Process Level Fault Probability Distribution Model of Intelligent Substation Based on Tree Structure

Diqiu Shen\textsuperscript{1*}, Kefeng Zhang\textsuperscript{2}, Liang Wang\textsuperscript{1}, Wenxing Lu\textsuperscript{1} and Qiaojie Zhu\textsuperscript{1}

\textsuperscript{1}China Southern Power Grid EHV Power Transmission Company Liuzhou Bureau, Liuzhou, Guangxi 545006, China
\textsuperscript{2}China Southern Power Grid EHV Power Transmission Company, Guangzhou, Guangdong 510530, China
\textsuperscript{*}E-mail: sundigger@163.com

Abstract. In order to estimate the probability distribution of process level fault in Smart Substation, a tree based probability distribution model of process level fault in Smart Substation is proposed. Taking 14 years’ process level fault data of Smart Substation in a regional power grid as sample data, this paper compares and analyzes various probability models. The probability distribution of process level fault scale of Smart Substation in this area is calculated by using the Boley Tanner branching process model in tree structure. The experimental results show that the model can estimate the probability distribution of process level fault size of Smart Substation well. Under the same confidence requirement, the sample data needed by the method based on the Boley Tanner model to estimate the fault probability distribution is one order of magnitude less than that directly based on the actual fault data.

1. Introduction

Today’s power system is developing to the direction of large capacity, ultra-high voltage, long distance, large-scale interconnection, which brings good economic benefits, but also challenges the safe and reliable operation of power grid \cite{1}. In recent years, the continuous blackouts at home and abroad have caused immeasurable economic losses and social impact, which has aroused widespread concern of the international community.

Since 2001, scholars at home and abroad (involving power system, mathematics, statistical physics, nonlinear dynamics and other fields) began to study the evolution and intervention of process level faults in Smart Substation from different perspectives \cite{2}. These researches mainly focus on: 1) the characteristics and evolution of process level fault in Smart Substation from a macro perspective; 2) the intervention of process level fault in Smart Substation, the role of a certain kind of disturbance in the process of process level fault evolution or the identification of fault chain; 3) the supplement of traditional risk management methods, using probability model, based on simulation data and actual fault in this paper, the fault probability distribution of intelligent substation is analyzed\cite{3}.

Wu Xia et al. \cite{4} considered the seasonal characteristics of wind power and load and built a random fuzzy model of net active power distribution of high wind power penetration substation bus at the same time in the day. In this method, the historical data of the net active power of the bus bar of a specific high-speed wind power penetration substation are analyzed by curve fitting, and the probability density function is obtained, and the specific location is obtained based on the fuzziness of the probability distribution parameters. Li Junfeng et al.\cite{5} pointed out that the traditional fault...
diagnosis model of the low voltage side of the main transformer needs a lot of fault feature data, and the fault feature data are not independent, so it is difficult to find the most suitable fault feature data. It is an important part of traditional power system reliability evaluation to calculate historical fault statistics. Based on the above research, this paper applies the tree structure to the design of process level fault probability distribution model of intelligent substation, so as to estimate the probability distribution of process level fault in intelligent substation.

2. Design of process level fault probability distribution model for Smart Substation

2.1 The analysis of the characteristics of process level fault data in intelligent substation

One of the common characteristics of blackouts is the successive faults of transmission lines. Therefore, to a certain extent, the number of line faults can be used to measure the severity of blackouts and characterize the development process of blackouts [6-7]. Therefore, this paper only analyzes the process level fault of intelligent substation, and does not consider other types of fault, such as generator fault, transformer fault and so on.

Due to the small proportion of 110kV and below faults, and some 110kV faults do not provide accurate time scale information, this paper only analyzes the fault data of 220kV and above, a total of 733 faults. There are different fault forms in the data, including three-phase short circuit, single-phase short circuit, automatic reclosing, and accelerated tripping fault after reclosing in the fault line [8]. According to the characteristics of the stage propagation of process layer fault in Smart Substation, this study classifies different faults into different stages of process layer fault in Smart Substation according to the time of fault occurrence [9]. Table 1 shows the total number of failures in each stage of these smart substation process level failures.

| stage | Number of failures |
|-------|--------------------|
| Z₀    | 556                |
| Z₁    | 83                 |
| Z₂    | 31                 |
| Z₃    | 20                 |
| Z₄    | 14                 |
| Z₅    | 6                  |
| Z₆    | 5                  |
| Z₇    | 3                  |
| Z₈    | 3                  |
| Z₉    | 3                  |
| Z₁₀   | 3                  |
| Z₁₁   | 2                  |
| Z₁₂   | 2                  |
| Z₁₃   | 1                  |
| Z₁₄   | 1                  |
| Z₁₅   | 0                  |

It can be seen from the Table 1 that the maximum number of fault propagation stages is 14. The fault scale is measured by the number of trip lines in the process level fault and initial fault of an intelligent substation, and 459 process level faults of intelligent substation are analyzed. The analysis results are given in Table 2.

| Number of trip lines | Initial failure times | Initial failure probability | Number of process level failures in Smart Substation | Process level fault probability of Intelligent Substation |
|----------------------|-----------------------|-----------------------------|-----------------------------------------------------|-------------------------------------------------------|
| 1                    | 402                   | 0.8758                      | 341                                                 | 0.7429                                                |
| 2                    | 35                    | 0.07625                     | 62                                                  | 0.1351                                                |
| 3                    | 13                    | 0.02832                     | 27                                                  | 0.05882                                               |
| 4                    | 5                     | 0.01089                     | 10                                                  | 0.02179                                               |
It can be seen from Table 2 that the maximum number of trip lines is 7 in the initial fault, and the maximum number of trip lines is 19 in the process level fault of an intelligent substation. The probability of a process level fault of an intelligent substation with a fault scale of 7 or less trip lines accounts for about 99%, which indicates that it is difficult to accurately estimate the probability of large-scale fault by using traditional analysis methods.

2.2 Tree structure branching process model

Branching process is a random process, which describes the process of splitting or perishing of a group of particles [10]. According to the characteristics of the branching process, the fault evolution process of the process layer of the smart substation can be described as follows: at the beginning, there are \( Z_0 (Z_0 > 0) \) lines disconnected, because these lines are disconnected, the reliability of the power grid is reduced, resulting in \( Z_1 \) lines disconnected in the next stage; and then there are \( Z_2 \) lines disconnected in the next stage. This propagation, until no fault occurs at a certain stage, indicates the end of the process layer fault process of the smart substation.

2.2.1 Branching probability

One of the important parameters of the branching process model is the branching probability \( \hat{\lambda} \), that is, the probability that one fault in any stage will produce \( r \) \((r = 0, 1, 2, \ldots)\) in the next stage. All the faults in the second stage to the last stage are caused by the faults in the previous stage respectively. Therefore, we can use the following method to estimate \( \hat{\lambda} \).

Assuming that there are \( J \) Smart Substation process layer faults in total, and let \( Z_k^{(j)} (j = 1, 2, \ldots, J; k = 0, 1, \ldots) \) represent the number of faults in the \( k \)-th stage of the \( j \)-th Smart Substation process layer fault, then the estimated value of branching probability \( \hat{\lambda} \) is:

\[
\hat{\lambda} = \sum_{i=1}^{J} \frac{\sum_{j=1}^{J} (Z_{00}^{(i)} + Z_{01}^{(i)} + \cdots + Z_{N(i)}^{(i)})}{\sum_{j=1}^{J} (Z_{00}^{(i)} + Z_{01}^{(i)} + \cdots + Z_{N(i)}^{(i)})} \quad (1)
\]

In this formula, \( J \) is the total number of faults and \( N(i) \) is the maximum number of stages where the number of line faults is not zero.

By substituting the data in Table 1 into equation (1), it can be concluded that:

\[
\hat{\lambda} = \frac{Z_1 + Z_2 + \cdots + Z_{15}}{Z_0 + Z_1 + \cdots + Z_{15}} \approx 0.24 \quad (2)
\]

When \( \hat{\lambda} \) value is greater than 1, it means that the process layer fault of intelligent substation will continue to develop until the process layer fault scale of intelligent substation reaches the maximum scale specified by the system; when \( \hat{\lambda} \) value is less than 1, the fault scale gradually decays to 0, and the process layer fault of Intelligent Substation stops. According to the calculation results, \( \hat{\lambda} < 1 \) indicates that all process level faults of Smart Substation will eventually end.
2.2.2 Process level fault probability distribution model of Intelligent Substation

Another important part of the tree structure branching process model is the probability distribution of initial failure [13]. If the initial fault distribution is assumed to be Poisson distribution, then the process level fault probability distribution of Smart Substation is generalized Poisson distribution; if the initial fault distribution is arbitrary, then the process level fault probability distribution of Smart Substation is Boley Tanner distribution.

(1) Generalized poisson model

If the initial fault distribution obeys the Poisson distribution with the parameter and does not consider the case that the initial fault is 0, then the initial fault probability distribution is:

$$P[Z_0 = r] = \frac{e^{-\theta} \theta^r}{(1-e^{-\theta}) r!}, r = 1, 2, \ldots (3)$$

In this formula, $r$ is the number of faulty lines. $\overline{Z_0}$ is the mean value. In this case, the process level fault size of Smart Substation follows the generalized Poisson distribution, as shown in equation (4).

$$H[Z = r] = \hat{\theta}(r \hat{\lambda} + \hat{\theta})^{r-1} \frac{e^{-r \hat{\lambda}}}{(1-e^{-\theta}) r!}, r = 1, 2, \ldots (4)$$

The value of $\hat{\lambda}$ can be obtained from equation (2). According to table 2, we can see $\overline{Z_0} \approx 1.211$, so from equation (4) we can get $\hat{\theta} \approx 0.3965$.

(2) Boley Tanner model

If the initial fault has arbitrary distribution, the process level fault size of Smart Substation follows the Boley Tanner distribution.

$$P[Z = r] = \sum_{z_0} P[Z_0 = z_0] z_0 \hat{\lambda}(r \hat{\lambda})^{r-1-z_0} \frac{e^{-r \hat{\lambda}}}{(r-z_0)!} (5)$$

In this equation, $P[Z_0 = z_0]$ can be calculated from the actual fault data, and the value of parameter $\hat{\lambda}$ can be calculated from equation (2).

3. Experimental results and analysis

3.1 Probability estimation of fault size

Taking the process level fault data of Smart Substation as the sample data, the branch process based probability models (including generalized Poisson model and Boley Tanner model), traditional Poisson model and power-law model are used to estimate and compare the process level fault probability distribution of Smart Substation. By substituting parameters, values and values of different fault lines into equation (4), equation (5), the probability distributions estimated by different models can be obtained. Figure 1 shows the estimated results of different probability models.
It can be seen from Figure 1 that when the number of small-scale faults is less than 5, the four models can well estimate the fault probability. With the increase of fault scale, the fitting degree of the Boletan model and the actual fault data is obviously better than the other three models. Among them, the power-law model overestimates the probability of blackout. In practical problems, whether it is power system blackout fault data or data in other fields, its power-law distribution characteristics generally appear in the tail of the probability density function curve. Taking blackout fault as an example, the traditional reliability research considers that the probability distribution of fault is exponential distribution, while the actual fault data and research methods are different. The results show that the probability distribution of failure presents partial power law, that is, large-scale failure presents power law.

3.2 Goodness of fit test

Chi square goodness of fit test is one of the methods to judge the similarity between the theoretical distribution obtained by probability model and the sample distribution obtained by actual data.

Chi square goodness of fit method is described as follows:

Supposing $X$ is a discrete distribution with the value of $1, 2, \cdots, k$ and $P(X = i) = p_i, 1 \leq i \leq k$. $X$ is observed $n$ times, $X_i (i = 1, 2, \cdots, k)$ is the number of times $X$ takes $i$ in $n$ times. Then we can compare the difference between the expected frequency and the observed frequency of $X$ each value by statistics $\chi^2$.

$$\chi^2 = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i} \quad (6)$$

When $n \to \infty$, the limit distribution of $\chi^2$ obeys the $\chi^2$ distribution of degree of freedom $k - 1$.

It can be seen from equation (6) that the larger the $\chi^2$ is, the larger the sample data deviates from the probability model. Therefore, we divide the sample space into five disjoint subspaces: $M_1 = \{1\}$, $M_2 = \{2\}$, $M_3 = \{3\}$, $M_4 = \{4\}$, $M_5 = \{5, 6, \cdots\}$.

In this equation, $M_i = \{i\} (i = 1, 2, 3, 4)$ indicates that the space contains the case that the number of faults is $i$; $M_5 = \{5, 6, \cdots\}$ indicates that the space contains the case that the number of faults is $5, 6, \ldots$. After the above space partition, the number of process level faults $X_i (i = 1, 2, 3, 4, 5)$ of intelligent substation with total faults in each subspace follows polynomial distribution. The results of chi square test are shown in Table 3.
### Tab. 3 Chi square values of different models

| $r$ | $X_i$ | Boley Tanner | Generalized Poisson | Power law | Power law |
|-----|-------|--------------|---------------------|-----------|-----------|
|     |       | $p_i$ | $np_i$ | $p_i$ | $np_i$ | $p_i$ | $np_i$ |
| 1   | 341   | 0.6877 | 315.7 | 0.6398 | 293.7 | 0.6450 | 296.1 | 0.5505 | 252.7 |
| 2   | 62    | 0.1776 | 81.5  | 0.2211 | 101.5 | 0.1613 | 74.04 | 0.3286 | 150.8 |
| 3   | 27    | 0.06875 | 31.56 | 0.08276 | 37.99 | 0.07167 | 32.90 | 0.09809 | 45.02 |
| 4   | 10    | 0.03128 | 14.36 | 0.03273 | 15.02 | 0.04031 | 18.50 | 0.01952 | 8.960 |
| ≥5  | 19    | 0.03467 | 15.91 | 0.02361 | 10.84 | 0.08172 | 37.51 | 0.00329 | 1.510 |

It can be seen from Table 3 that the $\chi^2$ value calculated by the Boley Tanner model is obviously smaller than other models, that is, the Boley Tanner model has the best fitting degree with the sample data. Making $\chi^2_\alpha(m)$ be the significance level $\alpha$ and the chi square value under the degree of freedom $m$, then when $\alpha$ is 0.01, the $\chi^2 < \chi^2_{0.01}(3) = 11.34$ of the Boley Tanner model, that is, when the significance level is 0.01, the Boley Tanner model can well describe the distribution of the actual fault data.

### 4. Conclusion

In this paper, the applicability of traditional Poisson, power-law, generalized Poisson and Boley Tanner probability models to process level fault probability distribution estimation of Smart Substation is compared based on 14 years of line fault data of a regional power grid. By analyzing the characteristics of protection and automatic devices and dispatching behavior, this paper analyzes and preprocesses the fault record data of power grid in 14 years, and applies it to the calculation of branch process model parameters. The analysis results show that compared with the traditional probability model, the Boley Tanner model can better estimate the probability distribution of fault size, and has a high degree of fitting with the actual sample data distribution. For the same set of sample data, power-law model overestimates the probability of large-scale failure, while Poisson model and generalized Poisson branching process model underestimate the probability.

### Reference

[1] Maanaoui H.E., Gehrmann O., Yaagoubi M.E. Lifetime prediction of simple shear loaded filled elastomers based on the probability distribution of particles[J]. Polymer Testing, 2019, 75:229-236.

[2] Krivoruchenko, M. I. The probability distribution of the number of electron-positron pairs produced in a uniform electric field[J]. Physical Review D, 2012, 86(2):27704-27704.

[3] Noel, N, Nemeth. Probability density distribution of the orientation of strength-controlling flaws from multiaxial loading using the unit-sphere stochastic strength model for anisotropy[J]. International Journal of Fracture, 2014, 185(1/2):97-114.

[4] Karimi A, Aminifar F, Fereidunian A, et al. Energy storage allocation in wind integrated distribution networks: An MILP-Based approach[J]. Renewable energy, 2019, 134: 1042-1055.

[5] Liao X, Wang H, Niu J, et al. Research on Simulation Training System of Immersive Substation Based on Virtual Reality[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing. 2019, 486(1): 012112.

[6] Asce J L L M, Jan Eliá, and Zdeněk P. Baant Hon.M.ASCE. Computation of Probability Distribution of Strength of Quasibrittle Structures Failing at Macrorack Initiation[J]. Journal of Engineering Mechanics, 2012, 138(7):888-899.
[7] Zhang Y C , Zhao H Q , Jiang W , et al. Time dependent failure probability estimation of the solid oxide fuel cell by a creep-damage related Weibull distribution model[J]. International Journal of Hydrogen Energy, 2018, 43(29):13532-13542.

[8] He H , Lee W J , Luo D S , et al. Insulator Infrared Image Denoising Method Based on Wavelet Generic Gaussian Distribution and MAP Estimation[J]. IEEE Transactions on Industry Applications, 2017, 53(4):3279-3284.

[9] Amjad, Saeed, Khan, et al. Improved Bounds on the Decoding Failure Probability of Network Coding Over Multi-Source Multi-Relay Networks[J]. IEEE Communications Letters, 2016, 20(10):2035-2038.

[10] Ben El-Shanawany A , Ardron K H , Walker S P . Lognormal Approximations of Fault Tree Uncertainty Distributions[J]. Risk Analysis, 2018, 38(8):1576-1584.