IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 20, 2021, accepted April 28, 2021, date of publication April 30, 2021, date of current version May 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3076892

Universal Statistics of Redistribution Factors and Large Scale Cascades in Power Grids

FRANZ KAISER AND DIRK WITTHAUT
Forschungszentrum Jülich, Institute for Energy and Climate Research–Systems Analysis and Technology Evaluation (IEK-STE), 52428 Jülich, Germany
Institute for Theoretical Physics, University of Cologne, 50937 Köln, Germany

Corresponding author: Franz Kaiser (f.kaiser@fz-juelich.de)

This work was supported in part by the German Federal Ministry of Education and Research [Bundesministerium für Bildung und Forschung (BMBF)] under Grant 03EK3055B, and in part by the Helmholtz Association through the joint initiative Energy System 2050–A Contribution of the Research Field Energy and through the Grant VH-NG-1025.

ABSTRACT Cascades of failures are among the biggest threats to supply networks such as power grids: An initially failing element may trigger the failure of other elements, thereby eventually causing the entire network to collapse. Here, we analyse the statistics of Line Outage Distribution Factors (LODFs), which describe the rerouting of electric power flows after a line failure. In particular, we demonstrate that absolute LODFs are approximately log-normally distributed throughout network topologies. We then illustrate that this log-normal distribution of redistribution factors results in a heavy tailed distribution of outage sizes in a simplified, stochastic cascade model over a certain range of parameters. This cascade model extends previous stochastic cascade models by adding more realistic redistribution mechanisms as well as including more realistic initial trigger events. Our results demonstrate that the statistics of redistribution factors is a fundamental trait throughout different networks and presents a possible explanation for the vast occurrence of heavy tailed distributions in real-world reanalyses of power outage sizes.

INDEX TERMS Transmission lines, network theory (graphs), graph theory, cascading failures, transmission line outages, power grids.

I. INTRODUCTION

In our daily lives, we depend on a reliable supply with electrical power. Large scale power outages can have a catastrophic impact on society, economy and other infrastructure networks as recent examples demonstrate [1], [2]. Remarkably, empirical reanalyses of historic power grid blackouts have revealed the scale-free nature of outage sizes: large scale outages are not rare, but the size of outage sizes decays algebraically [3], [4]. The reason for this scaling is still not fully understood, but different possible explanations have been put forward [5]–[8].

Blackouts are in most cases initiated by the failure of a single or only very few transmission or generation elements which cause the failure of other elements and so forth – eventually leading to a cascade of failures where a large part of the system breaks down [9]. A single step in a cascade is essentially governed by two variables: The initial flows in the network and the flow rerouting in the network after a failure. The latter may be compactly summarised in terms of the Line Outage Distribution Factors (LODFs) which arise from a linearisation of power flows and relate the initial flows to the flow changes after a failure [10]. Remarkably, LODFs are a purely topological property of a network, i.e. they do not depend on the power injections. Thus, a key indicator of a given network's resilience is their distribution. Previous work on spatial aspects of flow rerouting has mainly focused on a microscopic perspective on link failures, studying different properties of individual failures such as the distribution of flow changes after line outages [11] with a particular focus on the decay with distance [12]–[16]. Here, we adopt a statistical perspective on the distribution of LODFs for an entire network which yields a structural indicator of a complex network's resilience with respect to perturbations.

The access to real-world power grid data such as network topologies is limited due to the sensitive nature of the information – power grids are considered to be a critical infrastructure. However, recent efforts increase the availability of openly available power grid datasets that typically rely on OpenStreetMap [17]–[19]. Different synthetic power grids that are based on real-world grids have been designed for power flow studies [20]–[24]. In addition to that, different algorithms have been developed to generate synthetic networks that display the main topological properties of real

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

world power grids [25]–[29]. Besides, several studies have addressed the statistical properties of real world transmission grids and the corresponding transmission lines [30].

In this article, we analyse the distributions of LODFs for various real-world and synthetic grids systematically. To the best of our knowledge, this is the first systematic analysis of the distribution of LODFs for different real-world and synthetic transmission grids. In fact, the distribution of LODFs is purely based on the network topology and may thus be considered as a network observable, similarly to the degree distribution or betweenness measures that have been considered in previous analyses of power grids [26], [31], [32].

II. LINEAR FLOW NETWORKS AND LINE OUTAGE DISTRIBUTION FACTORS

In most cases, cascades of failures are well-described by a linearised approach to the power flow equations known as the DC approximation. Here, we briefly review the mathematical aspects and the derivation of LODFS using a more general language that applies to power grids as well as to other types of networks to facilitate a translation of our results.

A. THEORY OF LINEAR FLOW NETWORKS

Consider a linear flow network on a simple, connected graph $G(E, V)$ with $\mathcal{M} = |E|$ edges and $\mathcal{N} = |V|$ vertices. Assume that each edge $\ell = (j, k)$ in the graph is assigned a weight $b_\ell \in \mathbb{R}$ and each node has a potential $\vartheta_n \in \mathbb{R}, n \in \{1, \ldots, \mathcal{N}\}$. In a linear flow network, the flow $F_\ell \in \mathbb{R}, \ell \in \{1, \ldots, \mathcal{M}\}$, on an edge $\ell = (j, k)$ connecting nodes $j, k \in V$ scales linearly with the potential drop along the line such that [14]

$$F_\ell = b_\ell \cdot (\vartheta_j - \vartheta_k). \tag{1}$$

Next, we assign an orientation to each edge $\ell = (j, k)$ in the graph and say that the edge is oriented from node j to node k such that $F_\ell > 0$ is a flow from node j to node k and a negative sign indicates a flow in the opposite direction. This setup applies for example to power transmission grids [14], [34], where F_ℓ is the flow of real power on a transmission line ℓ, ϑ_n denotes the nodal voltage phase angle and b_ℓ is the line susceptance. This corresponds to the so-called 'DC approximation' of AC power flows that typically offers a good description of the power flows if lines are lossless and not too heavily loaded [34]. An equivalent description is also used for hydraulic and vascular networks [35], where F_ℓ is the flow of water or nutrients, ϑ_n is the local pressure and b_ℓ the edge's capacity.

Now assume that each node m has an in- or outflow p_m. Then the edge flows are related to the inflows by Kirchhoff's current law

$$p_m = \sum_{\ell \in \Gamma(m)} F_\ell. \tag{2}$$

Here, $\Gamma(m) \subset E(G)$ is the set of all edges connected to vertex m with each edge sorted according to its orientation. Combining Eqs. (1) and (2), we arrive at the following set of equations

$$p_m = \sum_{\ell \in \Gamma(m)} b_\ell \cdot (\vartheta_m - \vartheta_k), \tag{3}$$

where the sum runs again over all edges $\ell = (m, k)$ whose start point or terminal end is node m. As a next step, we can define the graph Laplacian $\boldsymbol{L} \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}}$ that encodes the topology of the graph in a compact form and is defined by its entries as follows [36]

$$L_{jk} = \begin{cases} -b_\ell & \text{if } \ell = (j, k) \in E(G) \\ \sum_{m \in \Gamma(j)} b_m & \text{if } j = k \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

We will see in the following that this matrix is crucial to describe link failures in linear flow networks. To write the above set of equations more compactly, we define a vector of potentials $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_\mathcal{N})^\top \in \mathbb{R}^N$ and a vector of in- and outflows $\boldsymbol{p} = (p_1, \ldots, p_\mathcal{N})^\top \in \mathbb{R}^N$ to write Equation (3) compactly [14], [37]

$$\boldsymbol{L\vartheta} = \boldsymbol{p}. \tag{5}$$

The nodal potentials are thus subject to a Poisson-type equations. To solve for the vector of potentials, this equation needs to be inverted. However, the Laplacian always has a vanishing eigenvalue $\lambda_1 = 0$ and is thus not invertible. This problem is typically overcome by making use of the matrix's *Moore-Penrose pseudoinverse* $\boldsymbol{L}^\dagger$ which has properties similar to the actual matrix inverse [38], [39].

B. SINGLE LINK FAILURES IN LINEAR FLOW NETWORKS

Assume that a single link e fails that carries the initial flow $F_e^{(0)}$. Then the flow change ΔF_ℓ on another link ℓ can be calculated as follows [10], [14]

$$\Delta F_\ell = \mathrm{LODF}_{\ell,e} F_e^{(0)}.$$

The factor $\mathrm{LODF}_{\ell,e}$ connecting the initial flows and the flow changes is known as *Line Outage Distribution Factor* (LODF) and measures the change in flow on a link ℓ when a link e fails. Thus, the flow F_ℓ on a link ℓ after the failure may be calculated as

$$F_\ell = F_\ell^{(0)} + \Delta F_\ell = F_\ell^{(0)} + \mathrm{LODF}_{\ell,e} F_e^{(0)}. \tag{6}$$

Summarising the LODF for all possible failing links e and all possible monitoring links ℓ, we can define an LODF matrix, $\mathrm{LODF} \in \mathbb{R}^{\mathcal{M} \times \mathcal{M}}$. Its entries may then be expressed in a purely topological manner [10], [14]

$$\mathrm{LODF}_{\ell,e} = b_\ell \frac{\boldsymbol{q}_\ell^t \boldsymbol{L}^\dagger \boldsymbol{q}_e}{1 - b_\ell \boldsymbol{q}_e^t \boldsymbol{L}^\dagger \boldsymbol{q}_e}. \tag{7}$$

Here, $e = (r, s) \in E(G)$ is an edge, $\boldsymbol{q}_e \in \mathbb{R}^\mathcal{N}$ is a vector with entry one at position r and entry minus one at position s and t denotes the transposed vector. The LODF assumes values between minus one and one, $\mathrm{LODF}_{\ell,e} \in [-1, 1]$, i.e. only the

amount of flow that was present initially may be redistributed. The diagonal elements are typically set to minus one for consistency, i.e. $\mathrm{LODF}_{\ell,\ell} = -1, \forall \ell \in E(G)$. Furthermore, there are several cases where the LODF vanishes, for example if two parts of the network are only connected via a bridge [14], [40] or via a network isolator [37], [41].

III. LOGARITHMIC LODFs ARE APPROXIMATELY NORMALLY DISTRIBUTED

In this section, we analyse the distribution of LODFs for real-world and synthetic power grids in detail. We use the term distribution synonymous to the probability density function here and in the following. Since the absolute LODFs are bounded by unity, they may be naturally studied on a logarithmic scale. For this analysis, we neglect the cases where the LODF vanishes since these are typically rare in large networks. We also do not consider the diagonal elements $\mathrm{LODF}_{\ell,\ell} = -1$. The distribution of LODFs is mainly governed by two factors: Firstly, the distribution of edge weights b which we denote by P_B in the following. Secondly, it is governed the distribution of entries of the Laplacian matrix's Moore-Penrose pseudoinverse $\boldsymbol{L}^\dagger$ which we denote by $P_{L^\dagger}$. Importantly, the former distribution is in some sense incorporated into the latter one since the off-diagonal elements of the Laplacian matrix are again (summed) elements from the distribution of weights P_B.

The study of the elements of random matrices has led to the development of random matrix theory. Typically, the distribution of these elements is analysed using its eigenvalues [42]. On the other hand, research has addressed the spectra of complex networks as encoded in the graph's adjacency matrix or its Laplacian matrix [43]–[45]. The Moore-Penrose pseudoinverse $\boldsymbol{L}^\dagger$ - as the actual inverse - has eigenvalues inverse to the eigenvalues of the Laplacian $\boldsymbol{L}$ except for the zero eigenvalue. Thus, diagonalizing both matrices using the eigenvalues $\lambda_1 = 0, \lambda_2, \ldots \lambda_\mathcal{N}$ ordered by magnitude and corresponding eigenvectors $\vec{v}_1 = \vec{1}/\sqrt{\mathcal{N}}, \vec{v}_2, \ldots, \vec{v}_\mathcal{N}$ of the Laplacian matrix, we may write [39]

$$\boldsymbol{L} = (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_\mathcal{N}) \begin{pmatrix} 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \lambda_\mathcal{N} \end{pmatrix} \begin{pmatrix} \vec{v}_1^\top \\ \vec{v}_2^\top \\ \ldots \\ \vec{v}_\mathcal{N}^\top \end{pmatrix}$$

$$\Rightarrow \boldsymbol{L}^\dagger = (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_\mathcal{N}) \begin{pmatrix} 0 & 0 & \ldots & 0 \\ 0 & \lambda_2^{-1} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \lambda_\mathcal{N}^{-1} \end{pmatrix} \begin{pmatrix} \vec{v}_1^\top \\ \vec{v}_2^\top \\ \ldots \\ \vec{v}_\mathcal{N}^\top \end{pmatrix}.$$

Understanding the spectrum of the graph Laplacian and thus the topology of the underlying graph is key to understanding the spectrum of the pseudo-inverse $\boldsymbol{L}^\dagger$ and thus the distribution of LODFs. In addition to that, the Laplacian eigenvalues determine the dynamical properties of power grids [46].

A. DISTRIBUTION OF LINE SUSCEPTANCES

In this section, we analyse the distribution of transmission line susceptances P_B. It has been demonstrated that the distribution of line reactances x_e follows approximately an exponential distribution $P_x(x_e) = \lambda e^{-\lambda x_e}$ by making use of the Kullback-Leibler divergence [30]. The DC approximation of the power flow is based on the assumption that transmission lines are purely inductive [34]. In this case, the line reactance and susceptance are related by $b_e \approx -x_e^{-1}$, i.e. we can obtain the distribution of line susceptances by taking the inverse distribution of the distribution of line reactances P_x. We will make use of this fact to compare the distribution of LODFs in networks with unit line susceptances and sucseptances following an inverse exponential distribution.

B. DISTRIBUTION OF INITIAL LINE LOADINGS

Here, we consider the distribution of line flows as it appears in a dispatch of the open energy system model 'PyPSA-EUR' which models the European energy system [33]. To this end, we evaluate the absolute flow $|F_i^{(0)}|$ on a line i and divide it by the maximal flow $F_i^{\max}$ on the line to evaluate the relative loading

$$L_i^{(0)} = \frac{|F_i^{(0)}|}{F_i^{\max}}. \tag{8}$$

We then examine the statistics of relative loadings for a dispatch spanning an entire year in hourly resolution. The network has 4428 lines and 3037 nodes and, as a result, there is a detailed statistics of relative loadings. For the given dataset, the flow on a line is limited to 80% of the maximal flow to incorporate a security constraint.

We find that the relative loadings are approximately exponentially distributed, i.e. they are described by the probability density function $P_L(L^{(0)}) = \lambda e^{-\lambda L^{(0)}}$ (see Figure 1). The maximum likelihood estimator for an exponential distribution is calculated as

$$\hat{\lambda} = \langle L^{(0)} \rangle^{-1}, \tag{9}$$

where $\langle \cdot \rangle$ denotes the average. For the empirical distribution of line loadings in 'PyPSA-EUR', we observe an estimate of $\hat{\lambda} \approx 5$ with small monthly variations.

C. DISTRIBUTION OF LODFS

To preprocess the data for evaluating the logarithmic distribution, we first create the set of all pairs of edges for which the LODF has non-zero entries

$$\mathcal{L} = \{\ell, k \in E(G) | \mathrm{LODF}_{\ell,e} \neq 0 \wedge \ell \neq k\}.$$

In Figure 2, we present the distribution of LODFs for two different topologies: The MATPOWER test case '2736sp' that represents the Polish power grid during peak conditions in summer of 2004 [21] (top row) and a 100×100 square grid with unit line susceptances (bottom row). Whereas the distribution of logarithmic LODFs closely corresponds to a log-normal distribution (a,d) with significantly stronger tails in both cases (b,e), the distribution of Laplacian eigenvalues differs greatly for the two topologies (c,f). Thus, there is a surprising similarity between the distribution of LODFs for vastly different topologies. Here, Gaussian fits are based

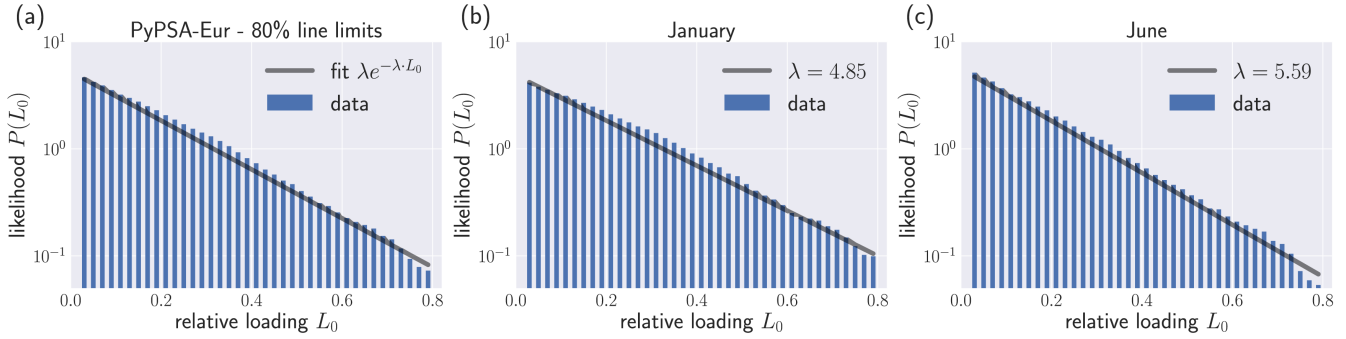

FIGURE 1. Distribution of absolute line loadings in a high-resolution energy system model. (a) We evaluate the magnitude of the relative line loadings $L^{(0)}$ over a year of demand and generation patterns in hourly resolution occurring in the European energy system model 'PyPSA-EUR' [33] described in table 1. The dispatch is calculated via an optimal power flow algorithm including a security margin, $|L_i^{(0)}| \leq L^{\text{thresh}} = 0.8$, as a proxy for $N-1$ security (cf. Eq. (17)). The probability density function is well approximated by an exponential distribution with parameter $\hat{\lambda} \approx 5.26$ estimated using the maximum likelihood estimator in Eq. (9). (b,c) The given dataset contains a weak seasonal effect, displaying slightly higher relative loading in winter months (b) than in summer months (c) that result in a steeper exponent of the distribution of relative loadings in the latter case.

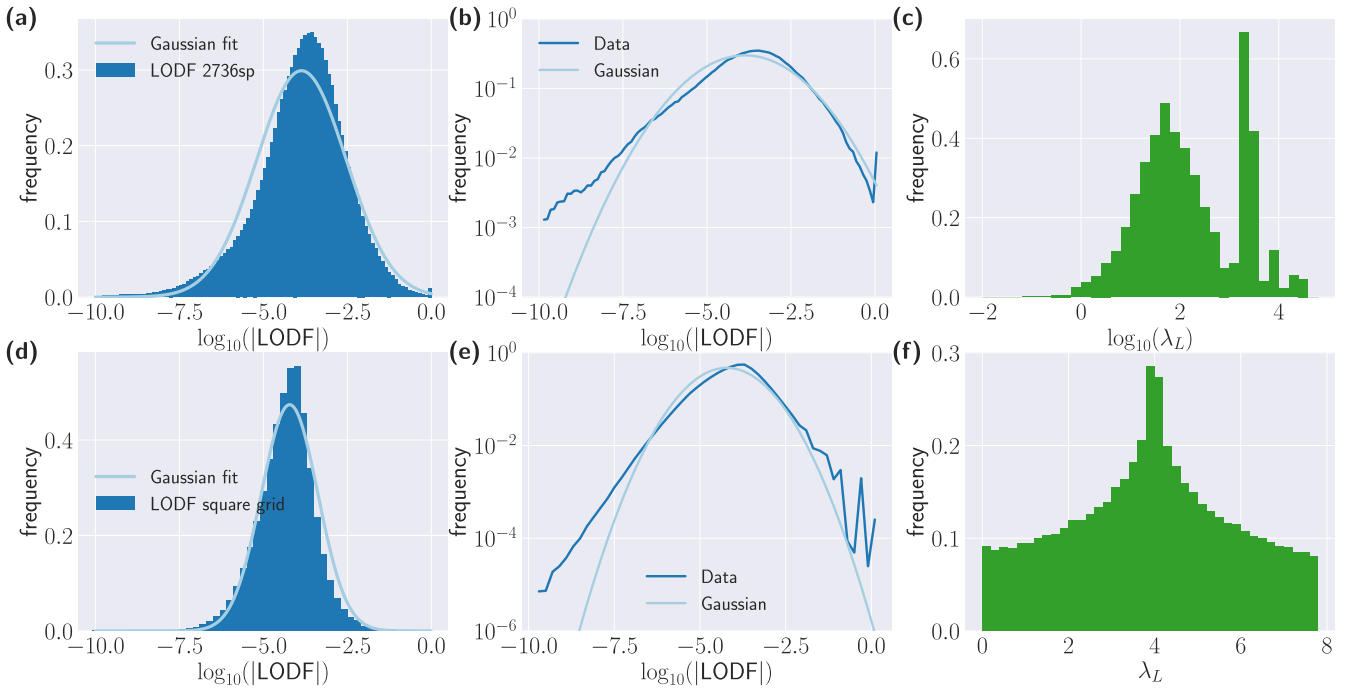

FIGURE 2. The distribution of absolute LODFs is approximately log-normal for both real-world and artificial network structure. We analyse the distribution of LODFs and Laplacian eigenvalues for (a-c) the real-world power grid '2736sp' that corresponds to the Polish power grid during summer peak 2004 with edge weights representing the link susceptance and (d-f) a regular square lattice of size 100 × 100 with unit edge weights. (a,d) The distribution of LODFs (dark blue histogram) follows approximately a log-normal distribution (light blue) except for the tails at low values for both the '2736sp' grid (a) and the square grid (d). (b,e) The heavy tails (dark blue) – as compared to the Gaussian distribution (light blue) – become clearly visible when analysing the distributions on a log-log scale. (c,f) Even though the distributions of LODFs have a similar shape for both networks, the spectra of the Graph Laplacian differ significantly between the two topologies. For the real-world grid, the distribution is clearly bimodal and spreads over 4 orders of magnitude – note the logarithmic x-scale. In contrast, for the regular lattice, the distribution differs considerably and spans only 1 order of magnitude. Thus, although both quantities are purely topological, the similarity in the log-normal distribution of LODFs cannot easily be understood in terms of the Laplacian spectrum alone.

on the maximum likelihood estimates of the mean μ_{LN} and variance σ^2_{LN} for a log-normal distribution which are given by [47]

$$\hat{\mu}_{\text{LN}} = \frac{\sum_{i=1}^{N} \log(X_i)}{N}, \tag{10}$$

$$\hat{\sigma}^2_{\text{LN}} = \frac{\sum_{i=1}^{N} \left(\log(X_i) - \hat{\mu}_{\text{LN}}\right)^2}{N}, \tag{11}$$

where X_i are the realizations of the random variable "X" under consideration.

To further examine the characteristics of the distribution of LODFs, we systematically evaluate different moments of the underlying distribution for different grids [48]. First, we calculate the mean of the absolute logarithmic LODFs

$$\mu = \frac{1}{|\mathcal{L}|} \sum_{\ell,k \in \mathcal{L}} \log(|\text{LODF}_{\ell,e}|). \tag{12}$$

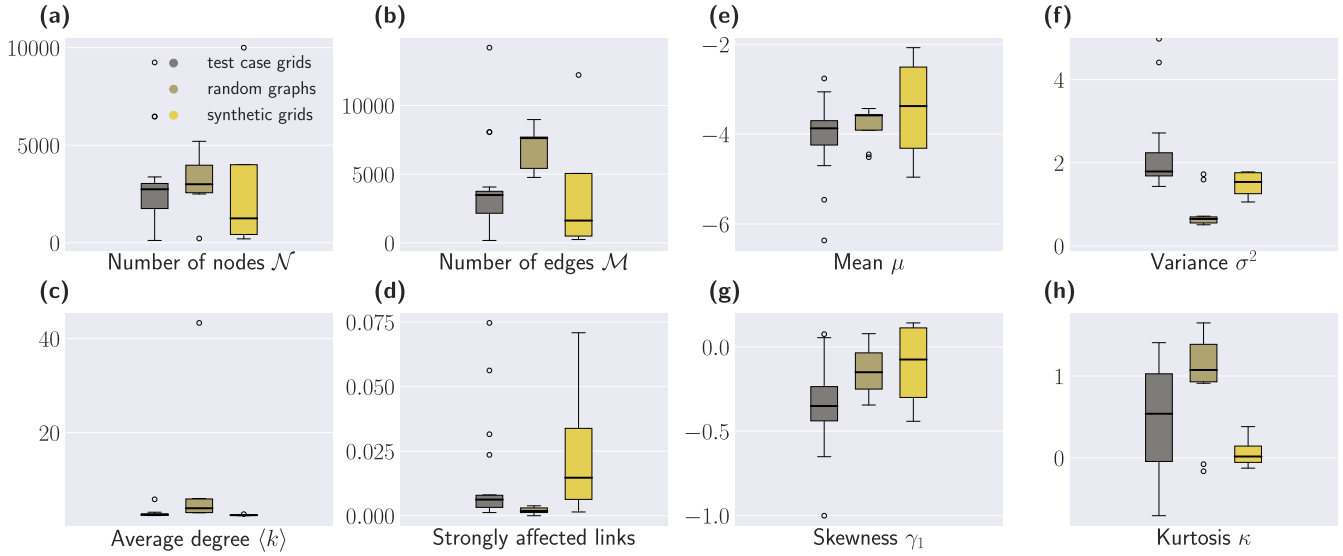

FIGURE 3. Statistical properties of the logarithmic LODF distribution and the network structure for synthetic grids, real world grids and random graphs. We analyse 11 random graphs, 4 synthetic grids and 20 test case grids that are inspired or correspond to real world grids (see tables 1 and 2 for details). (a,b,c) Attributes of the network topologies underlying the grids: The topological attributes span a wide range in terms of their number of nodes $\mathcal{N}$, their number of edges $\mathcal{M}$ and their average degree $\langle k \rangle$. (c,d,g,h) We analyse the first, second, third and fourth moment (see sec. III) of different real-world, random and synthetic grids (see tables 1 and 2). Although the grids are of different sizes (d) and different connectivities (g,h), the statistical properties are similar: the (logarithmic) mean lies approximately at $\mu \approx -4$, the variance around $\sigma^2 \approx 2$ except for the random grids which are much more regular, the skewness γ_1 is slightly negative and the excess kurtosis κ positive for almost all grids, indicating that large deviations are more likely than for Gaussian distributions. Test cases are taken from Refs. [20]–[24], [33] (see table 1 for details).

Second, we calculate the variance σ^2

$$\sigma^2 = \frac{1}{|\mathcal{L}|} \sum_{\ell,k \in \mathcal{L}} (\log(|\mathrm{LODF}_{\ell,e}|) - \mu)^2. \tag{13}$$

To specifically compare the distributions to log-normal distributions, we also calculate the normalized third and fourth moment, namely the skewness γ_1

$$\gamma_1 = \frac{1}{|\mathcal{L}|} \sum_{\ell,k \in \mathcal{L}} \left(\frac{\log(|\mathrm{LODF}_{\ell,e}|) - \mu}{\sigma} \right)^3, \tag{14}$$

and the excess kurtosis

$$\kappa = \frac{1}{|\mathcal{L}|} \sum_{\ell,k \in \mathcal{L}} \left(\frac{\log(|\mathrm{LODF}_{\ell,e}|) - \mu}{\sigma} \right)^4 - 3. \tag{15}$$

The skewness vanishes for a Gaussian distribution due to its symmetry. The excess kurtosis measures the deviation from a kurtosis of three observed for the Gaussian distribution and indicates if rare events happen more ($\kappa > 0$) or less ($\kappa < 0$) frequently than for a Gaussian distribution [49].

Finally, we make use of another indicator that is related to the upper tail of the distribution of LODFs and gives a measure of a grid's vulnerability: We calculate the relative number of LODFs exceeding a threshold of 0.1

$$\frac{\left|\{\ell, k \in \mathcal{L} || \mathrm{LODF}_{\ell,e}| > 0.1\}\right|}{\mathcal{M}(\mathcal{M}-1)}, \tag{16}$$

and refer to this measure as the 'strongly affected links'. The measure may be interpreted as the probability that the failure of a randomly chosen link results in the increase of the flow on another randomly chosen link by more than 10% of the flow carried initially by the failing link.

We analyse these properties of the distribution of logarithmic LODFs for different power grids in Figure 3(e-h). In particular, we consider test case grids that are based on power system test cases and synthetic grids that are created using a synthetic power grid algorithm (see Table 1 in the Appendix). To get a better statistics and benchmark the results, we also consider random graphs that are generated either from regular grids or random network models (see Table 2 in the Appendix).

All grids are similar in terms of their average degree $\langle k \rangle$ except for the random graphs that – in some cases – display a much higher number of edges which results in larger average degrees (see panels (a-c)). The mean of the logarithmic absolute LODFs is $\overline{\mu} = -3.91$ for all grids tested and the variance is $\overline{\sigma^2} = 1.73$, except for the random graphs where a much lower value of a variance close to unity may be observed. This is likely due to the fact that the random graphs considered here are in most cases very regular in terms of the graph degree and thus much more homogeneous than realistic power grids. The skewness is negative for almost all distributions tested with a mean skewness of $\overline{\gamma_1} = -0.26$ evaluated over all grids, indicating distributions with a peak located at values larger than the mean value. For the excess kurtosis, we observe almost exclusively values larger than zero, in most cases exceeding unity, with a mean of $\overline{\kappa} = 0.60$. This indicates that almost all LODF distributions have heavier tails than a log-normal distribution. In terms of the network vulnerability as measured by the strongly affected links, we observe a mean

of $9.5 \cdot 10^{-3}$ with the synthetic grid models displaying a much higher vulnerability with a mean of $2.5 \cdot 10^{-2}$. Thus, we conclude that the distribution of LODFs displays a high degree of similarity for different power grids and synthetic networks. In Figures 7,8,9 and 10 in the Appendix, we show the actual statistics of LODFs for 24 test case grids and synthetic grids, for which aggregated statistical properties are summarized in Figure 3.

IV. APPLICATION: A CASCADE MODEL WITH LOG-NORMAL LOAD REDISTRIBUTION

Based on our finding that LODFs are log-normally distributed over a wide range of topologies, we will now discuss a simple probabilistic cascade model that incorporates this effect. To this end, we will study a modification of the 'CASCADE' model due to Dobson *et al.* [7], [8].

A. THE CASCADE MODEL AND A POSSIBLE EXTENSION

Consider a simple network consisting of N components. Initially, each component j is assumed to have a load $L_j^{(0)}$ that is smaller than its maximal load $L_j^{\max}$. If a component exceeds its maximal load, the component breaks down and a redistribution mechanism is triggered that distributes the load to the other components in the network. These may in turn trigger further breakdowns, resulting in a cascade of breakdowns that eventually stops if the network has broken down entirely or if no further overloads occur.

In the original setup by Dobson *et al.*, the initial loads $L^{(0)}$ are drawn from a uniform distribution, i.e. $L^{(0)} \in \mathcal{U}(L^{\min}, L^{\text{thresh}})$ where $L^{\min}$ is the minimum loading in the distribution and $L^{\text{thresh}} \leq L^{\max}$ is the threshold loading, potentially incorporating a security margin to the maximal loading $L^{\max}$. Redistribution after failures is incorporated by increasing the load on all components in the network by a constant addend D_1. Furthermore, the cascade of failures is triggered by an initial shock that increases the load on all components by an added D_0. Thus, the loading on a component i in a network after the failure of M components is calculated as

$$L_i^{(1)} = L_i^{(0)} + D_0 + MD_1.$$

Choosing critical values of these parameters that depend on the system size N, Dobson *et al.* demonstrate that this model yields a power law of the number of components failing – in close correspondence with power laws of blackout sizes observed empirically in historic power blackout sizes [3].

Inspired by the redistribution of real power flow after line failures in power transmission grids as described in Eq. (6), we suggest extending this mechanism as follows:

1. Security margin and distribution of line loadings: Typically, real-world power grids are operated using the $N-1$ security criterion which means that upon the failure of any transmission or generation element, no other line becomes overloaded. This is approximately taken into account in the model by a security margin

$$L^{\text{thresh}} = c \cdot L^{\max} \tag{17}$$

which limits the maximal initial loading to a share $c \in [0, 1]$ of the maximal $L^{\max}$. In the following, we simply set $L^{\max} = 1$ for all components such that our model emulates relative loading of components.

We then consider either a uniform distribution of component loadings as in the original 'CASCADE' model such $L^{(0)} \in \mathcal{U}(-L^{\text{thresh}}, L^{\text{thresh}})$, or an exponential distribution that we have found empirically in a large-scale energy system model (see Sec. III-B). In the latter case, we initially draw all loadings from an exponential distribution $P_L(L^{(0)}) = \lambda e^{-\lambda L^{(0)}}$ with $\lambda \approx 5$. If the initial loading on an element i exceeds the threshold value, we simply reset it with the threshold value $L_i^{(0)} := L^{\text{thresh}}$.

2. Redistribution after failures: Inspired by our findings on log-normally distributed LODFs, we adopt a redistribution scheme after failures in the spirit of the redistribution of real power flow on transmission lines as introduced in Eq. (6). Assume that the component k with initial load $L_k^{(0)}$ fails. We suggest updating the load on another component i by

$$L_i^{(1)} = L_i^{(0)} + (-1)^m \mathfrak{L} L_k^{(0)}, \quad \forall i \neq k \in \{1, \ldots, N\}. \tag{18}$$

Here, $\mathfrak{L} \in \text{Lognormal}(\mu, \sigma)$ is drawn from a log-normal distribution with mean μ and standard deviation σ and m assumes the values one or zero with equal probability to model the randomly chosen sign. This update rule thus corresponds to a probabilistic version of the update rule (6) if we insert the relative loadings (see Eq. (8)) and assume that all lines have the same maximal loading. Thus, our model adopts the redistribution scheme observed in power transmission grids in the DC approximation and takes into account the fact

Algorithm 1 Stochastic cascade model

```
% Randomly choose initial trigger element k ∈ {1, ..., N} and add it to set of failing components C
C ← k
repeat
  has_overloads ← 0
  % Redistribute load from all failed components:
  for all remaining components do
    % Update ∀i ∈ {1, ..., N}, i ∉ C:
    L_i ← L_i + Σ_{k∈C} (−1)^m 𝔏 L_k,   .
    % Reset set of current failures and remove failed components
    C ← {}
    if |L_i| > L^max then
      Add i to set of current failures C
      has_overloads ← 1
    end if
  end for
until has_overloads = 0
```

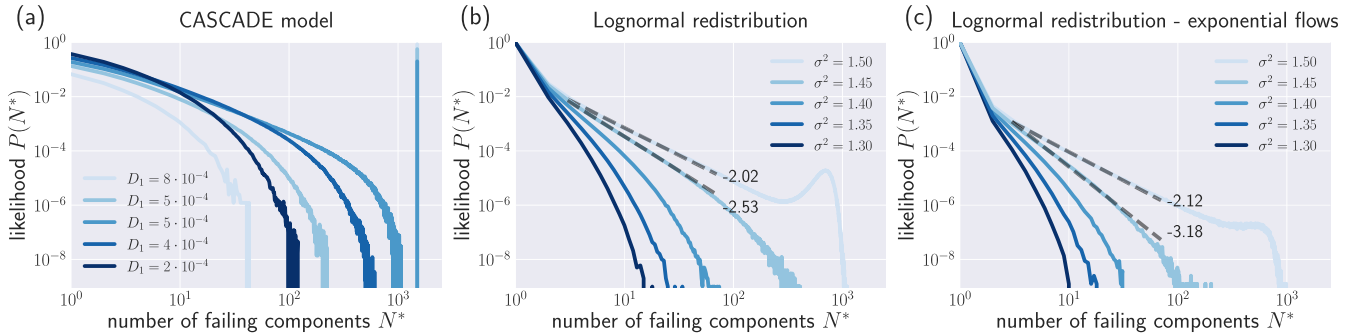

FIGURE 4. Distribution of outage sizes for different random redistribution models. We show the likelihood $P(N^*)$ of an outage of size N^* for (a) the CASCADE model, (b) the log-normal redistribution model with uniform initial loads and (c) with exponentially distributed initial loads. The effective parameters in the CASCADE model are chosen to be consistent to the parameters in the other models or their averages, respectively (see Eq. (19)). Colour code (from dark blue to light blue) represents increasing values of redistribution constant D_1 for the CASCADE model (a) and variance of the Gaussian distribution of logarithmic redistribution factors σ^2 for the other models (b,c). While the CASCADE model displays an abrupt transition point at which either less or all components in the system fail, the transition is smoother for the suggested models. Note the peak in the histogram at the system size $N = 1500$ for the CASCADE model (a), which indicates that the entire system fails with a high likelihood. The system considered here has $N = 1500$ elements for all panels, the mean of the distribution is set to $\mu = -5$ for panels (b) and (c) and parameters for CASCADE model are calculated using the parameters shown in b) and Eqs. (19) and (20). Grey dotted lines represent least-square fits of linear functions on the intermediate range of failure sizes performed on the log-log scales for the two top curves along with the resulting scaling exponent. The range of parameters is chosen such that it agrees with typical values found in statistics of LODFs in test case and synthetic grids (see Table 1).

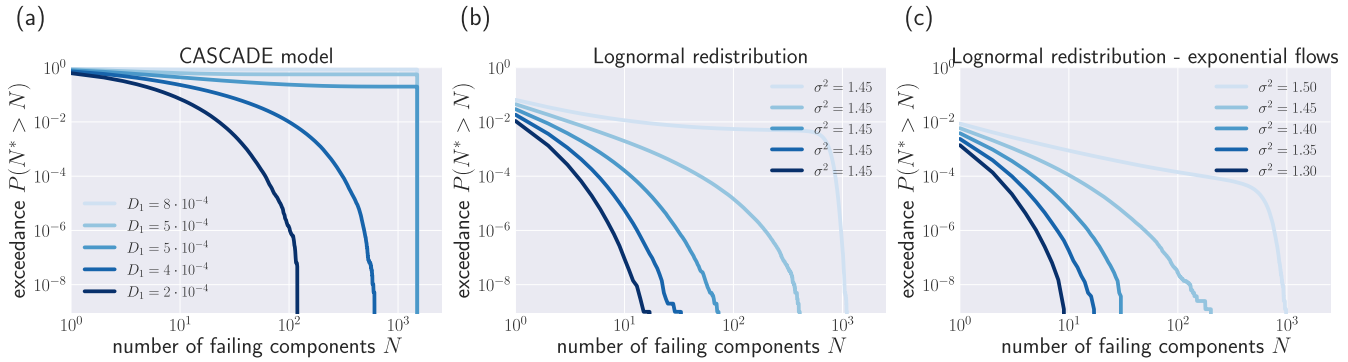

FIGURE 5. Distribution of outage sizes for different random redistribution models. We show the exceedance $P(N^* > N)$ which is the likelihood that an outage will result in the failure of N or more elements. The distributions shown here are the same as in Figure 4.

that LODFs are log-normally distributed throughout different topologies.

3. Initial trigger event: In contrast to the 'CASCADE' model, we assume that the initial event triggering the cascade of failures is the failure of a single element as well, i.e. the cascade is triggered by the same mechanism that makes it propagate. This corresponds to the mechanism for cascading failures in real power transmission grid, where large scale cascades are often triggered by a single link failure that triggers other link failures and so forth [9], [14], [15].

We summarise this cascade model in the Algorithm 1.

The proposed model thus incorporates crucial aspects of flow rerouting and cascading failures in power flow models while remaining entirely probabilistic and, in this regard, extending previous models. Similar to the 'CASCADE' model, this has the advantage that there is no need to consider a particular grid topology and allows approaching the statistics of cascading failures purely from a probabilistic viewpoint. Thus, the model fills a gap between realistic, but non-probabilistic models and purely probabilistic models that are ruled by less realistic redistribution schemes [4], [50]–[52].

B. HEAVY TAILS OCCUR OVER WIDE RANGE OF PARAMETERS

Outage sizes in empirical data have been demonstrated to have heavy tails [53]. Different explanations for this scaling law have been put forward, ranging from an interpretation of the power system being in a critical state [4], [7] to relating the power law to power laws in city size distributions [5]. A recent analysis of the probability distribution of the number of customers affected per outage in the U.S. has found a load dependency of the scaling exponent with typical values ranging from -2.1 to -2.8 [54]. Here, we demonstrate that power laws of outage sizes occur over a wide range of parameters in our extended CASCADE model. Our model incorporates essential properties of failure cascade in linear flow models while being entirely probabilistic.

TABLE 1. Distribution of topological parameters and moments of the logarithmic LODF distribution for 24 different test grids. We refer to the 20 first grids as 'test case grids' since they are mostly based on power system cases except for the two grids taken from PyPSA-Eur and to the latter four as 'synthetic grids'. Test grids are either taken from the publicly available test case archive of MATPOWER [20]–[22] or taken from the University of Washington power systems test case archive [24]. The Scandinavian grid data and the central European topology where extracted from the open energy system model PyPSA-Eur [33] which are based on the publicly available network data by the transparency platform of the European Network of Transmission System Operators (ENTSO-E). 'case_ACTIVSg' are synthetic power grids inspired by real-world North American power grids [23].

Test case	Mean μ	Variance σ^2	Skewness γ_1	Kurtosis κ	Number of nodes $\mathcal{N}$	Number of edges $\mathcal{M}$	Average degree $\langle k \rangle$
IEEE case118	-2.762	1.698	-0.090	-0.707	118	179	3.034
case145	-3.481	2.631	-1.002	1.407	145	422	5.821
IEEE case300	-3.060	2.069	-0.401	-0.040	300	409	2.727
case1354pegase	-3.544	1.431	0.012	0.042	1354	1710	2.526
case1888rte	-3.680	1.689	-0.330	0.604	1888	2308	2.445
case1951rte	-3.737	1.791	-0.412	0.827	1951	2375	2.435
case2383wp	-3.816	1.584	-0.324	1.072	2383	2886	2.422
case2737sop	-3.875	1.810	-0.651	1.329	2737	3497	2.555
case2746wop	-3.868	1.790	-0.651	1.398	2746	3505	2.553
case2848rte	-4.081	1.894	-0.157	0.215	2848	3442	2.417
case2868rte	-4.157	2.110	-0.429	1.012	2868	3471	2.421
case2869pegase	-5.461	4.982	-0.466	-0.200	2869	3968	2.766
case3012wp	-3.841	1.574	-0.261	0.882	3012	3566	2.368
case3120sp	-3.893	1.668	-0.293	0.827	3120	3684	2.362
case3375wp	-3.907	1.627	-0.382	1.196	3374	4068	2.411
case6468rte	-4.702	1.713	0.055	0.474	6468	8065	2.494
case6470rte	-4.691	1.689	0.076	0.459	6470	8066	2.493
case9241pegase	-6.366	5.117	-0.306	-0.410	9241	14207	3.075
Scandinavia_PyPSA	-3.709	4.412	-0.623	-0.310	272	373	2.743
Central_Europe_PyPSA	-4.502	2.717	-0.371	-0.057	2440	3494	2.864
case_ACTIVSg200	-2.077	1.057	-0.441	0.065	200	245	2.450
case_ACTIVSg500	-2.653	1.323	-0.252	-0.032	500	584	2.336
case_ACTIVSg2000	-4.100	1.754	0.143	-0.126	2000	2667	2.667
case_ACTIVSg10k	-4.958	1.780	0.104	0.382	10000	12217	2.443

TABLE 2. Distribution of topological parameters and moments of the logarithmic LODF distribution for different regular and random graphs. Link weights were either set to unity or calculated based on the inverse parameters of an exponential distribution with $\lambda = 0.02$. To produce the Voronoi lattics, we distributed 2000 points randomly in the unit square [0, 1] × [0, 1] and calculated their Voronoi tessellation. For the Erdős–Rényi (ER) random graph we used a connection probability of $p = 0.2$.

Test case	Mean μ	Variance σ^2	Skewness γ_1	Kurtosis κ	Number of nodes $\mathcal{N}$	Number of edges $\mathcal{M}$	Average degree $\langle k \rangle$
ER graph - unweighted	-3.819	0.565	-0.335	0.911	220	4766	43.327
Square grid - unweighted	-3.584	0.663	-0.170	0.947	2500	4900	3.920
Square grid - weighted	-3.582	0.715	-0.150	0.978	2500	4900	3.920
Triangular grid - unweighted	-4.451	1.595	-0.344	-0.164	2626	7625	5.807
Triangular grid - weighted	-4.514	1.725	-0.330	-0.078	2626	7625	5.807
Delaunay lattice - unweighted	-3.909	0.509	0.071	1.649	3000	8974	5.983
Delaunay lattice - weighted	-3.914	0.515	0.079	1.594	3000	8975	5.983
Voronoi lattice - weighted	-3.469	0.649	-0.062	1.074	3973	5947	2.994
Voronoi lattice - unweighted	-3.431	0.540	-0.008	1.491	3976	5953	2.994
Hexagonal grid - unweighted	-3.558	0.601	-0.126	1.284	5200	7699	2.961
Hexagonal grid - weighted	-3.575	0.681	-0.162	1.135	5200	7699	2.961

To be able to compare the proposed model to the CASCADE model, we choose the redistribution parameter D_1 in the CASCADE model as the expected value of the product probability distribution between initial loadings $L^{(0)}$ and redistribution factors $\mathfrak{L}$, such that

$$D_1 := E\left(|\mathfrak{L} \cdot L^{(0)}|\right). \tag{19}$$

If a large number of components fails, we thus have the formal equivalence in the update equations

$$L_i^{(1)} = L_i^{(0)} + \sum_{k=1}^{M} |\mathfrak{L} \cdot L_k^{(0)}| \approx L_i^{(0)} + MD_1.$$

Furthermore, we choose the initial trigger parameter D_0 to be equal to the initial security margin to which we add a small value

$$D_0 := (L^{\max} - L^{\text{thresh}}) + E\left(|\mathfrak{L} \cdot L^{(0)}|\right), \tag{20}$$

since below this value, the initial trigger event cannot result in a cascade and this value simulates a behaviour close to criticality, where power-laws of cascade sizes have been observed.

In Figure 4, we compare the resulting cascade sizes obtained for a large number of simulations of the model with the parameters indicated. We analyse the likelihood $P(N^*)$ that a given number of components N^* fails, calculated over 10^8–10^9 realisations of the initial conditions and randomly chosen trigger elements. We consider (a) the CASCADE model, (b) the stochastic load redistribution model suggested at the beginning of this section IV with uniform initial

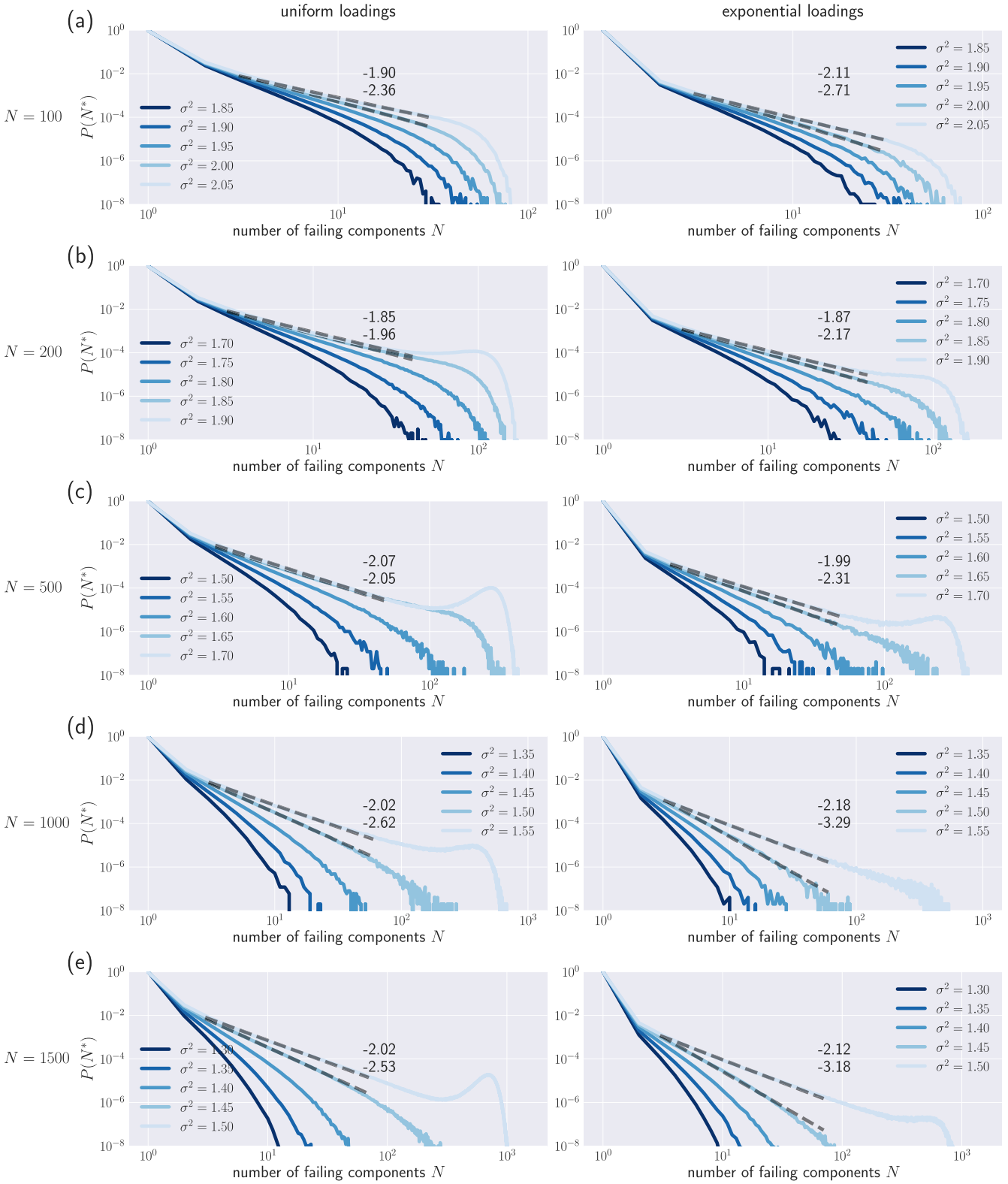

FIGURE 6. Distribution of outage sizes in the stochastic load redistribution model for systems of different sizes. We compare the likelihood $P(N^*)$ of an outage of size N^* for systems with initial element loadings drawn uniformly (left) and exponentially (right) for systems with different number of elements (a-e). We fix the mean of redistribution factors to $\mu = -5$ and choose the range of variances σ^2 for each system size such that power laws of outage sizes occur. The range of parameters considered here matches typical values found in redistribution factors for synthetic and test case grids (see Table 1).

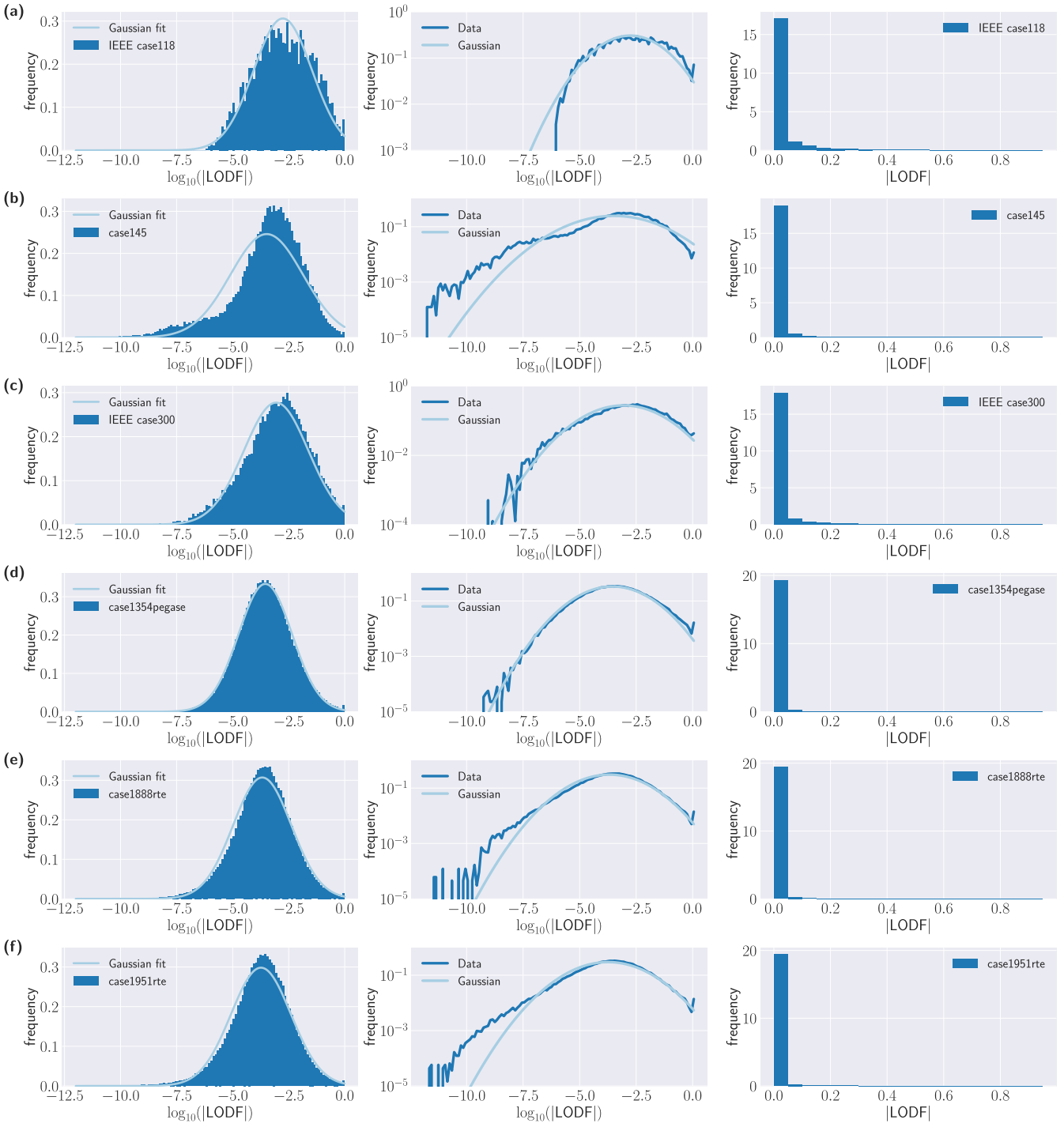

FIGURE 7. Distribution of absolute LODFs for the first six test grids listed in table 1. We plot the frequency of occurence of logarithmic absolute LODFs and a Gaussian fit (see sec. III) with a linear y-scale (left) and a logarithmic y-scale (center). In most cases, the frequency of occurence of logarithmic absolute LODFs is well-approximated by the Gaussian fit. Furthermore, we plot the frequency of occurrence of absolute LODFs (right).

loadings and (c) with exponential initial loadings. Typical values for the log-normal statistics of redistribution factors used in the stochastic load redistribution model are extracted from the parameters obtained for test case and synthetic grids listed in Table 1 and discussed in section III-C: We choose a mean of $\mu = -5$ which is in the typical range of $\mu \in [-3, -6]$ observed for the logarithmic mean and a variance of $\sigma^2 \in [1.3, 1.5]$ which also matches the typical range of $\sigma^2 \in [1, 3]$ observed for actual grid datasets. In Figure 6 in the Appendix, we analyse the sensitivity of these results for varying system sizes and varying variance σ^2 chosen in the critical range. Note that values differing from the critical range will result in either only a small share of the components or the entire system failing due to a limited system size. Furthermore, we fix the security margin on the relative loading to $c = 0.7$. Note that this "70%-rule" is

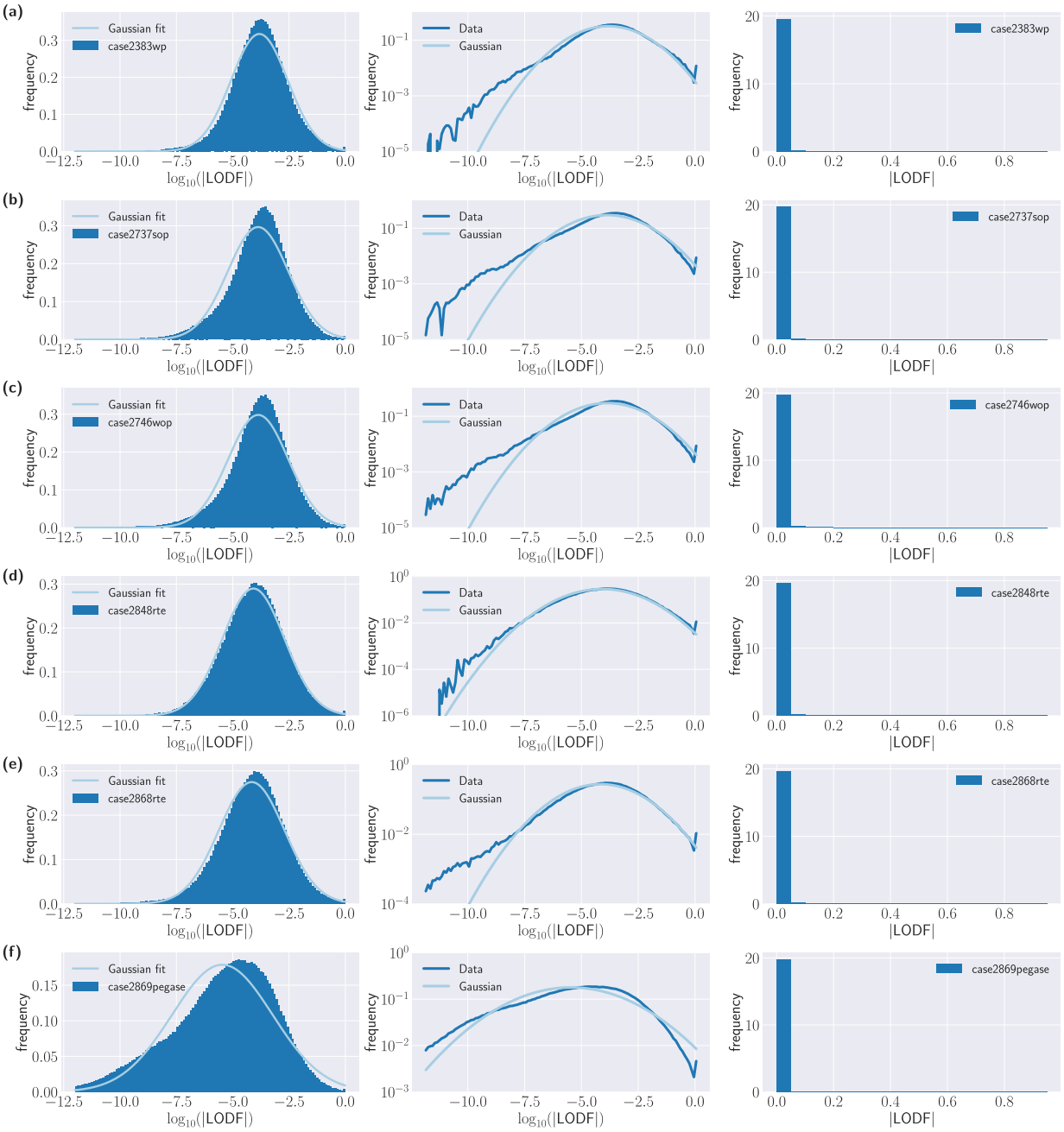

FIGURE 8. Distribution of absolute LODFs for the sixth up to the twelfth test grids listed in table 1. We plot the frequency of occurence of logarithmic absolute LODFs and a Gaussian fit (see sec. III) with a linear y-scale (left) and a logarithmic y-scale (center). In most cases, the frequency of occurence of logarithmic absolute LODFs is well-approximated by the Gaussian fit. Furthermore, we plot the frequency of occurrence of absolute LODFs (right).

a common way to ensure approximate $N-1$ security also when operating and modelling real-world power transmission grids [55]–[57].[1]

[1] Note that in the dataset shown in Figure 1 the security margin is set to $L^{\text{thresh}} = 0.8$, i.e. 80% of the maximal loading. We make use of this dataset to estimate the scaling exponent since flows are not strongly affected by the threshold in this case which would otherwise result in a peak at $L^{(0)} = 0.7$ as an indication of positive shadow prices, i.e. a possible economic optimum with higher line flows [58]. Thus, we estimate the exponent from this distribution to be able to use the entire range of loadings for estimation of the scaling exponent.

Whereas the CASCADE model (a) displays a rather abrupt transition point for which either fewer components or the entire system fail, the curves are smoother for all parameters in the models suggested here. In particular, the curves in the suggested models are more flat for all parameters under consideration indicating a power law over a wider range of parameters. Thus, we conclude that our model can reproduce essential features observed in the CASCADE model while presenting power laws over a wide range of parameters. For the critical cases where power laws occur, we find that the

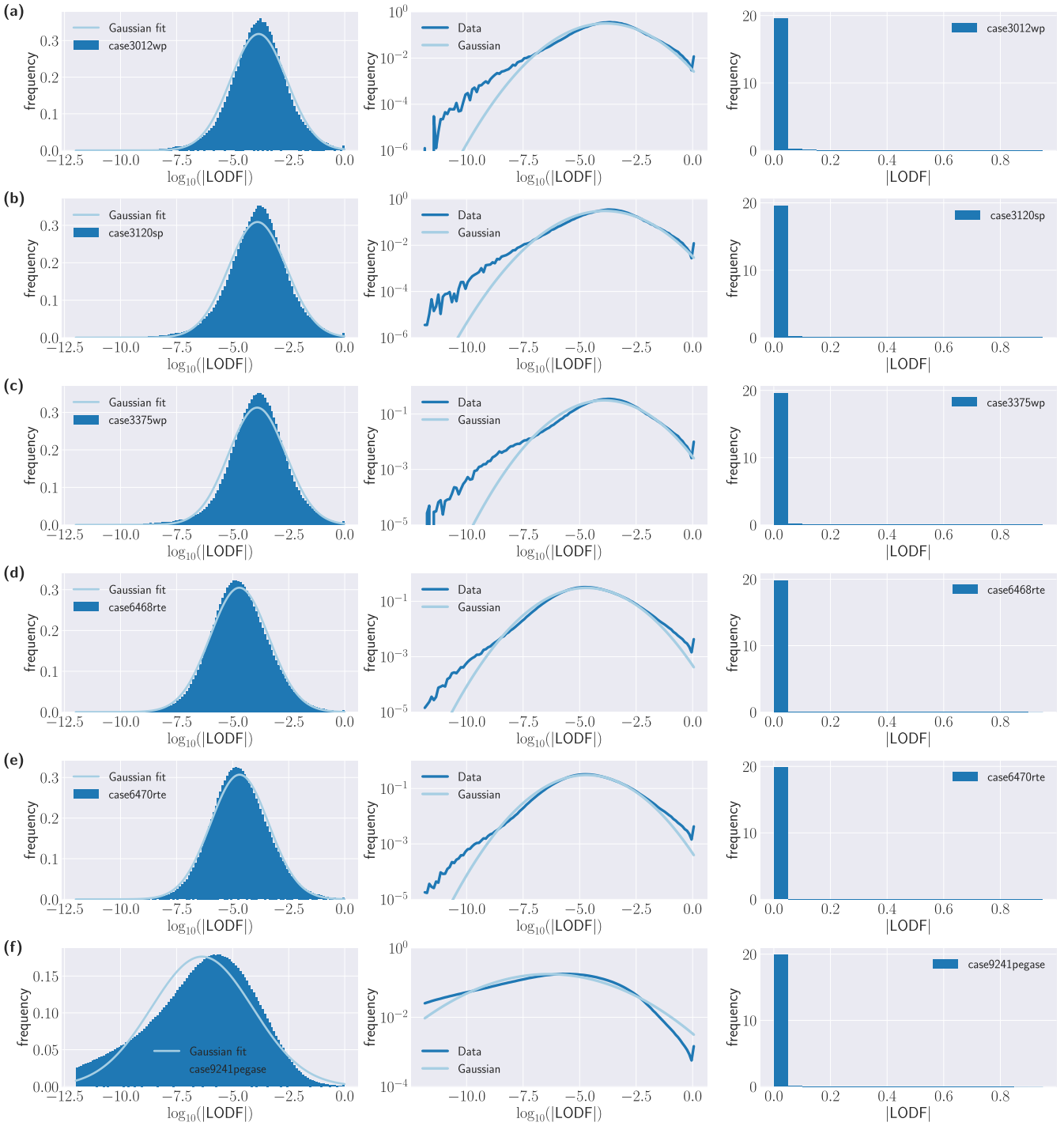

FIGURE 9. Distribution of absolute LODFs for the twelfth up to the eighteenth test grids listed in table 1. We plot the frequency of occurence of logarithmic absolute LODFs and a Gaussian fit (see sec. III) with a linear y-scale (left) and a logarithmic y-scale (center). In most cases, the frequency of occurence of logarithmic absolute LODFs is well-approximated by the Gaussian fit. Furthermore, we plot the frequency of occurrence of absolute LODFs (right).

scaling exponents (dotted lines, Figure 4) matches the scaling exponents found in empirical data with values in the range of -2 to -3. To confirm this result, we also analyse the likelihood of exceedance $P(N^* > N)$, which is the likelihood that the outage size exceeds N elements, for the same distributions of outage sizes in Figure 5 in the Appendix and analyse the scaling of the probability of outage sizes for systems with different number of elements in Figure 6 in the Appendix.

V. DISCUSSION AND CONCLUSION

In this manuscript, we analysed the distribution of Line Outage Distribution Factors for different real-world- and

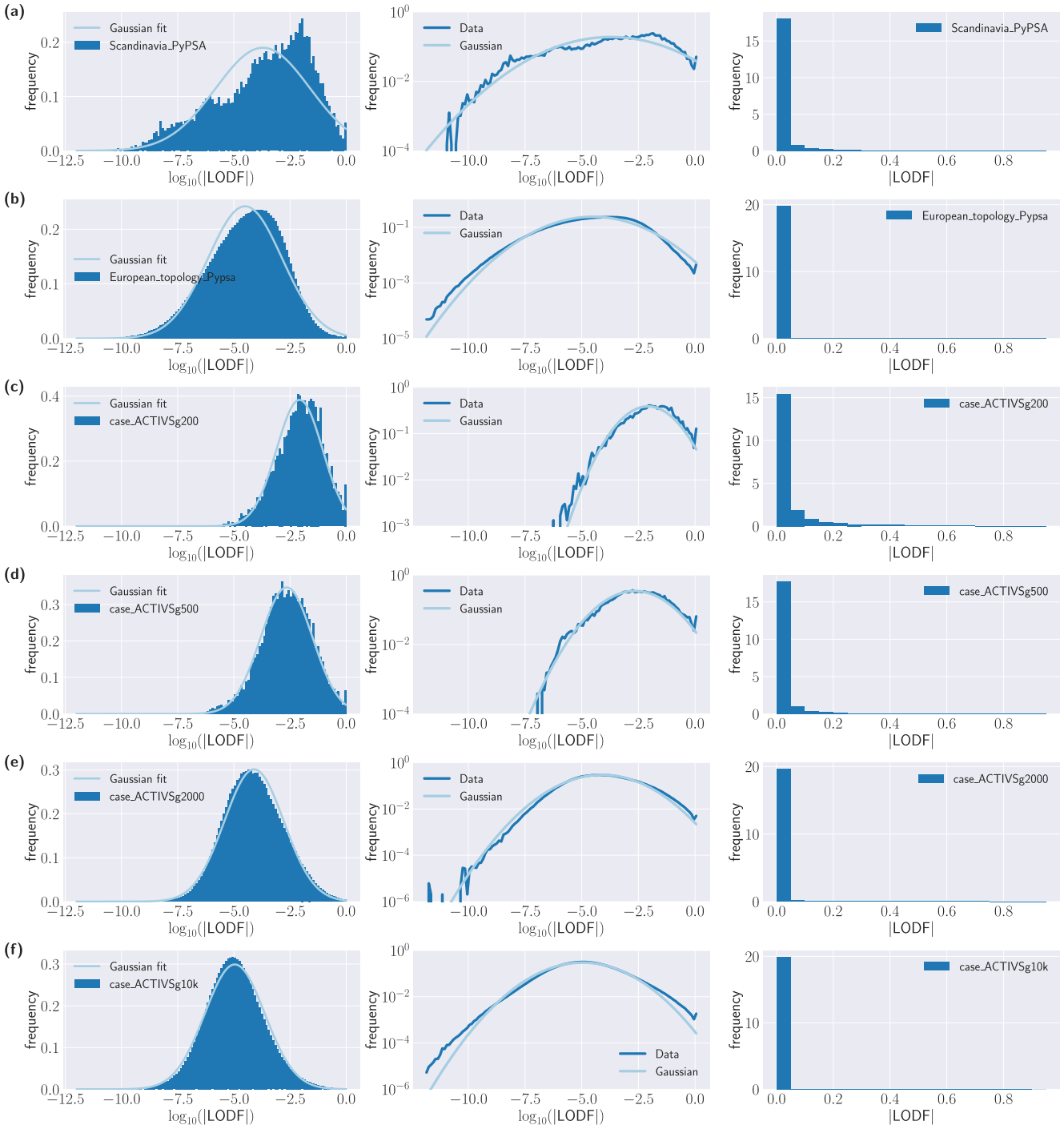

FIGURE 10. Distribution of absolute LODFs for the final six test grids listed in table 1. We plot the frequency of occurence of logarithmic absolute LODFs and a Gaussian fit (see sec. III) with a linear y-scale (left) and a logarithmic y-scale (center). In most cases, the frequency of occurence of logarithmic absolute LODFs is well-approximated by the Gaussian fit. Furthermore, we plot the frequency of occurrence of absolute LODFs (right).

synthetic grids. In particular, we analysed how this distribution changes throughout different synthetic and real-world topologies: We found that the distribution of the magnitude of LODFs is approximately log-normal, but additionally shows heavy tails throughout the topologies analysed here. We made use of this finding to introduce a stochastic load redistribution model for cascading failures that incorporates essential mechanisms of link failures in linear flow models – such as the aforementioned log-normal distribution of redistribution factors. The model, as a result of the log-normal distribution of LODFs, offers a potential explanation for the widespread occurrence of power laws in empirical data of power outage sizes.

In contrast to microscopic studies that analyse the impact of individual failures, our approach is a macroscopic one, focusing on the statistics of redistribution factors.

This approach has been demonstrated to be fruitful in many regards. On the one hand, it allows shifting the focus from the small-scale network structure to the vulnerability of a network as a whole. The distribution may thus be used to characterise network resilience, and potentially come up with a single index for an entire network – similar to the relative number of ''strongly affected links'' suggested here. Furthermore, this distribution could be used to evaluate whether a given synthetic network topology corresponds to realistic power grid topology in terms of network resilience. Our results demonstrate that there is a strong correspondence between the distributions of LODFs for vastly different topologies and even random graphs. Although the calculation of LODFs is based purely on the network topology via the pseudoinverse of the graph Laplacian, we could not find a simple theoretical explanation for this universal scaling.

On the other hand, a focus on the statistics of LODFs rather than individual values allows studying cascade models that feature more realistic flow redistribution. These models remain entirely probabilistic and thus do not require any assumptions about network topology. In this manuscript, we laid a first foundation by proposing such a cascade model that we study numerically. However, in principle, the a priori knowledge of the statistics of loadings and redistribution will also allow estimating cascade statistics analytically. This could help to shed further light on empirical observation of cascade sizes in real-world outages.

The stochastic load redistribution model for cascading failures introduced here focuses on the number of the failing components as an indicator for the severity of blackout sizes. This is due to the fact that the statistics of LODFs can be most easily and most directly related to the failure of individual components and thus outage sizes. Nevertheless, a number of other indicators has been proposed that aim to classify the severity of blackouts such as the number of customers affected, the unserved power or the value of lost load to name but a few [5], [7], [59], [60]. However, in order to extend our model with one or several other indicators, we would have to tune additional parameters to estimate e.g. the statistics of power consumption, which is why we leave this question open for future studies.

APPENDIX
TABLE OF POWER GRID TEST CASES

See Table 1 and 2.

APPENDIX
ADDITIONAL FIGURES

See Figure (5)–(10).

ACKNOWLEDGMENT

The authors would like to thank Tom Brown for providing the power grid data and dispatch of the European power system and Jonas Wassmer and Philipp Böttcher for helpful comments.

REFERENCES

[1] K. Everhart and G. Molnar. (2021). *International Energy Agency*. Accessed: Mar. 1, 2021. [Online]. Available: https://www.iea.org/commentaries/severe-power-cuts-in-texas-highlight-energy-security-risks-related-to-extreme-weather-events

[2] Union for the Coordination of Transmission of Electricity. (2007). *Final Report on the System Disturbance on 4 November 2006*. Accessed: Mar. 1, 2021. [Online]. Available: https://www.entsoe.eu/fileadmin/user_upload/_library/publications/ce/otherreports/Final-Report-20070130.pdf

[3] P. Hines, K. Balasubramaniam, and E. C. Sanchez, ''Cascading failures in power grids,'' *IEEE Potentials*, vol. 28, no. 5, pp. 24–30, Sep. 2009.

[4] B. A. Carreras, D. E. Newman, I. Dobson, and A. B. Poole, ''Evidence for self-organized criticality in a time series of electric power system blackouts,'' *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 9, pp. 1733–1740, Sep. 2004.

[5] T. Nesti, F. Sloothaak, and B. Zwart, ''Emergence of scale-free blackout sizes in power grids,'' *Phys. Rev. Lett.*, vol. 125, no. 5, Jul. 2020, Art. no. 058301.

[6] J. Lehmann and J. Bernasconi, ''Stochastic load-redistribution model for cascading failure propagation,'' *Phys. Rev. E*, vol. 81, no. 3, Mar. 2010, Art. no. 031129.

[7] 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, Interdiscipl. J. Nonlinear Sci.*, vol. 17, no. 2, Jun. 2007, Art. no. 026103.

[8] 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.

[9] P. Pourbeik, P. S. Kundur, and C. W. Taylor, ''The anatomy of a power grid blackout–root causes and dynamics of recent major blackouts,'' *IEEE Power Energy Mag.*, vol. 4, no. 5, pp. 22–29, Sep. 2006.

[10] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power Generation, Operation and Control*. New York, NY, USA: Wiley, 2014.

[11] H. Cetinay, F. A. Kuipers, and P. Van Mieghem, ''A topological investigation of power flow,'' *IEEE Syst. J.*, vol. 12, no. 3, pp. 2524–2532, Sep. 2018.

[12] S. Kettemann, ''Delocalization of disturbances and the stability of AC electricity grids,'' *Phys. Rev. E*, vol. 94, no. 6, Dec. 2016, Art. no. 062311.

[13] D. Labavić, R. Suciu, H. Meyer-Ortmanns, and S. Kettemann, ''Long-range response to transmission line disturbances in DC electricity grids,'' *Eur. Phys. J. Special Topics*, vol. 223, no. 12, pp. 2517–2525, Oct. 2014.

[14] J. Strake, F. Kaiser, F. Basiri, H. Ronellenfitsch, and D. Witthaut, ''Non-local impact of link failures in linear flow networks,'' *New J. Phys.*, vol. 21, no. 5, May 2019, Art. no. 053009.

[15] F. Kaiser, J. Strake, and D. Witthaut, ''Collective effects of link failures in linear flow networks,'' *New J. Phys.*, vol. 22, no. 1, Jan. 2020, Art. no. 013053.

[16] J. Lehmann and J. Bernasconi, ''Current redistribution in resistor networks: Fat-tail statistics in regular and small-world networks,'' *Phys. Rev. E*, vol. 95, no. 3, Mar. 2017, Art. no. 032310.

[17] W. Medjroubi, U. P. Müller, M. Scharf, C. Matke, and D. Kleinhans, ''Open data in power grid modelling: New approaches towards transparent grid models,'' *Energy Rep.*, vol. 3, pp. 14–21, Nov. 2017.

[18] F. Wiese, I. Schlecht, W.-D. Bunke, C. Gerbaulet, L. Hirth, M. Jahn, F. Kunz, C. Lorenz, J. Mühlenpfordt, J. Reimann, and W.-P. Schill, ''Open power system data—Frictionless data for electricity system modelling,'' *Appl. Energy*, vol. 236, pp. 401–409, Feb. 2019.

[19] J. Bennett, *OpenStreetMap*. Birmingham, U.K.: Packt, 2010.

[20] C. Josz, S. Fliscounakis, J. Maeght, and P. Panciatici, ''AC power flow data in MATPOWER and QCQP format: ITesla, RTE snapshots, and PEGASE,'' 2016, *arXiv:1603.01533*. [Online]. Available: http://arxiv.org/abs/1603.01533

[21] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, ''Matpower: Steady-state operations, planning and analysis tools for power systems research and education,'' *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[22] S. Fliscounakis, P. Panciatici, F. Capitanescu, and L. Wehenkel, ''Contingency ranking with respect to overloads in very large power systems taking into account uncertainty, preventive, and corrective actions,'' *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4909–4917, Nov. 2013.

[23] 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.

[24] R. D. Christie. (1999). *Power Systems Test Case Archive*. Accessed: Mar. 1, 2021. [Online]. Available: http://www.ee.washington.edu/research/pstca/
[25] S. Soltan and G. Zussman, "Generation of synthetic spatially embedded power grid networks," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Jul. 2016, pp. 1–5.
[26] P. Schultz, J. Heitzig, and J. Kurths, "A random growth model for power grids and other spatially embedded infrastructure networks," *Eur. Phys. J. Special Topics*, vol. 223, no. 12, pp. 2593–2610, Oct. 2014.
[27] Z. Wang, A. Scaglione, and R. J. Thomas, "Generating statistically correct random topologies for testing smart grid communication and control networks," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 28–39, Jun. 2010.
[28] R. Espejo, S. Lumbreras, and A. Ramos, "A complex-network approach to the generation of synthetic power transmission networks," *IEEE Syst. J.*, vol. 13, no. 3, pp. 3050–3058, Sep. 2019.
[29] Z. Wang, R. J. Thomas, and A. Scaglione, "Generating random topology power grids," in *Proc. 41st Annu. Hawaii Int. Conf. Syst. Sci. (HICSS)*, Waikoloa, HI, USA, Jan. 2008, p. 183.
[30] M. H. Athari and Z. Wang, "Statistically characterizing the electrical parameters of the grid transformers and transmission lines," 2017, p. 7, *arXiv:1706.02754*. [Online]. Available: http://arxiv.org/abs/1706.02754
[31] E. Bompard, D. Wu, and F. Xue, "The concept of betweenness in the analysis of power grid vulnerability," in *Proc. Complex. Eng.*, Feb. 2010, pp. 52–54.
[32] B. Liu, Z. Li, X. Chen, Y. Huang, and X. Liu, "Recognition and vulnerability analysis of key nodes in power grid based on complex network centrality," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 65, no. 3, pp. 346–350, Mar. 2018.
[33] J. Hörsch, F. Hofmann, D. Schlachtberger, and T. Brown, "PyPSA-Eur: An open optimisation model of the European transmission system," *Energy Strategy Rev.*, vol. 22, pp. 207–215, Nov. 2018.
[34] K. Purchala, L. Meeus, D. V. Dommelen, and R. Belmans, "Usefulness of DC power flow for active power flow analysis," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, vol. 1, Jun. 2005, pp. 454–459.
[35] E. Katifori, G. J. Szöllősi, and M. O. Magnasco, "Damage and fluctuations induce loops in optimal transport networks," *Phys. Rev. Lett.*, vol. 104, no. 4, Jan. 2010, Art. no. 048704.
[36] M. E. J. Newman, *Networks: An Introduction*. London, U.K.: Oxford Univ. Press, 2010.
[37] F. Kaiser, V. Latora, and D. Witthaut, "Inhibiting failure spreading in complex networks," 2020, *arXiv:2009.02910*. [Online]. Available: http://arxiv.org/abs/2009.02910
[38] P. Van Mieghem, K. Devriendt, and H. Cetinay, "Pseudoinverse of the Laplacian and best spreader node in a network," *Phys. Rev. E*, vol. 96, no. 3, 2017, Art. no. 032311.
[39] F. Dörfler, J. W. Simpson-Porco, and F. Bullo, "Electrical networks and algebraic graph theory: Models, properties, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 977–1005, May 2018.
[40] H. Ronellenfitsch, D. Manik, J. Hörsch, T. Brown, and D. Witthaut, "Dual theory of transmission line outages," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 4060–4068, Sep. 2017.
[41] F. Kaiser and D. Witthaut, "Topological theory of resilience and failure spreading in flow networks," Sep. 2020, *arXiv:2009.10349*. [Online]. Available: http://arxiv.org/abs/2009.10349
[42] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices* (Cambridge Studies in Advanced Mathematics). Cambridge, U.K.: Cambridge Univ. Press, 2009.
[43] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, and A. N. Samukhin, "Spectra of complex networks," *Phys. Rev. E*, vol. 68, no. 4, Oct. 2003, Art. no. 046109.
[44] P. N. McGraw and M. Menzinger, "Laplacian spectra as a diagnostic tool for network structure and dynamics," *Phys. Rev. E*, vol. 77, no. 3, Mar. 2008, Art. no. 031102.
[45] P. Van Mieghem, *Graph Spectra for Complex Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[46] M. Rohden, A. Sorge, M. Timme, and D. Witthaut, "Self-organized synchronization in decentralized power grids," *Phys. Rev. Lett.*, vol. 109, no. 6, Aug. 2012, Art. no. 064101.
[47] E. Limpert, W. A. Stahel, and M. Abbt, "Log-normal distributions across the sciences: Keys and clues," *BioScience*, vol. 51, no. 5, p. 341, 2001.
[48] D. Zwillinger and S. Kokoska, *CRC Standard Probability and Statistics Tables and Formulae*. Boca Raton, FL, USA: CRC Press, Dec. 1999.
[49] P. H. Westfall, "Kurtosis as peakedness, 1905–2014.R.I.P," *Amer. Statistician*, vol. 68, no. 3, pp. 191–195, Jul. 2014.
[50] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman, "Complex dynamics of blackouts in power transmission systems," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 14, no. 3, p. 11, 2004.
[51] S. Mei, F. He, X. Zhang, S. Wu, and G. Wang, "An improved OPA model and blackout risk assessment," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 814–823, May 2009.
[52] 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.
[53] B. A. Carreras, D. E. Newman, and I. Dobson, "North American blackout time series statistics and implications for blackout risk," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4406–4414, Nov. 2016.
[54] S. Biswas and L. Goehring, "Load dependence of power outage statistics," *Europhys. Lett.*, vol. 126, no. 4, p. 44002, Jun. 2019.
[55] D. P. Schlachtberger, T. Brown, S. Schramm, and M. Greiner, "The benefits of cooperation in a highly renewable European electricity network," *Energy*, vol. 134, pp. 469–481, Sep. 2017. [Online]. Available: http://arxiv.org/abs/1704.05492
[56] M. Gathmann. (2019). *World's First Dynamic Control Center in Germany Automates Load Balancing and Outage Prevention*. Accessed: Mar. 1, 2021. [Online]. Available: https://new.siemens.com/global/en/company/stories/infrastructure/2019/ dynamic-power-grid.html
[57] T. Brown, P. Schierhorn, E. Tröster, and T. Ackermann, "Optimising the European transmission system for 77% renewable electricity by 2030," *IET Renew. Power Gener.*, vol. 10, no. 1, pp. 3–9, Jan. 2016.
[58] M. Frysztacki and T. Brown, "Modeling curtailment in germany: How spatial resolution impacts line congestion," in *Proc. 17th Int. Conf. Eur. Energy Market (EEM)*, Sep. 2020, pp. 1–7.
[59] T. Schröder and W. Kuckshinrichs, "Value of lost load: An efficient economic indicator for power supply security? A literature review," *Frontiers Energy Res.*, vol. 3, p. 55, Dec. 2015.
[60] B. Johnson, V. Chalishazar, E. Cotilla-Sanchez, and T. K. A. Brekken, "A Monte Carlo methodology for earthquake impact analysis on the electrical grid," *Electr. Power Syst. Res.*, vol. 184, Jul. 2020, Art. no. 106332.

FRANZ KAISER received the B.Sc. degree in physics from the University of Göttingen, Germany, in 2015, and the M.Sc. degree in physics from the University of Göttingen and with the Max Planck Institute for Dynamics and Self-Organization, Göttingen, in 2018. He is currently pursuing the Ph.D. degree in physics with the Forschungszentrum Jülich and the University of Cologne, Germany.

DIRK WITTHAUT received the M.Sc. and Ph.D. degrees in physics from the Technical University of Kaiserslautern, Kaiserslautern, Germany, in 2004 and 2007, respectively. He worked as a Postdoctoral Researcher with the Niels Bohr Institute, Copenhagen, Denmark, and the Max Planck Institute for Dynamics and Self-Organization, Gottingen, Germany. He has been a Guest Lecturer with the Kigali Institute of Science and Technology, Rwanda. Since 2014, he has been leading a Research Group with the Forschungszentrum Jülich, Germany. He is currently an Assistant Professor with the University of Cologne.