Data-driven forced oscillation localization using inferred impulse responses☆

Shaohui Liu [a,*], Hao Zhu [a], Vassilis Kekatos [b]

[a] *Chandra Family Department of ECE, The University of Texas at Austin, United States of America*
[b] *Elmore Family School of ECE, Purdue University, United States of America*

ARTICLE INFO

ABSTRACT

Poorly damped oscillations pose threats to the stability and reliability of interconnected power systems. In this work, we propose a comprehensive data-driven framework for inferring the sources of forced oscillation (FO) using solely synchrophasor measurements. During normal grid operations, fast-rate ambient data are collected to recover the impulse responses in the small-signal regime, without requiring the system model. When FO events occur, the source is estimated based on the frequency domain analysis by fitting the least-squares (LS) error for the FO data using the impulse responses recovered previously. Although the proposed framework is purely data-driven, the result has been established theoretically via model-based analysis of linearized dynamics under a few realistic assumptions. Numerical validations demonstrate its applicability to realistic power systems including nonlinear, higher-order dynamics with control effects using the IEEE 68-bus system, and the 240-bus system from the IEEE-NASPI FO source location contest. The generalizability of the proposed methodology has been validated using different types of measurements and partial sensor coverage conditions.

1. Introduction

Forced oscillations (FOs) such as the 2019 Florida event [1] pose threats to the security and reliability of interconnected power systems, as the potential resonance with a natural system mode could induce oscillations within the interconnection. Among the critical challenges in mitigating FOs, determining the source location using Phasor Measurement Unit (PMU) measurements in real-time has attracted significant attention in recent years [2].

Thanks to widespread installations of PMUs, synchronized measurements with high sampling rates make it possible to develop data-driven FO localization methods. By calculating the energy function, *dissipating energy flow* (DEF)-based methods have been proposed in [3–5], and adopted by the ISO-NE [6]. However, the success of DEF-based methods relies upon measurements at the FO source generator and presumes accurate system topology information. Recently, a model-enhanced method has been proposed to address the lack of measurements at the FO sources by utilizing a model of linearized grid dynamics [7]. Nevertheless, acquiring accurate system parameters remains a challenge for large interconnections. Other data-driven FO localization methods include a Robust Principal Component Analysis (RPCA)-based method, which extracts the low-rank FO components from PMU measurements [8], and two recent works suggesting system identification

approaches [9,10]. However, the RPCA method lacks theoretical connections between the power system model and low-rank FO components, as pointed out by [9]. While the system identification-based approaches are explainable from a theoretical perspective, the requirement of collecting measurements at most buses may be impractical and incur high computational complexity.

To address the aforementioned issues in the state-of-the-art, this work puts forth a purely data-driven FO localization framework, solidly founded upon linear system analysis of the power system in the small-signal regime. This analysis suggests that the FO input can be fitted in the frequency (Fourier) domain with the FO measurements if the system impulse responses between candidate input–output buses are available. To this end, we genuinely consider using the impulse responses recovered from ambient measurements collected during normal operation of the power system. Our earlier works assert that the response recovery is analytically guaranteed for both internal system states at generators, as well as output variables at buses and transmission lines [11,12]. By *output variables*, we refer to the algebraic variables of the differential–algebraic equations governing grid dynamics, such as terminal voltage angles/frequencies and active line flows. The analysis conducted in this work suggests that when direct measurements of generator states are not available, the *relative phase shifts* between output data could still be accurately estimated and, subsequently, used to identify the FO source.

We also show that the proposed framework can be readily generalized to handle FO scenarios with multiple sources and frequencies.

A two-stage algorithm is accordingly developed to recover the FO sources. Before the FO events, ambient PMU measurements are collected to infer impulse responses. When an FO event is detected, the recovered responses are used to identify FO sources. Heed that our algorithm requires measurements only at a nearby bus for each possible FO source. Moreover, the algorithm can leverage diverse types of measurements (including line flows) at any location in contrast to energy-based methods such as [4]. Finally, the proposed algorithm is computationally efficient, as most of the computational burden is on the offline recovery of impulse responses.

The main contribution of this work is two-fold. First, we analyze the FO responses for both generator states and output variables, and formulate localization problems for both cases accordingly. The proposed FO localization framework utilizes the impulse responses recovered using ambient measurements, and builds upon linearized grid dynamics under reasonable assumptions for wide-area interconnections. Second, we develop a computationally efficient data-driven FO localization algorithm that utilizes PMU data collected before and during the FO event. The algorithm requires no knowledge of the actual system model or dynamics parameters and is flexible regarding measurement types and locations.

The rest of this paper is organized as follows. Section 2 introduces the linear system dynamics and formulates the FO localization problem. Section 3 deals with source localization first from generator state FO data using impulse responses recovered from ambient data. It then generalizes results to collecting FO data from output variables and to settings of FOs having multiple sources. A two-stage localization algorithm is developed in Section 4. Numerical tests on the IEEE 68-bus system and the IEEE-NASPI FO contest dataset in Section 5 demonstrate the effectiveness of the novel algorithm under different input and measurement conditions. The paper is concluded in Section 6.

2. Problem formulation

In the vicinity of an operating point, grid dynamics can be approximated by a linearized state-space model [13, Ch. 14]:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t). \tag{1}$$

Here vector $\mathbf{x} \in \mathbb{R}^m$ collects all system state variables, such as generators' internal states, while $\mathbf{u} \in \mathbb{R}^n$ collects system inputs corresponding to disturbances to nodal power balance. Both $\mathbf{x}$ and $\mathbf{u}$ represent deviations from the operating point. Matrices $\mathbf{A}$ and $\mathbf{B}$ are the state matrix and input matrix, respectively. The linearized model in (1) approximates well the power system dynamics of interest to our work focusing on the small-signal oscillation regime [10][13, Ch. 14]. Note that (1) is a multiple-input multiple-output (MIMO), and linear time-invariant (LTI) system. Given the system inputs $\mathbf{u}(t)$, the states $\mathbf{x}(t)$ can be fully determined using the system's impulse response with the subscripts here indicating the input and output. For a single input $u_\ell(t)$ at location ℓ, the output $\mathbf{x}(t)$ can be found as the convolution of $u_\ell(t)$ with the ℓth column of the impulse response matrix $T_{\mathbf{u},\mathbf{x}}(t)$. Expressed in the Fourier domain per the convolution theorem, we get that

$$\hat{\mathbf{x}}(\xi) = \hat{T}_{u_\ell,\mathbf{x}}(\xi) \cdot \hat{u}_\ell(\xi), \ \forall \ell = 1, \ldots, n \tag{2}$$

where the notation $\hat{\cdot}(\xi)$ stands for the Fourier-domain representation of a signal evaluated at frequency $\xi \in \mathbb{R}$.

Forced oscillations (FOs) are sustained oscillations caused by periodic inputs to power systems. Possible FO sources include cyclic loads, malfunctioning equipment, and controller failures; see [1] for more discussions. Without loss of generality, this work considers FO inputs due to generator controller failures. Typically, turbine governor control directly affects the mechanical power balance in the form of perturbing input $u_\ell(t)$. Recent work [14] has found that exciter control can also

be modeled with a similar effect on $u_\ell(t)$. Thus, we focus on these two types of FO sources due to generator controls.

If $u_\ell(t)$ is a periodic sinusoidal signal with F frequency components, its Fourier-domain representation becomes

$$\hat{u}_\ell(\xi) \triangleq \sum_{j=1}^{F} \frac{A_j}{2} \left[e^{i\phi_j} \Delta(\xi - \xi_j) + e^{-i\phi_j} \Delta(\xi + \xi_j) \right] \tag{3}$$

where $\Delta(\cdot)$ denotes the Dirac delta function. Per component j, symbols ξ_j, A_j, and ϕ_j denote its frequency, amplitude, and phase angle, respectively. The number of frequency components F is known to be small with a typical bound of $F \leq 2$ [1]. Although a single-frequency input could induce multiple harmonics of odd orders, most of these harmonics are likely to lie outside the [0.1,0.8] Hz range of inter-area oscillation modes. We first focus on single-frequency oscillations and handle the case $F = 2$ in Section 3.2.

For $F = 1$, the single frequency ξ_1 in (3) will be simply denoted by ξ^*. If ξ^* is close to the frequency of a system mode, the *resonance condition* could lead to large-amplitude oscillations observed at locations away from the source, which greatly complicates FO localization [15]. Under resonance, the frequency ξ^* can be easily estimated by selecting the largest frequency component in the recorded data, that is $\xi^* := \arg\max_\xi \|\hat{\mathbf{x}}(\xi)\|_2$.

Knowing ξ^* and the system's impulse response, we can now estimate the input mode $\hat{u}_\ell$ for each possible source location ℓ using the *least-squares (LS)* error criterion applied to (2). This way, the task of FO source localization amounts to selecting the index ℓ minimizing

$$\min_{1 \leq \ell \leq n} \ \min_{\hat{u}_\ell \in \mathbb{C}} \left\| \hat{T}_{u_\ell,\mathbf{x}}(\xi^*) \cdot \hat{u}_\ell - \hat{\mathbf{x}}(\xi^*) \right\|_2^2 . \tag{4}$$

For each ℓ, the inner LS problem enjoys a closed-form solution. This makes it very fast to search for the location with the smallest LS fitting error, as detailed later.

Remark 1 (*Generalizability*). The linearized system model in (1) has been introduced to simplify the analysis; the tests of Section 5 are conducted on realistic settings with nonlinear dynamics and controllers in place. In addition, while only generator controllers have been considered as the FO sources, our approach could be extended to other FO types. As pointed out in [1], cyclic outputs from loads and batteries can arise from controller issues in their power-electronic interfaces. However, these controllers tend to have much faster timescales, and thus are not likely to resonate with inter-area modes that are of higher interest to the FO problem. Moreover, the assumptions of single-mode and single-source for the FO taken in (4) represent most of FO cases observed in practice. The generalized scenario of FOs with multiple modes and/or sources is discussed in Section 3.2 and tested later on, as well.

3. FO localization via recovered responses

While the system's impulse responses enable an efficient FO localization per (4), it can be challenging to obtain them without having an accurate model of the power system and the dynamic parameters of all generators. To tackle this predicament, we leverage an approach to estimate impulse responses using synchrophasor data under ambient conditions alone [11,12]. The approach recovers impulse responses by cross-correlating ambient PMU data between pairs of system inputs-outputs, and enjoys guaranteed recovery performance under some reasonable assumptions listed below. For the rest of the paper, we use the superscript $\cdot^a$ to denote recorded signals under ambient conditions and the superscript $\cdot^f$ to denote signals recorded under FO conditions.

AS 1 (*Swing Dynamics*). Grid dynamics follow the second-order model $\mathbf{M}\ddot{\delta} + \mathbf{D}\dot{\delta} + \mathbf{K}\delta = \mathbf{u}$, with the state including generator rotor angles in δ and speed (frequency) deviations in $\omega = \dot{\delta}$.

AS 2 (*Homogeneous Damping*). The ratio between inertia and damping constants is homogeneous across generators, namely $\mathbf{D} = \gamma\mathbf{M}$ with a constant $\gamma > 0$.

AS 3 (*Lossless System*). Under lossless lines, the power flow Jacobian $\mathbf{K}$ is a symmetric Laplacian matrix.

AS 4 (*Ambient Condition*). The ambient system input is a zero-mean white-noise random process with locational variance proportional to the inertia constant, i.e., having covariance matrix $\boldsymbol{\Sigma} = \alpha\mathbf{M}$ with $\alpha > 0$.

While these assumptions are postulated to simplify the analysis, they tend to hold pretty well in large interconnections. Swing dynamics in (AS 1) has been popularly used for small-signal analysis to study key oscillation modes [13, Ch. 14]. The proportional damping in (AS 2) for approximating the swing dynamics has been considered in [11, 12,16,17], as a simplification of practical damping settings. For high-voltage transmission lines, power losses are negligible leading to (AS 3) even with most common load models like constant-current ones [18]. Finally, the ambient condition in (AS 4) implies that areas of larger perturbations have more inertia deployed. While this assumption may not be very realistic, our analysis in [11,12] essentially used it to show that all ambient modes of interest are equally excited.

Under these assumptions, the pair-wise impulse responses can be recovered by cross-correlating the corresponding ambient PMU measurements as formally stated next [11,12].

Lemma 1 (*State's Impulse Responses*). *Under (AS 1)–(AS 4), the impulse responses of rotor angle or frequency are equivalent to the cross-correlation of the related ambient signals:*

$$T_{u_\ell,\omega_k}(t) = \frac{2\gamma}{\alpha}C_{\omega_\ell^a,\omega_k^a}(t) = -\frac{2\gamma}{\alpha}\frac{d}{dt}C_{\omega_\ell^a,\delta_k^a}(t) \tag{5}$$

$$T_{u_\ell,\delta_k}(t) = -\frac{2\gamma}{\alpha}C_{\omega_\ell^a,\delta_k^a}(t) = -\frac{2\gamma}{\alpha}\frac{d}{dt}C_{\delta_\ell^a,\delta_k^a}(t) \tag{6}$$

for $t \geq 0$. Here $C_{\cdot,\cdot}$ denotes the cross-correlation operator and the subscripts indicate its two arguments. Because $T_{u_\ell,\omega_k}(t)$ is the derivative of $T_{u_\ell,\delta_k}(t)$, each one of them can be recovered from the other using time differentiation or integration.

Lemma 1 asserts that the cross-correlation of ambient data can recover the impulse responses. For simplicity, we will use $C_{\omega_\ell^a,\mathbf{x}^a}(t)$ to represent the recovered $T_{u_\ell,\mathbf{x}}(t)$, assuming suitable operations have been taken per Lemma 1. The former enables us to establish a data-driven FO localization approach in the following.

Proposition 1 (*Data-driven FO Localization*). *Using the recovered impulse responses, the FO source can be identified based on the estimated FO frequency ξ^* by solving:*

$$\min_{1\leq\ell\leq n}\ \min_{\hat{u}_\ell\in\mathbb{C}}\left\|\hat{C}_{\omega_\ell^a,\mathbf{x}^a}(\xi^*)\cdot\hat{u}_\ell - \hat{\mathbf{x}}^f(\xi^*)\right\|_2^2. \tag{7}$$

Proposition 1 builds upon the recovered responses as the feature vectors for each possible source. Hence, any recovery accuracy issue faced by the data-driven approach in Lemma 1 could lead to errors in FO localization, as shown by our numerical tests. Heed that there is an unknown scaling factor of $(2\gamma/\alpha)$ in Lemma 1. Nonetheless, this factor is the same amongst all locations under (AS 2) and (AS 4), and can thus, be omitted upon solving (7). The proposed data-driven localization approach can benefit from the fact that each possible source location would lead to very different system-wide responses. It is also known that the participation factors of an oscillation mode at individual generators can be very different, especially from one area to another [19, Sec. 8.3]. This renders the proposed FO approach more robust against possible recovery errors as demonstrated in Section 5. In general, the more state variables are observed, the better the localization performance.

Back to (7), the solution of the inner LS problem is simply:

$$\hat{u}_\ell \leftarrow \left[\hat{C}_{\omega_\ell^a,\mathbf{x}^a}(\xi^*)\right]^+ \hat{\mathbf{x}}^f(\xi^*) \tag{8}$$

where superscript $^+$ denotes the matrix's pseudo-inverse. Accordingly, the LS fitting error for location ℓ becomes

$$\left\|\hat{\mathbf{x}}^f(\xi^*) - \hat{C}_{\omega_\ell^a,\mathbf{x}^a}(\xi^*)\left[\hat{C}_{\omega_\ell^a,\mathbf{x}^a}(\xi^*)\right]^+\hat{\mathbf{x}}^f(\xi^*)\right\|_2^2.$$

This closed-form solution of LS makes it very convenient to evaluate the fitting error for each possible input location and thus to search for the best one.

3.1. Localization based on output observations

Proposition 1 assumes synchrophasor readings of generators' internal states, such as δ and ω, under both ambient and oscillatory conditions. However, PMUs can only measure algebraic variable outputs at the power system level, at buses and lines. We next extend the proposed approach to consider measurements of system outputs. Specifically, we consider measurements of bus voltage phase angle θ_i and frequency f_i per bus i, as well as the active line flow p_{ij} for the line (i, j) connecting buses i and j. Other measurement types such as voltage magnitudes and reactive power could be included.

By linearizing the power flow at a given operating point, system outputs $\mathbf{y}$ can be linearly related to system states through an observation matrix $\mathbf{C}$ as

$$\mathbf{y} \simeq \mathbf{Cx} \tag{9}$$

where $\mathbf{y} \in \mathbb{R}^{m'}$ collects all the aforementioned bus and line outputs; recall that the dynamic model in (1) was also derived upon linearization. Our earlier work [12] has utilized the linearized model (9) to extend Lemma 1 to consider output responses for bus and line variables, as stated next.

Lemma 2 (*Output's Impulse Responses*). *Under (AS 1)–(AS 4) and (9), the impulse response of bus angle θ_i and frequency f_i have the following equivalence with $t \geq 0$:*

$$T_{u_\ell,f_i}(t) = \frac{2\gamma}{\alpha}C_{\omega_\ell^a,f_i^a}(t) \tag{10}$$

$$T_{u_\ell,\theta_i}(t) = -\frac{2\gamma}{\alpha}C_{\omega_\ell^a,\theta_i^a}(t). \tag{11}$$

For line flow p_{ij}, it similarly holds that:

$$T_{u_\ell,p_{ij}}(t) = -\frac{2\gamma}{\alpha}C_{\omega_\ell^a,p_{ij}^a}(t). \tag{12}$$

Similar to Proposition 1, Lemma 2 enables us to use the recovered responses for bus angles/frequencies and line flows to localize FOs. Nonetheless, recovering these impulse responses still relies on measurements of state variables, i.e., the input generator speed $\omega_\ell^a(t)$ in (10)–(12). Therefore, Lemma 2 is not practically useful. To tackle this issue, we could use the frequency at the nearest bus for each candidate FO source as a surrogate, i.e., using $f_k^a(t)$ at the bus k that is closest to input ℓ to replace $\omega_\ell^a(t)$. This naive approach could expand Proposition 1 to using all system output measurements, but may introduce inaccuracies to FO localization in (7). The nearest-bus frequency $f_k^a(t)$ does not exactly match the generator speed $\omega_\ell^a(t)$, because per (9) the former is in fact a linear combination of all generator states, not just the closest one. This mismatch has also been confirmed numerically, leading to high FO localization error for the naive approach.

To address this observability predicament, we develop a formal approach to account for the inaccuracy in the recovered responses. Consider the Fourier transform of the cross-correlation $C_{y_i^a,y_j^a}(t)$ between two outputs at buses i and j. Its Fourier-domain representation is called the *cross power spectral density (CPSD)* and will be denoted by $\hat{C}_{y_i^a,y_j^a}(\xi)$.

Interestingly, the phase angle of CPSD can be shown to reflect the relative phase shift from $\hat{T}_{u_\ell, y_i}$ to $\hat{T}_{u_\ell, y_j}$ as follows [20]:

$$\angle \hat{C}_{y_i^a, y_j^a}(\xi) = \angle \hat{T}_{u_\ell, y_j}(\xi) - \angle \hat{T}_{u_\ell, y_i}(\xi), \forall \xi. \tag{13}$$

Hence, the phase angles of the terms in (2) relate as:

$$\angle \hat{y}_j^f(\xi) = \angle \hat{u}_\ell(\xi) + \angle \hat{T}_{u_\ell, y_j}(\xi)$$
$$= \angle \hat{u}_\ell(\xi) + \angle \hat{T}_{u_\ell, y_i}(\xi) + \angle \hat{C}_{y_i^a, y_j^a}(\xi), \forall \xi. \tag{14}$$

Eq. (14) asserts that we can use the measured output signal $y_i^a(t)$ as a surrogate for the non-measured candidate input signal $\omega_\ell^a(t)$. Nonetheless, by doing so, we introduce an unknown angle $\angle \hat{T}_{u_\ell, y_i}(\xi)$ for all output FO measurements in $\hat{\mathbf{y}}$. Because this angle ambiguity is the same for all output locations, it can be eliminated by fitting the angle relations in (14) with an unknown input angle, instead of the input frequency component in Proposition 1. In other words, for each candidate location ℓ and surrogate location i, we can alternatively estimate $\angle \hat{u}_i(\xi) = \angle \hat{u}_\ell(\xi) + \angle \hat{T}_{u_\ell, y_i}(\xi)$ instead of $\angle \hat{u}_\ell(\xi)$ alone; recall each candidate location ℓ is paired up with a single surrogate location i. This allows us to establish a data-driven FO localization approach by altering the LS error fitting procedure as shown below.

Proposition 2 (*FO Localization Using Output Data*). *By using the ambient signal y_i^a at the closest bus i for any input location, the FO source can be identified using the phases of the CPSD terms at the estimated FO frequency ξ^* by solving:*

$$\min_{1 \leq i \leq n} \min_{\angle \hat{u}_i} \| \angle \hat{u}_i + \angle \hat{C}_{y_i^a, \mathbf{y}^a}(\xi^*) - \angle \hat{\mathbf{y}}^f(\xi^*) \|_2^2. \tag{15}$$

Similar to (7), the inner LS problem in (15) features a simple closed-form solution given by

$$\angle \hat{u}_i \leftarrow \frac{1}{m'} \sum_{j=1}^{m'} \left(\angle \hat{y}_j^f(\xi^*) - \angle \hat{C}_{y_i^a, y_j^a}(\xi^*) \right). \tag{16}$$

This easy solution makes it very efficient to search for the best input location. Note that (15) is based solely on output measurements, recorded under ambient conditions and FO conditions. Moreover, it only requires to identify the surrogate output y_i at bus i for each possible input location. Assuming a reasonable level of PMU deployment, the size of output vector $\mathbf{y}$ could be sufficiently large to allow for distinguishing each input location. Additionally, it can incorporate other types of outputs at any location, thanks to the flexibility of Lemma 2. Hence, the proposed FO localization method can readily incorporate output measurements. In the numerical tests, we will further demonstrate the flexibility of our approach when output measurements closest to the FO source are unavailable.

3.2. Multiple FO sources and modes

Although most of the analysis thus far focuses on a single FO source with one oscillation frequency, it is possible to extend our results to multiple sources/modes based on the data-driven framework for recovering output responses. Note that the case of single source with multiple frequencies is straightforward to handle. This is because a frequency-domain analysis of the oscillation output $\hat{\mathbf{y}}^f$ can estimate the number and accordingly the frequencies of underlying oscillation modes by using a prescribed threshold. However, the case of multiple, unknown number of, source locations could greatly complicate the FO localization, as it would become a combinatorial search over all the possible scenarios. Fortunately, it is practically reasonable to assume at most two input source locations for the FO problem, as the chance of having more than two coincidental sources would diminish to almost zero in practice [1]. Hence, we consider the extension to at most two FO sources with a total of F estimated oscillation frequencies. Accordingly, each frequency ξ_j^* has been obtained by comparing its component with a prescribed threshold $\mathcal{T}$, such that $\| \hat{\mathbf{y}}^f(\xi_j^*) \|_2 > \mathcal{T}$.

Based on the identified frequencies, we extend (15) to search over any possible pair of input locations $\{i_1, i_2\}$, by solving

$$\min_{1 \leq i_1, i_2 \leq n} \sum_{j=1}^{F} \min_{\angle \hat{u}_{i_1, i_2}} \left\| \angle \hat{u}_{i_1, i_2} - \angle \hat{\mathbf{y}}^f(\xi_j^*) \right.$$
$$\left. + \angle \hat{C}_{y_{i_1}^a, \mathbf{y}^a}(\xi_j^*) + \angle \hat{C}_{y_{i_2}^a, \mathbf{y}^a}(\xi_j^*) \right\|_2^2. \tag{17}$$

Clearly, each subproblem still follows the same structure as the inner LS minimization in (15) with the averaging solution. This allows us to quickly compute the minimal fitting error for a candidate pair $\{i_1, i_2\}$, and then pick the pair with the best fit. For the possibility of two input locations, the number of candidate choices becomes $n(n-1)/2$ as compared to the n choices for the case of a single source. This increased computational complexity could be addressed by using a more efficient search method based on e.g., sparse signal recovery or compressed sensing [21]. We do not consider the latter here due to the scope limit but will investigate this computation issue with numerical tests later in Section 4.

4. FO localization algorithm

The proposed scheme is implemented as a two-phase algorithm tabulated as Algorithm 1. We consider the realistic condition that PMUs are installed at critical substations with large generation and lines with high power flows connecting control areas. We also assume there is a PMU installed at one nearby bus for each possible FO source generator. During *Phase 1*, pair-wise impulse responses are inferred from ambient measurements. First, we apply a bandpass filter with passband [0.1,0.8] Hz on ambient data $\mathbf{x}^a(\mathbf{y}^a)$ to find the detrended signals [1]. For simplicity, both state vector $\mathbf{x}$ and output vector $\mathbf{y}$ are denoted by $\mathbf{x}$ in this section. Next, the detrended data streams are cross-correlated for all pairs of input/output locations per Lemmas 1 and 2. Depending on the type of impulse responses of interest and ambient data, numerical differentiation may be taken to recover $T_{u_\ell, \mathbf{x}}$ for all possible input locations ℓ. The Fast Fourier Transform (FFT) will be taken to generate $\{ \hat{T}_{u_\ell, \mathbf{x}}(\xi) \}_{\ell=1}^n$.

When an FO event is detected, *Phase 2* of the proposed algorithm is initiated to localize the FO source. First, the algorithm runs FFT analysis on $\mathbf{x}^f$ to generate $\hat{\mathbf{x}}^f(\xi)$, and identify the oscillation frequency ξ^*. Next, for each possible ℓ, the LS solution for input $\hat{u}_\ell$ is calculated using the recovered $\hat{T}_{u_\ell, \mathbf{x}}(\xi^*)$ by (4) or (15) depending on the measurement locations. Finally, the algorithm reports the location ℓ^* with the smallest fitting error of (4) or (15). Note that the neighboring set $\mathcal{N}_{\ell^*}$ of location ℓ^* per graph connectivity will also be reported, as the estimation error of recovered impulse responses may lead the identified source to a nearby location.

To show the computational efficiency of the algorithm, we consider *Phase 1* and *Phase 2* separately. We consider a system with n measurement locations (PMUs) for simplicity. In *Phase 1*, the computational complexity is dominated by cross-correlation. Assume each PMU takes $\mathcal{M}_1$ ambient samples. The complexity for each pair of cross-correlation is $\mathcal{O}(\mathcal{M}_1^2)$ [22]. Thus, the complexity of *Phase 1* is $\mathcal{O}(n^2 \mathcal{M}_1^2)$. In *Phase 2*, assume each PMU takes $\mathcal{M}_2$ FO samples. The complexity of the FFT step is $\mathcal{O}(n \mathcal{M}_2 \log \mathcal{M}_2)$. Additionally, the complexity of calculating LS solutions and report the FO source is $\mathcal{O}(n^2)$. In general, the number of FO samples $\mathcal{M}_2$ is larger than the number of locations n. Thus, the computational complexity of *Phase 2* is $\mathcal{O}(n \mathcal{M}_2 \log \mathcal{M}_2)$.

To implement Algorithm 1, we recommend using at least 10 min of ambient data to recover impulse responses, and 20 s of FO data for localization. The length of recommended measurements indicate that $n \ll \mathcal{M}_2 \ll \mathcal{M}_1$. Thus, the computational burden is mostly at the off-line stage. For on-line FO source localization, *Phase 2* is very efficient thanks to the closed-form LS solution, and the small number of required FO samples. Note that for the two-source scenario, the complexity of solving (17) becomes $\mathcal{O}(n^3)$, which does not alter the overall complexity as $n \ll \mathcal{M}_2$.

Algorithm 1 FO Localization

1: **Input:** Ambient data $\mathbf{x}^a \in \mathbb{R}^{n \times t1}$, oscillation data $\mathbf{x}^f \in \mathbb{R}^{n \times t2}$, number of sources n, passband $[f_1, f_2]$.
2: **Output:** The candidate oscillation source ℓ^* by LS solution, and set $\mathcal{N}_{\ell^*}$ per graph connectivity.
3: **Initialize**: Set the iteration index $i, k = 1$, inferred impulse response set $\{C_{x_i^a, x_k^a}\}_{i,k=1}^n \in \mathbb{R}^{t1}$, variables in frequency domain: $\{\hat{C}_{x_i^a, x_k^a}(\xi)\}_{i,k=1}^n \in \mathbb{C}^{t1}$, $\{\hat{x}_i^f(\xi)\}_{i=1}^n \in \mathbb{C}^{t2}$, recovered oscillation input $\hat{\mathbf{u}} \in \mathbb{C}^n$, oscillation mode ξ^*.

Phase 1 - Impulse Responses Inference (off-line)

4: **for** i in range(n) **do**
5:     Filter and update $\mathbf{x}_i^a$:
      $\mathbf{x}_i^a \leftarrow bandpass\left(\mathbf{x}_i^a - mean(\mathbf{x}_i^a), [f_1, f_2]\right)$.
6: **end for**
7: **for** i in range(n) **do**
8:     **for** k in range(n) **do**
9:         Cross-correlate ambient data streams: $C_{x_i^a, x_k^a}$.
10:         **if** Derivative required for $C_{x_i^a, x_k^a}$ **then**
11:             Update $C_{x_i^a, x_k^a}$: $C_{x_i^a, x_k^a} \leftarrow \frac{d}{dt} C_{x_i^a, x_k^a}$.
12:         **end if**
13:     **end for**
14:     Run FFT and update $\hat{C}_{x_i, \mathbf{x}}(\xi)$: $\hat{C}_{x_i, \mathbf{x}}(\xi) \xleftarrow{\text{fft}} C_{x_i, \mathbf{x}}$.
15: **end for**
16: **Return:** inferred impulse response set $\{\hat{C}_{x_i, \mathbf{x}}(\xi)\}_{i=1}^n$.

Phase 2 - FO solver (on-line)

17: Run FFT and update $\hat{\mathbf{x}}^f$: $\hat{x}^f(\xi) \xleftarrow{\text{fft}} \mathbf{x}^f$.
18: Identify oscillation mode and update ξ^*:
    $\xi^* \leftarrow \arg\max_\xi |\hat{\mathbf{x}}^f(\xi)|$.
19: **for** i in range(n) **do**
20:     **if** $\mathbf{x}^a$ taken at generators **then**
21:         Solve LS problem and update $\hat{u}_i$: $\hat{u}_i \leftarrow$ (8).
22:     **else**
23:         Solve LS problem and update $\angle \hat{u}_i$: $\angle \hat{u}_i \leftarrow$ (16).
24:     **end if**
25: **end for**
26: Pick ℓ^* with the smallest fitting error, and generate its neighboring set $\mathcal{N}_{\ell^*}$.
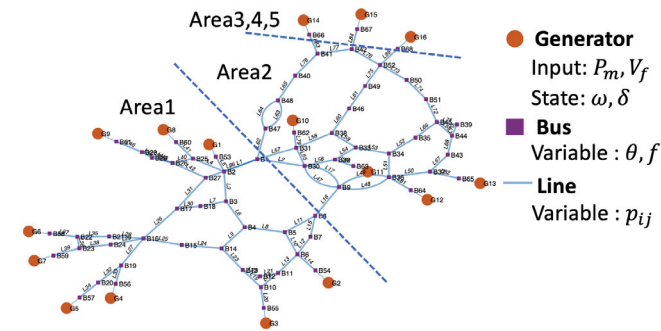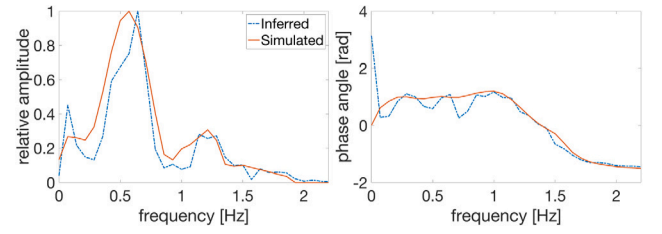27: **Return:** ℓ^*, $\mathcal{N}_{\ell^*}$.

Fig. 1. Diagram of the 68-bus system with 16 generators.

Fig. 2. Spectrum of rotor speed ω_{13} response from u_1: simulated vs. inferred from ambient data.

5. Numerical results

The proposed method has been numerically evaluated on the IEEE 68-bus system, and the 240-bus system from the IEEE-NASPI oscillation source location contest [23]. As shown in Fig. 1, the 68-bus system has 16 generators, where the 9 generators in Area 1 are equipped with exciters. For this system, impulse responses, ambient data, and forced oscillations were generated in MATLAB using PST [24], as detailed soon. We considered *all* generators as FO input locations. We commenced with the single-source, single-FO scenario, with the following cases: *(i)* Direct measurements of generator states with full observability; *(ii)* Measurements of generators' terminal variables, again with full observability; and *(iii)* Measurements only at surrogate buses along with buses and lines as outputs. We subsequently explored the cases of two FO sources and two FO frequencies.[1] For the 240-bus system, we have used the FO dataset from the IEEE-NASPI contest [23], and an ambient dataset generated by the TSAT simulator with random load perturbations for the same system.

For the 68-bus system, we simulated nonlinear generator models with control effects and set system inputs at generator mechanical power and exciter voltage references. For impulse responses, we ran time-domain simulations with an impulse-like function (very short rectangular wave?) as input. To generate ambient signals, the mechanical power and exciter references of all generators were perturbed by zero-mean white Gaussian noise. To generate FO signals, the generator's mechanical power or exciter's reference was perturbed by a sinusoidal signal with a frequency close to a system mode. We selected the four modes 0.143 Hz, 0.43 Hz, 0.57 Hz, and 0.714 Hz (identified as system modes from the simulated impulse responses) as FO input frequencies. The sampling rate for all tests was set to 200 Hz (time-step of $dt = 0.005$ s). Note that the simulated system is nonlinear; thus, the LTI assumption in (AS 1) has some mismatch. The line resistance-to-reactance ratio was small, and damping constants $\mathbf{D}$ were proportional to inertia constants $\mathbf{M}$, making (AS 2)–(AS 3) hold pretty well.

5.1. Single FO source with single FO mode

We first validated the algorithm using measurements of generator states. Despite violations of model assumptions, the Fourier-domain responses matched very well with the model-based responses in terms of amplitude and phase angle within the frequency band of interest. Fig. 2 compares the two responses for rotor speed ω. Clearly, the match between the two is very close, except for some small deviations, primarily due to model approximations. By using the recovered responses, we were able to localize most FOs via the LS solutions in (4). Among all 64 scenarios, the only error of LS solutions comes from the less significant system mode 0.57 Hz. However, the source can still be recovered from the neighboring set of the identified source, showing the validity of the proposed model and algorithm. The experiment setup and test results are shown in Table 1.

Next, we simulated a scenario with measurements of the generators' terminal buses. As mentioned before, due to the power flow coupling, the approximation error between generator speed and terminal bus frequency manifests itself in estimated responses, and leads to errors in FO source recovery. The numerical tests suggest that despite the reduced accuracy in inferred responses, the proposed algorithm can still recover most FO sources correctly from LS solutions and their neighboring sets, as shown in Table 1.

---

[1] Codes and results are available at: https://github.com/ShaohuiLiu/fo_local.
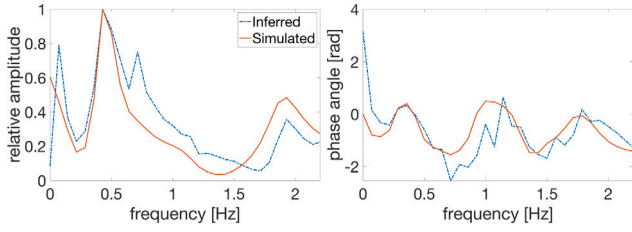
Fig. 3. Spectrum of rotor speed ω_{13} (measured by f_{37}) response from u_{11}: simulated vs. inferred from ambient data.

Table 1
Localization accuracy for single-source FOs.

Input	Sensor coverage	LS	Graph
Governor	Rotor (Full)	98.40%	100%
Governor	Bus (Full)	92.20%	96.90%
Governor	Bus (Partial)	87.53%	98.50%
Governor	Bus&Line (Partial)	82.80%	92.20%
Exciter	Bus (Full)	30.60%	97.20%
Exciter	Bus (Partial)	36.10%	94.50%

Table 2
Localization accuracy for 2 FO sources.

Sensor coverage	LS	Graph	Major mode
Rotor (Full)	90.23%	98.05%	99.61%
Bus (Partial)	43.95%	91.21%	96.48%

We also tested the proposed algorithm under partial sensor coverage with no measurements at generator buses. Each generator was surrogated by a measurement at a nearby bus. In the 68-bus system, the assumed PMUs were placed within at most 2 hops away as depicted in Fig. 1. We considered measurements of both types, bus frequency f_i and line flow p_{ij}. As suggested in Section 3.2, relative phase shifts are used for FO localization when measurements for input locations are not near generators. The results advocate that FO sources recovered from LS solutions are less accurate due to inaccurate response estimations, as shown by the example in Fig. 3. The proposed algorithm still points out the correct FO source area, as the correct source can be recovered from the reported neighboring set in most cases, as reported in Table 1.

Finally, disturbances originating from the generator's excitation system were considered. We perturbed the voltage reference at 9 generators equipped with exciters to generate simulated impulse responses and FOs. For FOs, we simulated 4 identified modes, so there were 36 scenarios. In this case, excitation system outputs are dominated by the dynamics of the excitation system. Thus, the responses of states and variables are delayed, and the linearized system may not reflect the actual responses well. Under the realistic partial sensor coverage setup, we noticed that the FO sources recovered from LS solutions were inaccurate. Nevertheless, the correct FO source area is still pointed out correctly in the neighboring set in most cases, as reported in Table 1.

5.2. Two FO sources with multiple modes

We also considered the case of two FO sources at two distinct input frequencies. The input frequencies were selected at the identified system modes, with 0.143 Hz set as the major mode with twice the input amplitude as 0.43 Hz. This case is more challenging as the major mode may dominate system oscillations. As detailed in Table 2, the proposed algorithm can still identify both FO frequencies and locate the sources correctly for almost all 256 possible scenarios. With full observability on rotor states, Algorithm 1 missed the actual major FO source in a single scenario and successfully identified both sources for 251 scenarios. Under the more realistic condition with measurements only at proxy buses, Algorithm 1 still achieves high accuracy in locating the major FO
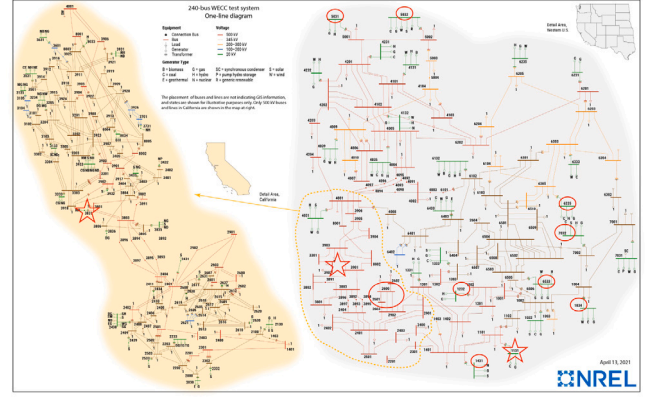

Fig. 4. Diagram of the 240-bus system with 26 monitored generators [25]. Circled locations are monitored generators with both name and rating matched at the ambient and FO datasets. Locations with stars are FO sources in Case 3 and Case 4.

source. Both sources can be recovered from the reported neighboring set for the vast majority of scenarios, showing the effectiveness and robustness of the proposed FO localization algorithm.

5.3. Validation with the IEEE-NASPI FO contest dataset

This 240-bus system was used as a benchmark by the FO localization contest [23]. The power system includes a total of 13 test cases. Here, we focus on Case 3 and Case 4, which are considered to be the most challenging cases with both resonance conditions and unmonitored sources. A 15-min ambient dataset was used to infer impulse responses, while FO signals lasting for 15 s from the contest dataset were subsequently used for the localization. Due to some mismatches in the naming and data types between the available ambient and FO datasets, we were only able to utilize the rotor frequency data at 9 out of the 26 measured generators to evaluate Algorithm 1, as marked in Fig. 4. For Case 3, Algorithm 1 has successfully identified bus 1431 as the FO source regardless of the presence of fault-induced transients. Indeed, bus 1431 is the closest bus to bus 1131, the actual source, among all monitored locations. For the 3 scenarios under Case 4 with mostly pre-fault FO signals, Algorithm 1 identified bus 2634, which is also the closest one to the actual source at bus 3831. If the post-fault FO signals were considered, the source was identified as bus 1232, which is also very close to the actual source. This numerical test has shown promising results in using the proposed algorithm for a realistic large test system, even if the FO source is not directly monitored.

6. Conclusions

This work has developed a data-driven framework for FO localization with dynamics responses recovered from ambient synchrophasor data. FO responses have been analyzed for both generator states and output variables as buses and lines. Leveraging the recovered response, the FO localization task has been posed as an LS fitting problem. A computationally efficient algorithm has been proposed that uses ambient and oscillatory phasor measurements to identify FO sources. The algorithm is flexible regarding location and types of PMU data, and can provide a system operator with a neighboring set of possible FO locations. The proposed method has been numerically validated using the IEEE 68-bus system under realistic nonlinear dynamics with control effects, as well as a dataset on a 240-bus system generated for the IEEE-NASPI contest [23]. The method could pinpoint FO sources for the vast majority of scenarios, under full and partial observability, governor or exciter oscillations. Future research directions include model-based analysis for different types of FO sources, performance analysis for various measurement types and operation conditions, as well as numerical validation using real-world PMU data.

CRediT authorship contribution statement

Shaohui Liu: Data curation, Methodology, Software, Writing – original draft. **Hao Zhu:** Funding acquisition, Methodology, Supervision, Writing – review & editing. **Vassilis Kekatos:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Hao Zhu, Vassilis Kekatos reports financial support was provided by National Science Foundation (US).

Data availability

Data will be made available on request.

Acknowledgments

References

[1] NERC, Recommended Oscillation Analysis for Monitoring and Mitigation Reference Document, NERC, Atlanta, GA, USA, 2021.

[2] L. Chen, D. Trudnowski, L. Dosiek, S. Kamalasadan, Y. Xu, X. Wang, Forced oscillations in power systems, in: IEEE TR110, IEEE, 2023.

[3] S. Maslennikov, B. Wang, E. Litvinov, Dissipating energy flow method for locating the source of sustained oscillations, Int. J. Electr. Power Energy Syst. 88 (2017) 55–62.

[4] D. Osipov, S. Konstantinopoulos, J.H. Chow, A cross-power spectral density method for locating oscillation sources using synchrophasor measurements, IEEE Trans. Power Syst. (2022) 1–9.

[5] L. Chen, Y. Min, W. Hu, An energy-based method for location of power system oscillation source, IEEE Trans. Power Syst. 28 (2) (2013) 828–836.

[6] S. Maslennikov, E. Litvinov, ISO New England experience in locating the source of oscillations online, IEEE Trans. Power Syst. 36 (1) (2021) 495–503.

[7] B.C. Lesieutre, Y. Abdennadher, S. Roy, Model-enhanced localization of forced oscillation using pmu data, in: Allerton Conf. on Communication, Control, and Computing, IEEE, 2022, pp. 1–8.

[8] T. Huang, N.M. Freris, P.R. Kumar, L. Xie, A synchrophasor data-driven method for forced oscillation localization under resonance conditions, IEEE Trans. Power Syst. 35 (5) (2020) 3927–3939.

[9] R. Delabays, A.Y. Lokhov, M. Tyloo, M. Vuffray, Locating the source of forced oscillations in transmission power grids, PRX Energy 2 (2023) 023009.

[10] Y. Cai, X. Wang, G. Joós, I. Kamwa, An online data-driven method to locate forced oscillation sources from power plants based on sparse identification of nonlinear dynamics (SINDy), IEEE Trans. Power Syst. 38 (3) (2022) 2085–2099.

[11] P. Huynh, H. Zhu, Q. Chen, A.E. Elbanna, Data-driven estimation of frequency response from ambient synchrophasor measurements, IEEE Trans. Power Syst. 33 (6) (2018) 6590–6599.

[12] S. Liu, H. Zhu, V. Kekatos, Dynamic response recovery using ambient synchrophasor data: A synthetic Texas Interconnection case study, in: Hawaii Intl. Conf. on System Sciences, HICSS, Maui, HI, 2023.

[13] R.B. Arthur, V. Vittal, Power System Analysis, second ed., London: UK, 2000, pp. 532–538.

[14] S.C. Chevalier, P.D. Hines, Mitigating the risk of voltage collapse using statistical measures from PMU data, IEEE Trans. Power Syst. 34 (1) (2018) 120–128.

[15] S.A.N. Sarmadi, V. Venkatasubramanian, Inter-area resonance in power systems from forced oscillations, IEEE Trans. Power Syst. 31 (1) (2016) 378–386.

[16] M. Jalali, V. Kekatos, S. Bhela, H. Zhu, Inferring power system frequency oscillations using Gaussian processes, in: IEEE Conf. on Decision and Control, Austin, TX, 2021, pp. 3670–3676.

[17] F. Paganini, E. Mallada, Global analysis of synchronization performance for power systems: Bridging the theory-practice gap, IEEE Trans. Autom. Control 65 (7) (2020) 3007–3022.

[18] P. Mackin, R. Daschmans, B. Williams, B. Haney, R. Hung, J. Ellis, Dynamic Simulation Studies of the Frequency Response of the Three U.S. Interconnections with Increased Wind Generation, LBNL Technical Report, 2010, [Online]. Available: https://www.osti.gov/biblio/1004165.

[19] P.W. Sauer, M.A. Pai, J.H. Chow, Power System Dynamics and Stability: with Synchrophasor Measurement and Power System Toolbox, John Wiley & Sons, 2017.

[20] D.J. Trudnowski, Estimating electromechanical mode shape from synchrophasor measurements, IEEE Trans. Power Syst. 23 (3) (2008) 1188–1195.

[21] H. Zhu, G.B. Giannakis, Sparse overcomplete representations for efficient identification of power line outages, IEEE Trans. Power Syst. 27 (4) (2012) 2215–2224.

[22] D. Hale, An Efficient Method for Computing Local Cross-Correlations of Multi-Dimensional Signals, CWP Report, Vol. 656, Citeseer, 2006.

[23] B. Wang, S. Maslennikov, IEEE-NASPI Oscillation Source Location Contest-Case Development and Results, Tech. Rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2021.

[24] J. Chow, K. Cheung, A toolbox for power system dynamics and control engineering education and research, IEEE Trans. Power Syst. 7 (4) (1992) 1559–1564.

[25] H. Yuan, R.S. Biswas, J. Tan, Y. Zhang, Developing a reduced 240-bus wecc dynamic model for frequency response study of high renewable integration, in: 2020 IEEE/PES Transmission and Distribution Conference and Exposition, T&D, IEEE, 2020, pp. 1–5.