A Method for Variance-based Sensitivity Analysis of Cascading Failures

Aaron S. C. Leavy, Georgios A. Nakas, and Panagiotis N. Papadopoulos, *Member, IEEE*

Abstract—Cascading failures of relay operations in power systems are inherently linked with the propagation of wide-area power system blackouts. In this paper, we consider a power system cascading failure as an indicator matrix encoding: what power system relays operated within a cascading failure inherently capturing the component and the sequence of tripping events. We propose that this matrix may then be used with extended forms of variance-based sensitivity estimators to quantitatively rank how sensitive observed power system cascading failures are to power system variables, considering overall system cascading failures as well as cascading failures grouped by network area and relay types. We demonstrate our proposed method by investigating the sensitivity of cascading failures to relay parameters, system conditions, and fault location using a version of the IEEE 39 bus model modified to include protection relays, wind farms, and tap-changing transformers. Input power system variables included: system operational scenario, disturbance location, relay parameters or thresholds. The Case Studies' results confirm the method's utility by successfully generating relative rankings of input variables' importance with respect to cascading failure propagation. The results also show cascading failures' sensitivity to input variables to be high due to non-linear relationships between input variables and cascading failures.

Index Terms—Cascading failures, power system cascading failures, protection relay thresholds, sensitivity analysis, sensitivity indices, variance-based sensitivity analysis.

I. INTRODUCTION

MANY observed wide-area blackouts have historically occurred after chains of successive control and protection relay operations before the afflicted power system eventually collapsed, as covered in [1], [2]. The profession has therefore converged towards the concept of *cascading failures*. This term-of-art is defined in [2], which reviews methods and analysis approaches applied to addressing the cascading failures problem. Cascading failures in electrical power systems are high impact low probability events, where the impact can be extremely detrimental to societal functioning especially where these cascading failures lead to wide-area blackouts [1]. An active area of cascading failure research focuses on quantifying the effect of power system variables or parameters on cascading failure propagation. However, a

A. S. C. Leavy, G. A. Nakas, and P. N. Papadopoulos are with the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, United Kingdom. E-mails: {aaron.leavy, georgios.nakas, panagiotis.papadopoulos}@strath.ac.uk.

variance-based sensitivity analysis method to quantify the variance of cascading failures with respect to power system variables has not been proposed in prior works. This paper proposes such a variance-based sensitivity method.

Many previous works on cascading failures apply analysis methods which are fundamentally steady-state in nature, *i.e.* they do not model system dynamics. This is a simplification of realistic analysis of cascading failures, and therefore engenders new avenues of research into tractable cascading failures analysis methods explicitly incorporating dynamics [2]. Typically, these dynamic approaches are in the form of root mean square (RMS) phasor time domain simulations [2]. In addition, [3] highlights the lack of understanding and complex nature of mechanisms participating in cascading failures as well as the importance and need of cascading failure analysis tools that capture such mechanisms. Ref. [4] provides an overview of to cascading failure modelling and simulation issues.

A particular modelling framework incorporating system dynamics specifically orientated towards cascading failures analysis is proposed in [5], which explicitly allows the modelling of typical controller dynamics and protection relays. The work in [5] compares a proposed dynamics-based framework with a quasi steady-state framework, and demonstrates the substantial deviation in observed cascading failures depending on the choice of dynamics versus steady-state framework. The authors of [5] also indicate the importance of dynamic assumptions related to load models employed in the dynamics framework, which shows cascading failures to be sensitive to dynamic model approximations. The issue of protection relay modelling in dynamics studies for the purpose of cascading failure research is addressed in [6]. The proposal of [6] is a steady-state approach for generating lists of protective relays which may be critical to cascading failure propagation. This highlights these listed relays to be explicitly modelled in dynamic simulations assessing cascading failures, whilst ignoring relays deemed to be less important. The authors of [6] state the critical importance of including protection relays in cascading failures analysis, yet they show by implication that not all protective relays may be necessary by the very demonstration of their proposed method.

The results in [5], [6] indicate that there is the potential for complex sensitivities when considering modelling assumptions. Some formal effort has been undertaken to rigorously address sensitivity analysis in dynamic models in [7], which reviews a range of sensitivity analysis methods applied to dynamic simulations. The authors of [7] state that a variance-based sensitivity method [8]—often referred to as

with the term *Sobol method*—qualitatively represents a good balance between computational tractability and result utility for their motivating problem. However, this approach did not specifically consider cascading failures nor the implementation of protection devices.

A successive steady-state approach is proposed in [9] to model cascading failures, incorporating modelling of a range of protection relays. Although not the main thrust of their work, the authors perform a sensitivity analysis of some input parameters on two measures of resilience. First-order effects are calculated, which quantify the share of the variance in the resilience measures which can be individually apportioned to each of the input parameters, *i.e.* ignoring interactions between input parameters.

An investigation into the effects of power re-dispatch policy, and network line and generator topology hyper-parameters on both power loss and line failures during cascading failures is performed in [10]. The work in [10] uses a steady-state simulation approach, and uses graphical means to assess input-output relationships. In [11], authors review categories of approaches to under-frequency load shedding (UFLS) scheme conceptual design and implementation. UFLS relays are used to disconnect system load on detection of a low system frequency condition. Such a condition will typically occur during an acute lack of generation within the system; hence, the UFLS relays can disconnect load in order to reduce generation-load imbalance. A key theme of [11] is the complexity of setting UFLS relay thresholds in a qualitative manner in order to avoid deleterious propagation of blackouts. Although [11] reviews approaches in the literature for setting relay thresholds, another common implied theme is lack of attention paid as part of the analyses to the actual power system cascading failure itself which UFLS relay operations themselves are part of. In [12], cascading failures are investigated using dynamic simulations. Hidden failures of protection devices are also represented, by investigating the possibility of the protection devices tripping before reaching their preset thresholds.

In [13], a modified version of the IEEE 39 bus model is extended with: demand transformers and tap changers; automatic generation control; the addition of some wind farms; and the addition of protection relays. The modified model is then used to generate power system cascading failures via dynamic simulations including relay modelling whilst considering input power system operational scenarios and three-phase fault location. The contribution of [13] are descriptive statistics results of observed power system relay operations during cascading failures which show: firstly, inclusion of demand transformer tap changers in dynamic simulations decreases the number of cascading failures, and secondly; inclusion of automatic generation control decreases the number of UFLS operations. Ref. [13] only provides descriptive statistics results of observed power system relay operations during cascading failures which highlight the impact of relevant mechanisms (tap changers and automatic generation control) to the number of cascading failures and relay trips. Our proposed method is demonstrated in this paper using the power system model developed and used within [13]. However, [13] does not allow for quantitative assessment to enable the linking of observed cascading failure relay operations with input power system operational scenarios and disturbance locations in a manner which could detect the relative importance of those input variables with respect to cascading failure propagation. Here lies the difference between the contributions of [13] and our paper's contributions: we propose a method which can be used to perform variance based sensitivity analysis which allows ranking of input variables importance with respect to cascading failure propagation.

In [14], sequences of circuit outages are considered explicitly by forming influence graphs from sequence data to create a Markov chain model. This Markov chain model can then be used for statistical analysis purposes. Protective relay devices are intrinsically linked with cascading failure propagation as it is often protective relays which trip circuits causing circuit outages of the type studied by [14]. The work in [14] therefore emphasises the utility of focussing on cascading failure sequences themselves whether modelled as generic equipment outages such as the circuit outages in [14], or our proposed method of considering relay operations explicitly.

We shall now explain the motivation of our work, with reference to the aforementioned prior works in the literature.

It is clear from [1], [2] that cascading failures represent a significant risk to successful power system operation. Yet it is also clear that much work remains to incorporate reasonable consideration of system dynamics into cascading failure studies, a task principally complicated by the implied computational burden. Nonetheless, works such as [5], [6], [9] add strong emphasis to the need to model dynamics, particularly protective relays. Both [9], [10] imply that performing sensitivity analysis with respect to cascading failure studies could provide useful results to engineers. However, both works are primarily concerned with measures of the outcomes or impacts of cascading failures rather than the cascading failure sequences themselves. Explicitly modelling protective devices in dynamic simulations adds significant model complexity as evidenced by [5], [6]. This implies that rigorous sensitivity analysis should be a method available to engineers to identify how cascading failures may be affected, and moreover what parameters are most important with respect to dynamic modelling assumptions; this is possible via our proposed method. The contributions of [14] imply that cascading failures in power systems are intrinsically linked by the unique sequence of relay operations observed during the cascading failure if one assumes that circuit outages are being driven by relay operations. Considering cascading failures in this manner allows a quantitative and detailed representation of cascading failures amenable to performing sensitivity analysis.

The ability to perform sensitivity analysis on cascading failures themselves—as opposed to the outcomes of cascading failures such as energy not served or customers interrupted—would allow engineers to rank the importance of variables of interest on cascading failures, and therefore offer valuable information for more targeted planning and operation of the power system in consideration of these variables. We do not believe that sensitivity analyses of cascading failures themselves with respect to relay parameters have been performed explicitly in the literature.

Our contribution is as follows.

1) Power system cascading failures are treated as a novel indicator matrix of unique ordered relay operations capturing the exact components involved, and the order in which the components are tripping. This enables the application of the proposed variance-based sensitivity analysis estimators that allows quantitative assessment on power system cascading failures sequences and not just the impact of the cascade, offering more informative assessment. Three Case Studies to show the methods utility for identifying cascading failures sensitivity to various input variables or parameters, including operational scenario variables and protection device thresholds are presented. Three Aggregations, grouping cascading failures by relay type and network area are also proposed, to analyse the impact in a meaningful manner, e.g. identify specific areas that are affected or particular mechanisms related to voltage or frequency related phenomena.

2) Sensitivity analysis indices are therefore estimated for any input variables to the power system; we specifically investigate the variables: system demand, wind farm generation output, disturbance locations, and implicit and explicit relay threshold settings. Our results show these variables to non-linearly influence the propagation of cascading failures.

In consideration of [6], [11] and [13], we perform three Case Studies in order to demonstrate the utility of our proposed method for the purpose of systematically determining the sensitivity of cascading failures to power system variables. We perform these three Case Studies over three Aggregations, which are three different ways of aggregating observed power system cascading failures by characteristic.

II. PROPOSED METHODOLOGY

We shall now explain our proposed method in detail: how we consider cascading failures as indicator matrices of operated protection relays capturing the sequence of specific relay operation (*e.g.* over-voltage relay operation of a wind farm) and therefore providing a way to quantitatively characterise cascading failures. This method therefore allows the engineer to investigate which power system variables—out of a set specified by the engineer—cause most of the variance in power system cascading failures. The we shall explain our estimators to calculate first order and total effects sensitivity indices given a cascading failure indicator matrix. This approach enables the identification of the ranking of impact a set of input variables have on the appearance and propagation of cascading failures defined in the detailed manner of the indicator matrices.

The relative ranking of input variables allows identification of variables that are relatively more important for causing variance within power system cascading failures. This allows further studies to be performed predicated on this ranking to generate policies to mitigate power system cascading failures. For example, if the parameter thresholds for UFLS relays are relatively more important for cascading failure propagation compared to other types of relays within a study, this means that cascading failure mitigation schemes could be designed by focussing upon UFLS parameter value setting as a top priority over all other relay types. Our method's utility is during planning timescales and offline analysis and aims at informing protection scheme design as well as planning or operational interventions to mitigate the risk of propagation of cascading failures. For example, identifying important operational parameters or protection relay thresholds that have an important role in the propagation of cascades and consequently particular attention or measures need to be put in place to address this. In addition, it allows for targeted sensitivity analysis with respect to particular network areas or mechanisms (e.g. due to voltage related phenomena).

A. Summary of Methodology

Our method can be summarized via the following list of steps. We also indicate in brackets which Sections within Sec. II are relevant to each step.

1) Input power system variables to be investigated for their effects on cascading failure variance are defined a priori for the specific case study at hand; upper and lower bounds for these variables must also be defined (Sec. II-E).

2) A block of dynamic power system simulations initialised with initial conditions sampled from these input power system variables, are run to generate cascading failures (Sec. II-B, II-E and II-J).

3) The sequences of operated relays are transformed to cascading failure indicator matrices for a particular Aggregation, and concatenated to a sequential running tally of previous cascading failure indicator matrices (Sec. II-C, II-D and II-I).

4) All cascading failure indicator matrices tallied so far are then used to estimate first order and total effects indices (Sec. II-F, II-G, II-H, and II-I).

5) The indices are ranked to show relative importance of power system variables with respect to cascading failure variance.

6) Steps 2-4 can be repeated with new independent simulation blocks drawn from the engineer's defined input variables, whilst keeping a running tally of all cascading failures with each new independent simulation block. Indices and their rankings will converge towards their population values with increasing numbers of simulation blocks.

The input variables to be studied as well as their upper and lower bounds need to be defined before performing the sensitivity analysis. This choice can be informed by historic or forecasted data (e.g. for operational variables like system demand) or by specific design considerations (e.g. for investigating the sensitivity of protection device thresholds). In our work, we used engineering judgement to choose upper and lower bounds depending on the specific case study at hand, as described in detail in Sec. III-C. In order to consider system operational conditionsand therefore initial conditions for the underlying power system simulationsa set of operational power system variables (e.g. system demand, wind generation, etc.),

are specified. For example: system demand could be represented by one variable; or a set of many variables (one per load location) with individually engineer-specified prior sampling distributions, depending on the level of detail required in the analysis.

The simulations we perform do not inherently incorporate any system or relay uncertainties; the only source of uncertainty in our Case Studies which we use to demonstrate our proposed method is due to the quasi-Monte Carlo sampling of input variables as part of the variance-based sensitivity calculations. For a particular point value of input variables, the corresponding power system simulations are deterministic. Uncertainty in model and relay modelling can implicitly be considered with the choice of input variables and their associated range, e.g. using the threshold of a protection device activation as an input variable as shown in the case studies presented in Sec. III.

Note that our method is agnostic to the specific dynamic models used to represent power system components within the underlying power system simulations; the method relies upon the protection relay operation sequence. Therefore, any level of detail within for example relay models can be captured (potentially limited by the simulation framework used). This will be implicitly considered via the influence of these dynamic models on the order of protection relay operations within the cascading failures.

The proposed method is concerned with quantifying the effect on cascading failure variance due to a set of defined input variables' variances, while capturing the detailed dynamic behaviour of the system including the effect of protection devices. This allows for more informative design of countermeasures with respect to typical contingency analysis studies. Although we demonstrate our method using three Case Studies, the method is more interchangeable than just those Case Studies shown in this paper: the choice of input variables, and cascading failure Aggregations can be defined as needed, thus allowing quantitative investigation of the impact of various parameters on the propagation of cascading failures.

B. Cascading Failures as a Protection Relay Sequence

We consider a *cascading failure* to be an ordered *sequence* of unique protection relays observed to have operated as a consequence of a disturbance applied in a particular power system simulation.

Using dummy variables, we then transform a cascading failure into a two-dimensional matrix format with the matrix columns indexed by relay name whilst the matrix rows are indexed by event number; event number represents the ordering in which unique relays operated. Elements within this matrix are binary-valued, indicating which unique protection relays have operated in what order during a time domain simulation. Where a matrix element is equal to 1, this indicates that the specific relay r operated at the specific event number e; where as a matrix element with value 0 indicates that a relay r did not operate at that event number e. This means that only a single 1 may exist in each row and in each column of this matrix representation of a particular cascading failure sequence. This

TABLE I
AN EXAMPLE OF OBSERVED CASCADING FAILURE SEQUENCES

Sim. No.	Sequence		
1	Relay A	Relay B	
2	*no relays operated*		
3	Relay C	Relay A	Relay D

TABLE II
AN INDICATOR MATRIX REPRESENTATION OF THE CASCADING FAILURE
FROM SIMULATION NUMBER 1 IN TABLE I

		Relay Name	
Sim No.	Event No.	Relay A	Relay B
1	1	1	0
1	2	0	1

TABLE III
AN INDICATOR MATRIX REPRESENTATION OF THE CASCADING FAILURES
FROM SIMULATIONS NUMBER 1 AND 2 IN TABLE I

		Relay Name	
Sim No.	Event No.	Relay A	Relay B
1	1	1	0
1	2	0	1
2	1	0	0
2	2	0	0

allows the detailed evolution of relay events to be captured, *i.e.* what relays trip in what order with respect to other relays.

C. Cascading Failure Sequence to Cascading Failure Indicator Matrix Transformation

This encoding process can be explained through an example. Let us assume three separate power system simulations are performed, resulting in the observed cascading failures as in Table I

The cascading failures in Table I include four unique relays labelled: Relay A, Relay B, Relay C, and Relay D. Note that no relays were observed with the second simulation, thus the cascading failure from the second simulation is an empty sequence.

We may then transform the first cascading failure—which has two events—into an indicator matrix representation as in Table II.

If we transform the second simulation cascading failure—even though it is an empty sequence—into an indicator matrix representation and append with the indicator matrix representation of the cascading failure result from the first simulation in Table II, we then have Table III.

As a cascading failure—with three events as well as two new previously unseen unique relays—was observed in the cascading failure from Simulation Number 3 in Table I, appending the indicator matrix representation of the cascading failure observed in Simulation Number 3 to the indicator matrix in Table III increases the Event Number and Relay Name dimension lengths, giving Table IV.

TABLE IV
AN INDICATOR MATRIX REPRESENTATION OF THE CASCADING FAILURES
FROM ALL SIMULATIONS IN TABLE I

		Relay Name			
Sim No.	Event No.	Relay A	Relay B	Relay C	Relay D
1	1	1	0	0	0
1	2	0	1	0	0
1	3	0	0	0	0
2	1	0	0	0	0
2	2	0	0	0	0
2	3	0	0	0	0
3	1	0	0	1	0
3	2	1	0	0	0
3	3	0	0	0	1

D. Overall Cascading Failures Indicator Matrix

We append multiple cascading failure matrices together for a total of $N(K+2)$ separate time domain simulations observed from N separate blocks of simulations and K input variables, into a results matrix $\mathbf{C}_N$ with elements $c_{s_N,r_N} \in \{0,1\}$. The $N(K+2)$ separate time domain simulations stem from simulation block matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A}_i^{\mathbf{B}}$ and their vectors of results $f(\mathbf{A})$, $f(\mathbf{B})$, and $f(\mathbf{A}_i^{\mathbf{B}})$ as explained in Sec. II-E.

$$\mathbf{C}_N = [c_{s_N,r_N}], \ s_N \in \mathcal{S}_N, r_N \in \mathcal{R}_N \quad (1)$$

$$\mathcal{S}_N = \{(n,k,e_N)\,|\,n \in \mathcal{N}_N, k \in \mathcal{K}, e_N \in \mathcal{E}_N\} \quad (2)$$

$$\mathcal{N}_N = \{1,2,\cdots,N\} \quad (3)$$

$$\mathcal{K} = \{1,2,\cdots,K,K+1,K+2\} \quad (4)$$

$$\mathcal{R}_N = \{1,2,\cdots,R_N\} \quad (5)$$

$$\mathcal{E}_N = \{1,2,\cdots,E_N\} \quad (6)$$

In (1)-(6), K is the number of independent input variables to which we are interested in determining the sensitivity of cascading failures. Note in (1)-(6) the subscript N; this indicates that the subscripted quantity is after N simulations blocks have been performed. Therefore, $\mathcal{R}_N$ is the set of all unique relays observed to have operated during the N simulation blocks; similarly, $\mathcal{E}_N$ is the set of event numbers observed from the longest cascading failure seen during the N simulation blocks, i.e. equal in length to the cascading failure with the most relay operations seen in N simulation blocks. R_N is the number of unique relays observed to have operated during the N simulation blocks; E_N is the largest number of cascading failure events within any of the sequences observed to have occurred during the N simulation blocks.

As independent simulation blocks are simulated via time domain simulations, observed cascading failure sequences (if any) are transformed to a matrix representation and appended together. Unique relay columns are appended with any relays which have not already been observed in previous simulation blocks. Event number rows are appended such that the number of unique relay event rows is equal to the longest observed cascading failure seen in the simulated blocks so far. This means that all cascading failures from earlier simulation blocks

are included when additional blocks are simulated, so that the cascading failures result indicator matrix $\mathbf{C}_{N+1}$ may be defined as in (7)-(12).

$$\mathbf{C}_{N+1} = [c_{s_{N+1},r_{N+1}}], \ s_{N+1} \in \mathcal{S}_{N+1}, r_{N+1} \in \mathcal{R}_{N+1} \quad (7)$$

$$\mathcal{S}_{N+1} = \{(n,k,e_{N+1})\,|\,n \in \mathcal{N}_{N+1}, k \in \mathcal{K}, e_{N+1} \in \mathcal{E}_{N+1}\} \quad (8)$$

$$\mathcal{N}_{N+1} = \{1,2,\cdots,N,N+1\} \quad (9)$$

$$\mathcal{K} = \{1,2,\cdots,K,K+1,K+2\} \quad (10)$$

$$\mathcal{R}_{N+1} = \{1,2,\cdots,R_N,\cdots,R_{N+1}\} \quad (11)$$

$$\mathcal{E}_{N+1} = \{1,2,\cdots,E_N,\cdots,E_{N+1}\} \quad (12)$$

Note that even though $\mathcal{N}_N$—the simulation block index set after N simulation blocks—is a subset of $\mathcal{N}_{N+1}$—the simulation block index set after $N+1$ simulation blocks—no additional unique relays or a greater number of unique events may necessarily have been observed in the $(N+1)$-th simulation block compared to the preceding (N)-th simulation block. Therefore the number of unique relays may be the same between the N simulation block and the $N+1$ simulation block; similarly, the longest cascading failure observed in the $N+1$ simulation block may be the same length as the longest cascading failure observed in the N simulation block.

In the example in Sec. II-C, $R_1 = 2$, and $E_1 = 2$ since $N = 1$ for the cascading failure shown in Table II. In Table III, $N = 2$ but $R_2 = 2$ and $E_2 = 2$ since there are no additional relays observed and no longer cascading failure than what is observed in the preceding Table II. However after another simulation block sample, Table IV gives $N = 3$, $R_3 = 4$, and $E_3 = 3$ as there are two new relays and a longer cascading failure of three relay operations observed.

This sequential approach of generating a new simulation block, observing any cascading failures from the new simulation block, and appending the results to all previously observed cascading failures results (including results with no observed cascading failures) from previous simulation blocks allows sequential estimation of sensitivity indices after each simulation block.

Given a cascading failure indicator matrix including the results of at least one simulation block, we then apply an extended form of variance-based sensitivity index estimators to calculate scalar values of sensitivity index estimates for each input variable; these estimators are shown in (23) and (24) and are explained in Sec. II-H. After estimating sensitivity indices, we may rank them from largest to smallest thus indicating the relative importance of the input variables with respect to cascading failure sensitivity, i.e. a very large estimate for a particular input variable implies that cascading failures are very sensitive to this variable and therefore this would be ranked highly relative to the other input variables.

E. Experimental Design for Estimation of Variance-based Sensitivity Indices

1) Simulation Block Matrices: A matrix of size $N \times 2K$ is created by generating N sample points within a $2K$-dimension unit hypercube, either completely randomly as in a true Monte Carlo approach or using a quasi random sequence generator such as a Sobol sequence. The first K columns of the matrix are then used after rescaling to construct a matrix $\mathbf{A}$, and the last K columns of the matrix are used after rescaling to construct a matrix $\mathbf{B}$. Note that the rescaling of the $N \times K$ matrices used to form $\mathbf{A}$ and $\mathbf{B}$ is to ensure that the i-th columns in $\mathbf{A}$ and $\mathbf{B}$ comply with the upper and lower bounds of the actual input variable X_i.

The two constructed matrices $\mathbf{A}$ and $\mathbf{B}$ are K columns wide and N rows long, where K is the number of input variables under study and N is the number of simulation blocks we wish to simulate. A set of matrices $\mathbf{A}_i^{\mathbf{B}}$ is formed from $\mathbf{A}$ and $\mathbf{B}$, where all columns in $\mathbf{A}_i^{\mathbf{B}}$ are from $\mathbf{A}$ except from the i-th column which is from $\mathbf{B}$. Therefore there are K different $\mathbf{A}_i^{\mathbf{B}}$ matrices, where each $\mathbf{A}_i^{\mathbf{B}}$ matrix is N rows long and K columns wide. The matrix elements in $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A}_i^{\mathbf{B}}$ represent scalar realizations of input variables X_i.

The $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A}_i^{\mathbf{B}}$ matrices represent sampling matrices specially designed for estimating certain integrals using Monte Carlo schemes. Interested readers are referred to [8], [15] and the references therein for a more involved discussion and conceptual reasoning behind the $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A}_i^{\mathbf{B}}$ matrices.

2) Result Matrices: Power system simulations are performed using the simulation block matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A}_i^{\mathbf{B}}$ as inputs. Each row of matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A}_i^{\mathbf{B}}$ are separately evaluated via a power system simulation: each row is used to set the input variables X_i for the multivariate function f (*i.e.* the simulation) for all rows in $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A}_i^{\mathbf{B}}$. This results in a total of $N(K + 2)$ simulations being performed: N simulations from $\mathbf{A}$; another N simulations from $\mathbf{B}$; and KN simulations due to all of the $\mathbf{A}_i^{\mathbf{B}}$ matrices. Results vectors are therefore given by $f(\mathbf{A})$, $f(\mathbf{B})$, and $f(\mathbf{A}_i^{\mathbf{B}})$, where $f(\mathbf{A})$, $f(\mathbf{B})$, and $f(\mathbf{A}_i^{\mathbf{B}})$ are all N elements long and there are K separate $f(\mathbf{A}_i^{\mathbf{B}})$ vectors.

F. Estimators for Scalar-valued Outputs

Given a set of result vectors $f(\mathbf{A})$, $f(\mathbf{B})$, and $f(\mathbf{A}_i^{\mathbf{B}})$, the first order effects sensitivity index $\hat{S}_i$ and total effect sensitivity index $\hat{S}_{T,i}$ estimators for the i-th input variable are calculated as shown in (14) and (15), based upon numerators *Formula (b)* and *Formula (f)* respectively in [8, Table II].

$$V[f(\mathbf{A}), f(\mathbf{B})] = \frac{1}{2N} \left[\sum_{n=1}^{N} \left(f(\mathbf{A})_n^2 + f(\mathbf{B})_n^2 \right) - \frac{1}{2N} \left(\sum_{n=1}^{N} (f(\mathbf{A})_n + f(\mathbf{B})_n) \right)^2 \right] \quad (13)$$

$$\hat{S}_i = \frac{\frac{1}{N} \sum_{n=1}^{N} f(\mathbf{B})_n \left(f(\mathbf{A}_i^{\mathbf{B}})_n - f(\mathbf{A})_n \right)}{V[f(\mathbf{A}), f(\mathbf{B})]} \quad (14)$$

$$\hat{S}_{T,i} = \frac{\frac{1}{2N} \sum_{n=1}^{N} \left(f(\mathbf{A})_n - f(\mathbf{A}_i^{\mathbf{B}})_n \right)^2}{V[f(\mathbf{A}), f(\mathbf{B})]} \quad (15)$$

In (13)-(15), subscript n indicates the n-th simulation block out of N simulation blocks studied. The quantity $V[f(\mathbf{A}), f(\mathbf{B})]$ represents the total variance of all (quasi) independent simulation block results in vectors $f(\mathbf{A})$ and $f(\mathbf{B})$.

G. Estimators for Vector-valued Outputs

The estimators $\hat{S}_i$ and $\hat{S}_{T,i}$ in Sec. II-F are only for scalar outputs of f. Some attention has been paid in the literature to variance-based sensitivity indices for vector-valued experiments or models, *i.e.* $\mathbf{Y} = \mathbf{f}(X_1, X_2, \cdots, X_i)$. A first order effects sensitivity index is conceptualized and a corresponding estimator provided for a vector-valued output in [16]. The first order effects sensitivity index S_i for a vector-valued output $\mathbf{Y}$ with J elements of a vector-valued function is shown in (16), where $S_{i,j}$ are the first order effects sensitivity indices calculated individually for the i-th input variable and the j-th output in $\mathbf{Y}$; and $V(\mathbf{Y}_j)$ is the total variance of the j-th output in $\mathbf{Y}$.

$$S_i = \frac{\sum_{j=1}^{J} V(\mathbf{Y}_j) S_{i,j}}{\sum_{j=1}^{J} V(\mathbf{Y}_j)} \quad (16)$$

Although not explicitly addressed in [16], it has been accepted by other researchers that a similar approach to (16) may be used to determine the total effects sensitivity index as in (17); although not an exhaustive paper on the subject, authors in [17, Sec. 2.1] state an equation similar to (17). Note that in (17), $S_{T,i,j}$ are the total effects sensitivity indices calculated individually for the i-th input variable and the j-th output in $\mathbf{Y}$; and $V(\mathbf{Y}_j)$ is the total variance of the j-th output in $\mathbf{Y}$.

$$S_{T,i} = \frac{\sum_{j=1}^{J} V(\mathbf{Y}_j) S_{T,i,j}}{\sum_{j=1}^{J} V(\mathbf{Y}_j)} \quad (17)$$

We assume that the estimators $\hat{S}_i$ and $\hat{S}_{T,i}$ for S_i and $S_{T,i}$ respectively for a vector-valued output are simply calculated from the estimators $\hat{S}_{i,j}$ and $\hat{S}_{T,i,j}$ calculated separately for each scalar-valued output within $\mathbf{Y}$ using the estimators in (14) and (15); this approach is shown in (18) and (19).

$$\hat{S}_i = \frac{\sum_{j=1}^{J} V(\mathbf{Y}_j) \hat{S}_{i,j}}{\sum_{j=1}^{J} V(\mathbf{Y}_j)} \quad (18)$$

$$\hat{S}_{T,i} = \frac{\sum_{j=1}^{J} V(\mathbf{Y}_j) \hat{S}_{T,i,j}}{\sum_{j=1}^{J} V(\mathbf{Y}_j)} \quad (19)$$

H. Extended Variance-based Sensitivity Index Estimators

We propose that the estimators in (18) and (19) may be extended to another dimension which for the purposes of this example we shall assume is of length M, giving (20) and (21). Thus, in (20) and (21) the simulations are assumed to produce

matrix-valued output $\mathbf{Y}$ with length M in one dimension and of length J in the other dimension.

$$\hat{S}_i = \frac{\sum_{m=1}^{M} \sum_{j=1}^{J} \mathrm{V}\left(\mathbf{Y}_{m,j}\right) \hat{S}_{i,m,j}}{\sum_{m=1}^{M} \sum_{j=1}^{J} \mathrm{V}\left(\mathbf{Y}_{m,j}\right)} \tag{20}$$

$$\hat{S}_{T,i} = \frac{\sum_{m=1}^{M} \sum_{j=1}^{J} \mathrm{V}\left(\mathbf{Y}_{m,j}\right) \hat{S}_{T,i,m,j}}{\sum_{m=1}^{M} \sum_{j=1}^{J} \mathrm{V}\left(\mathbf{Y}_{m,j}\right)} \tag{21}$$

We assume that the concatenation of power system simulations $\mathbf{f}$—where $\mathbf{f}$ is constructed from results $f\left(\mathbf{A}\right)$, $f\left(\mathbf{B}\right)$, and $f\left(\mathbf{A}_i^{\mathbf{B}}\right)$—takes in as inputs simulation block matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A}_i^{\mathbf{B}}$ of size $N \times K$ where there are N (quasi) independent simulation blocks for a K variable problem, and returns an aggregated cascading failures indicator matrix $\mathbf{C}_N$, *i.e.* $\mathbf{C}_N = \mathbf{f}\left(\mathbf{A}, \mathbf{B}, \mathbf{A}_i^{\mathbf{B}}, \dots, \mathbf{A}_K^{\mathbf{B}}\right)$ and $\mathbf{C}_N$ is defined as in (1)-(6).

Therefore, all $c_{(n,1,e_N),r_N}$ in $\mathbf{C}_N$ correspond to the cascading failure indication of n-th simulation block, e_N-th unique event number, and r_N-th unique relay due to the $\mathbf{A}$ matrix inputs; similarly, all $c_{(n,K+2,e_N),r_N}$ in $\mathbf{C}_N$ correspond to the cascading failure indication of n-th simulation block, e_N-th unique event number, and r_N-th unique relay due to the $\mathbf{B}$ matrix inputs. And lastly, all $c_{(n,i+1,e_N),r_N}, i \in \{1, 2, \cdots, K\}$ in $\mathbf{C}_N$ correspond to the cascading failure indication of n-th simulation block, e_N-th unique event number, and r_N-th unique relay due to the i-th matrix $\mathbf{A}_i^{\mathbf{B}}$ inputs.

Based upon (20) and (21), we propose estimators $\hat{S}_{i,N}$ and $\hat{S}_{T,i,N}$ after N simulation blocks for the indices S_i and $S_{T,i}$ in (23) and (24) respectively. In (22), (23) and (24), the elements $c_{(n,k,e_N),r_N}$ are within $\mathbf{C}_N$. Estimators $\hat{S}_{i,N}$ and $\hat{S}_{T,i,N}$ are calculated by averaging over sensitivity indices individually calculated for each unique relay and unique event number, weighted by their corresponding unique relay variance and unique event number variance.

$$V_N = \frac{1}{2N} \sum_{e_N=1}^{E_N} \sum_{r_N=1}^{R_N} \left[\sum_{n=1}^{N} \left(c_{(n,1,e_N),r_N}^2 + c_{(n,K+2,e_N),r_N}^2 \right) \right.$$
$$\left. - \frac{1}{2N} \left(\sum_{n=1}^{N} \left(c_{(n,1,e_N),r_N} + c_{(n,K+2,e_N),r_N} \right) \right)^2 \right] \tag{22}$$

$$\hat{S}_{i,N} = \frac{1}{NV_N} \sum_{e_N=1}^{E_N} \sum_{r_N=1}^{R_N} \left[\sum_{n=1}^{N} c_{(n,K+2,e_N),r_N} \right.$$
$$\left. \cdot \left(c_{(n,i,e_N),r_N} - c_{(n,1,e_N),r_N} \right) \right] \tag{23}$$

$$\hat{S}_{T,i,N} = \frac{1}{2NV_N} \sum_{e_N=1}^{E_N} \sum_{r_N=1}^{R_N} \left[\sum_{n=1}^{N} \right.$$
$$\left. \left(c_{(n,1,e_N),r_N} - c_{(n,i,e_N),r_N} \right)^2 \right] \tag{24}$$

In (22), (23) and (24), all variables and parameters have the same meaning as those in (1)-(6). The quantity V_N in (22) represents the total variance calculated across all N (quasi) independent simulation blocks simulated so far. The indices calculated by (23) and (24) can be used to rank power system variables with respect to power system cascading failures.

I. Output Aggregations

In this paper, we also consider three of what we refer to as Aggregations: Overall Cascading Failure Aggregation; Relay Type Aggregation; and Network Area Aggregation.

These Aggregations represent different ways of aggregating cascading relay operations. The Overall Cascading Failure Aggregation aggregates all relay operations within the entire model, and can be considered as a general measure of cascading failure sensitivity for the entire modelled power system. The Relay Type Aggregation aggregates different types of relays into different sets of cascading results, therefore allowing sensitivity of cascading failures of specific relay types to be measured separately. Lastly, the Network Area Aggregation aggregates relays by network area rather than relay type, and so can be viewed as cascading failure sensitivity within individual network areas.

These Aggregations allow the comparison of sensitivities of cascading failures for the same system input variable values. For example, cascading failures in one part of a network may have different sensitivities to a set of input variables than cascading failures in another part of the network for the same set of input variables. A similar comparison can be made with relay types; for example, voltage relay cascading failures are likely to have higher sensitivity to voltage relay thresholds and this can be seen via the Relay Type Aggregation.

Note that in the preceding discussion the aggregation of relay indication is assumed to be performed across all unique relays and cascading failure event numbers, such as in (22), (23) and (24). We use this aggregation to perform sensitivity analysis on the total cascading failures within the entire power system.

However it is possible to consider the cascading failure indicator matrix results as being assembled from groups of cascading failure indicator matrix results. It is therefore straightforward to aggregate unique relays and event numbers within each of these groups of output cascading failure indicator matrices, then apply (22), (23) and (24) to each of the groups of indicator matrix results.

These alternative aggregations (aside from the total system cascading failures aggregation implied until now) could be applied to different kinds of groups depending upon the engineer's specific needs.

J. Quasi Monte Carlo Sampling

We use a quasi Monte Carlo approach using Sobol sequences to sequentially generate the (quasi) independent simulation blocks. After each simulation block, the point estimators for first order $\hat{S}_{i,N}$ and total $\hat{S}_{T,i,N}$ effects sensitivity indices are calculated. We use implementations of functions from the `SALib` library [18] which are themselves based on [19] to sequentially generate the simulation block matrices for subsequent use in time domain simulations. Sobol sequences

are an example of a low-discrepancy sequence, which may be used to generate points within a hypercube such that the points are highly dispersed to efficiently explore the hypercube volume.

The quasi Monte Carlo approach allows representative sampling of the input variables such that the estimated first order and total effects indices will converge towards population parameters with increasing number of simulation blocks. This means that the estimates will be representative whilst avoiding computational intractability of exhaustively simulating all possible scenarios from the defined input variables.

K. Scalability of Method

The simulation blocks require $(K + 2)$ simulations per block. This simulations are not dependant on each other, therefore the are embarrassingly parallel. This means that a computing platform with multiple processing cores can be used, with each simulation being run on a separate processor core. Hence, total wall clock time of our proposed method increases with: increasing simulation block number, and increasing number of input variables. Total time can be significantly decreased by increasing number of processor cores available for parallel power system simulations per simulation block.

In our Case Studies, we have simulated 2000 blocks on a PC with 16 GiB RAM and an Intel Core i7-4790 processor utilising 4 parallel processes. The simulation time is the most time consuming part of the methodology. The Base Case Study took 95.5 hours in total, where the power system simulations took 90.5 hours while the calculation of first order and total effect estimates took 5 hours. The UFLS Relay Frequency Threshold Case Study took 148 hours in total, where the power system simulations took 139 hours while the calculation of first order and total effect estimates took 9 hours. Lastly, the Protection Relay Threshold Case Study took 144 hours in total, where the power system simulations took 138 hours while the calculation of first order and total effect estimates took 6 hours.

III. Test System Case Studies and Aggregations

We now demonstrate our proposed method on three Case Studies and three Aggregations using a modified form of the IEEE 39 bus model [20]. This modified IEEE 39 bus model is the same model as used within the work in [13]. The reason for using a modified version of the IEEE 39 bus network is twofold: i) to include the action of protection devices and relevant mechanisms (e.g. actions of transformer tap changers) that are important for the representation of cascading failures, and ii) to include the dynamics of converter interfaced units (by adding wind generation in this case) that can affect overall system dynamic behaviour and consequently the evolution of cascades in power systems with high penetration of converter interfaced generation.

A. Modified IEEE 39 bus model

Our time domain power system model is identical to the IEEE 39 bus model in *Base Scenario* in [13]. We shall explain how this modified model differs from the IEEE 39 bus model [20].

1) Demand Splitting: All demands were split into five separate demand blocks; where four of these blocks were equal to 10% of total demand at that demand location, and the fifth block equal to 60% of total demand at that demand location.

2) Demand Block UFLS: UFLS protection relays were added to each of the four 10% demand blocks at each demand location.

3) Wind Farms: Three aggregated wind farms were added at busbars 5, 14, and 16 (as numbered in the original IEEE 39 bus model [20]). Each wind farm was modelled as an aggregated single power converter interfaced machine.

4) Synchronous Machine Dynamics: All synchronous machines used sixth-order generator models. Each generator was equipped with: an automatic voltage regulator; a power system stabiliser; a governor; and an over-excitation limiter.

5) Transformers: Distribution transformers were included with tap-changers, which were active during the time domain simulations. The distribution transformers interconnected demand blocks with the transmission network.

6) Protection Relays:

All synchronous generators were modelled with: under-voltage, over-voltage, under-speed, over-speed, and out-of-step protection relays. All three wind farms were modelled with: over-voltage, under-voltage, over-frequency, and under-frequency protection relays. We refer readers to [13] for more detail on how these protection relays' parameters were set.

B. Simulation Setup

We performed 120 s simulations in PowerFactory 2019 SP5 build 19.0.7. Each simulation was initialised from an optimal power flow (OPF). Each OPF was itself initialised by setting: total system demand level; and, individual wind farm productions from each of the three wind farms. The OPFs were used to: commit synchronous generators in the rest of the model to supply the remaining system demand which was not already met by the wind farms' productions; and, to calculate steady-state network variable values in preparation for the phasor time domain simulations.

In addition, the synchronous machines were assumed to be comprised of four equally-sized aggregated units acting in a unified manner and loaded sequentially. Hence, the model (aggregated) machines' base powers were rescaled based upon the total generation committed by the OPF, and the assumed number of sequentially loaded machines required to meet the total from the aggregated machine.

On running each time domain simulation, a three phase short circuit fault was applied to the midpoint of one of the 34 circuits within the model. This fault was applied at $t = 1$ s, and cleared at $t = 1.07$ s by disconnecting the fault and switching out the circuit. This approach models a combined ideal fault and fault clearance of 70 ms. Following fault clearance, the time domain simulation was run until $t = 120$ s. If any protection relay operation events were recorded during the simulation, these were stored for later conversion into a cascading failure indicator matrix as explained in Sec. II.

We performed 2000 simulation blocks for all Case Studies and all Aggregations, *i.e.* substituting $N = 2000$ in Sec. II.

Note that we did not consider relays which would be expected to clear the fault, to be part of any cascading failure, i.e. the actions of ideal transmission line relays to remove the ideal in-zone applied fault are explicitly excluded from the cascading failure sequence detection and therefore the cascading failures indicator matrix.

C. Case Studies

In this paper, we study three Case Studies which use the same model construction and data as in [13]. Our three Case Studies are as follows.

1) A Base Case Study, where system demand, wind farm generation, and disturbance location are considered.
2) A UFLS Relay Frequency Threshold Case Study, where UFLS tripping threshold design function variables are investigated in conjunction with the same power system variables in the Base Case Study.
3) A Protection Relay Threshold Case Study, where the effects of individual wind farm over-voltage relay tripping thresholds are explicitly quantified simultaneously in addition to the same power system variables in the Base Case Study.

1) Base Case Study: Our Base Case was a five variable sensitivity analysis of cascading failures with respect to system operational conditions and short circuit fault disturbance location. The system operational conditions represented system demand and generation, and in particular the output of the three added wind farms. The total system demand was apportioned in equal percentage share to all demands in the model. Non wind-farm generation was committed to satisfy the shortfall between wind farm output and total system demand using an OPF. The aim of this Case Study was to initially determine the extent of the impact of power system variables on the evolution of cascading failures. In addition, it was also used to determine if the ranking of the Base Case Study variables was qualitatively different from the variable rankings in the UFLS Relay Frequency Threshold and Protection Relay Threshold Case Study Case Studies.

The system condition and disturbance location were five input variables, listed as follows using a short name identifying label and a description: Fault Loc. fault location; Demand total system demand and WF 1, WF 2, and WF 3 which were the first, second and third wind farms productions as a percentage of each wind farms' installed capacity respectively. Bounds were 1 and 34 for Fault Loc., 0.7 and 1.2 for Demand, and 0.0 and 0.1 for WF 1, WF 2, and WF 3.

All variables apart from Fault Loc. were assumed to be real-valued, and so the transformation from the Sobol samples between $[0, 1]$ in these variables to be within the actual variable bounds was trivial. Real-valued Sobol samples between $[0, 1]$ for Fault Loc. were mapped to the integer-valued interval $[1, 34]$ using the inverse transform method assuming all integers in Fault Loc. were equally probable, which therefore specified which of the 34 circuits to apply the short circuit fault.

2) UFLS Relay Frequency Threshold Case Study: We also studied the sensitivity of cascading failures to the under-frequency tripping threshold on the UFLS protection relays spread throughout the model at demand locations. The choice to study UFLS threshold is motivated by a design challenge, i.e. the specific threshold of UFLS can be seen a design parameter and understanding the sensitivities with respect to it is of use to system operators and planners. Hence, this Case Study had seven input variables: all five input variables from the Base Case Study, and an additional two variables as follows: $F^{\text{coeff.}}$ coefficient of linear UFLS frequency threshold design equation; and $F^{\text{const.}}$ constant of linear UFLS frequency threshold design equation. Bounds were 0.0 and 0.1 for $F^{\text{coeff.}}$, and 0.1 and 0.9 for $F^{\text{const.}}$ respectively.

As each demand had five blocks where four were equipped with UFLS relays, we assumed the following design equation to set frequency thresholds f_l^{th} for block relay l based upon two free design variables $F^{\text{coeff.}}$ and $F^{\text{const.}}$. Block relays were numbered 1, 2, 3, and 4 corresponding to the four demand blocks per demand location. The design equation (25) was motivated by [11], which amongst other work indicates typical UFLS relay tripping thresholds.

$$f_l^{\text{th}} = 60 - F^{\text{coeff.}} \cdot (l - 1) - F^{\text{const.}}, \quad l \in \{1, 2, 3, 4\} \quad (25)$$

3) Protection Relay Threshold Case Study: Our last Case Study investigates the sensitivity of cascading failures to the setting of over-voltage tripping thresholds on each of the three aggregated wind farms. The justification for investigating this is due to the preponderance of wind farm over-voltage tripping events observed in the cascading failure results in [13]. The choice to study activation thresholds of a protection devices is motivated by understanding sensitivities related to hidden failures of protection devices, or the impact of incorrect settings. This Case Study had eight input variables: all five input variables from the Base Case Study, and an additional three variables with bounds 1.045 and 1.21 as follows: W1 OV, W2 OV, and W3 OV; representing first, second and third wind farm over-voltage relay tripping thresholds respectively.

IV. DISCUSSION AND COMPARATIVE ASSESSMENT OF RESULTS

The point estimates for all input variables in all three Case Studies for the Overall Cascading Failure Aggregation are included in Table V. Only a subset of results are included here for the Relay Type and Network Area Aggregations. The point estimates for all input variables for the UFLS Relay Frequency Threshold Case Study for the Relay Type Aggregation are given in Table VI. The point estimates for all input variables for the Base Case Study for the Network Area Aggregation are given in Table VII. In Table VII, the area aggregations are as follows based upon [20]: Area 1: all relays at busbars 16, 19–24, and 33–36; Area 2: all relays at busbars 1–3, 17–18, 25–30, and 37–38; and Area 3: all relays at busbars 4–15, 31–32, and 39.

We found that the estimated first order effects sensitivity indices for all input variables over all three Case Studies

TABLE V
FIRST ORDER AND TOTAL EFFECTS SENSITIVITY INDEX POINT ESTIMATES AGGREGATED ACROSS CASCADING FAILURES

Input Variable	First Order Effects						Total Effects					
	Base		UFLS Design Params.		Prot. Relay Thres.		Base		UFLS Design Params.		Prot. Relay Thres.	
	Mag.	Rank	Mag.	Rank	Mag.	Rank	Mag.	Rank	Mag.	Rank	Mag.	Rank
Fault Loc.	0.10319	1	0.036383	1	0.081407	1	0.94748	1	0.94154	1	0.99618	1
Demand	0.0095928	2	0.001314	4	0.003392	3	0.82575	2	0.7586	4	0.89187	2
WF 1	0.0018148	4	-0.00003413[a]	7	0.000424	4	0.67888	5	0.53242	6	0.7068	4
WF 2	0.0018148	4	0.00030717	6	0.000424	4	0.69185	4	0.58325	5	0.69387	5
WF 3	0.0023334	3	0.00051195	5	-0.00084799[a]	7	0.72309	3	0.5258	7	0.7422	3
$F^{\text{coeff.}}$	—	—	0.0034813	3	—	—	—	—	0.76952	3	—	—
$F^{\text{const.}}$	—	—	0.014028	2	—	—	—	—	0.83578	2	—	—
W1 OV	—	—	—	—	0[b]	5	—	—	—	—	0.18635	7
W2 OV	—	—	—	—	0.0046639	2	—	—	—	—	0.36994	6
W3 OV	—	—	—	—	-0.000424[a]	6	—	—	—	—	0.10812	8

[a] Point estimate outwith the closed interval $[0, 1]$.

TABLE VI
FIRST ORDER AND TOTAL EFFECTS SENSITIVITY INDEX POINT ESTIMATES AGGREGATED BY RELAY TYPES ON THE UFLS RELAY
FREQUENCY THRESHOLD CASE STUDY

Input Variable	First Order Effects						Total Effects					
	Freq. Relays		OOS Relays		Volt. Relays		Freq. Relays		OOS Relays		Volt. Relays	
	Mag.	Rank	Mag.	Rank	Mag.	Rank	Mag.	Rank	Mag.	Rank	Mag.	Rank
Fault Loc.	0.080722	2	0.17865	1	0.037758	1	0.42497	2	0.94197	1	0.91459	2
Demand	0.0092425	3	0	3	-0.0020977[a]	5	0.23789	3	0.7349	2	0.73668	3
WF 1	0.0008644	7	0	3	-0.0034087[a]	7	0.085177	7	0.36542	5	0.36618	7
WF 2	0.0045547	6	0	3	-0.0028843[a]	6	0.090014	5	0.35324	6	0.40433	5
WF 3	0.0046545	5	0	3	0.00026221	3	0.089383	6	0.38166	4	0.38545	6
$F^{\text{coeff.}}$	0.0081121	4	0	3	-0.00078663[a]	4	0.15908	4	0.27609	7	0.72711	4
$F^{\text{const.}}$	0.518044	1	0.016241	2	0.034612	2	0.85551	1	0.41008	3	0.92324	1

[a] Point estimate outwith the closed interval $[0, 1]$.

TABLE VII
FIRST ORDER AND TOTAL EFFECTS SENSITIVITY INDEX POINT ESTIMATES AGGREGATED BY NETWORK AREA ON THE BASE CASE STUDY.

Input Variable	First Order Effects						Total Effects					
	Area 1		Area 2		Area 3		Area 1		Area 2		Area 3	
	Mag.	Rank	Mag.	Rank	Mag.	Rank	Mag.	Rank	Mag.	Rank	Mag.	Rank
Fault Loc.	0.22357	1	0.10036	1	0.18618	1	0.95326	1	0.94819	1	0.92693	1
Demand	0.014749	2	0.0048563	4	0.0079225	2	0.72232	2	0.79643	2	0.73243	2
WF 1	0.0038814	4	0.011331	2	-0.0015845[a]	4	0.50768	5	0.63698	5	0.54863	4
WF 2	0.0031051	5	0.0048563	4	-0.003169[a]	5	0.52243	3	0.64791	3	0.53477	5
WF 3	0.0077627	3	0.0080938	3	0.0007922	3	0.51079	4	0.64265	4	0.55061	3

[a] Point estimate outwith the closed interval $[0, 1]$.

were very small and similarly valued near zero, apart from: the $F^{\text{const.}}$ from Sec. III-C2 in the UFLS Threshold Case Study with Frequency Relay Grouping in the Relay Type Aggregation; and, Fault Loc. for all Case Studies and Aggregations This means that with the exception Fault Loc. and the aforementioned $F^{\text{const.}}$ variable, the cascading failures produced during the Cascading Failures, Relay Types, and Network Area Aggregations are insensitive to the sole effects of each of the individual input variables for all Case Studies. The cascading failures produced in all three Aggregations can be regarded as having a small sensitivity to the individual effect of the Fault Loc.

Many of the point estimates first order effects indices are slightly negative; this is indicative of the population parameters being very small positive numbers or zero, and there are insufficient samples to cause these specific point estimates to converge to either 0 or a slightly positive number. Note that the numerator in (23) is not guaranteed to always return non-negative values even if the underlying population indices are zero or small positive numbers. Therefore, negative sample index values should be interpreted by the engineer as indicating zero or small positive valued population indices.

In the UFLS Threshold Case Study with Frequency Relay Grouping in the Relay Type Aggregation, the $F^{\text{const.}}$ variable

has a large first order index. This is expected since this variable is interpreted as a constant term in (25) which is therefore a first order term in the sensitivity analysis.

However all input variables had large total effect sensitivity indices much further towards unity over all three Case Studies. This coupled with near zero first order index results implies that cascading failures are sensitive to interactions between the input variables in all three Case Studies and over all three Aggregations. In the context of power systems, this means that these input variables cannot be considered in isolation with respect to their influence on cascading failures; the influence in all variables at once must be considered due to the interaction between the input variables.

Fault Loc. total effect sensitivity index is often very close to unity showing consistently that the fault location is the most impactful variable considered in our studies. System demand total effect sensitivity index is also large for all three Case Studies over all three Aggregations, but not as large as the Fault Loc. total effect sensitivity index tended to be. The wind farm generation variables total effect sensitivity indices have some, but relatively lower total order sensitivity impact on cascading failures in all three Case Studies and Aggregations; the wind farms' generation tended to have similar index estimates across all three Case Studies and Aggregations. These observations imply to power system engineers that cascading failures themselves in power systems are very sensitive to short circuit fault location and system demand.

One notable example of Fault Loc. not being the ranked input variable for total order effects is the Relay Type Aggregation for the UFLS Relay Frequency Threshold Case Study, with results in Table VI. For the Frequency relay and Voltage relay groupings, the highest ranked input variable is $F^{\text{const.}}$, but this variable is only ranked third in the Out-of-step relay grouping. These results are noticeably different from the Overall Cascading Failure Aggregation results for the same UFLS Relay Frequency Threshold Case Study (results in Table V), where Fault Loc. is ranked first and $F^{\text{const.}}$ second. Our method makes these changes in input variable rankings apparent by allowing consideration of different Aggregations.

Note that in the UFLS Relay Frequency Threshold Case Study, the $F^{\text{coeff.}}$ and $F^{\text{const.}}$ variables had quite significant index estimates. This makes intuitive sense as UFLS relay operation is seen very often in cascading failures in this particular modified IEEE 39 bus model as observed in [13]; changing the thresholds of UFLS relays influences many of the relays within our IEEE 39 bus model and therefore if and when they will trip during cascading failures.

The total effects sensitivity indices for the over-voltage relay thresholds in the Protection Relay Threshold Case Study are ranked smallest out of the eight input variables even though wind farm over-voltage tripping is observed to occur frequently in [13].

These results in Tables V, VI, and VII imply that the variance in the observed cascading failures—*i.e.* the observed operated relays—is qualitatively affected non-linearly by the input variables variances in all three Case Studies and all three Aggregations. The cascading failures are primarily influenced by Fault. Loc. and somewhat highly by Demand. The system

Demand affects generators' commitments and thus impacts system dynamics and therefore system variables observed by relays. Moving fault location will also change the relays which operate during a cascading failure, hence why cascading failures have high sensitivity to Fault Loc.

While Fault Loc. and System Demand seem to be generally important inputs in our Case Studies, it should be noted that this cannot necessarily be generalised for different protection functions investigated or for different systems. In general, our method provides relative rankings for the input variables and needs to be applied in the particular context. The discrepancy of very high total effects sensitivity indices but very small first order effects sensitivity indices indicates that the power system cascading failure is a highly non-linear event. This therefore provides evidence that modelling cascading failures using linear additive models is likely to lead to simplistic assumptions and therefore neglect of important aspects related to how cascading failures evolve. Thus, methods to more efficiently and tractably perform global sensitivity analysis incorporating consideration of non-linearity should be investigated preferentially over attempts to provide contributions on cascading failures reliant on assumed linear additive models.

V. CONCLUSION

Power system cascading failures caused by chains of protection relays operating and eventually leading to blackouts remain high impact low probability events. The capability to assess cascading failures sensitivity to power system variables would offer engineers a useful tool in design of planning or operational schemes to limit cascading failure risk, as well as understand important system parameters and variables when performing studies. Our method effectively determines the cascading failure in terms of relay tripping and the sequence of relay tripping events. This is an advantage of our method as it allows the quantification of the impact of input power system variables on this detailed sequence.

It is well known that cascading failure propagation is dependant on system operational conditions, including: generator commitment and dispatch, system demand, fault location, and relay thresholds. Investigating how cascading failures themselves may be affected by these influences can allow engineers to create targeted mitigations. This is an improvement on previous sensitivity approaches, which typically look at quantities such as lost load due to a cascading failure, rather than the cascading failure itself.

We consider cascading failures to be encoded by an ordered progression of operated protection relays which captures the evolution of cascading failures in detail, which allows us to apply variance-based sensitivity analysis using our proposed method to the matrix to quantify the sensitivity of observed cascading failures to input system variables.

We applied our proposed method to a modified version of the IEEE 39 Bus model via three Case Studies and three Aggregations. The Aggregations are: Overall Cascading Failures, Relay Type, and Network Area. Overall Cascading Failures aggregation provides cascading failure sensitivity indices for all cascading failures in the entire network; Relay Type Aggregation aggregates cascading failures from different relay types

separately; and Network Area aggregates cascading failures separately based upon the area of the network which cascading failure events occur within. Previous literature has indicated possible influences on cascading failure sensitivity from UFLS relays, and wind farm over-voltage relays. Therefore these three Case Studies assess cascading failure sensitivity to: fault location and operational background; design equation of UFLS relay tripping thresholds; and wind farm over-voltage relay tripping thresholds. The three Aggregations indicate the flexibility of the method for grouping cascading failures for different purposes, such as total network cascading failures, cascading failures by relay types, and cascading failures by network area. This flexibility indicates that the method can be deployed to find cascading failure sensitivity with respect to any kind of user specified cascading failure groupings, as well as user-specified system variables or parameters acting as input variables to the sensitivity analysis.

The Case Study results highlight the utility of the proposed method to generate point estimates of sensitivity indices which can be used by engineers to rank the relative importance of power system variables with respect to power system cascading failures. The results also highlight the non-linear effect of cascading failure mechanisms with respect to studied variables, which indicates to engineers that cascading failures should not be studied using additive linear models.

REFERENCES

[1] H. M. Merrill, M. A. Hossain, and M. Bodson, "Nipping Blackouts in the Bud: Introducing a Novel Cascading Failure Network," *IEEE Power and Energy Magazine*, vol. 18, no. 4, pp. 64–75, 2020.

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

[3] J. Bialek, E. Ciapessoni, D. Cirio, E. Cotilla-Sanchez, C. Dent, I. Dobson, P. Henneaux, P. Hines, J. Jardim, S. Miller, M. Panteli, M. Papic, A. Pitto, J. Quiros-Tortos, and D. Wu, "Benchmarking and Validation of Cascading Failure Analysis Tools," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4887–4900, 2016.

[4] A. J. Flueck, I. Dobson, Z. Huang, N. E. Wu, R. Yao, and G. Zweigle, "Dynamics and Protection in Cascading Outages," in *2020 IEEE Power Energy Society General Meeting (PESGM)*, 2020, pp. 1–5.

[5] J. Song, E. Cotilla-Sanchez, G. Ghanavati, and P. D. H. Hines, "Dynamic Modeling of Cascading Failure in Power Systems," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2085–2095, 2016.

[6] M. Abdi-Khorsand and V. Vittal, "Identification of Critical Protection Functions for Transient Stability Studies," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 2940–2948, 2018.

[7] R. Preece and J. Milanovic, "Assessing the Applicability of Uncertainty Importance Measures for Power System Studies," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2076–2084, 2016.

[8] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index," *Computer Physics Communications*, vol. 181, no. 2, pp. 259–270, 2010.

[9] M. Noebels, R. Preece, and M. Panteli, "AC Cascading Failure Model for Resilience Analysis in Power Networks," *IEEE Systems Journal*, pp. 1–12, 2021.

[10] A. Azzolin, L. Dueas-Osorio, F. Cadini, and E. Zio, "Electrical and topological drivers of the cascading failure dynamics in power transmission networks," *Reliability Engineering & System Safety*, vol. 175, pp. 196–206, Jul. 2018.

[11] H. H. Alhelou, M. H. Golshan, T. Njenda, and N. Hatziargyriou, "An Overview of UFLS in Conventional, Modern, and Future Smart Power Systems: Challenges and Opportunities," *Electric Power Systems Research*, vol. 179, 2020.

[12] T. Weckesser and T. Van Cutsem, "Identifying Plausible Harmful N-K Contingencies: A Practical Approach Based on Dynamic Simulations," in *2018 Power Systems Computation Conference (PSCC)*, 2018, pp. 1–8.

[13] G. A. Nakas and P. N. Papadopoulos, "Investigation of the Impact of Load Tap Changers and Automatic Generation Control on Cascading Events," in *2021 IEEE Madrid PowerTech*, 2021, pp. 1–6.

[14] K. Zhou, I. Dobson, Z. Wang, A. Roitershtein, and A. P. Ghosh, "A Markovian Influence Graph Formed From Utility Line Outage Data to Mitigate Large Cascades," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 3224–3235, Jul. 2020.

[15] A. Saltelli, "Making best use of model evaluations to compute sensitivity indices," *Computer Physics Communications*, vol. 145, no. 2, pp. 280–297, May 2002.

[16] F. Gamboa, A. Janon, T. Klein, and A. Lagnoux, "Sensitivity indices for multivariate outputs," *Comptes Rendus Mathematique*, vol. 351, no. 7-8, pp. 307–310, 2013.

[17] S. Xiao, Z. Lu, and P. Wang, "Multivariate global sensitivity analysis for dynamic models based on energy distance," *Structural and Multidisciplinary Optimization*, vol. 57, no. 1, pp. 279–291, 2018.

[18] W. Usher, J. Herman, T. Iwanaga, L. Teixeira, N. Cellier, C. Whealton, D. Hadka, xantares, bernardoct, F. Rios, C. Mutel, E. Cederstrand, TobiasKAndersen, J. v. Engelen, ANtlord, H. Kranas, V. K. Dixit, and bsteubing, "SALib/SALib: SALib v1.3.12," Nov. 2020.

[19] S. Joe and F. Kuo, "Constructing Sobol' sequences with better two-dimensional projections," *SIAM Journal on Scientific Computing*, vol. 30, no. 5, pp. 2635–2654, 2007.

[20] A. Pai, "Appendix A 10 Machine 39 Bus System Data," in *Energy Function Analysis for Power System Stability*, ser. Kluwer International Series in Engineering and Computer Science. Boston, MA: Kluwer Academic Publishers, 1989, vol. Power Electronics and Power Systems, pp. 223–227.

Aaron S. C. Leavy received his M.Eng. degree in Electrical and Mechanical Engineering from the Department of Electronic and Electrical Engineering at the University of Strathclyde, Glasgow, U.K. in 2015. He is currently pursuing his Ph.D. degree and is employed as a Research Assistant, both within the Department of Electronic and Electrical Engineering at the University of Strathclyde. His interests in power systems engineering include: analysis, planning, operation, design, control, and stability.

Georgios A. Nakas received the Dipl. Eng. degree from the Department of Electrical and Computer Engineering at the Aristotle University of Thessaloniki, Greece, in 2017 and he is currently pursuing his Ph.D. degree at the Department of Electronic and Electrical Engineering at the University of Strathclyde, Glasgow, U.K. His main research interests include modelling of electric power systems and machine learning applications on power system dynamics.

Panagiotis N. Papadopoulos (S'05-M'14) received the Dipl. Eng. and Ph.D. degrees from the Department of Electrical and Computer Engineering at the Aristotle University of Thessaloniki, in 2007 and 2014, respectively. Currently, he is a Senior Lecturer and a UKRI Future Leaders Fellow in the Department of Electronic and Electrical Engineering at the University of Strathclyde. His research interests are in power system stability and dynamics, focusing on tackling the increasing uncertainty and complexity in the system dynamic behaviour.