21	3	1	High	9
Input State 22	1	2	High	0
Input State 23	2	2	High	6
Input State 24	3	2	High	34
Input State 25	1	3	High	0
Input State 26	2	3	High	2
Input State 27	3	3	High	69

Table 6.24 The Conditional Probability Table with 27 Input States for Wichita

Outage Level	1	2	3	4	5	6	7	8	9
Input State 1	0.65	0.19	0.13	0.02	0.00	0.00	0.00	0.00	0.00
Input State 2	0.21	0.32	0.16	0.26	0.00	0.00	0.05	0.00	0.00
Input State 3	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 4	0.68	0.20	0.04	0.04	0.04	0.00	0.00	0.00	0.00
Input State 5	0.13	0.33	0.20	0.20	0.07	0.07	0.00	0.00	0.00
Input State 6	0.00	0.00	0.00	0.50	0.00	0.50	0.00	0.00	0.00
Input State 7	0.20	0.30	0.40	0.10	0.00	0.00	0.00	0.00	0.00
Input State 8	0.07	0.27	0.13	0.33	0.07	0.13	0.00	0.00	0.00
Input State 9	0.00	0.17	0.00	0.17	0.33	0.33	0.00	0.00	0.00
Input State 10	0.52	0.14	0.19	0.10	0.00	0.05	0.00	0.00	0.00
Input State 11	0.10	0.18	0.31	0.15	0.15	0.08	0.03	0.00	0.00
Input State 12	0.00	0.20	0.20	0.00	0.00	0.20	0.40	0.00	0.00
Input State 13	0.22	0.00	0.56	0.11	0.00	0.11	0.00	0.00	0.00
Input State 14	0.06	0.19	0.23	0.10	0.10	0.23	0.10	0.00	0.00
Input State 15	0.04	0.08	0.12	0.15	0.15	0.23	0.08	0.15	0.00
Input State 16	0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 17	0.00	0.15	0.30	0.25	0.15	0.15	0.00	0.00	0.00
Input State 18	0.00	0.05	0.05	0.15	0.15	0.29	0.12	0.20	0.00
Input State 19	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 20	0.00	0.08	0.00	0.23	0.31	0.15	0.08	0.15	0.00
Input State 21	0.00	0.00	0.22	0.00	0.11	0.11	0.33	0.22	0.00
Input State 22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 23	0.00	0.33	0.17	0.17	0.00	0.33	0.00	0.00	0.00
Input State 24	0.00	0.03	0.00	0.06	0.06	0.15	0.29	0.24	0.18
Input State 25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 26	0.00	0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00
Input State 27	0.00	0.00	0.00	0.01	0.07	0.07	0.22	0.38	0.25

Table 6.25 All 27 Input States and Number of Observations in Each State for Wichita

	Month Type	Fair ay Level	Previous Week Outage	Expected Number of Outages	Expected Level of Outages
Input 1State 1	1	1	Low	8.90	1
Input State 2	2	1	Low	14.52	3
Input State 3	3	1	Low	15.52	3
Input State 4	1	2	Low	9.02	1
Input State 5	2	2	Low	15.15	3
Input State 6	3	2	Low	24.32	5
Input State 7	1	3	Low	12.87	2
Input State 8	2	3	Low	17.47	3
Input State 9	3	3	Low	22.58	5
Input State 10	1	1	Medium	11.23	2
Input State 11	2	1	Medium	17.25	3
Input State 12	3	1	Medium	26.21	6
Input State 13	1	2	Medium	15.47	3
Input State 14	2	2	Medium	20.47	4
Input State 15	3	2	Medium	27.09	6
Input State 16	1	3	Medium	8.94	1
Input State 17	2	3	Medium	19.08	4
Input State 18	3	3	Medium	30.81	6
Input State 19	1	1	High	11.39	2
Input State 20	2	1	High	27.87	6
Input State 21	3	1	High	33.31	7
Input State 22	1	2	High	0.00	1
Input State 23	2	2	High	19.35	4
Input State 24	3	2	High	44.60	8
Input State 25	1	3	High	0.00	1
Input State 26	2	3	High	13.45	2
Input State 27	3	3	High	51.66	8

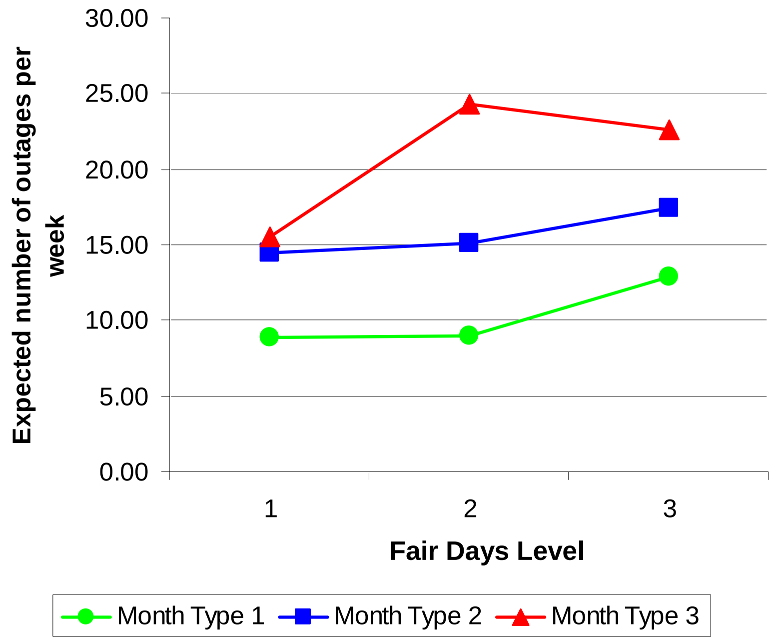

Figure 6.24 The Trends in Expected Values of Outages with Previous Week Outage as Low for Wichita

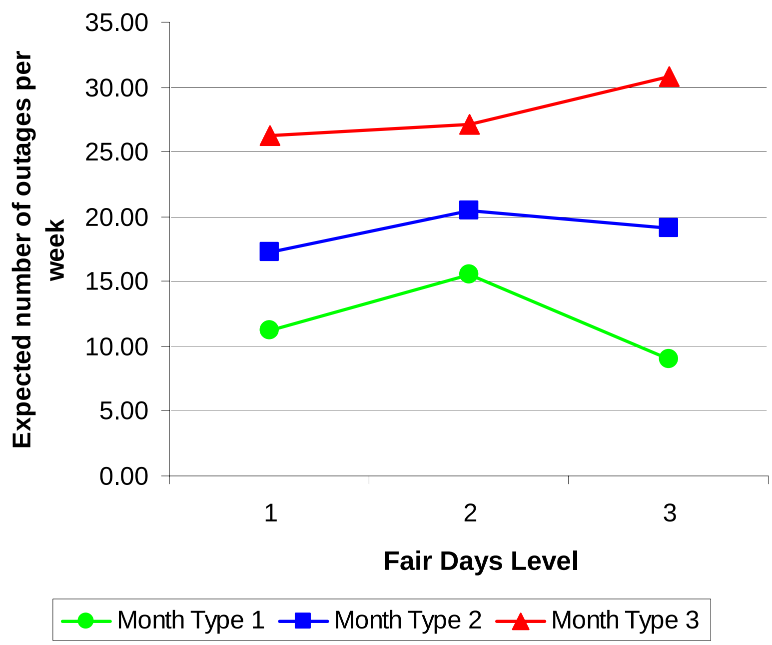

Figure 6.25 The Trends in Expected Values of Outages with Previous Week Outage as Medium for Wichita

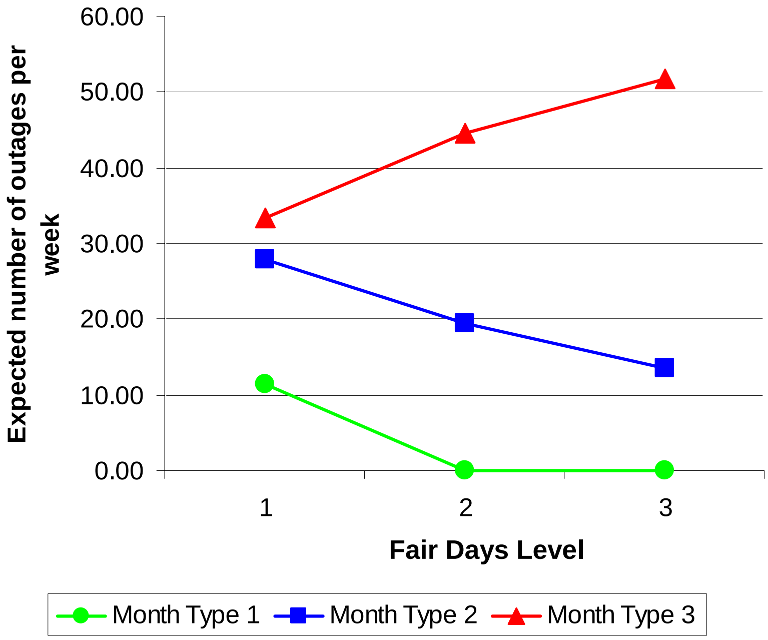

Figure 6.26 The Trends in Expected values of Outages with Previous Week Outage as High for Wichita

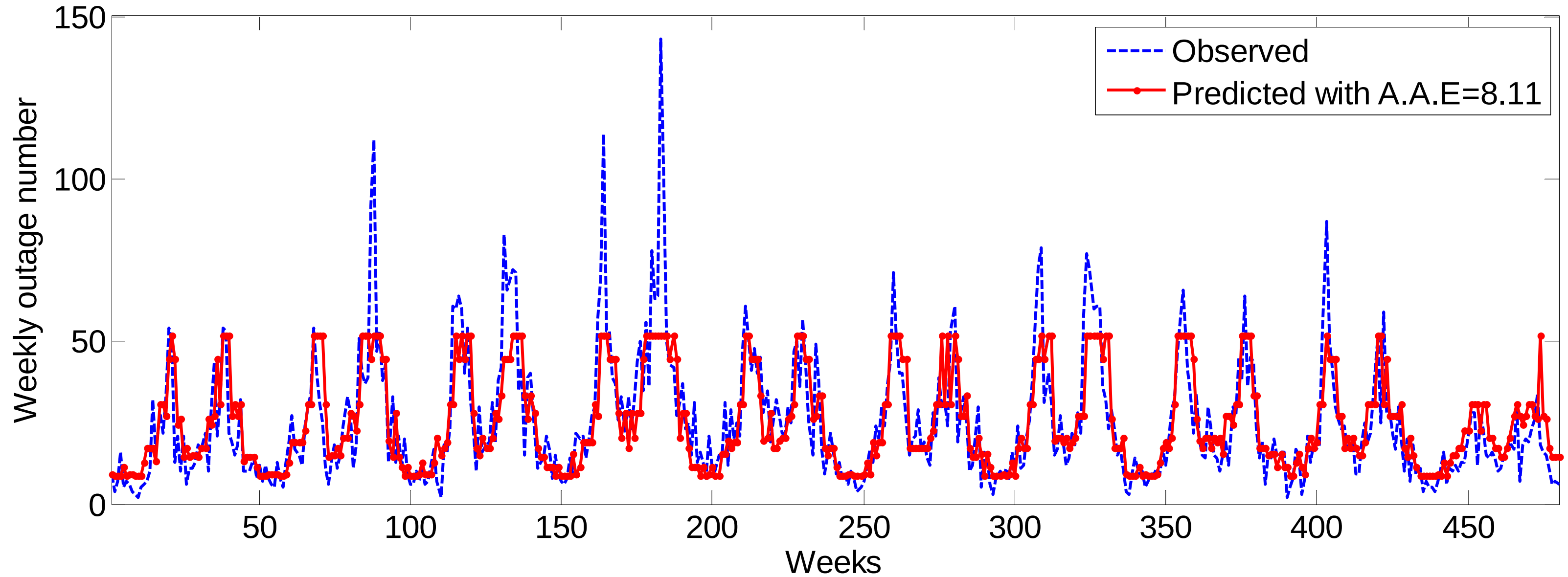

Figure 6.27 Outages Predicted and Observed by the Bayesian Model with 27 Input States for Wichita

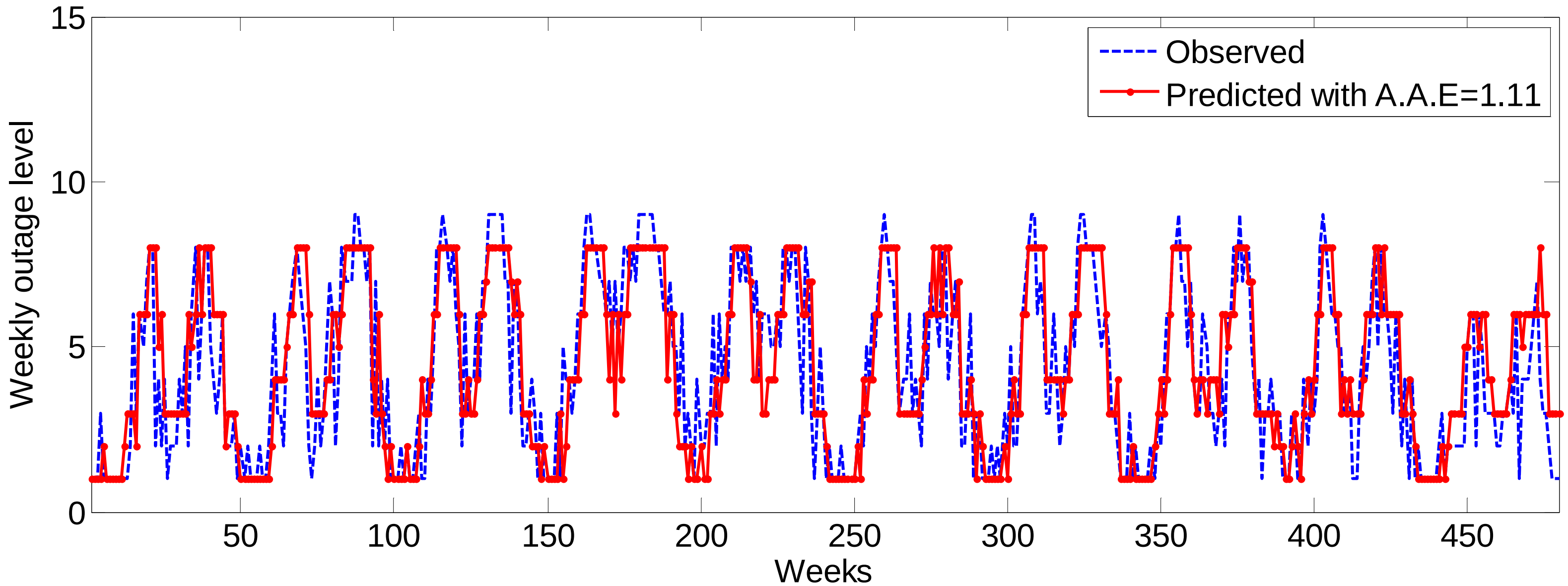

Figure 6.28 Outage Levels Predicted and Observed by the Bayesian Model with 27 Input States for Wichita

6.3.4 Models with 18 Input States and Predictions for More Cities

6.3.4.1 Model and Performance for Topeka

The numbers of training cases for each input state are given in Table 6.25 and the CPT is given in Table 6.26. Like Wichita, there are four input states (10, 13, 16 and 17) with only zero or a few observations.

Experiments were done to find the optimal value for the threshold used to determine the outage level in the previous week. AAE computed for predictions of animal-related outages with different values of threshold for Topeka are shown in Table 6.27. These results match that of Wichita, that is, 70% is the most optimal value with the smallest AAE.

The expected number and levels of outages for each input states are computed and listed in Table 6.28. The trends in expected outages in each state are shown in Figure 6.29 and 6.30. These trends are similar to those observed for Wichita.

Table 6.26 All 18 Possible Input States and Number of Observations in Each State for Topeka

	Month Type	Fair Day Level	Previous Week Outage	Number of Observations
Input State 1	1	1	Low	71
Input State 2	2	1	Low	48
Input State 3	3	1	Low	7
Input State 4	1	2	Low	39
Input State 5	2	2	Low	56
Input State 6	3	2	Low	20
Input State 7	1	3	Low	8
Input State 8	2	3	Low	39
Input State 9	3	3	Low	49
Input State 10	1	1	High	2
Input State 11	2	1	High	7
Input State 12	3	1	High	8

Input State 13	1	2	High	0
Input State 14	2	2	High	8
Input State 15	3	2	High	42
Input State 16	1	3	High	0
Input State 17	2	3	High	2
Input State 18	3	3	High	74

Table 6.27 The Conditional Probability Table with 18 Input States for Topeka

Outage Level	1	2	3	4	5	6	7	8	9
Input State 1	0.58	0.23	0.08	0.03	0.04	0.03	0.00	0.01	0.00
Input State 2	0.23	0.17	0.19	0.15	0.08	0.15	0.02	0.02	0.00
Input State 3	0.00	0.14	0.29	0.14	0.00	0.29	0.14	0.00	0.00
Input State 4	0.56	0.18	0.08	0.15	0.00	0.03	0.00	0.00	0.00
Input State 5	0.07	0.18	0.18	0.18	0.14	0.16	0.05	0.04	0.00
Input State 6	0.10	0.10	0.20	0.20	0.10	0.10	0.10	0.10	0.00
Input State 7	0.00	0.25	0.50	0.13	0.00	0.00	0.13	0.00	0.00
Input State 8	0.13	0.18	0.28	0.13	0.18	0.05	0.03	0.03	0.00
Input State 9	0.02	0.04	0.04	0.18	0.04	0.12	0.29	0.16	0.10
Input State 10	0.00	0.00	0.50	0.50	0.00	0.00	0.00	0.00	0.00
Input State 11	0.00	0.00	0.29	0.14	0.43	0.14	0.00	0.00	0.00
Input State 12	0.00	0.00	0.00	0.00	0.00	0.25	0.13	0.38	0.25
Input State 13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 14	0.13	0.13	0.13	0.38	0.13	0.13	0.00	0.00	0.00
Input State 15	0.00	0.02	0.00	0.05	0.05	0.14	0.19	0.21	0.33
Input State 16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 17	0.00	0.00	0.00	0.50	0.50	0.00	0.00	0.00	0.00
Input State 18	0.00	0.00	0.00	0.00	0.05	0.09	0.18	0.20	0.47

Table 6.28 The Performance of the Model with Different Values for the Threshold in Model with 18 Input States for Topeka

Threshold	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
AAE	7.15	6.99	6.98	6.84	6.75	6.65	6.66	6.48	6.55	6.71

Table 6.29 All 18 Possible Input States and Number of Observations in Each State for Topeka

	Month Type	Fair Day Level	Previous Week Outage	Expected Values of Outages	Expected Outage Level
Input State 1	1	1	Low	7.95	1
Input State 2	2	1	Low	12.91	3
Input State 3	3	1	Low	16.92	4
Input State 4	1	2	Low	8.00	1
Input State 5	2	2	Low	15.42	4
Input State 6	3	2	Low	16.67	4
Input State 7	1	3	Low	13.48	3
Input State 8	2	3	Low	13.37	3
Input State 9	3	3	Low	25.08	7
Input State 10	1	1	High	13.50	3
Input State 11	2	1	High	16.36	4
Input State 12	3	1	High	33.32	8
Input State 13	1	2	High	0.00	1
Input State 14	2	2	High	13.79	3
Input State 15	3	2	High	33.36	8
Input State 16	1	3	High	0.00	1
Input State 17	2	3	High	16.44	4
Input State 18	3	3	High	37.58	9

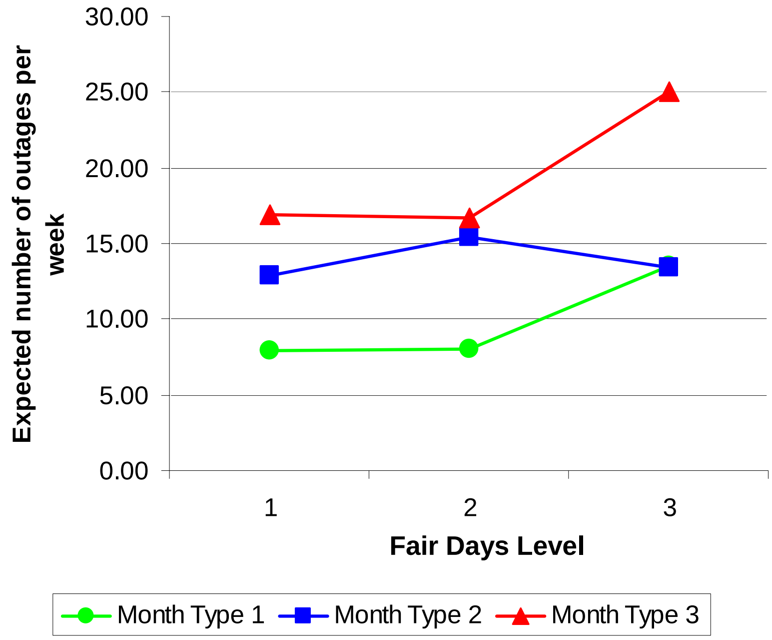

Figure 6.29 The Trends in Expected Values of Outages with Previous Week Outage as Low for Topeka

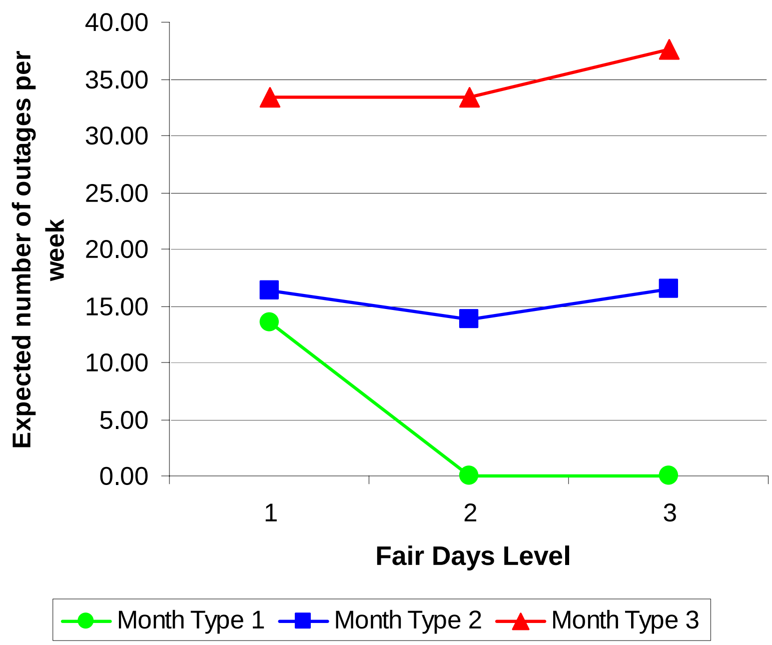

Figure 6.30 The Trends in Expected Values of Outages with Previous Week Outage as High for Topeka

Figure 6.31 shows the time series of predicted number of outages and Figure 6.32 shows the predictions of outage levels for Topeka from the year 1998 to 2007. For Topeka, the model with 18 input states outperforms the one with 9 input states. The addition of the third input brings down the AAE from 7.17 to 6.48. In Figure 6.31, there are not as many underestimations in months with high outages as in Figure 6.8 which shows the predictions by the model with 9 input states for Topeka. Also there is no overestimation for the year 2007. Once again, we have observed better performance of predictions when the predictions are represented as the outage levels instead of the numbers of outages.

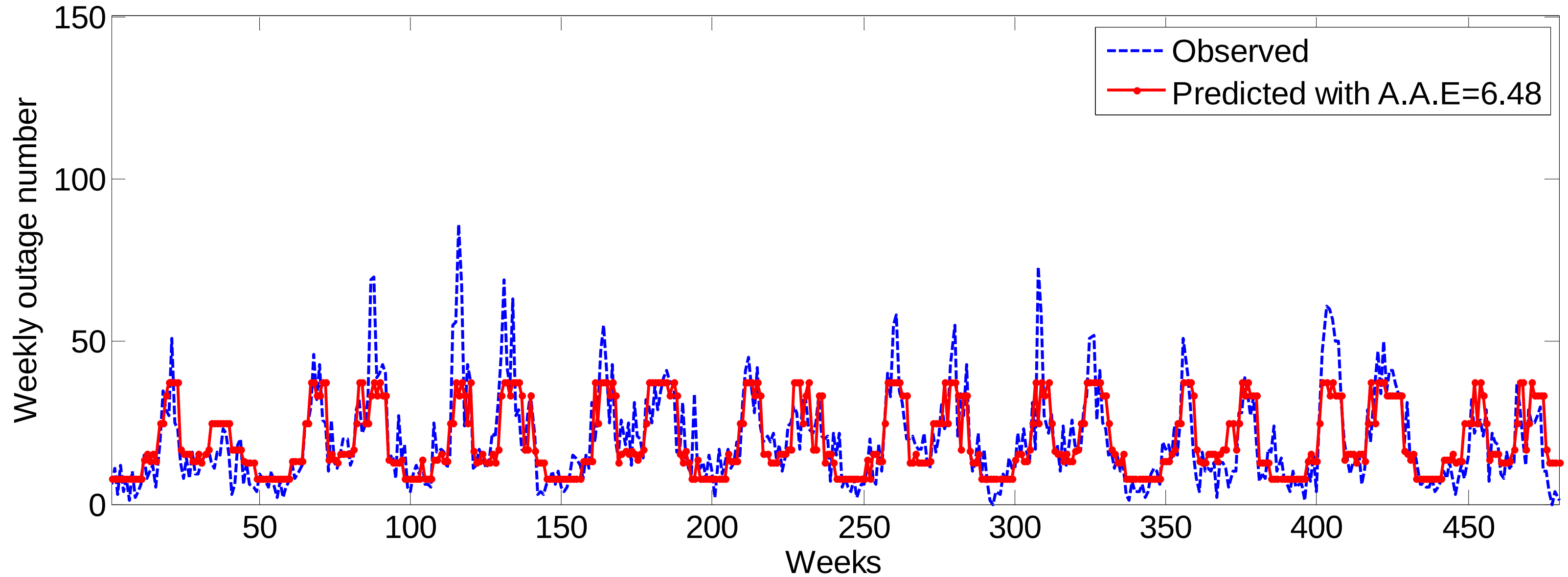

Figure 6.31 Outages Predicted and Observed by the Bayesian Model with 18 Input States for Topeka

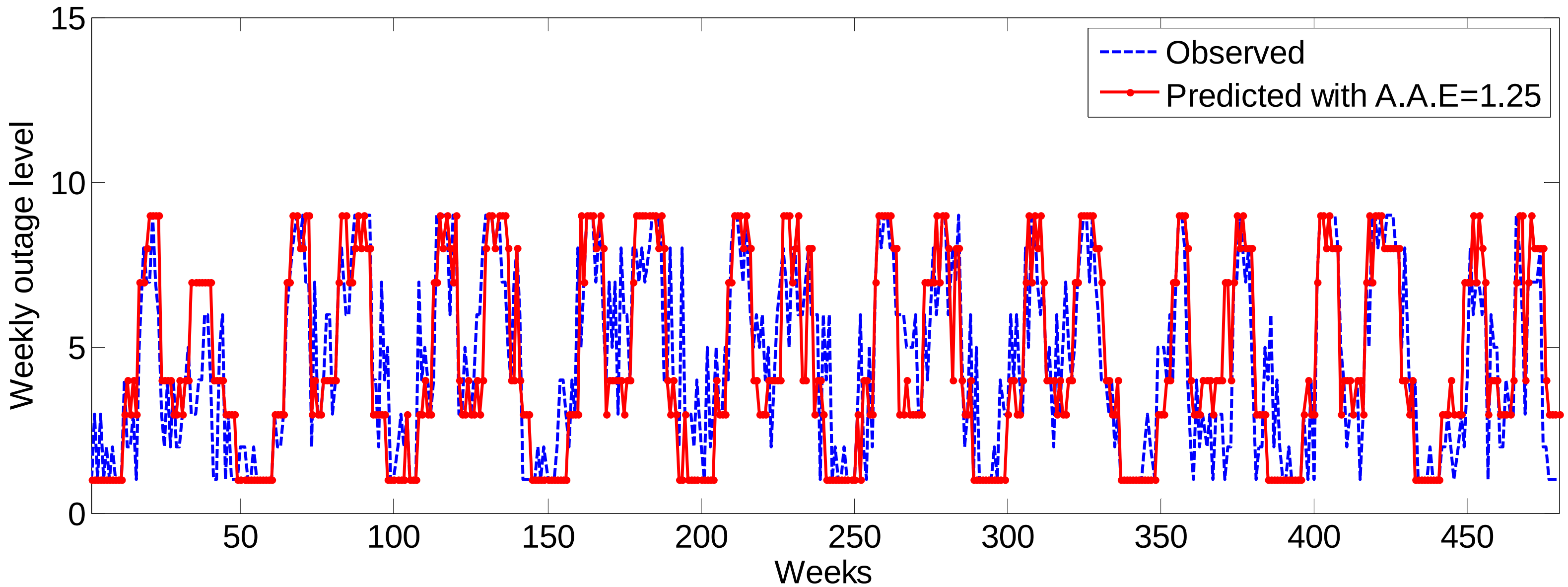

Figure 6.32 Outage Levels Predicted and Observed by the Bayesian Model with 18 Input States for Topeka

6.3.4.2 Model and Performance for Lawrence

Similarly, we have done predictions for animal-related outages with different values for the threshold and computed the average absolute error for Lawrence as shown in Table 6.29. Again, 70% is found to be the optimal value for threshold. The numbers of training cases for each input state are given in Table 6.30 and the CPT is given in Table 6.31. Similar to Topeka, the four input states (10, 13, 16 and 17) have only zero or a few observations. The expected number and levels of outages for each input states are computed and listed in Table 6.32. The trends in expected outages in each state are shown in Figure 6.33 and 6.34. In Figure 6.33, animal-related outages per week increase when the month type increases from 1 to 3. Also we can observe increasing trends in animal-related outages when the fair days level increases from 1 to 2. However, the expected values of outages decrease when the fair days level increases from 2 to 3. This observation is similar to the one in Figure 6.11, which shows the trends in expected outages of the model with 9 input states. This deficiency in predictions is not fixed by adding one more input to the Bayesian model.

Table 6.30 The Performance of the Model with Different Values for the Threshold for Lawrence

Threshold	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
AAE	3.153	3.063	3.056	3.045	2.939	2.938	2.940	2.936	2.994	3.143

Table 6.31 All 18 Possible Input States and Number of Observations in Each State for Lawrence

	Month Type	Fair Day Level	Previous Week Outage	Number of Observations
Input State 1	1	1	Low	72
Input State 2	2	1	Low	44
Input State 3	3	1	Low	6
Input State 4	1	2	Low	38
Input State 5	2	2	Low	55

Input State 6	3	2	Low	32
Input State 7	1	3	Low	8
Input State 8	2	3	Low	40
Input State 9	3	3	Low	64
Input State 10	1	1	High	2
Input State 11	2	1	High	12
Input State 12	3	1	High	12
Input State 13	1	2	High	0
Input State 14	2	2	High	9
Input State 15	3	2	High	35
Input State 16	1	3	High	0
Input State 17	2	3	High	0
Input State 18	3	3	High	51

Table 6.32 The Conditional Probability Table with 18 Input States for Lawrence

Outage Level	1	2	3	4	5	6	7	8	9
Input State 1	0.63	0.21	0.04	0.07	0.04	0.00	0.01	0.00	0.00
Input State 2	0.43	0.18	0.16	0.14	0.07	0.00	0.02	0.00	0.00
Input State 3	0.00	0.17	0.17	0.17	0.33	0.17	0.00	0.00	0.00
Input State 4	0.50	0.24	0.08	0.13	0.03	0.00	0.03	0.00	0.00
Input State 5	0.15	0.25	0.13	0.29	0.09	0.07	0.02	0.00	0.00
Input State 6	0.09	0.16	0.09	0.09	0.22	0.13	0.06	0.16	0.00
Input State 7	0.75	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 8	0.33	0.35	0.05	0.18	0.08	0.03	0.00	0.00	0.00
Input State 9	0.16	0.13	0.09	0.20	0.09	0.14	0.08	0.11	0.00
Input State 10	0.00	0.00	0.00	0.50	0.50	0.00	0.00	0.00	0.00
Input State 11	0.17	0.17	0.25	0.25	0.08	0.08	0.00	0.00	0.00
Input State 12	0.00	0.08	0.00	0.17	0.25	0.17	0.17	0.17	0.00
Input State 13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 14	0.11	0.22	0.11	0.33	0.11	0.00	0.11	0.00	0.00

Input State 15	0.00	0.03	0.03	0.14	0.20	0.09	0.20	0.17	0.14
Input State 16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 18	0.00	0.02	0.02	0.20	0.10	0.12	0.12	0.33	0.10

Table 6.33 All 18 Possible Input States and Number of Observations in Each State for Lawrence

	Month Type	Fair Day Level	Previous Week Outage	Expected Number of Outages	Expected Outage Level
Input State 1	1	1	Low	3.40	1
Input State 2	2	1	Low	4.51	2
Input State 3	3	1	Low	8.04	4
Input State 4	1	2	Low	4.02	1
Input State 5	2	2	Low	6.29	3
Input State 6	3	2	Low	9.22	5
Input State 7	1	3	Low	2.46	1
Input State 8	2	3	Low	4.75	2
Input State 9	3	3	Low	8.42	4
Input State 10	1	1	High	8.41	5
Input State 11	2	1	High	6.16	3
Input State 12	3	1	High	11.10	6
Input State 13	1	2	High	0.00	1
Input State 14	2	2	High	6.92	3
Input State 15	3	2	High	13.35	7
Input State 16	1	3	High	0.00	1
Input State 17	2	3	High	0.00	1
Input State 18	3	3	High	13.72	7

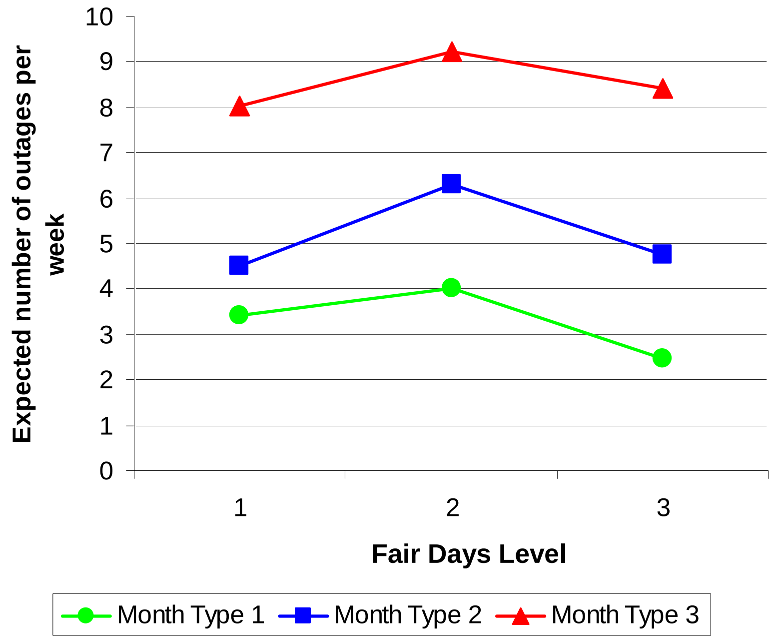

Figure 6.33 The Trends in Expected values of Outages with Previous Week Outage as Low for Lawrence

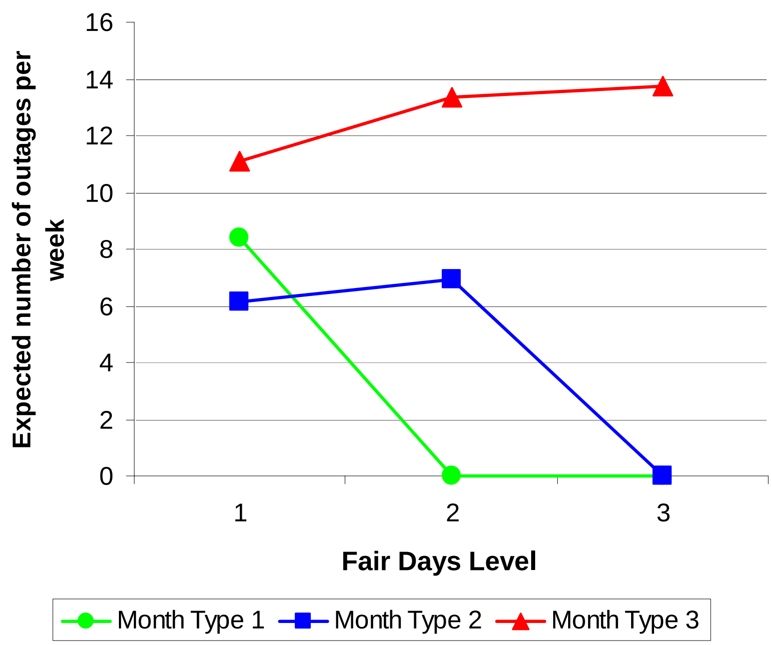

Figure 6.34 The Trends in Expected values of Outages with Previous Week Outage as High for Lawrence

Figure 6.35 shows the time series of predicted number of outages and Figure 6.36 shows the predictions of outage level for Lawrence from year 1998 to 2007. The addition of the third input brings down the average absolute error from 3.17 to 2.94.

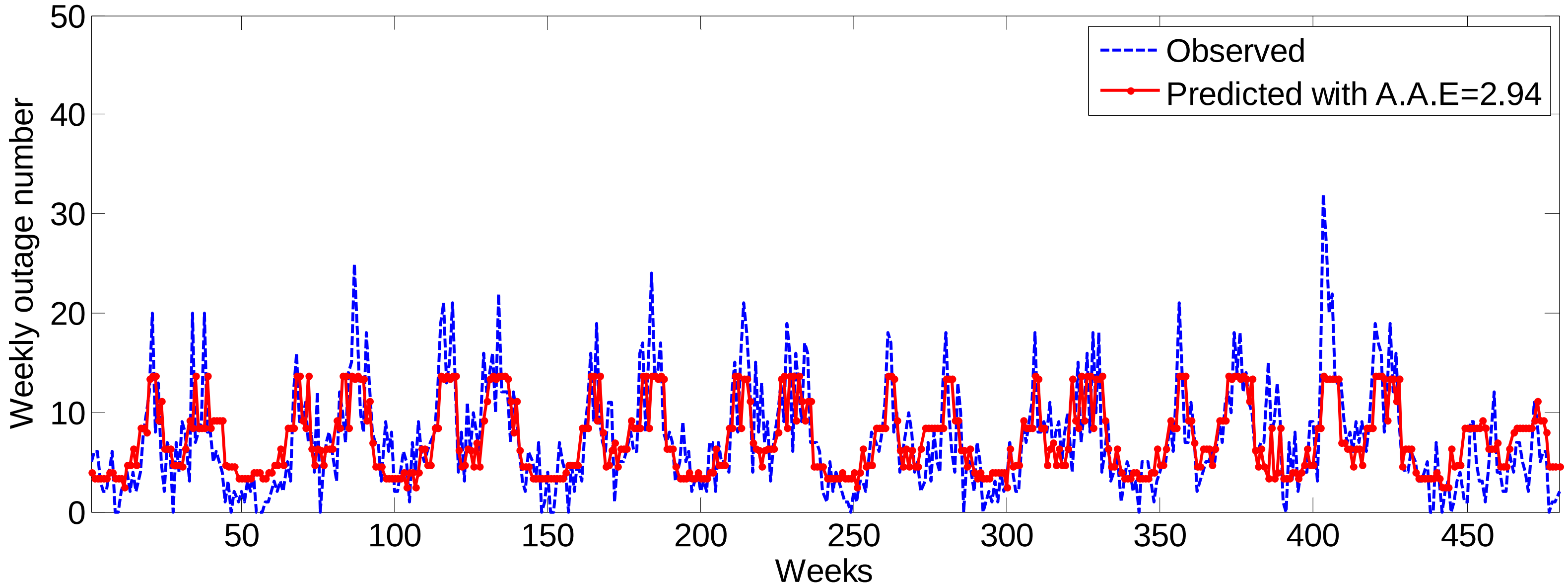

Figure 6.35 Outages Predicted and Observed by the Bayesian Model with 18 Input States for Lawrence

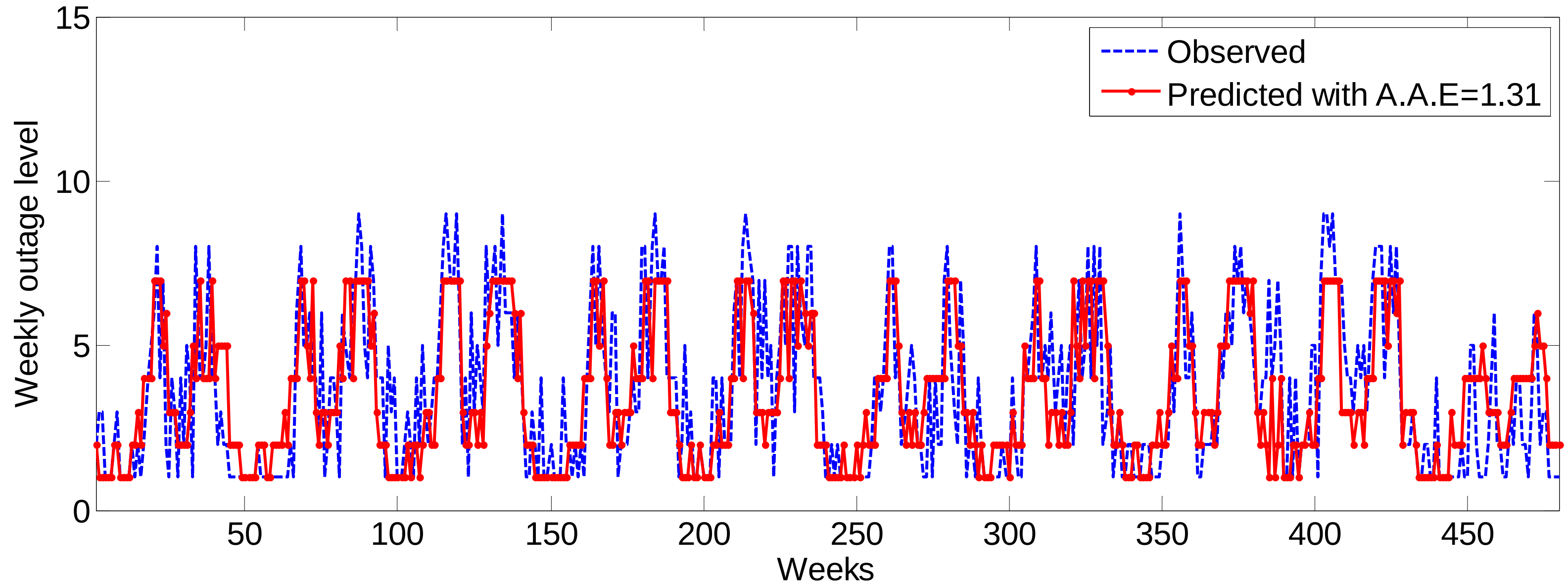

Figure 6.36 Outage Levels Predicted and Observed by the Bayesian Model with 18 Input States for Lawrence

6.3.4.3 Model and Performance for Manhattan

AAE for predictions of animal-related outages with different values of threshold are shown in Table 6.33. According to this table, the AAE does not change much, which implies that adding an extra input might not change the results. However to keep things consistent with other cities, 70% is used as the threshold value.

The numbers of training cases for each input state are given in Table 6.34 and the CPT is given in Table 6.35. The expected values and levels of outages for each input states are computed and listed in Table 6.36. The trends in expected values in each state are shown in Figure 6.37 and 6.38. There are only two input states which lack training data: input state 16 has zero observations and input state 13 has two observations. It results in decreasing trends in the expected outages as shown in Figure 6.38 when the month type is 1 and the fair days level increases from 1 to 3. Other than that, we have observed increasing trends in expected outages with increasing the month type and the fair days level. In Figure 6.37, we can observe increasing trends in expected outages when the month type increases but it is not the case when the fair days level increases. As we can see, the Bayesian model with 18 input states works better for Topeka and Wichita than for Lawrence and Manhattan. Topeka and Wichita are the two biggest cities in the service area of Westar Energy and both have sufficient observations of animal-related outages. Manhattan and Lawrence are the smaller cities and thus there are not as many observations of animal-related outages. Therefore, we can see that the Bayesian models with 9 input states and 18 input states have better performance with sufficient data for bigger cities. In other word, the insufficient data for small cities causes challenges for the Bayesian models to capture the correct trends in the outages.

Again, we take the expected value in each input state as the predictions for the weeks with the same input state. Figure 6.39 shows the time series of predicted number of outages and Figure 6.40 shows the predictions of outage level for Manhattan from the year 1998 to 2007. The addition of the third input brings down the AAE from 2.18 to 2.12.

Table 6.34 The Performance of the Model with Different Values for the Threshold in the Model with 18 Input States for Manhattan

Threshold	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
AAE	2.14	2.14	2.12	2.13	2.13	2.10	2.14	2.12	2.12	2.10

Table 6.35 All 18 Possible Input States and Number of Observations in Each State for Manhattan

	Month Type	Fair Day Level	Previous Week Outage	Number of Observations
Input State 1	1	1	Low	88
Input State 2	2	1	Low	56
Input State 3	3	1	Low	14
Input State 4	1	2	Low	22
Input State 5	2	2	Low	53
Input State 6	3	2	Low	45
Input State 7	1	3	Low	4
Input State 8	2	3	Low	30
Input State 9	3	3	Low	65
Input State 10	1	1	High	4
Input State 11	2	1	High	13
Input State 12	3	1	High	7
Input State 13	1	2	High	2
Input State 14	2	2	High	4
Input State 15	3	2	High	34
Input State 16	1	3	High	0
Input State 17	2	3	High	4
Input State 18	3	3	High	35

Table 6.36 The Conditional Probability Table with 18 Input States for Manhattan

Outage Level	1	2	3	4	5	6	7	8	9
Input State 1	0.25	0.18	0.26	0.17	0.11	0.02	0.00	0.00	0.00
Input State 2	0.14	0.32	0.13	0.18	0.18	0.04	0.02	0.00	0.00
Input State 3	0.00	0.00	0.21	0.14	0.29	0.36	0.00	0.00	0.00
Input State 4	0.27	0.18	0.41	0.09	0.05	0.00	0.00	0.00	0.00
Input State 5	0.06	0.23	0.17	0.25	0.23	0.06	0.02	0.00	0.00
Input State 6	0.16	0.11	0.11	0.04	0.29	0.18	0.09	0.00	0.02
Input State 7	0.25	0.25	0.25	0.00	0.25	0.00	0.00	0.00	0.00
Input State 8	0.20	0.13	0.30	0.10	0.27	0.00	0.00	0.00	0.00
Input State 9	0.05	0.09	0.14	0.15	0.42	0.12	0.02	0.02	0.00
Input State 10	0.00	0.25	0.50	0.00	0.25	0.00	0.00	0.00	0.00
Input State 11	0.08	0.08	0.31	0.08	0.23	0.23	0.00	0.00	0.00
Input State 12	0.00	0.00	0.14	0.00	0.57	0.14	0.00	0.14	0.00
Input State 13	0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 14	0.00	0.00	0.25	0.25	0.25	0.25	0.00	0.00	0.00
Input State 15	0.00	0.00	0.06	0.15	0.41	0.18	0.12	0.09	0.00
Input State 16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Input State 17	0.00	0.00	0.00	0.75	0.00	0.25	0.00	0.00	0.00
Input State 18	0.00	0.06	0.03	0.06	0.37	0.11	0.20	0.17	0.00

Table 6.37 All 18 Possible Input States and Number of Observations in Each State for Manhattan

	Month Type	Fair Day Level	Previous Week Outage	Expected Number of Outages	Expected Outage Level
Input State 1	1	1	Low	1.93	2
Input State 2	2	1	Low	2.43	3
Input State 3	3	1	Low	4.91	5

Input State 4	1	2	Low	1.49	2
Input State 5	2	2	Low	3.02	4
Input State 6	3	2	Low	4.81	5
Input State 7	1	3	Low	1.96	2
Input State 8	2	3	Low	2.32	3
Input State 9	3	3	Low	4.20	5
Input State 10	1	1	High	2.46	3
Input State 11	2	1	High	3.76	4
Input State 12	3	1	High	6.74	5
Input State 13	1	2	High	0.50	1
Input State 14	2	2	High	4.33	5
Input State 15	3	2	High	6.74	5
Input State 16	1	3	High	0.00	1
Input State 17	2	3	High	4.12	5
Input State 18	3	3	High	8.21	6

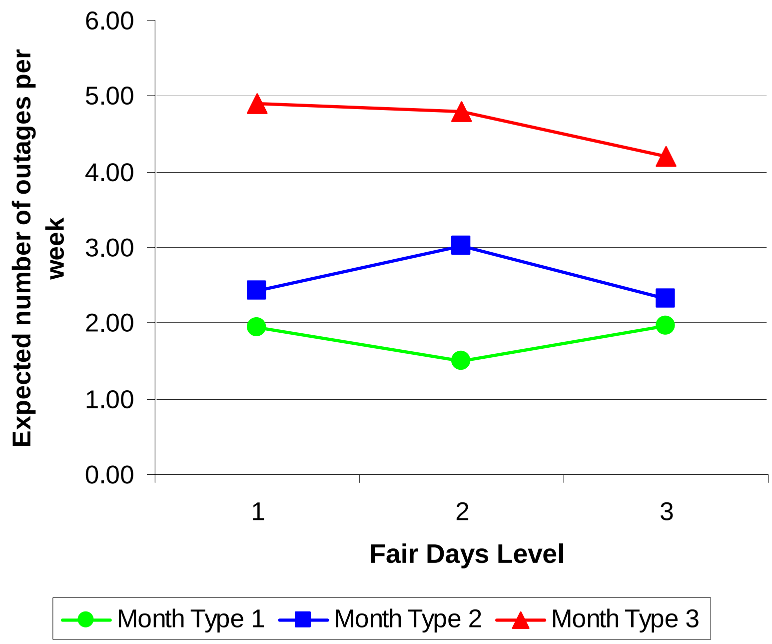

Figure 6.37 The Trends in Expected values of Outages with Previous Week Outage as Low for Manhattan

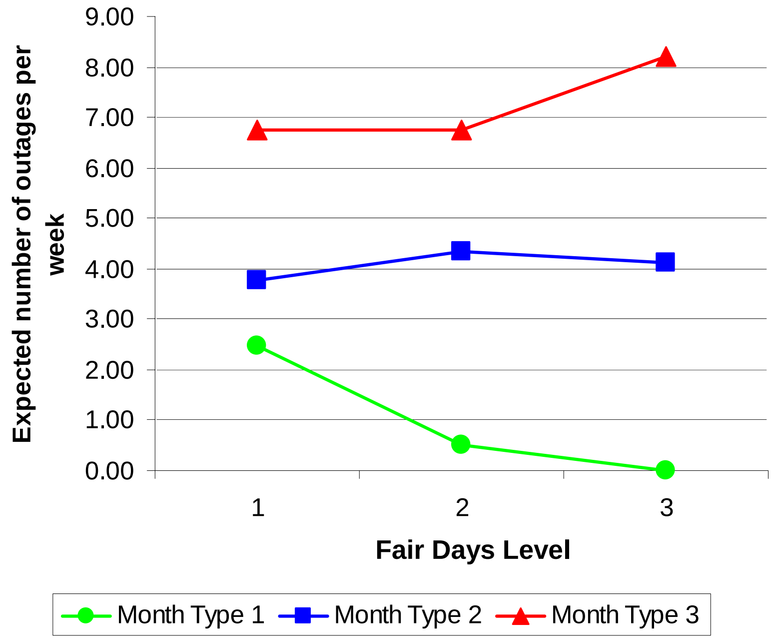

Figure 6.38 The Trends in Expected Values of Outages with Previous Week Outage as High for Manhattan

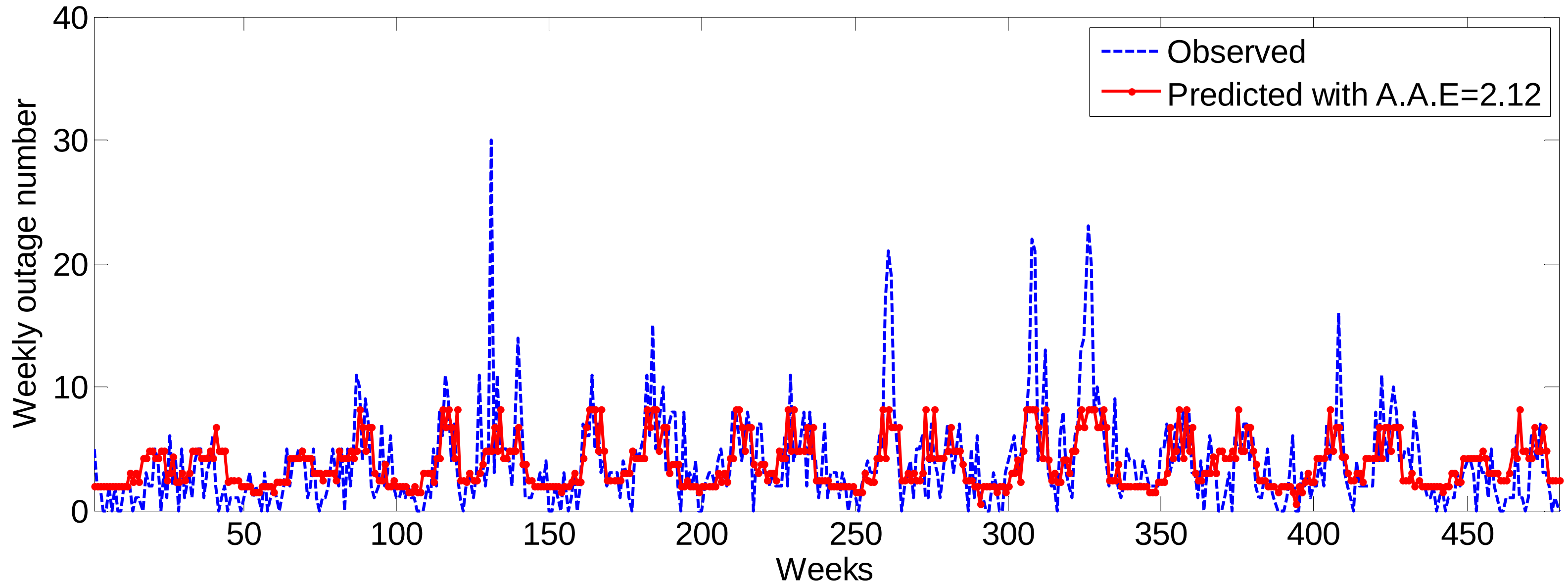

Figure 6.39 Outages Predicted and Observed by the Bayesian Model with 18 Input States for Manhattan

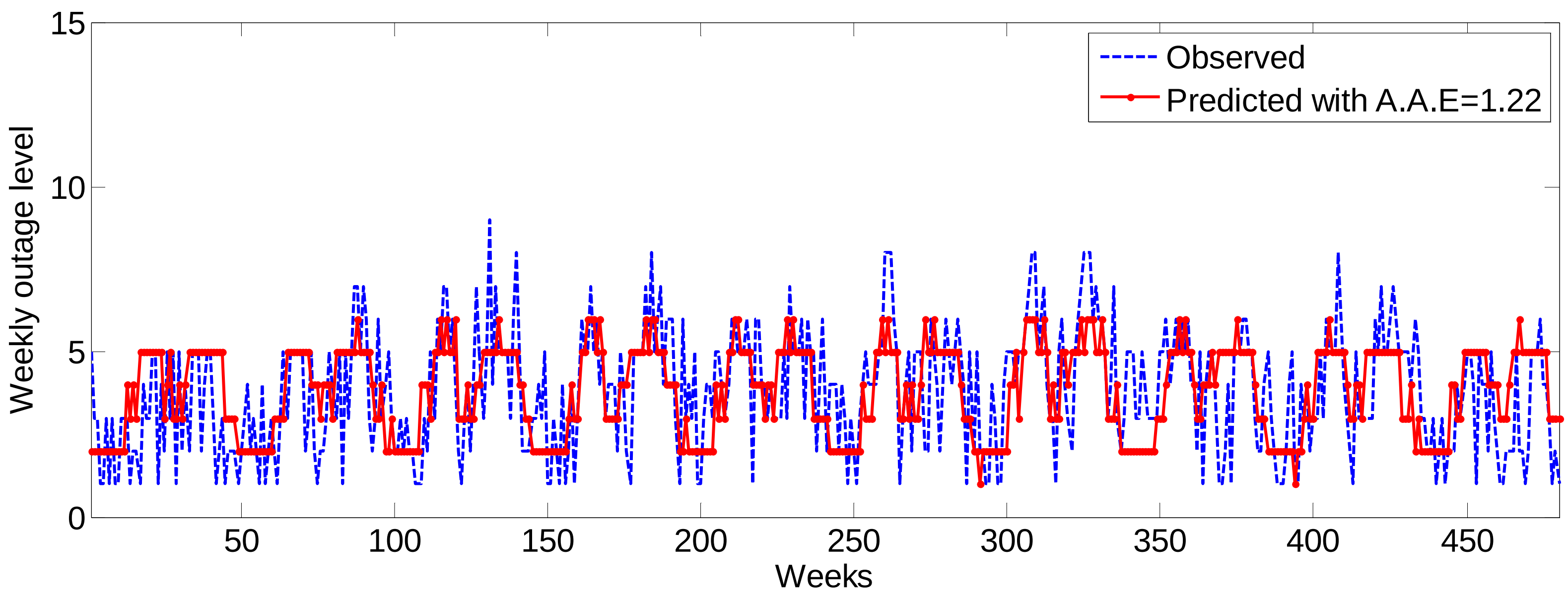

Figure 6.40 Outage Levels Predicted and Observed by the Bayesian Model with 18 Input States for Manhattan

6.3.4.4 Comparison between the Bayesian Models with 18 and 9 Input States

Plots of weekly observed and predicted outages by the Bayesian model with 18 and 9 input states for four cities are shown in Figure 6.41-6.44 with slopes of the best-fit lines (S) and their correlation coefficients (R). The AAE, values of S and R are listed in Table 6.38. Values of one for S and R would indicate best fit around the ideal (y=x) line. The model with 18 input states gives better results than the one with 9 input states for all the cities. The values of overall S for ten years are smaller than one which implies underestimations in the prediction. However, this is not the case for the prediction in 2007 for Wichita, Topeka and Lawrence. The big values of S for these three cities, 1.63, 1.19 and 1.31, imply overestimations in the predictions by the model with 9 input states. The overestimations in the predictions in 2007 are due to the significant decrease in observed outages in 2007 for these three cities. The model with 18 input states brings down the overestimations. There is no overestimation observed in the predictions in 2007 for Manhattan because the observed outages in 2007 have a range close to the ones in the other years. The correlation coefficients range from 0.26 to 0.77 with lower values for smaller cities. The fact that the correlation coefficients are smaller than 1 means the degree of linear relationships between predicted outages and observed outages is not high. This is because of the insufficiency of linear model and loss of information in the classification of the outages. The Bayesian models with 18 input states have higher overall R than the Bayesian models with 9 input states.

Table 6.38 Results of the Bayesian Models for Four Cities

Model	City	Overall AAE	Overall S	Overall R	AAE for 2007	S for 2007	R for 2007
Bayesian with 18 Input States	Wichita	8.31	0.84	0.77	6.80	1.30	0.66
	Topeka	6.48	0.86	0.77	6.70	1.16	0.74
	Lawrence	2.98	0.82	0.69	2.87	1.13	0.46
	Manhattan	2.12	0.64	0.53	1.90	0.94	0.26
Bayesian with 9 Input States	Wichita	9.53	0.80	0.68	11.14	1.63	0.71
	Topeka	7.17	0.82	0.70	6.61	1.19	0.76
	Lawrence	3.17	0.78	0.62	3.69	1.31	0.40
	Manhattan	2.18	0.60	0.45	2.11	1.06	0.33

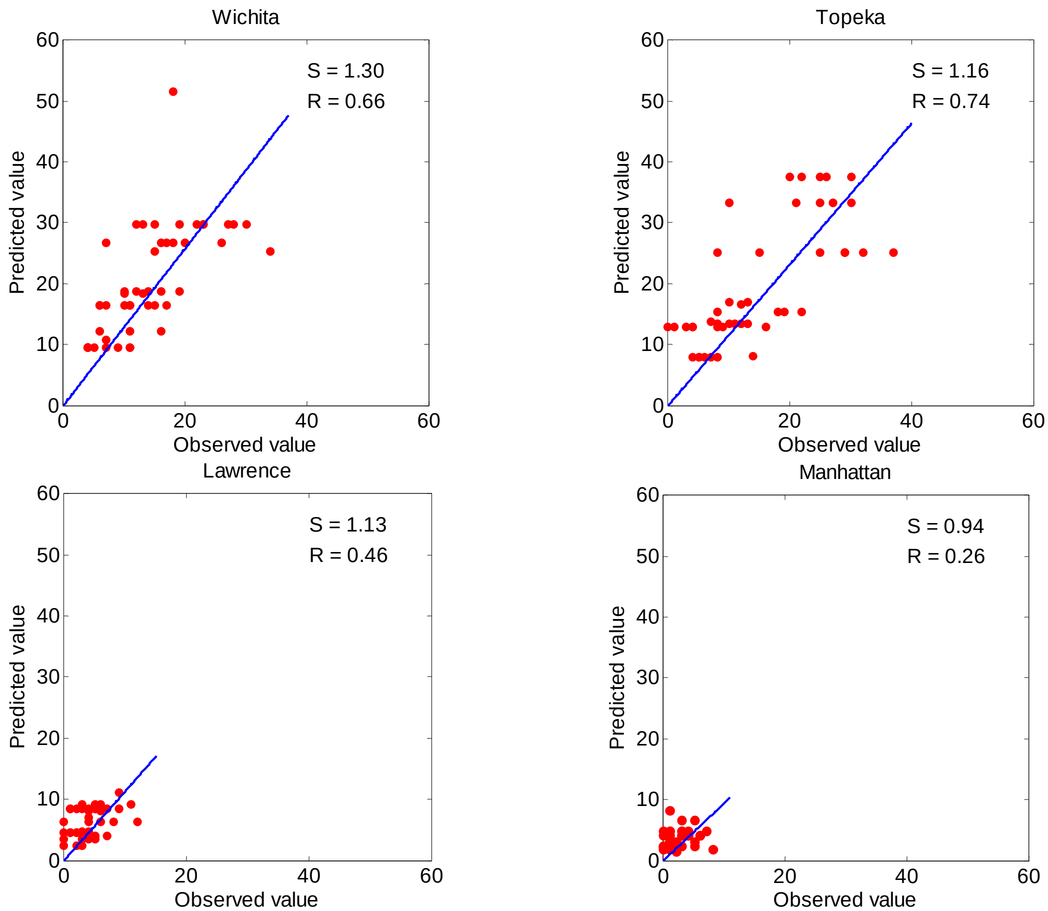

Figure 6.41 Scatter Plot of Outages Observed and Predicted in 2007 for Four Cities by the Bayesian Model with 18 Input States

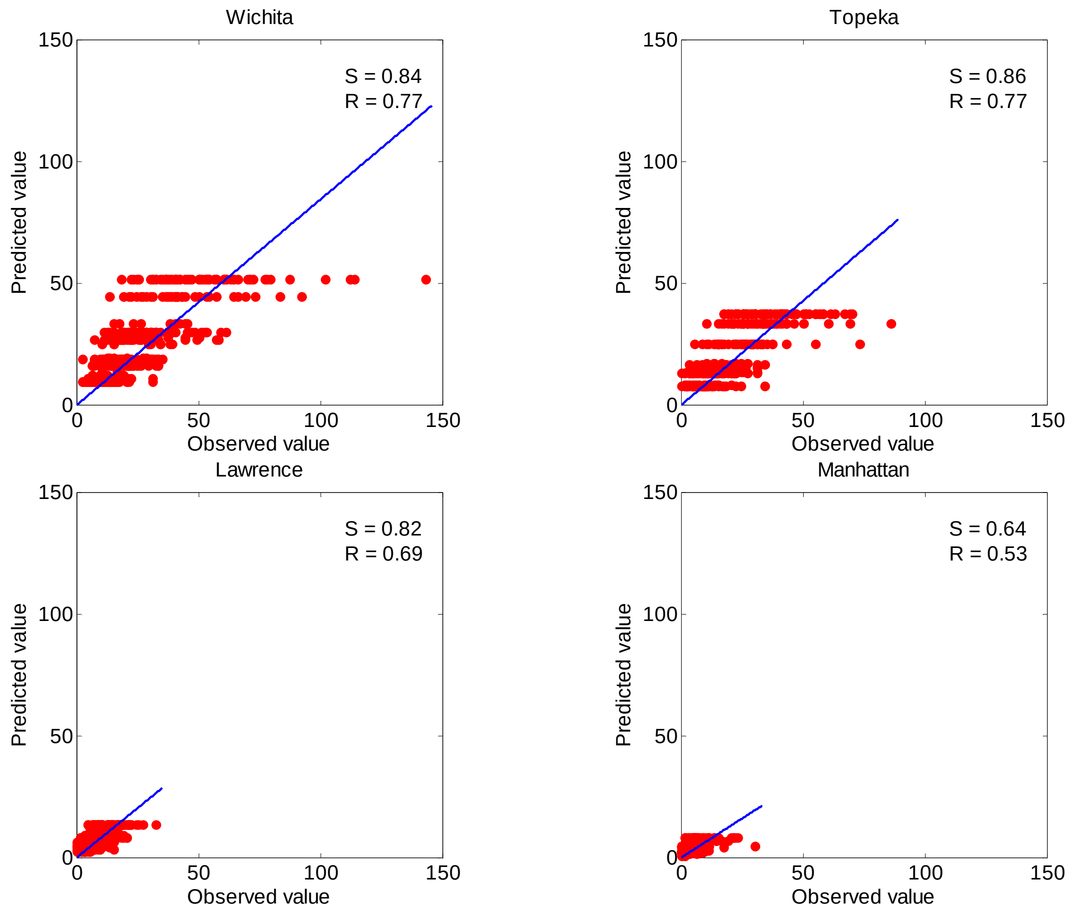

Figure 6.42 Scatter Plot of Outages Observed and Predicted for Four Cities from 1998 to 2007 by the Bayesian Model with 18 Input States

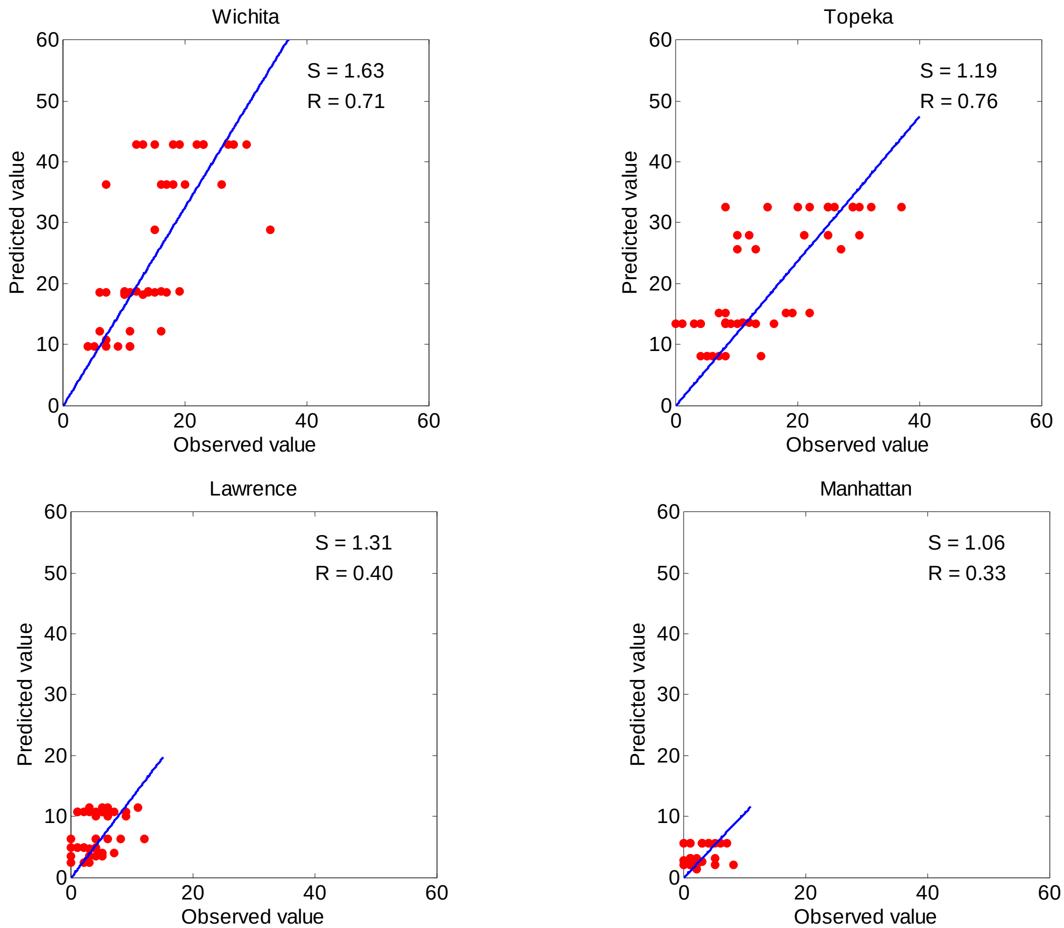

Figure 6.43 Scatter Plot of Outages Observed and Predicted in 2007 for Four Cities by the Bayesian Model with 9 Input States

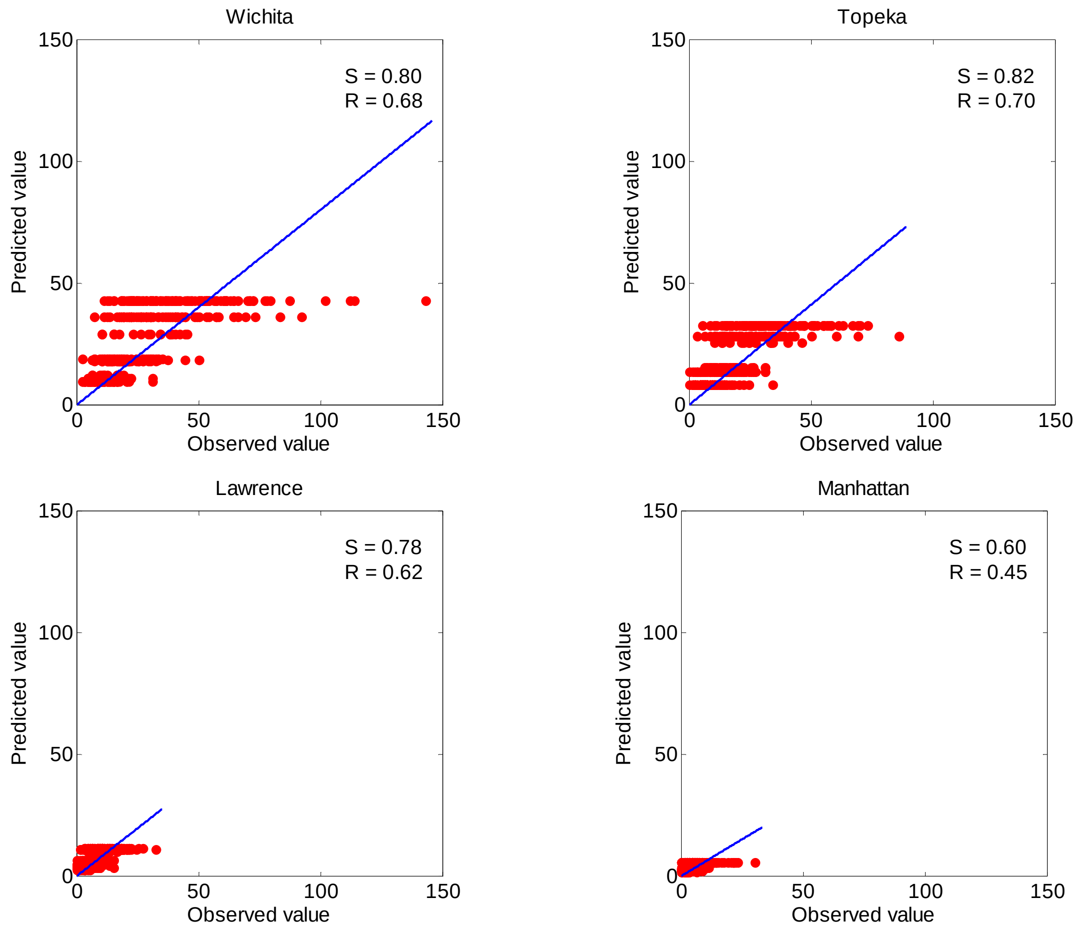

Figure 6.44 Scatter Plot of Outages Observed and Predicted for Four Cities from 1998 to 2007 by the Bayesian Model with 9 Input States

6.4 Monte Carlo Simulations

In section 6.2 and 6.3, we assume that the predicted value for each state is the expected value, which represents a point estimate for the number of outages. But since a particular month type, a particular level of fair weather days per week and a particular previous week outage level themselves are composed of a number of entities, an input state represents a range of different values of factors, and is only a rough classification of the effects of month and fair weather days on animal-caused outages. Thus, the model is expected to have errors in prediction and we should find a range of values within which the observed numbers of outages are expected to lie. Monte Carlo simulation is a common method to find out the confidence intervals. Moreover, classifying of the input data into discrete levels makes the model prone to inaccuracies in predictions because all the outages in one level are represented by an average value, which causes underestimations clearly observed in the predictions for all the cities. Outages higher than the average in an outage level are ignored while computing average. To overcome this insufficiency, we use Monte Carlo simulations to get a range for predicted outages.

Monte Carlo simulation uses random numbers to resample a system and gives the distributions of the output. Monte Carlo methods tend to be used when it is infeasible or impossible to compute an exact result with a deterministic algorithm [106]. The results of a Monte Carlo simulation are distributions of possible outcomes instead of one predicted outcome. In other word, Monte Carlo simulations give the range of possible outcomes that could occur and the likelihood of any outcome occurring. This is as with given the same weather conditions, the occurrences of animal-related outages are observed for hundreds or thousands times instead of the limited and sometime insufficient training cases. Even though Monte Carlo Simulations are an approximate technique, any degree of precision can be achieved by simply increasing the number of iterations [107]. Monte Carlo simulations have had a great impact in many different fields of computational science, especially in reliability assessment of power system [108-110].

6.4.1 The Selection of Probability Distribution Functions

Monte Carlo simulations are categorized as a sampling method because the inputs are randomly generated from probability distributions to simulate the process of sampling from an actual population. Therefore, we have tried to choose a distribution for the inputs that most closely match data we already have, or best represents our current state of knowledge. We thought of three methods to find the probability distribution function of the occurrences of outages in each input state. At first, the counting property of outage events makes it intuitive to consider Poisson distributions for calculating the number of outages in a given week. Unfortunately, when we tried to fit Poisson distributions to the normalized histograms of outages in each input state, we found out that for most input states the actual histograms are more scattered than Poisson distributions, as shown in Figure 6.45. Although in some cases the fit was good, overall it was found to be not a good approach. The difference will cause underestimations for the weekly outages since the probabilities that are higher than the Poisson distribution function are discarded during Monte Carlo simulations.

The second approach is to take the conditional probability table as the PDF instead of Poisson distribution function. The conditional probability table in each input state is normalized by the bin size in that outage level, which allows all outages in an outage level have an equal chance of occurrence instead of only taking the average value to represent the whole outage level. The reason behind equal chance for every outage in one outage level is that we think all the outages in one outage level should be treated evenly. As shown in Figure 6.46, the conditional probability table normalized by the bin size in each outage level is always able to cover the real range of scattered data in each input state, which would overcome possible under-estimations due to Poisson as the distribution function for each input state in Monte Carlo simulations.

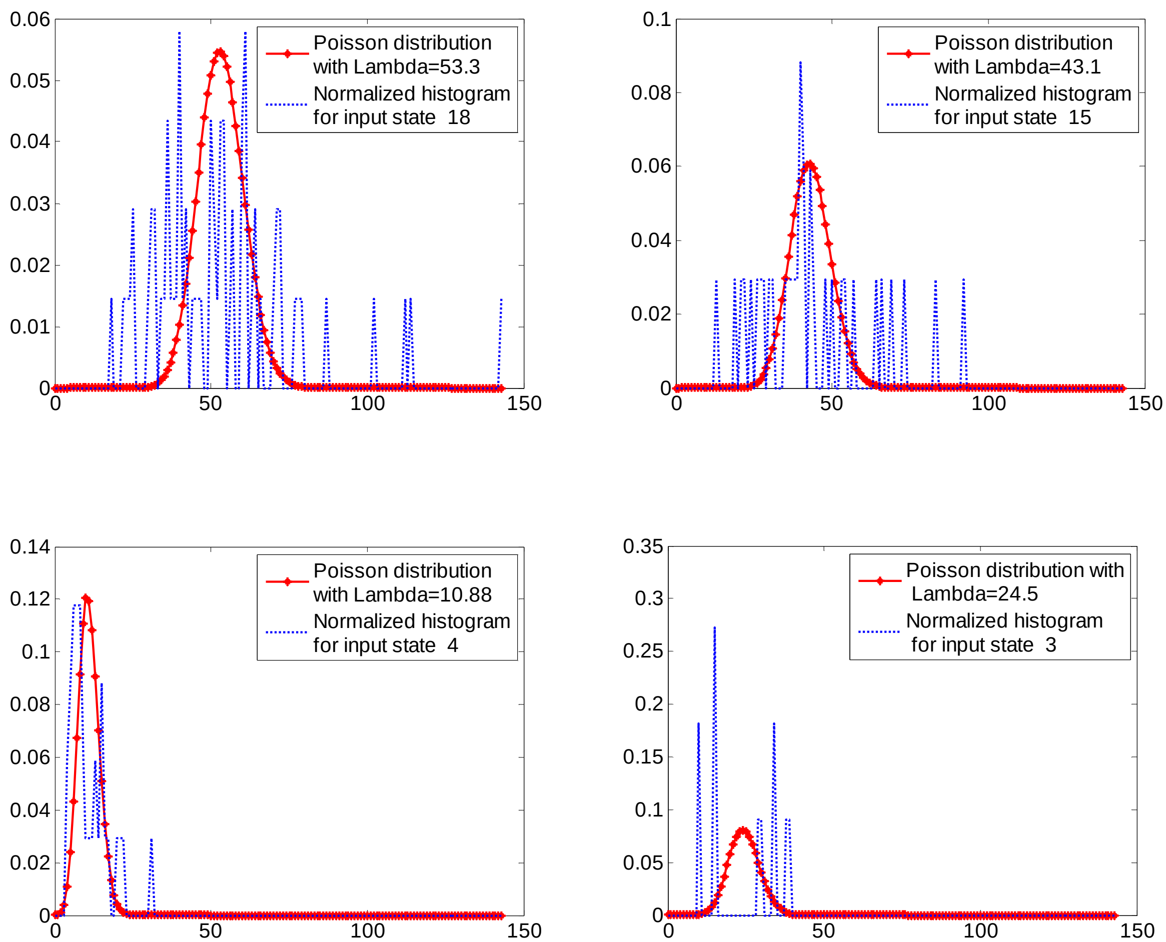

Figure 6.45 Examples of Fitting Poisson Distribution to the Histogram of Each Input State of the Bayesian Model with 18 Input States

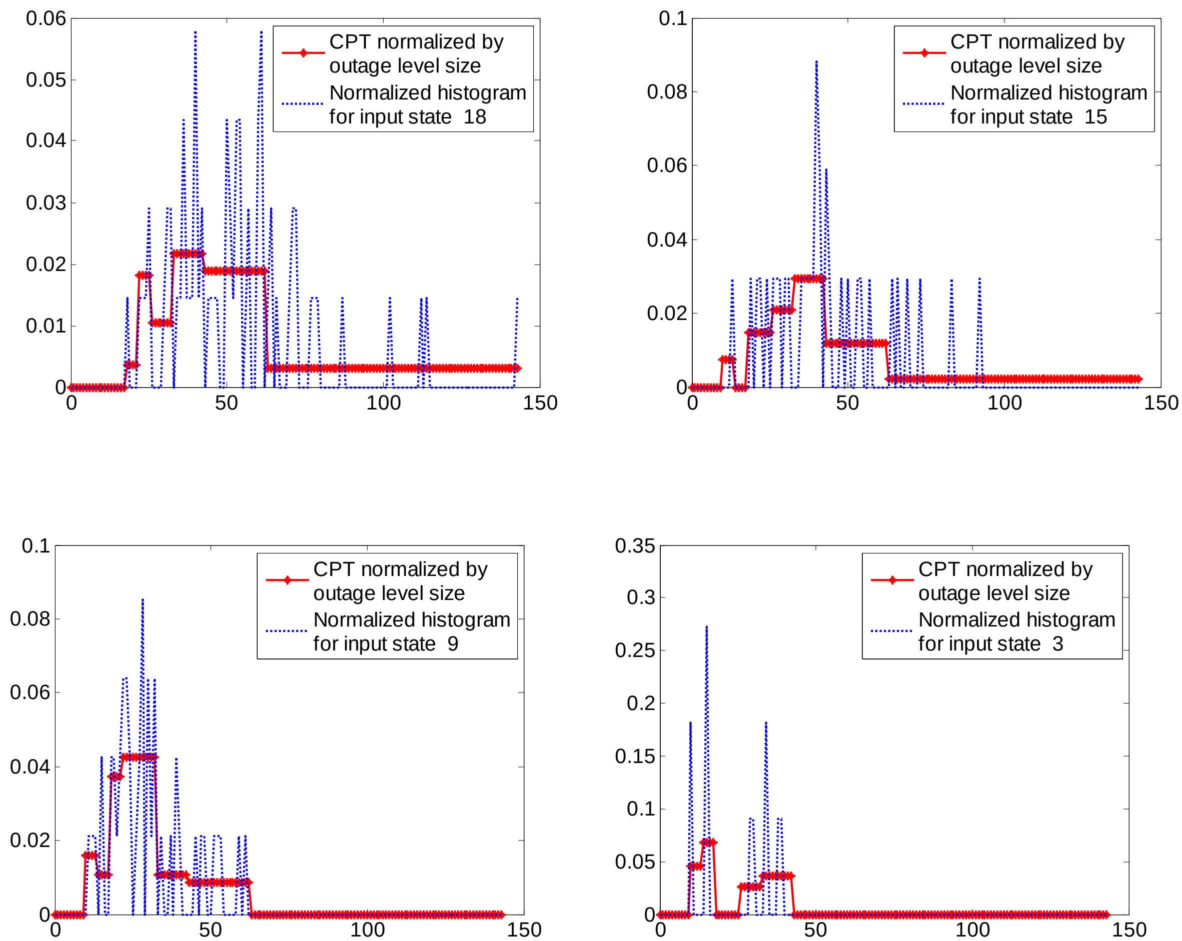

Figure 6.46 Examples of Normalized CPT and the Histogram of Each Input State of the Bayesian Model with 18 Input States

The third approach is to take the histogram in each input state as the probability distribution function. The histogram in each input state is normalized by the total occurrences of outages in that input state. To eliminate the gaps in the histogram, we have applied a smoothing algorithm in which the probability at point *i* is related to not only the occurrences at point *i* but also the ones of points *i-1* and *i+1* as shown in the following equation:

$$p(1) = \frac{x_1 + x_1 + x_1 + x_2}{4 \times \sum_{j=1}^{N} x_j} \qquad (6.5)$$

$$p(i) = \frac{x_i + x_{i-1} + x_i + x_{i+1}}{4 \times \sum_{j=1}^{N} x_j} \qquad 2 \le i \le N-1 \tag{6.6}$$

$$p(N) = \frac{x_N + x_N + x_N + x_{N-1}}{4 \times \sum_{j=1}^{N} x_j} \tag{6.7}$$

where x_i is the occurrences at point i and N is the total number of points.

Figure 6.47 shows the normalized histogram and smoothed histogram for input states 18,15,9 and 3.

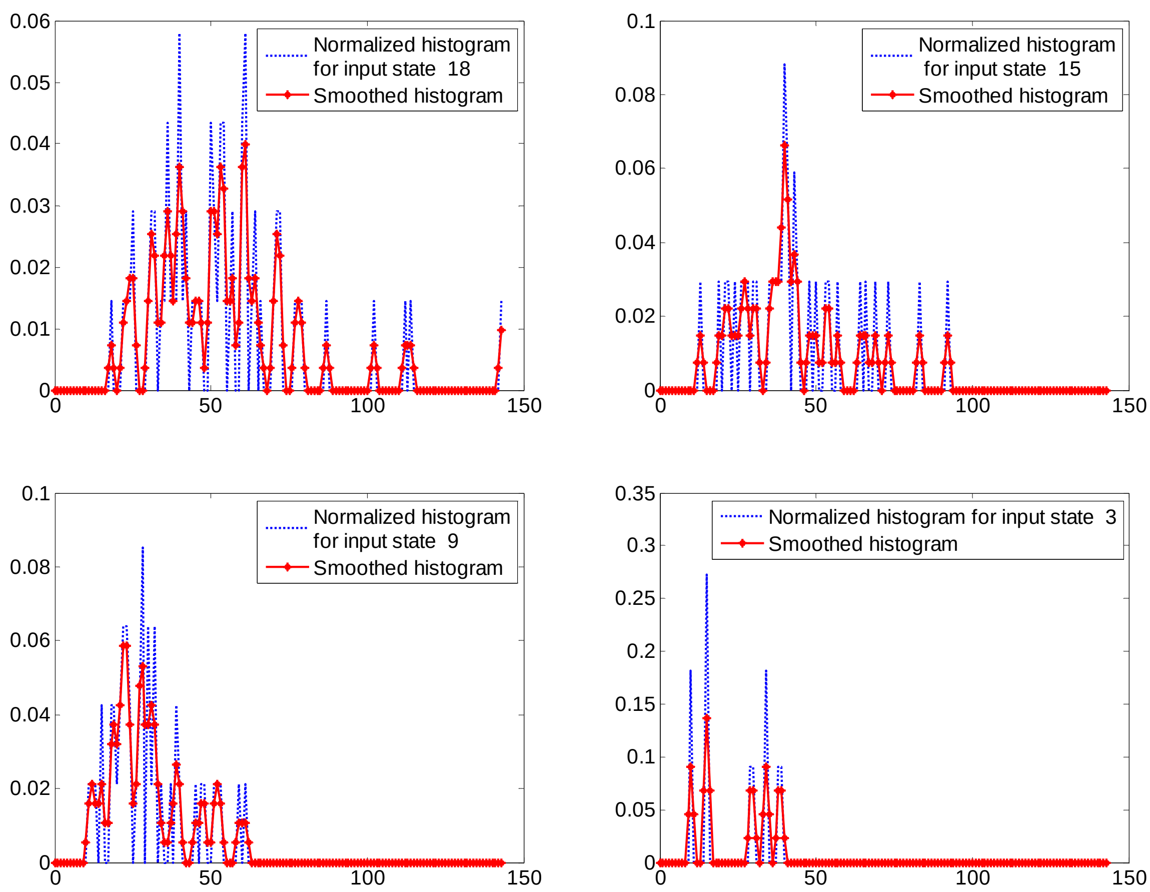

Figure 6.47 Examples of Normalized Histogram and Smoothed Histogram of Each Input States of the Bayesian Model with 18 Input States

6.4.2 Monte Carlo Algorithm Implementation

We have implemented Monte Carlo simulations based on normalized CPT of Bayesian model with 18 input states (MCS CPT18), smoothed histogram of Bayesian model with 18 input states (MCS H18), normalized CPT of Bayesian model with 9 input states (MCS CPT9) and smoothed histogram of Bayesian model with 9 input states (MCS H9).

The outline of the algorithm for MCS CPT18 and MCS CPT9 is provided below:

- Find out the input state for a given week
- Generate a uniform random number
- Using roulette wheel selection with this random number to select an outage level based on conditional probability table (not normalized by the bin sizes in outage levels)
- Generate another uniform random number
- Using roulette wheel selection with this random number select a value of outage from each outage level. The outages follow uniform distribution within one outage level.
- Repeat the simulation 10000 times for each week

The algorithm for MCS H18 and H9 is similar but requires less steps as given below:

- Find out the input state for a given week
- Generate a uniform random number
- Using roulette wheel selection with this random number to select an outage based on smoothed histogram
- Repeat the simulation 10000 times for each week

The animal-related data and weather data from 1998 to 2007 for Wichita, Topeka, Lawrence and Manhattan has a total of 480 weeks. Each week has a given input state, which determines the PDF for that week. Using this information, the algorithm generates one outage value for that week. Since the simulation is repeated for 10000 iterations, we get 10000 simulated sample points for each week. For each week, the expected outage is the mean of it's 10000 sample points. By simply adding up the sample points of 4 weeks in the same month in an iteration, we get 10000 sample points for monthly outages and by adding up the sample points of 48 weeks in an iteration we get 10000 sample points for the yearly outages. The mean of the

10000 simulations is taken as prediction instead of using the expected value computed by equation 6.1, which improves the performance since every outage has a chance to be generated instead of representing one outage level by only one value, the average value.

6.4.3 Confidence Interval

With 10000 sample points for every week, we can easily find out the confidence interval. The upper limit for 95% confidence is the smallest integer X such that the percentage of all the numbers below X exceeds 97.5% of the 10000 data points. The lower limits are assumed to be the biggest integer which makes the percentage of all the number below it smaller than 2.5%. For confidence intervals of the monthly and the yearly prediction, we do not aggregate the confidence intervals of the weekly prediction because we do not know if the probability distribution has additive property. We compute the confidence intervals based on the 10000 aggregated the monthly and the yearly data points in the same way as we do for the weekly data.

The upper limits give us a range in which the actual observed values are expected to lie, given the combination of the month type and the number of fair weather per week. As the amount of confidence is reduced, the range allowed for the predicted value decreases. With a lower confidence, more observed values might lie outside the predicted range of values. We didn't show the lower limits because the upper limits are more of concern. The upper limits are significant because they provide a benchmark for the utilities on animal-caused outages that can take place in the system. The utilities can take preventive actions based on these upper limits.

6.5 Simulations for Four Cities

The results for the weekly, the monthly and the yearly predictions by MCS CPT18 are shown in Figure 6.48-6.55. In predictions of 480 weeks as given by MCS CPT18, only these three weeks (the 183rd, 194th and 204th) for Wichita, five weeks (the 116th, 160th, 190th, 194th and 386th) for Topeka and five weeks (127th, 131st, 194th, 260th and 433rd) for Manhattan had observed values outside of the upper limits of 95% confidence. And all the observations for Lawrence lie within the upper limits of 95% confidence. For the monthly predictions, out of 120 months there are only four months (the 37th, 46th, 49th and 51st) for Wichita, two months (49th and 97th) for Topeka and one month (the 82nd) for Manhattan exceeding the 95% upper confidence limits. Again there is no monthly observation in Lawrence beyond 95% upper confidence limits. For the yearly predictions, there is one year with observed values outside of the upper limits for

Wichita and Manhattan. In other words, no matter for the weekly, the monthly or the yearly predictions, there are at least 90% of the observed values found to lie within the upper limits of 95% confidence limits for all the cities. Even though the sizes of the four cities vary and the patterns in outages change from city to city, the method based on Bayesian models and Monte Carlo simulations provides good performance for every city.

We observed that the predictions catch the pattern in the observed time series better when the base of aggregation increases. The reason behind this is that bigger aggregation somewhat even outs the time series and results in more consistent pattern in the data. Therefore the monthly predictions approximate the observed values better than weekly prediction and there are more observed values in the monthly predictions within the 95% upper confidence limits than in the weekly predictions. We may expect better performance of the yearly predictions than the monthly prediction but it's not the case. When the base for aggregation is too big, 48 weeks in the yearly predictions as shown in Figure 6.56, too much information is ignored and the predictions tend to flatten out over the years since the weather conditions and the month type are similar from year to year. The weekly and the monthly predictions by Monte Carlo simulations based on other three approaches are shown in Figure A.1-16 in Appendix A, which gives similar observations to the ones by MCS CPT18. The yearly predictions by MCS CPT9, MCS H18 and MCS H9 are shown in Figure 6.57-6.59. The yearly predictions by MCS CPT18 and MCS H18 can both catch the variances in the observed outages while the ones by MCS CPT9 and MCS H9 have failed to do so. Therefore, for the rest of the analysis we have not used the results of MCS CPT9 and MCS H9 models. Some years like 2001 for Wichita, 2000 for Topeka, 2006 for Lawrence and 2004 for Manhattan fall out of the upper limits of 95% confidence. The further investigation for these years is given in the next section.

Figure 6.60 and 6.61 show the observed outages and the expected outages in year 2007 for four cities by MCS CPT18 and MCS H18. The expected outages are the mean of 10,000 simulations. Figure 6.62-6.65 shows examples of results of MCS CPT18 in one week, one month and one year for each city. The weekly simulations tend to regenerate the CPT used as probability distribution. The monthly and the yearly simulations develop their own distributions which are similar to Gaussian distribution.

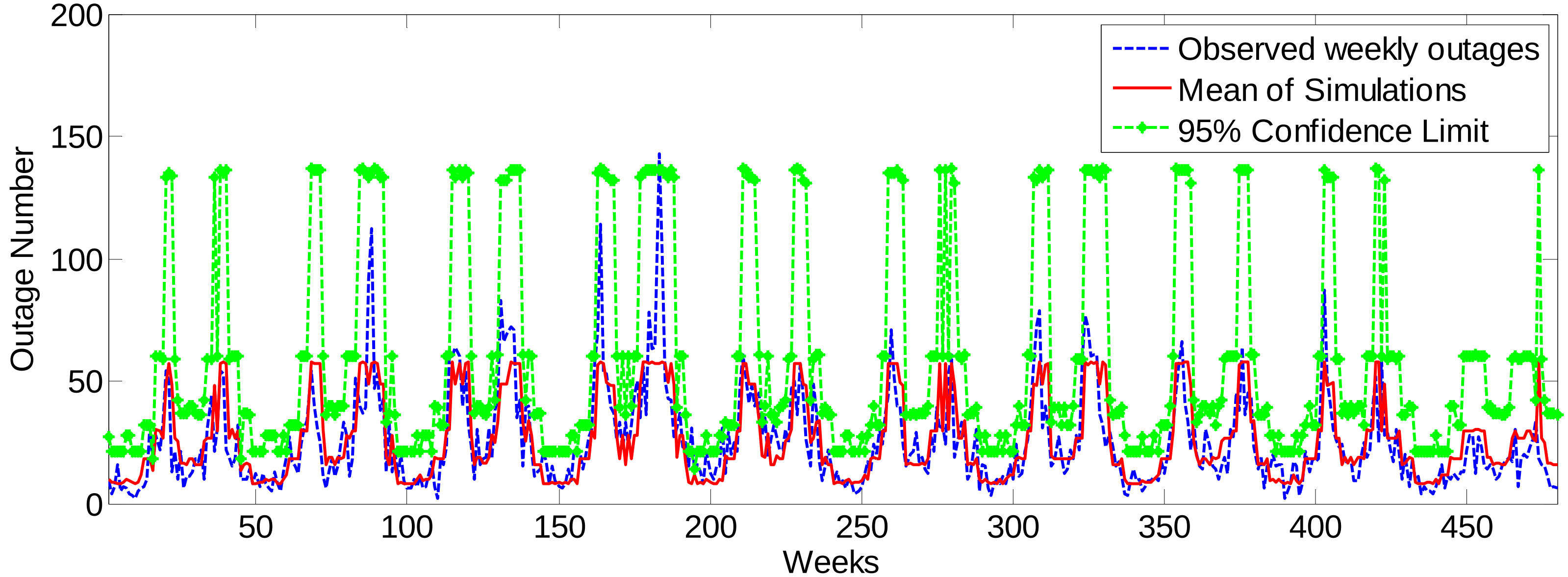

Figure 6.48 Weekly Prediction and 95% Confidence Limit by MCS CPT18 for Wichita

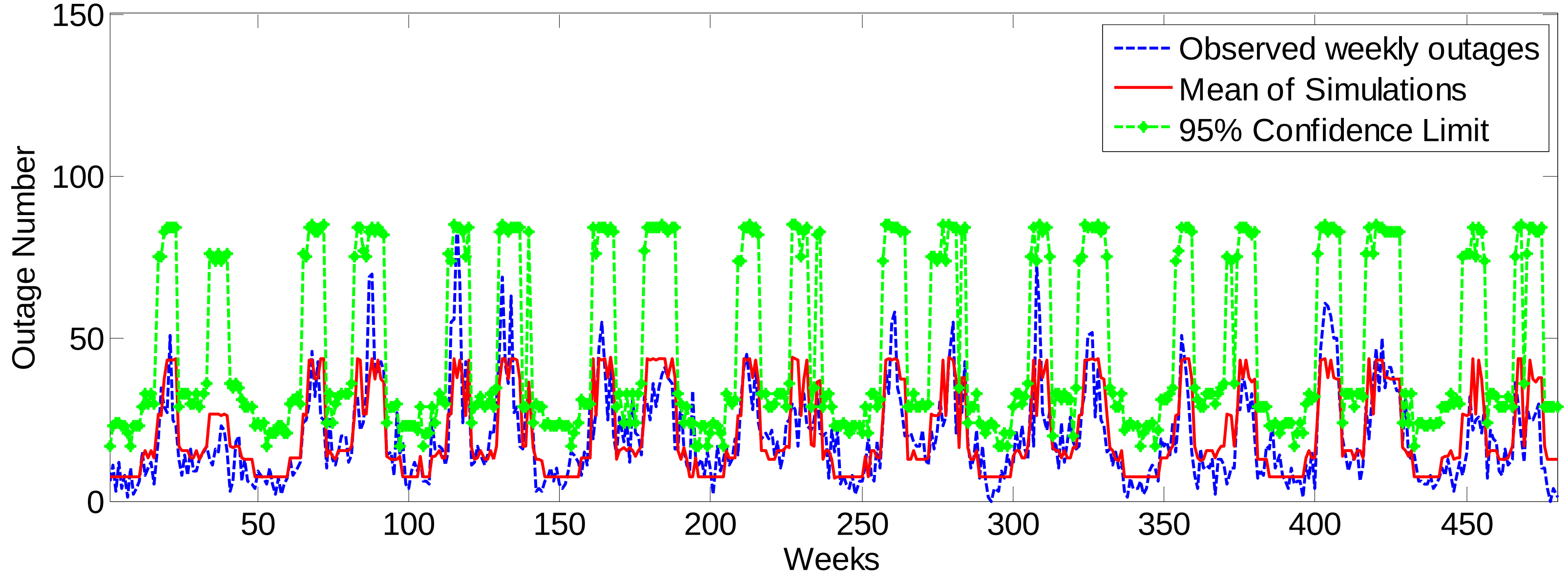

Figure 6.49 Weekly Prediction and 95% Confidence Limit by MCS CPT18 for Topeka

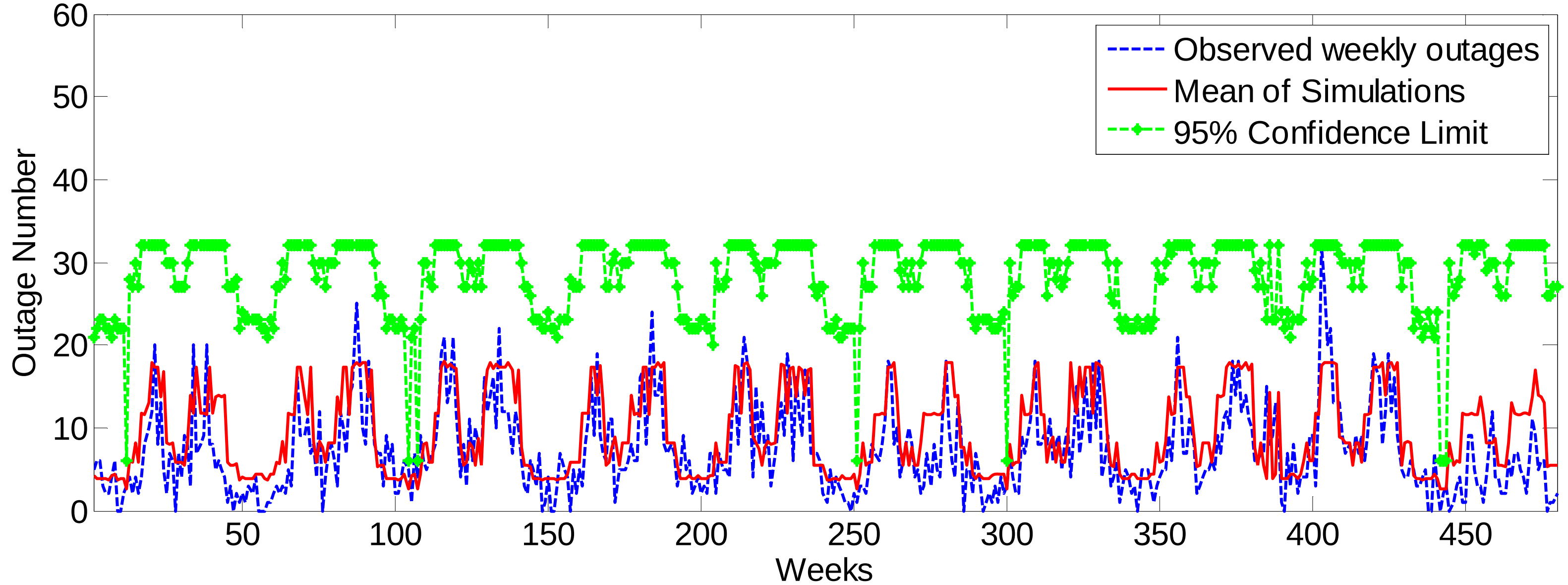

Figure 6.50 Weekly Prediction and 95% Confidence Limit by MCS CPT18 for Lawrence

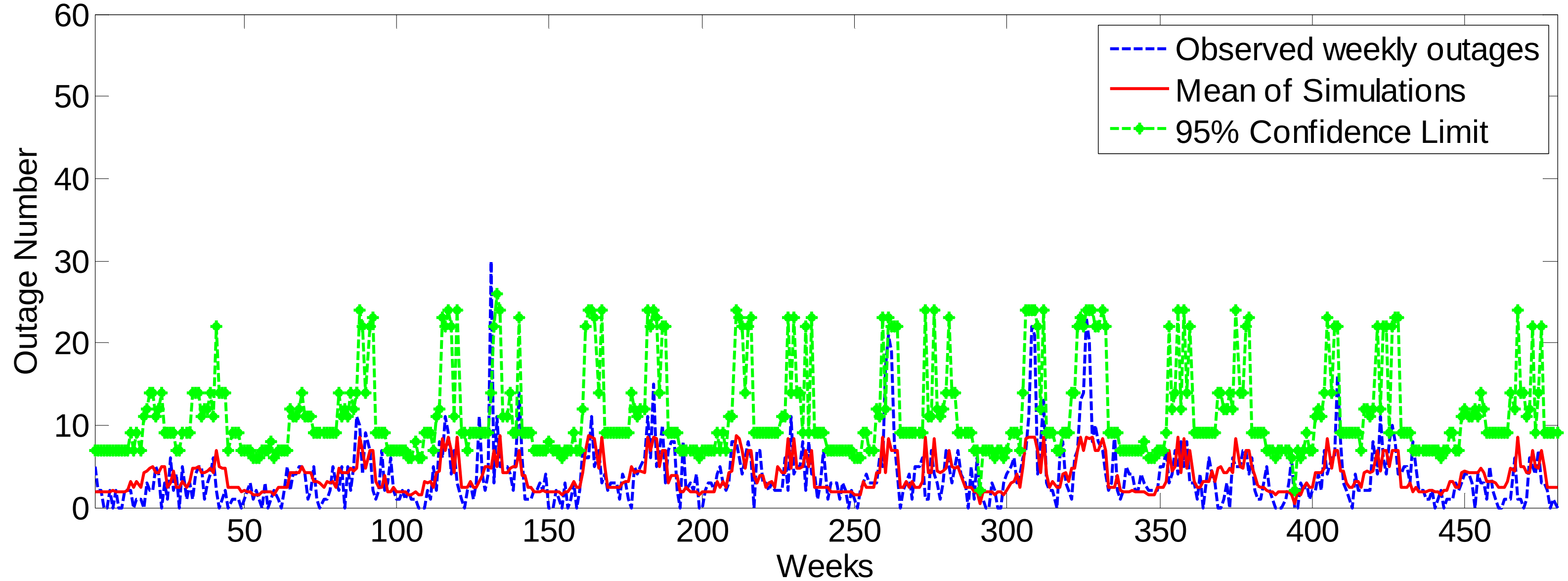

Figure 6.51 Weekly Prediction and 95% Confidence Limit by MCS CPT18 for Manhattan

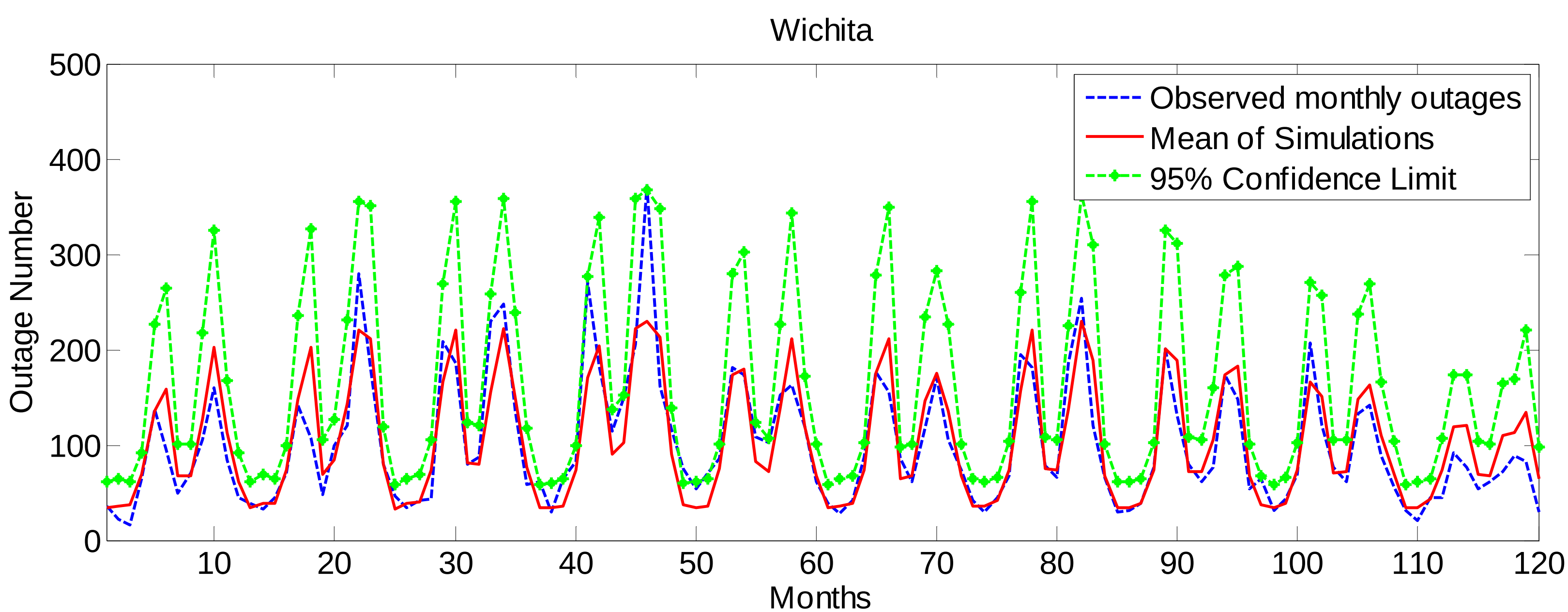

Figure 6.52 Monthly Prediction and 95% Confidence Limit by MCS CPT18 for Wichita

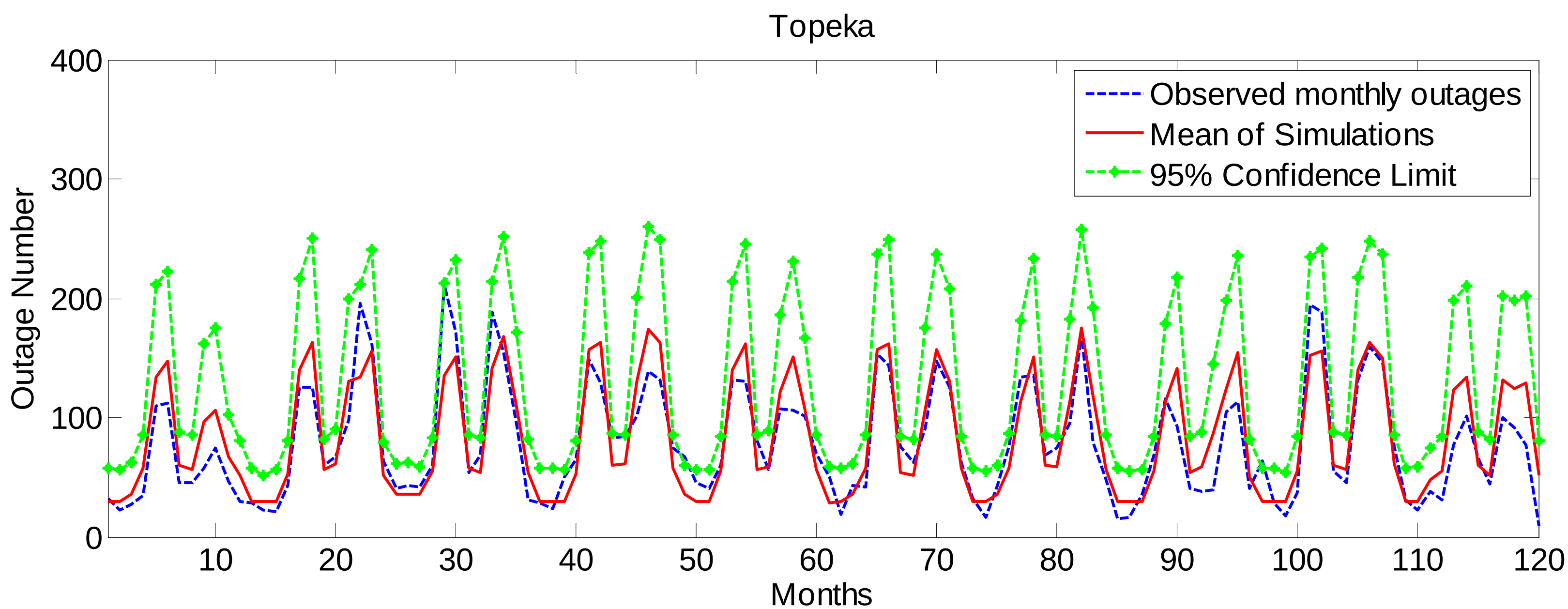

Figure 6.53 Monthly Prediction and 95% Confidence Limit by MCS CPT18 for Topeka

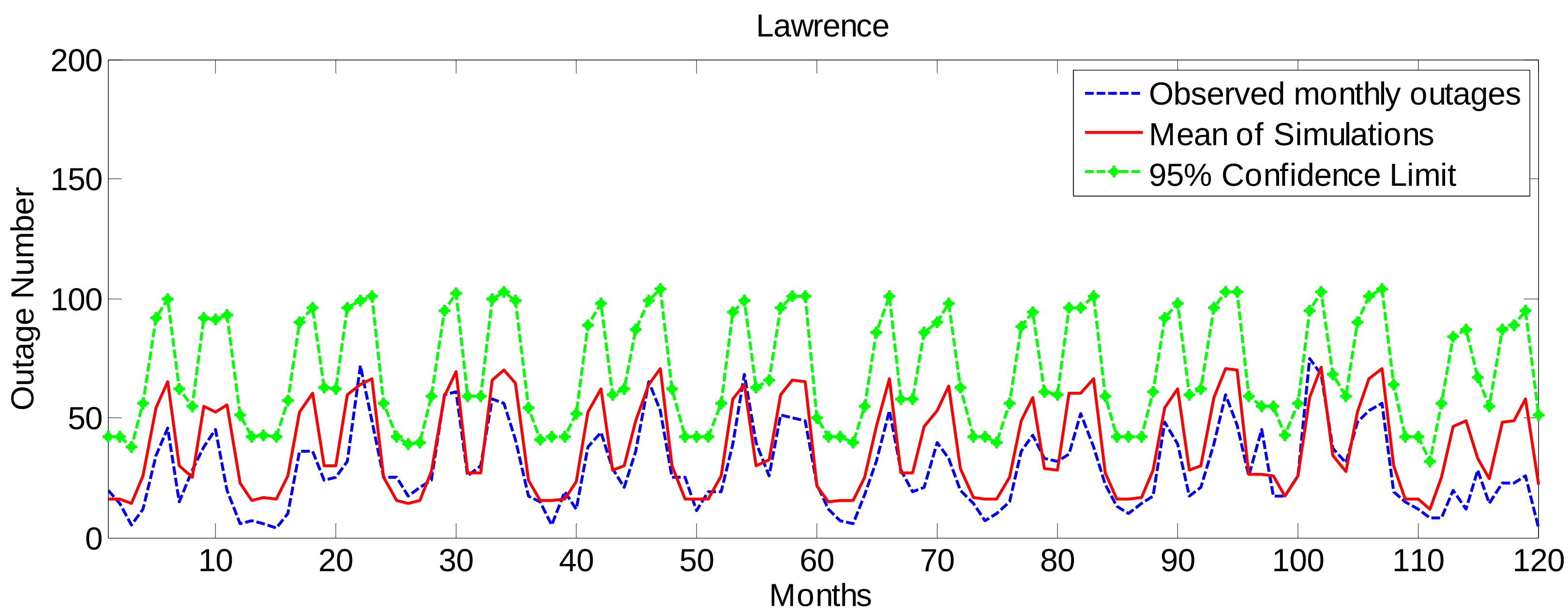

Figure 6.54 Monthly Prediction and 95% Confidence Limit by MCS CPT18 for Lawrence

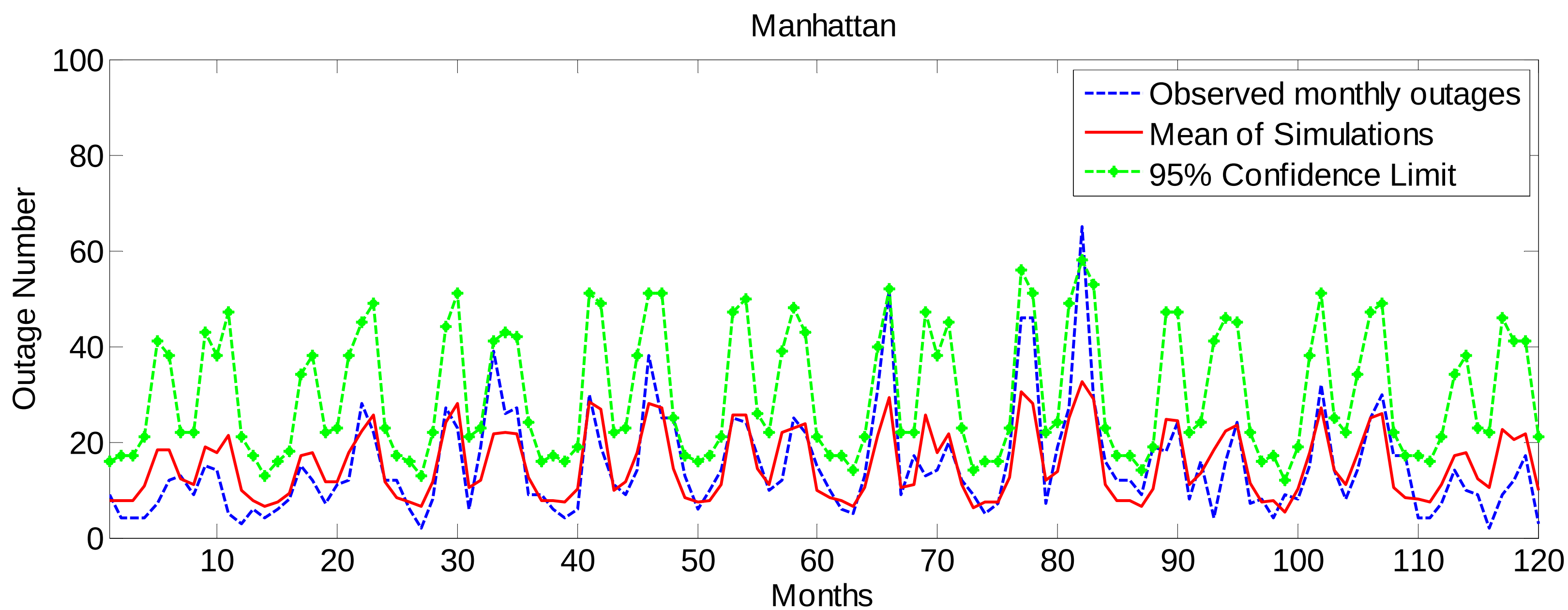

Figure 6.55 Monthly Prediction and 95% Confidence Limit by MCS CPT18 for Manhattan

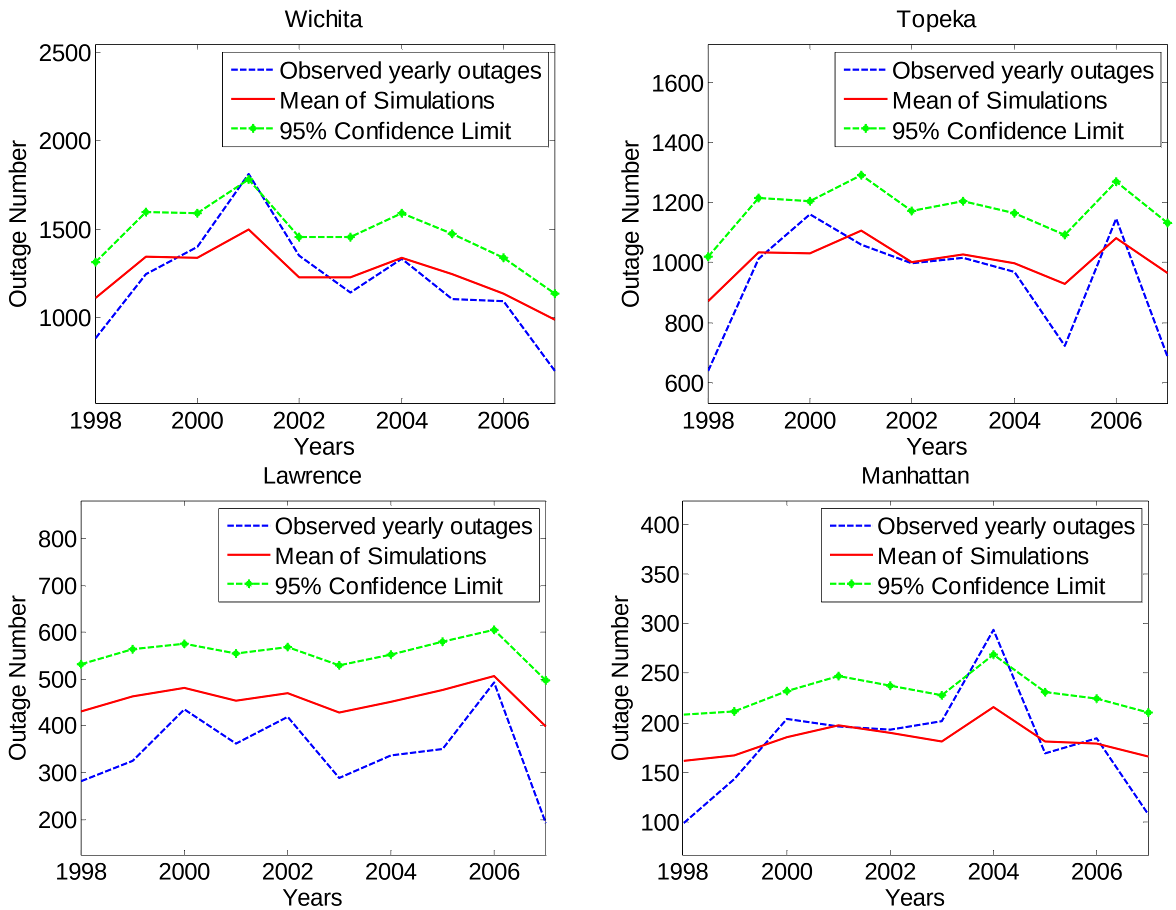

Figure 6.56 Yearly Predictions and 95% Confidence Limits by MCS CPT18 for Four Cities

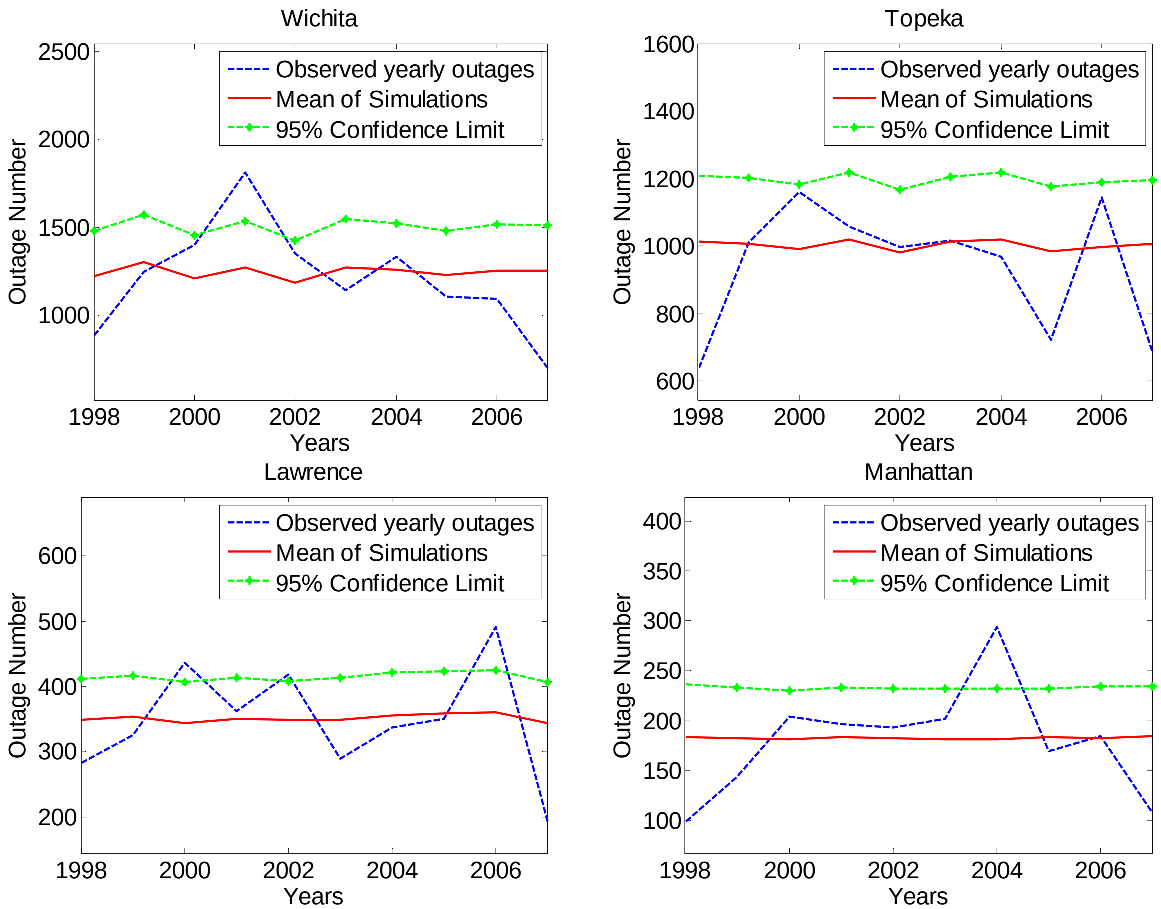

Figure 6.57 Yearly Predictions and 95% Confidence Limits by MCS CPT9 for Four Cities

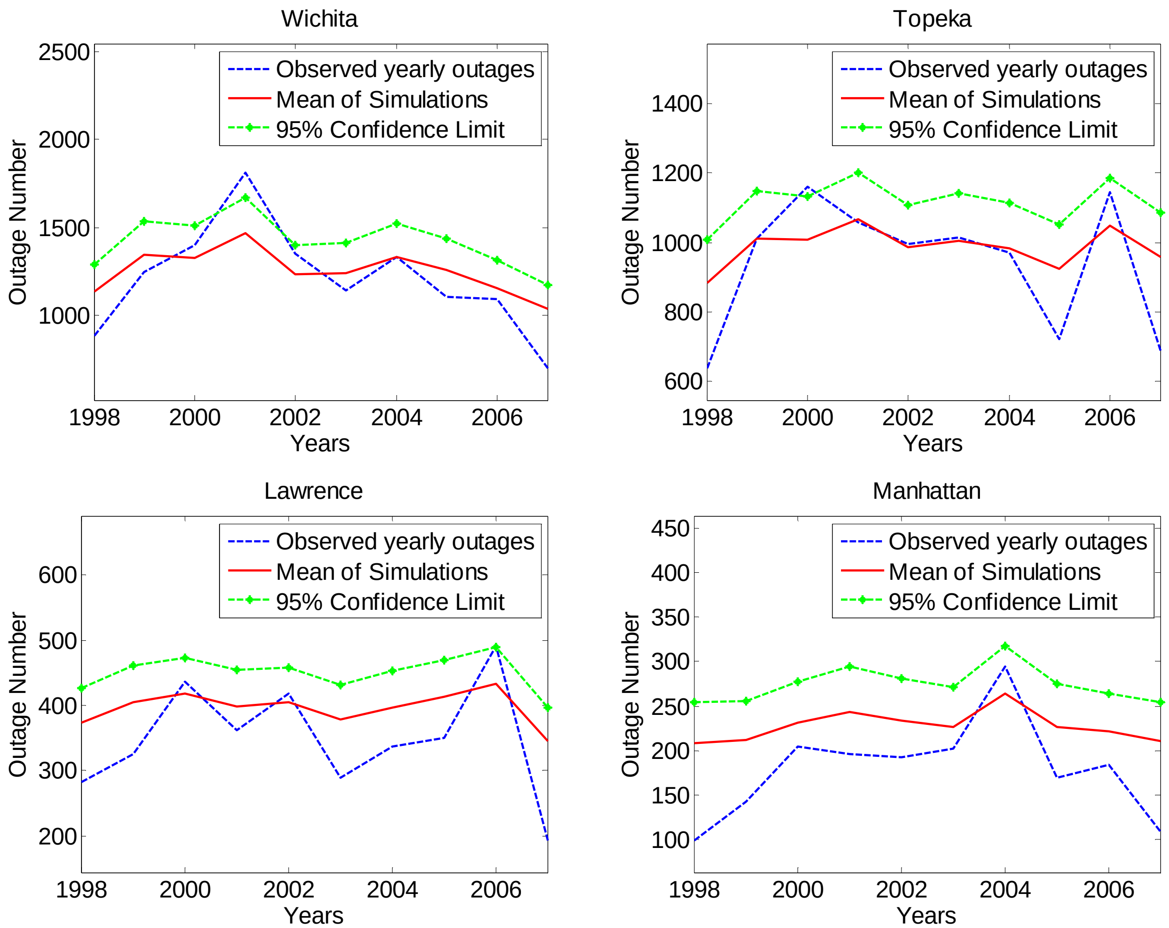

Figure 6.58 Yearly Predictions and 95% Confidence Limits by MCS H18 for Four Cities

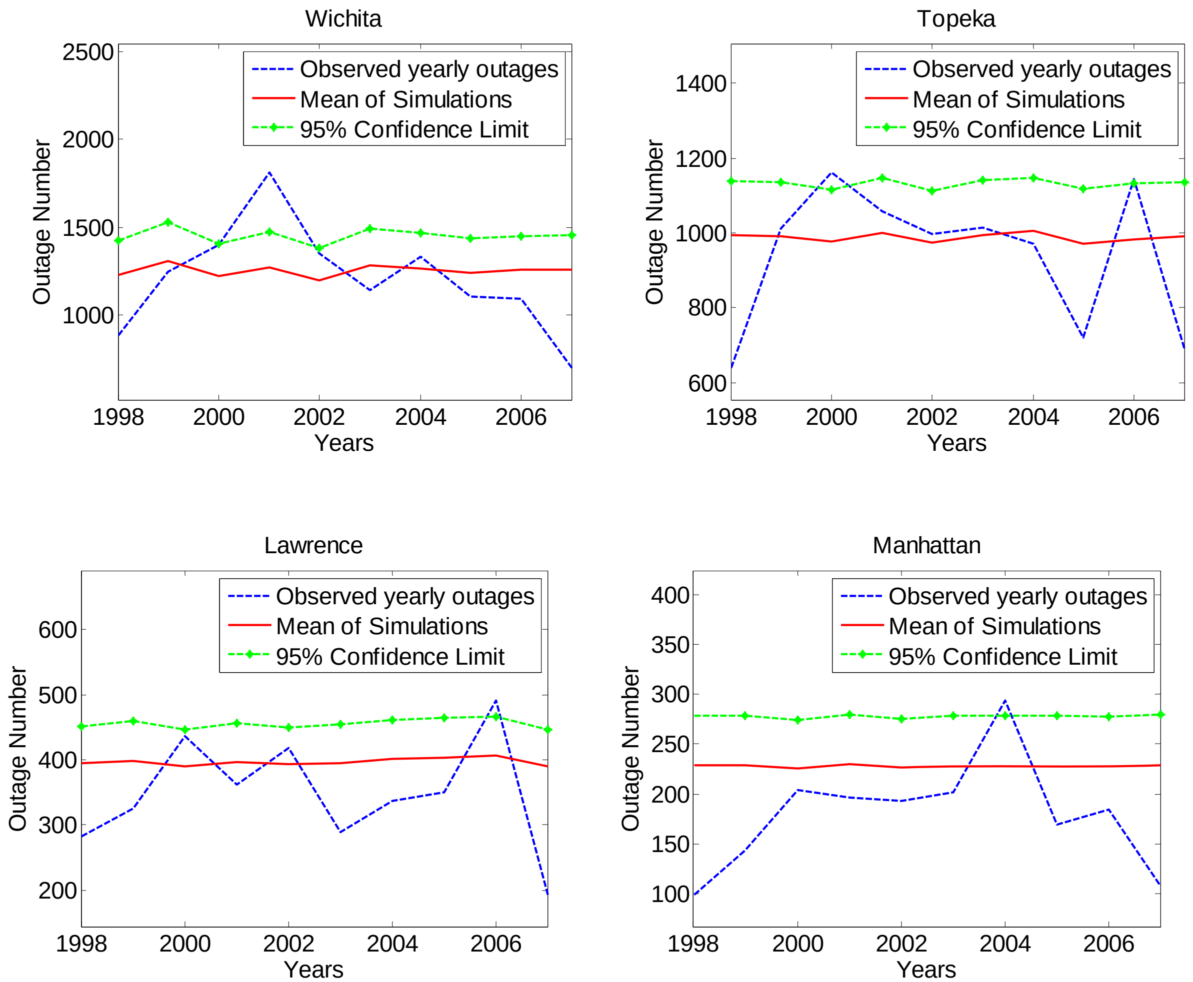

Figure 6.59 Yearly Predictions and 95% Confidence Limits by MCS H9 f or Four Cities

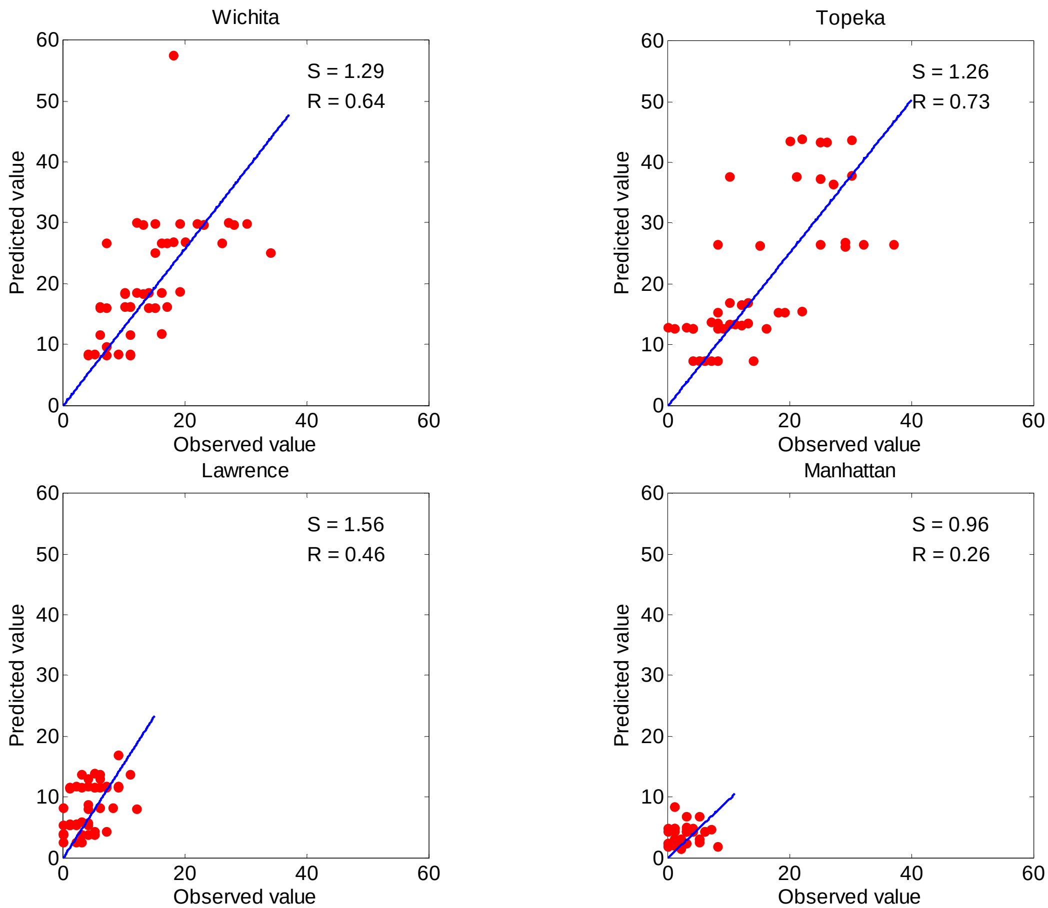

Figure 6.60 Outages Predicted and Observed by MCS CPT18 in 2007 for Four Cities

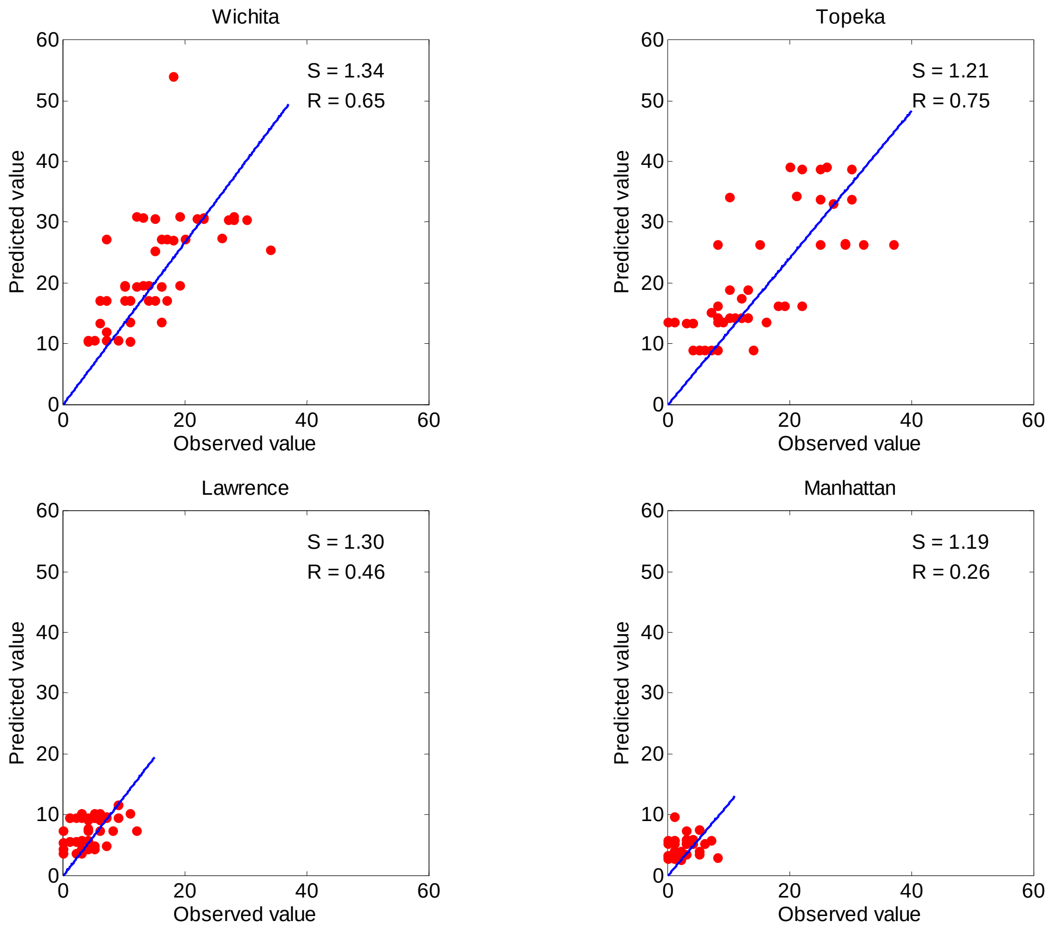

Figure 6.61 Outages Predicted and Observed by MCS H18 in 2007 for Four Cities

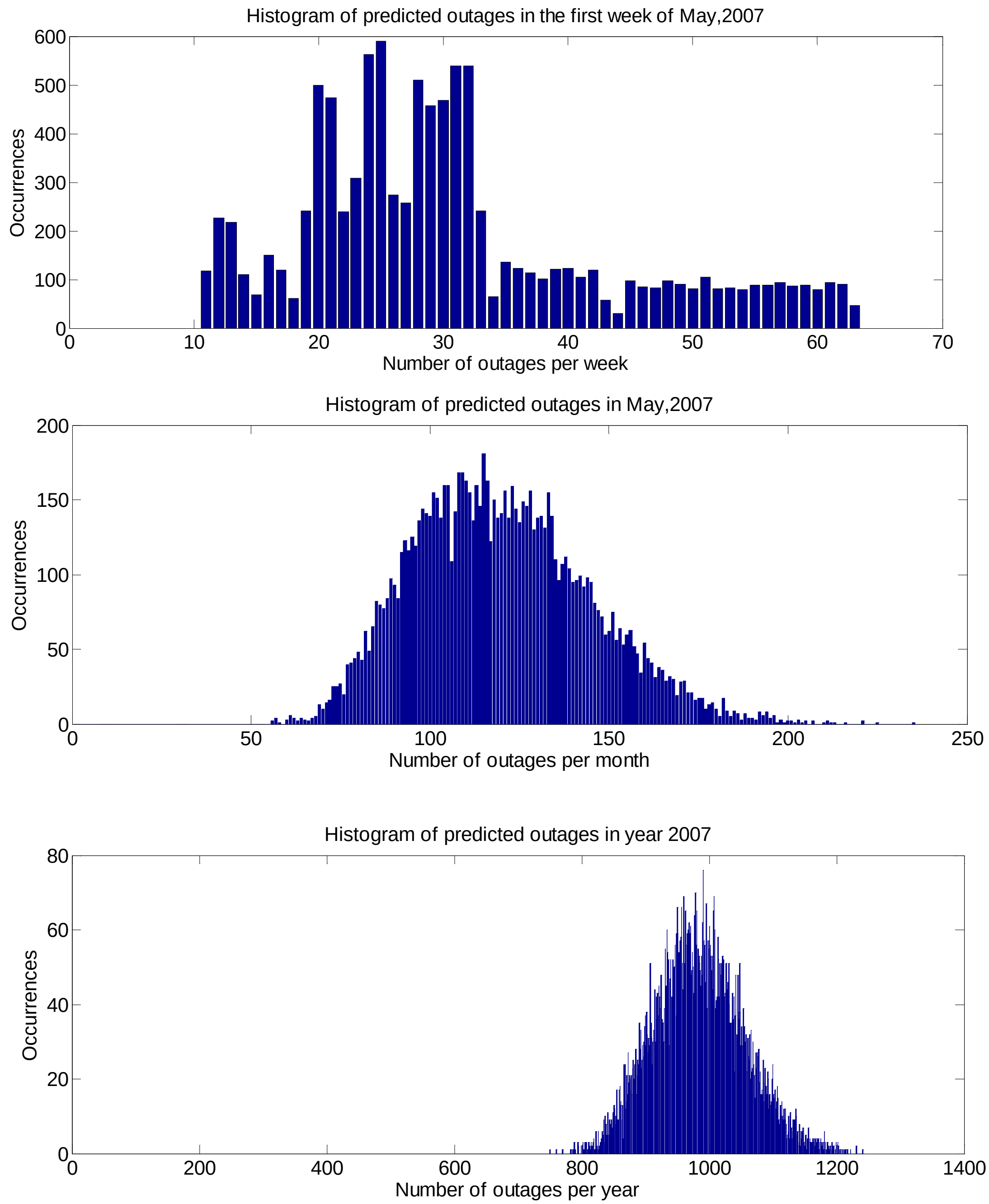

Figure 6.62 Examples of Predicted Weekly, Monthly and Yearly Outages for Wichita by MCS CPT18

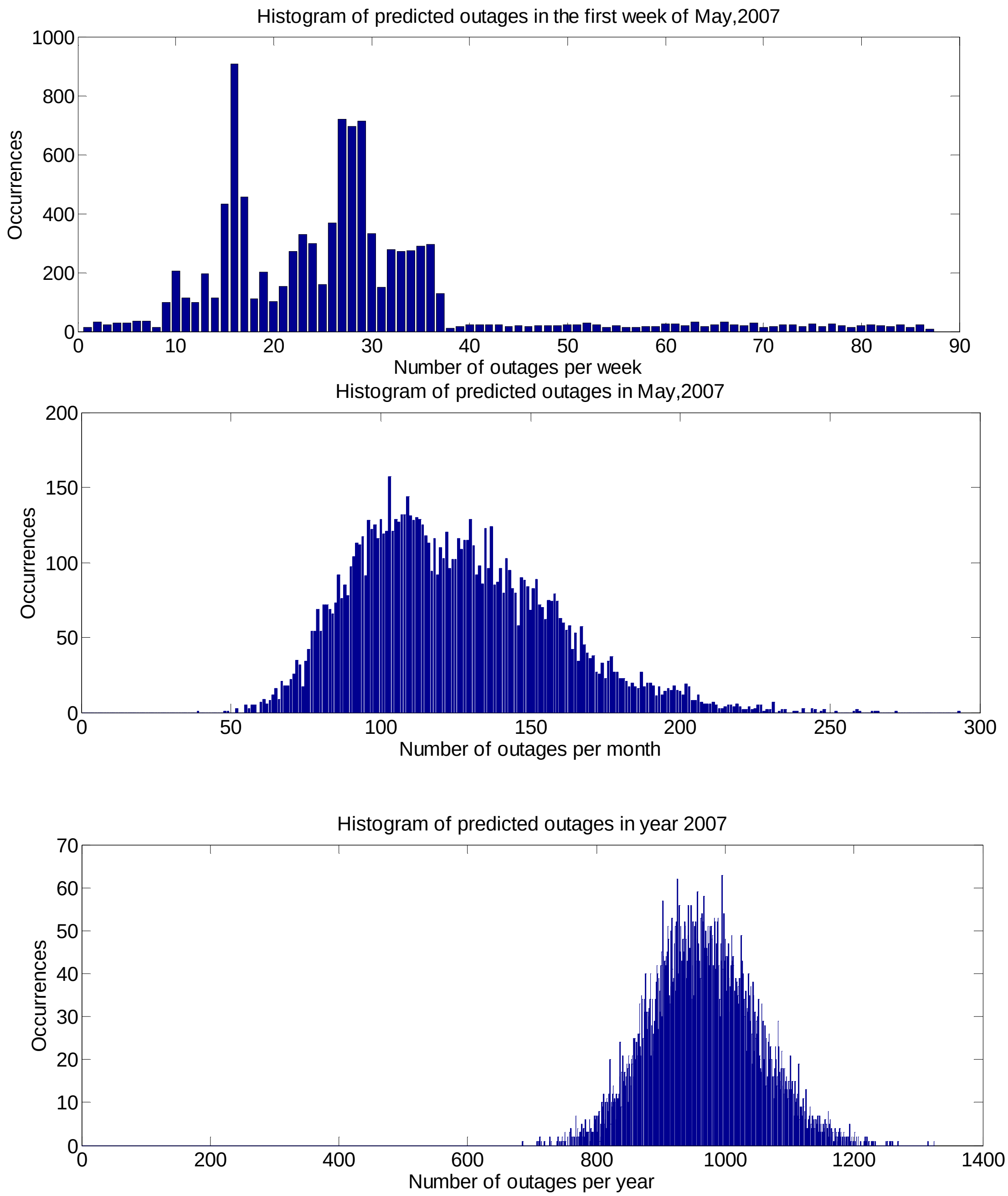

Figure 6.63 Examples of Predicted Weekly, Monthly and Yearly Outages for Topeka by MCS CPT18

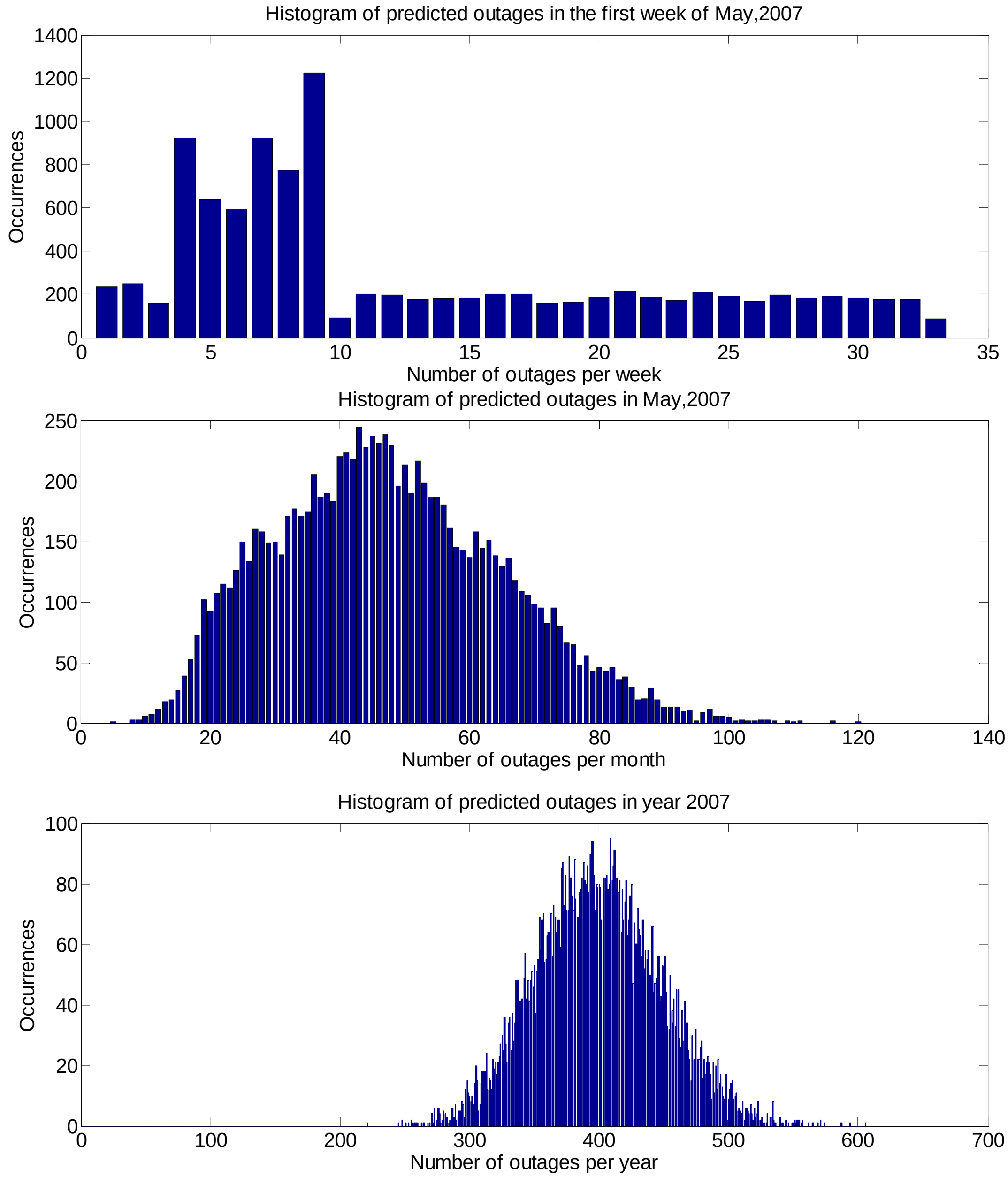

Figure 6.64 Examples of Predicted Weekly, Monthly and Yearly Outages for Lawrence by MCS CPT18

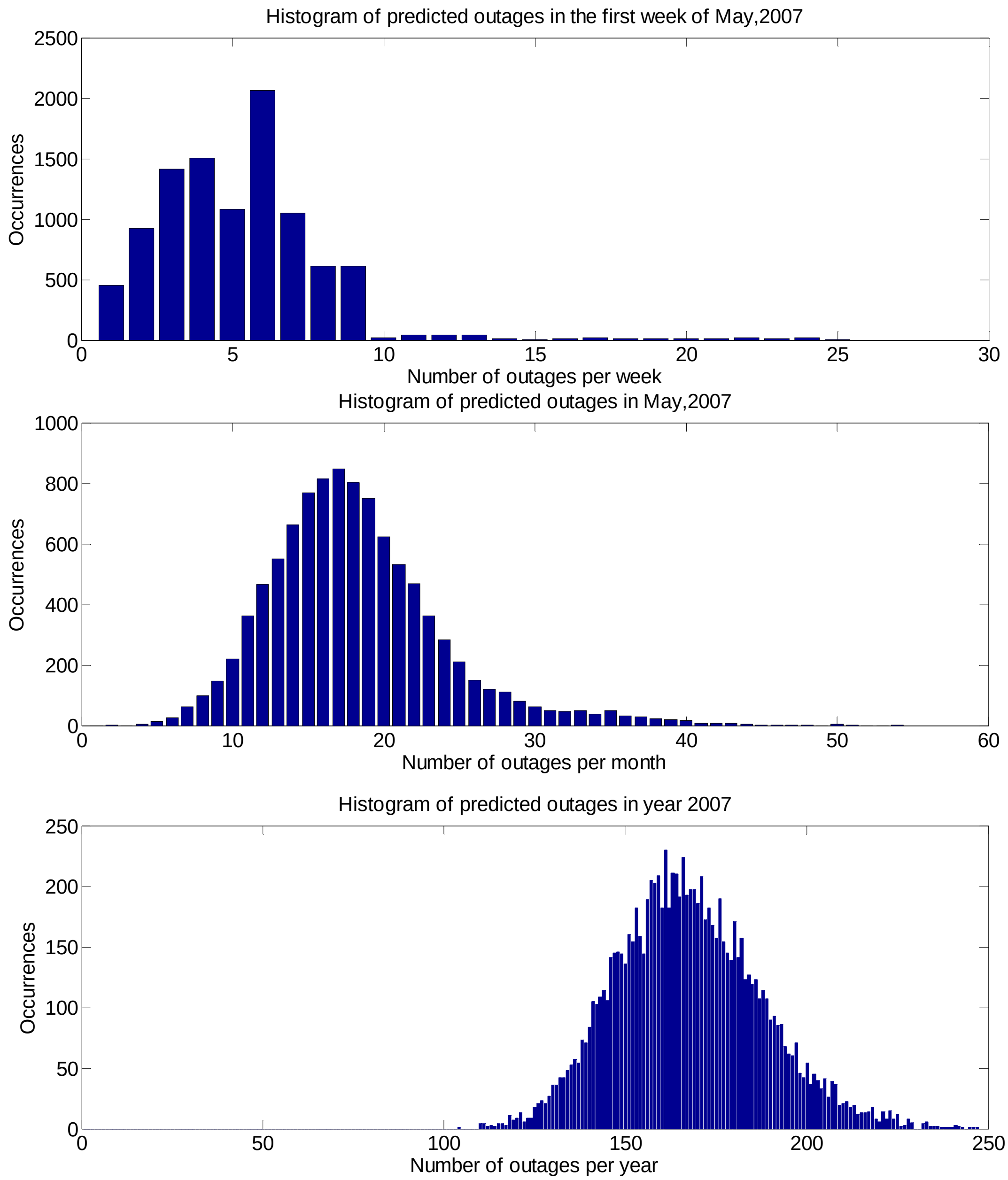

Figure 6.65 Examples of Predicted Weekly, Monthly and Yearly Outages for Manhattan by MCS CPT18

6.6 Investigation of Outliers in Yearly Predictions

Since the two models MCS CPT18 and MCS H18 showed different results, we need to further investigate the years which had yearly observations higher than the upper limits shown by either of the two models. We have used the MWNN model to investigate the outliers in the yearly observations due to its high performance in predictions. Therefore, the years which are outliers are taken as testing data and the rest are taken as training data. Hence, the outage data is rearranged by putting the outlier year as the last year in the time series. The training error in the weekly prediction for all the cities are shown Figure 6.66. These graphs approximate Gaussian distribution with mean close to zero for each city. Based on the distribution of training error we can estimate the confidence interval for the testing year. The summation of variances of 48 weeks is the variance of one year and for Gaussian distribution, two-Sigma gives the 95% confidence interval. The results of computation are shown in Table 6.30. Only the year 2001 for Wichita lies outside of upper limit of 95% confidence which indicates that this may be the only outlier. The plots for predicted and observed outages are shown in Figure A.17-20 in Appendix A.

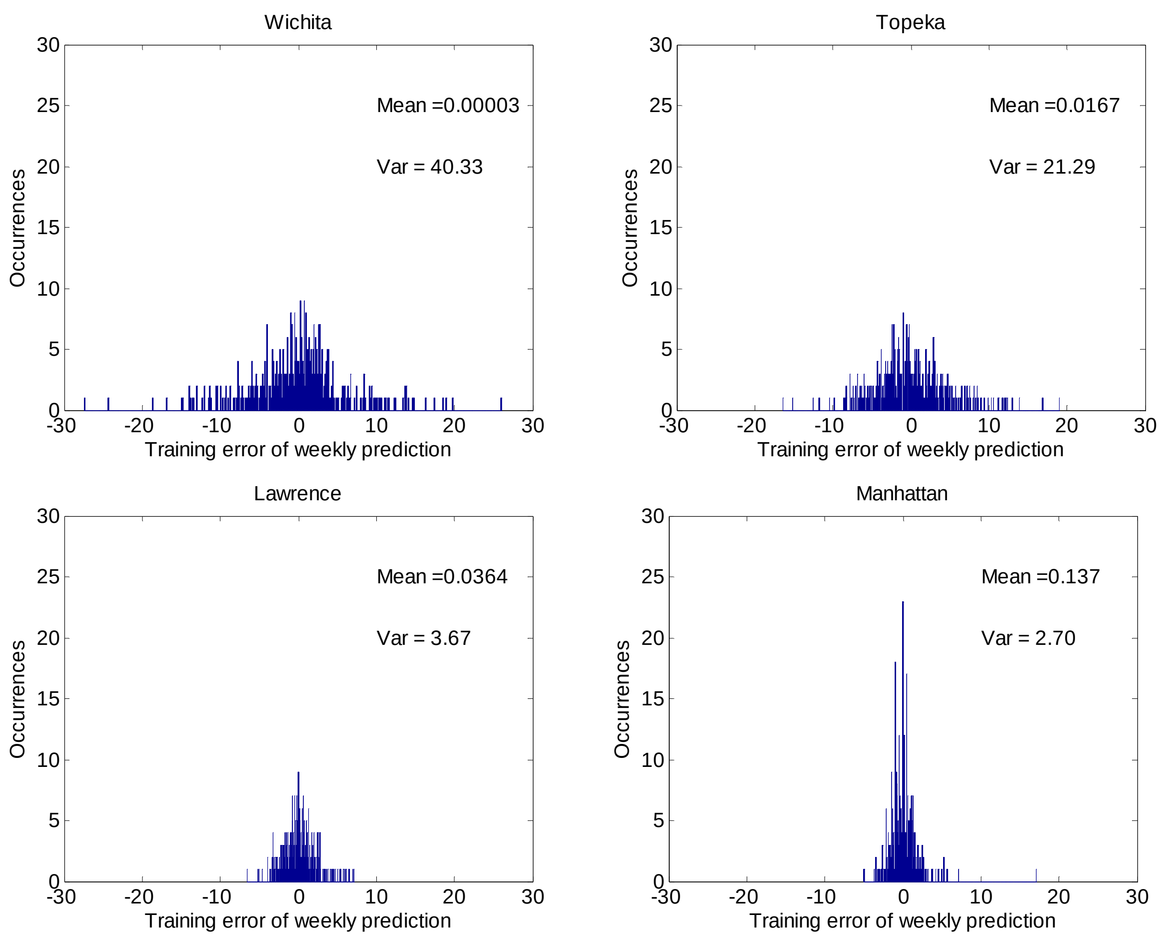

Figure 6.66 Training Errors of Weekly Predictions by MWNN Model for Four Cities

Table 6.39 Predicted Yearly Outages and 95% Confidence Limits for Outliers by MWNN Model

	Wichita 2001	Topeka 2000	Lawrence 2006	Manhattan 2004
Observed Outages	1809	1160	492	294
Predicted Outages	1463	1099	480	274
Two-Sigma	88	64	27	23
95% Upper Limit	1551	1163	507	297
Outlier(Y/N)	Y	N	N	N

CHAPTER 7 - Conclusions and Future Work

7.1 Comparison of Models

This dissertation presented four different methods including the Poisson regression, the Neural Network, the wavelet based Neural Network and the Bayesian model combined with Monte Carlo simulations to predict animal-related outages on the overhead distribution feeders. The performance of every model was tested with data from four cities in Kansas, Wichita, Topeka, Lawrence and Manhattan which represent the two biggest cities and two smaller cities in Kansas. The data was aggregated on a weekly basis since bigger sample size could even out the randomness in the daily data. Simulations showed that all the methods were able to catch the patterns in the time series of weekly data. The AAE was used to evaluate the predictions of different models as showed in Table 7.1. Note that the AAE for NN models are overall AAE for the whole data set which includes training data and testing data. The slopes of best-fit lines between predicted and observed outages are shown in Table 7.2 and the correlation coefficients between them are shown in Table 7.3. Because the random nature and the fast fluctuations in the data, the traditional linear models, the Poisson regression models and the Bayesian models faced a bigger challenge in catching the high pikes in the time series. The NN models which have the ability to approximate highly complex functions outperformed the traditional linear models. However, the NN models still could not give accurate predictions in the extreme cases which demanded techniques like wavelet decomposition to preprocess the fluctuating data. Wavelet based NN models decomposed the data into approximate and detail subseries, constructed NN predictors for the subseries and summed up the outputs from the NN predictors to get the final prediction for the original data. The Artificial Immune System was applied to overcome the overtraining problem in the application of NNs. The simulations showed wavelet decomposition and AIS techniques greatly improved the performance of the NN predictors: the AAE for Wichita dropped from 7.54 events per week to 4.67, for Topeka it dropped from 6.35 to 3.7, for Lawrence it dropped from 2.8 to 1.5 and for Manhattan it dropped from 1.95 to 1.22.

The Poisson regression model is intuitive in the way that it considers the counting nature of outage events and assumes Poisson distributed response variables. The assumption of Poisson distributed outage events is based on the condition of rare events of large numbers which was proven to be true from analysis of the historical data. The linear terms in the Poisson regression model (3.19) consist of variables for the number of fair days per week, the month type and an interaction term. This construction gives the Poisson regression model the ability to predict the future outages without dependence on the historical outages once the model has been trained, which explained the bigger errors in the Poisson regression model than the other models since given the same weather conditions the outage level varies from year to year. In all the other models the outages in previous weeks are taken as one explanatory variable.

Table 7.1 The AAE of Four Models for Four Cities

Models	Wichita	Topeka	Lawrence	Manhattan
MCS CPT18	8.52	6.80	3.67	2.14
MCS CPT9	9.67	7.51	3.18	2.19
MCS H18	8.46	6.64	3.14	2.44
MCS H9	9.70	7.36	3.35	2.53
Bayesian with 18 Input States	8.31	6.480	2.98	2.12
Bayesian with 9 Input States	9.53	7.17	3.17	2.18
AIS	4.67	3.70	1.50	1.22
MWNN	4.71	3.77	1.53	1.25
WNN	6.75	5.62	2.39	1.51
NN	7.54	6.35	2.80	1.95
Poisson	9.89	8.55	3.28	2.23

Table 7.2 The Slopes of Best-fit Line for Four Cities by Different Models

Models	Wichita	Topeka	Lawrence	Manhattan
MCS CPT18	0.90	0.94	1.08	0.66
MCS CPT9	0.84	0.90	0.80	0.62
MCS H18	0.87	0.89	0.91	0.78
MCS H9	0.83	0.86	0.88	0.74
Bayesian with 18 Input States	0.84	0.86	0.82	0.64
Bayesian with 9 Input States	0.80	0.82	0.78	0.60
AIS	1.00	1.00	1.00	1.00

MWNN	1.00	1.00	0.99	1.00
WNN	1.02	1.02	1.03	1.06
NN	0.90	0.90	0.86	0.76
Poisson	0.79	0.76	0.78	0.60

Table 7.3 The Correlation Coefficients for Four Cities by Different Models

Models	Wichita	Topeka	Lawrence	Manhattan
MCS CPT18	0.77	0.77	0.68	0.53
MCS CPT9	0.68	0.70	0.62	0.45
MCS H18	0.77	0.77	0.69	0.54
MCS H9	0.68	0.70	0.62	0.45
Bayesian with 18 Input States	0.77	0.77	0.69	0.53
Bayesian with 9 Input States	0.68	0.70	0.62	0.45
AIS	0.94	0.94	0.94	0.89
MWNN	0.94	0.94	0.94	0.89
WNN	0.91	0.90	0.87	0.89
NN	0.82	0.80	0.72	0.68
Poisson	0.66	0.54	0.59	0.43

The NN models with the number of fair days per week, the month type, the outages from previous 4 weeks as inputs, are able to approximate the complex relations between the weather conditions and the animal-related outages and learn the patterns in the historical data, which makes it outperform the traditional linear method such as Poisson regression. The number of inputs and the network structure were optimized by experimentation.

The NN models were greatly enhanced by wavelet techniques. Application of the wavelet decomposition to the time series of weekly animal-related outages provided a deeper insight into the pattern of these outages. Wavelet decomposition with resolution level 1 was found out to be the optimal. To overcome the overtraining problem associated with neural networks, this dissertation proposed a hybrid approach with AIS used for hypermutating and retraining the networks during the testing stage. As shown in Table 7.2 and 7.3, this hybrid approach gave significantly high performance of predictions. For the predictions for all the cities, the slopes of best fit linear functions are extremely close to the ideal number 1 which means the predicted values follow the observed values very well. And the correlation coefficients are close to 1 too which indicates the high degree of linear pattern between predicted values and observed values. Also the results suggest that spatial aggregation increases the accuracy of the models. Overall, the wavelet based NN models outperformed all other models and gave significantly good performance.

The method of Bayesian Model and Monte Carlo simulations are proposed for further analysis of the outages. The Bayesian model categorized numerous combinations of the month type, the number of fair days per week and the previous week outage level into discrete input states and tried to capture the probabilistic relationships between each input state and outage levels in the conditional probability table. The Bayesian model allowed computation of the expected number of outages in each weather state. Based on the conditional probability table and histogram of outages as probability distribution functions, Monte Carlo simulations were carried out to find out the confidence intervals of the predictions. Simulations showed that Monte Carlo Simulation based on CPT with 18 input states (MCS CPT18) outperformed Monte Carlo Simulation based on CPT with 9 input states (MCS CPT9), Monte Carlo Simulation based on smoothed histogram with 18 input states (MCS H18) and Monte Carlo Simulation based on smoothed histogram with 9 input states (MCS H9). For MCS CPT18, at least 90% percent of the observed values are within the upper limits of 95% confidence of the predicted values for every city which indicates the prediction is reliable. Even though it was prone to underestimate in the extreme cases, this model successfully tracked the monthly fluctuations from year 1998 to 2007 and approximated the observed values better than the weekly prediction and there are more observed values in the monthly predictions within the 95% upper confidence limits than in the weekly predictions. When the base for aggregation is too big, 48 weeks in the yearly predictions, too much information is ignored and the predictions tend to flatten out over the years since the weather conditions are similar from year to year. The upper limits are significant because they provide a way for the utilities to reduce the number of animal-caused outages that can take place in the system. The utilities can take preventive actions if outages are higher than based on these upper limits.

The approach presented in this paper is useful to utilities for end of the year analysis of past year's reliability performance of the distribution systems. Performance of a specific year can be compared with the past performance to identify deviations. Significant increase in outages would require the utility to do further analysis and take remedial actions. And the confidence limits on the number of outages caused by animals allow the utilities to define a cutoff point beyond which further action would be needed.

In the yearly analysis, when the observed outages in a certain year lie outside of the upper limits, we call this year an outlier. Since the two models MCS CPT18 and MCS H18 showed different results for outliers for four cities, we further investigated the years which had yearly observations higher than the upper limits shown by either of the two models. We have used the MWNN model to investigate the outliers in the yearly predictions due to its high performance in predictions. It found only year 2001 for Wichita as an outlier and no outlier in other three cities.

7.2 Future Work

The data we have only carry basic information, daily incident number for a whole distribution network in one city and weather conditions for a whole city. More information in the data is needed for precise treatments of localization of the outages to help utilities in locating the weak points of the network and take actions on that location to lower the outage rate. Also tree density information along the overhead feeders will greatly improve the predictions of animal-related outages since tree attracts animals and animals can cause outages indirectly through trees. For improvement of modeling, tree density can be taken as one input to the Bayesian model. Since the wavelet based neural network model performed well, it can also be applied to study the lighting and wind effects.

References

[1] E. Miranda, "Animal Intrusion in Substations and Transmission Towers," Southern California Edison, Rosemead, California, IEEE. Rep. 0-7803-5569-5/99, 1999.

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

[3] Y. Zhou and A. Pahwa, "Prediction of Weather-related Failures of Overhead Distribution Feeders," Thesis, May 2004, pp. 61 – 63.

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

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

[6] R. N. Allan, R. Billington, A. M. Breipohl and C. H. Grigg, "Bibliography on the Application of Probability Methods in Power System Reliability Evaluation: 1988 – 1991," *IEEE Transactions on Power Systems,* Feb. 1994, pp. 41 – 49.

[7] R. N. Allan, R. Billington, S. M. Shahidehpour and C. Singh, "Bibliography on the Application of Probability Methods in Power System Reliability Evaluation: 1982 – 1987," *IEEE Transactions on Power Systems*, Nov. 1988, pp. 1555 – 1564.

[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] 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.

[15] B. P. Lang and A. Pahwa, "Power Distribution System Reliability Planning Using a Fuzzy Knowledge-Based Approach," *IEEE Transactions on Power Delivery*, Vol. 15, No. 1, Jan 2000, pp. 279 –2 84.

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

[17] 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.

[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] R. L. Chen, K. Allen and R. Billington, "Value-Based Distribution Reliability Assessment and Planning," *IEEE Transactions on Power Delivery*, Jan. 1995, pp. 421 – 429.

[22] J. Momoh, "Value-Based Distribution System Reliability Analysis," *IEEE International Conference on Computational Cybernetics and Simulation*, 1997, pp. 3452 – 3457.

[23] R. E. Brown, S. Gupta, R. D. Christie, S. S. Venkata, and R. Fletcher, " Automated Primary Distribution System Design: Reliability and Cost Optimization," *IEEE Transactions on Power Delivery,* Apr. 1997, pp. 1017 – 1022.

[24] A. Chowdhury and D. O. Koval, "Current Practices and Customer Value-Based Distribution System Reliability Planning," *IEEE Industry Applications Conference*, 2000, pp. 909 – 916.

[25] V. Longo and W. R. Puntel, "Evaluation of Distribution System Enhancement Using Value-Based Reliability Planning Procedures," *IEEE Transactions on Power Systems,* Aug. 2000, pp. 1148 – 1153.

[26] J. H. Teng and C. N. Lu, "Feeder-Switch Relocation for Customer Interruption Cost Minimization," *IEEE Transactions on Power Delivery,* Jan. 2002, pp.254 – 259.

[27] R. Billinton and G. Wacker, " Customer Cost of Electric Service Interruptions," *Proceedings of the IEEE*, June 1989, pp. 919 – 930.

[28] R. Billinton, S. A. Ali and G. Wacker, " Reliability Worth Comparisons", *IEEE Power Engineering Review,* May 2001, pp. 3 – 5.

[29] R. Billinton, S. A. Ali and G. Wacker, "Determination and Use of Sector Composite Customer Damage Functions", *IEEE Canadian Conference on Electrical and Computer Engineering*, 1999, pp. 1483 – 1488.

[30] 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.

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

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

[33] Y. Y. Hsu, L. M. Chen, J. L. Chen et al, "Application of a Microcomputer-Based Database Management System to Distribution System Reliability Evaluation," *IEEE Transactions on Power Delivery,* Jan. 1990, pp. 343 – 350.

[34] W. F. Horton, S. Goldberg and R. A. Hartwell, "A Cost/Benefit Analysis in Feeder Reliability Studies," *IEEE Transactions on Power Delivery,* Jan. 1989, pp. 446 – 452.

[35] J. S. Lawler, J. S. Lai, L. D. Monteen, J. B. Patton, and D. T. Rizy, "Impact of Automation on the Reliability of the Athens Utilities Board's Distribution System," *IEEE Transactions on Power Delivery,* Jan. 1989, pp. 770 – 778.

[36] 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.

[37] S. Sahai and A. Pahwa, "Failures of Overhead Distribution System Lines Due to Animals," *Proceedings of the 36th Annual North American Power Symposium,* University of Idaho, Moscow, Idaho, Aug. 2004.

[38] S. Gupta, A. Pahwa and R. E. Brown, "Predicting the Failure Rates of Overhead Distribution Lines Using an Adaptive Fuzzy Technique," *Proceedings of the 33rd North*

American Power Symposium, Texas A&M University, College Station, USA, Oct. 2001, pp. 193 – 197.

[39] S. R. Gilligan, “A Method for Estimating the Reliability of Distribution Circuits,” *IEEE Transactions on Power Delivery*, Vol. 7, Issue 2, Apr. 1992, pp. 694-698.

[40] G. Kjolle and K. Sand, “RELRAD - An Analytical Approach for Distribution System Reliability Assessment,” *IEEE Transactions on Power Delivery*, Vol. 7, issue 2, Apr. 1992, pp. 809 – 814.

[41] R. E. Brown, S. Gupta, S. S. Venkata, R. D. Christie and R. Fletcher, “Distribution System Reliability Assessment Using Hierarchical Markov Modeling”, *IEEE Transactions on Power Delivery*, Vol. 11, issue 4, Oct. 1996, pp. 1929-1934.

[42] C. A. Volkmann, “A Probabilistic Approach to Distribution System Reliability Assessment,” *Third International Conference on Probabilistic Methods Applied to Electric Power Systems*, Jul. 1991, pp.169 – 173.

[43] N. Balijepalli, S. S. Venkata and R. D. Christie, “Modeling and Analysis of Distribution Reliability Indices,” *IEEE Transactions on Power Delivery*, Vol.19, Issue 4, Oct. 2004, pp.1950 – 1955.

[44] R. Billington and R. Goel, “An Analytical Approach to Evaluate Probability Distribution Associated with the Reliability Indices of Electrical Distribution Systems,” *IEEE Transactions on Power Delivery*, July 1986, pp. 245-251.

[45] 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.

[46] R. Billinton and P. Weng, “Teaching Distribution System Reliability Evaluation Using Monte Carlo Simulation,” *IEEE Transactions on Power Systems*, May 1999, pp. 397-403.

[47] R. E. Brown and T. M. Taylor, “Modeling the Impact of Substations on Distribution Reliability,” *IEEE Transactions on Power Systems*, Feb. 1999, pp. 349 – 354.

[48] R. E. Brown and J. R. Ochoa, “Impact of Subcycle Transfer Switches on Distribution System Reliability, *IEEE Transactions on Power Systems,* Feb. 2000, pp. 442 – 447.

[49] P. A. Kuntz, R. D. Christie and S. S. Venkata, “A Reliability Centered Optimal Visual Inspection Model for Distribution Feeders,” *IEEE Transactions on Power Delivery*, Oct. 2001, pp. 718 – 723.

[50] R. Billington and W. Li, *Reliability Assessment of Electric Power Systems Using Monte Carlo Methods*, Plenum Press, New York, 1994.

[51] R. N. Allan and M. G. DaSilva, "Evaluation of Reliability Indices and Outage Cost for Distribution Systems," *IEEE Transactions on Power Systems,* Feb. 1995, pp. 413 – 419.

[52] S. Asgarpoor and M. J. Mathine, "Reliability Evaluation of Distribution Systems with Nonexponential Down Times," *IEEE Transactions on Power Systems,* May 1997, pp. 579 – 584.

[53] R. Billington and P. Wang, "Distribution System Reliability Cost/Worth Analysis Using Analytical and Sequential Simulation Technique," *IEEE Transactions on Power Systems,* Nov. 1998, pp. 1245 – 1250.

[54] J. B. Bowles, "Commentry-caution: Constant Failure-rate Models May be Hazardous to your Design," *IEEE Transactions on Reliability*, Sep. 2002, pp. 375 – 377.

[55] R. E. Brown, G. Frimpong, and H.L. Willis, "Failure Rate Modeling Using Equipment Inspection Data," *IEEE Transactions on Power Systems*, May 2004, pp. 782 – 787.

[56] J. Burke, "Using Outage Data to Improve Reliability," *IEEE Computer Applications in Power*, Apr. 2000, pp. 57 – 60.

[57] P. Wang and R. Billington, "Reliability Cost/Worth Assessment of Distribution Systems Incorporating Time-Varying Weather Conditions and Restoration Resources," *IEEE Transactions on Power Delivery*, Jan. 2002, pp. 260 – 265.

[58] B. Retterath, S. S. Venkata, and A. A. Chowdhury, "Impact of Time-Varying Rates on Distribution Reliability," *Proceedings of the 8th International Conference on Probabilistic Methods Applied to Power Systems,* Ames Iowa, Sep. 2004.

[59] M.-Y. Chow, S. O. Yee and L.S. Taylor, "Recognizing Animal-Caused Faults in Power Distribution Systems Using Artificial Neural Networks," *IEEE Transactions on Power Delivery,* July 1993, pp. 1268 – 1274.

[60] D. T. Radmer, P. A. Kuntz, R. D. Christie, S. S. Venkata, and R. H. Fletcher, "Predicting Vegetation-Related Failure Rates for Overhead Distribution Feeders," *IEEE Transactions on Power Delivery,* Oct. 2002, pp. 1170 – 1175.

[61] P. A. Kuntz, R. D. Christie and S. S. Venkata, "Optimal Vegetation Maintenance Scheduling of Overhead Electric Power Distribution Systems," *IEEE Transactions on Power Delivery,* Oct. 2002, pp. 1164 – 1169.

[62] R. E. Brown, S. Gupta, R.D. Christie, S. S. Venkata, and R. Fletcher, "Distribution System Reliability Assessment: Momentary Interruptions and Storms," *IEEE Transactions on Power Delivery,* Oct. 1997, pp. 1569 – 1575.

[63] IEEE Standard 1410 – 1997, IEEE Guide for Improving the Lightning Performance of Electric Power Overhead Distribution Lines, IEEE Press, Piscataway, NJ, 1997.

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

[65] C. W. Williams, "Weather Normalization of Power System Reliability Indices," *Proceedings of IEEE PES T&D Conference and Expo*, Dallas, TX, Sept. 2003.

[66] 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.

[67] L. Xu, M.Y. Chow, and L. S. Taylor, "Analysis of Tree-Caused Faults in Power Distribution Systems," *to be presented at the 35th North American Power Symposium*, University of Missouri – Rolla, Rolla, MO, Oct. 2003.

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

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

[70] C. A. Warren, R. Ammon and G. Welch, "A Nationwide Survey of Distribution Reliability Measurement Practices," *IEEE Transactions on Power Delivery*, Jan. 1999, pp. 250 – 257.

[71] R. E. Brown, "Identifying the Worst Performing Feeders," *Proceedings of the 8th International Conference on Probabilistic Methods Applied to Power Systems*, Ames Iowa, Sept. 2004.

[72] 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.

[73] R. D. Christie, "Statistical Classification of Major Event Days in Distribution System Reliability," *IEEE Transactions on Power Delivery*, Oct. 2003, pp. 1336 – 1341.

[74] 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.

[75] R. Harness, "Animal-caused Outages," EDM International, Inc. 2007.

[76] S. Sahai and A. Pahwa, "Prediction of Animal-caused Outages in Overhead Distribution System ," Thesis, May 2005.

[77] J. A. Chapman and G. A. Feldhamer, "Fox and Gray Squirrels," *Wild Mammals of North America: Biology, Management and Economics*, Johns Hopkins University Press, 1982, pp. 209 – 229.

[78] M. Gui, A. Pahwa, and S. Das, "Prediction of Animal Related Failures in Overhead Distribution Systems Using Neural Networks and Wavelet Transform Techniques", *Proceeding of the 37th Annual North American Power Symposium*, Carbondale, Illinois, Sep. 2006.

[79] http://en.wikipedia.org/wiki/Poisson_distribution as retrieved on Mar. 20, 2009 at 12:23:45 GMT.

[80] J. H. Ahrens and U. Dieter, "Computer Generation of Poisson Deviates," *ACM Transactions on Mathematical Software*, Vol. 8, Issue 2, 1982, pp. 163 – 179.

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

[82] P. McCullagh, and J. A. Nelder, Generalized Linear Models, 2 Ed, Chapman and Hall, New York, 1989.

[83] R. H. Loh, "Time Series Forecast With Neural Network and Wavelet Techniques," Thesis, University of Queensland, St Lucia, Oct. 2003.

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

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

[86] Y. F. Low and Z. Y. Zheng, "Enhancing Neural Network Electricity Load Forecast with Wavelet Techniques," *Proceeding of the 4th Asian Control Conference*, Singapore, Sep. 2002, pp. 203 – 208.

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

[88] D. J. C. MacKay, "Bayesian non-linear modeling for the 1993 energy prediction competition", *Maximum Entropy and Bayesian Methods*, Santa Barbara 1993, G. Heidbreder (Ed.), Dordrecht: Kluwer, 1995.

[89] N. Satio and G. Beylkin "Multiresolution representations using the auto-correlation functions of compactly supported wavelets", *IEEE Transactions on Signal Processing*, 1992.

[90] M. Azoff, "Neural network time series forecasting of financial markets", John Wiley & Sons, 1994.

[91] M. Gui, A. Luo, "Short Term Load Forecast Using Wavelet Neutral Network", *Electricity*, Vol.18, No.16, Aug. 2004, pp. 11 – 15.

[92] G. Beylkin and N. Satio. "Wavelets, their autocorrelation functions and multi-resolution representation of signals", *IEEE Transactions on Signal Processing*, Vol. 7, 1997, pp. 147 – 164.

[93] A. Graps "A introduce to wavelets," *IEEE Transactions on Computational Science and Engineering*, Vol. 2, No. 2, Summer 1995, pp. 50 – 61.

[94] D. B. Percival and A. T. Walden. *Wavelet Methods for Time Series Analysis*, Cambridge University Press, Cambridge, United Kingdom, 2000, pp. 56 – 158.

[95] G. Strang and T. Nguyen, Wavelets and Filter Banks, Cambridge Press, 1997, pp. 189 – 193.

[96] G. Beylkin and N. Satio. "Wavelets, their autocorrelation functions and multi-resolution representation of signals", *IEEE Transactions on Signal Processing*, Vol. 7, 1997, pp. 147 – 164.

[97] P. Yu, A. Goldberg, and Z. Bi, "Time series forecasting using wavelets with predictor-corrector boundary treatment," in *Proceedings of 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, Aug. 2001.

[98] L. N. D Castro and F. J. V. Zuben, "The clonal selection algorithm with engineering applications," in *Proceedings of GECCO*, Las Vegas, NV, Jul. 2000, pp. 36 – 37.

[99] L. N. D. Castro and J. Timmis, Artificial Immune Systems: A New Computational Intelligence Approach, Springer, New York:Heidelberg, NY, 2001.

[100] L. N. D. Castro and Fernando J. V. Zuben, “Learning and Optimization Using the Clonal Selection Principle*,”* *IEEE Transactions on Evolutionary Computation*, *Special Issue on Artificial Immune Systems*, Vol. 6, 2002, pp. 239 – 251.

[101] Thomas Bayes, "An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S.", *Philosophical Transactions on on Giving Some Account of the Present Undertakings, Studies and Labours of the Ingenious in Many Considerable Parts of the World,* Vol. 53, 1763, pp.370 – 418.

[102] S. Z. Zhang, N. H. Yang and X. K. Wang, “Construction and Application of Bayesian Networks in Flood Decision Supporting System,” *Proceedings of the First International Conference on Machine Learning and Cybernetics,* Nov. 2002.

[103] D. Heckerman and J. S. Breese, “ Causal Independence for Probability Assessment and Inference using Bayesian Networks,” *IEEE Transactions on Systems, Man and Cybernetics*, Part A, Vol. 26, Issue 6, Nov. 1996, pp. 826 – 831.

[104] S. Z. Zhang, H. Yu, H. Ding, N. H. Yang and X. K. Wang, “An Application of Online Learning Algorithm for Bayesian Network Parameter”, *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics*, Vol.1, Nov. 2003, pp. 153–156.

[105] K. Murphy, “A Brief Introduction to Graphical Models and Bayesian Networks,” 1998, http://www.cs.ubc.ca/~murphyk/Bayes/bayes.html as retrieved on Mar. 21, 2009 at 2:30:40 GMT.

[106] N. Tropolis and S. Ulam, “The Monte Carlo Method,” *Journal of the American Statistical Association*, Vol. 44, No. 247, 1949, pp.335 – 341.

[107] R. Y. Rubinstein and D. P. Kroese, Simulation and the Monte Carlo Method (2nd ed.). New York: John Wiley & Sons, 2007.

[108] R. Billinton and W. Li, Reliability Assessment of Electric Power Systems Using Monte Carlo Methods, New York: Plenum, 1994.

[109] R. Billinton and X. Tang, “Selected Considerations in Utilizing Monte Carlo Simulation in Quantitative Reliability Evaluation of Composite Power Systems,” *Electric Power System Research*, Vol. 69, Issue 2-3 , May 2004, pp. 205 – 211.

[110] 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.

Appendix A - Additional Results of Simulations

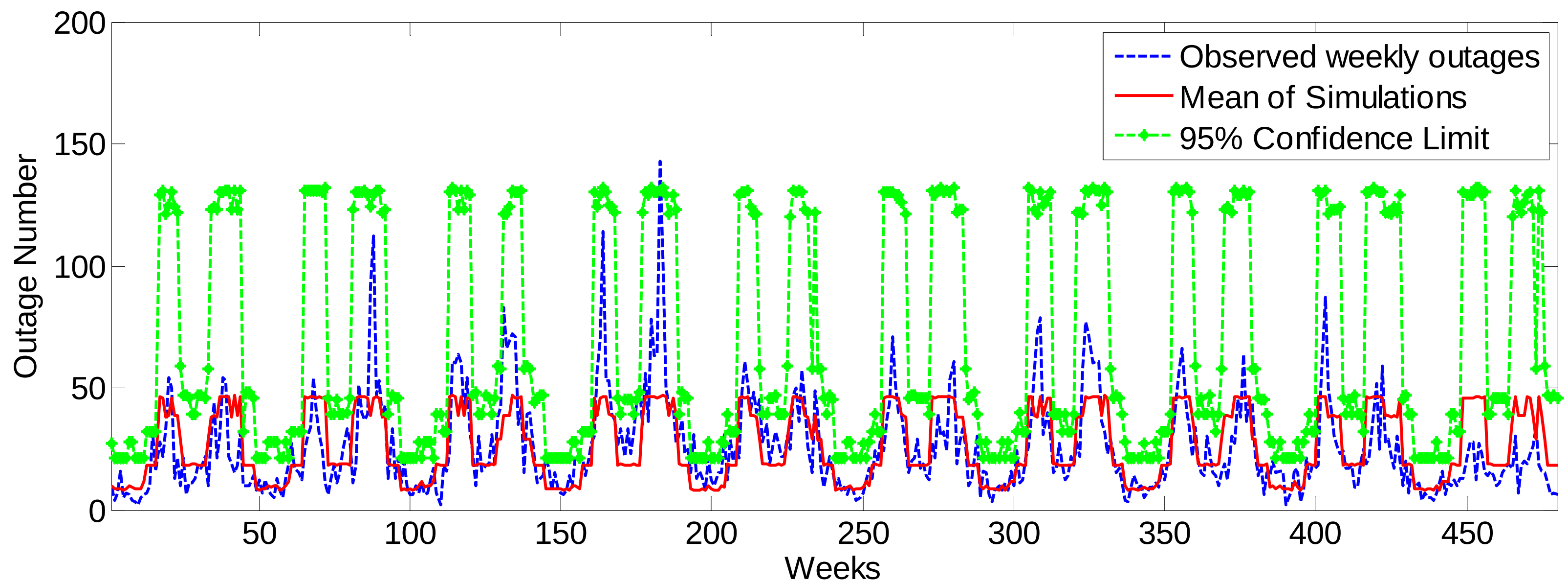

Figure A.1 Weekly Prediction and 95% Confidence Limit for Wichita by MCS CPT9

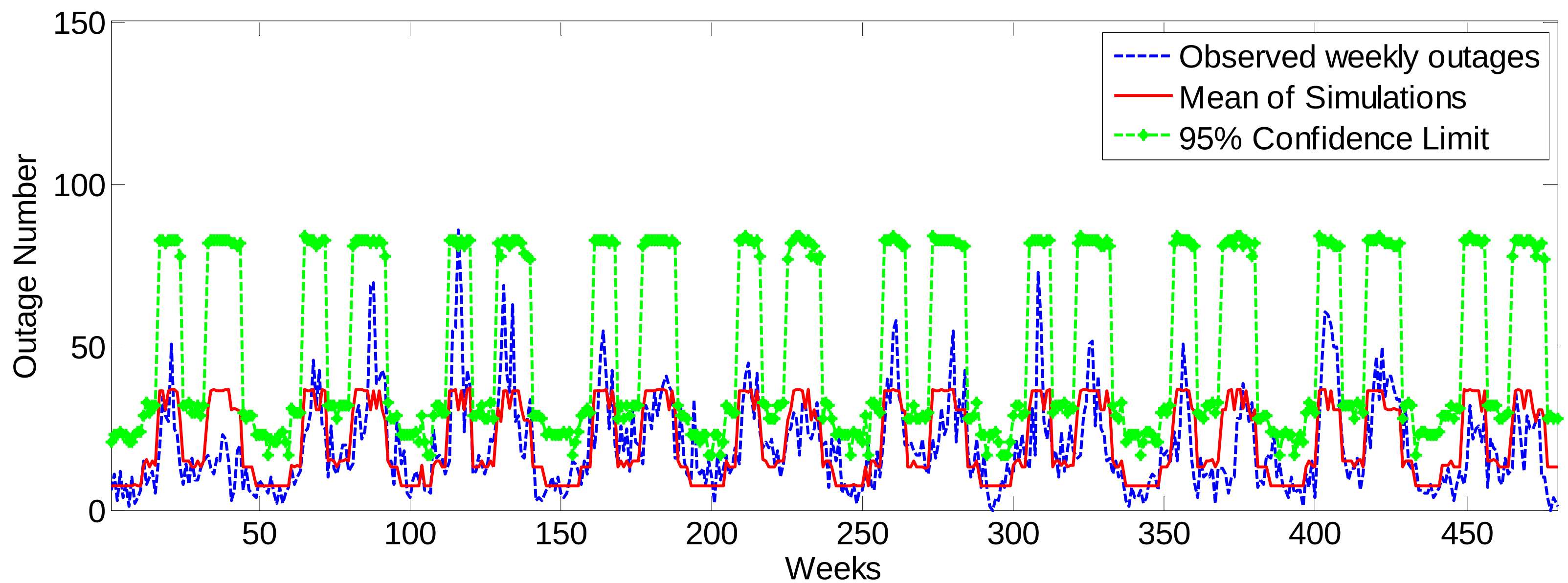

Figure A.2 Weekly Prediction and 95% Confidence Limit for Topeka by MCS CPT9

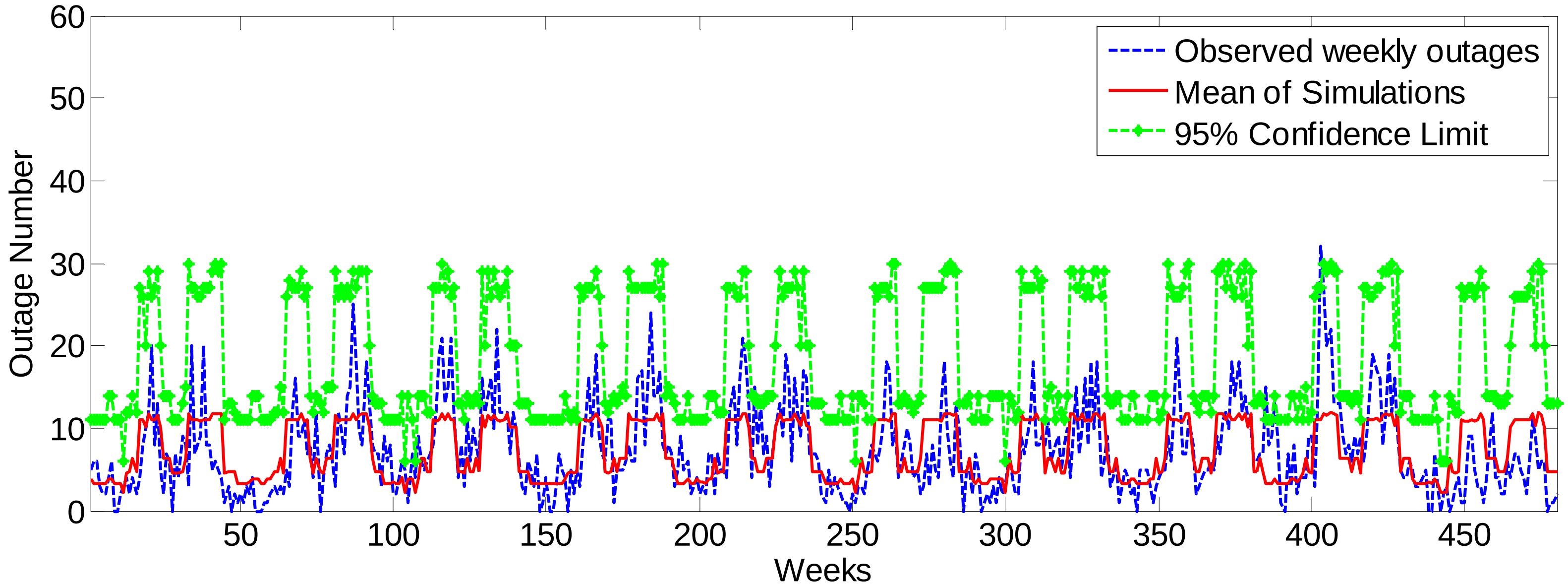

Figure A.3 Weekly Prediction and 95% Confidence Limit for Lawrence by MCS CPT9

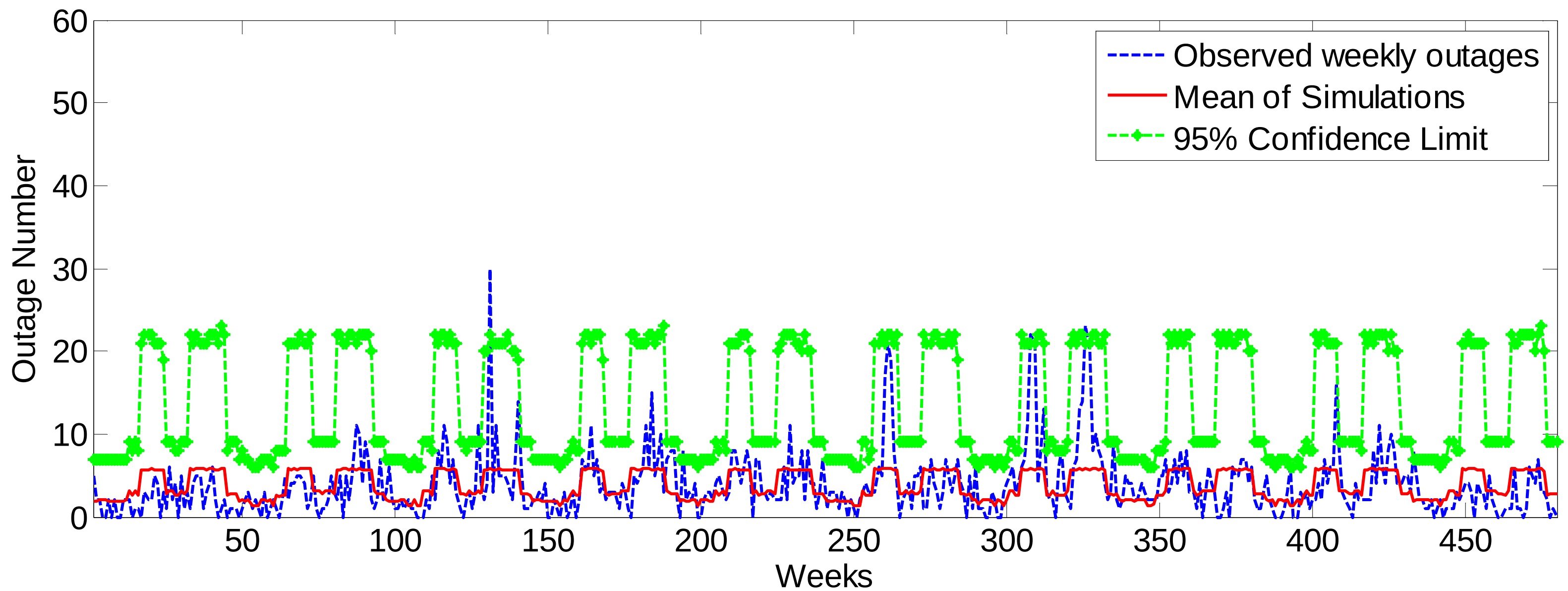

Figure A.4 Weekly Prediction and 95% Confidence Limit for Manhattan by MCS CPT9

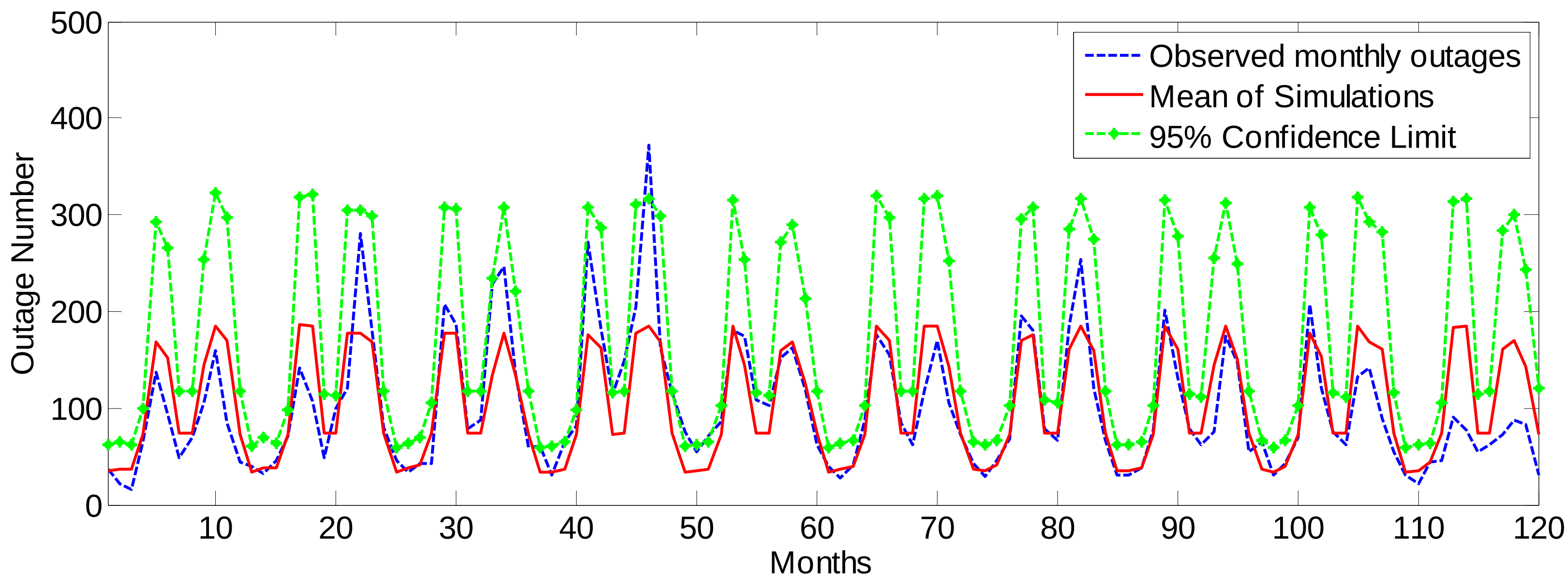

Figure A.5 Monthly Prediction and 95% Confidence Limit for Wichita by MCS CPT9

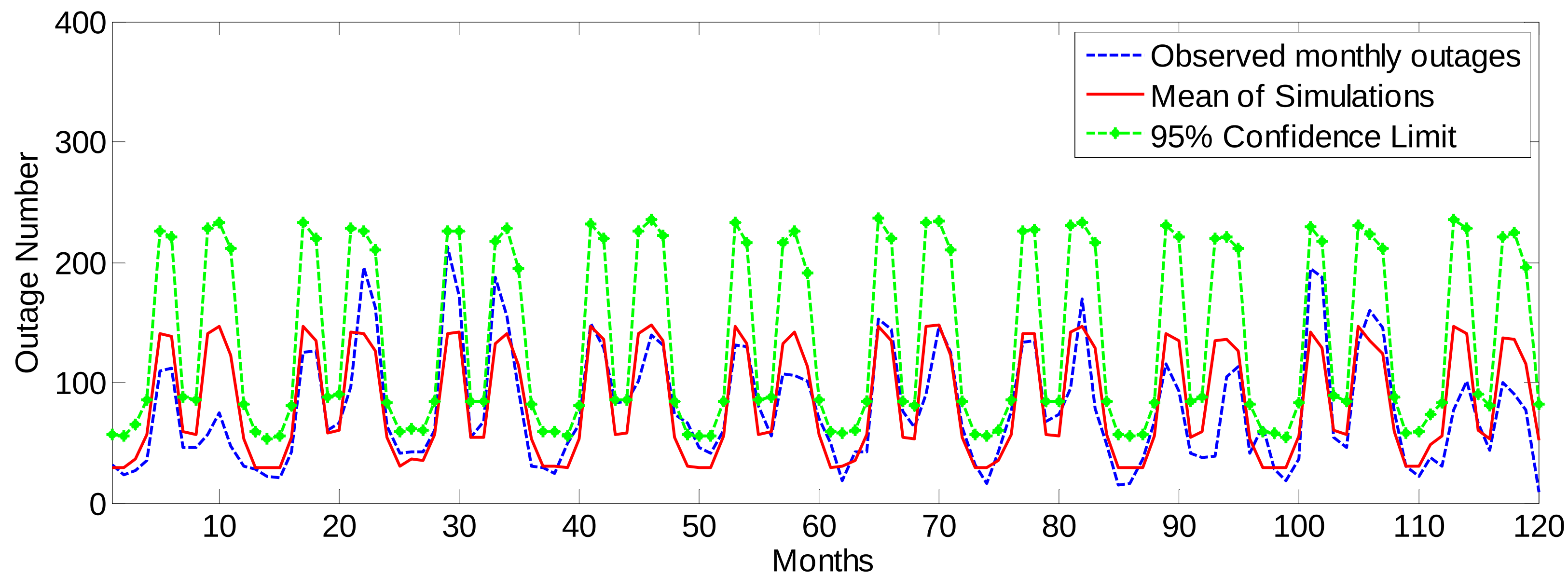

Figure A.6 Monthly Prediction and 95% Confidence Limit for Topeka by MCS CPT9

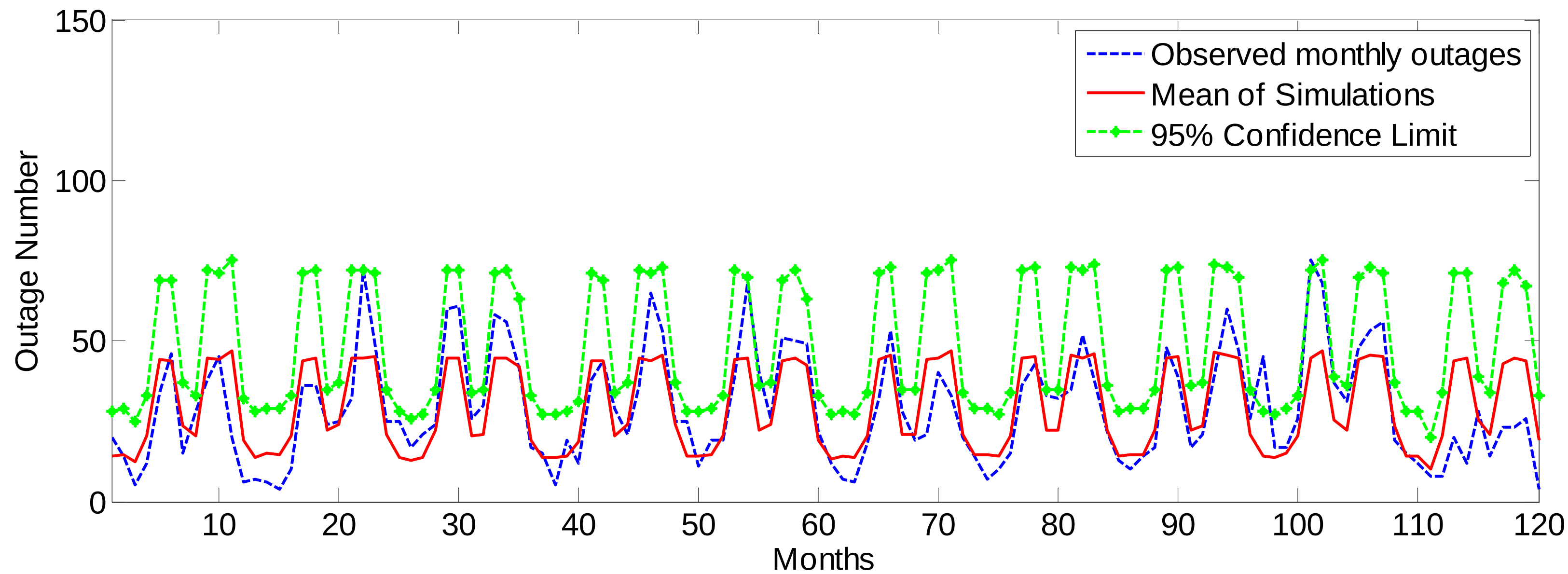

Figure A.7 Monthly Prediction and 95% Confidence Limit for Lawrence by MCS CPT9

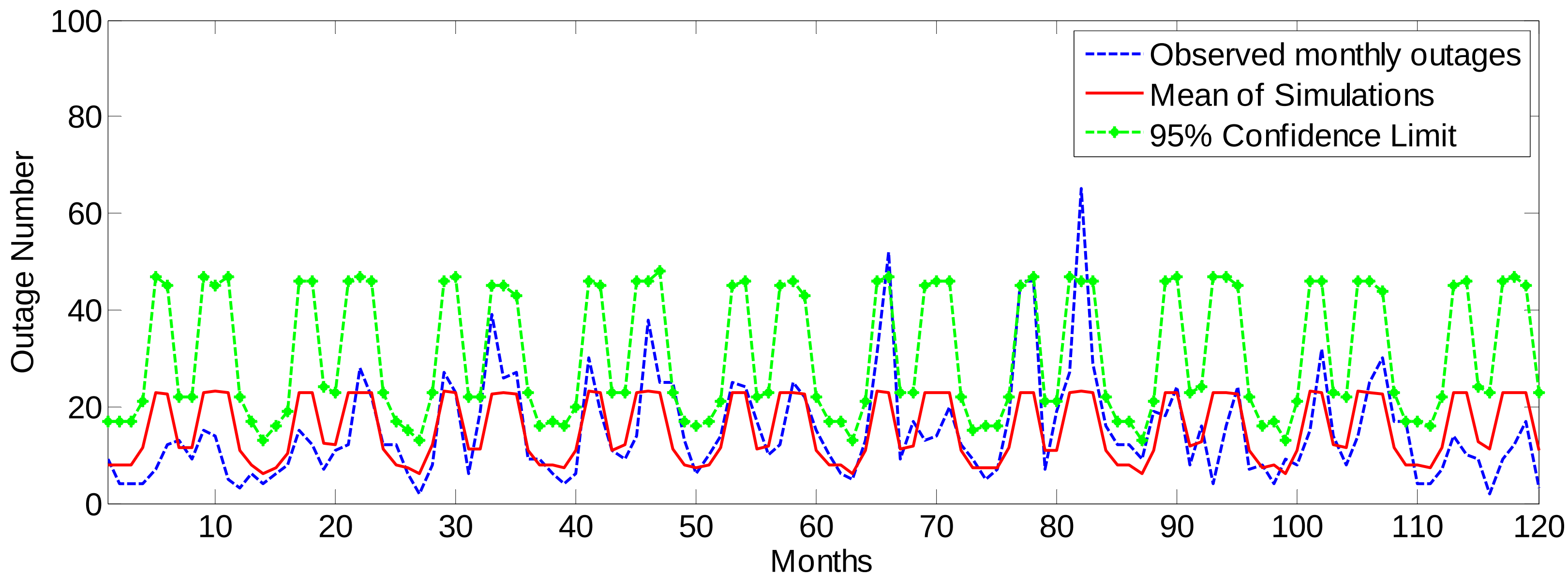

Figure A.8 Monthly Prediction and 95% Confidence Limit for Manhattan by MCS CPT9

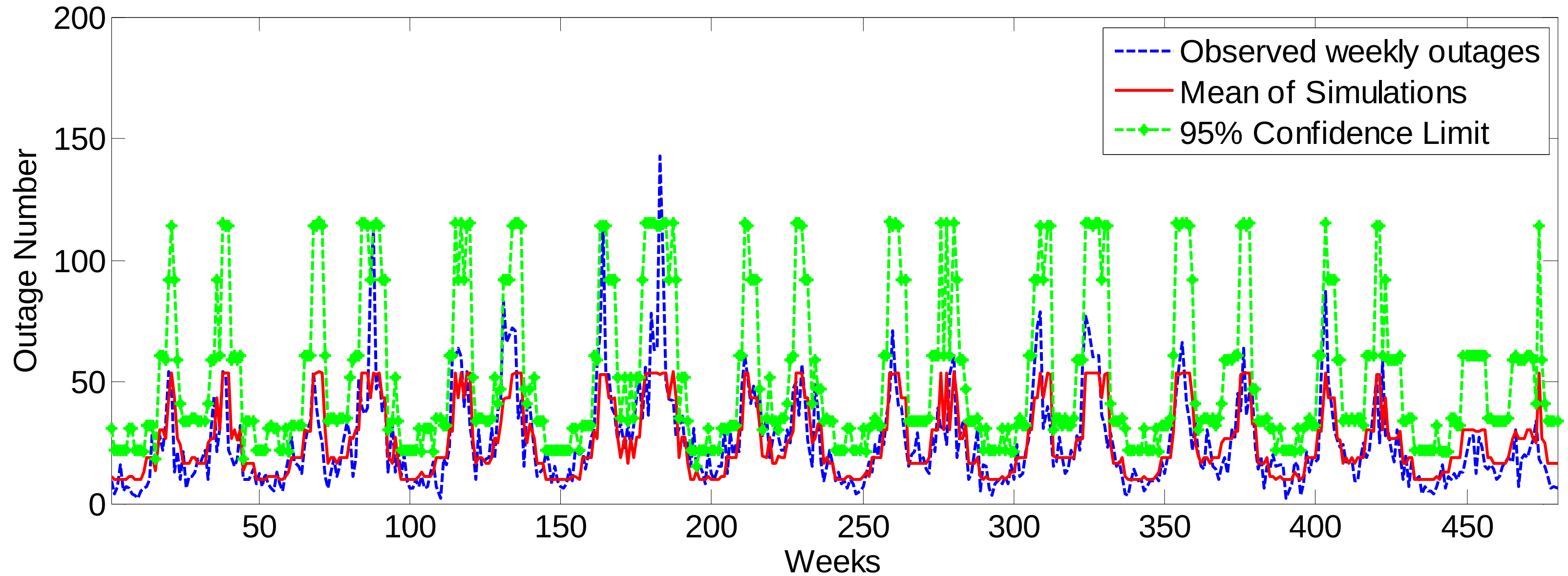

Figure A.9 Weekly Prediction and 95% Confidence Limit for Wichita by MCS H18

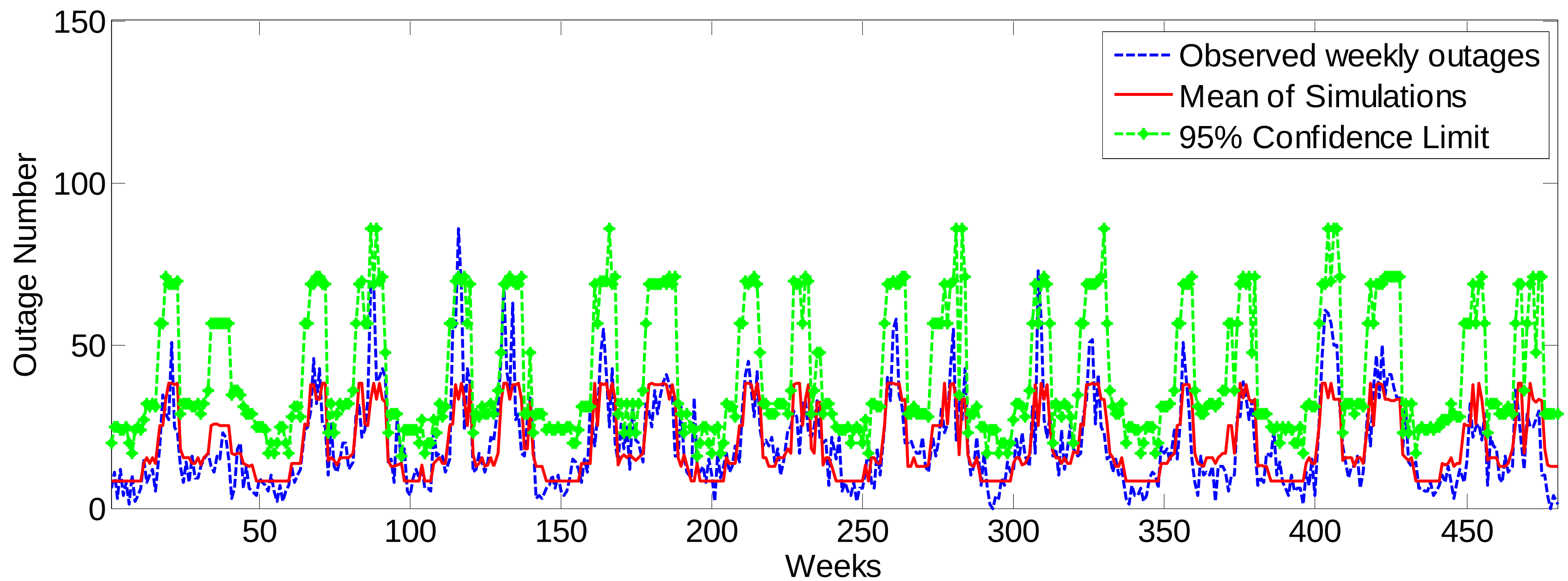

Figure A.10 Weekly Prediction and 95% Confidence Limit for Topeka by MCS H18

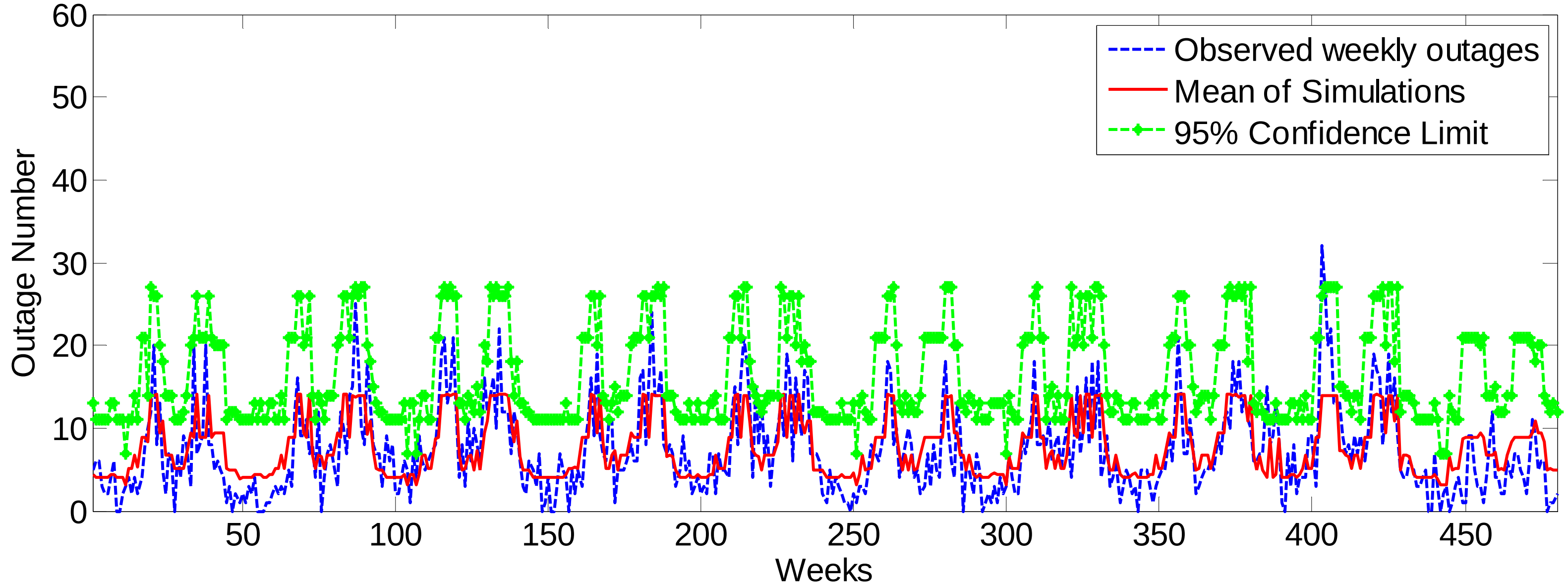

Figure A.11 Weekly Prediction and 95% Confidence Limit for Lawrence by MCS H18

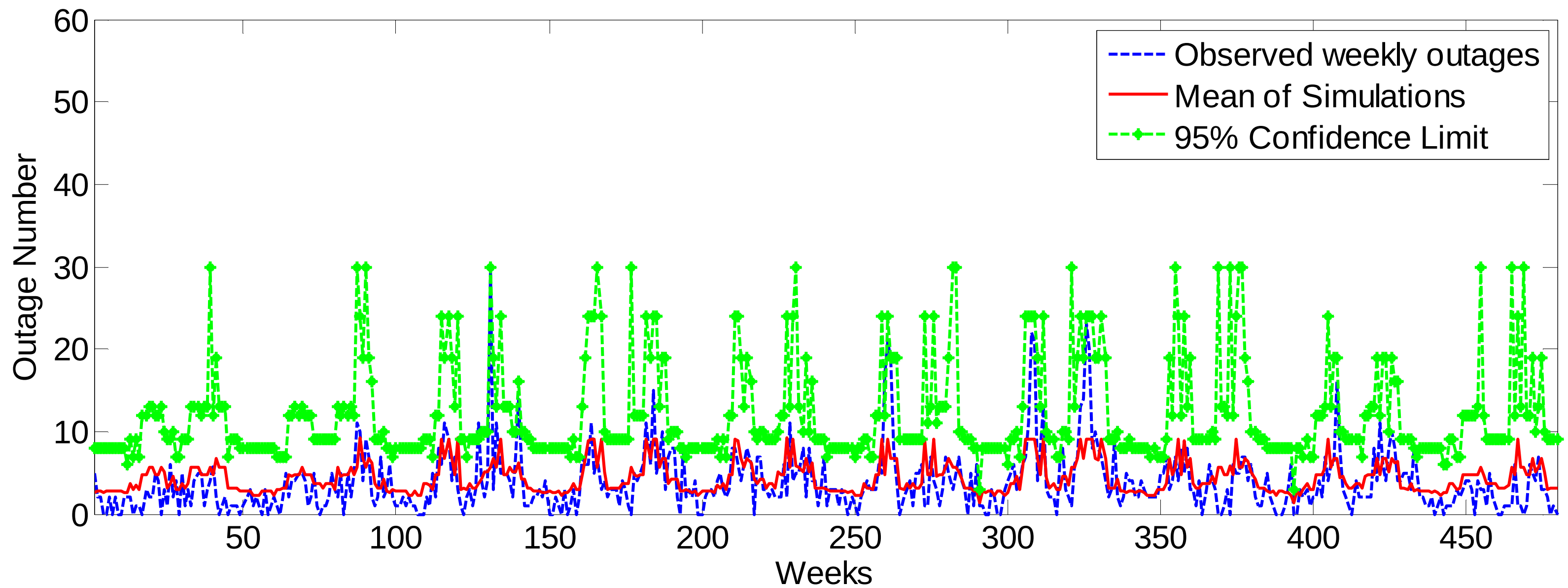

Figure A.12 Weekly Prediction and 95% Confidence Limit for Manhattan by MCS H18

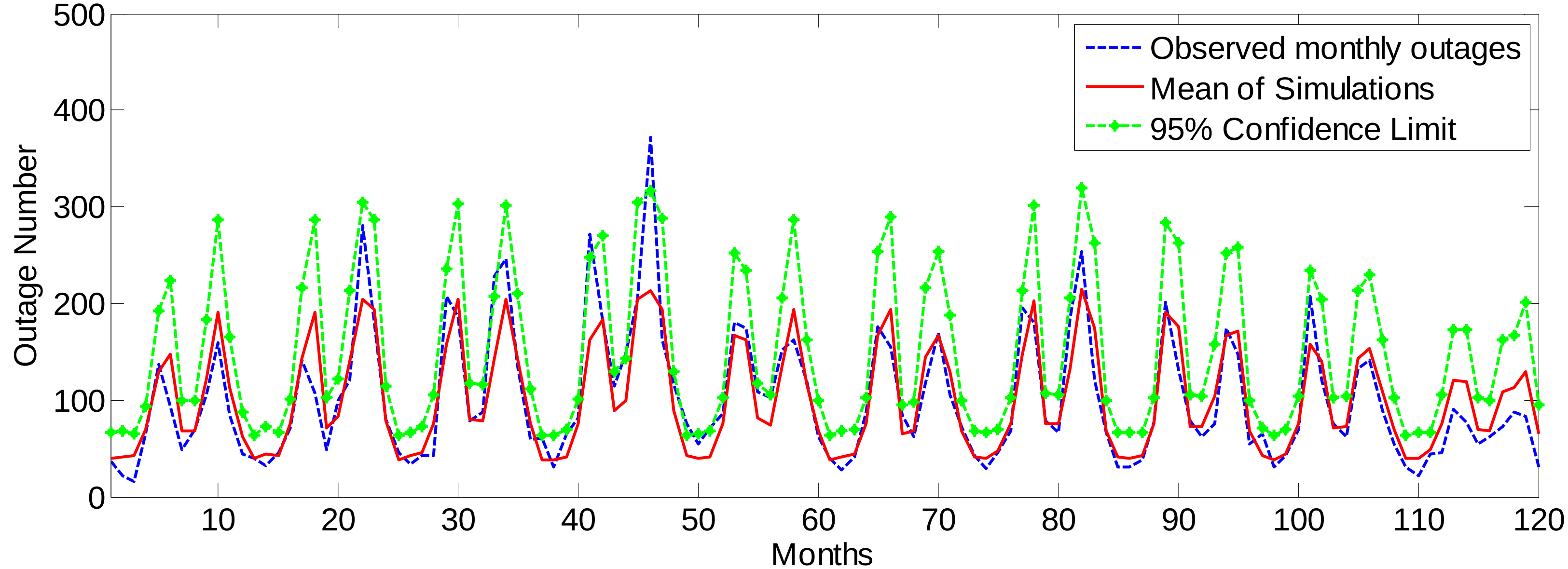

Figure A.13 Monthly Prediction and 95% Confidence Limit for Wichita by MCS H18

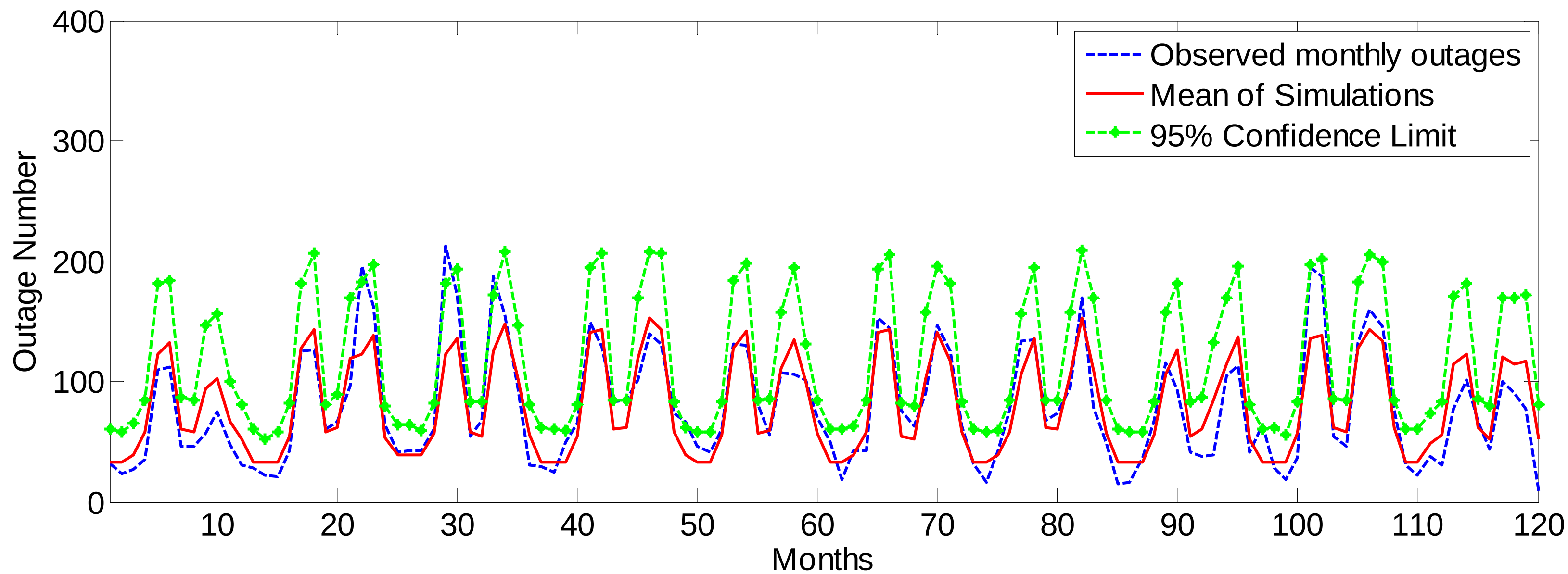

Figure A.14 Monthly Prediction and 95% Confidence Limit for Topeka by MCS H18

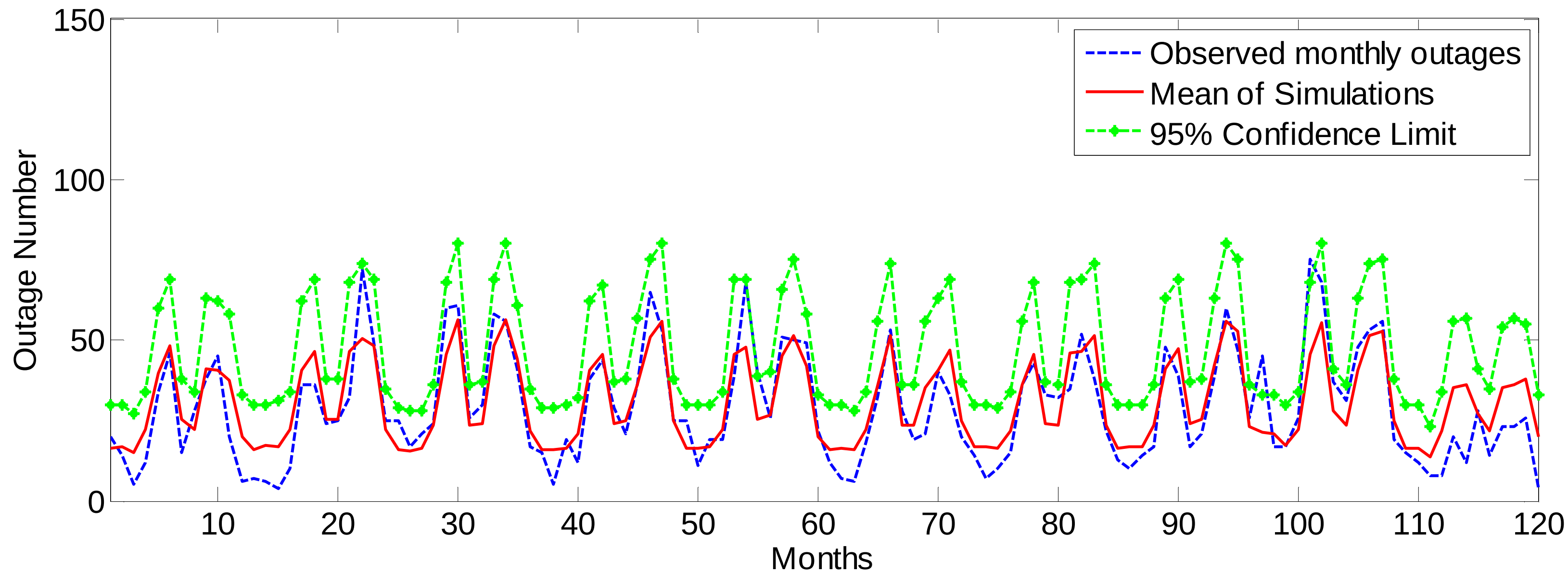

Figure A.15 Monthly Prediction and 95% Confidence Limit for Lawrence by MCS H18

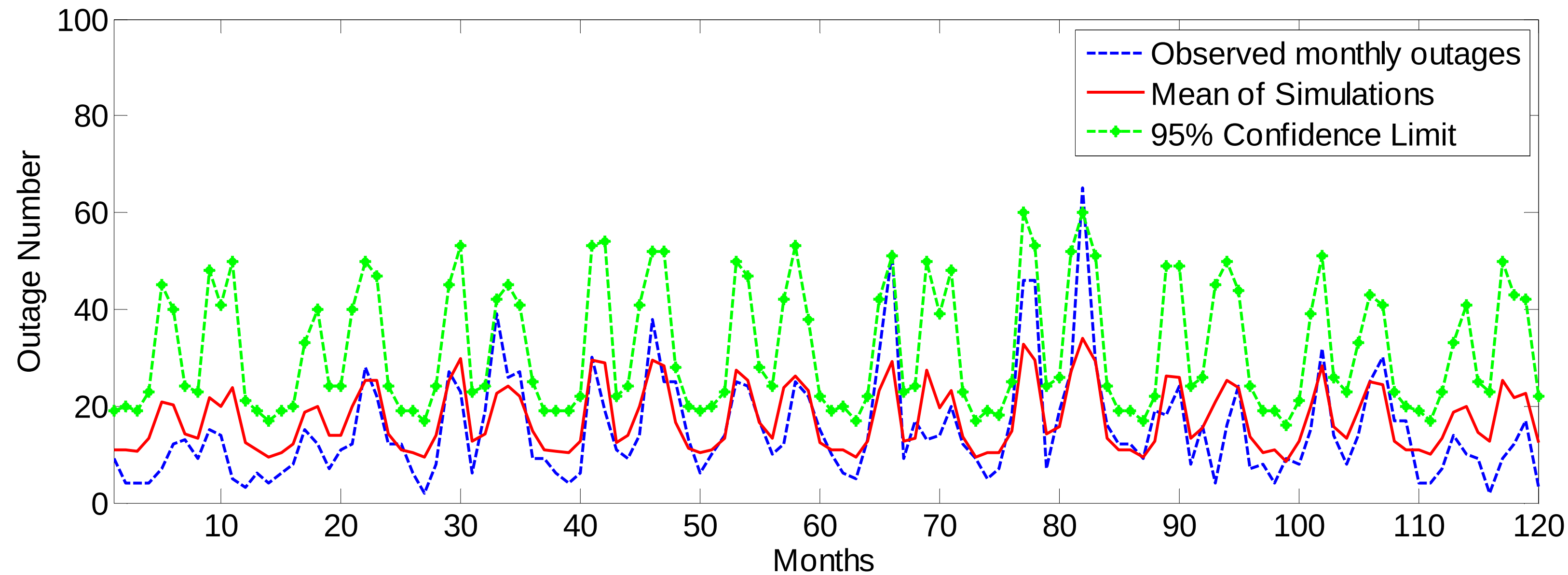

Figure A.16 Monthly Prediction and 95% Confidence Limit for Manhattan by MCS H18

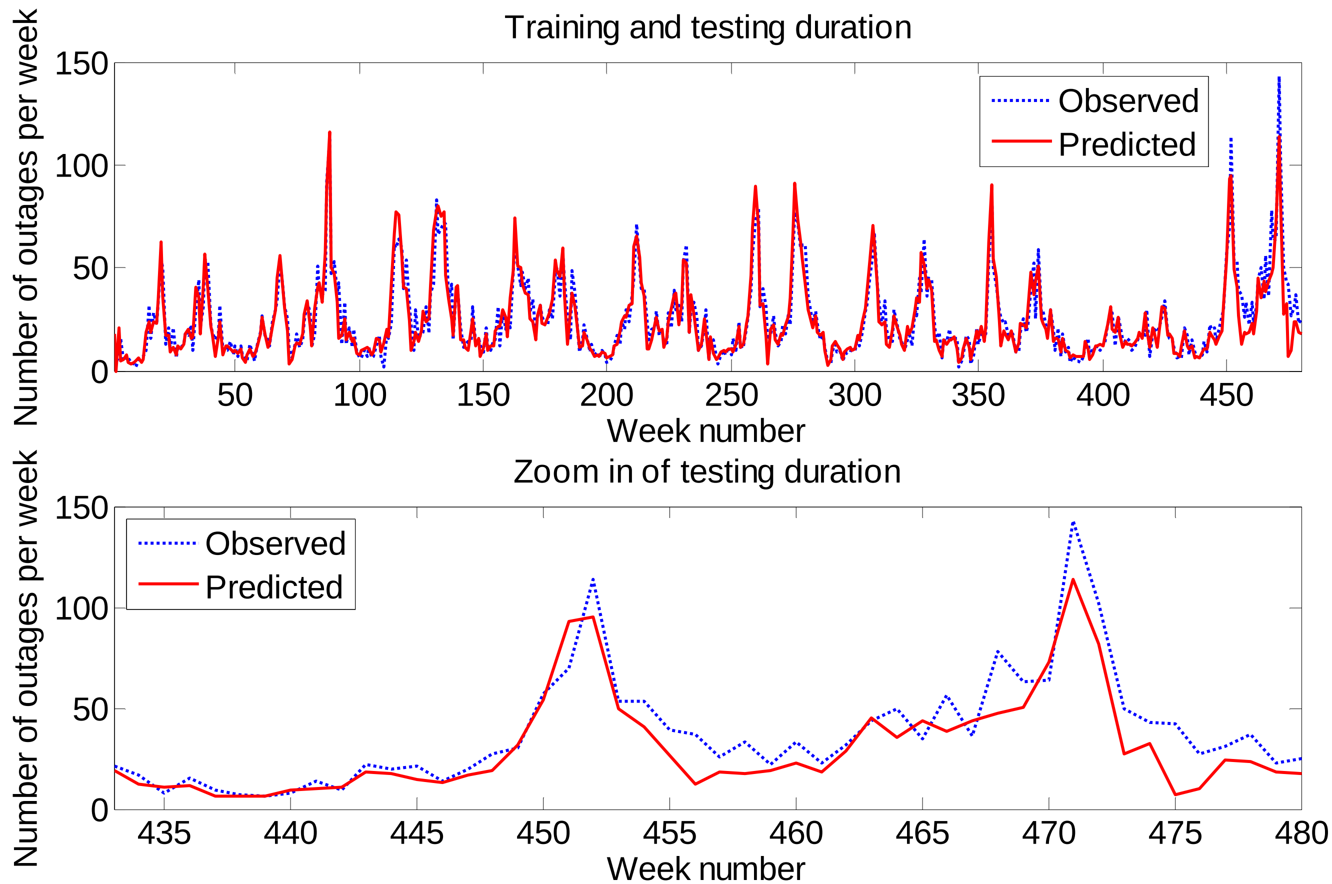

Figure A.17 Weekly Prediction by MWNN Model with 2001 as Testing Data for Wichita

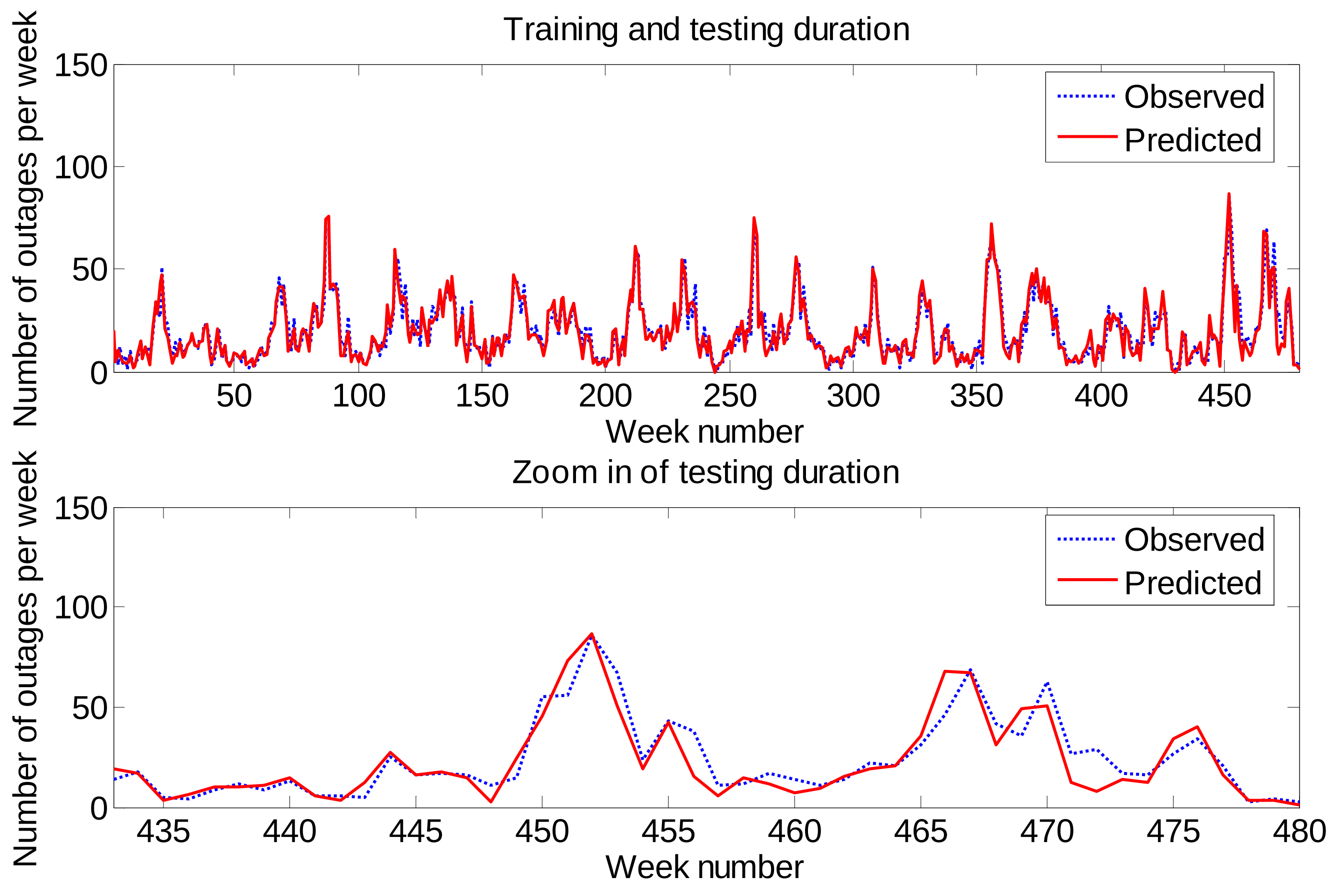

Figure A.18 Weekly Prediction by MWNN Model with 2000 as Testing Data for Topeka

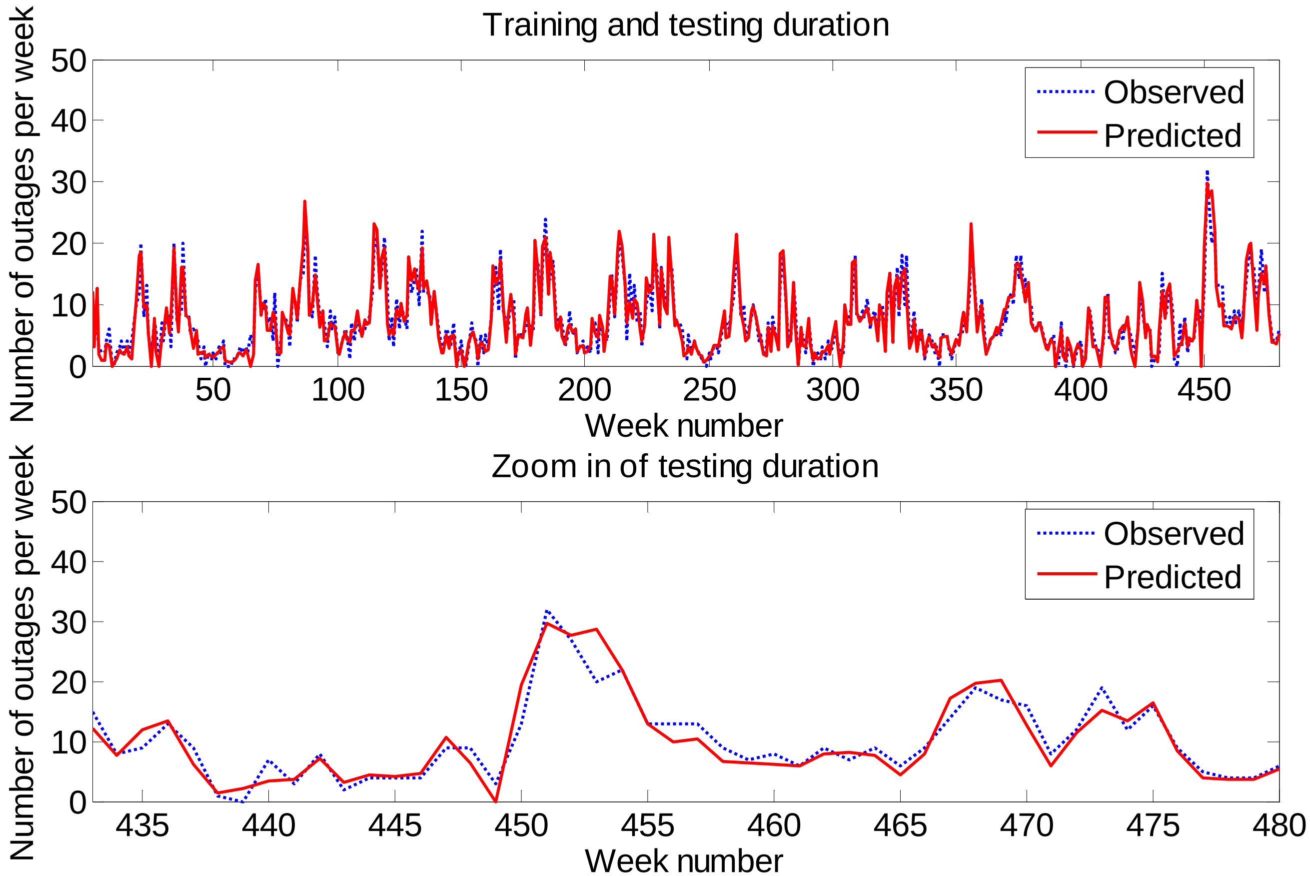

Figure A.19 Weekly Prediction by MWNN Model with 2006 as Testing Data for Lawrence

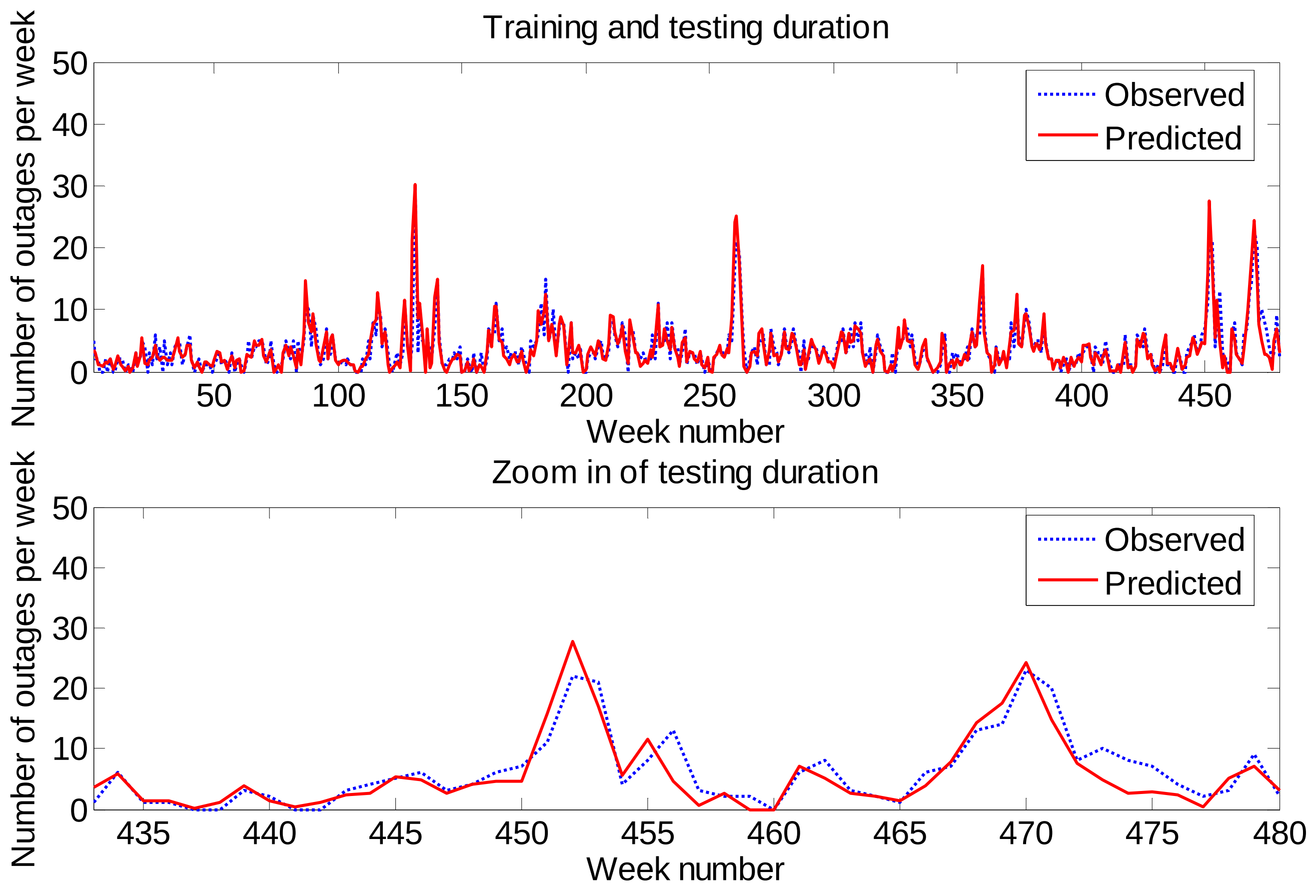

Figure A.20 Weekly Prediction by MWNN Model with 2004 as Testing Data for Manhattan