Assessing Risk from Cascading Blackouts Given Correlated Component Failures

Laurence A. Clarfeld, Margaret J. Eppstein, Paul D.H. Hines and Eric M. Hernandez
University of Vermont, Burlington, VT, USA

Abstract—Despite the infrequent occurrence of cascading power failures, their large sizes and enormous social costs mean that they contribute substantially to the overall risk to society from power failures in the grid. Therefore it is important to accurately understand the risk associated with such events. A cascading event may be triggered by a small subset of \( k \) components failing simultaneously or in rapid succession. While most prior work, including our own work into an efficient “Random Chemistry” method for risk analysis, has assumed that components fail independently, this paper proposes a method for deriving correlated outage probabilities such that pairs of branches that are proximate in space are more likely to fail together than distant ones. Combining Random Chemistry risk analysis with this approach to correlated outage probabilities shows that overall blackout risk can greatly increase with even small amounts of correlation. Results from the 2383-bus Polish test case under various load levels illustrate the substantial impact that correlation has on blackout risk.

Index Terms—blackout risk, cascading failure, cascading outage, correlated outages, Random Chemistry

I. INTRODUCTION

A cascading power failure occurs when a small number of components in a power grid fail, setting off a chain reaction of subsequent component failures that can lead to large blackouts. Cascading power failures are rare events, but their vast size means they pose a significant risk to power grids [1]–[3]. Reliability regulations require that power systems be operated to be robust to single component failures (\( N - 1 \) security) and increasingly require that grid operators make plans to ensure \( N - k \) security [4]. There is, however, no guarantee that sets of two or more components failing together will not cause a cascade. Sets of \( k \) simultaneous outages are typically referred to as \( N - k \) contingencies. Furthermore, mechanisms such as “hidden failures” can exacerbate the risk and impact of cascades [5]–[7]. In this paper, we consider only branch outages. We refer to sets of \( k \) branch outages that initiate a cascading failure as \( N - k \) malignancies whereas sets of \( k \) branch outages that do not cause a cascade are referred to as benign contingencies.

The combinatorial search space of \( N - k \) contingencies makes it difficult to estimate risk in a computationally tractable manner. A number of existing papers propose methods for quantifying the risk of cascading failure [8]–[13]. A limitation of most prior approaches (including our own) is the assumption that branch outages are independent events [1], [11], [12], [14], [15]. In reality, branch failures are unlikely to be independent when a common cause is responsible for the outages. For example, damage caused by weather-related disturbances may be spatially correlated [16]. Protection system failures can sometimes cause multiple outages within a small geographic region [17]. Similarly, terrorist attacks may be spatially localized. This type of geographical correlation was handled in [18] by assuming 100% correlation of outages within a fixed radius. In [16] spatial correlation was achieved by probabilistically determining failure rates of lines adjacent to initial failures according to a Poisson process. In [19], a random field with spatial autocorrelation was used in a cascade model to assess risk from common-cause events. Correlation between outages can also be associated with non-spatial attributes such as component age [20].

Another way to correlate component failures is through copula analysis. Copulas have been applied in many fields, such as finance [21], neuroscience [22], and climate research [23]. Within the realm of power systems, copulas are a popular tool for uncertainty analysis, such as in [24]. In [25], Li suggests copulas as a useful way to incorporate correlation between random variables in power systems risk analysis. Here, we will present the use of a Gaussian copula and show that incorporating even modest levels of correlation can greatly increase the risk associated with cascading failures.

A. Finding Triggering Events with “Random Chemistry”

In [26] we introduced an efficient \( O(\log N) \) stochastic set size reduction algorithm referred to as “Random Chemistry” (RC) for identifying small minimal sets of initiating events that trigger some outcome of interest. In [27] we applied RC to the problem of identifying minimal \( N - k \) malignancies that lead to cascading failures in power grids (Fig. 1), where \( N \) is the total number of branches in the power system. In [11], [12], [15] we showed that RC can be used to efficiently estimate the system-wide risk of large cascading failures in power grids. A comparison of risk estimation to Monte Carlo (MC)
B. Estimating Risk

For any given set of branches $\omega$ that cause a cascading failure, risk can be calculated as:

$$R_{\omega} = p_{\omega} s_{\omega}$$

(1)

where $p_{\omega}$ is the joint probability of the branches in $\omega$ failing and $s_{\omega}$ is the size of the resultant blackout. Note that $p_{\omega}$ is itself a function of $p_i$, the independent outage probability for each branch $i$, and any effect of correlation among branch outage probabilities. In this paper, $s_{\omega}$ is quantified as the total power (MW) unserved due to load shedding. The total risk posed to the system by $N - k$ malignancies, for a given $k$, is then:

$$R_k = \sum_{\omega \in \Omega_k} R_{\omega}$$

(2)

where $\Omega_k$ is the complete set of all $N - k$ malignancies for each $k \geq 2$. The overall risk $R$ of cascading failure due to all $k \leq k_{max}$ is thus:

$$R = \sum_{k \in \{2, k_{max}\}} R_k$$

(3)

The RC (or any other) algorithm is unlikely to find the entire set $\Omega_k$ for $k > 2$ on realistically sized systems, because of the computationally intractable sizes of these sets. However, various approaches have been used to estimate the number of $N - 3$ malignancies $|\Omega_3|$ from changes in the rate of identification of new unique $N - 3$ malignancies. This can then be used to project system risk for $k_{max} = 3$ from the identified sets of $N - 2$ and $N - 3$ malignancies [15], [27].
Using (8) to find each element of the covariance matrix \( C \), we can then use the probability density function of the multivariate normal distribution (9) to form our copula.

\[
f(x) = \frac{1}{\sqrt{(2\pi)^k |C|}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu)^\top C^{-1} (\mathbf{x} - \mu) \right\} \tag{9}
\]

Using our copula function, \( F_X(0) \) will represent the joint probability of system failure \( \Pr(\mathbf{X} \leq 0) \), when all \( k \) components fail together. To calculate \( F_X(0) \) we integrate over the region in the joint distribution that represents failure of all system components.

\[
F_X(x) = \int_0^0 \cdots \int_{-\infty}^{0} f(x_1, x_2, \ldots, x_k) \, dx_1 \, dx_2 \ldots \, dx_k
\]

The multiple-integral in (10) represents the generalized solution for arbitrary \( k \). In this work, we consider only \( k = 2 \) and solve the resultant double-integral numerically using the vectorized adaptive quadrature method [31].

### III. Case Study

We extend our previous work assessing cascading failure risk [11], [12], which also used the Polish test case at peak winter load [28], to account for spatially correlated outages. As previously noted, in this proof-of-concept study we only consider \( N - 2 \) malignancies and assume Gaussian copulas. However, the approach is readily generalizable to greater values of \( k \) (assuming \( |\Omega_k| \) can be estimated) and/or alternative distribution functions.

Simulations were conducted using the 2383-bus, 2896-branch Polish power system, at the 1999 peak winter load, which is available via the MATPOWER simulation package [28]. As described in [11], [12], we made several modifications to this test case, including an increase in line limits by a factor of 1.05 above the pre-contingency line flows that occur when the system is at 1.10 times actual load. This change was made to ensure that this “base case” is \( N - 1 \) secure. We then examined loads that were 55\% to 115\% of the base case, to assess how risk changes under varying load conditions.

The true spatial locations of branches and buses are not publicly available for this test case, so hypothetical locations were inferred based on a graph layout of the grid topology, assuming branches are straight lines between buses (Fig. 4).
as described in [11], [12], we assigned branch failure rates randomly from a normal distribution with the same mean and variance as those provided by the RTS-96 test case [32], since these rates were unavailable for the Polish grid.

A. Distance Metric

The appropriate definition of “distance” may vary depending on the type of common cause threatening the system. Without loss of generality, we employ a proximity-based metric that assumes branches are straight lines. Consider branches $U$ with endpoints $(u_1, u_2)$ and $V$ with endpoints $(v_1, v_2)$. Let the distance from $U$ to $V$ be defined as

$$\text{Dist}(U, V) = \frac{\sum_{i=1}^{2} d(u_i, V) + \sum_{i=1}^{2} d(v_i, U)}{2} \quad (11)$$

where $d(u_i, V)$ is the minimum euclidean distance from the point $u_i$ to the line segment $V = (v_1, v_2)$, as illustrated in Fig. 5.

![Figure 5. Visual example for calculating the distance between branches $U$ and $V$ with endpoints $(u_1, u_2)$ and $(v_1, v_2)$, respectively.](image)

In this formulation, it is worth noting that $d(u_i, V) \neq d(v_i, U)$. This makes sense when considering branches of different lengths. For example, consider branches $A$ and $B$ in Fig. 6. All of $B$’s span overlaps with $A$ while only a portion of $A$’s span overlaps with $B$, so it follows that $B$ is in some sense closer to $A$ than $A$ is to $B$. This asymmetry is handled by averaging $d(u_i, V)$ and $d(v_i, U)$ in (11). This distance metric conforms with what would be intuitively expected when considering spatially correlated damage, as seen in Fig. 6, where $\text{Dist}(A, B) > \text{Dist}(C, D)$ and $\text{Dist}(E, F) > \text{Dist}(G, H)$.

![Figure 6. Branch pairs used for pairwise distance examples described in the text.](image)

Using this metric, a pair of branches will have distance 0 only if they are parallel branches between the same buses. This definition of distance can be extended to larger subsets of branches by taking the average of the pairwise distances:

$$\text{Dist}(U_1, U_2, \ldots, U_k) = \frac{2 \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \text{Dist}(U_i, U_j)}{k(k-1)} \quad (12)$$

The distribution of pairwise branch distances in the 2896-branch Polish grid, according to (11) and assuming the topology shown in Fig. 4, is shown in Fig. 7. Branch pairs that cause cascading failures are relatively rare, with only 0.013% of branch pairs comprising $N-2$ malignancies. While there is a significant relationship between branch distance and blackout size ($p < 0.01$), the amount of difference explained by branch distance is very low ($R^2 = 0.019$). However malignant pairs tend to be much closer to each other than benign pairs (Fig. 8), as supported by a two-sample Kolmogorov-Smirnov test on the two distributions ($p \ll 0.01$). This tendency for branch pairs in $N-2$ malignancies to be close together will exacerbate the impact of spatial correlation to our cascading failure risk calculations.

![Figure 7. Distribution of pairwise branch distances in the Polish grid using the proposed proximity-based distance method, in arbitrary units.](image)

B. Results

The total system risk contributed by $N-2$ malignancies was calculated for spatially correlated branch outages across a range of scenarios. Risk was calculated for varying load levels from 55%-115% of 2004 peak winter load in the adjusted Polish grid, for all combinations of $L \in \{0, 500, 1000, 1500, 2000\}$ and $\rho_o \in \{0, 0.05, 0.10, 0.15\}$.

Changes in the total system risk as a function of load at $L = 2000$ (the longest characteristic correlation length tested), over the different values tested for $\rho_o$, are shown in Fig. 9. As noted in [12], risk varies non-monotonically with load, in part due to variations in the proximity of generation to demand that result from optimal power flow dispatch at different load levels. With this characteristic length, even at $\rho_o = 0.05$, risk in the correlated case from load levels 98%-111% surpass the maximum risk seen in the uncorrelated case at any load.
level tested (up to 115%). The highest overall system risk found occurred at 105% load with $L = 2000$ and $\rho_o = 0.15$, where risk increased 225% over the uncorrelated estimate. The greatest relative increase in risk, as a function of $\rho_o$, occurred at the lowest load levels, where overall system risk was lowest. However, the greatest absolute increases in risk, as a function of $\rho_o$, occurred at load levels of 95%-112%, where there were the most $N - 2$ malignancies. For a given load level, risk increases faster than linearly as $\rho_o$ increases (Fig. 10).

Just as $\rho_o$ can influence risk, so too can the characteristic length ($L$), as shown in Fig. 11. Results are included from load levels 80%-115% for $\rho_o = 0.15$. As expected, increasing $L$ for a given $\rho_o$ increases risk. However, in this case the rate of increase is non-monotonic, with the largest increases at intermediate values of $L$ (Fig. 10), because the effect of increasing $L$ diminishes as $L$ approaches the diameter of the grid topography.

It is also informative to investigate the degree to which each branch contributes to overall risk as a function of spatial correlation. Given branch $i$ with independent outage probability of $p_i$, we can measure $i$’s contribution to total risk posed by $N - k$ malignancies by finding the sensitivity, $S_k(i)$ of risk to $p_i$. As discussed in [11], [12], this equates to a partial derivative of risk with respect to $p_i$. Here, we estimate these sensitivities using a finite difference approximation:

$$S_k(i) = \frac{\partial R_k(p_i)}{\partial p_i} \approx \frac{R_k(p_i + \Delta p_i) - R_k(p_i)}{\Delta p_i}$$

(13)

where $R_k(p_i + \Delta p_i)$ is the total risk of the system posed by $N - k$ malignancies when $p_i$ is increased by a small amount $\Delta p_i$. In our calculations, we used $\Delta p_i = 10^{-15}$; empirical tests showed that further decreasing $\Delta p_i$ did not substantially change the results. Branch sensitivities were calculated at the 100% load level with $\rho_o = 0.15$ and $L = 500$. Branch sensitivities in the correlated case described above are approximately 1.4 times that of the uncorrelated case (Fig. 12). The overall relative order of branch sensitivity is largely, but not entirely, preserved (Spearman’s rank correlation $r_s = 0.947$). For example, if we look just at the ten most sensitive branches (i.e., those of greatest concern) there are notable changes in the relative ordering of branch sensitivity between the uncorrelated and correlated case ($r_s = 0.758$).

IV. DISCUSSION

A number of recent papers on cascading failure risk assume that branch outages are statistically independent events. This assumption neglects the possibility of common cause failures such as relay failures, weather events, or terrorist attacks. This paper presents a systematic method to account for spatial correlations among branch outages. The copula-based method
described and demonstrated here is general in that it can be tailored to the details of a specific power system and disturbance category. Parametric choices include the correlation function and associated constants, the distance metric, and the distribution of the copula function.

The application of the method to a large power systems test case shows that even small correlations between component failures can lead to significant increases in system risk posed by \( N - k \) malignancies. This increase in risk is exacerbated by the fact that branch pairs that are close together in this system are more likely to cause cascading failures than are branch pairs that are farther apart.

Prior studies have shown that, in the spatially uncorrelated case, the sensitivity of risk to individual branch outages exhibits a very heavy-tailed distribution, with a few branches contributing disproportionately to risk [12]. Adding spatial correlation to branch outages magnifies this disparity in the relative contributions of each branch to the overall system risk. Furthermore, the relative sensitivity of risk to different branch outages can differ between the uncorrelated and correlated cases, especially in the branches with the highest sensitivities. These observations may have important implications for proposed strategies to mitigate risk by reducing the flow on the most sensitive branches [11].

The results presented in this paper suggest that practical approaches to \( N - k \) cascading failure risk is possible. Reliability regulations (e.g., NERC TOP-004-1.R3) increasingly require that transmission system operators study and protect their systems against the risk of cascading failures triggered by \( N - k \) outages. The method presented here has sufficient computational efficiency to be useful in operations planning tools that quantify the risk of \( N - k \) events and quickly identify important sources of that risk. A key element needed to make this sort of tool practical is better data about the probability of transmission branch outages and the ways in which different types of common causes impact those probabilities. Some data that would be helpful for tuning this model are available to industry through systems such as the NERC TADS database, but these data are not typically available for research.

Future work will study the impact of parametric choices and design details on risk in a variety of test cases, including those with more accurate geographical data (e.g., [33]), and will apply a more sophisticated AC cascading failure simulator. In addition, we will extend this work to analyze spatially correlated \( N - k \) malignancies for \( k > 2 \). This will yield insights as to whether spatial correlation increases or decreases the relative importance of higher-order \( N - k \) malignancies on risk, with important practical implications for methods designed to estimate the risk of cascading failures.

REFERENCES
