

This is the author's final, peer-reviewed manuscript as accepted for publication. The publisher-formatted version may be available through the publisher's web site or your institution's library.

Bayesian network model with Monte Carlo simulations for analysis of animal-related outages in overhead distribution systems.

Min Gui, Anil Pahwa, and Sanjoy Das

### How to cite this manuscript

If you make reference to this version of the manuscript, use the following information:

Gui, M., Pahwa, A., & Das, S. (2011). Bayesian network model with Monte Carlo simulations for analysis of animal-related outages in overhead distribution systems. Retrieved from <http://krex.ksu.edu>

### Published Version Information

**Citation:** Gui, M., Pahwa, A., & Das, S. (2011). Bayesian network model with Monte Carlo simulations for analysis of animal-related outages in overhead distribution systems. IEEE Transactions on Power Systems, 26(3), 1618-1624.

**Copyright:** Copyright © 2011 IEEE

**Digital Object Identifier (DOI):** doi: 10.1109/TPWRS.2010.2101619

**Publisher's Link:** [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5708195&tag=1](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5708195&tag=1)

This item was retrieved from the K-State Research Exchange (K-REx), the institutional repository of Kansas State University. K-REx is available at <http://krex.ksu.edu>

# Bayesian Network Model with Monte Carlo Simulations for Analysis of Animal-Related Outages in Overhead Distribution Systems

Min Gui, *Member, IEEE*, Anil Pahwa, *Fellow, IEEE* and Sanjoy Das

**Abstract**—This paper extends previous research on using a Bayesian network model to investigate impacts of time (month) and weather (number of fair weather days in a week) on animal-related outages in distribution systems. Outage history (outages in the previous week) is included as an additional input to the model, and inputs and outputs are classified systematically to reduce errors in estimates of outputs. Conditional probability table obtained from the historical data are used to estimate weekly animal-related outages which is followed by a Monte Carlo simulation to find estimates of mean and confidence limits for monthly animal-related outages. Comparison of results obtained for four cities of different sizes in Kansas with those obtained using a hybrid wavelet/neural network model shows consistency between the two models. The methodology presented in this paper is simple to implement and useful for the utilities for year-end analysis of the outage data to identify specific reliability related concerns.

**Index Terms**—Animal-related failures, Bayesian network, Monte Carlo Simulation, Power distribution systems, Power system reliability.

## I. INTRODUCTION

Although animals cause significant number of outages in overhead distribution systems [1-7], the exact causal relationship between them has not been addressed adequately in literature. Practical techniques for mitigating animal-caused outages have been presented [1, 2]. Chow et al have focused on identifying and classifying animal-caused outages [3, 4]. Models for estimating outages caused by animals with time of the year and weather conditions as inputs have been proposed previously by the authors of this paper [5-6]. A discrete Bayesian network model with two inputs to study these effects was presented in [5]. Division of inputs and outputs into different discrete levels was based on an ad-hoc approach and outages were assumed to follow Poisson distribution for estimating the statistical upper bound of outages in the specified time duration. A hybrid model with wavelet

decomposition and neural networks for estimation of animal-related was presented in [6]. This study illustrates that incorporating outage history in the model results in significantly enhanced performance.

Simplicity of applying Bayesian network models make them very attractive for representing effects of animals on outages in distribution systems. Therefore, additional research was conducted to refine the model presented in [5]. A systematic approach was used to classify inputs and outputs into different discrete levels for the Bayesian network model to reduce errors in estimates of outputs. Details of this approach are available in [7].

The main focus of this paper is to apply the modified Bayesian network model to study animal-related outages over a period of ten years in four cities in Kansas and to compare the results with those obtained with wavelet/neural network hybrid model [6]. A Monte Carlo simulation is implemented to estimate the monthly outages and to determine their upper and lower confidence bounds. The results of this research are consistent with those of the published hybrid model, highlighting the effectiveness of the proposed method.

The four cities included in this study are Manhattan (7 distribution substations with 176 miles of distribution feeders at 12.47 kV), Lawrence (7 distribution substations with 193 miles of distribution feeders at 12.47 kV), Topeka (22 distribution substations with 560 miles of distribution feeders mostly at 12.47 kV and a very small portion at 4 kV), and Wichita (42 distribution substations with 1165 miles of distribution feeders mostly at 12.47 kV and a very small portion at 4 kV). Although the study covers a rather protracted period of ten years, prior discussions with utility engineers revealed that the grid topology changed little during this interval. Thus we have assumed that the grid structure remained the same throughout our analysis.

The methodology presented in this paper is simple to implement and is a useful tool for utilities in their year-end analysis of outage data, to identify specific reliability related concerns. Comparison of the observed outages with the estimated upper limit gives an indication of the reliability of the distribution system over the specified period. Observed outages exceeding the upper limit may warrant corrective actions to be taken by the utility. The results can also be used by utilities to justify higher than usual outages, as long as they are below the upper limit, in their reports to the state utility commissions.

This work was supported in part by the National Science Foundation under Grant ECS-0501288.

Min Gui (e-mail: csugm@hotmail.com) graduated with the Ph.D. degree in Electrical Engineering at Kansas State University and is with Luminant Energy, Dallas, TX 75201.

Anil Pahwa (e-mail: pahwa@ksu.edu) is Professor and Sanjoy Das (e-mail: sdas@ksu.edu) is Associate Professor in Electrical and Computer Engineering department at Kansas State University, Manhattan, KS 66506.

## II. BAYESIAN NETWORK MODEL

A Bayesian network is a probabilistic graphical model that represents a set of random variables and their conditional interdependencies by means of a directed acyclic graph [8-12]. The nodes of this graph are the random variables. A directed edge from one node (parent) to another (child) indicates a direct causal relationship between the corresponding random variables. The probabilistic nature of the child node's dependence on its parents, is quantified by a conditional probability table present at that node [10].

A two-layer Bayesian network with *Time*, *Weather*, and *Outage History* as inputs and *Outages* in the week as output is shown in Fig. 1.

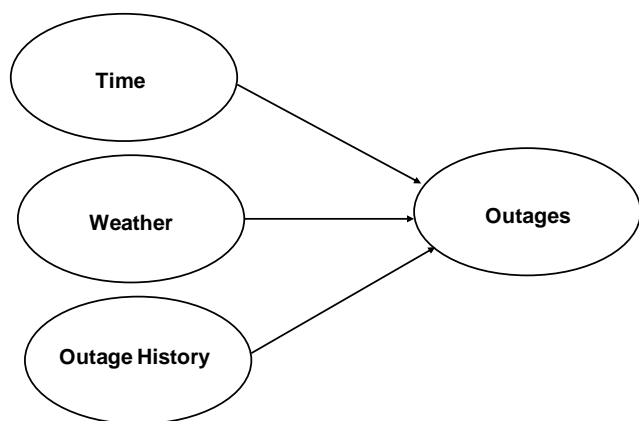


Fig. 1. A two-layer Bayesian network with three inputs for estimation of weekly animal-related outages

### A. Model Variables

#### 1) Time

Time, defined by the month of the year, is classified into three discrete levels similar to that in [6], which are *Low* (January, February and March), *Medium* (April, July, August, and December), and *High* (May, June, September, October, November). This grouping is based on the expected level of animal activity.

#### 2) Weather

Since animals are more active during fair weather (temperature between 40 and 85 F and no other weather activity), weather for a week is classified into three levels based on the number fair weather days in the week. These three levels are *Low* (0 fair weather days), *Medium* (1 to 3 fair weather days), and *High* (4 or higher fair weather days) in the week representing low, medium, and high probability of outages based on animal activity. With all possible combinations of time and weather, there are totally 9 input states for the Bayesian network model with these two inputs.

#### 3) Outage History

Given the same month and same weather conditions, the outages vary in a certain range due to the probabilistic nature of the outages. In the previous study based on wavelet/neural network hybrid model [6], it was found that using outages in weeks prior to the current week as additional inputs improves

the model performance. To capture this feature, previous week's outage level is used as the as the third input in the Bayesian network model. Dividing the previous week's outages into two levels (*High* and *Low*) as well as three levels (*High*, *Medium* and *Low*) were investigated. It was found that two levels are better suited for modeling [7], as it improves the model performance while preventing the conditional probability table from becoming needlessly large. The marginal improvement obtained from three levels is more than offset by a significant computational overload. With two levels for previous week's outages there are a total of 18 possible combinations of inputs for the model that are henceforth referred to as 'states' in this paper. The cutoff for *High* outage level in the previous week is set at 70th percentile, which means weeks with outages higher than those occurring in 70% of the 480 weeks are defined as *High*.

#### 4) Outage Level

Histograms (number of weeks with outages in the given range) of weekly animal-caused outages in the four cities considered for the study for the past ten years or a total of 480 weeks are shown in Fig. 2. Analysis based on different levels for outages [7] showed that classification with nine outage levels is the most suitable for all cities. It was observed that the average absolute error in the Bayesian network model's outage estimates decreased as the number of discrete outage levels was increased from one to nine. However, no further improvement could be obtained beyond this point. Therefore, a total of nine discrete outage levels have been used uniformly in the present study.

Due to the differences in sizes, the disparity in the outages occurring in each city was high, even under similar input conditions. Unfortunately, early attempts at normalization based on size of the city as well as length of feeders did not yield satisfactory results. Therefore, in this paper, the outage levels were discretized separately for each city, such that each outage level contained roughly the same number of outages. These ranges are shown, separately for each city, in Table I.

TABLE I  
OUTAGE LEVEL RANGES FOR EACH CITY

Outage Level	Wichita	Topeka	Lawrence	Manhattan
1	0 to 9	0 to 7	0 to 3	0
2	10 to 13	8 to 10	4 to 5	1
3	14 to 17	11 to 13	6	2
4	18 to 21	14 to 16	7 to 8	3
5	22 to 25	17 to 19	9 to 10	4 to 6
6	26 to 32	20 to 24	11 to 12	7 to 8
7	33 to 42	25 to 29	13 to 15	9 to 13
8	43 to 63	30 to 36	16 to 20	14 to 24
9	64 to 143	37 to 86	21 to 33	25 to 30

## III. MODEL IMPLEMENTATION

### A. Conditional Probability Table (CPT)

The historical data is used to learn the parameters of the model, which are the entries in the conditional probability table, i.e. the conditional probability of each outage level given the *Time* (month type), *Weather* (the level of fair weather days),

and *Outage History* (the level of outages in the previous week), that is,

$$P(O_L = i \mid M_T = j, FWD_L = k, PWO_L = l)$$

where  $O_L$  is the outage level,  $M_T$  is the month type,  $FWD_L$  is the level of fair weather days, and  $PWO_L$  is the level of outages in the previous week, and

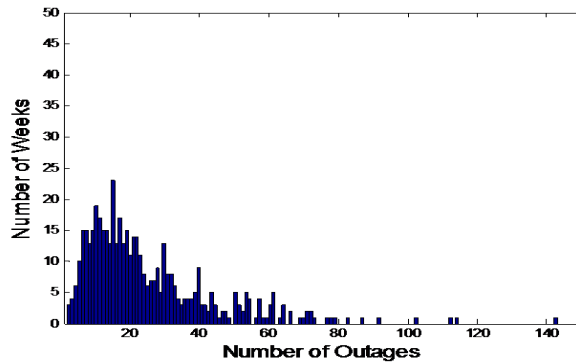
$$i = 1, \dots, 9; j = 1, 2, 3; k = 1, 2, 3; l = \text{Low or High.}$$

The 18 input states represent all possible combinations of three input variables. The number of weeks in the historical data belonging to each state are shown separately for each city in Table II. Some input states such as those numbered 13 and 16 have nearly all zeros or mostly zero entries, which implies that no or very few weeks matched conditions of these states. These two states represent conditions where the month type is 1 (or low animal activity), previous week outage level is high, and the fair weather day level is medium and high, respectively. Since animal activity is low in these months, even higher level of fair weather days is unlikely to produce many outages. Therefore, these combinations are very unlikely to occur in real life, which explains the lack of sufficient number weeks in these states. Similarly, states 10 and 17, which have fewer weeks, are very unlikely to occur. Therefore, even though we have limited data for these states, their impact on determination of expected number of outages in a time period would be minimal.

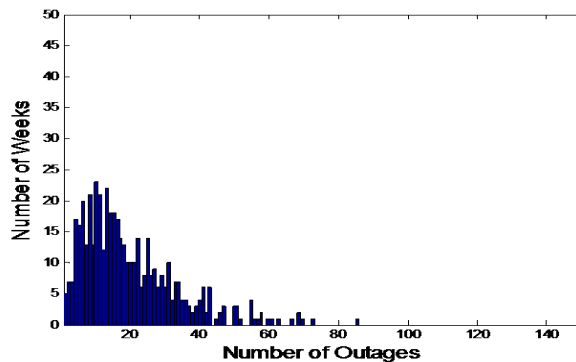
TABLE II  
NUMBER OF WEEKS PER STATE FOR FOUR CITIES

Input State	$M_T$	$FWD_L$	$PWO_L$	Number of Weeks			
				WCT*	TPK*	LRC*	MHN*
1	1	1	Low	73	71	72	88
2	2	1	Low	58	48	44	56
3	3	1	Low	11	7	6	14
4	1	2	Low	34	39	38	22
5	2	2	Low	46	56	55	53
6	3	2	Low	30	20	32	45
7	1	3	Low	12	8	8	4
8	2	3	Low	35	39	40	30
9	3	3	Low	47	49	64	65
10	1	1	High	1	2	2	4
11	2	1	High	13	7	12	13
12	3	1	High	9	8	12	7
13	1	2	High	0	0	0	2
14	2	2	High	6	8	9	4
15	3	2	High	34	42	35	34
16	1	3	High	0	0	0	0
17	2	3	High	2	2	0	4
18	3	3	High	69	74	51	35

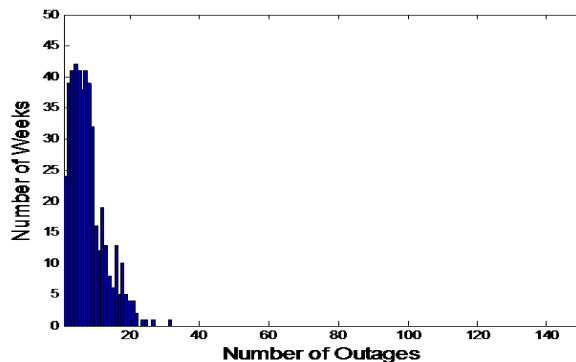
\*WCT: Wichita, TPK: Topeka, LRC: Lawrence, MHN: Manhattan



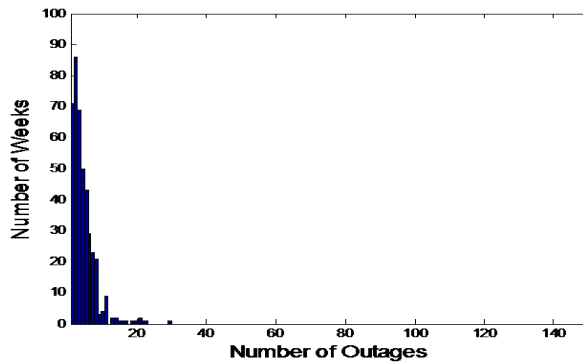
(a) Wichita



(b) Topeka



(c) Lawrence



(d) Manhattan

Fig. 2. Histograms of weekly animal-related outages in different cities from 1998 to 2007

Since the graph structure is fully known with fully observed historical data, Maximum Likelihood Estimation was used to learn values in the CPT. Hence, the equation to compute the conditional probabilities for input state  $m$  is:

$$P(O_L = i \mid \text{Input State} = m) = N_i / N_m$$

Where  $N_i$  is the number of occurrences in outage level  $i$  in state  $m$  and  $N_m$  is the total number of occurrences in state  $m$ .

### B. Estimation of Animal-related Outages

In order to get the expected number of outages for a given week with a given state, weighted sum of average value or median value of outages in each level weighted by conditional probability, has to be obtained. In the previous work, median values were used [3], but we have chosen the average values as they better represent the historical outage data. The median values are based on range of outages in each outage level, but the average values take account of the distribution of outages within the outage levels and thus can provide better characterization of the outage levels.

The expected number of animal-caused outages in each input state can thus be computed by the following equation :

$$E(\text{Number of outages} \mid \text{Input state} = m) = \sum_{i=1}^9 P(O_L = i \mid \text{Input State} = m) \text{Avg}(O_L = i)$$

for  $m = 1, \dots, 18$ .

Where  $P(O_L = i \mid \text{Input State} = m)$  is the conditional probability of occurrence of outage level  $i$ , given input state  $m$  and  $\text{Avg}(O_L = i)$  is the average value of outages in the outage level  $i$ .

The expected number of outages at each input state is computed and listed in Table III. This value is considered as the estimate of outages for the weeks with this state. Estimating outages over a larger time period, such as a month, can be readily obtained by summing all the weekly estimated values for that month. However, since no prior probability distribution of the outages for each state is assumed, it is not possible to compute the variance and confidence limits directly for a meaningful comparison of computed values with observed values of outages. We attempted to fit different probability distributions to the outage data, but that did not provide consistent results. Therefore, Monte Carlo simulation as detailed in the next section was used to obtain the variance and the confidence limits.

## IV. MONTE CARLO SIMULATION

### A. Probability Distribution Functions

To implement Monte Carlo simulation, we have to determine the probability distribution function (pdf) of outages in each input state. Therefore, the entries in the CPT for each state are normalized by the size of the bin related to each outage level to obtain pdf for each state. A sample pdf thus obtained is shown in Fig. 3. It corresponds to Input State 18 for Wichita. Also shown in this figure are the normalized probabilities for specific outage values. The pdf gives the correct trend as state 18 is expected to have the highest number of outages. As this

graph shows, probability of outages is zero for very low values, it is low for very high values, and it is high for in between values.

TABLE III  
THE EXPECTED NUMBER OF OUTAGES IN EACH STATE FOR FOUR CITIES

Input State	$M_T$	$FWD_L$	$PWO_L$	Expected Number of Outages			
				WCT	TPK	LRC	MHN
1	1	1	Low	9.57	7.95	3.40	1.93
2	2	1	Low	16.35	12.91	4.51	2.43
3	3	1	Low	25.24	16.92	8.04	4.91
4	1	2	Low	10.73	8.00	4.02	1.49
5	2	2	Low	18.73	15.42	6.29	3.02
6	3	2	Low	26.72	16.67	9.22	4.81
7	1	3	Low	12.22	13.48	2.46	1.96
8	2	3	Low	18.39	13.37	4.75	2.32
9	3	3	Low	29.76	25.08	8.42	4.20
10	1	1	High	11.39	13.50	8.41	2.46
11	2	1	High	27.87	16.36	6.16	3.76
12	3	1	High	33.31	33.32	11.10	6.74
13	1	2	High	0.00	0.00	0.00	0.50
14	2	2	High	19.35	13.79	6.92	4.33
15	3	2	High	44.60	33.36	13.35	6.74
16	1	3	High	0.00	0.00	0.00	0.00
17	2	3	High	13.45	16.44	0.00	4.12
18	3	3	High	51.66	37.58	13.72	8.21

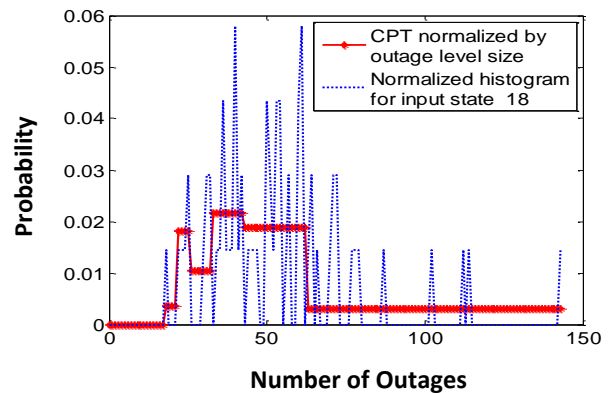


Fig. 3. The normalized histogram and CPT for Input State 18 for Wichita

### B. Simulation Procedure

Stepwise implementation of Monte Carlo simulations is provided below:

1. Find the input state for a given week.
2. Generate a uniform random number.
3. Using roulette wheel with this random number select an outage level based on the pdf for that state.
4. Generate another uniform random number.
5. Using roulette wheel with this random number select a value for outage from the selected outage level. The outages follow uniform distribution within one outage level.
6. Repeat the simulation 10000 times for each week.

Since the simulation is repeated 10000 times, we get 10000 simulated sample points for each week. By simply adding up the sample points of four weeks in the same month in an iteration, we get 10000 sample points for monthly outages. The mean, variance, and the corresponding 95% confidence limit for monthly outages are then computed from the 10000 samples. Although the same approach can be used to determine yearly outages, examination of the results showed that many details are lost if yearly aggregation is considered. On the other hand, weekly observations showed too much noise and fluctuations.

V. RESULTS

The Monte Carlo simulation methodology presented in the previous sections was applied to all the four cities of this study to estimate outages and the associated 95% upper limit for every month of the 10 year duration. Fig. 4 and 5 show examples of histogram for selected months obtained from the Monte Carlo simulation with the Bayesian network model for Wichita and Manhattan. In both cases, the results closely resemble Gaussian distribution. Similar results were obtained for all the cities for all the months. From these plots estimated monthly mean and 95% percent limits can be easily computed. Fig. 6 gives the monthly observed outages, estimated outages and the associated 95% upper limit for Wichita for all the months computed with the Monte Carlo simulations. Fig. 7 gives similar results obtained with wavelet/neural network

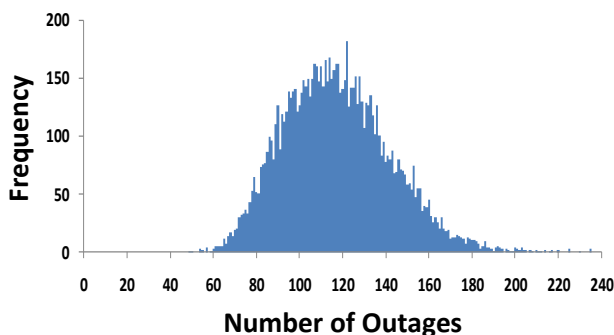


Fig. 4. Histogram of outages in May 2007 in Wichita based on Monte Carlo Simulation

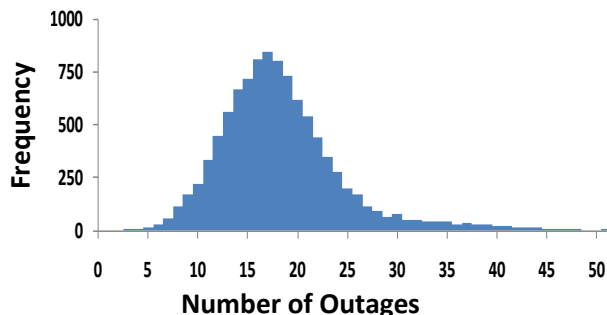


Fig. 5. Histogram of outages in May 2007 in Manhattan based on Monte Carlo Simulation

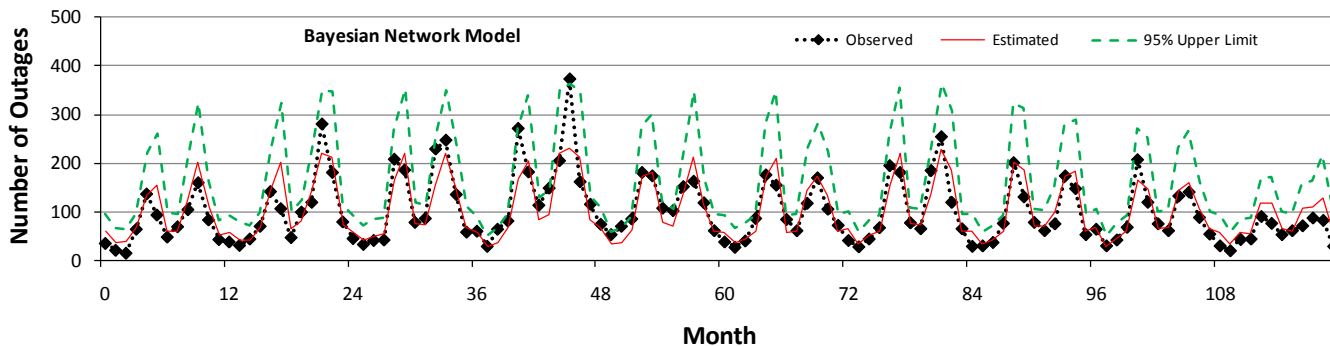


Fig. 6. Observed values of outages and monthly estimates with 95% limits by Monte Carlo simulation using the Bayesian network model for Wichita

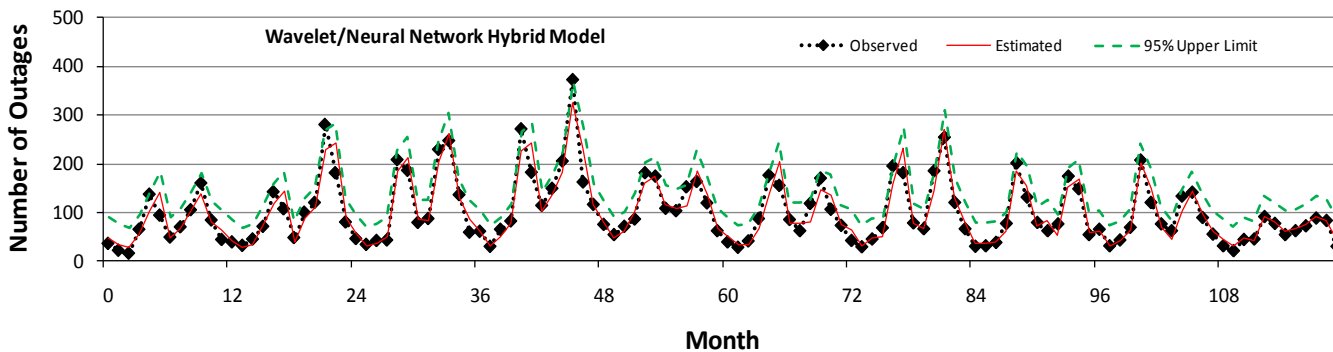


Fig. 7. Observed values of outages and monthly estimates with 95% limits by wavelet/neural network hybrid model for Wichita

hybrid model [6] over the same duration for Wichita. It can be clearly seen that the estimated values follow the observed values in both cases. However, the results of the wavelet/neural network hybrid model follow the outages more closely with a smaller variance compared to the Bayesian network model. Similar results were obtained for the other cities. Absolute Average Error (AAE) and the maximum error between the estimated and observed values of outages with the two models for the four cities are shown in Table IV. Both the AAE and the maximum error are lower for all the cities for the wavelet/neural network model. The maximum error decreases with the size of the city except that it is higher for Manhattan than Lawrence. However, this is true for both the models, which could be due to uncertainties in the data.

TABLE IV  
AAE AND MAXIMUM ERROR FOR THE TWO MODELS

	Bayesian Model		Wavelet/ANN Model	
	AAE	Max Error	AAE	Max Error
Wichita	22	142	15	70
Topeka	18	80	14	76
Lawrence	6	29	5	19
Manhattan	5	37	4	28

The monthly observed values were found to be below the 95% upper limit in all but four cases (May 2001, August 2001, October 2001, and March 2002) in Wichita, three months (May 2000, August 2001, April 2004) in Topeka, two months in Lawrence (January 2006 and May 2006), and two months (December 2001 and October 2004) in Manhattan with the Bayesian network model. With the wavelet/neural network hybrid model observations for four months (October 1999, May 2001, October 2001, and September 2004) in Wichita, five months in Topeka (October 1999, May 2000, May 2003, May 2004, and May 2006) five months in Lawrence (May 1999, September 2001, February 2004, May 2005, and May 2006), and three months in Manhattan (May 2003, May 2004, and September 2004) were found to be above the 95% limit. Only a few observations (May 2001 and October 2001 for Wichita, May 2000 for Topeka, and May 2006 for Lawrence) were higher than the upper limit of both the Bayesian network model and the wavelet neural network model. Table V shows selected (some of the months with observed outages higher than the estimated outages from either models) results for each of the four cities. Rows with observed outages higher than the upper limit of either of the models are shaded with their entries shown in bold. Note that in several cases the observed outages are only slightly higher than the upper limit. Results for other years were very similar and thus are not included in the paper. Estimation of yearly outages using these methods did not yield meaningful results because month to month temporal variations cancelled out in the yearly aggregate.

## VI. CONCLUSIONS

The main focus of this paper is to present a modified Bayesian network model and apply it to study animal-related outages over a period of ten years in four cities in Kansas and compare the results with those obtained with wavelet/neural network hybrid model [6]. The Bayesian network model

presented in this paper is able to capture the time-based pattern in animal-related outages.

TABLE V  
OBSERVED VALUES, ESTIMATED VALUES, AND 95% UPPER LIMITS FOR  
OUTAGES IN DIFFERENT CITIES FOR SELECTED YEARS

	Observed Outages	Bayesian Model		Wavelet/ANN Model	
		Estimated	Upper Limit	Estimated	Upper Limit
<b>Wichita (2001)</b>					
Jan	61	56.89	91	61.70	102.96
Feb	30	30.35	52	32.17	73.43
Mar	65	36.56	67	49.67	90.93
Apr	82	67.63	97	77.81	119.07
<b>May</b>	<b>271</b>	<b>169.18</b>	<b>270</b>	<b>225.92</b>	<b>267.18</b>
Jun	182	204.30	339	242.13	283.39
Jul	114	84.08	131	102.14	143.40
<b>Aug</b>	<b>149</b>	<b>95.24</b>	<b>145</b>	<b>136.09</b>	<b>177.35</b>
Sep	205	220.12	353	183.25	224.51
<b>Oct</b>	<b>372</b>	<b>230.37</b>	<b>364</b>	<b>326.18</b>	<b>367.43</b>
Nov	162	211.54	346	232.45	273.70
Dec	116	84.03	132	116.53	157.79
<b>Topeka (2000)</b>					
Jan	41	49.44	79	68037	105.02
Feb	43	33.11	55	28.49	65.14
Mar	42	38.45	63	27.72	64.37
Apr	60	49.31	71	81.46	118.13
<b>May</b>	<b>212</b>	<b>131.77</b>	<b>205</b>	<b>135.52</b>	<b>172.17</b>
Jun	172	149.94	229	226.04	262.69
Jul	54	54.74	82	61.71	98.36
Aug	68	48.72	80	64.46	101.11
Sep	188	142.01	214	167.70	204.35
Oct	155	168.29	253	170.00	206.65
Nov	94	109.63	169	90.03	126.68
Dec	31	49.57	79	64.28	100.93
<b>Lawrence (2006)</b>					
<b>Jan</b>	<b>45</b>	<b>23.45</b>	<b>37</b>	<b>44.20</b>	<b>59.43</b>
Feb	17	16.67	25	22.20	37.43
Mar	17	13.41	24	19.69	34.92
Apr	26	15.12	26	21.97	37.20
<b>May</b>	<b>75</b>	<b>46.43</b>	<b>72</b>	<b>58.55</b>	<b>73.78</b>
Jun	68	59.56	89	78.88	94.11
Jul	37	28.65	43	42.48	57.71
Aug	31	24.30	38	31.88	47.11
Sep	48	43.89	70	35.06	50.29
Oct	53	52.24	78	59.60	74.83
Nov	56	56.14	84	65.87	81.10
Dec	19	26.15	40	15.45	30.68
<b>Manhattan (2004)</b>					
Jan	9	10.83	21	12.45	23.45
Feb	5	6.16	13	5.30	16.09
Mar	7	6.85	14	5.03	15.82
Apr	18	12.96	23	15.81	26.60
<b>May</b>	<b>46</b>	<b>28.05</b>	<b>54</b>	<b>33.78</b>	<b>44.57</b>
Jun	46	28.06	53	60.66	71.45
Jul	7	11.48	23	14.13	24.92
Aug	19	12.43	24	17.40	28.19
<b>Sep</b>	<b>27</b>	<b>19.85</b>	<b>41</b>	<b>14.65</b>	<b>25.44</b>
<b>Oct</b>	<b>65</b>	<b>27.94</b>	<b>53</b>	<b>81.01</b>	<b>91.81</b>
Nov	29	27.65	53	28.74	39.27
Dec	16	11.28	22	16.84	27.74

Monte Carlo simulations with the Bayesian network model enable determination with great accuracy of the mean and confidence limits for the monthly estimates of outages. The upper and the lower limits provide a range within which the total outages are expected to lie 95% of the time. The upper limits are particularly useful to utility companies as they provide a benchmark on animal-caused outages for each month of the given year. The utilities would need to do field evaluations should the observed outage counts exceed the upper limit.

Comparison with results obtained from the wavelet/neural network hybrid model show consistency in performance of both models. Although the wavelet/neural network hybrid model tracks the outages more closely and has lower variance, both models were equally effective in screening the outages to determine months with observed outages higher than the upper limit. The Bayesian network model is attractive because of its simplicity and ease of implementation.

The methodology presented in this paper is designed for year-end screening of past year's reliability performance of the distribution systems. Only if the observed values for a given month are higher than the estimated upper limit, the utilities would have to do additional investigations to locate the causes of problems and devise remedial measures. Further, the results would allow utilities to justify relatively large outages occurring in their systems in their annual reports to the state utility commissions as long as they remain below the model's upper limit.

#### ACKNOWLEDGMENT

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

#### REFERENCES

- [1] S. D. Frazier and C. Bonham, "Suggested Practices for Reducing Animal-caused Outages," *IEEE Industry Applications Magazine*, Vol. 2, Issue 4, 1996, pp. 25-31.
- [2] S. D. Frazier, "Animals, Power Systems, and Reliability in a Deregulated Environment," *IEEE PES Summer Meeting*, Vol. 1, 1999, pp. 302-307.
- [3] 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*, Vol. 8, Issue 3, 1993, pp. 1268-1274.
- [4] L. Xu, M.-Y. Chow, "A Classification Approach for Power Distribution Systems Fault Cause Identification," *IEEE Transactions on Power Systems*, Vol. 21, Issue 1, 2006, pp. 53 – 60.
- [5] S. Sahai and A. Pahwa, "A Probabilistic Approach for Animal-Caused Outages in Overhead Distribution Systems," *International Conference on Probabilistic Methods Applied to Power Systems*, Stockholm, Sweden, June 2006, pp. 1-7.
- [6] M. Gui, A. Pahwa, and S. Das, "Analysis of Animal-Related Outages in Overhead Distribution Systems with Wavelet Decomposition and Immune Systems Based Neural Networks," *IEEE Transactions on Power Systems*, Vol. 24, Issue 4, pp. 1765-1771.

- [7] M. Gui, A. Pahwa, and S. Das, "Classification of Input and Output Variables for a Bayesian Model to Analyze Animal-Related Outages in Overhead Distribution Systems," *International Conference on Probabilistic Methods Applied to Power Systems*, Singapore, June 2010, pp. 469-474.
- [8] S. Z. Zhang, N. H. Yang and X. K. Wang, "Construction and Application of Bayesian Networks in Flood Decision Supporting System," *Proc. of the First International Conference on Machine Learning and Cybernetics*, Vol. 2, Nov. 2002, pp. 718 – 722.
- [9] 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, 1996, pp. 826 – 831.
- [10] 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.
- [11] 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 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.
- [12] 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.

#### BIOGRAPHIES

**Min Gui** (M'09) received the B.E. degree in Information Engineering and the M.E. degree in Pattern Recognition and Intelligence System from Central South University, Changsha, China, in 2000 and 2003 and the Ph.D. degree in Electrical Engineering from Kansas State University, Manhattan, USA, in 2009. Her employment experience includes HuNan XiangNengXuJi High Technique Company (2003 to 2005) and Quanta Technology, Raleigh, NC as a Senior Engineer (June 2009 to May 2010). Currently, she is working at Luminant Energy as a Senior Market Risk Analyst in Dallas, TX. Her fields of interest include energy market, generation optimization and planning, distribution system reliability, and quantitative analytics for power systems. Dr. Gui received the second prize in the student poster-paper contest at the IEEE Power Engineering Society General Meeting in Tampa, FL, in June 2007.

**Anil Pahwa** (F'03) received the B.E. (honors) degree in electrical engineering from Birla Institute of Technology and Science, Pilani, India, in 1975, the M.S. degree in electrical engineering from University of Maine, Orono, in 1979, and the Ph.D. degree in electrical engineering from Texas A&M University, College Station, in 1983. Since 1983, he has been with Kansas State University, Manhattan, where presently he is Professor in the Electrical and Computer Engineering Department. His research interests include distribution automation, distribution system planning and analysis, distribution system reliability, and intelligent computational methods for distribution system applications. Dr. Pahwa is a member of Eta Kappa Nu, Tau Beta Pi, and ASEE.

**Sanjoy Das** received his B.S. in Electrical Engineering from Sambalpur University, Orissa, India in 1987 and the M.S. and Ph.D. in Electrical Engineering from Louisiana State University, Baton Rouge, in 1994. Between 1994 and 1997 he received his postdoctoral training from the University of California, Berkeley and the Smith-Kettlewell Institute, San Francisco. He has also held research positions in the industry. Since 2001, he has been in the Electrical and Computer Engineering Department at Kansas State University, where he is currently Associate Professor. His research interests include multi-agent systems, machine learning, bio-inspired computing, game theory, quantum computing, and multi-objective optimization, and their applications to power systems and smart grids. Dr. Das is a member of ACM.