External System Generator Outage Localization Based on Tie-Line Synchrophasor Measurements

Zhen Dai ⓘ, *Student Member, IEEE*, and Joseph Euzebe Tate, *Member, IEEE*

Abstract—An identification algorithm is proposed for generator outages in external systems given tie-line flow measurements and pre-outage linear sensitivity factors. The problem is challenging due to limited information available to operators in interconnected systems. To overcome the underdetermined nature of the problem (the number of tie-line measurements is smaller than the number of generators), a clustering method based on pivoted QR decomposition is implemented so that the outage location can be identified to the area of origination. Two test systems, the 68-bus New England system and the 500-bus synthetic system, were used for validation. The results demonstrate that with limited knowledge of the external system, the algorithm is able to identify the correct generator cluster in all cases. Another advantage of the proposed algorithm is its high efficiency, which enables detection within sub-milliseconds. In addition, an estimation of the cluster injection change is also provided by the algorithm.

Index Terms—Generator outage, event location estimation, phasor measurement unit (PMU), interconnected power systems, linear sensitivity factor.

I. INTRODUCTION

OUTAGE identification is crucial to prevent stability problems and cascading blackouts, because loss of generation generally leads to frequency deviations and voltage decreases. For interconnected systems, identification of external generator outages is more difficult due to limited knowledge of the external system. In [1], the geographical coordinates of the generator outage are estimated based on the time of arrival of the frequency drop due to generator tripping. The key assumption underlying this work is that a disturbance in frequency will propagate at a constant speed throughout the system. The method is simple to implement, and it requires only frequency measurements of the system. However, the accuracy depends on the measurement device coverage and a correct estimation of the propagation speed, which may vary in different regions. The authors in [2] proposed an adaptive solution to overcome the constant speed assumption. However, it requires extensive off-line simulation

to form a database of different propagation speeds in each partitioned region.

Both of these previous methods of identifying generator outages rely on a network of novel monitoring devices (the FNET system [3]) capable of measuring frequency at a high resolution (in comparison to traditional SCADA monitoring systems). Phasor measurement units (PMUs), which are being increasingly deployed throughout the power grid, can be used for these methods because they provide accurate frequency measurements. In addition to high-resolution frequency measurements, PMUs can also provide time-synchronized voltage and current phasors at a higher temporal resolution than traditional SCADA systems (e.g., one sample per cycle, compared to traditional SCADA scan rates of one sample every few seconds). We propose to use high-resolution boundary tie-line measurements, captured via PMUs, to perform generator outage identification within the external system. By relying on an alternative data set, this approach is complementary to the methods previously introduced for wide-area generator outage identification [1], [2].

The main challenge in using tie-line measurements to identify external generator outages is the lack of a comprehensive, up-to-date model for the external system, which would be needed to (for example) perform real-time dynamic simulation of the external generator outages. Instead, only limited data are typically disclosed between control areas—usually in the form of linear sensitivity factors (LSFs). For example, in North America, the NERC Interchange Distribution Calculator (IDC) uses shared linear sensitivity factors (LSFs) to solve interconnection-wide congestion problems in the North American Eastern Interconnection [4]. While LSFs do not represent a full model of the external system's behavior, they do provide critical information; for example, two commonly used LSFs—the injection shift factor (ISF) and the power transfer distribution factor (PTDF)—are used in contingency analyses and market studies [5]–[7] to relate bus injection changes to flow changes. Traditionally, ISFs represent a relationship between power flows on transmission lines and injected power at system buses and can be either derived based on the dc or ac power flow model. The ac approach requires linearization around an operating point of the power flow equations. By contrast, the dc ISF is completely determined by the network model.

We assume the only information available for use in the generator outage identification are tie-line measurements between the internal and external system and a matrix of LSFs relating tie-line flows to external generator injections. The LSFs are assumed accurate, meaning that there is no unknown topological

errors prior to the investigated generator outages. Compared to the number of generators in the external system, the number of measurements is usually very limited, resulting in an underdetermined problem. One method of addressing this is to reduce the dimensionality of the search space by clustering similar generators together. Using clusters has been a widely adopted approach for model order reduction in large scale power systems. For example, by replacing coherent generators with dynamic aggregated equivalents [8], [9], transient stability simulation can be sped up significantly. However, unlike in coherency studies, we focus on the LSF matrix based on network models rather than dynamic models. There is a vast literature on techniques for matrix dimension reduction, subset selection and their applications [10]. A method based on QR decomposition is chosen in this paper because, as a traditional factorization method, it can be computed with great efficiency, which makes it suitable for online application. In addition, QR decomposition with column pivoting is widely adopted in numerical computation tools for its simplicity and practical reliability [11]. The permutation matrix of the pivoted QR decomposition provides a natural ordering to pick a representative generator within each group. In summary, the clustering technique is proposed to overcome limited observability (not enough PMUs) and potential identifiability problems (nearly identical generators). A few examples in which the QR decomposition has been successfully used in other domains are: rule base reduction in fuzzy modeling [12], model order reduction for multi-agent systems [13], video summarization based on hierarchical keyframes [14], and spectral clustering of graphs [15].

Based on the QR decomposition of the LSF matrix, only N_l (usually, equal to the number of monitored tie-lines) representative generators are preserved. These generators' LSFs are then used to approximate the original LSF matrix; the remaining generators are categorized into one of N_l groups based on the similarity of their LSFs to the representative generators' LSFs. The estimated bus injection changes can then be approximated by inverting the reduced-order model that only relates the tie-line flows to the representative generators' injections.

One important characteristic of generator outages is that the outaged generator will have the largest injection change, while the remaining generators will increase power outputs to compensate the loss. In modeling this response, the effects of governor and machine dynamics are neglected, since the proposed method uses synchrophasor measurements sampled within half a cycle (i.e. the first sample after detection) following an outage—when the system responses are not determined by control dynamics or inertia [16]. Thus, the group with the largest injection change should be the group in which the outage occurred, and this is the criteria used in the outage identification algorithm.

Fig. 1 demonstrates how the proposed algorithm can be integrated into the energy management system (EMS). PMU measurements (from tie-lines and possibly locations in the external system) are collected and passed onto detection algorithms. Once an outage is detected, the LSF-based identification algorithm is initiated. The results—outage cluster location and injection change estimation—can be used for various applications, e.g., sending customized alarms, initiating coordination between
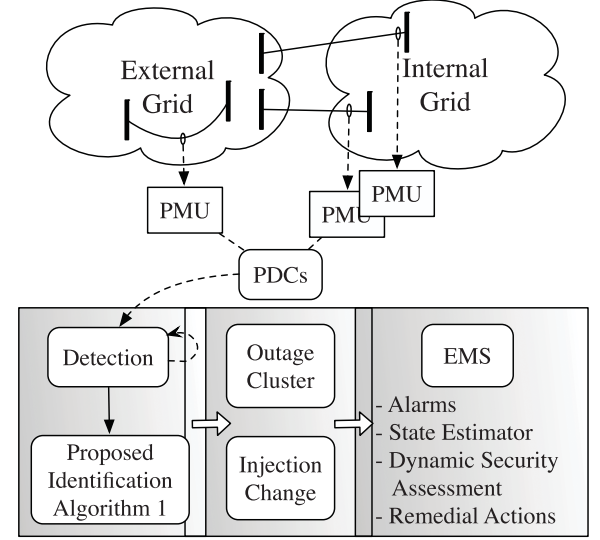

Fig. 1. Integration of the proposed algorithm into utility control centers.

operators, updating models for state estimator and dynamic security assessment, and initiating remedial actions.

The remainder of the paper is organized as follows. A review of LSFs is presented in Section II, followed by the proposed identification algorithm in Section III. Experimental results from testing the method on two systems—the 68-bus New England system [17], [18] and the synthetic 500-bus system [19]—are provided in Section IV. The concluding remarks and future work are presented in Section V. The Appendix A includes a detailed demonstration of the identification algorithm for a 5-bus system.

II. REVIEW OF LINEAR SENSITIVITY FACTORS

In this section, a review of LSFs is provided. It should be noted that historically the term linear sensitivity methods are sometimes used interchangeably with dc power flow methods in security analysis studies [7], with the focus being active power. We consider two linear sensitivity factors relating changes in bus injections to changes in tie-line flows. The main focus is on current injection sensitivity factors (CISFs), which (linearly) relate the tie-line currents ($\boldsymbol{I}_l \in \mathbb{C}^{N_l}$) to the bus current injections ($\boldsymbol{I}_b \in \mathbb{C}^{N_b}$):

$$
\boldsymbol{I}_l = \begin{bmatrix} I_{l_1} \\ I_{l_2} \\ \vdots \\ I_{l_{N_l}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_{l_1}^T \\ \boldsymbol{\alpha}_{l_2}^T \\ \vdots \\ \boldsymbol{\alpha}_{l_{N_l}}^T \end{bmatrix} \begin{bmatrix} I_{b_1} \\ I_{b_2} \\ \vdots \\ I_{b_{N_b}} \end{bmatrix} = \boldsymbol{A}\boldsymbol{I}_b, \qquad (1)
$$

$$
\boldsymbol{\alpha}_{l_i}^T = ((y_{l_i} + y_{f_i}^s)\boldsymbol{e}_{f_i}^T - y_{l_i}\boldsymbol{e}_{t_i}^T)\boldsymbol{Y}^{-1}, \qquad (2)
$$

where $\boldsymbol{\alpha}_{l_i}^T$ is the current injection sensitivity factor for line i between buses f_i and t_i; y_{l_i} is the line admittance; $y_{f_i}^s$ is the shunt admittance at bus f_i; and $\boldsymbol{Y}$ is the system admittance matrix. The vector $\boldsymbol{e}_{f_i} \in \mathbb{R}^{N_b}$ is a vector with a one for the f_i-th element and a zero for all other elements. Given the linear relation, when the load and the internal generator outputs are assumed to be

constant, the changes in line currents after a generator outage can be approximated by the injection changes at external generators' buses and the corresponding CISFs $\boldsymbol{A}_l$:

$$\Delta \boldsymbol{I}_l = \boldsymbol{A}\Delta \boldsymbol{I}_b \approx \boldsymbol{A}_l \Delta \boldsymbol{I} \tag{3}$$

where $\Delta \boldsymbol{I} \in \mathbb{C}^N$, and N is the number of generators in the external system.

Under the dc power flow assumptions, the traditional injection shift factor (ISF) can be seen as a special case derived by ignoring resistance and assuming a flat voltage profile. Analogous to (3), a linear relationship between tie-line active power flow and active power injection holds:

$$\Delta \boldsymbol{P}_l \approx \boldsymbol{\Psi}_l \Delta \boldsymbol{P}, \tag{4}$$

where $\boldsymbol{\Psi}_l$ is the ISF matrix [6]; $\boldsymbol{P}_l$ and $\boldsymbol{P}$ are the tie-line power flow and generator active power output vector, respectively. Note that the network topology is assumed to remain fixed immediately following the generator outage; therefore, $\boldsymbol{A}_l$ and $\boldsymbol{\Psi}_l$ remain constant.

III. CLUSTERING OF GENERATORS BASED ON LSFs

In order to identify the outage location, there are two approaches given the linear sensitivity factors (3) or (4) of the system:

1) Both $\Delta \boldsymbol{I}$ and $\boldsymbol{A}_l$ are assumed to be known for each potential event. $\Delta \boldsymbol{I}_l$ can be precomputed and compared to measurements. The case that best fits the measurements is identified. A similar methodology, applied to line outage detection, can be found in [20].

2) Only $\boldsymbol{A}_l$ (or $\boldsymbol{\Psi}_l$) is known. $\Delta \boldsymbol{I}$ can be estimated based on $\Delta \boldsymbol{I}_l$ measurements. The outage is identified by locating the most significant loss of injection.

Method (1) requires precise knowledge of $\Delta \boldsymbol{I}$, the change in injection currents, for each potential external generator outage. Such information is usually not available from external system operators. One solution is to approximate the redistribution of injection using participation factors computed with inertia or governor droop constants of generators. Even if such information is made available, the underlying assumptions may undermine the accuracy of identification. For example, inertia-based participation factors are derived assuming the rotor speeds of all the generators decrease at the same rate. However, studies have shown that disturbances do not affect all generators instantaneously on the system [1]. The method will be further challenged given the presence of oscillations after an outage, which when appearing in measurements will increase the discrepancy between the expected and the observed behavior (see case studies in [6]).

By contrast, the second method only requires the linear sensitivity factor matrix ($\boldsymbol{A}_l$ or $\boldsymbol{\Psi}_l$) and the changes in tie-line measurements ($\Delta \boldsymbol{I}_l$ or $\Delta \boldsymbol{P}_l$). The main challenge of this method is how to estimate the injection change based on a limited number of measurements; in other words, how to solve a underdetermined linear system. As mentioned in Section I, dimension reduction methods can be used to group generators into clusters based on the similarity of LSFs so that $\boldsymbol{A}_l$ can be approximated by $\tilde{\boldsymbol{A}}$ (a square matrix of smaller dimension

---

Algorithm 1: External Outage Identification Based on CISFs.

Input: the CISF matrix $\boldsymbol{A}_l$, the tie-line current change $\Delta \boldsymbol{I}_l$
Output: outage cluster location $\mathcal{G}^*$

    Form generator clusters:

1:   Column pivoted QR factorization on $\boldsymbol{A}_l$:

$$\boldsymbol{A}_l \boldsymbol{\Pi} = \boldsymbol{QR} \tag{5}$$

$$= \boldsymbol{Q}[\boldsymbol{R}_1 | \boldsymbol{R}_2] \tag{6}$$

$$\boldsymbol{M} = [\boldsymbol{\mathcal{I}}_{N_l} | \boldsymbol{R}_1^{-1} \boldsymbol{R}_2] \boldsymbol{\Pi}^T, \tag{7}$$

    where $\boldsymbol{\mathcal{I}}_{N_l}$ is the identity matrix of size N_l.

2:   Each generator j is assigned to cluster $\mathcal{G}_j$ where $\mathcal{G}_j = \arg\max_i |m_{i,j}|$, using the notation that $m_{i,j}$ represents the element of $\boldsymbol{M}$ at row i, column j.

3:   $\tilde{\boldsymbol{A}} = \boldsymbol{A}_l \boldsymbol{\Pi}_l$ where $\boldsymbol{\Pi}_l$ is the first N_l columns of $\boldsymbol{\Pi}$.

    Identify outage:

4:   Injection change estimation

$$\Delta \tilde{\boldsymbol{I}} = \tilde{\boldsymbol{A}}^{-1} \Delta \boldsymbol{I}_l \tag{8}$$

5:   Outage cluster location

$$\mathcal{G}^* = \arg\max_i |\Delta \tilde{I}_i| \tag{9}$$

N_l). The net injection of a generator group (rather than each generator) can then be estimated by inverting the reduced-order system.

A. Algorithm for Generator Clustering and Outage Identification

To overcome the limitation on measurements (assuming $N_l < N$, as is usually the case), a clustering-based identification algorithm is proposed in Algorithm 1. A demonstration using a simple 5-bus system is also provided in Appendix A.

The algorithm consists of two main parts:

1) *Form generator clusters:* Inspired by [13] and [21], a clustering algorithm based on the QR decomposition of the CISF matrix is used. (More details about the QR decomposition with column pivoting can be found in [22]). The CISF matrix $\boldsymbol{A}_l$ is assumed to be known (e.g., provided by the external system operator). The first step is to categorize the N generators into N_l clusters. QR factorization with column pivoting on $\boldsymbol{A}_l$ (5) is conducted to determined the representative generators of each cluster (by the first N_l columns of the permutation matrix $\boldsymbol{\Pi}$). Matrix $\boldsymbol{R}$ consists of two parts: the square matrix $\boldsymbol{R}_1$ and the remainder $\boldsymbol{R}_2$ (6). The membership matrix $\boldsymbol{M}$ is formed based on $\boldsymbol{R}$ as defined in (7) [13]. Note that the columns of $\boldsymbol{\Pi}$ that correspond to $\boldsymbol{R}_1$ as well as the identity matrix in $\boldsymbol{M}$, i.e., the first N_l columns, are the representatives. More specifically, the representative generator of cluster j, denoted as g_j^r, is determined by finding the row number of the (only) non-zero element in column j of $\boldsymbol{\Pi}$ ($j = 1, \ldots, N_l$).

$$g_j^r = \arg\max_i \Pi_{i,j} \tag{10}$$

where $\Pi_{i,j}$ is the element of row i and column j of $\mathbf{\Pi}$. The cluster membership of generator j is determined by the row with the largest absolute value, i.e., $\mathcal{G}_j = \arg\max_i |m_{i,j}|$. If the CISFs of generator j are identical to that of the representative generator in the cluster (i.e., $g_{\mathcal{G}_j}^r$), all elements of the j-th column of $\mathbf{M}$ will be zero except the $\mathcal{G}_j$-th element is 1. Strictly speaking, when $\mathbf{M}$ is formed, the representative generators are determined implicitly, thus (10) is not necessary, and it will not directly alter identification results in (9). Therefore it is not included in the main algorithm but here to allow a general analysis of membership and its impact on results in subsequent sections. Note that here generator j refers to the generator corresponding to the j-th column of the CISF matrix. If needed, a one-to-one mapping from the original generator names (or numbers) to the new order $1, 2, \ldots, N$ can be done so that the algorithm can be directly applied (see Appendix for an example). However, to eliminate ambiguity introduced by an arbitrary mapping, we use the original generator IDs in the case study section.

2) *Identify Outage Location:* The original CISF matrix is replaced by $\tilde{\mathbf{A}}$, which is formed from the representative columns of $\mathbf{A}_l$. The net injection change of each cluster can be approximated by solving the reduced-order system (8). The most severe injection change in terms of magnitude is identified as the outage source (9). If the actual outage generator $g_o \in \mathcal{G}^*$, then it is considered correctly identified.

The assumption that the most severe injection change indicates the location of outage is valid since we focus on the following scenario. When a single generator in the external system goes off-line thus resulting in a large (negative) power injection change, the loss is shared by the remaining generators by increasing their outputs (depending on the electrical distances with respect to the outaged generator). If loads and power losses are assumed to be constant, the total generation changes should be zero, with only one negative injection change from the outaged generator, which also is the largest change in terms of magnitude. Regardless of the clustering strategy, as long as more than two clusters exist, the net injection of the group containing the outage will be the largest in magnitude. As a result, by examining the estimated injection of each group and locating the largest generation change, the outage can be localized. Note that when the CISFs are used, the injection current instead of the power is estimated. The injection power can be computed given external generator terminal voltage phasors, which are not generally available. By assuming the terminal voltages of generators being 1 p.u. in magnitude, the apparent power injection equals the magnitude of injection current. Therefore instead of finding out the most severe loss of active power, the outage location is identified as the group with the most significant change in apparent power. More detailed discussion is provided in the case study section.

If the ISFs are used, then replace $\mathbf{A}_l$ with $\mathbf{\Psi}_l$ in the algorithm, and $\Delta \mathbf{I}_l$ with $\Delta \mathbf{P}_l$. The estimated injection power change $\Delta\tilde{\mathbf{P}}$ will be used instead of $\Delta\tilde{\mathbf{I}}$ at the final step (9).

B. Estimation of Cluster Injection Change

In addition to the location of the outage, the proposed method can also approximate the net injection change (8). Note that since generators are categorized into groups and represented by one representative generator (in terms of CISFs), the estimated injection change is the approximate net change of the group. As mentioned in the previous section, when CISFs are used, the current rather than power injections are computed. To estimate the change in injected power, based on the change in injected current of the identified generator group, the following method is considered.

Treat the real part of the CISF matrix (referred to as $\tilde{\mathbf{A}}_{Re}$) as the ISF matrix and compute the power injections using tie-line power measurements. In other words, the power-flow sensitivity is approximated by the current sensitivity assuming a flat voltage profile (see [6]);

$$\Delta\hat{\mathbf{P}}_r = \tilde{\mathbf{A}}_{Re}^{-1}\Delta\mathbf{P}_l = \begin{bmatrix} \Delta\hat{P}_{r,1} \\ \Delta\hat{P}_{r,2} \\ \vdots \\ \Delta\hat{P}_{r,N_l} \end{bmatrix}. \tag{11}$$

$\Delta\hat{P}_{r,i}$ represents the estimated net injection change in group i, $i = 1 \ldots N_l$. As a result, given the estimated net injection change vector computed by (11), the change of group $\mathcal{G}^*$ (in which the outage is located) is $\Delta\hat{P}_{r,\mathcal{G}^*}$.

By comparing ΔP_o to the estimated net injection $\Delta\hat{P}_{r,\mathcal{G}^*}$, the error is defined as:

$$\varepsilon_r\% = \left| \frac{\Delta P_o - \Delta\hat{P}_{r,\mathcal{G}^*}}{\Delta P_o} \right| \times 100\%, \tag{12}$$

where ΔP_o is the actual net injection change of group $\mathcal{G}^*$

$$\Delta P_o = \sum_{i \in \{i | \mathcal{G}_i = \mathcal{G}^*\}} \Delta P_i. \tag{13}$$

Event detection is achieved by constantly monitoring the tie-line current or power flow with synchrophasor measurements. If the change between two samples of any tie-line exceeds a pre-determined threshold (e.g. 10%), then the identification algorithm is initiated. The choice of the threshold that leads to no misidentification (with minimal false positives or negatives) may be dependent on the system operating conditions and topology. For example, a generator with small output power and/or located far from the tie-lines will lead to small changes on the flows to the extent that the change may not be distinguishable from fluctuations under normal operations. However, such cases are generally not of great concern due to their small impact. For the 68-bus New England system, since most generators are equivalents with large outputs, the threshold can be set as high as 40% which will lead to no misidentification even when transient measurement errors and load fluctuations are considered. For the 500-bus system, the threshold is 5%, which is determined based on the outage of a particularly small generator at bus 351 with an output of 6.2 MW (0.08% of the total load). The threshold can be further raised to around 10%, if this outage case is discarded. Note that frequency decrease can also be

used as an event indicator [1], [2], which can be used in a complementary fashion in addition to the flow change. When an event is detected, the tie-line flow changes ΔI_l (computed using the first measurement after the event and the last measurement before the event) are calculated and then used as inputs for the identification algorithm.

IV. CASE STUDIES

The identification method has been tested on two systems: the 68-bus New England test system [17] and a synthetic 500-bus system representing part of the South Carolina network (ACTIVSg500)[19]. Dynamic models are available for both multi-regional systems. Generator outage cases were simulated using PSS/E with a solution step of 1/480 s (i.e., 8 samples per cycle with a nominal system frequency of 60 Hz). Similar to the down-sampling stage in synchrophasor computation in SEL relays [23], the simulated phasors are directly down-sampled to the reporting resolution (i.e. 1/120 s, 2 samples per cycle) and treated as synchrophasor measurements without further filtering. The tie-line flow changes (ΔI and ΔP) are computed as the difference between two samples with an inter-sample interval of 1/120 s. The identification algorithm was implemented in MATLAB on a desktop computer with an Intel Xeon E5-1607 processor and 8 GB RAM. Due to the proprietary acceleration technique used by MATLAB compiler, the first time a piece of code is complied, it takes much longer. The time will vary on a different platform possibly due to usage of different compilers. The reported time below is calculated as the average of 500 runs (discarding the first run) with the Java virtual machine disabled. The QR decomposition can be computed efficiently; e.g., for the 500-bus system, the cluster forming (equation (5) to (7) in Algorithm 1) takes 0.16 ms (first run 5.5 ms). While the LU decomposition and finding the maximum injection change (through sorting) in the identification stage (equation (8) and (9)) takes 0.03 ms (first run 1.43 ms). If $\widetilde{A}^{-1}$ is pre-computed given a fixed topology, the identification only takes 0.012 ms (first run 1.40 ms) in real-time. Although the computation time may vary depending on devices and platforms, the algorithm is efficient enough for online application.

A. 68-Bus New England System

This test system represents an equivalent of the Northeast Power Coordinating Council system with the New England Test System (NETS) and the New York Power System (NYPS) modeled in more detail. Each of the remaining regions is represented by a single equivalent generator. The system consists of 68 buses, 66 lines, 16 generators and 35 loads. The power exchange between NYPS and NETS is about 700 MW. There are four tie-lines between NETS and NYPS including two parallel lines. Since the two parallel lines are identical, only one is considered. Tie-line flow measurements have been generated with dynamic simulation of generator outages in the external system (NETS). The power outputs of generators range from 500 MW to 800 MW. Since there are only three ties effectively, three clusters are determined using either the traditional ISF or CISF matrix. Fig. 2 shows the external system (NETS) and the tie-lines. Three
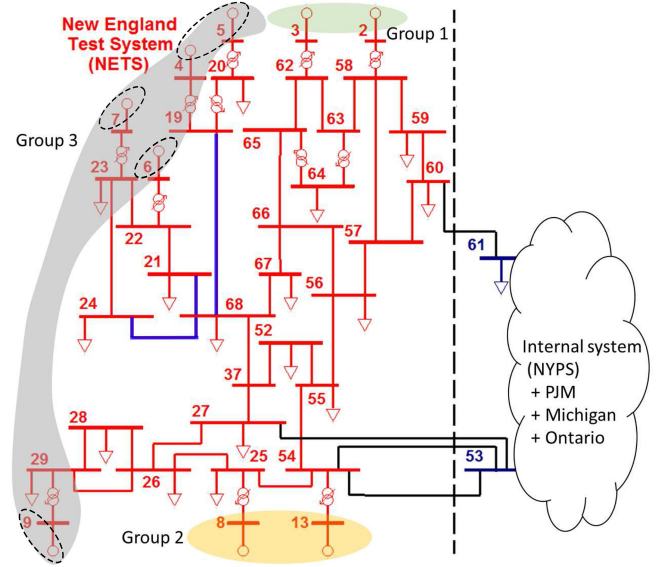

Fig. 2. The New England system [17] with 3 generator groups determined by Algorithm 1. Blue lines represent extra monitored lines, while dashed circles represent resulting additional subgroups as discussed in Section IV-A.

TABLE I
NEW ENGLAND SYSTEM GENERATION CHANGE ESTIMATES

g_o	G	$|\Delta M_g|$	ΔP_o	$\Delta \hat{P}_{r,\mathcal{G}*}$	$\varepsilon_r\%$	$|\Delta M_g^1|$	$\varepsilon_r^1\%$
2	1	0.0856	-456.83	-556.71	21.86	0.0808	14.38
3			-547.62	-533.60	2.56		6.94
8	2	0.1853	-399.48	-257.93	35.43	0.1825	30.28
13			-216.00	-227.57	5.36		12.02
4	3	0.4531	-326.95	-282.41	13.62	2.15e-4	1.93
5			-279.92	-196.81	29.69		7.14
6			-377.72	-323.56	14.34	0	3.86
7			-257.46	-237.23	7.86	0	5.10
9			-620.19	-1039.26	67.57	0	8.93
mean	-	-	-	-	22.0	-	10.06

ΔP_o is actual total generation change in the group. $\Delta \hat{P}_{r,\mathcal{G}*}$ is the estimate based on (11)–(13). $|\Delta M_g|$ is defined in (14). The last two columns (with superscript 1) represent results generated based on measurements from the three tie-lines and three extra lines.

clusters are formed with representative generators being g_2, g_{13} and g_9 respectively (10).

Experiments were conducted by applying Algorithm 1 using current or active power methods (i.e., with A_l or Ψ_l). When the CISF matrix (A_l) is used, the algorithm is capable of identifying the correct generator group with 100% accuracy. By contrast, the identification accuracy based on ISFs (Ψ_l) is very low—only two cases ($g_o = g_2, g_3$) are correctly identified out of 9. The misidentification is mainly due to the error introduced by the dc power flow model. Similar to the errors resulting in normal dc power flow computations, the error may be case dependent, which suggests the accuracy of identification based on such a model may vary given different topologies or systems. Based on this performance, only the CISF method is recommended and the ISF-based approach was not tested on the second study system.

Table I shows the estimation of generation change under different outage scenarios. First, the results generated based on measurements taken from the three tie-lines are presented,
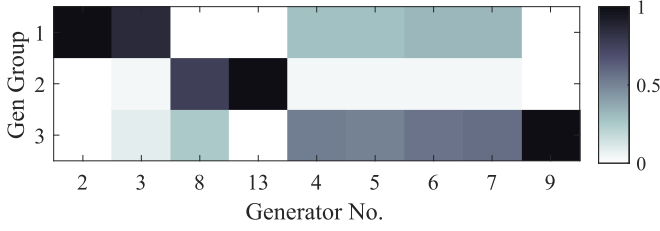
Fig. 3.   Visualization of column permuted $|\boldsymbol{M}|$ (7) of the New England system.

with the column headings $\Delta\hat{P}_{r,\mathcal{G}^*}$ and $\varepsilon_r\%$ corresponding to the estimated injection changes and their relative errors with respect to the actual changes ΔP_o (defined in (11)–(13)). We notice that group 3 (g_4–g_7 and g_9) is particularly inaccurate with the largest $\varepsilon_r\%$, 67.57%. The estimation error stems from the discrepancy between the representative CISFs and the actual CISFs of the generators. A closer investigation of the corresponding columns of $\boldsymbol{M}$ shows that $\max_i |m_{i,j}| = |m_{3,j}|$ is around 0.5 for generator 4 to 7 (compared to 1 of the representative generator g_9). The result is similar with group 2 (g_8 and g_{13}). To better represent the discrepancy of the CISFs within a particular group i, a measure $|\Delta M_g(i)|$ is defined as follows:

$$|\Delta M_g(i)| = \frac{1}{N_i} \sum_{j \in \{j | \mathcal{G}_j = i\}} \|\boldsymbol{m}_j - \boldsymbol{e}_i\|, \qquad (14)$$

where N_i is the number of generators inside group i, $\boldsymbol{m}_j$ denotes the j-th column of $\boldsymbol{M}$, and $\boldsymbol{e}_i$ is the column of $\boldsymbol{M}$ corresponding to the representative generator of group i (i.e., a vector with all zeros except the i-th element, which is set to 1). $|\Delta M_g(i)|$ reflects the average Euclidean distances of M within a group with respect to the representative generator. The values of $|\Delta M_g(i)|$ of each group is listed in the third column of Table I. A smaller value suggests the generators are more similar in CISFs within a group (e.g., a value of zero indicates that all the generators within the group have identical CISFs). The heterogeneous nature of group 3 (with the biggest $|\Delta M_g| = 0.4531$) suggests a reduced estimation accuracy, which is borne out by the relatively high values of $\varepsilon_r\%$ in the table.

Another way to represent the discrepancy in CISFs is through visualization of the absolute values of elements in $\boldsymbol{M}$ (see Fig. 3). Note that column permutation is performed so that generators of the same group are presented together. The permutation order follows the first column of Table I. As mentioned previously, for a representative generator of group i, the corresponding column is $\boldsymbol{e}_i$. This means only the i-th row is black while all remaining elements are white in the colormap. For each generator, the row number of the darkest element indicates its group membership. If generator j belongs to group $\mathcal{G}_j$, the lighter $m_{i,j}$, $i \neq \mathcal{G}_j$, the better, since it means there is less similarity to another representative generator. The figure shows that, for group 3, compared to the representative (g_9) CISFs, the remaining generators are also similar to the first generator group (which is represented by generator g_2), since the first row of columns g_4 to g_7 are of medium grey color. Thus, using the CISFs of g_9 to approximate those of g_4 to g_7 is expected to cause error. Similarly, g_{13} is a poor representation for group 2

due to the similarity between g_8 and group 3. In summary, the color patterns provide a visual interpretation of similarities within each group, which promotes analysis of results.

As established, low estimation accuracy of injection change is associated with poor representation of the original CISFs as a result of limited number of tie-lines. By increasing the number of PMUs (i.e., N_l), the results may be improved. If we augment $\boldsymbol{A}_l$ by adding more measurements (corresponding to more rows) in addition to the existing tie-lines, more groups (than the current three) may be generated. Unfortunately, the two systems are weakly linked. Assuming all lines in the internal system are monitored, the rank of the new $\boldsymbol{A}_l$ is still three. This indicates that having all measurements in the internal system would not have improved observability beyond tie-lines in terms of generator outages in the external system. Since widespread sharing of PMU data is expected among neighboring utilities[24], measurements from PMUs located in the external system may be used instead. As mentioned previously, the localization precision in group 3 (consisting of 5 generators) is poor. If three extra lines (highlighted in blue in Fig. 2) sharing the same terminal (bus 68) are monitored by PMUs, the results can be improved significantly. Bus 68 and the connected lines are critical since they bridge different generator groups. The column $|\Delta M_g^1|$ in Table I shows that group 3 is divided into four groups now: (1) g_4 and g_5, (2) g_6, (3) g_7, and (4) g_9. The proposed algorithm is able to identify the outage cluster successfully with a reduced mean error $\varepsilon_r^1\% = 10.06\%$ (see the last column in Table I). In particular, the errors in the original group 3 are decreased greatly especially for g_9, which is consistent with much smaller $|\Delta M_g^1|$ (with the largest being 2.15e-4). Note that further improvements may be achieved using a different placement (with more PMUs and/or PMUs closer to the external generators). Here we merely provide an example demonstrating augmented PMU measurements from the external system can improve the identification accuracy. In summary, the estimation accuracy depends on a good approximation of the original CISF matrix, which loosely translates to sufficient, representative measurements.

In addition to the error introduced by the approximation $\tilde{\boldsymbol{A}}$, it is difficult to analyze the error analytically due to the nonlinear relations between the injection power and the line flows (11). Although the voltage magnitude across the system is between 0.98 pu and 1.0815 pu (generator voltage 0.98 pu to 1.063 pu), the angles vary significantly (about 50 degrees within the interconnected systems). This violates the flat voltages assumption (to derive $\boldsymbol{\Psi} \approx \boldsymbol{A}_{Re}$), which contributes to the estimation error.

B. 500-Bus System

Although the New England test system is commonly used for stability studies, it is an equivalent representation with regional power plants modeled by a single generator. The highly aggregated representation may not reflect the system dynamics accurately. To better validate the proposed method, we also ran experiments on a larger test system with more realistic generator models. The 500-bus system is a synthetic case based on the footprint of the South Carolina system [19] and consists of two zones (Zone 1, representing the upstate portion of the network,
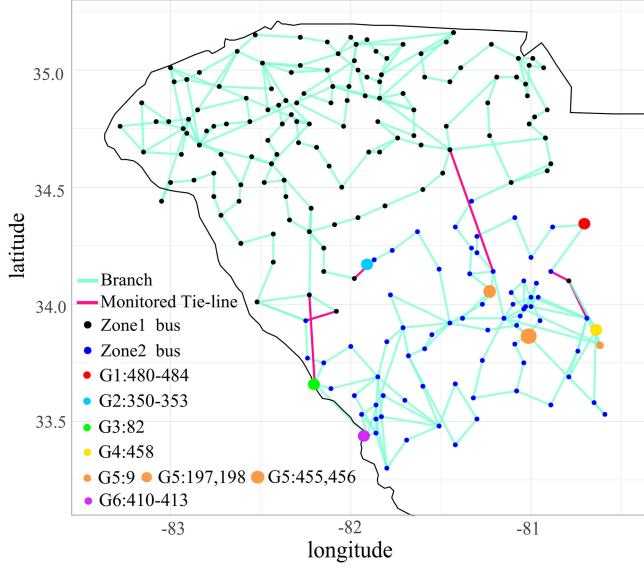
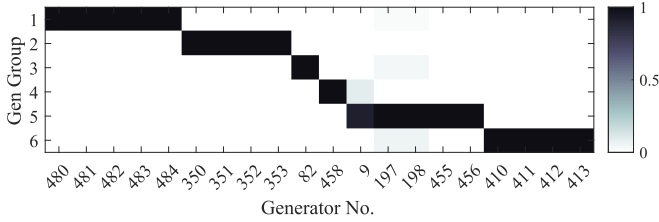Fig. 4. The South Carolina 500-bus system (geographical information available at [25] generated using R [26]).

Fig. 5. Visualization of column permuted $|\boldsymbol{M}|$ (7) of the 500-bus system.

and Zone 2, representing the rest of the system). The base case has 56 active machines and 7750 MW load, and there are 9 tie-lines connecting the two zones. Some of the tie-lines are very close together (nearly parallel), resulting in CISF values that are almost identical. After eliminating parallel, or nearly parallel, tie-lines, six were considered for the purposes of external generator outage identification (the magenta branches in Fig. 4). 20 generators with non-zero active power outputs (from 6.2 MW to 771.8 MW) are located in the external system (Zone 2). In our study, all generator outages in Zone 2 were considered.

Fig. 5 provides a visualization of the absolute value of $\boldsymbol{M}$ (column permuted following the order of g_o in Table II), after the 20 external generators are separated into 6 groups. Each black block indicates one generator group (with the row number as the group number). Note that the elements $m_{i,j}$ (where $i \neq \mathcal{G}_j$ for all j) are all zero except in group 5 ($i = 5$), which means all the generators within groups 1–4 and 6 are identical from the perspective of CISF similarity (characterized by $|\Delta M_g(i)| = 0$ for $i = 1, \dots, 4, 6$). The columns corresponding to generators 9, 197 and 198, which are in group 5, have non-zero elements (around 0.1) and result in $|\Delta M_g(5) = 0.0425|$. The three different color patterns are consistent with the fact that there are three natural clusters (power plants) in group 5 (see Fig. 4). As a result of increased number of measurements and higher similarity in

TABLE II
500-BUS SYSTEM GENERATION CHANGE ESTIMATES

Group	g_o	$|\Delta M_g(i)|$	ΔP_o	$\Delta \hat{P}_{r,\mathcal{G}*}$	ε_r %
1	480	0	-30.3	-30.0	1.0
	481		-22.1	-21.8	1.6
	482		-25.2	-24.1	4.4
	483		-11.5	-11.3	1.4
	484		-15.0	-14.5	3.4
2	350	0	-11.1	-11.3	1.2
	351		-4.4	-4.5	1.3
	352		-10.7	-10.8	0.8
	353		-61.3	-61.5	0.3
3	82	0	-360.7	-314.1	12.9
4	458	0	-108.5	-102.6	5.5
5	9	0.0425	-446.8	-428.3	4.1
	197		-153.4	-191.5	24.8
	198		-110.5	-129.9	17.5
	455		-35.7	-30.6	14.4
	456		-67.4	-69.3	2.9
6	410	0	-53.4	-54.7	2.3
	411		-53.1	-50.5	4.9
	412		-53.2	-51.8	2.7
	413		-39.3	-39.2	0.1
mean	-	-	-	-	5.358

ΔP_o is actual total generation change in the group. $\Delta \hat{P}_{r,\mathcal{G}*}$ is the estimate based on (11)–(13).

CISFs between generators within a group, $|\Delta M_g|$ within each group in this system is much smaller indicating each group is more homogeneous. In contrast, the smallest $|\Delta M_g|$ in the New England system is 0.0856 given only three tie-lines.

Similar to the results in Table I, Table II shows the approximated cluster generation changes and relative errors given different outage cases using Algorithm 1 and $\boldsymbol{A}_l$. The outage is correctly identified to the group in all cases. The net injection change results are better on average compared to the New England system (e.g. ε_r% is 5.36% versus 22% previously), since the differences in CISFs within a group are less compared to the New England system. If the locations and the number of tie-lines allow revelation of natural clusters of the system (ideally to individual plants, e.g. in the 500-bus system), the estimation results are expected to be more accurate. In this case, the proposed algorithm is even able to identify the outage location and estimate the loss of small generation with high accuracy, e.g., generator outage at bus 351. Other factors may contribute to the difference in estimation error between systems, including: the differences between power angles of the generators, the X/R ratios of the lines, the severity of voltage drops after disturbances, and the magnitude of (neglected) load fluctuations and internal system generator output changes.

V. CONCLUSION

An identification algorithm for generator outage localization in external systems is proposed, which utilizes measured tie-line flow measurements and known LSFs. A clustering technique based on QR decomposition is implemented to divide the external system generators into different groups to overcome the limitation on the number of tie-line measurements as well as the potential problem of indistinguishable generators. By estimating the group generation change and identifying the group injection loss, the outage can be localized to the originating cluster. The

results demonstrate that the proposed method based on current injection sensitivity factors (CISFs) can accurately identify the generator group experiencing an outage. The net injection power change within a group of generators can also be estimated. When given sufficient number of measurements, which helps reveal the natural clusters in the system (e.g., in the 500-bus system), the average error is around 5.36%. Given shared synchrophasor data from neighboring regions, the injection change accuracy is expected to improve (e.g., by about 12% in the New England system). These results provide incentives for sharing CISFs (not just ISFs) and synchronized measurements within an interconnection. Given the high reporting speed of synchrophasor measurements and the high computational efficiency of matrix decomposition, the proposed algorithm can be used for online applications. It can also be used as a complementary approach to frequency-based identification methods previously proposed (e.g., by using abnormal frequency changes to initiate the identification algorithm, or restricting the set of potential generators considered in geographic localization techniques). In summary, by correctly identifying the outage location and estimating the loss of generation, the proposed identification method will help to raise situational awareness in interconnected systems. Future work may include: validation of the proposed method given out-dated system information (e.g., errors in the LSF matrix) and given noisy measurements.

APPENDIX A
DEMONSTRATION OF ALGORITHM 1

We use a simple 5-bus system (Fig. 6) to demonstrate the construction of $\tilde{A}$ and M. The branch resistance, reactance and shunt susceptance in per unit are also given. The shunt conductance is assumed to be zero for all five lines.

Internal | External System

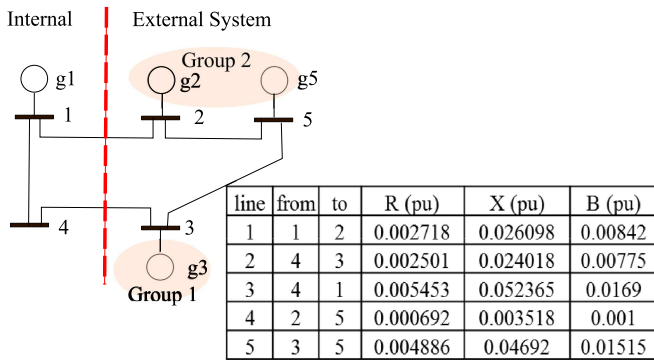| line	from	to	R (pu)	X (pu)	B (pu)
1	1	2	0.002718	0.026098	0.00842
2	4	3	0.002501	0.024018	0.00775
3	4	1	0.005453	0.052365	0.0169
4	2	5	0.000692	0.003518	0.001
5	3	5	0.004886	0.04692	0.01515

Fig. 6.   5-bus system (branch data included).

Table III shows the CISFs of the external system. The rows correspond to the boundary tie-lines and the columns to the three external generators (using the notation from earlier in the paper, $N_l = 2$ and $N = 3$). The order of the columns suggests the mapping $g_2 \mapsto 1$, $g_3 \mapsto 2$, $g_5 \mapsto 3$, which is used to determine the representative generators. Given the permutation matrix Π
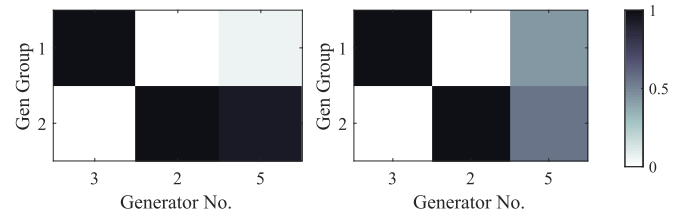

Fig. 7.   Visualization of column permuted $|M|$ of the 5-bus system (left: (16), right: $|M'|$ with parameters R, X and B of l_4 $10\times$ larger than the original).

in (5), only two clusters ($N_l = 2$) will be formed.

$$\Pi = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{15}$$

The representative generator of the first cluster is $g_1^r = 2$, since only $\Pi_{2,1} = 1$ in the first column, suggesting the representative generator of cluster 1 is the generator corresponding to the second column (i.e., g_3). Analogously, we have $g_2^r = 1$. The representative generator of cluster 2 is generator g_2.

TABLE III
CISF MATRIX A_l OF THE EXTERNAL SYSTEM

CISF	Gen 2	Gen 3	Gen 5
l_1: 1->2	-0.0395 - 0.0002i	-0.4061 + 0.0014i	-0.0653 + 0.0021i
l_2: 4->3	-0.1919 - 0.0003i	-0.0393 - 0.0006i	-0.1812 - 0.0013i

The first part of the algorithm is to form generator clusters based on the CISF matrix. The matrix M is computed based on (7):

$$M = \begin{bmatrix} 0 & 1 & 0.0704 - 0.0059i \\ 1 & 0 & 0.9296 + 0.0059i \end{bmatrix}. \tag{16}$$

Due to the permutation matrix, the columns of M corresponds to the original order defined in A_l (g_2, g_3 and g_5). By examining each column, the cluster membership of each generator is determined. For example, the first column correspond to generator 2, and the maximum element is on the second row, i.e. $\mathcal{G}_2 = \arg\max_i |m_{i,1}| = 2$. This means that g_2 belongs to generator group 2. In addition, the column is unitary with the second element being 1 indicating this generator is a representative, which is corroborated by $g_2^r = 1$. The non-representative generator (g_5) is considered to belong to group 2 (the same group as g_2) due to $\max_i |m_{i,3}| = |m_{2,3}| = 0.9297$.

Fig. 7 shows a visualization of column permuted $|M|$ (the left subplot). M is essentially column permuted $\tilde{A}^{-1}A_l$, the elements of which are in a way normalized to allow easier comparison of different CISFs. Due to the small impedance of line 4 connecting generator g_2 and g_5, the two generators are very similar in terms of CISFs with $|\Delta M_g(2)| = 0.0499$.

If the parameters (R, X and B) of line 4 are increased by a factor of 10, the visualization of the new M' is shown in the

right subplot of Fig. 7.

$$M' = \begin{bmatrix} 0 & 1 & 0.4318 - 0.0222i \\ 1 & 0 & 0.5684 + 0.0222i \end{bmatrix} \quad (17)$$

g_5 still belongs to the same group as g_2 due to $\max_i |m_{i,3}| = |m_{2,3}| = 0.5689$. However, the colormap clearly suggests the CISFs of g_2 and g_5 are very different. In particular, when comparing the third columns in both cases, we can see that $m_{2,3}$ (the darkest element suggesting membership) in the right subplot becomes lighter (from 0.9297 to 0.5689), while $m_{1,3}$ becomes darker (from 0.0706 to 0.4324). This suggests g_5 is no longer close to g_2 in terms of electrical distance. In fact, it becomes much closer to g_3 now. This results in a new and much larger $|\Delta M_g(2)| = 0.3056$. Consequently, $\tilde{A}$ will be a poorer representation of A_l.

REFERENCES

[1] G. Zheng, Y. Liu, and G. Radman, "Wide area frequency based generation trip event location estimation," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, San Diego, CA, USA, Jul. 2012, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6344827

[2] H. Zhang, F. Shi, Y. Liu, and V. Terzija, "Adaptive online disturbance location considering anisotropy of frequency propagation speeds," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 931–941, Mar. 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7105965/

[3] Y. Zhang *et al.*, "Wide-area frequency monitoring network (FNET) architecture and applications," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 159–167, Sep. 2010. [Online]. Available: http://ieeexplore.ieee.org/document/5504177/

[4] N. Muller *et al.*, "NERC IDC: Managing congestion in the North American Eastern Interconnection," in *Proc. IEEE/PES Power Syst. Conf. Expo.*, Seattle, WA, USA, Mar. 2009, pp. 1–8. [Online]. Available: http://ieeexplore.ieee.org/document/4840168/

[5] X. Cheng and T. Overbye, "PTDF-based power system equivalents," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 1868–1876, Nov. 2005. [Online]. Available: http://ieeexplore.ieee.org/document/1525117/

[6] Y. C. Chen, S. V. Dhople, A. D. Dominguez-Garcia, and P. W. Sauer, "Generalized injection shift factors," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2071–2080, Sep. 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7420704/

[7] A. J. Wood, B. F. Wollenberg, and G. B. Shebl, *Power Generation, Operation, and Control*. Hoboken, NJ, USA: Wiley, 2014.

[8] R. Podmore, "Identification of coherent generators for dynamic equivalents," *IEEE Trans. Power App. Syst.*, vol. PAS-97, no. 4, pp. 1344–1354, Jul. 1978. [Online]. Available: http://ieeexplore.ieee.org/document/4181567/

[9] J. Chow, R. Galarza, P. Accari, and W. Price, "Inertial and slow coherency aggregation algorithms for power system dynamic model reduction," *IEEE Trans. Power Syst.*, vol. 10, no. 2, pp. 680–685, May 1995. [Online]. Available: http://ieeexplore.ieee.org/document/387903/

[10] I. Guyon, M. Nikravesh, S. Gunn, L. A. Zadeh, and J. Kacprzyk, Eds., *Feature Extraction* (Studies in Fuzziness and Soft Computing Series), vol. 207. Berlin, Germany: Springer, 2006. [Online]. Available: http://link.springer.com/10.1007/978-3-540-35488-8

[11] G. Quintana-Ort, X. Sun, and C. H. Bischof, "A BLAS-3 version of the QR factorization with column pivoting," *SIAM J. Sci. Comput.*, vol. 19, no. 5, pp. 1486–1494, Sep. 1998. [Online]. Available: http://epubs.siam.org/doi/10.1137/S1064827595296732

[12] M. Setnes and R. Babuska, "Rule base reduction: Some comments on the use of orthogonal transforms," *IEEE Trans. Syst., Man, Cybern., Part C (Appl. Rev.)*, vol. 31, no. 2, pp. 199–206, May 2001. [Online]. Available: http://ieeexplore.ieee.org/document/941843/

[13] P. Mlinaric, S. Grundel, and P. Benner, "Efficient model order reduction for multi-agent systems using QR decomposition-based clustering." in *Proc. 54th IEEE Conf. Decis. Control*, Dec. 2015, pp. 4794–4799. [Online]. Available: http://ieeexplore.ieee.org/document/7402967/

[14] A. Amiri and M. Fathy, "Hierarchical keyframe-based video summarization using QR-decomposition and modified-means clustering," *EURASIP J. Advances Signal Process.*, vol. 2010, no. 1, Dec. 2010, Art. no. 892124. [Online]. Available: https://asp-eurasipjournals.springeropen.com/articles/10.1155/2010/892124

[15] A. Damle, V. Minden, and L. Ying, "Robust and efficient multi-way spectral clustering," 2016, *arXiv:1609.08251*.

[16] P. Kundur, N. J. Balu, and M. G. Lauby, *Power System Stability and Control*, vol. 7. New York, NY, USA: McGraw-Hill, 1994.

[17] M. Mohammed and A. Mostafa, "Impacts of high-depth of wind power penetration on interconnected power systems dynamics," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Toronto, Toronto, ON, Canada, Jun. 2016. [Online]. Available: http://hdl.handle.net/1807/77862

[18] B. Pal and B. Chaudhuri, *Robust Control in Power Systems*. Berlin, Germany: Springer, 2006.

[19] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3258–3265, Jul. 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7725528/

[20] J. E. Tate and T. J. Overbye, "Line outage detection using phasor angle measurements," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1644–1652, Nov. 2008. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4652583

[21] H. Zha, X. He, C. Ding, H. Simon, and M. Gu, "Spectral relaxation for k-means clustering," in *Proc. 14th Int. Conf. Neural Inf. Process. Syst., Natural Synthetic*. Cambridge, MA, USA, 2001, pp. 1057–1064. [Online]. Available: http://dl.acm.org/citation.cfm?id=2980539.2980675

[22] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2012.

[23] "SEL-451-5 relay protection, automation, and control system instruction manual," Schweitzer Engineering Laboratories, Inc., Pullman, WA, USA, Jan 2015.

[24] "NASPI 2014 survey of synchrophasor system networksResults and findings," NASPI Data and Network Management Task Team, Network Systems Group, Tech. Rep. NASPI-2015-TR-015, Jul. 2015. [Online]. Available: https://www.naspi.org/sites/default/files/reference_documents/8.pdf

[25] South Carolina 500-bus system: ACTIVSg500. [Online]. Available: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg500/. Accessed on: Sep. 24, 2019.

[26] R Core Team, R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org/

Zhen Dai (S'12) received the B.E. degree from Tsinghua University, Beijing, China, in 2011, and the M.A.Sc. and Ph.D. degrees in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 2014 and 2019, respectively.

Joseph Euzebe Tate (S'03–M'08) received the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 2005 and 2008, respectively. He is currently an Associate Professor with the University of Toronto, Toronto, ON, Canada. His research interests include combining advanced telemetry, data processing, and visualization techniques to facilitate renewable integration, and improve power system reliability.