IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 5, 2020, accepted March 26, 2020, date of publication March 30, 2020, date of current version April 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2984240

A Probabilistic Cascading Failure Model for Dynamic Operating Conditions

RUI MA, (Student Member, IEEE), SHENGMIN JIN, (Student Member, IEEE), SARA EFTEKHARNEJAD, (Senior Member, IEEE), REZA ZAFARANI, (Member, IEEE), AND WOLF PETER JEAN PHILIPPE, (Student Member, IEEE)

Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244, USA

Corresponding author: Sara Eftekharnejad (seftekha@syr.edu)

This work was supported by the National Science Foundation (NSF) under Grant 1600058.

ABSTRACT Failure propagation in power systems, and the possibility of becoming a cascading event, depend significantly on power system operating conditions. To make informed operating decisions that aim at preventing cascading failures, it is crucial to know the most probable failures based on operating conditions that are close to real-time conditions. In this paper, this need is addressed by developing a cascading failure model that is adaptive to different operating conditions and can quantify the impact of failed grid components on other components. With a three-step approach, the developed model enables predicting potential sequence of failures in a cascading failure, given system operating conditions. First, the interactions between system components under various operating conditions are quantified using the data collected offline, from a simulation-based failure model. Next, given measured line power flows, the most probable interactions corresponding to the system operating conditions are identified. Finally, these interactions are used to predict potential sequence of failures with a propagation tree model. The performance of the developed model under a specific operating condition is evaluated on both IEEE 30-bus and Illinois 200-bus systems, using various evaluation metrics such as Jaccard coefficient, F_1 score, Precision@K, and Kendall's tau.

INDEX TERMS Cascading failures, power system reliability, propagation of cascades.

I. INTRODUCTION

The IEEE Power and Energy Society Cascading Failure Working Group (CFWG) has defined a cascading failure as a sequence of dependent component failures in which the failure of one or more component leads to the failure of others, continuing with further subsequent failures [1]. The ability to predict cascading failures enables more effective operating decisions, and proper mitigation strategies that will eventually reduce large-scale blackouts. The unprecedented increase in power system uncertainties, due to factors such as increased penetration of wind and solar generation and load uncertainties, makes real-time power system conditions less predictable. This unpredictability demands a dynamic cascading failure model, which can predict probable failures based on conditions that are closer to real-time conditions [2]–[4]. In response to this pressing need, we propose a new efficient and effective cascading failure model. The developed model addresses a main drawback of the existing models, i.e. applicability to online power grid operations.

The concept of developing a cascading failure model has been given significant attention in the power engineering research, and a variety of cascading failure models have been developed. Examples include the ORNL-PSerc-Alaska (OPA) model [5]–[7], Manchester model [8], Hidden failure model [9], topology-based model [10]–[14], CASCADE model [15] and the branching process model [16] to simulate and analyze the impact of cascading failures on power systems. The developed failure models can be grouped into three different categories: (1) *topology-based* [10]–[14], (2) *DC/AC-based* [5]–[9], [17]–[19], and (3) *statistical* models [15], [16]. Topology-based models, motivated by the complex network methods [20]–[22], represent power systems as a large-scale network in which failures spread from a node to its neighbors. With a proper network representation, various methods such as centrality measures (e.g., betweenness) can be used to assess a system's vulnerability to various failures. What limits the application of topology-based methods is the fact that a cascading failure is only treated as a local

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

phenomenon, i.e., loss of a node only affects the neighboring nodes. However, in a real power system, cascading outages propagate globally [23]: a failure of one component can cause another failure in a distant location. Additionally, it is demonstrated in [24] that these topological cascading failure models could lead to erroneous conclusions regarding vulnerable system components, which may result in wrong mitigation strategies. DC/AC-based models such as the OPA model and the Manchester model address this issue by modeling failures with a power law distribution. Based on the DC/AC power flow, the propagation of cascades can be simulated to gain insights into the physics of cascade propagation in the power grids [25]. Hence, the information of cascading outages at each point in time (such as power flows and load losses) can be used to understand the propagation and impacts of cascades [19]. These models, however, are computationally expensive and generate large-scale data [25]. Statistical models, such as the CASCADE model and Branching process model, enable fast generation of cascade data, but neglect power system information such as topology, power flows and power injections [15], [16]. While these models reveal the cascade size, the contribution of each component to a cascading failure, a necessary information for online operations, e.g., generation ramping, cannot be derived. To address the drawbacks of the statistical models, data-driven statistical models such as the influence graph model [25], [26] and the interaction model [27]–[29] are developed to simulate cascading failures using the *interactions* among system components. An interaction is defined as the probability of one component's failure, given the failure of other components. This failure probability can be estimated from historical cascading failure data or data generated from the DC/AC-based models. These interaction and influence graph method, however, do not predict potential failures in near real-time. Recent studies have used learning-based method to efficiently analyze cascading failures. Authors in [30] propose to apply reinforcement learning to efficiently identify critical fault chains. In another study [17], Artificial neural network (ANN) is used to promptly estimate Energy-Not-Supplied in real time. However, the failure probability of each system component, an important important information for system protection, cannot be derived.

To summarize, while current cascading failure models can analyze cascades, these models cannot provide the potential propagation of cascades as well as the failure probabilities of components for system operating conditions that are close enough to real-time, (1) topology-based models do not capture the global propagation of cascades; (2) DC/AC-based models are computationally expensive; (3) statistical models disregard the differences between system components; (4) interaction and influence graph models yield the failure probability but do not consider the effect of different operating conditions on this probability; and (5) learning-based method cannot provide the failure probability of each system component;

To predict potential cascading failures for online operating conditions, a *dynamic cascading failure model* that is adaptive to changing system operating conditions is developed in this paper. The model enables fast estimation of the interactions between system components given present transmission line power flows. Concretely, the model first extracts failure propagation patterns from large-scale simulated cascade data, under different grid operating conditions. Next, specific interactions that correspond to the present transmission line power flows, can be generated. Finally, these interactions are used to forecast potential subsequent failures in cascades under the up-to-date conditions. Having the knowledge of the potential failures, operators could block the corresponding relays [31], i.e., stop tripping of the components and still keep these components in service in power systems, to buy time to identify appropriate mitigative actions and prevent or reduce the impact of cascades.

The contributions of this work over the state-of-the-art in modeling cascading failures can be summarized as:

1) A dynamic cascading failure model, that is adaptive to changing operating conditions, is developed to accurately, and promptly predict potential cascading failures;
2) Bayesian framework is utilized to estimate power grid component interactions for *a specific operating condition*, resulting in more accurate estimation of the interactions for near real-time operations;
3) A *propagation tree* method is developed, that enables predicting the propagation of failures and provides the failure probability for each system component, and *at each point in time*. This information can help initiate actions on the components with high failure probability, e.g. relay blocking, so that the propagation of cascades is prevented. Hence, the operators could have more time to identify optimal mitigative actions;
4) The developed model can forecast impending failures in a significantly shorter time than the DC/AC-based cascading failure models, and thus is more useful for power system monitoring and control.

The paper is organized as follows: Section II introduces the dynamic failure model, including how it quantifies dynamic failure interactions. Section III presents case studies in which the effectiveness of the proposed model in predicting potential cascading failures is evaluated. Conclusions are presented in Section IV.

II. DYNAMIC FAILURE MODEL

Analyzing potential cascades using a simulation-based cascading failure model is time consuming for near real-time applications. Thus, an alternative statistical model that (1) extracts the *failure propagation patterns* from historical or simulated cascade data (collected offline), and (2) enables fast prediction of cascades online using such patterns is needed. Recently developed statistical failure models, namely Qi's interaction model [27], [28] and Hines's influence graph model [25] are efforts towards addressing this critical need in power systems. However, the extracted failure propagation patterns by these two statistical models do not consider the impact of various operating conditions on the failure interactions between components. As shown by Ju *et al.* [29], the interactions significantly vary under different loading conditions. To address the aforementioned problem, it has been suggested in [29] to revise the interaction model with additional large-scale simulated cascade data, an approach that is time-consuming and inapplicable to online operations. This problem will be addressed in this paper.

In this section, a dynamic failure model is developed to enable (1) fast estimation of failure patterns under online operating conditions, and (2) to predict the propagation of cascades, as illustrated in Fig. 1. Offline analysis uses historical or simulated cascade data under different operating conditions to learn all the failure interactions among components under different system states. The real-time analysis, using Phasor Measurement Unit (PMU) measurements, identifies only those interactions (among all) that apply to real-time conditions, hence referred to as *dynamic interactions*. Once the updated interaction model is generated, a methodology developed in this work, referred to as the *propagation tree method*, is used to predict potential propagation of cascades. Eventually, this model helps power system operators better understand the outcomes of the potential failures under the present system operating condition.

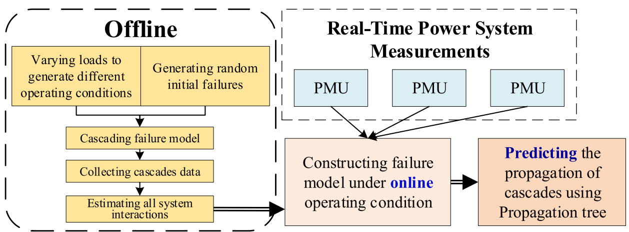


FIGURE 1. Development of a dynamic failure model for predicting cascades.

A. CASCADE DATA

To estimate the interactions between system components, large-scale historical cascade data is needed. As cascades are rare events, real cascade data are not adequate for this purpose. Therefore, simulated cascade data that contain (a) different stages of a cascade, and (b) the initial power loading conditions should be generated as an alternative. In this study, we consider transmission lines and transformers among power system components, while the methodology can be extended to other components. The initiating event (i.e., *generation* 1) of each case in the cascade data is the failure of one or several system components, followed by subsequent failures (*generations*) until no more failures take place, or the system becomes unstable [26]. An example of the cascade data is shown in Fig. 2, where the initial event in generation 1 produces two subsequent failures (i.e., two *children*) in generation 2. The failure of these two components (lines 4 and 5) causes line 2 to fail in generation 3 due to overloads. As no more components are outaged after the outage of line 2, this cascade stops after three generations. The simulation model (discussed in detail in Section III-A) utilizes the mitigative control actions of generation ramping and load shedding to eliminate system limit violations. The generation ramping is first used to adjust the power flow of lines while load shedding acts as the last resort. The other protection strategies such as frequency regulation have not been considered for generation of scenarios. While this may have simplified the scenario generation, in future, the same developed methodology can be applied to a more comprehensive cascading failure database. In validating the results we have made sure that all methods compared are consistent in terms of the available protection/regulation services. As explained next, upon collection of cascade data and using a Bayesian framework, the dynamic interactions between components can be estimated from cascade generations and the initial loading conditions of components (before failures).

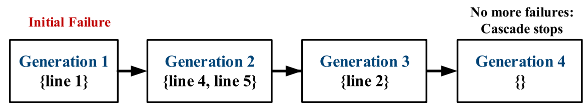


FIGURE 2. Example of a cascade propagation.

B. ESTIMATING DYNAMIC INTERACTIONS

Conventional interaction model is firstly introduced in [27] to identify critical system components that have large contribution to cascading failures. The same authors have further improved the conventional interaction model in [28] to use an efficient estimation technique, i.e., Expectation Maximum (EM) algorithm, to estimate component failure interactions with less cascade data. While the main idea of calculating failure interactions is the same, the difference is in the estimation algorithm. Although conventional interaction models can capture the propagation patterns from cascade data, they neglect how component failure interactions vary with different loading levels of system components. Intuitively, a transmission line with a higher loading ratio of $\frac{\text{Power Flow}}{\text{Thermal Capacity}}$ is more likely to fail compared to a line with a lower loading ratio. To capture the effect of line loading on the propagation of cascades, we develop a dynamic interaction model that incorporates the system loading conditions into the failure interactions. Hence, the model enables updating the interactions based on the up-to-date power flows. A dynamic interaction between failures of two components can be thought of as a conditional failure probability $P(j \mid i, r_{\mathrm{j}})$:

$$P(j \mid i, r_{\mathrm{j}}) = P(j \text{ fails in generation } k+1 \mid i \text{ fails in generation } k, \text{loading ratio of line } j), \quad (1)$$

where $P(j \mid i, r_{\mathrm{j}})$ is the probability of line j failing, given that line i has failed and loading ratio for line j is r_{j}. The loading ratio (or state) r_j for line j is defined as,

$$r_j = \frac{pf_j}{pf_j^{max}} \tag{2}$$

where pf_j and pf_j^{max} are the apparent power flow and maximum apparent power flow capacity of line j. With the obtained cascade data under different operating conditions, Bayes' theorem is used to estimate the conditional probability $P(j \mid i, r_j)$ [32],

$$P(j \mid i, r_j) = \frac{P(j, r_j \mid i)}{P(r_j \mid i)} \tag{3}$$

where $P(j, r_j \mid i)$ is the joint probability of the failure of line j and the state of line j, given line i has failed in the previous generation. $P(r_j \mid i)$ is the conditional probability of the state of line j given the failure of line i. In this study, the loading conditions of system components in a cascade are the states before the occurrence of the initial failures, as in a statistical model it is infeasible to predict the changes in the loading conditions during the propagation of cascades.

To estimate the conditional failure probability $P(j, \ r_j \mid i)$ in (3), the methodology introduced by Qi *et al.* [27] is adopted in this paper. Under a specific loading ratio r, an *interaction matrix* $A \in \mathbb{R}^{n\times n}$ can be obtained, where n is the number of system components. The interaction matrix summarizes the interactions between system components using the stages of cascade data. The elements of A, A_{ij}, represent the number of times, among all generations of all cascades, component j fails subsequent to the failure of component i in the previous generation, when the loading ratio of component j is r_j.

The conventional interaction model in [27] assumes that the failure of component j in a generation is only subsequent to one component's outage in the previous generation; and that component is the one that has led component j to fail the most among all the cascades. This is due to the difficulty in identifying the *complete* cause of subsequent outages in cascade data. After obtaining A, the empirical failure interaction probability between two failed components can be computed and recorded in an *interaction probability matrix* B, with its elements $B_{ij} = \frac{A_{ij}}{n_i}$. B_{ij} is the interaction probability between lines i and j, i.e. the empirical joint failure probability of the state of line j and the failure of line j caused by line i, and n_i is the number of times component i fails among all cascades. The probability matrix B quantifies the interaction between any two lines during the propagation of cascades under a specific line loading ratio. B_{ij} can be thought of as the conditional failure probability $P(j, \ r_j \mid i)$.

The initial failures in the first generation are often caused by exogenous events, e.g., a tree falling on a line, while failures in the subsequent generations are caused by the outages of other system components [33]. Hence, the dynamic interaction between two components is separated into two different interactions: (1) initial interaction $P_0(j \mid i, r_i, r_j)$ and (2) subsequent interaction $P_{1+}(j \mid i, r_j)$, defined as

$$P_0(j \mid i, r_i, r_j) = \frac{P_0(j, r_j, r_i \mid i)}{P_0(r_j, r_i \mid i)}, \tag{4}$$

$$P_{1+}(j \mid i, r_j) = \frac{P_{1+}(j, r_j \mid i)}{P_{1+}(r_j \mid i)}, \tag{5}$$

where 0 and 1+ denote the initial generation and subsequent generations of cascades, respectively. The initial interaction is extracted from the outaged components between the first and second generations, while the subsequent interactions are obtained from all cascade generations except the first generation. The benefits achieved by dividing the initial and subsequent interactions separately will be discussed further in Section III-C. For simplicity, we use D to denote *dynamic* interaction matrices, and D_0 and D_{1+} to denote the initial and subsequent interaction matrices for the rest of the paper.

It can be observed from (2) that the loading ratio r_j is in range [0, 1]. It is impractical to estimate the failure probability of line j for any r_j, as r_j is a continuous value. Therefore, in this paper, instead of estimating failure probability $P_0(j \mid i, r_i, r_j)$ and $P_{1+}(j \mid i, r_j)$ for any loading ratio r, based on Bayes' theorem, we discretize the two interaction probabilities as [32],

$$P_0(j \mid i, r_i \in [\frac{n}{b}, \frac{n+1}{b}), r_j \in [\frac{m}{b}, \frac{m+1}{b})), \tag{6}$$

$$P_{1+}(j \mid i, r_j \in [\frac{m}{b}, \frac{m+1}{b}), \tag{7}$$

where b is the number of bins used to discretize continuous variable r and $m, n \in \{0, 1, \ldots, b-1\}$. For instance, if we use two bins, i.e., $b = 2$, there are four initial interactions: $P_0(j \mid i, r_i \in [0, 0.5), r_j \in [0, 0.5))$, $P_0(j \mid i, r_i \in [0, 0.5), r_j \in [0.5, 1.0))$, $P_0(j \mid i, r_i \in [0.5, 1.0), r_j \in [0, 0.5))$ and $P_0(j \mid i, r_i \in [0.5, 1.0), r_j \in [0.5, 1.0))$. Also, there are two subsequent interactions: $P_{1+}(j \mid i, r_j \in [0, 0.5))$ and $P_{1+}(j \mid i, r_j \in [0.5, 1.0))$.

C. CASCADE PROPAGATION ANALYSIS

In the previous section, the dynamic interactions between system components were estimated and a methodology to incorporate online measurements into these interactions was introduced. To predict potential cascades online, we develop a methodology to calculate the failure probability of each component in each generation of a cascade. In this study, as the causes of the initial events are exogenous, we only focus on analyzing the propagation of cascades. Calculation of failure probabilities is conducted by constructing a *Propagation Tree*, an example of which is shown in Fig. 3. Each layer of the tree corresponds to a generation of a cascade, and each node in a layer represents a component that might fail in the corresponding generation. An edge (i.e., branch) in a propagation tree denotes the interaction between two nodes in two consecutive layers. When the failure probability of a component in a generation is lower than a threshold ε, it is assumed that this component would not generate subsequent child failures. Thus, the corresponding node in the propagation tree becomes a *leaf node*, such as node 3 in generations 3 and 5 in Fig. 3. The steps to construct a propagation tree are:

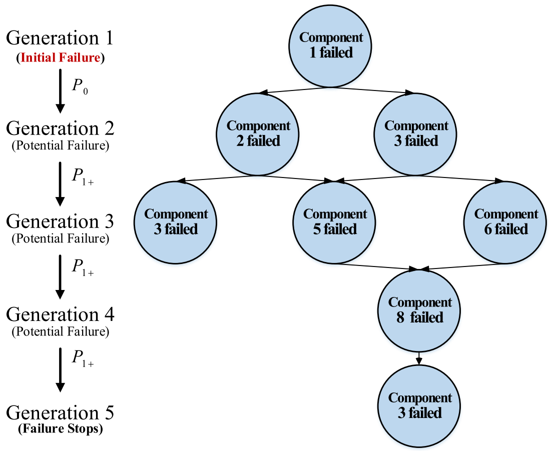


FIGURE 3. An example of a propagation tree.

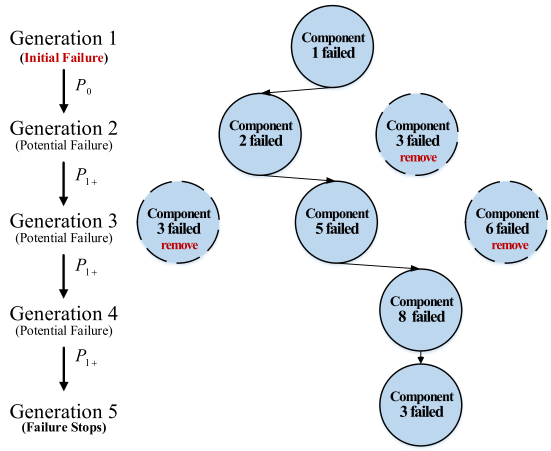


FIGURE 4. An illustration of the subtree of component 3 in generation 5.

Step 1 Given the online operating conditions of each system component, $r_q, q \in \{1, 2, \ldots, n\}$, where n is the total number of system components, the corresponding initial P_0^{current} and subsequent P_{1+}^{current} interactions in (6) and (7) are found among the interactions studied offline;

Step 2 Given the initial failed components, column vector V_1 is constructed to represent the failure probability of each component at the beginning of a cascade. $V_1(i) = 1$ denotes component i is part of the initial failures while $V_1(i) = 0$ denotes component i did not fail at the beginning of a cascade. Define $V_1' = e - V_1$ to describe the survival probability of each component where e is an all-ones column vector.

Step 3 Set $G = 2$ (second generation). Given the initial interaction P_0^{current}, the failure probability of each component in the second generation is calculated:

$$V_2(i) = 1 - \prod_{j=1}^{N_1}(1 - V_1(j)P_0^{\text{current}}(i \mid j)), \tag{8}$$

$$V_2'(i) = V_1'(i)(1 - V_2(i)), \tag{9}$$

where V_2 and V_2' store the failure and survival probability of each component in the second generation, respectively. N_1 is the total number of initially failed components in the first generation. Following past literature, we assume that component failure probability follows a geometric distribution [34]. If no element in V_2 is larger than a user defined ε, go to Step 6; otherwise, go to Step 4.

Step 4 Set $G = G + 1$. Using the subsequent interaction P_{1+}^{current}, the failure probability of each component in the Gth generation is calculated. Note that a system component may be predicted as a potential failure in more than one generation of the propagation tree, e.g., component 3 is predicted to fail in generations 2, 3 and 5 in Fig. 3. As a component cannot fail more than once in a cascade, the probability that a component only fails in generation G and not fail in any previous generations is calculated by extracting a subtree from the propagation tree for that component. In this subtree, the component in question only appears in generation G and the potential occurrences of this component in *any* previous generations are removed. For example, for the propagation tree in Fig. 3, to calculate the failure probability of component 3 in generation 5, a subtree is generated and shown in Fig. 4. It can be observed that the component 3 only occurs in generation 5 and the potential failures of component 3 in generations 2 and 3 are removed in the subtree. Given the subtree for each component in generation G, its failure probability in generation G is:

$$V_{\text{G}}(i) = 1 - \prod_{j=1}^{N_{\text{G}-1}^{\text{sub}}}(1 - V_{\text{G}-1}^{\text{sub}}(j)P_{1+}^{\text{current}}(i \mid j)), \tag{10}$$

$$V_{\text{G}}'(i) = V_{\text{G}-1}'(i)(1 - V_{\text{G}}(i)), \tag{11}$$

where V_{G} records the component failure probability at generation G, while V_{G}' stores the survival probability of each component until generation G, i.e., the probability that a component does not fail in the first G generations. $V_{\text{G}-1}^{\text{sub}}(j)$ is the failure probability of component j in generation $G-1$ of the subtree. $N_{\text{G}-1}^{\text{sub}}$ is the total number of parent nodes of component i in generation $G - 1$ of the constructed subtree.

Step 5 If no element in V_G is larger than ε, go to Step 6; otherwise, return to Step 4.

Step 6 Construction of the propagation tree is stopped.

Once all components stop generating child failures, the propagation tree for a given operating condition is constructed. The failure probability of a component in a cascade can be calculated as,

$$V_{\text{final}}(i) = 1 - \prod_{k=1}^{N_{\text{G}}}(1 - V_{\text{k}}(i)) \tag{12}$$

where $V_{\text{final}}(i)$ is the probability of component i fails in this this cascade. N_{G} is the total number of generations after constructing the propagation tree. It can be observed in the construction steps of the propagation tree that the developed method enables calculating the failure probability of a component in different generations, while the interaction model in [27], its improvement in [28] and the event tree model cannot provide this information. This is due to the fact that the interaction model ignores the component failure uncertainty that a component might fail in different generations with different probabilities and only allows components to fail in a single generation. The propagation tree can be used to analyze cascades and identify the most vulnerable areas of a system to cascading failures in different generations. Notice that, when the topology changes due to maintenance, this method can still be applied by updating the initial failure probability of other components that are in service using (8).

D. CASCADE DATA SUFFICIENCY

So far, the dynamic interaction model assumed that cascade data is sufficiently available for constructing the initial and subsequent interaction matrices and ensuring accurate cascade prediction. Intuitively, more cascade data tends to provide more interaction information between components. However, the question of "how much data is sufficient for creating interaction matrices?" is yet to be answered.

Recent research studies [27]–[29] have introduced two criteria to assess the adequacy of cascade data for constructing conventional interaction models. It is shown that with an increase in the number of the cascades in the data, the total number of different interactions between components, i.e., the number of non-zeros in the interaction matrix B increases. However, the total number of different interactions does not significantly change after the number of cascades reaches a threshold. Thus, to obtain most of the interactions between components, the change in the number of non-zeros is used to determine the number of cascades required for building the interaction matrix B. Another lower bound for the number of required cascade data is the number of cascades that can be used to obtain the dominant interactions. By comparing the mismatch of propagation capacity, i.e., the average number of failures in one cascade, between the original cascades and the predicted cascades that are from interaction model, this lower bound of the number of cascades can be determined. The objective of the conventional interaction model is to identify key interactions that largely contribute to the propagation of cascades, while the dynamic interaction model developed in this work is to predict the failure probability of each component. In other words, the entries of the interaction matrix are more important to the dynamic interaction model than the number of non-zeros when deciding on data adequacy for constructing the dynamic interaction model.

The increased number of cascades will change the entries of the dynamic interactions but the change would be limited when the total number of cascades is above a threshold. Here, Frobenius norm is chosen to measure the change in the entries of the interaction matrices and to determine the required number of cascades for building the dynamic interaction model,

$$E^i = \frac{\left\| D^{\text{i}} - D^{\text{i}-1} \right\|_{\text{F}}}{N_{\text{nz}}} = \frac{\sqrt{\sum_{p=1}^{N_{\text{branch}}} \sum_{q=1}^{N_{\text{branch}}} \left| d_{\text{pq}}^{i} - d_{\text{pq}}^{i-1} \right|^2}}{N_{\text{nz}}}, \quad i = 2, \ldots, N_{\text{total}} \tag{13}$$

where D^{i} and $D^{\text{i}-1}$ are the dynamic interaction matrices, and i is the number of cascades. N_{nz} is the number of non-zero elements in D^{i}. $\|\cdot\|_{\text{F}}$ denotes the Frobenius norm, E^{i} indicates the change of the dynamic interaction matrix when the number of cascades is i, d_{pq}^{i} and d_{pq}^{i-1} are the elements in the pth row and qth column of matrix D^{i} and $D^{\text{i}-1}$, respectively, N_{branch} is the number of components, and N_{total} is the total number of cascades.

III. CASE STUDIES

The developed dynamic interaction model is evaluated on two test systems. The first test system is the widely used IEEE 30-bus system, with 30 buses and 41 transmission lines, representing a portion of the American Electric Power system [35]. The second system is a synthetic electric grid case, i.e. Illinois 200-bus system, that is statistically and functionally similar to real-world electric grids [36]. This system has a total of 200 buses and 245 lines. The diagrams of these two test systems are given in Fig. 5 and Fig. 6, respectively. The cascade data is generated from simulations using AC-OPA [7] model. To evaluate the performance of the dynamic interaction model, different performance analysis metrics are used. Also, the developed dynamic interaction model is compared with four baselines: Hines' influence model [25], Qi's interaction model [27], Qi's EM model [28] and dynamic interaction model without differentiating initial and subsequent failures. The goal is to demonstrate the effectiveness of the developed dynamic interaction model in predicting cascades, and highlight its contributions over the existing methodologies.

A. GENERATING CASCADE DATA

To generate cascade data, a simulation based model, i.e. AC OPA model is used. This model has been previously validated in [7], [37]. AC OPA model uses AC optimal power flow (AC OPF) to solve the power flow and determine the operator actions such as load shedding. The detailed steps followed for the AC OPA model used in this paper are as follows:

Step 1 Each load is initialized by multiplying a random number that is uniformly distributed in $[2-\gamma, \gamma,]$ to the nominal value of the load, where γ is the load variability and is set to 1.67. Notice that all loads vary asynchronously in this study. The choice for

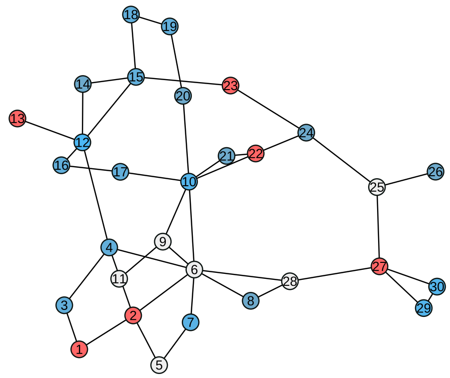


FIGURE 5. Diagram of IEEE 30-bus system that has 30 buses and 41 transmission lines. Red nodes represent generator buses and blue nodes represent load buses.

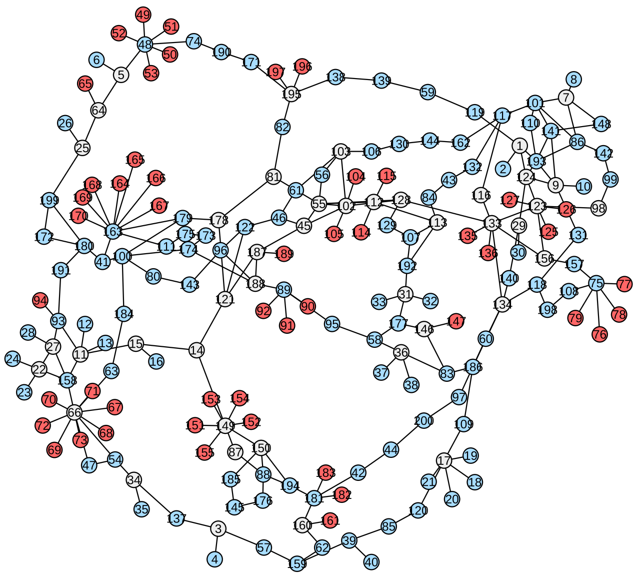


FIGURE 6. Diagram of Illinois 200-bus system that has 200 buses and 245 transmission lines. Red nodes represent generator buses and blue nodes represent load buses.

the load variability γ is inspired by [27], [28] to represent load variations throughout a year. The load and generation profiles of the two test systems can be obtained from [35] and [38];

Step 2 Line power flows are initialized using AC OPF. If AC OPF diverges, loads are shed until a solution is reached. If no solutions are obtained, back to step 1;

Step 3 An initial event is set. It is assumed that each system component fails independently with a failure probability of 0.001, as the initial events are rare in practice. Notice that the initial failure probability can be further improved by analyzing the historical failure data;

Step 4 All the islands within the system are identified. For each island, the supply and demand are balanced by ramping up/down generation or shedding load. After the generation and load in all islands are re-balanced, the AC OPF is calculated. If AC OPF diverges, loads are shed until a solution is reached;

Step 5 If any line flows violate their limits, the overloaded lines are tripped and Step 4 is repeated; otherwise, the simulation will stop.

B. DETERMINING CASCADE DATA ADEQUACY

With the aforementioned simulation-based cascade model, a total of 160,000 and 240,000 cascades are generated for the IEEE 30-bus and Illinois 200-bus systems, respectively. By gradually increasing the number of cascades, the changes in the dynamic interaction matrices, in particular, initial interaction and subsequent interactions, for both test cases are illustrated in Fig. 7 and Fig. 8. Setting the number of bins as two, four different initial interaction matrices that correspond to $P(j|i, r_\text{i} \in [0, 0.5), r_\text{j} \in [0, 0.5))$, $P(j|i, r_\text{i} \in [0, 0.5), r_\text{j} \in [0.5, 1.0))$, $P(j|i, r_\text{i} \in [0.5, 1.0), r_\text{j} \in [0, 0.5))$, $P(j|i, r_\text{i} \in [0.5, 1.0), r_\text{j} \in [0.5, 1.0))$, and two different subsequent interaction matrices $P(j|i, r_\text{j} \in [0, 0.5))$ and $P(j|i, r_\text{j} \in [0.5, 1.0))$ are extracted from the cascade data. It can be observed that the change in dynamic interaction matrices for both the initial and subsequent interactions becomes smaller with an increase in the number of cascades, particularly when the number of cascades is small. In other words, more cascade data provides more interaction information. However, when the number of cascades is above a threshold, the Frobenius norm E is close to zero, which means interactions do not change with more cascade data. The change in the Frobenius norm, determines the number of cascade data required for the IEEE 30-bus and Illinois 200-bus systems to be approximately 140,000 and 200,000, respectively.

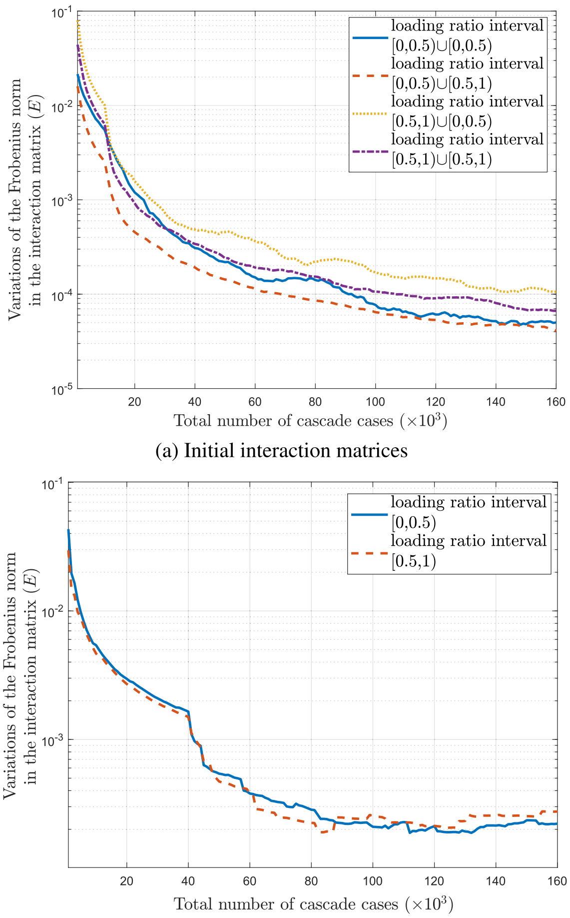
C. PERFORMANCE ANALYSIS

Once the minimum number of required cascade data is determined, the initial and subsequent interactions can be generated and used to predict potential cascades. If a component failure probability in a generation, calculated by the propagation tree, is larger than a threshold ε, this component is regarded as a potential failure in the corresponding generation of a cascade. To evaluate the prediction performance of the developed dynamic interaction model, a set of metrics are used, as discussed later. The performance of the developed method is analyzed in terms of (1) how accurately all failures in a cascade, referred to as *total failures*, are predicted (not considering which generation each failure happened), and (2) how accurately failures in each generation of a cascade, referred to as *temporal failures*, are predicted.

Prediction Evaluation for Total Failures is performed using three metrics: *Jaccard coefficient*, F_1 *score*, and *Precision@K*.

(1) *Jaccard coefficient* is used to measure the similarity between two sets A_{Truth} and $A_{\text{Predicted}}$ [39]:

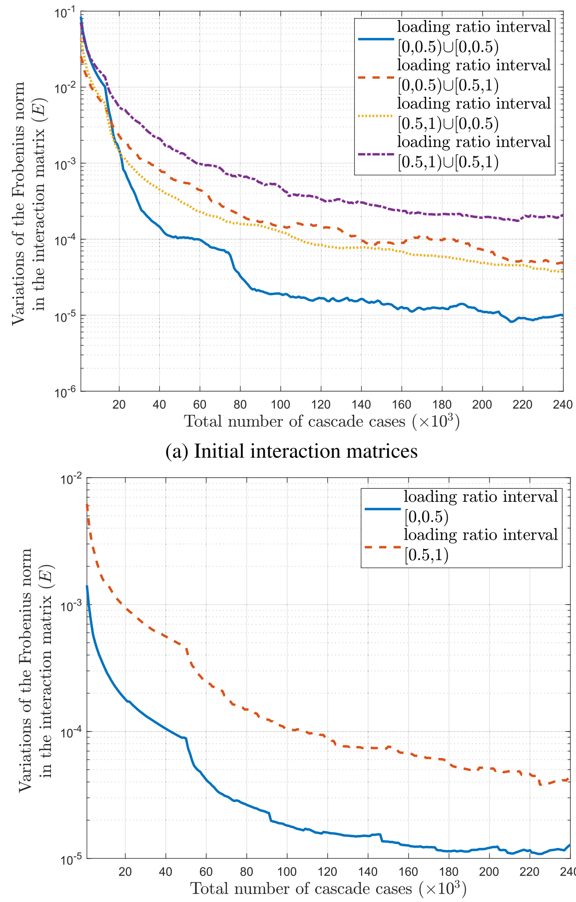
$$J\left(A_{\text{Truth}}, A_{\text{Predicted}}\right) = \frac{|A_{\text{Truth}} \cap A_{\text{Predicted}}|}{|A_{\text{Truth}} \cup A_{\text{Predicted}}|}, \tag{14}$$



(a) Initial interaction matrices

(b) Subsequent interaction matrices

FIGURE 7. Variations of the Frobenius norm calculated for the dynamic interaction matrix for IEEE 30-bus system.



(a) Initial interaction matrices

(b) Subsequent interaction matrices

FIGURE 8. Variations of the Frobenius norm calculated for the dynamic interaction matrix for Illinois 200-bus system.

where A_{Truth} and $A_{\text{Predicted}}$ are the set of truly failed and predicted failed components in a cascade, respectively. A_{Truth} can be obtained from the original cascade data, while $A_{\text{predicted}}$ is determined by comparing the total failure probability of each component, i.e., V_{final} in (12), with a threshold ε. If the total failure probability of a component is above ε, this component is regarded as a potential failure in a cascade. For temporal failures, A_{Truth} can be obtained from the original data in each generation of a cascade, while $A_{\text{predicted}}$ is determined by comparing the failure probability of each component in a generation, V_{G} in (10), with a threshold ε.

(2) F_1 *score* is a weighted average of the *Precision* and *Recall*, which measures the prediction accuracy of the dynamic interaction model [40]. Here, Precision is the fraction of potential failures that indeed fail in a cascade, while Recall is the fraction of failed components that are successfully predicted as impending failures [41].

$$\text{Precision} = \frac{|A_{\text{Truth}} \cap A_{\text{Predicted}}|}{|A_{\text{Predicted}}|}, \tag{15}$$

$$\text{Recall} = \frac{|A_{\text{Truth}} \cap A_{\text{Predicted}}|}{|A_{\text{Truth}}|}, \tag{16}$$

$$F_1 \text{ score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{17}$$

(3) *Precision@K* (Precision at K), inspired by [42], is the fraction of components that have failed among the top K predicted potential failures:

$$\text{Precision@}K = \frac{\left|A_{\text{Truth}} \cap A^{K}_{\text{Predict, Sort}}\right|}{K}, \tag{18}$$

where K is user-defined, and set to the number of components that indeed fail in a cascade. $A^{K}_{\text{Predict, Sort}}$ is the top K predicted failures in the sorted set of potential failures $A_{\text{Predict, Sort}}$, where potential failures are sorted based on their corresponding failure probabilities.

Prediction Evaluation for Temporal Failures is performed using Jaccard coefficient, F_1 score, and two other metrics to assess failure prediction accuracy in each cascade generation:

(1) *Kendall's Tau* (τ) is a correlation coefficient used to measure the association between two ranking methods, where τ is in the range of $[-1, 1]$ [43]. $\tau = 1$ shows

a complete agreement, and $\tau = -1$ shows a complete disagreement between the two ranking methods. $\tau = 0$ means the two ranking methods are independent [43]. Given two rankings (orders) $A : a_1, \ldots, a_n$ and $B : b_1, \ldots, b_n$, consider pair (a_i, b_i) and (a_j, b_j). If $a_i > a_j$ and $b_i > b_j$ or $a_i < a_j$ and $b_i < b_j$, the pair is concordant; If $a_i > a_j$ and $b_i < b_j$ or $a_i < a_j$ and $b_i > b_j$, the pair is discordant; and If $a_i = a_j$ or $b_i = b_j$, the pair is a tie [44]. τ is formulated as,

$$\tau = \frac{n_c - n_d}{\sqrt{[n_{all} - \sum_{i=1}^{t} \frac{t_i(t_i-1)}{2}] \cdot [n_{all} - \sum_{j=1}^{u} \frac{u_j(u_j-1)}{2}]}},$$
$$n_{all} = \frac{n(n-1)}{2}, \tag{19}$$

where n_c and n_d are the number of concordant and discordant pairs, respectively. n is the number of elements in the ranking orders A and B, t and u are the number of different group of ties in rank A and B, and t_i and u_j are the number of elements in the i^{th} and j^{th} group of ties in A and B, respectively. For example, given the ranking order A that has seven elements ($n = 7$) as 1 2 2 3 4 4 4, the number of different group of ties are two ($t = 2$) since the rankings 2 and 4 appear more than once, that is, two elements are ranked with the same order 2 ($t_1 = 2$) and three elements are ranked with the order 4 ($t_2 = 3$).
As components in a generation of a cascade either fail or survive, they cannot be ranked; hence, a heuristic ranking method is used for components that have indeed failed at each generation. In this heuristic method, the rank for failed and survived components are set to 1 and 2, respectively. For those components that are predicted to fail, the rank is assigned based on the corresponding failure probabilities. Hence, Kendall's Tau can be used to evaluate the similarity between the predicted and the real failed components in each generation of a cascade.

(2) *The average precision of the top N generations* is also assessed in this paper. Incorrectly predicted failures in one generation of a cascade will lead to wrong predictions in all subsequent generations. As the initial cascade generations are more influential in determining operator actions, the average precision of the top N generations, $Avg_N^{Precision}$, is used to evaluate the performance of the developed failure prediction methodology.

$$Avg_N^{Precision} = \frac{\sum_{i=1}^{N} \text{Precision}_i}{N} \tag{20}$$

where Precision$_i$ is the precision for the ith generation. Since most cascades stop within 10 generations, N is set to 10.

In addition to the previously collected cascade data for the IEEE 30-bus and the Illinois 200-bus systems, another 50,000 cascades for each test system is generated using the same cascade data generation process to validate the performance of the developed interaction model by using holdout cross validation method [45]. A threshold ε, is used to identify the potential failures in each generation of a cascade.

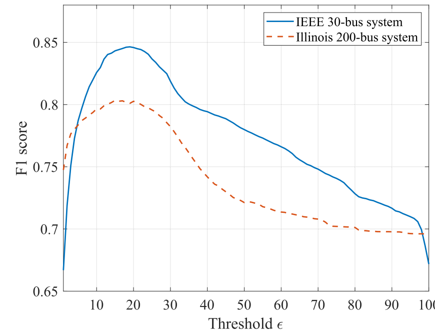


FIGURE 9. F_1 score of prediction accuracy of total failures in IEEE 30-bus and Illinois systems with different $\varepsilon \in [0\ 1]$.

This threshold is varied from zero to one to find the optimal ε that yields the highest prediction accuracy. The F_1 score of the prediction of total failures with different ε in IEEE 30-bus and Illinois 200-bus systems are given in Fig. 9. Here, these values are $\varepsilon = 0.18$ for the IEEE 30-bus system, and $\varepsilon = 0.15$ for the Illinois 200-bus system. We found that the prediction accuracy of the developed interaction model increases with the number of bins, b, that are used to discretize dynamic interactions. However, the prediction accuracy did not change as the number of bins increased beyond a certain value. By looking at the prediction accuracy while varying the numbers of bins from 1 to 100, which is given in Fig. 10, we have set the number of bins for the IEEE 30-bus and Illinois 200-bus systems to 35 and 40, respectively.

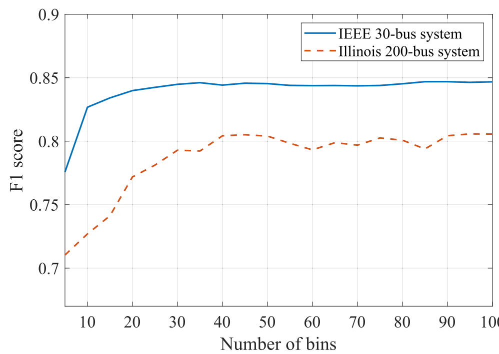


FIGURE 10. F_1 score of prediction accuracy of total failures in IEEE 30-bus and Illinois 200-bus systems with different bins (from 1 to 100).

The prediction performance based on both total failures and temporal failures for two test systems are provided in Table 1 and Table 2. It can be observed that, in most cases, the developed dynamic interaction model significantly outperforms the other four baselines in terms of prediction accuracy for both total failures and temporal failures. The performance results of the developed dynamic interaction model with and without separating the interactions also show that dividing the interactions into initial and subsequent interactions has significant improvement in predicting the two different type of failures. Specifically, for the total failures,

TABLE 1. Performance analysis of the proposed dynamic interaction model for the IEEE 30-bus system.

Metric	Qi's Model [27]	Qi's EM Model [28]	Hines's model [25]	Dynamic interaction model without separating initial and subsequent outages	Developed dynamic interaction model
Total Failures					
Jaccard Coefficient	0.5436	0.5598	0.3257	0.7108	**0.8159**
F_1 score	0.5631	0.5127	0.3344	0.7133	**0.8477**
Precision@K	0.6409	0.6058	0.3943	0.7713	**0.8972**
Temporal Failures					
Jaccard Coefficient	0.4203	0.3965	0.1815	0.5631	**0.7605**
F_1 score	0.4383	0.4058	0.2008	0.5864	**0.7866**
Kendall's tau	0.7888	0.7478	0.7976	0.8122	**0.8726**

TABLE 2. Performance analysis of the proposed dynamic interaction model for the Illinois 200-bus system.

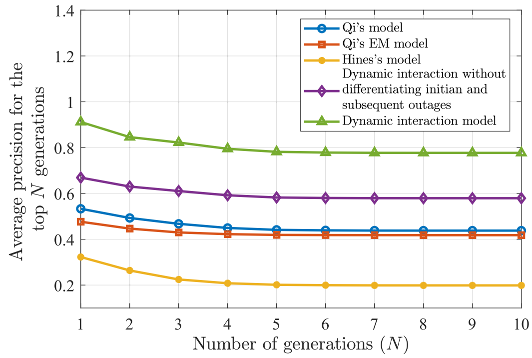
Metric	Qi's Model [27]	Qi's EM Model [28]	Hines's model [25]	Dynamic interaction model without separating initial and subsequent outages	Developed dynamic interaction model
Total Failures					
Jaccard Coefficient	0.6720	0.7126	0.6600	0.5959	**0.7988**
F_1 score	0.6589	0.6776	0.6649	0.5586	**0.8058**
Precision@K	0.7059	0.7690	0.7448	0.6274	**0.8526**
Temporal Failures					
Jaccard Coefficient	0.6196	0.6577	0.6316	0.5219	**0.7585**
F_1 score	0.6220	0.6580	0.6358	0.5273	**0.7711**
Kendall's tau	0.8216	0.8360	0.8504	0.8575	**0.8692**

the Jaccard coefficient of the dynamic interaction model is 0.8159 and 0.7988 for the IEEE 30-bus, and Illinois 200-bus systems, which means the similarity between the predicated failed components and real failed components is 81.59% and 79.88%, respectively. The F_1 score (a weighted average for prediction accuracy) for the two test systems reaches 84.77% and 80.58%, higher than that of the other four baselines. Similarly, the precision@K values show that 89.72% and 85.26% of the top K predicted failures are correctly predicted, where K is equivalent to the number of components that have indeed failed. Low precision means more lines are incorrectly recognized as risky lines, which would cause unnecessary actions such as relay blocking on those lines. These unnecessary actions are costly and may decrease the system reliability. It can be observed from the performance comparison results that the prediction precision is significantly improved by using the developed dynamic interaction model. In addition to the prediction performance of the total failures, the prediction for the temporal failures also reaches a high accuracy, although the accuracy is slightly decreased. The reason for this reduction in accuracy is that a wrong prediction of a failed component in a generation would cause wrong prediction of failed components in all subsequent generations of a cascade, which can be thought of as a prediction noise. With the propagation of cascades, the effect of this noise is amplified and results in reduction in the prediction accuracy. For a better illustration of the prediction accuracy in each generation of a cascade, and to decrease the effect of the noise, the average precision of the top N generations for the two test systems are given in Fig. 11. We observe that the prediction accuracy decreases with the increase in the number of top N generations. Compared to the other four baselines, the developed interaction model provides a more precise prediction of the failed components in a generation, especially for the top four generations. The values for Jaccard coefficient and Kendall's Tau also show a high degree of similarity and strong agreement between the predicted outages and the real outages in each generation.

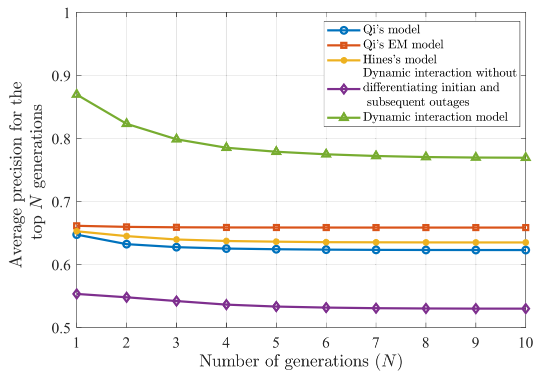
The computation time of the dynamic interaction model is also reported here to assess its suitability in near real-time. The time required to online analysis of 1,000 cascades for IEEE 30-bus and Illinois 200-bus systems using the dynamic interaction model are 8.1043 s and 250.2494 s, respectively, on a computer with an i7-7700 CPU, 4.2GHz core. Compared to the time it takes to analyze the same number of cascades using AC cascading failure model, which is 300.7678 s for the IEEE 30-bus system and 1371.4187 s for the Illinois 200-bus system, the time efficiency of the developed model is a clear improvement. Although the OPF approach in AC cascading failure model is fast if pre-calculated solutions of some specific operation conditions are stored, it is impractical to calculate and store solutions for all possible system operating conditions due to the numerous uncertainties present in today's power grids, particularly from renewable energy resources. It can be observed that the computation time would be increased with the size of the system. In terms of the type of the system, a well-connected network will resist more cascade propagation, as it is more robust. Hence, the number of generations in a cascade in a well-connected power system would be less than other loosely connected systems. As a consequence, the former would have shorter computation time using propagation tree due to the less number of cascade generations that need to be estimated. Furthermore, in order to make the developed dynamic interaction model more suitable for online applications, several strategies are proposed here to speed up the computation time: 1) Instead of evaluating all $N-1-1$ and $N-2$ contingencies, it is suggested to apply the dynamic interaction on the selected credible contingencies that can be obtained from various contingency selection methods [46]–[50]; 2) Only forecast the propagation of the first K_{predict}) generations of a cascade as the initial progress of the failure propagation is more important to system operators to identify potential failures; 3) High performance computing can further reduce the computation time.

D. POTENTIAL FAILURE IDENTIFICATION

Analyzing the performance of different statistical cascading failure models has shown that the developed dynamic interaction model provides more accurate prediction of the total and temporal failures in each generation of a cascade. With the dynamic interaction model, one is able to identify components that have high failure probability in the subsequent failures, particularly at the beginning of a cascade, where the progress of the failures is slow [51]. Therefore, operators can take targeted control actions that reduce the loading ratio of the identified potential components or disable the corresponding relays to stop the tripping of these components. The examples of cascade prediction for the IEEE 30-bus and Illinois 200-bus systems are illustrated in Fig. 12 and Fig. 13. The real sequence of the two cascades that are obtained from the AC-OPA model is given in Table 3.



(a) IEEE 30-bus system



(b) Illinois 200-bus system

FIGURE 11. Average precision with top *N* number of generations.

TABLE 3. Example of two real sequence of failures in two test systems.

	IEEE 30-bus system	Illinois 200-bus system
Generation	Failed components	
1	12, 36	41, 144, 208
2	10, 29, 30, 31, 33, 35	9, 10, 83, 112, 229, 230
3	16, 28, 34, 40, 41	8

It can be observed in Fig. 12 that, given the initial failures in generation one, i.e. lines 12 and 36 are out in IEEE 30-bus system, the obtained failure probabilities of lines 10, 29, 30, 31, 33 and 35 from the dynamic interaction model are close to one, which are in accordance with the real failed components in generation 2 in Table 3. It is also seen that the real failed components, i.e. components 16, 28, 34, 40 and 41, have high failure probability in generation three in Fig. 12. However, line 32 has high failure probability in the generation 3 in Fig. 12 while it does not fail in the real case. Further investigation shows that the power flow of line 32 is close to its limit. Comparing the real failure sequence in Illinois 200-bus system with Fig. 13, the same observation is made: the dynamic interaction model effectively identifies components that are prone to failure in each generation. The only exception is the wrongly predicted (False positive) failure of line 207 in the second generation. Although line 207 does not get overloaded, it is found that the power flow of line 207 is close to its thermal rating. It can be observed from the illustration results that the dynamic interaction model

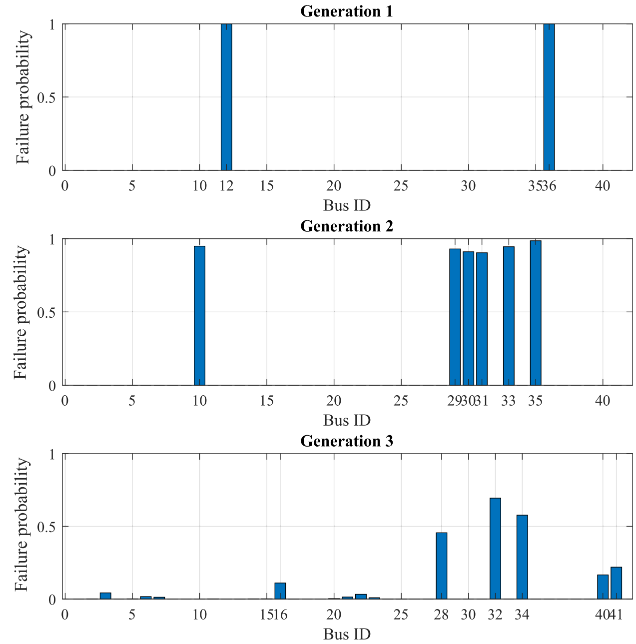


FIGURE 12. Potential failures in each generation of a cascade in IEEE 30-bus system.

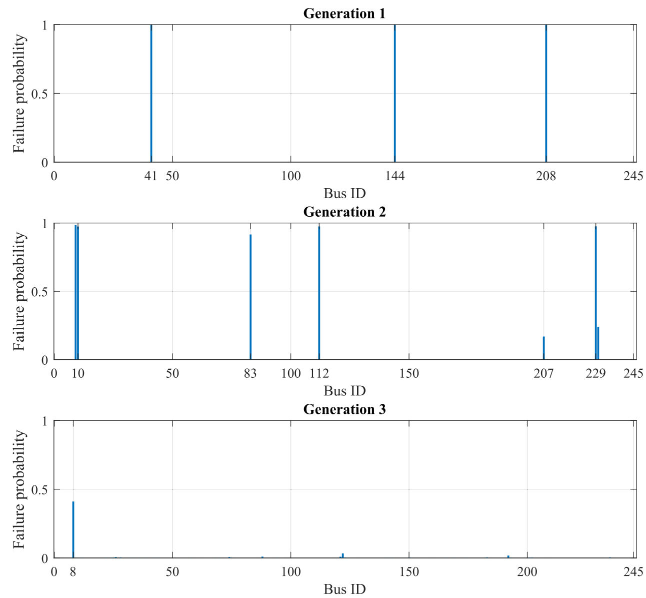


FIGURE 13. Potential failures in each generation of a cascade in Illinois 200-bus system.

provides useful insights into the potential failures in each generation of a cascade, that in turn can be used by the system operators to take effective actions that prevent the propagation of cascades, in a timely manner.

IV. CONCLUSION

This paper presents the *dynamic cascading failure model*, which facilitates prediction of impending cascading failures under a specific power system operating condition. The conventional statistical cascading failure models only capture the general propagation patterns of cascades. On the contrary, the dynamic failure model incorporates system operating conditions into the component failure interactions using a Bayesian approach, and thus enables identifying component failure interactions that correspond to the latest available line power flows. By analyzing simulated or historical cascade data, the interactions between system components are estimated offline, and used in online operations to calculate the failure probability of each system component using the introduced *propagation tree* model. The prediction efficiency of the developed dynamic failure model is evaluated by various metrics, showing a significant improvement over four other state-of-the-art methods in predicting total and temporal failures of a cascade. When an unexpected power system failure occurs, the system operators can promptly identify the most probable component failures with the developed failure model. Hence, proper actions can be taken to prevent further loss of system components. Furthermore, the research in this study can be extended to help determine proper failure mitigation strategies such as protective relay blocking.

Although our method outperforms the traditional simulation-based methods and conventional interaction models in terms of time and accuracy, for larger utility systems, it may need to be further enhanced in terms of computation time, as a sufficiently large amount of scenarios, in the order of millions, need to be simulated. For future study, we aim to use imprecise probability method and Expectation–maximization (EM) to reduce the number of cascade data required to construct the dynamic failure model, such that the developed method can become more suitable for larger utility systems.

REFERENCES

[1] R. Baldick *et al.*, "Initial review of methods for cascading failure analysis in electric power transmission systems IEEE PES CAMS task force on understanding, prediction, mitigation and restoration of cascading failures," in *Proc. IEEE Power Energy Soc. Gen. Meeting-Convers. Del. Electr. Energy 21st Century*, Jul. 2008, pp. 1–8.

[2] H. H. Alhelou, M. Hamedani-Golshan, T. Njenda, and P. Siano, "A survey on power system blackout and cascading events: Research motivations and challenges," *Energies*, vol. 12, no. 4, p. 682, 2019.

[3] *Grid 2030—A National Vision for Electricity's Second 100 Years*, US Department of Energy, Washington, DC, USA, 2003.

[4] H. Guo, C. Zheng, H. H.-C. Iu, and T. Fernando, "A critical review of cascading failure analysis and modeling of power system," *Renew. Sustain. Energy Rev.*, vol. 80, pp. 9–22, Dec. 2017.

[5] I. Dobson, B. Carreras, V. Lynch, and D. Newman, "An initial model for complex dynamics in electric power system blackouts," in *Proc. 34th Annu. Hawaii Int. Conf. Syst. Sci.*, 2001, p. 2017.

[6] S. Mei, F. He, X. Zhang, and D. Xia, "An improved OPA model and the evaluation of blackout risk," *Automat. Electr. Power Syst.*, vol. 32, no. 13, pp. 1–5, 2008.

[7] S. Mei, Y. Ni, G. Wang, and S. Wu, "A study of self-organized criticality of power system under cascading failures based on AC-OPF with voltage stability margin," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1719–1726, Nov. 2008.

[8] D. P. Nedic, I. Dobson, D. S. Kirschen, B. A. Carreras, and V. E. Lynch, "Criticality in a cascading failure blackout model," *Int. J. Elect. Power Energy Syst.*, vol. 28, no. 9, pp. 627–633, Nov. 2006.

[9] J. Chen, J. S. Thorp, and I. Dobson, "Cascading dynamics and mitigation assessment in power system disturbances via a hidden failure model," *Int. J. Elect. Power Energy Syst.*, vol. 27, no. 4, pp. 318–326, May 2005.

[10] P. Crucitti, V. Latora, and M. Marchiori, "A topological analysis of the italian electric power grid," *Phys. A, Stat. Mech. Appl.*, vol. 338, nos. 1–2, pp. 92–97, Jul. 2004.

[11] P. Walley, "Towards a unified theory of imprecise probability," *Int. J. Approx. Reasoning*, vol. 24, nos. 2–3, pp. 125–148, May 2000.

[12] C. D. Brummitt, R. M. D'Souza, and E. A. Leicht, "Suppressing cascades of load in interdependent networks," *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 12, pp. E680–E689, Mar. 2012.

[13] D. J. Watts, "A simple model of global cascades on random networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 9, pp. 5766–5771, Apr. 2002.

[14] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, pp. 1025–1028, Apr. 2010.

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

[16] I. Dobson, B. A. Carreras, and D. E. Newman, "Branching process models for the exponentially increasing portions of cascading failure blackouts," in *Proc. 38th Annu. Hawaii Int. Conf. Syst. Sci.*, 2005, p. 64a.

[17] R. Rocchetta, E. Zio, and E. Patelli, "A power-flow emulator approach for resilience assessment of repairable power grids subject to weather-induced failures and data deficiency," *Appl. Energy*, vol. 210, pp. 339–350, Jan. 2018.

[18] B. Schäfer, D. Witthaut, M. Timme, and V. Latora, "Dynamically induced cascading failures in power grids," *Nature Commun.*, vol. 9, no. 1, pp. 1–13, Dec. 2018.

[19] F. Cadini, G. L. Agliardi, and E. Zio, "A modeling and simulation framework for the reliability/availability assessment of a power transmission grid subject to cascading failures under extreme weather conditions," *Appl. Energy*, vol. 185, pp. 267–279, Jan. 2017.

[20] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, p. 47, 2002.

[21] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Phys. Rep.*, vol. 424, nos. 4–5, pp. 175–308, 2006.

[22] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, no. 6825, pp. 268–276, Mar. 2001.

[23] I. Dobson, B. A. Carreras, D. E. Newman, and J. M. Reynolds-Barredo, "Obtaining statistics of cascading line outages spreading in an electric transmission network from standard utility data," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4831–4841, Nov. 2016.

[24] P. Hines, E. Cotilla-Sanchez, and S. Blumsack, "Do topological models provide good information about electricity infrastructure vulnerability?" *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 20, no. 3, Sep. 2010, Art. no. 033122.

[25] P. Hines, I. Dobson, and P. Rezaei, "Cascading power outages propagate locally in an influence graph that is not the actual grid topology," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 958–967, Mar. 2017.

[26] K. Zhou, I. Dobson, Z. Wang, A. Roitershtein, and A. P. Ghosh, "A Markovian influence graph formed from utility line outage data to mitigate large cascades," 2019, *arXiv:1902.00686*. [Online]. Available: http://arxiv.org/abs/1902.00686

[27] J. Qi, K. Sun, and S. Mei, "An interaction model for simulation and mitigation of cascading failures," *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 804–819, Mar. 2015.

[28] J. Qi, J. Wang, and K. Sun, "Efficient estimation of component interactions for cascading failure analysis by EM algorithm," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3153–3161, May 2018.

[29] W. Ju, K. Sun, and J. Qi, "Multi-layer interaction graph for analysis and mitigation of cascading outages," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 7, no. 2, pp. 239–249, Jun. 2017.

[30] Z. Zhang, S. Huang, Y. Chen, S. Mei, R. Yao, and K. Sun, "An online search method for representative risky fault chains based on reinforcement learning and knowledge transfer," *IEEE Trans. Power Syst.*, early access, Nov. 4, 2019, doi: 10.1109/TPWRS.2019.2951171.

[31] S.-I. Lim, C.-C. Liu, S.-J. Lee, M.-S. Choi, and S.-J. Rim, "Blocking of zone 3 relays to prevent cascaded events," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 747–754, May 2008.

[32] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*. London, U.K.: Chapman & Hall, 2013.

[33] M. Vaiman, K. Bell, Y. Chen, B. Chowdhury, I. Dobson, P. Hines, M. Papic, S. Miller, and P. Zhang, "Risk assessment of cascading outages: Methodologies and challenges," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 631–641, May 2012.

[34] A. N. Philippou, C. Georghiou, and G. N. Philippou, "A generalized geometric distribution and some of its properties," *Statist. Probab. Lett.*, vol. 1, no. 4, pp. 171–175, Jun. 1983.

[35] R. Christie and I. Dabbagchi, "30 bus power flow test case," Univ. Washington, Seattle, WA, USA, Tech. Rep., 1993.

[36] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3258–3265, Jul. 2017.

[37] B. Carreras, J. M. R. Barredo, I. Dobson, and D. E. Newman, "Validating the OPA cascading blackout model on a 19402 bus transmission network with both mesh and tree structures," in *Proc. 52nd Hawaii Int. Conf. Syst. Sci.*, 2019, pp. 1–12.

[38] Electric Grid Test Case Repository. (2017). *Illinois 200-Bus System: Activsg200*. [Online]. Available: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg200

[39] P. Jaccard, "Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines," *Bull. Soc. Vaudoise Sci. Nat.*, vol. 37, pp. 241–272, Jan. 1901.

[40] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," School Inform. Eng., Flinders Univ., Adelaide, SA, Australia, Tech. Rep. SIE-07-001, 2007.

[41] D. L. Olson and D. Delen, *Advanced Data Mining Techniques*. Berlin, Germany: Springer, 2008.

[42] N. Craswell, "Precision at n," in *Encyclopedia of Database Systems*. Boston, MA, USA: Springer, 2009, pp. 2127–2128.

[43] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, nos. 1–2, pp. 81–93, Jun. 1938.

[44] A. Agresti, *Analysis of Ordinal Categorical Data*, vol. 656. Hoboken, NJ, USA: Wiley, 2010.

[45] S. Raschka, "Model evaluation, model selection, and algorithm selection in machine learning," 2018, *arXiv:1811.12808*. [Online]. Available: http://arxiv.org/abs/1811.12808

[46] 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.

[47] P. Kaplunovich and K. Turitsyn, "Fast and reliable screening of N-2 contingencies," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4243–4252, Nov. 2016.

[48] D. Bienstock and A. Verma, "The N-k problem in power grids: New models, formulations, and numerical experiments," *SIAM J. Optim.*, vol. 20, no. 5, pp. 2352–2380, 2010.

[49] V. Donde, V. Lopez, B. Lesieutre, A. Pinar, C. Yang, and J. Meza, "Severe multiple contingency screening in electric power systems," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 406–417, May 2008.

[50] S. Eftekharnejad, "Selection of multiple credible contingencies for real time contingency analysis," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2015, pp. 1–5.

[51] I. H. Mads and R. Almassalkhi. (2014). *Model-Predictive Cascade Mitigation in Electric Power Systems With Storage and Renewables*. [Online]. Available: https://www.ferc.gov/CalendarFiles/20140624104957-T3B%20-%201%20-%20Almassalkhi_Hiskens_Final.pdf

RUI MA (Student Member, IEEE) received the B.S. degree in electrical engineering from Central South University, Changsha, China, in 2014, and the M.S. degree from The Ohio State University, Columbus, OH, USA, in 2015. He is currently pursuing the Ph.D. degree with Syracuse University, Syracuse, NY, USA.

His research interests include phasor measurement unit (PMU) in smart grids, contingency analysis, cascading failure modeling and mitigation, and cyber-security analysis.

SHENGMIN JIN (Student Member, IEEE) received the B.S. degree in information and computing sciences from Fudan University, Shanghai, China, in 2010, and the M.S. degree in computer science from Syracuse University, Syracuse, NY, USA, in 2016, where he is currently pursuing the Ph.D. degree.

From 2010 to 2013, he was a Software Developer on data warehouse with eBay, China. From 2013 to 2014, he worked as an Infrastructure Specialist on big data solutions with the IBM System Technology Group, China. He was a Research Intern with the University of Rochester, NY, USA, in 2015. His work has been published in top data mining conferences, including ICDM and CIKM. His research interests include large scale graph mining, graph representation, and social network analysis.

Mr. Jin was a recipient of the 2016 Syracuse Outstanding Achievement Award in graduate study at Syracuse University.

SARA EFTEKHARNEJAD (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the University of Tehran, Tehran, Iran, in 2006, the M.Sc. degree from West Virginia University, Morgantown, WV, USA, in 2008, and the Ph.D. degree in electrical engineering from Arizona State University, Tempe, AZ, USA, in 2012.

From 2012 to 2014, she worked as a Transmission Planning Engineer with Tucson Electric Power Company, Tucson, AZ, USA. From 2014 to 2015, she was an Assistant Professor with the University of Idaho. Since 2015, she has been an Assistant Professor with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, USA.

REZA ZAFARANI (Member, IEEE) received the Ph.D. degree in computer science from Arizona State University, Tempe, AZ, USA, in 2015.

He is currently an Assistant Professor with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, USA. His research emphasis has been on addressing challenges in large-scale data analytics to enhance the scientific discovery process from big data. These challenges include evaluation without ground truth, fast identification of relevant information in massive datasets, learning with limited information, user behavior analysis and modeling at scale, and information integration and modeling across multiple data sources. His research has been published at major academic venues and highlighted in various scientific outlets. He is the Principal Author of *Social Media Mining: An Introduction*, a textbook by Cambridge University Press. His research interests are in data mining, machine learning, social media mining, and social network analysis.

Dr. Zafarani is the winner of the President's Award for Innovation and Outstanding Teaching Award at Arizona State University. He is an Associate Editor of *SIGKDD Explorations* and *Frontiers in Communication*.

WOLF PETER JEAN PHILIPPE (Student Member, IEEE) received the bachelor's degree in electromechanical engineering from the Faculty of Sciences, State University of Haiti (FDS-UEH), Haiti, in 2012, and the M.Sc. degree in electrical engineering from Syracuse University, NY, USA, in 2015, where he is currently pursuing the Ph.D. degree.

His main research interests include demand response, modeling of wind power variation, and power system operation and control under high wind power penetration.