Using Transmission Line Outage Data to Estimate Cascading Failure Propagation in an Electric Power System

Hui Ren, *Student Member, IEEE*, and Ian Dobson, *Fellow, IEEE*

***Abstract*—We study cascading transmission line outages recorded over nine years in an electric power system with approximately 200 lines. The average amount of propagation of the line outages is estimated from the data. The distribution of the total number of line outages is predicted from the propagation and the initial outages using a Galton–Watson branching process model of cascading failure.**

***Index Terms*—Failure analysis, power systems, reliability, risk analysis.**

I. Introduction

CASCADING failure is the process by which initial outages of electric power transmission system components can occasionally propagate to more widespread outages and large blackouts. The outages are dependent in that the outages that have already occurred weaken the system and make further outages more likely. Indeed, the empirical probability distributions of blackout size observed in several countries have approximate power law regions that cannot be produced by independent outages and are broadly consistent with cascading failure models [1].

Cascading failures that arise in practice can be very complicated chains of events and often include unanticipated interactions or rare events. One reason for this is that power system engineers work hard to mitigate the likely and anticipated cascades. We suggest that any practical approach to estimation of the overall probabilities of cascading failure must be a bulk statistical "top-down" approach that neglects some of the detail of the cascades.

A bulk statistical approach is different from and complementary to methods of risk analysis that rely on detailed analysis of enumerated interactions. A detailed analysis can describe the risk of a subset of cascades, particularly some of the likely, anticipated and shorter cascades so that high risk cascades can be mitigated [2]. Moreover, the detailed analysis can support specific mitigations of the interactions or component reliabilities. However, the detailed analysis cannot describe the overall risk of cascading failure because it does not account for the huge number of unlikely or complicated events, especially when the interactions due to network physics are augmented with the interactions due to software and human factors. The bulk statistical approach seeks to quantify the overall risk of cascading failure and the reliability benefits of improvements to the system.

In this paper, we analyze some power industry data on transmission line outages. Of course all cascading interactions, no matter whether rare and intricate or likely and simple, are accounted for in the observed data. We choose to examine outages of high voltage lines because lines typically outage during transmission system disturbances and outages of other components or operational, planning and maintenance errors tend to also cause line outages. In a sense we are using the line outages to monitor the more general cascading processes.

We summarize previous work on probabilistic models for cascading outages. Dobson and coworkers developed a probabilistic model of cascading [3], approximated this model with a branching process [4]–[6], and considered how branching processes could model the propagation of outages in data observed in single, large blackouts [7]. They applied Galton–Watson and continuous state branching processes to estimate propagation and the distributions of line outages and load shed in data produced by the OPA simulation of cascading line outages [8], [6], [9], [10]. Chen and McCalley *et al.* [11] proposed an exponentially accelerated cascading model for the number of line outages and fit this model, a generalized Poisson model (follows from a branching process), and a negative binomial model to the distribution of the total number of line outages observed in North America over 20 years [12].

In this paper, we demonstrate a way to test a bulk statistical model of cascading line outages on industry data. The bulk statistical model is a branching process that describes initial outages that then propagate in stages. We estimate the average amount of propagation λ and hence predict the probability distribution of the total number of line outages. This efficiently predicted probability distribution of the total number of line outages is compared to the empirical probability distribution of the total number of line outages to test whether the branching process model is valid for this prediction. We show how to quantify the propagation of cascading failures from observed data and efficiently predict the effect of the propagation on the risk of the cascade. A summary of some of the results of this paper appears in the conference paper [9].

Manuscript received November 13, 2007; revised March 18, 2008. First published June 10, 2008; current version published September 12, 2008. This work was supported in part by the Power Systems Engineering Research Center (PSERC) and the National Science Foundation (NSF) under Grant ECCS-0606003 and Grant SES-0623985. The work of H. Ren was supported in part by scholarships from the People's Republic of China and NCEPU. This paper was recommended by Associate Editor C. Nwankpa.

H. Ren was with Electrical and Computer Engineering Department, University of Wisconsin, Madison, WI 53706 USA. She is now with North China Electric Power University, Hebei 071003, China (e-mail: dahui.ren@gmail.com).

I. Dobson is with the Electrical and Computer Engineering Department, University of Wisconsin, Madison, WI 53706 USA (e-mail: dobson@engr.wisc.edu.)

Digital Object Identifier 10.1109/TCSII.2008.924365

II. Branching Process

Branching processes have long been used in a variety of applications to model cascading processes [13], [14], but their application to the risk of cascading failure is recent [4]. The Galton–Watson branching process gives a probabilistic model of the number of failures. There are a random number Z_0 of initial failures that then propagate randomly to produce subsequent failures in stages. Each failure in each stage (a "parent" failure) independently produces a random number $0, 1, 2, 3, \ldots$ of failures ("child" failures) in the next stage according to an offspring distribution that is a Poisson distribution of mean λ. The child failures then become parents to produce the next generation and so on. If the number of failures in a stage becomes zero, the cascade stops. The mean number of child failures for each parent is the parameter λ. λ quantifies the average tendency for the cascade to propagate. The intent of the modeling is not that each parent failure in some sense "causes" its child failures; the branching process simply produces random numbers of failures in each stage that can statistically match the outcome of cascading processes.

There are general arguments supporting the choice of a Poisson distribution for the offspring distribution [15]. The Poisson distribution is a good approximation when each failure propagates to a large number of components so that each parent failure has a small, fairly uniform probability of independently causing child failures in a large number of other components. This assumption seems reasonable for cascades in power systems, especially in the initial portions of the cascade when there are many unfailed components that are stressed by the failed components.

The power system cascades are observed until there are K nontrivial cascades. Each nontrivial cascade has a positive number of failures in stage zero $(Z_0 > 0)$ and all statistics are conditioned on $Z_0 > 0$. The failures in the kth cascade are written as $Z_0^{(k)}, Z_1^{(k)}, Z_2^{(k)}, Z_3^{(k)}, \ldots$ so that $Z_j^{(k)}$ is the number of outages in stage j of cascade k.

We assume an arbitrary distribution of nonzero initial failures $P[Z_0 = z_0]$ for $z_0 = 1, 2, 3, \ldots$. Then it is a standard result in branching processes that the total number of failures Y is distributed according to a mixture of Borel-Tanner distributions

$$P[Y = r] = \sum_{z_0=1}^{r} P[Z_0 = z_0] z_0 \lambda (r\lambda)^{r-z_0-1} \frac{e^{-r\lambda}}{(r-z_0)!}. \tag{1}$$

Distribution (1) is valid[1] for $0 \le \lambda < 1$ and shows that the distribution of the total number of failures Y depends on the distribution of the initial failures Z_0 and the propagation λ.

The standard Harris estimator for λ is

$$\lambda = \frac{\sum_{k=1}^{K} \left(Z_1^{(k)} + Z_2^{(k)} + \ldots \right)}{\sum_{k=1}^{K} \left(Z_0^{(k)} + Z_1^{(k)} + \ldots \right)} \tag{2}$$

[1]Distribution (1) can be generalized to $\lambda \ge 1$ by modeling saturation as explained in [8].

and is an asymptotically unbiased maximum likelihood estimator [13], [16], [17]. Estimator (2) is intuitive: Think of "parent" failures in each generation giving rise to "child" failures in the next generation. Then λ is the average family size; that is, the average number of child failures for each parent. Since $Z_0, Z_1, \ldots$ are all parent failures and $Z_1, Z_2, \ldots$ are all child failures, the estimator (2) is simply the total number of children in all the cascades divided by the total number of parents in all the cascades. A variant of (2) that accounts for saturation effects is used in [8].

III. Outage Data

The data is from a regional electric power transmission system with, approximately, 100 buses, 180 lines at 220 kV, and 20 lines at 500 kV. The data is recorded over about 9 years starting in 1997 and ending in 2006. The data for each transmission line outage includes the time (to the nearest minute), voltage level, and the auto-recloser's action. The voltage levels considered are 220 kV and 500 kV; outages at lower voltage levels are not considered because of the potential number of unrecorded cases. There are several types of line outages in the data, including three phase and single phase and outages with successful or unsuccessful auto-reclosing. In processing the data, both voltage levels and all types of line outages are regarded as the same and the detailed causes of the line outages (line fault, busbar fault, or other fault or operations) are neglected. Neglecting these distinctions in an initial, bulk statistical analysis is appropriate (future work may account for some of these distinctions). Large flashover events in the data with approximately 260 outages over two days are neglected because they lack time tags.

IV. Grouping Outages into Cascades and Stages

For our analysis it is necessary to group the line outages first into different cascades, and then into different stages within each cascade [7]. Here we use a simple method based on outages' timing. Since operator actions are usually completed within one hour, we assume that successive outages separated in time by more than one hour belong to different cascades. Since transients or auto-recloser actions are completed within one minute, we assume that successive outages in a given cascade separated in time by more than one minute are in different stages within that cascade. Much of the clustering of outages in stages can be seen in Fig. 1.

Table I is obtained by summing over all the 226 cascades the number of outages in each stage. That is, of the 396 outages, 296 are in stage 0 of a cascade, 45 are in stage 1 of a cascade, and so on. The initial outages are the 296 outages in stage 0. The probability distribution of the number of initial outages Z_0 is shown in Fig. 2(a). The distribution of Z_0 in Fig. 2(a) has a peak at 6 outages that prevents it being well approximated by a Poisson distribution. One reason for the peak is that some cascades are initiated by a bus outage, and the relay trips off all transmission lines connected to that bus simultaneously at the start of the cascade.

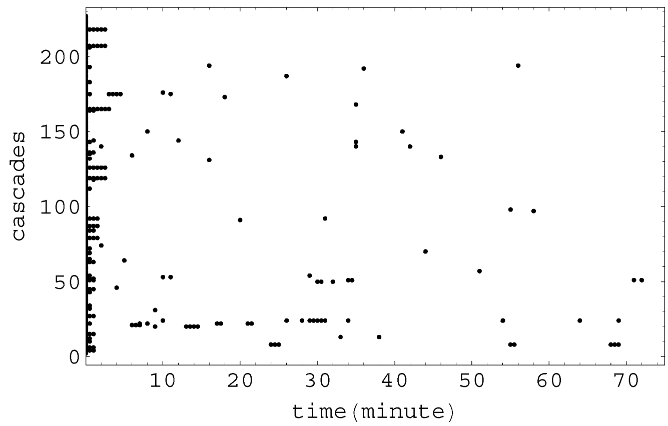

Fig. 1. Time since start of cascade for outages in each of the 226 cascades. The first 75 min of each cascade are shown (3 cascades exceed 75 min). Multiple outages at the same time are shown slightly displaced.

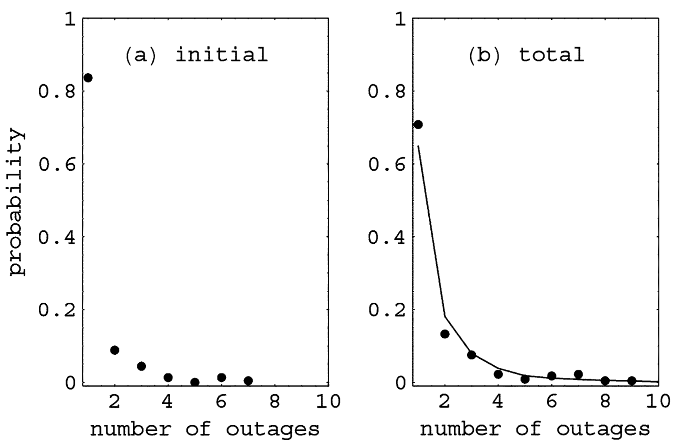

Fig. 2. Probability distributions of number of outages. (a) Initial outages Z_0 from data. (b) Total number of outages Y estimated using branching process (line) and from data (dots).

TABLE I
NUMBER OF OUTAGES IN EACH STAGE SUMMED OVER THE CASCADES

stage number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
no. of outages	296	45	18	14	10	3	1	1	1	1	1	1	2	1	1	0

V. ESTIMATING λ AND DISTRIBUTION OF TOTAL NUMBER OF OUTAGES Y

Applying estimator (2) to the line outage data in Table I yields the equation shown at the bottom of the page. That is, each outage produces an average of $\lambda = 0.25$ outages in the next stage. This result is insensitive to the grouping of outages into stages (redefining the minimum time between successive outages in different stages to be 2 min and recomputing λ yields $\lambda = 0.24$).

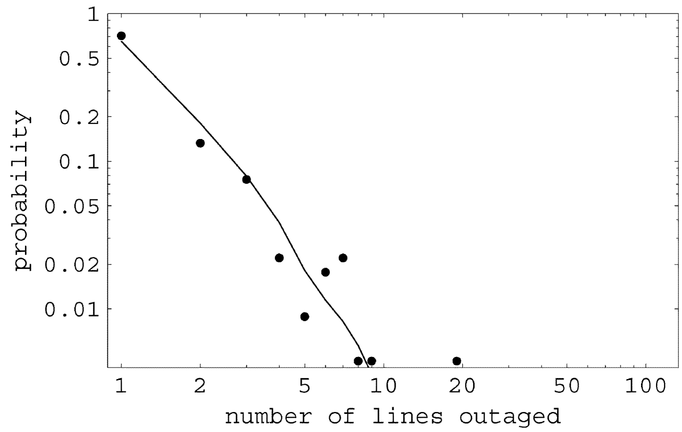

Fig. 3. Distribution of total number of outages Y estimated using branching process (line) and from data (dots); this log–log plot shows a heavy tail.

The empirical distribution of Y directly obtained from the data is

$$p_r = \frac{\text{number of cascades with } r \text{ outages}}{\text{number of cascades}}, \quad r = 0, 1, 2, \ldots. \quad (3)$$

To test how well the branching process model describes the data, we use (1) with $\lambda = 0.25$ and the distribution of initial outages to predict the distribution of the total number of outages Y, and compare this with the empirical distribution (3). This comparison is shown in Figs. 2(b) and 3. A chi-squared goodness-of-fit test shows that the distributions are consistent at the 5% confidence level (the test groups together 5 or more outages). A heavy tail in the distribution of the total number of line outages is also observed in North American data in [11], but our data has a heavier tail than [11].

VI. COMPARING ESTIMATION OF DISTRIBUTION OF Y VIA λ AND EMPIRICALLY

Suppose that the distribution of the initial failures Z_0 is known using standard methods of reliability analysis. Then the distribution of the total number of outages Y can be obtained by first estimating λ and then using the branching process model (1), or obtained empirically directly from the observed cascade data. We compare for these two approaches the number of cascades that need to be observed to yield a comparable standard deviation of the tail of the estimated distribution of Y. In particular, we obtain a result for the tail of more than 5 line outages. Although these tail events are rare, they can have high impact and substantial risk. The outage of many lines is more likely to lead to load shedding.

Observing outages for one year would yield an average of 25 cascades. To show how accurately λ could be estimated from

$$\lambda = \frac{45+18+14+10+3+1+1+1+1+1+1+2+1+1+0}{296+45+18+14+10+3+1+1+1+1+1+1+2+1+1} = 0.25$$

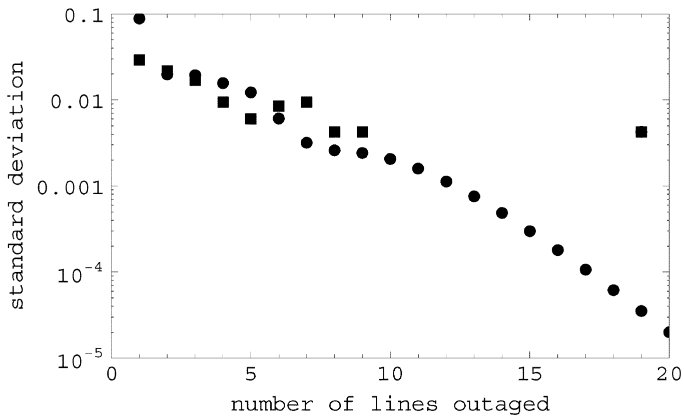

Fig. 4. Standard deviation σ_B for estimation via the branching process with one year of cascade data (circles) compared to the standard deviation σ_E for empirical estimation with 9 years of cascade data (squares). $\sigma_B < \sigma_E$ for more than 5 line outages.

one year of data, we took 9 non-overlapping random samples of 25 cascades and estimated λ for each sample of 25 cascades. A typical result is that the estimated λ has a standard deviation of 0.14. That is, assuming normality, an estimate of λ from one year of data lies within 0.14 of the true value about 68% of the time. This accuracy could be improved by collecting data over a longer time or over a larger region to increase the number of observed cascades.

The branching process probability of r total outages $P[Y = r]$ depends on λ according to (1) and, by linearizing (1), the standard deviation of $P[Y = r]$ due to a 0.14 standard deviation of λ is

$$\sigma_B(r) \approx \left| \frac{\partial P[Y = r]}{\partial \lambda} \right|_{\lambda=0.25} 0.14. \quad (4)$$

That is, (4) gives the standard deviation of the estimated distribution of Y for observing cascades in the 200 line power system for one year and using the branching process model.

The standard deviation of the empirical distribution (3) of the total number of outages for observing the 226 cascades in the 200 line power system for 9 years is

$$\sigma_E(r) \approx \sqrt{\frac{p_r(1 - p_r)}{226}}. \quad (5)$$

Fig. 4 compares the standard deviation σ_B for estimation via the branching process with the standard deviation σ_E for empirical estimation. Since for more than 5 line outages, σ_B computed from 1 year of data or 23 cascades is less than the σ_E computed for 9 years of data or 226 cascades, we can conclude that estimation of the tail of the distribution via the branching process requires an order of magnitude fewer cascades for the same uncertainty.

For the case studied, this conclusion of efficient estimation of the tail of the distribution via the branching process holds assuming that the distribution of initial failures Z_0 is known. If it is required to estimate the distribution of Z_0 empirically from only one year of data, then the resulting uncertainty in the estimated distribution of Z_0 would introduce significant uncertainty into the estimate of the distribution of Y.

We now consider how the size of the network and time for which it is observed generally affect the interpretation and accuracy of λ and the estimated distribution of failures. A bulk statistical approach inherently describes cascading with some "averaging" over space and time. In particular, estimating the propagation λ over a given part of a network yields a branching process model for the propagation of failures in that particular part of the network that is spatially homogeneous over that part of the network. Similarly, estimating the propagation λ over a given time period yields a branching process model for the propagation of failures for that particular time period that is temporally homogeneous over the time period. (Note that there could be systematic variations in λ at a fast time scale during cascades as well as slower time variations in λ as the average network stress changed. Variations at both the slow and fast timescales are averaged in obtaining the branching process model that is homogeneous over the time period.) The distributions of total number of failures predicted using λ are similarly averaged over a particular network and a particular time period. The averaging over space and time is routine in bulk statistical analyses, but needs to be kept firmly in mind in choosing the extents of the observations in space and time and in interpreting the results obtained.

It is desirable to increase the "resolution" of the results so that they could be obtained for and apply to smaller portions of networks observed over shorter periods of time. But a sufficient number of cascades need to be observed to get accurate enough results and the number of cascades observed is proportional to the size of the portion of the network observed and to the time period. In particular, the standard deviation of the estimate of λ is proportional to $1/\sqrt{K}$ where K is the number of cascades observed. Thus, increasing the accuracy of the estimate of λ by a factor of 2 requires observation of a network four times larger for the same time or observation of the same network for four times the time period. Alternatively, for a given accuracy of the estimate of λ, the time resolution can be doubled and the spatial resolution can be halved by halving the observation time and doubling the size of the network observed. One way to express the thrust of this paper is to note that empirical estimation of the tail of the probability distribution of the total number of failures has poor resolution because of the large number of cascades that need to observed and that much better resolution in time or space can be obtained by estimating parameters of models of cascading failure. In particular, assuming that the initial failure distribution is known, the results show that estimating λ of a branching process model for one year can yield a more accurate estimate of the tail of the probability distribution of the total number of failures than observing the same network for 9 years.

VII. Conclusion

We consider cascading transmission line outages observed in about 9 years of operation of a 200 line power system by describing their bulk statistical behavior rather than the details of individual cascades. We group the line outages into cascades and stages according to their outage times and then estimate the average propagation λ of the outages. For this data, the empirical distribution of the total number of line outages is well approximated by the initial line outages propagating according to a branching process with propagation parameter λ. In particular, this data supports the validity of the branching process model for prediction of the distribution of the total number of line outages. We hope that additional industry data sets will become available to engineers and risk analysts to test and confirm this new application of branching process models to cascading line outages in power systems. Establishing the practical validity of a branching process model in this way would contribute foundational knowledge for the statistical analysis of blackouts and cascading failure.

Estimating λ requires much less data than directly estimating the heavy tail of the empirical distribution, and this can enable the distribution of the tail of the total number of line outages to be estimated from line outage data observed over a much shorter time. In particular, suppose that the distribution of initial line outages is computed by standard methods of reliability analysis and that the propagation λ has been estimated from observations by the method of the paper. Then the branching process model efficiently predicts the tail of the distribution of the total number of line outages. This is a novel and practical way to compute the consequences of propagation of outages in cascading failure.

We have demonstrated a new method to predict from industry data the probability distribution of the size of cascading outages when the initial outage distribution is known. The success of the prediction for this data indicates that further testing of branching process models of cascading failure is appropriate.

Acknowledgment

I. Dobson would like to thank J. Rundle and the Center for Computational Science and Engineering, University of California-Davis, for their generous hospitality during a sabbatical leave.

References

[1] I. Dobson, B. A. Carreras, V. E. Lynch, and D. E. Newman, "Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization," *Chaos*, vol. 17, no. 2, Jun. 2007, 026103 (13 pages).

[2] Q. Chen and J. D. McCalley, "Identifying high risk n-k contingencies for online security assessment," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 823–834, May 2005.

[3] I. Dobson, B. A. Carreras, and D. E. Newman, "A loading-dependent model of probabilistic cascading failure," *Prob. Eng. Inf. Sci.*, vol. 19, no. 1, pp. 15–32, 2005.

[4] I. Dobson, B. A. Carreras, and D. E. Newman, "A branching process approximation to cascading load-dependent system failure," in *Proc. 37th Hawaii Int. Conf. Syst. Sci.*, Big Island, HI, Jan. 2004.

[5] I. Dobson, B. A. Carreras, and D. E. Newman, "Probabilistic load-dependent cascading failure with limited component interactions," in *Proc. IEEE Int. Symp. Circuits Syst.*, Vancouver, Canada, May 2004, vol. V, pp. 912–915.

[6] I. Dobson, B. A. Carreras, V. E. Lynch, B. Nkei, and D. E. Newman, "Estimating failure propagation in models of cascading blackouts," *Prob. Eng. Inf. Sci.*, vol. 19, no. 4, pp. 475–488, Oct. 2005.

[7] I. Dobson, B. A. Carreras, and D. E. Newman, "Branching process models for the exponentially increasing portions of cascading failure blackouts," in *Proc. 38th Hawaii Int. Conf. Syst. Sci.*, Big Island, HI, Jan. 2005.

[8] I. Dobson, K. R. Wierzbicki, B. A. Carreras, V. E. Lynch, and D. E. Newman, "An estimator of propagation of cascading failure," in *Proc. 39th Hawaii Int. Conf. Syst. Sci.*, Kauai, HI, Jan. 2006.

[9] I. Dobson, K. R. Wierzbicki, J. Kim, and H. Ren, "Towards quantifying cascading blackout risk," in *Proc. Bulk Power Syst. Dyn. Contr. VII*, Charleston, SC, Aug. 2007.

[10] K. R. Wierzbicki and I. Dobson, "An approach to statistical estimation of cascading failure propagation in blackouts," in *Proc. 3d Int. Conf. Crit. Infrastruct.*, Alexandria, VA, Sep. 2006.

[11] Q. Chen, C. Jiang, W. Qiu, and J. D. McCalley, "Probability models for estimating the probabilities of cascading outages in high-voltage transmission network," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1423–1431, Aug. 2006.

[12] R. Adler, S. Daniel, C. Heising, M. Lauby, R. Ludorf, and T. White, "An IEEE survey of US and Canadian overhead transmission outages at 230 kV and above," *IEEE Trans. Power Delivery*, vol. 9, no. 1, pp. 21–39, Jan. 1994.

[13] T. E. Harris, *Theory of Branching Processes*. New York: Dover, 1989.

[14] K. B. Athreya and P. E. Ney, *Branching Processes*. New York: Dover, 2004.

[15] P. C. Consul and M. M. Shoukri, "Some chance mechanisms related to a generalized Poisson probability model," *Amer. J. Math. Manag. Sci.*, vol. 8, no. 1 and 2, pp. 181–202, 1988.

[16] N. M. Yanev, "On the statistics of branching processes," *Theory Prob. Appl.*, vol. 20, pp. 612–622, 1975.

[17] P. Guttorp, *Statistical Inference for Branching Processes*. New York: Wiley, 1991.