# Polynomial Line Outage Distribution Factors for Estimating Expected Congestion and Security

Cremer, Jochen Lorenz

**DOI**
10.1109/TPWRS.2024.3463410

**Publication date**
2024

**Document Version**
Final published version

**Published in**
IEEE Transactions on Power Systems

# Polynomial Line Outage Distribution Factors for Estimating Expected Congestion and Security

Jochen Lorenz Cremer, *Member, IEEE*

***Abstract*—Extreme weather events and simultaneous $k$ faults pose significant challenges to the security of the power system, leading to sudden line congestion. Conventionally, Line Outage Distribution Factors (LODFs) are used to compute post-fault line flows. However, as $k$ increases, the complexity and number of required matrix inversions make these computations impractical for large systems. This paper introduces a polynomial approximation for LODFs, a method that efficiently combines and multiplies the matrices corresponding to single-line faults using Taylor series expansion. This method is faster than performing matrix inversions for each fault scenario. Moreover, we apply polynomial LODFs to compute expected line flows and enhance probabilistic security, reducing computational demands by decomposing N-$k$ faults into repeating basis functions. Case studies on 118-, 300-, 1354- and 2328-bus systems demonstrate the accuracy and computational superiority of polynomial LODFs in assessing expected congestion and security. These findings are a first step towards managing the reliability and efficiency of power systems in the face of increasing extreme weather events.**

***Index Terms*—Congestion management, line outage distribution factors, power system operation, Taylor series.**

## I. Introduction

THE electricity network transports more renewable energy and increasingly operates at physical limitations, also due to extreme weather hazards [1]. These severe weather hazards may cause more frequent failures of multiple equipment simultaneously or cascading failures [2], [3]. Operators consider seldom multiple failures in operational planning leading to several (partial) power blackouts in the last decades [4]. Currently, the N-1 criterion considers the secure operation when a maximum of one equipment fails at a time. Although the likelihood of other failures and N-$k$ equipment failures can be modelled [5], [6] and inform system-design [7], assessing the impact on the grid of all possible failure combinations during planning and operation is computationally challenging and may trigger instabilities not assessed. Planning with failure-induced congestion [8] and considering the risks of failures [9], [10] shows benefits [11] but requires new computational methods to handle the combinatorial complexity of failures triggered by multiple extreme events.

Power Transfer Distribution Factors (PTDF) and Line Outage Distribution Factors (LODF) are used for assessment, control, market, operational [12], and expansion planning [13]. PTDFs compute the power distribution on transmission lines due to real power transfers between two regions and, for example, can determine transfer limits for energy markets [14]. When a line fails, the power is rerouted over the altered network topology. The PTDFs can compute this rerouting by inverting the large susceptance matrix considering the fault. The Sherman-Morrison-Woodbury formula reduced the dimension of the inverted matrix for low-rank alternations of the PTDFs, leading to speedups when inverting [15]. The LODFs compute the real power distribution over the network in response to multiple line outages (N-$k$ with $k > 1$) [16], and recently were modified to consider bus voltages [17]. However, it remains challenging to compute system properties for many possible faults, in a grid with $N$ assets, where multiple simultaneous faults $k$ occur, as these methods require individual computational steps for each line fault combination.

Two types of approaches address this computational challenge, either by reducing the number of contingency scenarios or by speeding up individual scenarios. The first type of approach reduces the large number of contingency scenarios, based on electrical distance [18], identifying high-risk contingencies [19], [20], contingency ranking [21], probabilistic assessments [22], machine learning [23], [24] or using motifs for cascading failures [25]. The limitation of the first type of approach is that some contingency (scenarios) may be missed. [18] identifies vulnerable $N-3$ fault combinations using the electrical distance to avoid solving power flows. However, this approach may miss weather-induced multiple faults as these don't necessarily correlate with electrical distance, and the approach does not consider all fault combinations. [19] maintains a high-risk contingency list where the total number of all possible contingencies is limited to a number linearly proportional to the size of the system, and not all risks can be considered. [20] searches for the worst N-$k$ contingencies in a contingency list by incrementally running simulations in the time domain on the most stressed equipment. Approaches that identify N-$k$ contingencies using a linear search type may miss potentially dangerous combinations of contingencies and are not very suitable when the expectation or risk-based assessment of all N-$k$ contingencies is of interest, as these would require evaluating the impact of all combinations beyond those in

Received 22 February 2024; revised 18 June 2024 and 29 August 2024; accepted 16 September 2024. Date of publication 18 September 2024; date of current version 21 April 2025. This work was supported by the Dutch Research Council, Veni Talent Program under Grant 19161. Paper no. TPWRS-00325-2024.

The authors is with the Department of Electrical Sustainable Energy, Delft University of Technology, Delft, 2628 CD Delft, The Netherlands, and also with the Center for Energy, Austrian Institute of Technology, 1210 Vienna, Austria (e-mail: J.L.Cremer@tudelft.nl).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRS.2024.3463410.

Digital Object Identifier 10.1109/TPWRS.2024.3463410

the search. [21] identifies the worst-case contingency out of N-$k$ contingencies based on the risk and the load loss, and recursively ranks the worst contingencies and assesses their severity (load loss). However, identifying all severities requires solving one optimization problem per contingency scenario, which is a large number. [22] proposes a probabilistic risk-based assessment of the contingencies; however, the severity must be individually calculated for each contingency scenario and, therefore, does not scale well for exploding combinations of N-$k$ faults. [23] ranks contingencies based on probabilistic scores from machine-learned classifiers considering only N-1 contingencies, however, does not also scale well to N-$k$ contingencies. [24] reduces the total number of cascading contingencies and works well for locally propagating cascading failures as a graph convolutional neural network is used, however, the approach does not consider all possible weather-induced N-$k$ contingencies. To include a more exhaustive list of N-$k$ contingencies, [25] identifies motifs capturing the likelihood of a combination of N-$k$ failures improving the list. However, reducing N-$k$ failures to a subset of contingencies through the above-mentioned approaches does not arguably result in the highest accurate estimation when used to compute the expected congestion and expected security as the full set of N-$k$ contingencies would need to be considered and not a reduced 'worst-case' subset.

The second type of approach speeds up the assessment of the impact (severity) of the contingencies so that more contingencies can be assessed with a given computational budget. The foundation for many approaches that perform impact assessment of static security are the LODFs relying on matrix inversions. The Sherman-Morrision Woodbury formula [26] can efficiently update the LODFs when $k$ is low [27], and even faster for $k = 2$ [28]. However, [27] and [28] may not be most suitable when $k$ is greater than 2 when operating a resilient grid. Considering the network cycles in PTDFs and LODFs showed numerical speedups for small cycles [29]. However, [29] was not specifically developed to consider N-$k$ contingencies, and power systems may have many loops. Despite these speedups, many approaches rely on computing the severity (impact) of the combination of the contingencies with the support of the LODF matrices for each N-$k$ contingency which is limiting. This limit means that considering a large number of N-$k$ contingencies for operational planning is currently not yet possible. For example, risk analysis, security assessment or operational constraint programming can not consider risks of 'all' N-$k$ faults yet and most of the time operates on a subset of selected contingency scenarios as the computational time of considering the combination of contingencies is too large.

As pointed out earlier, the required computational time relates to the matrix inversion of the LODFs for each combination. Matrix inversion iteratively finds a matrix through an algorithm such as Gaussian elimination; however, iterations render the approach unsuitable for several uses. For example, computing risks may require Riemann integrals over a function defining the severity of faults. However, the differentiability of an algorithm that inverts matrices is not specified; hence, no integrals could be easily computed. Additionally, the computation of statistical moments or applying Chebyshev's inequality (as in [30]) and using Bernoulli distributions of the faults for large $k$ would require a function to measure the fault's impacts instead of an iterative or recursive function. [31] consider distributionally robust reliability assessment considering N-$k$ security. Their work and the computing of preventive and corrective control actions often rely on constraint programming. However, constraint programming can not efficiently consider N-$k$ based LODFs to include security constraints [21], [32], as these programs would need an inverted matrix for each fault and a large number of constraints slows these computations.

This paper proposes polynomial LODFs for faults that are far from each other. The proposed method does not rely on inverting matrices for the combination of faults but adds and multiplies matrices for the single N-1 line faults (such as using them as basis functions). The proposed polynomial LODFs approach the Sherman-Morrision-Woodbury update with two Taylor series approximations and linearly combine multiple outages. This paper then develops use cases to compute expected congestion and security assessment. There, the proposed decomposition keeps the computations constant with $k$. The approach decomposes large summation and multiplication into basis terms appearing multiple times, saving significant computational time. The two contributions are

1) polynomial LODFs for N-$k$ faults that add and multiply N-1 matrices. These LODFs show high accuracy in faults far away from each other.
2) algorithm to compute expected line flows. The computations do not grow with the number of simultaneous faults.

Case studies use the 118-, 300-, 1354- and 2328-bus systems. The study investigates the approximation error of the proposed polynomial LODF compared to the Sherman-Morrision Woodbury LODF baseline for multiple line fault combinations $k$. The study also analyzes the convergence of the Taylor series and how to select the number of Taylor components. The study then assesses the impact of fault proximity on the approximation error and convergence. The study finally investigates the advantages of computing expected congestion and security assessment. Sec. II presents LODFs, and Sec. III the proposed polynomial LODFs. Sec. IV develops the polynomial LODFs for a use case, computing expected congestion. Sec. V studies, and VI concludes.

## II. Line Outage Distribution Factors

The power network has a set of buses $\Omega^B$ and lines $\Omega^L$. The cardinalities $|\Omega^B|$ and $|\Omega^L|$ are the number of buses and lines. $A \in \{-1, 0, 1\}^{|\Omega^L|\times|\Omega^B|}$ is the branch incidence matrix where 1 and $-1$ represent the 'from bus' and 'to bus' of a connecting line, respectively. Zero represents that a line does not connect to a bus. As in [33], we start by building the susceptance matrix $B \in \mathbb{R}^{|\Omega^B|\times|\Omega^B|}$. Line resistances are neglected. Hence, $B$'s off-diagonal and diagonal elements of $B$ are only computed with the line reactances $\chi_l$ for the lines $l \in \Omega^L$ by

$$B_{\tilde{b},b} = \sum_{l\in\Omega^L} A_{l,b}A_{l,\tilde{b}}\frac{1}{\chi_l} \qquad \forall b, \tilde{b} \in \Omega^B \qquad (1a)$$

$$B_{b,b} = -\sum_{l\in\Omega^L} A_{l,b}A_{l,b}\frac{1}{\chi_l} \qquad \forall b \in \Omega^B \tag{1b}$$

The row and column are set 0 at the slack bus $\hat{b}$

$$B_{\hat{b},b} = 0 \qquad \forall b \in \Omega^B \tag{2a}$$

$$B_{b,\hat{b}} = 0 \qquad \forall b \in \Omega^B \tag{2b}$$

We then compute the Moore-Penrose pseudoinverse matrix $X = B^{-1}$. The power transfer distribution factors are

$$PTDF = B_{br} \times A \times X \tag{3}$$

where the diagonal matrix $B_{br} \in \mathbb{R}^{|\Omega^L|\times|\Omega^L|}$ has the inverse line reactance values $\frac{1}{\chi_l}$ in the diagonal entries, and all other entries are zero. $B$ and $B_{br}$ are two different matrices.

### A. Challenge of Computing N-$k$ Line Outage Power Distribution

The matrix $\tilde{B} = B + D$ describes the susceptance matrix of the power network after a perturbation $D$. The perturbed matrix $D \in \mathbb{R}^{|\Omega^B|\times|\Omega^B|}$ can show the change in the susceptance matrix when lines are in an outage. The matrix $D$ has the elements

$$D_{\tilde{b},b} = \sum_{l\in\Omega^L} -y_l A_{l,b}A_{l,\tilde{b}}\frac{1}{\chi_l} \qquad \forall b,\tilde{b} \in \Omega^B \tag{4a}$$

$$D_{b,b} = -\sum_{l\in\Omega^L} -y_l A_{l,b}\frac{1}{\chi_l} \qquad \forall b,\tilde{b} \in \Omega^B \tag{4b}$$

$$D_{\hat{b},b} = 0, \qquad D_{b,\hat{b}} = 0 \qquad \forall b \in \Omega^B \tag{4c}$$

where $y_l = \{0,1\}$ shows the lines in outage. $y_l = 0$ means the line is intact and $y_l = 1$ means that the line is out of power. The set of lines simultaneously in an outage is $\Omega^{\hat{L}} = \{l \,|\, y_l = 1 \wedge l \in \Omega^L\}$ and the total number of lines in the outage is $k = \sum_{l\in\Omega^L} y_l = |\Omega^{\hat{L}}|$.

Challenging is to compute

$$\tilde{B}^{-1} = (B + D)^{-1} \tag{5}$$

for all possible combinations of $y_l$. Computing these matrices for $k$ line fault combinations requires

$$M(k) = \sum_{\hat{k}=1\ldots k} \frac{|\Omega^L|!}{\hat{k}!(|\Omega^L| - \hat{k})!} \tag{6}$$

matrix inversions. These computations are in big-$\mathcal{O}$ notation $\mathcal{O}(\frac{|\Omega^L|^k}{k!})$ and become a burden for memory and computational times for large systems. This becomes challenging when operators aim to study many fault scenarios simultaneously, for example, to study the expected impact of large extreme events that could possibly lead to many N-$k$ fault combinations. Current methods would require assessing many power flows to compute the risks and expectations of the power flows post-fault scenarios. Although a single metric, like risks, expected congestion or expected security is desired, the assessment of many $M(k)$ possible scenarios is needed which represents a limitation of current methodologies.

### B. Line Outage Distribution Factors

LODFs use the Woodbury matrix identity [26] for (5). This identity

$$(B + UIV)^{-1} = B^{-1} - B^{-1}U\left(I + VB^{-1}U\right)^{-1}VB^{-1} \tag{7}$$

updates the inverse of the original matrix $B$ changed by $D = UIV$ where $I$ is the identity matrix with dimensions $|\Omega^B| \times |\Omega^B|$. We define the matrix

$$U = I + \sum_{l\in\Omega^{\hat{L}}} U_l \tag{8a}$$

$$U_l = -e_{b_l^t}e_{b_l^t}^T \tag{8b}$$

where $b_l^t$ is the 'to' bus, $b_l^f$ is the 'from' bus of the line $l$. All lines in outages are $\Omega^{\hat{L}}$, hence $|\Omega^{\hat{L}}| = k$. $e_i$ is the standard basis vector, which means that the $i$ th element is one, and all other elements are zero. The matrix $V$ is

$$V = \sum_{l\in\Omega^{\hat{L}}} V_l \tag{9a}$$

$$V_l = -\frac{1}{\chi_l}e_{b_l^f}e_{b_l^f}^T \tag{9b}$$

The computational advantage of performing this Woodbury update and inverting $(I + VB^{-1}U)^{-1}$ over directly inverting $(B + UIV)^{-1}$ is when $(I + VB^{-1}U)$ has significantly fewer dimensions than $(B + UIV)$, in other words, when the update is of low-rank. Then, the N-$k$ LODF matrix is

$$LODF_{\text{N}-k} = B_{br} \times A \times (B + UIV)^{-1}A^T \tag{10}$$

that exactly contains the PTDF (3), is no approximation and is exact. The LODFs (10) using Woodbury matrix update (7) compute faster than PTDFs (3) which use the Moore-Penrose pseudoinverse of the full matrix $X$. However, a large number of matrix inversions are needed ($M(k)$) as (6) shows.

## III. Polynomial Approximation of N-$k$ LODFs

The following derives the polynomial approximation

$$\tilde{B}^{-1} \approx h\left(y_l,\, g(V_l, U_l)^{-1}\right) \tag{11}$$

assuming that $\tilde{B}^{-1}$ is non-singular. $g(V_l, U_l)^{-1}$ inverts the function $g$ based on the basis matrices $U_l$ and $V_l$ for each line $\forall\, l \in \Omega^L$. Note, $g$ considers $U_l$ and $V_l$ of only one line $l$ as input at a time and no combinations of lines. Therefore, no matrices corresponding to the combination of faults need inversion, which is the idea of this work. $h$ sums and multiplies the inverted matrices (where $y_l = 1$) to consider fault combinations. The objective of this work is to derive $g$ and $h$ so that the post-fault flows are superposing the power flows from multiple line failures. This proposed polynomial approximation requires only $|\Omega^L|$ matrix inversions of individual lines, and avoids the $M$ matrix inversions of all possible fault combinations, hence the number of inversions is largely reduced ($|\Omega^L| \ll M$).

### A. Approximating N-k LODFs Through Additives

This section derives the approximation

$$\left(I+VB^{-1}U\right)^{-1} \approx \sum_{l\in\Omega^{\hat{L}}} \left(I+V_lB^{-1}U\right)^{-1} - (k-1)I + \varepsilon \tag{12}$$

for the part that needs to be inverted in (7). $\varepsilon$ considers all higher-order terms. We write

$$\left(I+VB^{-1}U\right)^{-1} = \left(I+\sum_{l\in\Omega^{\hat{L}}} V_lB^{-1}U\right)^{-1} \tag{13}$$

and define

$$G_l := I + V_lB^{-1}U \tag{14a}$$

$$H_l := V_lB^{-1}U = V_lB^{-1}\left(I+\sum_{\hat{l}\in\Omega^{\hat{L}}} U_{\hat{l}}\right) \tag{14b}$$

$$\begin{aligned}\mathcal{H}_l &:= \sum_{\hat{l}\in\Omega^{\hat{L}}\setminus l} H_{\hat{l}} \\ &= \sum_{\hat{l}\in\Omega^{\hat{L}}\setminus l} V_{\hat{l}}B^{-1}I + \sum_{\hat{l}\in\Omega^{\hat{L}}\setminus l}\sum_{\bar{l}\in\Omega^{\hat{L}}} V_{\hat{l}}B^{-1}U_{\bar{l}}\end{aligned} \tag{14c}$$

where $G_l = I + H_l$. With (13) and (14), we obtain

$$\left(I+VB^{-1}U\right)^{-1} = \left(G_l+\sum_{\hat{l}\in\Omega^{\hat{L}}\setminus l} H_{\hat{l}}\right)^{-1} \tag{15}$$

We develop the Taylor series (the literature references Neumann series [34] when approximating linear matrix operation with the analogous series) around $G_l$ for any line $l\in\Omega^{\hat{L}}$. $\mathcal{H}_l$ is considered a perturbation of $G_l$. The inverse of the perturbed matrix can be expressed as the convergent series

$$\begin{aligned}&\left(I+VB^{-1}U\right)^{-1} \\ &\quad\approx G_l^{-1} - G_l^{-1}\mathcal{H}_lG_l^{-1} + \left(G_l^{-1}\mathcal{H}_l\right)^2 G_l^{-1} - \dots\end{aligned} \tag{16}$$

iff $||G_l^{-1}\mathcal{H}_l|| < 1$. This application of series expansion for matrix inversion with perturbations can be found in linear algebra and numerical analysis literature, for example in [35, chapter 9]) or matrix perturbation theory in [36, chapter III, section 2]. The error of this series has bounds. When truncating the series after the $t_1$ terms, the error is

$$\begin{aligned}E_{t_1} =\ &(G_l+\mathcal{H}_l)^{-1} \\ &- \left(G_l^{-1} - \dots + (-1)^{t_1}\left(G_l^{-1}\mathcal{H}_l\right)^{t_1} G_l^{-1}\right)\end{aligned} \tag{17}$$

and the error bound is

$$||E_{t_1}|| \le \frac{||G_l^{-1}|| \times ||(G_l^{-1}\mathcal{H}_l)^{t_1+1}||}{1-||G_l^{-1}\mathcal{H}_l||} \tag{18}$$

where $(G_l^{-1}\mathcal{H}_l)^{t_1+1}$ is the next term in the series. $V_l$ and $G_l$ correspond to the outage of the line $l$ changing the power from the bus $b_l^f$ to all other buses represented by one row. Intuitively, this series around $G_l$ may converge faster at the $b_l^f$th row than in the other rows as Taylor approximates locally. Therefore, one could linearly combine the Taylor series for all lines in outage $\Omega^{\hat{L}}$ by using $(e_{b_l^f}e_{b_l^f}^T)$ in

$$\begin{aligned}\left(I+VB^{-1}U\right)^{-1} \approx \sum_{l\in\Omega^{\hat{L}}} &\left(G_l^{-1} - (e_{b_l^f}e_{b_l^f}^T)G_l^{-1}\mathcal{H}_lG_l^{-1}\right) \\ &- (k-1)I\end{aligned} \tag{19}$$

with the first two components ($t_1 = 2$) of the Taylor series.

### B. Deriving First Order Impact

To compute the inverse $G_l^{-1}$, the 'first order impact' of the line outage $l$, we use (8a) and the distributive property

$$\begin{aligned}G_l^{-1} = (I+V_lB^{-1}U)^{-1} &= \left(I+V_lB^{-1}\left(I+\sum_{l\in\Omega^{\hat{L}}} U_l\right)\right)^{-1} \\ &= \left(I+V_lB^{-1}I+V_lB^{-1}U_l+\sum_{\hat{l}\in\Omega^{\hat{L}}\setminus l} V_lB^{-1}U_{\hat{l}}\right)^{-1}\end{aligned} \tag{20}$$

We define

$$E_l := I + V_lB^{-1}I + V_lB^{-1}U_l \tag{21a}$$

$$F_l := \sum_{\hat{l}\in\Omega^{\hat{L}}\setminus l} V_lB^{-1}U_{\hat{l}} \tag{21b}$$

and expand with Taylor around $E_l$ with $t_2$ components, where we consider $F_l$ a perturbation on $E_l$, hence

$$G_l^{-1} \approx E_l^{-1} - E_l^{-1}F_lE_l^{-1} + E_l^{-1}F_lE_l^{-1}F_lE_l^{-1}\dots \tag{22}$$

which converges iff $||E_l^{-1}F_l|| < 1$. $E_l^{-1}$ is the Moore-Penrose pseudoinverse. Error bounds can be derived in a similar way as (17)–(18).

Exemplary, we write the approximation (with Taylor components $t_1 = 2$ and $t_2 = 1$) as follows

$$\begin{aligned}\tilde{B}^{-1} \approx B^{-1} &- \sum_{\hat{l}\in\Omega^{\hat{L}}}\sum_{l\in\Omega^{\hat{L}}} B^{-1}IE_l^{-1}V_{\hat{l}}B^{-1} \\ &- \sum_{\check{l}\in\Omega^{\hat{L}}}\sum_{\hat{l}\in\Omega^{\hat{L}}}\sum_{l\in\Omega^{\hat{L}}} B^{-1}U_{\check{l}}E_l^{-1}V_{\hat{l}}B^{-1} \\ &+ \sum_{\bar{l}\in\Omega^{\hat{L}}}\sum_{l\in\Omega^{\hat{L}}}\sum_{\hat{l}\in\Omega^{\hat{L}}\setminus l} B^{-1}I\left(e_{b_l^f}e_{b_l^f}^T\right) \\ &\times E_l^{-1}V_{\hat{l}}B^{-1}IE_l^{-1}V_{\bar{l}}B^{-1} \\ &+ \sum_{\tilde{l}\in\Omega^{\hat{L}}}\sum_{l\in\Omega^{\hat{L}}}\sum_{\hat{l}\in\Omega^{\hat{L}}\setminus l}\sum_{\bar{l}\in\Omega^{\hat{L}}} B^{-1}I\left(e_{b_l^f}e_{b_l^f}^T\right) \\ &\times E_l^{-1}V_{\hat{l}}B^{-1}U_{\tilde{l}}E_l^{-1}V_{\bar{l}}B^{-1} \\ &+ \sum_{\check{l}\in\Omega^{\hat{L}}}\sum_{\tilde{l}\in\Omega^{\hat{L}}}\sum_{l\in\Omega^{\hat{L}}}\sum_{\hat{l}\in\Omega^{\hat{L}}\setminus l} B^{-1}U_{\check{l}}\left(e_{b_l^f}e_{b_l^f}^T\right) \\ &\times E_l^{-1}V_{\hat{l}}B^{-1}IE_l^{-1}V_{\bar{l}}B^{-1} \\ &+ \sum_{\check{l}\in\Omega^{\hat{L}}}\sum_{\tilde{l}\in\Omega^{\hat{L}}}\sum_{l\in\Omega^{\hat{L}}}\sum_{\hat{l}\in\Omega^{\hat{L}}\setminus l}\sum_{\bar{l}\in\Omega^{\hat{L}}} B^{-1}U_{\check{l}}\left(e_{b_l^f}e_{b_l^f}^T\right)\end{aligned}$$

$$\times E_l^{-1} V_{\hat{l}} B^{-1} U_{\bar{l}} E_l^{-1} V_{\bar{l}} B^{-1} + \sum_{l\in\Omega^{\hat{L}}} B^{-1} I(k-1) I V_l B^{-1} + \sum_{\bar{l}\in\Omega^{\hat{L}}} \sum_{l\in\Omega^{\hat{L}}} B^{-1} U_{\bar{l}} (k-1) I V_l B^{-1} \quad (23)$$

This equation approximates (7) by substituting $(I + VB^{-1}U)$ with (19) truncating after $t_1 = 2$ components, and subsequently substituting $G_l^{-1}$ with (22) truncating after the first component $t_2 = 1$. Then, use the distributive property to multiply all obtained terms with $VB^{-1}$. Finally, after reorganising the terms, we get (23). Evaluating (23) requires to pre-compute $E_l^{-1}$, $V_l$ and $U_l$ for all lines $\Omega^L$.

Note that the proposed approximation of the N-$k$ LODFs in (7), (19), (22) requires only $k = |\Omega^{\hat{L}}|$ matrix inversions and $k \ll M$. The state-of-the-art approaches require $M$ inversions of N-$k$ matrices as shown in (6). In this work, the only matrices to (pseudo-) invert are $E_l$ of the single outages $l \in \Omega^{\hat{L}}$, which was the objective to derive in this section. The same (pseudo-) inverted matrices can be used for another fault combination. However, our work requires several multiplications and additions and may not be faster than $M$ inversions. The calculations also increase with the number of Taylor series components $t$ that approximate $LODF^t_{\text{N}-k}$ in (16) and (22).

## IV. Expected Congestion and Security

### A. Security for k Faults

The equations

$$f^t = LODF^t_{\text{N}-k} \times f^0 \quad (24a)$$

$$f^t_l = 0 \qquad \forall l \in \Omega^{\hat{L}} \quad (24b)$$

compute the post-fault line flows $f^t$ given the pre-fault line flows $f^0 \in \mathbb{R}^{|\Omega^L|}$ for the lines in fault. The physical limits on the transmission lines are

$$-f_l^B \le f_l^t \le f_l^B \qquad l \in \Omega^L \quad (25)$$

where $f_l^B$ is the limit of the power flow over line $l$. This paper computes only line overloads and does not provide a methodology to compute voltages. The system is here considered statically secure for the N-$k$ fault if $f_l$ are constrained by (25) for all lines $\forall l \in \Omega^L$.

### B. Expected Congestion At Faults

The proposed polynomial estimates the expected line flow considering the N-$k$ grid given outage probabilities. Such probabilities may be estimated from historical observations and line dependencies [37]. We consider $\mathbb{P}(y_l = 1)$ as the marginal probability that the line $l$ is in an outage considering the Bernoulli distribution [38], e.g.

$$\mathbb{P}(y_l) = \begin{cases} r & \text{if } y_l = 1 \\ 1 - r & \text{if } y_l = 0 \end{cases} \quad (26)$$

where $r$ is the probability that the line is in an outage. The probability of two outages occurring can be expressed with the Bayes'rule as

$$\mathbb{P}(y_l = 1 \cap y_{\hat{l}} = 1) = \mathbb{P}(y_{\hat{l}} = 1)\mathbb{P}(y_l = 1 | y_{\hat{l}} = 1) \quad (27)$$

The chain rule gives the joint probability distribution

$$\mathbb{P}\left(\bigcap_{l\in\Omega^{\hat{L}}} y_l = 1\right) = \prod_{l\in\Omega^{\hat{L}}} \mathbb{P}\left(y_l = 1 \,\middle|\, \bigcap_{\hat{l}\in\Omega^{\hat{L}}\setminus l} y_{\hat{l}} = 1\right) \quad (28)$$

through the continuous application of Bayes'rule. Assuming independent line faults, then

$$\phi_{\Omega^{\hat{L}}} := \mathbb{P}\left(\bigcap_{l\in\Omega^{\hat{L}}} y_l = 1\right) = \prod_{l\in\Omega^{\hat{L}}} \mathbb{P}(y_l = 1), \quad (29)$$

where $|\Omega^{\hat{L}}| = k$ correspond to $k$ faults at the same time. $\Omega^O$ are the possible lines where $k$ outages may occur, where $|\Omega^O| \le |\Omega^L|$. $\Omega_c^{\hat{L}} \subset \Omega^O$ is one $k$-fault sample (realisation) $c$, where $c = 1, 2, \ldots M(k)$ with $M(k)$ from (6). Using (3)–(5) and (24), we can write the expected line flows as

$$\mathbb{E}[f_l] = B_{br}\, A\, \mathbb{E}[\tilde{B}^{-1}]\, A^T\, f^0 \qquad \forall l \in \Omega^L \setminus \Omega^O \quad (30)$$

This expectation considers only the lines not considered as possible outage ($\Omega^L \setminus \Omega^O$) as the (24b) is not considered. Considering $\hat{k} = 1, 2, \ldots k$ fault combinations and assuming the independence of faults, we can write this expectation in the state-of-the-art approach as

$$\mathbb{E}[f_l] = \sum_{c=1,2,\ldots M(k)} \phi_{\Omega_c^{\hat{L}}}\, B_{br}\, A\, \tilde{B}_c^{-1}\, A^T\, f^0 \quad \forall l \in \Omega^L \setminus \Omega^O \quad (31)$$

where $\tilde{B}_c^{-1}$ considers the lines in fault $\Omega_c^{\hat{L}}$ that has $\hat{k}$ simultaneous faults ($|\Omega_c^{\hat{L}}| = \hat{k}$). As pointed out earlier in (6), using the standard LODFs to compute $\tilde{B}_c^{-1}$ requires many $(M(k))$ inversions of the matrices $(\tilde{B}_c^{-1})$.

### C. Proposed Decomposed Probabilistic Assessment

We consider $t_1 = 1$, $t_2 = 1$ and $\Omega_c^{\hat{L}}$, and derive

$$\tilde{B}_c^{-1} \approx B^{-1} + (\hat{k} - 1) \sum_{l\in\Omega_c^{\hat{L}}} B^{-1} I V_l B^{-1} - \sum_{\hat{l}\in\Omega_c^{\hat{L}}} \sum_{l\in\Omega_c^{\hat{L}}} \left(B^{-1} I E_l^{-1} V_{\hat{l}} B^{-1} - (\hat{k} - 1) B^{-1} U_{\hat{l}} I V_l B^{-1}\right) - \sum_{\bar{l}\in\Omega_c^{\hat{L}}} \sum_{\hat{l}\in\Omega_c^{\hat{L}}} \sum_{l\in\Omega_c^{\hat{L}}} B^{-1} U_{\bar{l}} E_l^{-1} V_{\hat{l}} B^{-1} \quad (32)$$

in a similar way as (23). The right-hand side of (32) has terms that simultaneously involve 1, 2 or 3 lines from the lines in outage $\Omega_c^{\hat{L}}$. These terms are computed offline

$$\mathcal{P}_{1,l} := B^{-1} I V_l B^{-1} \quad (33a)$$

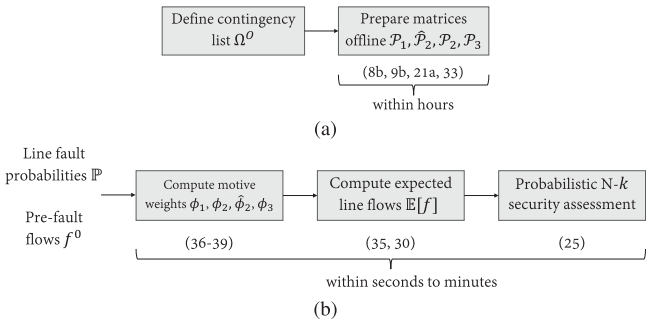



Fig. 1. Offline (a) and near-realtime (b) workflows of the proposed probabilistic security assessment.

$$\hat{\mathcal{P}}_{2,l,\hat{l}} := B^{-1} I E_l^{-1} V_{\hat{l}} B^{-1} \tag{33b}$$

$$\mathcal{P}_{2,\hat{l},l} := B^{-1} U_{\hat{l}} I V_l B^{-1} \tag{33c}$$

$$\mathcal{P}_{3,\check{l},l,\hat{l}} := B^{-1} U_{\check{l}} E_l^{-1} V_{\hat{l}} B^{-1} \tag{33d}$$

involving 1, 2, 2, or 3 lines in outage, respectively. These terms are computed well ahead of operation as shown in the offline workflow Fig. 1(a).

When assessing security for many combinations of $\hat{k}$ faults, the terms involving specific motifs of $\check{k} = 1, 2$, or 3 lines in an outage appear

$$N_{\check{k},\hat{k}} = \frac{(|\Omega^O| - \check{k})!}{(\hat{k} - \check{k})!(|\Omega^O| - \hat{k})!} \tag{34}$$

times. For example, a sample of $\check{k} = 2$ distinct lines could be in outage with another line ($\hat{k} = 3$ outage) or with two completely different lines ($\hat{k} = 4$ outage). Each $\check{k} = 1$ line motifs with the term $\mathcal{P}_{1,l}$ appears $N_{1,\hat{k}}$ times in all $\hat{k}$-fault combinations of the lines $\Omega^O$. For example, a single line outage appears $N_{1,2} = |\Omega^O| - 1$ times in all possible $\hat{k} = 2$ faults. The $\check{k} = 2$ line motifs of lines $l$ and $\hat{l}$ with the terms $\mathcal{P}_{2,l,\hat{l}}$ and $\hat{\mathcal{P}}_{2,l,\hat{l}}$ appear $N_{2,\hat{k}}$ times and the $\check{k} = 3$ line motifs with term $\mathcal{P}_{3,\check{l},l,\hat{l}}$ appears $N_{3,\hat{k}}$ times.

The proposed probabilistic security assessment exploits this repetitive appearance and decomposes the probabilistic assessment in motif-specific basis terms. The near realtime workflow (as shown in Fig. 1(b)), computes the expected LODFs with

$$\begin{aligned}\mathbb{E}\left[\tilde{B}^{-1}\right] \approx B^{-1} &+ \sum_{l\in\Omega^O} \phi_{1,l}\mathcal{P}_{1,l} \\ &+ \sum_{\hat{l}\in\Omega^O}\sum_{l\in\Omega^O} \phi_{2,l,\hat{l}}\mathcal{P}_{2,\hat{l},l}, -\hat{\phi}_{2,l,\hat{l}}\hat{\mathcal{P}}_{2,l,\hat{l}} \\ &- \sum_{\check{l}\in\Omega^O}\sum_{\hat{l}\in\Omega^O}\sum_{l\in\Omega^O} \phi_{3,l,\hat{l},\check{l}}\mathcal{P}_{3,\check{l},l,\hat{l}}\end{aligned} \tag{35}$$

considering the lines $\Omega^O$. Note that this summation considers the individual lines $\Omega^O$ and not all fault-combinations as in the state-of-the-art (31) that sums all combinations of faults $M(k)$, and note that $|\Omega^O| \ll M(k)$. $\phi_{1,l}$, $\phi_{2,l,\hat{l}}$, $\hat{\phi}_{2,l,\hat{l}}$ and $\phi_{3,l,\hat{l},\check{l}}$ are weights considering the probability that the motifs appear in $k$ fault combinations, as illustrated in Fig. 2. The weights are a function of the probabilities $\mathbb{P}_l$ and the frequency with which

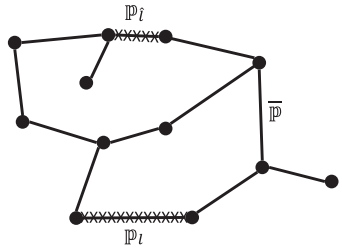



Fig. 2. A motif of two lines $l$ and $\hat{l}$ may be part of multiple different $k$-fault combinations. This motif may appear with a probability of $\mathbb{P}_l\mathbb{P}_{\hat{l}}\overline{\mathbb{P}}N_{2,3}$ in an $k = 3$ outage.

the motifs appear in the combination of faults ($N_{1,k}$, $N_{2,k}$, and $N_{3,k}$). We compute the weights as

$$\phi_{1,l} = \mathbb{P}_l \left(1 + \sum_{\hat{k}=2,\dots k} \overline{\mathbb{P}}^{\hat{k}-1}(\hat{k}-1)N_{1,\hat{k}}\right) \tag{36}$$

$$\phi_{2,l,\hat{l}} = \begin{cases} \mathbb{P}_l\mathbb{P}_{\hat{l}}\left(1 + \sum_{\hat{k}=3,\dots k} \overline{\mathbb{P}}^{\hat{k}-2}(\hat{k}-1)N_{2,\hat{k}}\right) & \text{if } l \neq \hat{l} \\ \mathbb{P}_l & \text{if } l = \hat{l} \end{cases} \tag{37}$$

$$\hat{\phi}_{2,l,\hat{l}} = \begin{cases} \mathbb{P}_l\mathbb{P}_{\hat{l}}\left(1 + \sum_{\hat{k}=3,\dots k} \overline{\mathbb{P}}^{\hat{k}-2}N_{2,\hat{k}}\right) & \text{if } l \neq \hat{l} \\ \mathbb{P}_l & \text{if } l = \hat{l} \end{cases} \tag{38}$$

and

$$\phi_{3,l,\hat{l},\check{l}} = \begin{cases} \mathbb{P}_l\mathbb{P}_{\hat{l}}\mathbb{P}_{\check{l}}(1 + \sum_{\hat{k}=4,\dots k} \overline{\mathbb{P}}^{\hat{k}-3}N_{3,\hat{k}}) & \text{if } l \neq \hat{l} \\ \mathbb{P}_{\check{l}}\phi_{2,l,\hat{l}} & \text{if } l = \check{l} \vee \hat{l} = \check{l} \\ \mathbb{P}_l & \text{if } l = \check{l} = \hat{l} \end{cases} \tag{39}$$

assuming the average outage probability

$$\overline{\mathbb{P}} = \frac{1}{|\Omega^O|}\sum_{l\in\Omega^O} \mathbb{P}(y_l = 1) \tag{40}$$

Subsequently, we can compute the expected line loading (30) using the expected LODFs from (35). Assessing the physical limits with (25) for the expected line flows provides the probability of whether the system is secure.

The primary benefit of this proposed approximation to probabilistic security assessment is that the computation of (35) has the complexity $\mathcal{O}(|\Omega^O|^3)$ (note the three nested sums in (35)) and, therefore, is independent of $k$. $k$ only appears in the scalar weights computed in (36)–(39). An additional advantage is that these weights are easily re-computed when the probabilities of lines $\mathbb{P}_l$ change.

## V. Case Study

The case study is on the 118-, 300-, 1354- and 2383-bus systems. The study tests $k = \{1, 2, \dots, 7\}$ fault combinations on the IEEE 118-bus system. 10 000 combinations were randomly sampled for each $k$ fault combination (total 70 000). The Taylor

series were tested for components $t = \{1, 2, \dots 10\}$. The inverses of matrices were computed with Moore-Penrose method with default tolerances in MATLAB and SCIPY. The Dijkstra method for the shortest paths [39] and the effective Thevenin impedance [40] quantified the 'proximity' of the faults. The pre-fault power flows $f^0$ were the active [MW] or apparent powers [MVA], respectively. The AC power flow (using Newton Raphson), DC power flow, and the DC Optimal Power Flow (OPF) computed pre-fault flows (and post-fault flows as baselines). Samples with post-fault flows that were unreasonably high were not considered, e.g. more than 200 times higher than the pre-fault flows; these resulted from islanding, system splits, or nonconvergence of the Taylor series. The errors to the baseline AC power flow were calculated on 1000 fault combinations for each $k$. The security was assessed for 1000 load power levels sampled around $\pm 0.25$ around the nominal load with DC assumptions. The samples were drawn from a Kumaraswamy(1.6,2.8) distribution with a Pearson correlation coefficient of 0.75. The 10 000 fault samples for each $k$-fault combination were $10\,000 \times 1000$ were 10 million fault samples. In the IEEE 118-bus system, 187 lines were assessed for static security whether they meet line limits, totalling $10\,000 \times 1000 \times 187 \times 7$ security assessment labels for each $k$ fault combination, where $k = \{1, 2, \dots 7\}$. An insecure, 'positive' sample is when the post-fault line flows are over the physical limitations, otherwise, the sample is considered 'negative'. Different metrics quantified the performance of estimating security with the proposed polynomial LODFs. Two types of accurate estimations (predictions) can occur, one is the true positives and the other is the true negatives. The true positive rate is the accurately predicted positive samples of all true positive labels. The true negative rate is the accurately predicted negative samples of all true negative labels. The precision is the rate of the true positives out of all, predicted, positive labels. The negative predictive value is the rate of the true negatives out of all, predicted, negative labels. The probabilistic security assessment assumed the failure probabilities $\mathbb{P}_l = 0.001$ for all lines $l \in \Omega^O$. $|\Omega^O| = \{5, 10, 50, 100, 150, 187\}$ different lines were selected where $k = \{1, 2, 3, 4, 5\}$ lines out of these $\Omega^O$ can possibly fail. The studies were performed on a standard laptop with i7, four cores CPU 1.80GHz and 16GB RAM. The AC power flows were run in MATPOWER 8.0 in MATLAB R2023b. The DC OPF was implemented in the CVXPY 1.3.1 library (with solver SCS 3.2.0) in Python 3.8.16. The graph metrics were computed using NETWORKX 2.8.4. The linear algebra was done in MATLAB or with SCIPY 1.11.3.

### A. Quantifying Approximation Error

This study investigates the approximation error of the proposed polynomial $LODF^t_{\mathrm{N}-k}$ to the conventional LODFs $LODF_{\mathrm{N}-k}$. The error was assessed by the mean of the element-wise squared errors

$$\mathrm{MSE} = \frac{1}{|\Omega^B|^2} \sum_{b \in \Omega^B} \sum_{\tilde{b} \in \Omega^B} \left( \mathcal{B}_{b,\tilde{b}} - \tilde{B}^{-1}_{b,\tilde{b}} \right)^2 \tag{41}$$

between the proposed polynomial approximation $\mathcal{B}$ and the actual $\tilde{B}^{-1}$ computed with the Woodbury-formula. The trivial baseline was the original, fully intact $B^{-1}$ that did not consider any N-$k$ fault changes in the matrix. The study calculated the mean of a subset of the 10 000 fault samples for each $k$ fault combination. The subset considered only disjoint fault combinations, so the fault combinations where each line in the fault shares no common bus. Table I shows this share of disjoint fault samples.

TABLE I
NUMBER OF FAULTS SAMPLES WHERE FAULTS HAVE A COMMON BUS (ADJOINT FAULTS) OR NOT (DISJOINT FAULTS)

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| adjoint [%] | 0.0 | 3.4 | 9.7 | 18.4 | 29.8 | 41.1 | 52.2 |
| disjoint [%] | 100.0 | 96.6 | 90.3 | 81.6 | 70.2 | 58.9 | 47.8 |

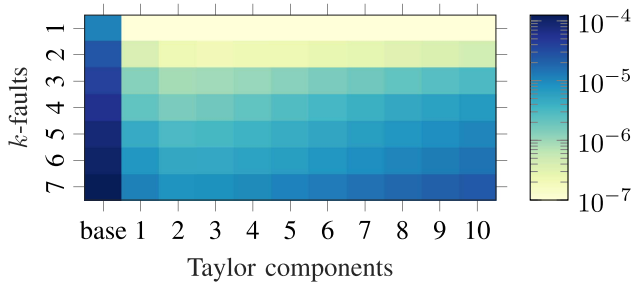



Fig. 3. MSE of all matrix elements between the proposed Taylor-series approximation with $t$ components and the actual $\tilde{B}^{-1}$ for disjoint faults. The trivial baseline $B^{-1}$ is shown as 'base'.

Fig. 3 shows that the proposed approximation improves over the trivial baseline in all cases by around 2-3 orders of magnitude. The error increases with increasing $k$ fault combinations. The error for a smaller number of Taylor components around $t = t_1 = t_2 = \{1, 2, 3, 4\}$ improved with increasing components; then, the error increased for increasing $t > 4$ as the Taylor series did not converge in many samples.

### B. Selecting Taylor Series Components

This study guides selecting the components $t$ for the two Taylor series approximations (16) and (22). The selection considers the error and convergence. The share of the converging Taylor series was investigated by computing the norms. When the norm is lower than 1, a series is considered converging; otherwise, diverging. The share of converging Taylor series was computed for the 10 000 samples per $k$ faults.

Fig. 4 shows the Taylor series in (16) converges at 90% of samples for all $k$-fault combinations. However, the convergence performance of the approximation in (22) is increasingly poor for larger $k$ combinations. Above $k > 4$ fault combinations, the Taylor series rarely converges. Therefore, a lower number of components $t_2$ may be preferable in (22) than $t_1$ in (16), particularly for higher $k$ faults.

Fig. 3 shows an improvement of roughly an order of magnitude in MSE when selecting $t = 2$ over $t = 1$ in all combinations of faults $k$. Therefore, we select $t_1 = 2$ for (16) and $t_2 = 1$ for (22) and test this selection further. For Taylor components

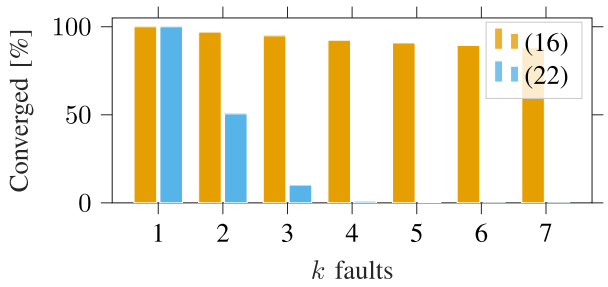



Fig. 4. Share of converged Taylor series.

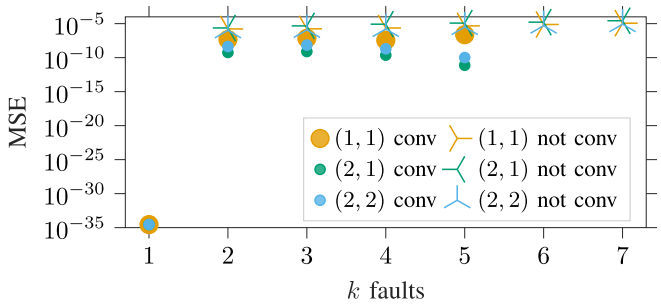



Fig. 5. MSE of converged (conv) and non-converged (not conv) Taylor series. Different Taylor series were studied with varying numbers of components. The parenthesis shows the number of components $(t_1, t_2)$ of the two Taylor series, (16) (22), respectively.

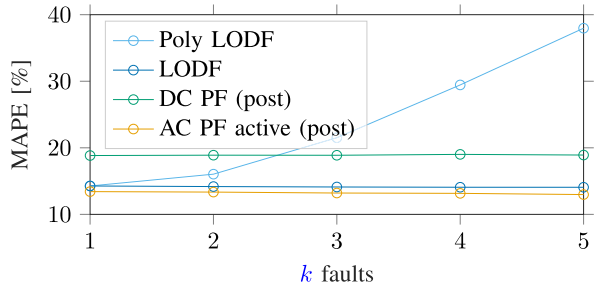



Fig. 6. Mean absolute percentage error to the apparent AC power flow for various $k$-fault combinations.

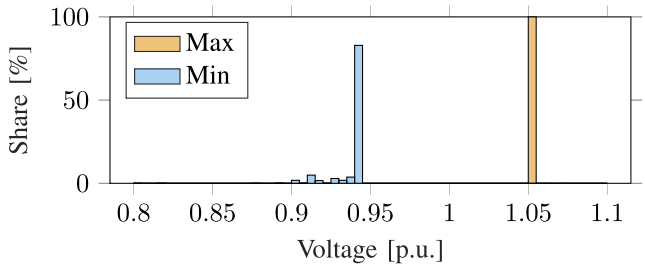



Fig. 7. Maximum and minimum voltages of the AC PF (post-fault) solution for various $k$-fault combinations considering converged cases only.

$t_1 = \{1, 2\}$ and $t_2 = \{1, 2\}$, we calculated MSE using samples where the two Taylor series converged and did not converge, respectively. 'converged' here means that both the two series inclusively converged, e.g. for $k > 4$ no sample had all Taylor series converging as Fig. 4 showed.

Fig. 5 shows the sensitivity of the convergence on the MSE. The MSE is lower for all $k$ fault combinations when the Taylor series successfully converges than when the Taylor series does not converge. The proposed selection of $(t_1 = 2, t_2 = 1)$ has the lowest MSE and confirms the validity of this selection for this case study. The MSE $\leq 10^{-7}$ for the samples when the series did not converge is still several orders of magnitude lower than the trivial baseline of using $B^{-1}$ as shown in Fig. 3. This also demonstrates that the Taylor series (16) (convergence above 90 % as shown in Fig. 4) has a significantly higher impact on MSE than the second Taylor series (22). The remaining case study considers these two selected components $t_1 = 2$ and $t_2 = 1$.

### C. Post-Fault Line Flows

This study investigates the accuracy of the proposed polynomial LODFs when computing post-fault line flows. 1000 fault samples for each $k = \{1, \dots 5\}$ were studied. The proposed polynomial LODFs compute the post-fault flows with (24) using the pre-fault apparent power [MVA] obtained from the pre-fault AC power flow. The mean absolute percentage error

$$\text{MAPE} = \frac{1}{|\Omega^L|} \sum_{l \in \Omega^L} \left| \frac{f_l^t - f_l^{AC}}{f_l^{AC}} \right| \tag{42}$$

references the apparent post-fault AC power flow $f_l^{AC}$ (considered as ground truth). Samples where no ground truth exists, as the post-fault sample then is assumed not physically feasible (e.g. when Newton Rapson did not converge). Around 75% of samples converged. Fig. 6 shows that the proposed polynomial LODFs result in errors similar to a DC approximation of post-fault power flows for $k = \{1, 2, 3\}$. However, the proposed polynomial LODFs become increasingly inaccurate for higher $k$. Additional baselines are the conventional LODF and the AC post-fault flows considering active power only. These results show that the polynomial LODFs need to be used with caution and if so, only below $k \leq 3$. Subsequently, we assessed the voltage errors for AC post-fault power flows using the AC power flow. Fig. 7 shows the minimal and maximal voltages over all buses per sample. As the figure shows for the converged samples, 100% of maximal voltages were close at 1.05. The minimal voltages varied between 0.9 and 1. LODFs (or polynomial LODFs) do not replace AC post-fault power flows as the latter provides these insights in the voltages helping to assess voltage stability.

### D. Faults Proximity

This study investigates the impact of the line faults' proximity on the MSE of the LODF matrix $\mathcal{B}$, the Taylor-convergence and the accuracy of the estimated post-fault power flows. The LODFs distribute more power from the line in an outage to the neighbouring proximity lines than to the lines 'far away' in the network. To avoid cross effects with $k$, this study first fixed $k = 2$ and analysed 10 000 samples; then, this study analysed all possible fault combinations $k = \{2, \dots 7\}$ with each 10 000

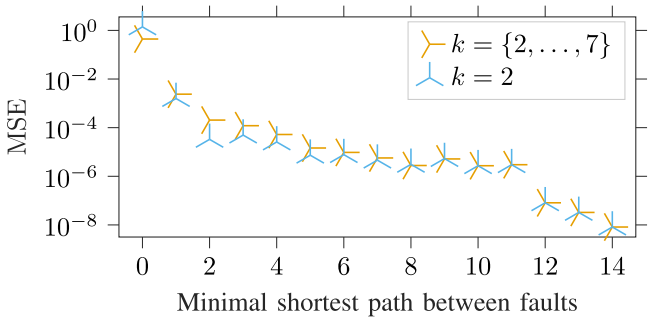



Fig. 8. MSE over the minimal distance between the faults.

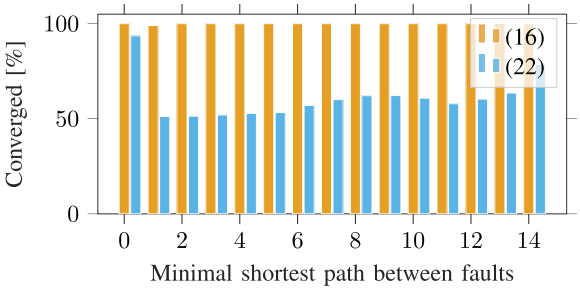



Fig. 9. Shares of converged Taylor series over the minimal fault distance.

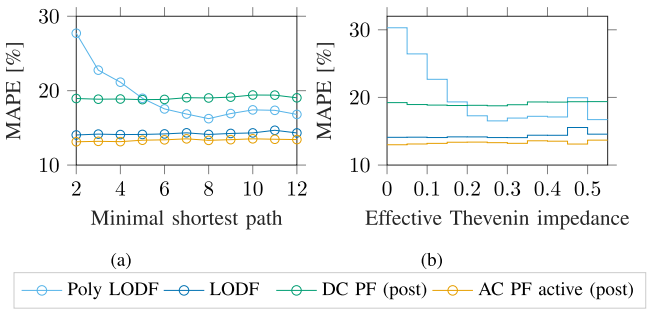
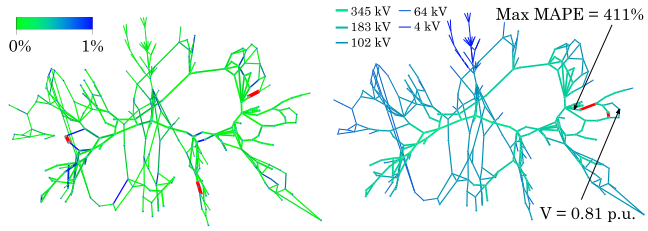



Fig. 10. Mean absolute percentage error to the apparent post-fault AC power flow for $k$-faults. Between the faults are different (a) shortest paths and (b) average effective Thevenin impedances (displayed in 0.05-steps bins).

samples. $k$ and the distance have cross-effects as the more faults coincide, the shorter their path as Table I shows.

Fig. 8 shows a strong decrease of the MSE in $\mathcal{B}$ with the minimal shortest path between the faults at $k = 2$ and $k = \{2, \ldots 7\}$. Adjoint faults showed an inferior performance. The MSE improves multiple orders of magnitude with increasing distance between the faults. As the MSE shows a similar relation to the proximity in $k = 2$ and $k = \{2, \ldots 7\}$, we conclude the main driver for the inaccuracy is the proximity. Particularly for a minimal path lower than 3 the proposed polynomial $LODF^{t}_{\mathrm{N}-k}$ is not recommended. Fig. 9 shows little impact of the minimal shortest paths on the convergence share. Interestingly, only a few samples did not converge in the adjoint faults (the shortest path is zero). However, their MSE is significantly higher than at longer minimal paths.




(a) Line colours show MAPEs [%] (b) Line col. show voltage levels [kV]

Fig. 11. Approximating line flows with polynomial LODFs is fairly accurate in the IEEE 300-bus system when (a) 3 outages are far away in different regions (in red), but is inaccurate when (b) 2 outages are close to each other. The line thickness represents voltage levels.

Fig. 10 shows the impact of the proximity of faults on the post-fault power flow computations (average for 1000 fault samples for each $k = \{1, \ldots 5\}$). The proposed poly LODFs was the only approach decreasing in error as the distances between faults increased. The approach has a lower MAPE than DC power flow in faults sufficiently far from each other, e.g. $\geq 5$ or a Thevenin impedance $\geq 0.2$. The conventional Woodbury LODFs, DC power flow and AC power flow using active power (for comparison) have lower errors than the proposed approach, however, are in a similar order of magnitude $10 - 20\%$.

### *E. Expected Congestion in Large Systems At Extreme Events*

This study investigates the performance of the proposed LODFs on large systems with AC and DC power flows. The pre-fault power flow is computed. Subsequently, the LODFs are used on the power flows to compute N-$k$ flows.

Fig. 11 shows an example of how to use (and not use) the polynomial approach. An appropriate use of the proposed approach is in cases where the faults are far away from each other as in Fig. 11(a) resulting in low MAPEs. Fig. 11(b) shows an $k = 2$ outages close to each other. This fault scenario led to a high MAPE at 411% in one connected line. There, the pre-fault flow was 221 MVA and the post-fault ground truth 0.4 MVA. As the Taylor series considered only one component, the difference from pre to post-fault was too large that the approach accurately approximated this change.

Fig. 12 simulated wind storms at two different locations of the power system which may require assessing $k = 2$ events, then, the proposed approach can quickly compute the expected line flows for all possible outage combinations (shown in red in the figure). This study was repeated for the 118-, 300-, 1354- and 2383-bus systems considering $|\Omega^{O}| = 20$ single line outages and $k = 2$ faults. The impact radius of the wind storms was assumed larger (so they tangengt). The LODFs and polynomial LODFs were applied to the pre-fault apparent AC power flow. The mean absolute error

$$\mathrm{MAE} = \frac{1}{|\Omega^{L}|} \sum_{l \in \Omega^{L}} \left| \mathbb{E}\left[f_l^{t}\right] - \mathbb{E}\left[f_l^{AC}\right] \right| \tag{43}$$

of the expected congestion references the expected apparent post-fault AC power flow $f_l^{AC}$ as ground truth. Table II shows

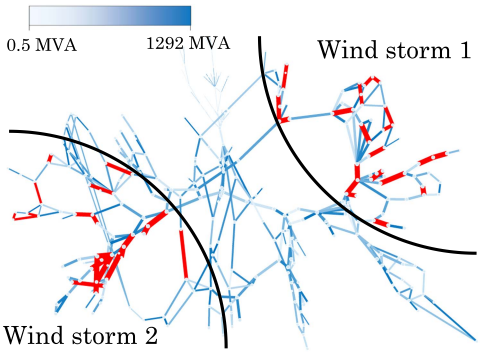



Fig. 12. Expected congestion in the IEEE 300-bus system for any $k = 2$ outages (marked in red) originating from two wind storms. The impact areas of the wind storms are shown as black-quarter circles.

TABLE II
MAE FOR EXPECTED *APPARENT* POWER FLOWS [MVA]

| System $\lvert\Omega^B\rvert$ | Poly. LODF | LODF | DC power flow |
|---|---|---|---|
| 118 | 4.4 | 4.4 | 5.9 |
| 300 | 12.8 | 12.7 | 18.7 |
| 1354 | 7.2 | 7.1 | 13.1 |
| 2383 | 4.5 | 4.4 | 5.4 |

TABLE III
$N - K$ SECURITY ASSESSMENT WITH PROPOSED LODFS

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Accuracy [%] | 100.0 | 99.8 | 99.4 | 98.9 | 98.2 | 97.5 | 96.6 |
| Flow error [%] | 0.0 | 2.5 | 6.0 | 10.1 | 14.7 | 19.2 | 24.8 |

the Polynomial LODFs and LODFs have similar errors on the expected congestion outperforming DC post-fault power flows.

### F. N-k Security Assessment

This study investigates the performance when using the LODFs for security assessment on the 118-bus system. Unreasonably higher congested lines (more than 200%) originated from inaccurate Taylor approximations or islanded grids and were removed from the assessment. The congestion assessment in Table III shows that the accuracy is above 96% in all combinations of faults $k$. For $k \leq 4$, the accuracy is above 98.9% which shows that the proposed method works well to approximate the LODFs. However, for $k > 4$, the proposed method becomes more inaccurate. When congestion was inaccurately predicted, the average flow error (only for incorrect fault samples) remained below 25% for all $k$ and increased with $k$. This trend shows the security prediction and flow estimations become more inaccurate with increasing $k$. Fig. 13 shows the same analysis for the minimal paths between faults. Interestingly, above the shortest distance of 2 between two faults, the approximation works well with almost 100% accurate security predictions. The flows had a small estimation error of $< 5\%$ on average if the predictions were inaccurate. For short times, the flow can exceed the line flow threshold, and bounding the estimation error around $< 5\%$ is an interesting insight. Fig. 14 shows the accuracy of the two types of predictions (positives and negatives). All four metrics decreased

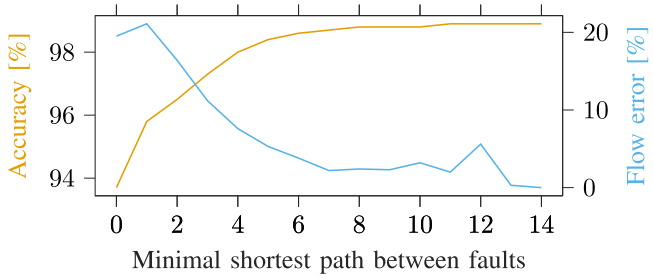



Fig. 13. Accuracy and relative error in flow predictions with proposed LODFs highly dependent on the minimal shortest path $d$ between faults.

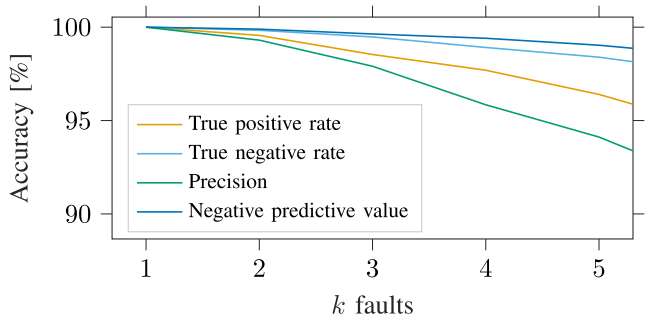



Fig. 14. Performance of security assessments using the proposed LODFs.

with $k$-faults. However, the precision and the true positive rate decreased faster (from 100% at $k = 1$ to 95% at $k = 7$) than the true negative rate and the negative predictive value (from 100% at $k = 1$ to 97% at $k = 7$). Therefore, the polynomial LODFs predicted negative samples better than positive samples, leading to an imbalance of accuracies. This imbalance was likely related to the selection of Taylor components at $t_1 = 2$ and $t_2 = 1$. The Taylor series expansion is a convergent alternating series. In any convergent alternating series, the truncations alternate between lower and upper bounds, so likely in the selection of $t_1 = 2$ and $t_2 = 1$ the truncations were more often at the bound of estimating lower line flows than flows higher than the true post-fault line flows. To more robustly assess the security under the assumption the Taylor series converges, someone may compute two consecutive Taylor series approximations, and then consider the supremum of the two line flows.

### G. Computational Analysis

Three studies assess the computational performance of the proposed polynomial LODFs, when applying them for congestion estimation, and for large systems. The first study assesses the computational times of the LODFs for varying Taylor series components (considering $t_1 = t_2$) and for varying fault conditions. The average time of 10 000 LODFs relative to the LODFs using the Woodbury formula is presented in Fig. 15. The computation of the proposed polynomial LODFs was slower than the Woodbury matrix inversion, as the polynomial LODFS involved many matrix multiplications and summations. The proposed LODFs outperformed the Woodbury LODFs for $k = \{1, 2\}$ and low components $t$.

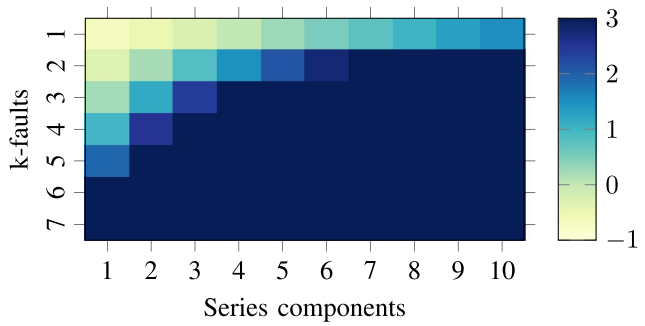



Fig. 15. Relative computational time of the proposed Taylor series approximation to the baseline Woodbury formula. The proposed approximation is only faster for low $k$ and low Taylor components $t$.

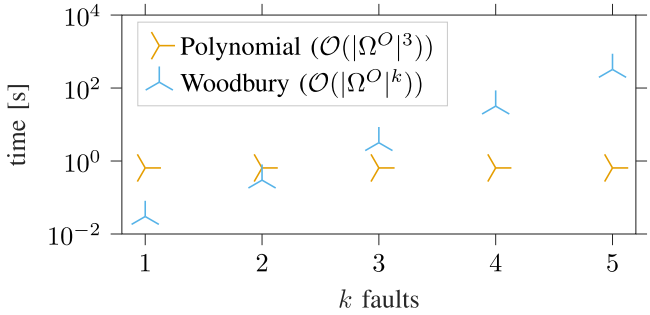



Fig. 16. Time computing the expected line flows with the proposed polynomial LODFs and baseline Woodbury LODFs for a constant number $\Omega^O = 10$ of contingency lines, out of which all $k$-fault combinations were considered.

The second study assesses the online computations of expected congestion and security assessment using the decomposing approach (online workflow in Fig. 1(b). Fig. 16 shows the computational time for $|\Omega^O| = 10$; the Woodbury formula takes too long for a standard laptop to study the higher numbers of $|\Omega^O|$ and $k$. Note that this figure demonstrates the proposed decomposing approach and that the polynomial LODFs for expected congestion are computationally constant with $k$, which is a main finding of this work. The proposed decomposing approach only computed the scalar weights $\phi_{1,l}$, $\phi_{2,l,\hat{l}}$, $\hat{\phi}_{2,l,\hat{l}}$ and $\phi_{3,l,\hat{l},\check{l}}$ for different $k$, and then replaced these weights in (35). The Woodbury formula, however, requires exponential growth in computational times with $k$. There, the summation in (31) considers $M(k)$ elements where $M(k)$ grows exponentially with $k$ as shown in (6). Therefore, the proposed polynomial LODF has strong computational benefits when $k > 3$ as the figure shows. Fig. 17 shows the computational time for different sizes of the power system, or here for increasing many lines $\Omega^O$ that can fail. In the example of $k = 4$, the Woodbury formula was intractable for $|\Omega^O| > 10$. The proposed polynomial LODFs were studied however for more $|\Omega^O|$ variations showing the LODFs improved the computations to around $\mathcal{O}(|\Omega^O|^3)$. For larger $k$ and larger $|\Omega^O|$, the computational benefits of the proposed approach over the Woodbury formula can be roughly estimated with $\frac{|\Omega^O|^k}{|\Omega^O|^3}$. Applying this rough estimate to the IEEE-118 bus system with $|\Omega^O| = 187$ lines showed a time benefit of one billion for $k = 7$. Hence, the proposed polynomial LODF has benefits for large systems and large $k$.

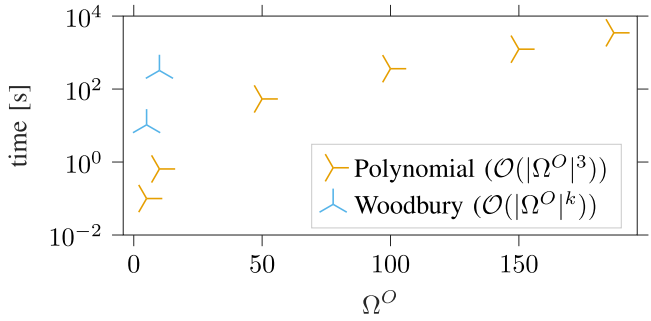



Fig. 17. Time computing the expected line flows with proposed polynomial LODFs and state-of-the-art Woodbury LODFs for a constant number $k = 4$ of outages. The Woodbury LODFs are unable to compute for larger $\Omega^O > 10$ in reasonable times.

TABLE IV
MEMORY AND COMPUTING REQUIREMENTS FOR THE OFFLINE WORKFLOW IN LARGE SYSTEMS WITH $|\Omega^O| = 20$ POSSIBLE FAULTS

| Systems $\lvert\Omega^B\rvert$ | Memory stored [MB] | Memory working [MB] | Computing time |
|---|---|---|---|
| 118 | 5 | 21 | 3 s |
| 300 | 11 | 137 | 9 s |
| 1354 | 54 | 2797 | 6 min |
| 2383 | 80 | 8665 | 42 min |

The third study in Table IV assesses memory and computational times (offline) in larger systems. Some auxiliary matrices can have sizes up to $|\Omega^B|^2 \times |\Omega^O|^2$ in the working memory and an implementation may need multiples of these. Table IV reports the maximal working memory across all single matrices. The stored memory considers $\sum_{i\in 1,2,3}\{B_{br}\, A\, \mathcal{P}_i\, A^T f^0\} + \{B_{br}\, A\, \hat{\mathcal{P}}_2\, A^T f^0\}$ for $|\Omega^O| = 20$ faults and were stored in single precision. The computational times consider the offline workflow (in Fig. 1(a)), e.g., the offline preparation of all matrices in Sections III and IV-C until (33) using Taylor series $t_1 = 1$ and $t_2 = 1$. The times and memory increase with system sizes which highly depends on an efficient implementation of the algorithm, here, reaching a working memory of 8GB in the 2383-bus system.

### H. Discussion

The proposed approach was accurate when the series converged and for failures far away from each other. The Taylor series of (16) converged for all $k$ but (22) only for low $k$. The approximation of (16) showed a higher impact on the accuracy, hence the overall approach showed reasonable accuracies also for higher $k$-faults, and security assessments, as inaccurate line flow estimations of 25% may be acceptable for short times, however, at higher $k$ typically other issues arises aside overloads, e.g. voltage instability which this approach not assesses. This study demonstrated that the proposed polynomial LODFs led to an approach to estimated expected congestion that decomposes the computation of expected line flows into repeated basis motifs, drastically reducing computations (for $k > 4$). There, we demonstrated computational reductions from $\mathcal{O}(|\Omega^O|^k)$ to $\mathcal{O}(|\Omega^O|^3)$. Applying parallel computing may further reduce the computational time.

The approach is limited to faults that are 'far' from each other, e.g., when the shortest distance between two faults is greater than a path of 4 lines (or with an effective Thevenin distance of greater than 0.2). This may only be at multiple extreme (weather) events over large areas, wind storms, floods, or coordinated attacks. For faults in proximity, the error of the polynomial LODFs increases drastically. This approach is limited to estimating line overloads, and can not provide insights to other stability issues more typically occurring at $k > 1$, e.g., voltage stability. The approach approximates the (nonlinear) superposition of LODFs with a linear superpositioning considering multiplications and summation, which is limiting specifically when the superposed terms are large (in the neighbourhood of the faults). Limiting is also that the approximation only works if the susceptance matrices of post-fault are non-singular, e.g. where the fault-combination does not create separate, islanded subsystems. Another limitation is that the polynomial LODFs do not always converge, but this can be checked by the norms. Numerical studies found the approach has a high convergence rate for low $k$ and is faster than Woodbury with a few Taylor components. Computational benefits are in the use case on expected line congestion. However, the working memory limits the number of single faults that can be combined.

## VI. Conclusion

This article investigated reducing the computational complexity of estimating expected post-fault line flows for multiple line faults. The proposed polynomial approximation of the LODFs simplifies the calculations, especially for scenarios involving multiple simultaneous line outages (N-$k$ faults). The approach computes expected line flows decomposing N-$k$ faults into repeating basis functions. Through case studies involving 118-, 300-, 1354- and 2328-bus systems, the article evaluated the accuracy and computational efficiency of the proposed polynomial LODFs showing the potential for reliability management. Future work can investigate the differentiability of the polynomial approximation, investigate the approach as constraints for N-$k$ security-constrained optimal power flow, methods to compute bus voltages based on the obtained post-fault power flows, the variance of post-fault power flows, the worst-case post-fault power flow and develop memory-efficient algorithms.

## References


[1] P. Panciatici, G. Bareux, and L. Wehenkel, "Operating in the fog: Security management under uncertainty," *IEEE Power Energy Mag.*, vol. 10, no. 5, pp. 40–49, Sep./Oct. 2012.

[2] 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, 2005.

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

[4] F. Capitanescu, "Are we prepared against blackouts during the energy transition?: Probabilistic risk-based decision making encompassing jointly security and resilience," *IEEE Power Energy Mag.*, vol. 21, no. 3, pp. 77–86, May/Jun. 2023.

[5] Q. Chen, "The probability, identification, and prevention of rare events in power systems," Ph.D. dissertation, Iowa State University, Ames, IA, USA, 2004.

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

[7] R. L.-Y. Chen, A. Cohn, N. Fan, and A. Pinar, "Contingency-risk informed power system design," *IEEE Trans. Power Syst.*, vol. 29, no. 5, pp. 2087–2096, Sep. 2014.

[8] R. Hemmati, H. Saboori, and M. A. Jirdehi, "Stochastic planning and scheduling of energy storage systems for congestion management in electric power systems including renewable energy resources," *Energy*, vol. 133, pp. 380–387, 2017.

[9] E. Karangelos and L. Wehenkel, "Probabilistic reliability management approach and criteria for power system real-time operation," in *Proc. Power Syst. Computation Conf.*, 2016, pp. 1–9.

[10] K. Koeck, "Probability based transmission system risk assessment," Ph.D. dissertation, Technical University Graz, Graz, Austria, Aug. 2016.

[11] D. Kirschen and D. Jayaweera, "Comparison of risk-based and deterministic security assessments," *IET Generation, Transmiss. Distrib.*, vol. 1, no. 4, pp. 527–533, 2007.

[12] I. Dobson et al., " Electric power transfer capability: Concepts, applications, sensitivity and uncertainty,"PSERC Publication, no. 01–34, 2001.

[13] M. Majidi-Qadikolai and R. Baldick, "Integration of $N-1$ contingency analysis with systematic transmission capacity expansion planning: Ercot case study," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 2234–2245, May 2016.

[14] A. Kumar, S. Srivastava, and S. Singh, "Available transfer capability (ATC) determination in a competitive electricity market using AC distribution factors," *Electric Power Compon. Syst.*, vol. 32, no. 9, pp. 927–939, 2004.

[15] M. Liu and G. Gross, "Role of distribution factors in congestion revenue rights applications," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 802–810, May 2004.

[16] T. Guler, G. Gross, and M. Liu, "Generalized line outage distribution factors," *IEEE Trans. Power Syst.*, vol. 22, no. 2, pp. 879–881, May 2007.

[17] H. Zhou, K. Yuan, and C. Lei, "Security constrained unit commitment based on modified line outage distribution factors," *IEEE Access*, vol. 10, pp. 25258–25266, 2022.

[18] S. Poudel, Z. Ni, and W. Sun, "Electrical distance approach for searching vulnerable branches during contingencies," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3373–3382, Jul. 2018.

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

[20] T. Van Cutsem and T. Weckesser, "Searching for plausible Nk contingencies endangering voltage stability," in *2017 IEEE PES Innov. Smart Grid Technol. Conf. Europe*, 2017, pp. 1–6.

[21] T. Ding, C. Li, C. Yan, F. Li, and Z. Bie, "A bilevel optimization model for risk assessment and contingency ranking in transmission system reliability evaluation," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3803–3813, Sep. 2017.

[22] J. McCalley et al., "Probabilistic security assessment for power system operations," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, 2004, pp. 212–220.

[23] J. L. Cremer and G. Strbac, "A machine-learning based probabilistic perspective on dynamic security assessment," *Int. J. Elect. Power Energy Syst.*, 2021, vol. 128, Art. no. 106571.

[24] Y. Liu, N. Zhang, D. Wu, A. Botterud, R. Yao, and C. Kang, "Searching for critical power system cascading failures with graph convolutional network," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 3, pp. 1304–1313, Sep. 2021.

[25] K. Zhou, I. Dobson, and Z. Wang, "The most frequent N-k line outages occur in motifs that can improve contingency selection," *IEEE Trans. Power Syst.*, vol. 39, no. 1 pp. 1785-1796, Jan. 2023.

[26] M. A. Woodbury, *Inverting Modified Matrices*. Princeton, NJ, USA: Department of Statistics, Princeton University, 1950.

[27] J. Guo, Y. Fu, Z. Li, and M. Shahidehpour, "Direct calculation of line outage distribution factors," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1633–1634, Aug. 2009.

[28] 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.

[29] H. Ronellenfitsch, M. Timme, and D. Witthaut, "A dual method for computing power transfer distribution factors," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1007–1015, Mar. 2016.

[30] T. Nesti, A. Zocca, and B. Zwart, "Emergent failures and cascades in power grids: A statistical physics perspective," *Phys. Rev. Lett.*, vol. 120, no. 25, 2018, Art. no. 258301.

[31] A. Bagheri and C. Zhao, "Distributionally robust reliability assessment for transmission system hardening plan under $N-K$ security criterion," *IEEE Trans. Rel.*, vol. 68, no. 2, pp. 653–662, Jun. 2019.
[32] D. A. Tejada-Arango, P. Sánchez-Martın, and A. Ramos, "Security constrained unit commitment using line outage distribution factors," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 329–337, Jan. 2018.
[33] C. S. Song, C. H. Park, M. Yoon, and G. Jang, "Implementation of PTDFs and LODFs for power system security," *J. Int. Council Elect. Eng.*, vol. 1, no. 1, pp. 49–53, 2011. [Online]. Available: https://doi.org/10.5370/JICEE.2011.1.1.049
[34] C. Neumann, *Uber die Methode des Arithmetischen Mittels*, vol. 13. Leipzig, Germany: S. Hirzel, 1887.
[35] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: JHU Press, 2013.
[36] G. W. Stewart and J.-g. Sun, "Matrix perturbation theory," *Comput. Sci. Sci. Comput.*, vol. 1, pp. 114–134, 1990.
[37] K. Zhou et al., "Bayesian estimates of transmission line outage rates that consider line dependencies," *IEEE Trans. Power Syst.*, vol. 36, no. 2, pp. 1095–1106, Mar. 2021.
[38] B. Dai, S. Ding, and G. Wahba, "Multivariate Bernoulli distribution," *Bernoulli*, vol. 19, no. 4, pp. 1465–1483, 2013.
[39] E. W. Dijkstra, "A note on two problems in connexion with graphs," in *Edsger Wybe Dijkstra: His Life, Work, Legacy*. San Rafael, CA, USA: Morgan & Claypool, 2022, pp. 287–290.
[40] P. Cuffe and A. Keane, "Visualizing the electrical structure of power systems," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1810–1821, Sep. 2017.

**Jochen Lorenz Cremer** (Member, IEEE) received the Ph.D. degree from Imperial College London, London, U.K., in 2020. He is currently an Assistant Professor with Delft AI Energy Lab, Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Delft, The Netherlands. His research interests include machine learning and mathematical programming applied to the operation and planning of power systems.