Analysis of Tree-Caused Faults in Power Distribution Systems

Le Xu, Student Member, IEEE, Mo-yuen Chow, Senior Member, IEEE and Leroy S. Taylor, Member, IEEE

Abstract—The reliability and quality of power distribution systems are usually affected by many different distribution faults. Tree-caused faults are one of the major fault causes. In this paper, four different measures: actual values, normalized values, relative values, and likelihood values are used to analyze the characteristics of tree-caused distribution faults. This paper also uses statistical techniques to analyze tree-caused faults with respect to several selected factors. The results can be used to assist power distribution engineers to design a more effective tree-fault prevention strategy and to provide a more effective fault restoration system.

Index Terms—power distribution systems, statistical analysis, tree-caused distribution faults.

I. INTRODUCTION

POWER systems play a very important role in our modern society, affecting virtually every person on a daily basis. Power distribution systems, which cover large geographical areas with substantial operation costs, aim at providing reliable, economical and safe supply of electricity to the customers of the power companies. These systems require fast response to changes in load demand, component failures and outage [6].

Distribution faults have been significantly affecting power systems’ reliability, security and quality, among other factors [1]. A lot of utility companies experience a large number of distribution faults in their systems. For example, Duke Energy Company has distribution systems in North Carolina and South Carolina. It has 1,787 circuits, 1,694,557 customers, 265,472 fuse locations, 4,295 recloser locations, 501,073 overhead transformers, 91,916 underground transformers, and 55,436 primary line miles. There are, on average, about 40,000 permanent distribution faults, 30,000 major temporary faults that trip main circuit breakers and 75,000 minor temporary faults that trip line breakers (reclosers) each year.

Power distribution systems are generally geographically dispersed, and have both global and local features under various dynamic operating environments. They are large-scale, nonlinear, time-varying systems. In addition, the exposure of some distribution systems to the uncertainty of the outdoor environment makes them highly vulnerable to natural disturbances and at the same time difficult to analyze and model. Therefore, a fault management system is desirable that can adapt to the environment and collect appropriate information to make the best decisions.

It is important to have a good understanding of the characteristics of the faults before developing effective fault management schemes. Better understanding of causes and consequences of distribution faults will be helpful to maintain distribution systems as well as to design future system in terms of reliability, protection devices installation and environment issues among others.

Power distribution systems can be affected by many different events, including equipment failure, animals, trees, winds, lightning, and others. Tree outage constitutes almost 17% of the total outages in Duke Power System and cost billions of dollars for its tree trimming program. It is critical to develop effective tools to reduce tree faults and the corresponding tree program cost.

Recent research has been focusing on the animal-contact caused faults [2], [4]. In this paper, we will address and analyze the characteristics of tree faults, one of the major fault causes, based on the historical outage data, and will investigate the influence of several selected factors on the tree faults occurrence.

II. POWER DISTRIBUTION FAULT DATA COLLECTION SYSTEM

Duke Energy has a detailed Power Distribution Fault Data Collection System. Every time a fault current is detected in the distribution system as a result of the activation of some protection device, like breaker relay, the distribution fault data are recorded into the Power Distribution Fault Data Collection System.

Each record contains many information fields, to name a few: circuit ID number, weather condition, and time of the fault occurrence etc. Some of the information items (e.g. the number of phases affected, protective devices activated) are directly based on data collected by the monitoring equipment of the distribution systems. Other the information items (e.g. weather condition) are based on a coding system, designed by Duke Power engineers, which describes the circumstances under which the outages occurred.
Selection of information for analysis

In this paper, we select seven regions of the Duke Power System: Chapel Hill, Clemson, Durham, Greenville, Hickory, Lancaster and Winston-Salem to perform tree analysis. These regions are reasonable representations of Duke Power’s different service areas: eastern region, central region, southern region and mountain area; these regions also cover metropolitan areas, cities, towns, rural areas and wooden areas; new systems and old systems.

There are thirty-three information items recorded in each outage data. Of these items, five factors that may contain essential information for tree caused faults are selected [1], [4]: circuit ID number, weather code, time off number, phases affected and the protective devices activated. The circuit ID number uniquely identifies a feeder in the distribution systems and provides the specific geographic information for that circuit. Weather code represents the weather condition at the time of outage using the coding system. The time off number gives the time information of the outage that includes month, day and time on the fault occurrence. The item of phases affected records the phases (X, Y, Z) affected by the fault. Finally, the protective devices activated item indicates which protective device activated as a result of the fault.

We further decompose the time off number into season, day of week and time of day individually. We use the number of phases affected to represent the item of phases affected. The seven factors are represented in sets as follows:

- CI (Circuit ID) = {all the circuit identification numbers under consideration}
- WE (Weather) = {fair, cold, rain, wind, wind & lightning, lightning, hail, snow, ice, hot}
- SE (Season) = {spring, summer, fall, winter}
- DW (Day of week) = {Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday}
- TD (Time of day) = {midnight, morning, afternoon, evening}
- NP (Number of Phases Affected) = {1, 2, 3, no info}
- PD (Protective Device) = {Transmission Device, Station Circuit Breaker, Line Recloser, Primary Fuse, Transformer Fuse, Transformer CSP, Panel Base, SEC/SVC Self Clearing, Manual Device, Primary Self Clearing}

Among the distribute fault data records, one of the fields is “fault cause” code, recorded by the crew during restoration service. For instance, a transformer failure was caused by an animal, the cause code is 4 (animal contact). Distribution fault causes are divided into 12 categories. In this paper, we will focus only on the tree fault, and the cause code is modified into binary form, if it is a tree caused fault, this item is set to 1; otherwise, it is set to 0. Cause code is represented in set form:

- CC (cause code) = \{1, tree fault \{0, non-tree fault

III. ANALYSIS OF TREE CAUSED FAULTS AT DUKE ENERGY POWER DISTRIBUTION SYSTEMS

We use four different measures as reported in [3] to analyze the tree caused distribution faults: (1) actual number of faults, (2) normalized number of faults, (3) relative number of faults, and (4) likelihood of faults.

The first measure, actual number of faults, is simply the actual number of tree faults in different regions, as displayed in Fig. 1.

Fig. 1 indicates that Greenville has the most tree faults among the selected seven regions, while Chapel Hill has the least tree fault. Does it mean that tree faults are more likely to happen in Greenville than other regions? We need to take the effects of regional factors such as circuit size into account to answer this question. The second measure, normalized measure, is investigated in equation (1) and is depicted in Fig. 2.

$$N_{\text{tree}} = \frac{\text{# of tree faults in region } R}{\text{# of total faults in region } R} \quad (\text{pu}) \quad (1)$$

Comparing Fig. 2 with Fig. 1, the normalized tree fault distribution is very different from the actual tree fault distribution. Fig. 2 implies that tree faults are more likely to happen in Chapel Hill than in Greenville, even though the actual number of the tree faults in the former region is much less than the latter. Greenville has a much larger distribution circuits than Chapel Hill, as a result, Greenville has more tree faults than Chapel Hill; on the other hand, Chapel Hill has more wooden area than Greenville, as a result, tree faults are more likely to happen in Chapel Hill than in Greenville.

To further analyze the data, we use the other two measures: relative number of faults and likelihood of faults as in equation (2) and (3) respectively. These two measures provide more detailed information for fault analysis of each selected factor.
The third measure is the relative measure of distribution faults of event \( x \) that belong to factory \( y \):

\[
R_{xy} = \frac{\text{# of tree faults at event } x \text{ in region } R}{\text{# of total faults in region } R \text{ (pu)}}. \tag{2}
\]

In this paper, \( y \in \{\text{CI, WE, SE, DW, TD, NP, PD}\} \), for \( y=SE \), event \( x \in \{\text{spring, summer, fall, winter}\} \), etc. This measure gives the relative frequency of event \( x \) that occurs in the factor \( y \) condition.

The fourth measure is the likelihood of a fault happening, given different conditions:

\[
L_{xy} = \frac{\text{# of tree faults at event } x \text{ in region } R \text{ (pu)}}{\text{# of total faults at event } x \text{ in region } R \text{ (pu)}}. \tag{3}
\]

The likelihood measure gives the probability of faults, given that a specific event \( x \) also occurs. The likelihood measure can be considered as the conditional probability of tree fault given the event \( x \).

Using these two measures, the data are further categorized according to weather condition, season, day of week, time of day, the number of phases affected, protective devices activated as shown in the following figures.

**A. Weather condition:**

![Fig. 3a Relative Value of Tree Fault with respect to Weather in the selected regions.](image)

![Fig. 3b Likelihood of Tree Fault with respect to Weather in the selected regions.](image)

Fig. 3a indicates that tree faults happen more often in fair weather, windy and lightning weather and windy weather, while only a few tree faults happen in snowy and icy weather.

Looking only at this measure, the result may contradict our intuition that snowy or icy weather would be more likely to cause the tree to fall down, and then lead to the fault of the power distribution system. However, our intuition is confirmed by the likelihood measure, presented in Fig. 3b. Fig.3b suggests that the likelihood of the tree faults in snowy and icy weather is much higher than that in fair weather.

One of the explanations of this phenomenon is that in the Duke Energy service areas, there are not much severe weather condition such as snow and ice during the winter. Thus, even if the snow and ice heavily influence the tree faults occurrence, not many tree faults in snowy and icy weather happen. Meanwhile, considering the other possible faults, tree faults are less likely to occur than other faults in fair weather condition in Duke Energy service area.

**B. Season**

![Fig. 4a Relative Value of Tree Fault with respect to Season in the selected regions.](image)

![Fig. 4b Likelihood of Tree Fault with respect to Season in the selected regions.](image)

Fig. 4a indicates that in Duke Energy Service areas, there are the most tree faults in the Summer and the least tree faults in the Fall. From the investigation of the likelihood of tree faults versus season, tree faults are found more likely to happen in the Winter than in the Summer (Fig. 4b).

Considering other faults, we can use similar arguments with the influence of weather condition in tree faults to normalize the tree faults with respect to those other faults. For example, animal faults are more likely to happen in the Summer than in the Winter, one of the reasons is that the hibernation of some animals, such as snake, in the Winter...
reduces the possibility of animal-contact caused faults. Also, there is a higher chance for snowy or icy weather, which has heavy influence on tree faults, to happen during winter time. Relatively speaking, the possibility of tree faults in winter increases compared with all other faults.

C. Day of Week

![Fig. 5a Relative Value of Tree Fault with respect to Day of week in the selected regions.](image)

![Fig. 5b Likelihood of Tree Fault with respect to Day of Week in the selected regions.](image)

Fig. 5a Relative Value of Tree Fault with respect to Day of week in the selected regions.

Fig. 5b Likelihood of Tree Fault with respect to Day of Week in the selected regions.

Basically, no obvious pattern of the tree fault with respect to day of week can be found from Fig. 5a and 5b, which suggests that the tree faults may not be greatly influenced by the day of week.

D. Time of Day

![Fig. 6a Relative Value of Tree Fault with respect to Time of Day in the selected regions.](image)

![Fig. 6b Likelihood of Tree Fault with respect to Time of Day in the selected regions.](image)

Fig. 6a Relative Value of Tree Fault with respect to Time of Day in the selected regions.

Fig. 6b Likelihood of Tree Fault with respect to Time of Day in the selected regions.

Fig. 6a implies that usually more tree faults happen in the afternoon than in the midnight. While Fig. 6b indicates that tree faults are more likely to happen during the evening and midnight than afternoon.

With consideration of one of the other possible factors: the human activities, there are more human activities during the day time. Some of the human activities (e.g. traffic accident, operation mistakes) may cause outage or faults in power distribution systems, thus increase the possibility of these faults, like Duke Accident Faults or Public Accident Faults. Hence tree faults have relatively lower possibility to happen during the day time compared with other faults. The contrary argument can apply to the higher probability of tree fault in the night time.

E. The Number of Phases Affected

![Fig. 7a Relative Value of Tree Fault with the factor of NP in the selected regions.](image)

![Fig. 7b Likelihood of Tree Fault with respect to NP in the selected regions.](image)

Fig. 7a Relative Value of Tree Fault with the factor of NP in the selected regions.

Fig. 7b Likelihood of Tree Fault with respect to NP in the selected regions.
Fig. 7a indicates that when a tree fault happens, it often affects only one phase, sometimes three phases, but rarely two phases. This conclusion is reasonable because distribution circuits have more single-phase lines than three-phase lines.

Fig. 7b (the likelihood measure) implies that it is more likely to be a tree fault when a fault affects two phases than the one affecting single phase. With consideration to other possible faults, the tree faults are normalized with respect to these faults. For instance, animal-contact caused faults usually cause single phase distribution faults or activate a three phase transformers. Yet, it is unlikely for an animal to create a fault involving two phases to ground. This decreases the possibility of tree caused one single phase fault and increases the possibility of tree caused two phases fault.

F. Protective Devices Activated

![Relative Value of Tree Fault with respect to PD in the selected regions](image1)

Fig. 8a. Relative Value of Tree Fault with respect to PD in the selected regions.

![Likelihood of Tree Fault with respect to PD in the selected regions](image2)

Fig. 8b. Likelihood of Tree Fault with respect to PD in the selected regions.

Fig. 8a indicates that the protective device activated by the tree faults is mainly the primary fuse, but on average, line reclosers are more likely activated by tree faults (Fig. 8b).

This information is useful for identifying the cause of distribution system faults. For example, if we have two similar faults as far as other six factors are concerned, but one fault activate liner recloser, the other one activate transformer fuse, then the first event is more likely to be caused by a tree than the second one.

Seven selected regions have different circuit ID numbers; therefore it is impossible to use the same figure axis labels to represent the fault distribution for every region in one 3-D figure as presented previously. Even for a specific region, there are around 100 circuits related to outage, hence the number of circuit ID is too large to display the relative measure or likelihood measure clearly in one single plot here. One possible way, however, is to categorize these outage related circuit ID numbers into several categories according to the geographical characteristic of each circuit, such as wooden, town. It needs more detailed information about each feeder in the distribution systems.

IV. LOGISTIC REGRESSION ANALYSIS OF TREE FAULTS

As from the previous analysis, some of the factors have obvious pattern (e.g. weather), while others do not (e.g. day of week). A logical question to ask is “Do all these seven factors have significant effect on the tree fault occurrence?” We will use statistical tools to answer this question.

We will use “cause code” as the dependent variable in the statistical analysis. Since the tree fault is coded as either 1 or 0, as mentioned in section II, CC is not normally distributed. In addition, all values of seven independent variables {CI, WE, SE, DW, TD, NP, PD} are nominal values instead of continuous values. We cannot use the conventional analysis of variance (ANOVA) technique to analyze the tree cause faults. Instead, we use logistic regression to analyze the tree fault with respect to each of the selected items, to estimate the effect of each of the independent variables on the categorical output, tree fault or non tree fault. Logistic regression does not assume the normally distribution of the independent variables and dependent variables. Logistic regression also makes no assumption about the linearity of relationship between the independent variables and dependent variables and the homogeneity of variance.

Logistic regression transforms the dependent variables into a logit variable, which is a natural log of the odds of the dependent occurring or not, and then applies maximum likelihood estimation to calculate the logit coefficient of the model. The odds that the dependent may be predicted based on the values of the independent variables is reflected from the log likelihood, which is maximized by maximum likelihood estimation [10].

Fig. 1 indicates that the Lancaster region has more tree faults than five other regions, only less than Greenville. Fig. 2 shows that the tree faults are most likely to happen in the circuits of Lancaster. Hence, in this following section, the outage data from Lancaster in Duke Power System are used for illustration purpose.

The log-likelihood test of the whole model, also called “the likelihood ratio test”, is conducted here as a significance test. The chi-square value, p-value and the R-square value will be listed in the Table I. The chi-square value measures the improvement of the fit of the model. A well fitting model, in general, will have a small p-value. The cutoff value for p-value is usually 0.05. R-square, actually a pseudo R-square, serves as an analog to the squared contingency coefficient [10].
Table I. Statistical Result of the Logistic Regression Analysis of Tree Faults

<table>
<thead>
<tr>
<th>Terms</th>
<th>Chi-square</th>
<th>p-value</th>
<th>R-square</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI</td>
<td>558.6895</td>
<td>&lt;0.0001</td>
<td>0.0427</td>
</tr>
<tr>
<td>WE</td>
<td>2084.697</td>
<td>0.0000</td>
<td>0.1594</td>
</tr>
<tr>
<td>SE</td>
<td>79.44719</td>
<td>&lt;0.0001</td>
<td>0.0061</td>
</tr>
<tr>
<td>DW</td>
<td>164.7656</td>
<td>&lt;0.0001</td>
<td>0.0126</td>
</tr>
<tr>
<td>TD</td>
<td>176.4331</td>
<td>&lt;0.0001</td>
<td>0.0135</td>
</tr>
<tr>
<td>PD</td>
<td>808.2838</td>
<td>&lt;0.0001</td>
<td>0.0618</td>
</tr>
<tr>
<td>NP</td>
<td>358.254</td>
<td>&lt;0.0001</td>
<td>0.0274</td>
</tr>
</tbody>
</table>

Table I shows that the factor of weather (WE) have a zero p-value, and all the other six p-values are less than 0.0001. All the p-values of seven selected influential factors are much smaller than the cutoff value of 0.05, which indicates that all of these seven factors are statistically significant to the tree fault occurrence. Thus, it is worthwhile to include these factors to help diagnose whether a power distribution fault is caused by trees.

We further use R-square value to evaluate and rank the relative importance of the seven selected factors. Table I shows that the R-square value of weather (WE) category is the highest. This result supports the previous conclusion that weather is the most influential factor for tree fault occurrence.

As discussed previously, no obvious pattern of the tree fault with respect to day of week (DW) can be found, so day of week (DW) probably has the lowest value of R-square. However, Table I indicates that the season (SE) category, instead of day of week (DW), has the lowest R-square value. If we take Greenville as another example, we can find that its R-square value for day of week (DW) is the lowest, just as we predicted. In fact, distribution faults have both global and local properties. We can get some general conclusions from the historical outage data, but some factors don’t lead to any general conclusion, but rather depend on individual region/circuit characteristics [1].

By examining the data from all other regions, we can find that all these seven factors are statistically significant to the tree fault occurrence for all regions; among these seven factors, weather (WE) category always has the highest R-square value, while the factors of season (SE) and day of week (DW) always have relatively low R-square among the seven selected factors.

V. CONCLUSION

Tree faults are one of the major causes of power distribution system outages. In this paper, we have discussed and analyzed the tree faults in power distribution systems, based on the actual outage data collected by Duke Energy Company. We use four different measures, namely, actual values, normalized values, relative values, and likelihood values to gain a better understanding of the characters of tree faults. In addition, we use logistic regression analysis to investigate the significance of several selected factors (e.g. weather) to the influence of tree fault. The results can be used in our future work to design an effective tree fault cause identification scheme.

VI. REFERENCES