

Data-Driven Estimation of Frequency Response from Ambient Synchrophasor Measurements

Phuc Huynh, *IEEE, Student Member*, Hao Zhu, *IEEE, Member*, Qianli Chen, and Ahmed Elbanna

Abstract—With the wide deployment of synchrophasor technology, measurement-based dynamic modeling and studies have been becoming increasingly useful for real-time grid operations. This paper considers the problem of estimating the power system frequency response from ambient synchrophasor measurements. Specifically, we develop the analytical conditions for establishing the equivalence between the cross-correlation of ambient generator speed data and the system frequency response between any two locations. Our conditions, relying on uniformly damped and equally excited oscillation modes, extend earlier work on electro-mechanical wave propagation modeling to non-homogeneous power networks. Numerical results suggest that the validity of the cross-correlation approach would hold for more realistic conditions such as non-uniform damping and high-order generator model. Its practical value is further corroborated by real data results, which closely match with the actual propagation time of electromechanical waves recorded during the 2008 Florida blackout in the Eastern Interconnection system.

I. INTRODUCTION

Power system frequency response to external disturbance is complex in nature and can often result in the so-termed electro-mechanical (EM) oscillations over the whole interconnection. These oscillations can be attributed to fast excitation systems and to weak tie lines; see e.g., [1, Ch. 1]. If the oscillation modes are poorly damped, a small disturbance input could trigger increasing level of oscillations and even wide-area outages, such as the 1996 US/Canada Western Interconnection blackout; see e.g., [2].

Traditionally, analyzing the oscillation modes relies on the detailed system dynamic models. Building these models, typically nonlinear, requires the full system information on synchronous generators, fast/slow exciters, and network components; see e.g., [3], [4]. *Small-signal analysis* of the linearized model around the operating point can provide the modal information and the primary frequency response. Nonetheless, this model-based framework is increasingly challenged by issues like outdated system information, bad data in state estimation, and numerical computation accuracy.

Manuscript received October 2, 2017; revised March 7, 2018; accepted April 22, 2018. This work was supported in part by the Siebel Energy Institute and the National Science Foundation (NSF) under Award Number ECCS-1802319. The authors would like to thank the University of Tennessee at Knoxville (UTK) for their support in providing the Frequency Disturbance Recorder (FDR) data.

P. Huynh is with the Department of Electrical & Computer Engineering, University of Illinois, 306 N. Wright Street, Urbana, IL 61801, USA; Email: pthuyh2@illinois.edu. H. Zhu is with the Department of Electrical & Computer Engineering, The University of Texas at Austin, 2501 Speedway, Austin, TX 78712, USA; Email: haozhu@utexas.edu. Q. Chen and A. Elbanna are with the Department of Civil & Environmental Engineering, University of Illinois, 205 North Mathews Ave., Urbana, IL 61801, USA; Emails: {qchen35, elbanna2}@illinois.edu

Measurement-based dynamic studies has become highly popular in the past decades thanks to the deployment of high-resolution and synchronized sensors in power systems such as phasor measurement units (PMUs). At little information regarding the underlying models, fast-sampled bus frequency/angle and voltage data could reveal important characteristics of the system dynamics. There exists a rich literature on oscillation mode estimation using either ambient data, ring-down signals, or probing responses; see e.g., [5]–[7] and references therein. Among these three types of system responses, ambient synchrophasor data has been increasingly explored due to its wide availability; see recent advances on fast subspace-based algorithms in [8], [9]. In addition, the statistical information in the ambient response has been exploited to estimate the power flow Jacobian matrix [10] or dynamic state Jacobian matrix [11], as well as to analyze sustained oscillations [12].

The present paper aims to develop a data-driven framework to estimate the grid frequency response from any input to output locations using ambient synchrophasor measurements. Different from existing methods focusing on temporal system responses, our approach explores the analytical conditions for recovering the spatial oscillation responses. This is related to the wave propagation effects of wide-area EM oscillations in interconnections [13]. By simplifying the grid using homogeneously placed generation/loads and uniform line parameters, a continuum modeling of planar EM waves was developed in [14] to estimate the wave propagation speed. Interestingly, this continuum EM wave model is analogous to that of seismic waves during earthquakes. One recent discovery in the discipline of seismology has related the cross-correlation of ambient noise fields at two recorded locations to the wave propagation response, or the so-termed *Green's function* from one location to the other; see e.g., [15], [16]. The cross-correlation approach has been successfully validated by experimental data for a variety of seismic waves. This technique has been suggested in [17] for the study of power system EM waves. Nonetheless, this earlier work only pointed out the potential application based on the connection between the EM and seismic waves, and did not establish the analytical validity for application to inter-connected power system. To the best of our knowledge, none of the earlier work has analytically investigated the problem of inferring the spatial grid frequency response from ambient synchrophasor data.

The contributions of the present paper lie in characterizing the analytical conditions and developing the practical implementations of the cross-correlation approach under the power system dynamic model. The linear swing dynamics based on

the classical generator model is considered, where the ambient condition is given by modeling the input power perturbation as a white-noise process. Based the system and data models, we have established a set of conditions under which the cross-correlation of ambient speed responses is equivalent to the input-output system response between the two generators, up to a scaling difference. Specifically, the equivalence result requires the oscillation modes to be uniformly damped, as well as to be uncorrelated and equally excited by the noise inputs. At a high level, the analytical conditions developed here can be thought as the generalization of the homogeneous grid model in continuum medium of [14] to the network of discrete generator nodes. By allowing the modes to be of different oscillation frequencies, the resultant model better represents realistic power systems than the planar wave one. While the modes of actual grids are not necessarily uniformly damped, our analysis suggests that these conditions can be relaxed such that only the main oscillation modes are sufficiently decoupled and uncorrelated. We suspect the latter would hold well for large-scale interconnections as validated by the real data collected for the Eastern Interconnection (EI) grid. Meanwhile, our synthetic data results generated using the Western System Coordinating Council (WSCC) 3-generator test case have also corroborated the validity of the cross-correlation approach with non-uniformly damped modes or even higher-order generator model.

The rest of this paper is organized as follows: Sec. II introduces the modeling of system dynamics and ambient data. Sec. III establishes the main equivalence results between the frequency response and the output cross-correlation, and presents the detailed algorithm for practical implementation as well as potential practical applications. Numerical validation results are presented in Sec. IV, using both synthetic data from WSCC 3-generator case and real frequency data of the EI grid. The paper is concluded in Sec. V.

Notation: Boldface letters denote column vectors or matrices. Vectors $\mathbf{0}$, $\mathbf{e}_\ell$ denote respectively the all-zero vector, the ℓ -th canonical vector of all zeros except for the ℓ -th entry being one. $\mathbf{I}$ denotes the identity matrix of suitable dimensions. Superscript $^\top$ stands for transpose, while $\text{diag}\{\cdot\}$ for diagonal matrix. Symbols $\dot{\delta}$ and $\ddot{\delta}$ denote the first- and second-order time derivative of δ , respectively. $\mathbb{E}$ is the expectation operator.

II. SYSTEM MODELING

Dynamics of a power system can be modeled by a set of differential and algebraic equations (DAEs). Consider the following one for a general non-linear dynamical system:

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{y}, \mathbf{u}) \\ \mathbf{0} = \mathbf{g}(\mathbf{x}, \mathbf{y}), \end{cases} \quad (1)$$

with time derivative of the state vector $\mathbf{x}$ also depending on the algebraic variables in $\mathbf{y}$ and input variables in $\mathbf{u}$. In the context of power system dynamics, $\mathbf{x}$ contains the internal angle and speed of generators and status of related equipment such as exciters and governors. In addition, $\mathbf{y}$ consists of the voltage magnitudes and angles at all buses to establish the power flow equations as algebraic equations, while the input $\mathbf{u}$ typically

includes the mechanical power injected to all generators. The set of DAEs in (1) can be linearized around an operating point, i.e., a solution to the steady-state power flow equations, in order to perform the small-signal stability analysis that focus on small ambient oscillations. Furthermore, vector $\mathbf{y}$ can be eliminated from the resultant linearized model to obtain a set of ordinary differential equations (ODEs) for the state vector $\mathbf{x}$.

We adopt the classical model for synchronous generators [3, Sec. 6.6] to specify the dynamic model (1) for ambient oscillations. Under this simplified generator model, the system states involve only the angle and speed of all the synchronous generators. For a system of n generators at an operating point, one can partition $\mathbf{x}$ into the angle deviation vector $\boldsymbol{\delta} = [\delta_1, \dots, \delta_n]^\top$ and the speed deviation vector $\boldsymbol{\omega} = [\omega_1, \dots, \omega_n]^\top$, from their respective synchronous values. Meanwhile, vector $\mathbf{u} = [u_1, \dots, u_n]^\top$ stands for mechanical-electrical power imbalance at all the generators. By linearizing the model and eliminating the algebraic variables, the system dynamical model can be written as so-called *swing equation* [3, Sec. 6.6]:

$$\begin{cases} \dot{\boldsymbol{\delta}} &= \boldsymbol{\omega} \\ \mathbf{M}\dot{\boldsymbol{\omega}} &= -\mathbf{K}\boldsymbol{\delta} - \mathbf{D}\boldsymbol{\omega} + \mathbf{u} \end{cases} \quad (2)$$

where $\mathbf{M} = \text{diag}\{M_1, \dots, M_n\}$ and $\mathbf{D} = \text{diag}\{D_1, \dots, D_n\}$ are the positive diagonal matrices containing the generator moment of inertia and damping coefficients, respectively. Also, the power flow Jacobian matrix $\mathbf{K} := \frac{\partial \mathbf{P}^e}{\partial \boldsymbol{\delta}}$ is the partial derivative for the generator electrical power output $\mathbf{P}^e = [P_1^e, \dots, P_n^e]^\top$ based on the power flow equations. With all generator parameters and network information given, one can construct the linearized model (2) using numerical tools such as the Power System Toolbox [18] or Power System Analysis Toolbox (PSAT) [19].

We are interested in obtaining the frequency response from the ℓ -th input to the k -th speed, as denoted by

$$T_{\ell k}(\tau) = \omega_k(\tau) \big|_{\mathbf{u}(t) = \delta(t)\mathbf{e}_\ell} \quad (3)$$

where $\delta(t)$ is the Dirac delta function. Our goal is to estimate $T_{\ell k}(\tau)$ directly from ambient synchrophasor data without requiring any knowledge of system model. Ambient operating conditions can be typically modeled by random load variations around nominal level; see e.g., [8], [11], [12]. The mechanical power mismatch input is modeled by a white noise process $\boldsymbol{\nu}(t)$ with the following properties:

$$\begin{aligned} \mathbb{E}[\boldsymbol{\nu}(t)] &= \mathbf{0}, \\ \mathbb{E}[\boldsymbol{\nu}(t)\boldsymbol{\nu}^\top(t-\tau)] &= \boldsymbol{\Sigma}\delta(\tau), \end{aligned} \quad (4)$$

which is zero-mean and uncorrelated across time. The corresponding ambient response of generator rotor speed is denoted by $\zeta(\tau) = \boldsymbol{\omega}(\tau) \big|_{\mathbf{u}(t) = \boldsymbol{\nu}(t)}$. The cross-correlation of the pair ζ_k and ζ_ℓ is given by

$$\begin{aligned} C_{\ell k}(\tau) &:= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \zeta_k(t)\zeta_\ell(t-\tau)dt \\ &= \mathbb{E}[\zeta_k(t)\zeta_\ell(t-\tau)], \end{aligned} \quad (5)$$

where the second equality is due to the ergodicity of $\zeta(t)$.

III. MAIN RESULTS

To establish the equivalence between $C_{\ell k}$ and $T_{\ell k}$, we will first perform linear system analysis for (2). To this end, consider its second-order equivalent form

$$\mathbf{M}\ddot{\boldsymbol{\delta}} + \mathbf{D}\dot{\boldsymbol{\delta}} + \mathbf{K}\boldsymbol{\delta} = \mathbf{u}, \quad (6)$$

and a few assumptions will be introduced to efficiently diagonalize (6) into decoupled scalar systems.

(as1) Matrix $\mathbf{M}$ is positive definite (symmetric with positive eigenvalues), while $\mathbf{K}$ is a symmetric Laplacian matrix of rank $(n - 1)$.

Since each M_i is always positive, the diagonal $\mathbf{M}$ is positive definite. Recall that matrix $\mathbf{K}$ is actually the power flow Jacobian for the reduced network consisting of only generator nodes; see e.g., [3, ch. 6]. If all the transmission lines are purely inductive, $\mathbf{K}$ would be a perfectly symmetric Laplacian matrix. As long as the network is fully connected, the rank of $\mathbf{K}$ equals to $(n - 1)$. Hence, for high-voltage transmission systems with minimal resistive losses, $\mathbf{K}$ would be nearly symmetric and Laplacian and (as1) would hold. As for (as2), it has been introduced to simplify the diagonalization process for (6). Nonetheless, larger generators would typically provide higher damping to the system, (as2) would hold approximately.

Consider the following generalized eigenvalue problem for the two matrices $\mathbf{M}$ and $\mathbf{K}$:

$$\mathbf{K}\mathbf{V} = \mathbf{M}\mathbf{V}\boldsymbol{\Lambda} \quad (7)$$

where matrix $\mathbf{V} = [\mathbf{v}_1 \cdots \mathbf{v}_n]$ consists of the generalized eigenvectors, while $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_n\}$ has the generalized eigenvalues. The following lemma can be established for $\mathbf{V}$ and $\boldsymbol{\Lambda}$; see e.g., [20, Ch. 15].

Lemma 1. *Under (as1), the generalized eigenvalues are all non-negative such that $\lambda_n \geq \dots \geq \lambda_2 > 0 = \lambda_1$. Moreover, the generalized eigenvectors are $\mathbf{M}$ -orthogonal; i.e.,*

$$\mathbf{V}^T \mathbf{M} \mathbf{V} = \mathbf{I}. \quad (8)$$

Clearly, matrix $\mathbf{V}$ would simultaneously diagonalize $\mathbf{M}$ and $\mathbf{K}$. To decouple (6) into scalar systems, the following assumption is needed for $\mathbf{D}$:

(as2) The damping matrix $\mathbf{D}$ can be diagonalized by $\mathbf{V}$; that is, $\mathbf{V}^T \mathbf{D} \mathbf{V} = \boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma} = \text{diag}\{\gamma_1, \dots, \gamma_n\}$.

Since $\mathbf{M}$ and $\mathbf{D}$ are both diagonal, under Lemma 1 the condition in (as2) is equivalent to having $\gamma_i = \gamma$ for any $i = 1, \dots, n$. Thus, (as2) boils down to a uniform damping condition among all generators. Under (as1) and (as2), one can decouple the dynamic system (6) using the transformation

$$\mathbf{z} = \mathbf{V}^{-1} \boldsymbol{\delta}. \quad (9)$$

Substituting (9) into (6) and using Lemma 1 and (as2) gives rise to

$$\ddot{\mathbf{z}} + \boldsymbol{\Gamma} \dot{\mathbf{z}} + \boldsymbol{\Lambda} \mathbf{z} = \mathbf{V}^T \mathbf{u} \quad (10)$$

where each element z_i of $\mathbf{z}$ is an oscillation mode. The second-order dynamics is decoupled into each mode as

$$\ddot{z}_i + \gamma_i \dot{z}_i + \lambda_i z_i = \mathbf{v}_i^T \mathbf{u}. \quad (11)$$

Under zero initial condition at $\tau = -\infty$, the trajectory $z_i(\tau)$ can be solved analytically. By converting (11) into a first-order vector system, standard technique from linear system theory is readily applicable using the eigen-decomposition of the resultant state matrix. As generator speed is directly related to the grid electrical frequency, we focus on time derivative of each mode, as given by

$$\dot{z}_i(\tau) = \int_0^\infty (a_i e^{c_i t} + b_i e^{d_i t}) \mathbf{v}_i^T \mathbf{u}(\tau - t) dt \quad (12)$$

where the following four parameters arise from the modal analysis:

$$\begin{aligned} a_i &= \frac{2\lambda_i}{\sqrt{\gamma_i^2 - 4\lambda_i}(-\gamma_i - \sqrt{\gamma_i^2 - 4\lambda_i})}, \\ b_i &= \frac{-2\lambda_i}{\sqrt{\gamma_i^2 - 4\lambda_i}(-\gamma_i + \sqrt{\gamma_i^2 - 4\lambda_i})}, \\ c_i &= \frac{-\gamma_i + \sqrt{\gamma_i^2 - 4\lambda_i}}{2}, \\ d_i &= \frac{-\gamma_i - \sqrt{\gamma_i^2 - 4\lambda_i}}{2}. \end{aligned}$$

Using the inverse transformation of (9), one can obtain

$$\begin{aligned} \omega_k(\tau) &= \sum_{i=1}^n v_{ki} \dot{z}_i(\tau) \\ &= \sum_{i=1}^n v_{ki} \int_0^\infty (a_i e^{c_i t} + b_i e^{d_i t}) \mathbf{v}_i^T \mathbf{u}(\tau - t) dt. \end{aligned} \quad (13)$$

where v_{ki} is the k -th element of vector $\mathbf{v}_i$. Thus, by setting the input $\mathbf{u}(t) = \delta(t) \mathbf{e}_\ell$, we find the frequency response

$$T_{\ell k}(\tau) = \sum_{i=1}^n v_{ki} v_{\ell i} (a_i e^{c_i \tau} + b_i e^{d_i \tau}). \quad (14)$$

This completes the analysis of system response for (2). Clearly, assumptions (as1)-(as2) are key for achieving the simplified second-order oscillation mode transformation, instead of a general eigen-analysis approach for the full system (2).

To proceed with the equivalence results, we need to further specify the ambient noise conditions as follows:

(as3) The input noise covariance is proportional to the damping matrix $\boldsymbol{\Sigma} = \alpha \mathbf{D}$ with α to be a positive constant.

This assumption would ensure that all mode z_i -s are equally excited. As (as2) leads to $\mathbf{D} = \gamma \mathbf{M}$, the proportional noise condition in (as3) is equivalent to $\boldsymbol{\Sigma} = \mu \mathbf{M}$ with $\alpha = \mu/\gamma$. Hence, this assumes that the noise variance is proportional to the locational inertia. This is justified as higher level of inertia would have been allocated to areas with higher level of disturbance through power system planning studies.

Proposition 1. *For the system (2), under assumptions (as1)-(as3), the cross-correlation between ambient responses ζ_k and ζ_ℓ is equivalent to the corresponding frequency response from u_ℓ to ω_k , up to a scaling difference, as given by*

$$C_{\ell k}(\tau) = \frac{\alpha}{2} T_{\ell k}(\tau).$$

Proof: Based on the analytical result in (13), one can write the ambient response $\zeta(t)$ with the noise input given by $\boldsymbol{\nu}$. Thus, the cross-correlation defined in (5) contains the components between any pair of modes, as given by

$$\begin{aligned} C_{\ell k}(\tau) &= \sum_{i=1}^n \sum_{j=1}^n v_{ki} v_{\ell j} \int_0^\infty dt_1 \int_0^\infty dt_2 (a_i e^{c_i t_1} + b_i e^{d_i t_1}) \\ &\quad (a_j e^{c_j t_2} + b_j e^{d_j t_2}) \mathbf{v}_i^T \boldsymbol{\Sigma} \mathbf{v}_j \delta(\tau + t_2 - t_1) \end{aligned} \quad (15)$$

where the uncorrelated property in (4) is used. Under (as3), the coefficient $\mathbf{v}_i^T \boldsymbol{\Sigma} \mathbf{v}_j = \alpha \mathbf{v}_i^T \mathbf{D} \mathbf{v}_j$, which is zero for any pair of modes with $i \neq j$ according to (as2). Thus, this leads to uncorrelated responses between any two different modes. Accordingly, the ambient cross-correlation response would only involve the auto-correlation components of every mode, as

$$C_{\ell k}(\tau) = \alpha \sum_{i=1}^n \gamma_i v_{ki} v_{\ell i} \int_0^\infty dt_1 \int_0^\infty dt_2 (a_i e^{c_i t_1} + b_i e^{d_i t_1}) (a_i e^{c_i t_2} + b_i e^{d_i t_2}) \delta(\tau + t_2 - t_1). \quad (16)$$

For stable systems, the integral evaluated at time ∞ is zero, giving rise to

$$C_{\ell k}(\tau) = \alpha \sum_{i=1}^n \gamma_i v_{ki} v_{\ell i} \left[-a_i e^{c_i \tau} \left(\frac{a_i}{2c_i} + \frac{b_i}{c_i + d_i} \right) - b_i e^{d_i \tau} \left(\frac{b_i}{2d_i} + \frac{a_i}{c_i + d_i} \right) \right]. \quad (17)$$

Through algebraic calculations, we find that for every mode i the four parameters satisfy

$$\frac{a_i}{2c_i} + \frac{b_i}{c_i + d_i} = -\frac{1}{2\gamma_i}, \text{ and } \frac{b_i}{2d_i} + \frac{a_i}{c_i + d_i} = -\frac{1}{2\gamma_i}. \quad (18)$$

Substituting (18) into (17) leads to

$$C_{\ell k}(\tau) = \frac{\alpha}{2} \sum_{i=1}^n v_{ki} v_{\ell i} (a_i e^{c_i \tau} + b_i e^{d_i \tau}), \quad (19)$$

which completes the proof for Proposition 1. ■

Remark 1: (*Fluctuation-Dissipation Theorem.*) There exist an interesting connection between our analysis and that of thermally driven equilibrium systems in statistical physics, known at the Fluctuation-Dissipation Theorem (FDT); see e.g. [21], [22]. The latter is based on a continuum medium model with the classical derivation on one-dimensional field in [22]. The scaling between $\mathbf{D}$ and $\boldsymbol{\Sigma}$ in (as3) is exactly related to the proportional relation between dissipation and noise covariance. Similar to FDT, the analytical conditions in (as2) could hold for more general damping scenarios. Nonetheless, the diagonal structure of both $\mathbf{M}$ and $\mathbf{D}$ under the power system dynamics leads to equivalently uniform damping condition and accordingly proportional ratio between inertia and noise level. Therefore, our analysis is connected to the general FDT conditions, while tackling the networked structure and dynamics in power grids.

A. The Proposed Algorithm

Proposition 1 has established the theoretical relation between the cross-correlation of ambient responses and the frequency response of interest. For practical implementations, one needs to select the measurement locations and estimate the frequency deviations. Typically PMUs are installed at almost all substations that generators are connected to, or at least one for each control area. To estimate the frequency response between any two generators in an interconnected system, we propose an algorithm, as summarized in Fig. 1, of the following four steps:

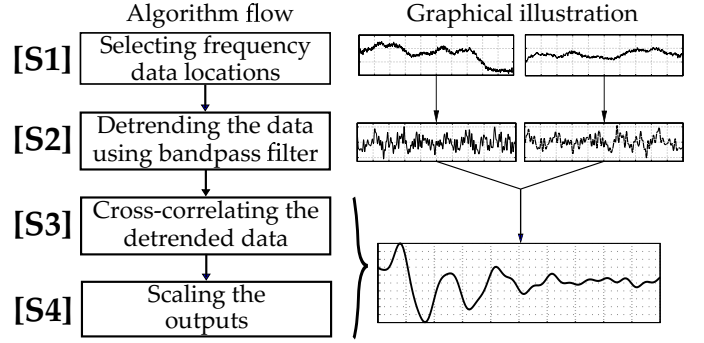


Fig. 1: The proposed 4-step algorithm to recover the frequency response between from ambient synchrophasor data.

[S1] For each generator, select the frequency measurement from the closest PMU in electrical distance (e.g., [23]). This could be at the substation that the generation is directly connected to, or a neighboring substation connected through very short transmission line.

[S2] Bandpass filtering is used to detrend the raw frequency data for the ambient response. As inter-area oscillation modes are of higher interest, the passband can be selected accordingly as [0.1, 0.7] Hz; see e.g., [4, ch.12]. This way, the slowly-varying mean component (close to 0 Hz) is filtered out, ensuring a zero-mean ambient response.

[S3] With the detrended outputs at sampling period Δ , one can compute the discrete-time counterpart of (5) to estimate the cross-correlation as

$$C_{\ell k}[\tau] = \frac{1}{\lfloor T/\Delta \rfloor} \sum_{m=1}^{\lfloor T/\Delta \rfloor} \zeta_k[m] \zeta_\ell[m - \tau]$$

where $\lfloor T/\Delta \rfloor$ is the rounded number of samples.

[S4] To recover $T_{k\ell}$, the estimated cross-correlation will be scaled to match the system frequency nadir point. This scaling step will be discussed in detail soon.

B. Practical Applications

Our proposed algorithm can be directly used to recover the spatial frequency response from any location to another in an interconnected grid. To this end, [S4] needs to resolve the magnitude ambiguity of the cross-correlation as in Proposition 1. The scaling factor can be obtained with the knowledge of frequency nadir value under a specific disturbance event through off-line transient stability studies. For example, the ratio between frequency nadir and power imbalance is available for the major interconnections [24]. Upon recovering the magnitude, the proposed algorithm can provide a fully measurement-based alternative to find the generator frequency response to any input disturbance.

Even if the magnitude information is unavailable, the proposed algorithm can still be used for applications that need the timing of EM wave propagation. As demonstrated through numerical tests, the cross-correlation output can accurately maintain the shape and thus the peak time of generator frequency response. This information is indispensable for locating the source of disturbance in a wide-area system; see e.g., [25].

Challenged by the same issues with model-based methods, the approach in [25] adopts the nominal propagation speed given by [13], [14]. Nonetheless, the latter was developed based on the homogeneous continuum model and would not be able to capture the characteristics of each individual system. Our data-driven approach can provide improved estimates of the propagation time to better locate the disturbance.

IV. CASE STUDIES

This section validates the proposed algorithm using synthetic data generated by the WSCC test case and real synchrophasor data collected at the Eastern Interconnection (EI) system. The synthetic data tests demonstrate the validity of our approach even if (as1)-(as3) fail to hold in practice, including non-symmetric $\mathbf{K}$, non-uniform damped modes, and higher-order generator dynamic model. Real data tests can verify the proposed method's capability in estimating the propagation time of frequency disturbance across the EI system. Since the scaling factor in [S4] requires the knowledge of actual system frequency nadir, the validations will not account for this step. Instead, the following normalized dissimilarity score is used:

$$S_{\ell k} = \frac{\|\bar{T}_{\ell k} - \bar{C}_{\ell k}\|_2}{\|\bar{T}_{\ell k}\|_2} \quad (20)$$

where $\bar{T}_{\ell k}$ ($\bar{C}_{\ell k}$) stands for the model-based (estimated) response normalized by its maximum absolute value, while $\|\cdot\|_2$ stands for the L^2 norm of a function. The normalized metric is effective for comparing the shape of the two time series. An almost zero $S_{\ell k}$ implies that the cross-correlation can accurately recover a scaled version of $T_{\ell k}$, while a higher value of $S_{\ell k}$ indicates a larger mismatch.

A. WSCC Case with Classical Generator Model

We use the WSCC 3-generator test case under the classical generator model with the parameters given by [19]. A one-line diagram of this system is shown in Fig. 2. Generators at buses 1, 2 and 3 are indexed by the associated bus number. First

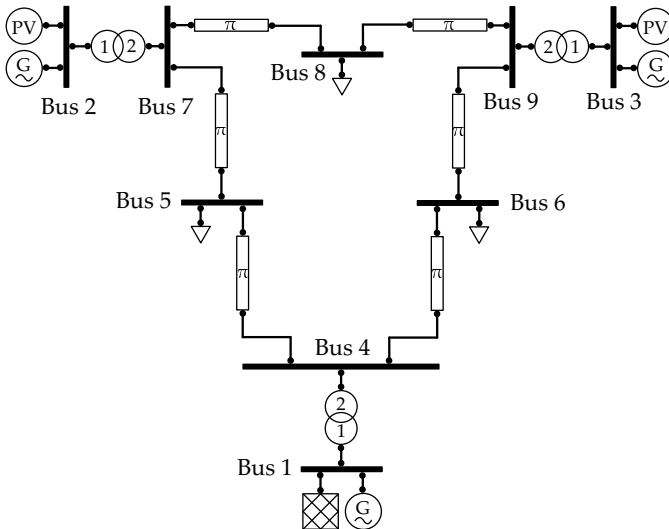


Fig. 2: One-line diagram of the WSCC test case [19].

TABLE I: Dissimilar scores for the WSCC test case under uniform damping condition

Linearized model				Time-domain simulation			
	u_1	u_2	u_3		u_1	u_2	u_3
ω_1	0.035	0.059	0.039	ω_1	0.113	0.259	0.183
ω_2	0.053	0.036	0.030	ω_2	0.249	0.332	0.226
ω_3	0.047	0.033	0.017	ω_3	0.180	0.229	0.158

TABLE II: Dissimilar scores for the WSCC test case under non-uniform damping condition

	u_1	u_2	u_3
ω_1	0.095	0.182	0.191
ω_2	0.177	0.172	0.311
ω_3	0.197	0.315	0.208

consider a uniformly-damped system satisfying (as2) with $\gamma = 0.2$; i.e., $D_i = 0.2M_i$ for $i = 1, 2, 3$. Synthetic ambient frequency data are generated using two methods: i) linearized model based on the system (2), and ii) time-domain simulation for the non-linear DAEs in (1). All the data computations are performed in the Matlab[®] environment.

For the linearized model method, matrices $\mathbf{M}$, $\mathbf{K}$ and $\mathbf{D}$ in (2) are first computed by the PSAT software [19]. Accordingly, the system state-space form can be easily constructed while ambient frequency deviation samples can be formed by simulating the system (using the *lsim* command) in response to the random input $\mathbf{u}$ generated by the *randn* function. Note that the resultant $\mathbf{K}$ is not perfectly symmetric as asserted in (as1) because the transmission lines are not purely inductive. Nonetheless, the line resistance-to-reactance (R/X) ratio is very small almost everywhere, and thus $\mathbf{K}$ is nearly symmetric. The highest R/X ratio has been observed to be around 20% for the lines connecting buses (5,7), and (4,6). The R/X ratio for the other lines is around 12%. As both lines (5,7) and (4,6) are connecting the generator at bus 1 to other buses, it has been observed that the asymmetry of $\mathbf{K}$ is slightly more evident at entries corresponding to bus 1.

Despite of an asymmetric $\mathbf{K}$ violating (as1), we observe that the estimated $\bar{C}_{\ell k}$ matches very well with the model-based result. Based on the linearized model, the actual frequency response can be calculated using the *impulse* command, while the discrete-time $\bar{C}_{\ell k}$ is obtained based on (5) of [S3]. Fig. 3 plots the comparisons between the model-based frequency response (in solid red lines) and the estimated one (in dashed blue lines), both normalized by the respective maximum absolute value, for each input-output pair. It has illustrated the high accuracy of the proposed algorithm in recovering the actual time series. This observation has also been corroborated by Table I, which lists the corresponding dissimilarity scores as defined by (20). The score values are all very close to zero, with slightly higher values related to the generator at bus 1; e.g., between the pairs (u_1, ω_2) , (u_1, ω_3) , (u_2, ω_1) and (u_3, ω_1) . These relatively higher mismatches could be related to the earlier observation that the asymmetry of $\mathbf{K}$ is more significantly affected at entries corresponding to generator 1.

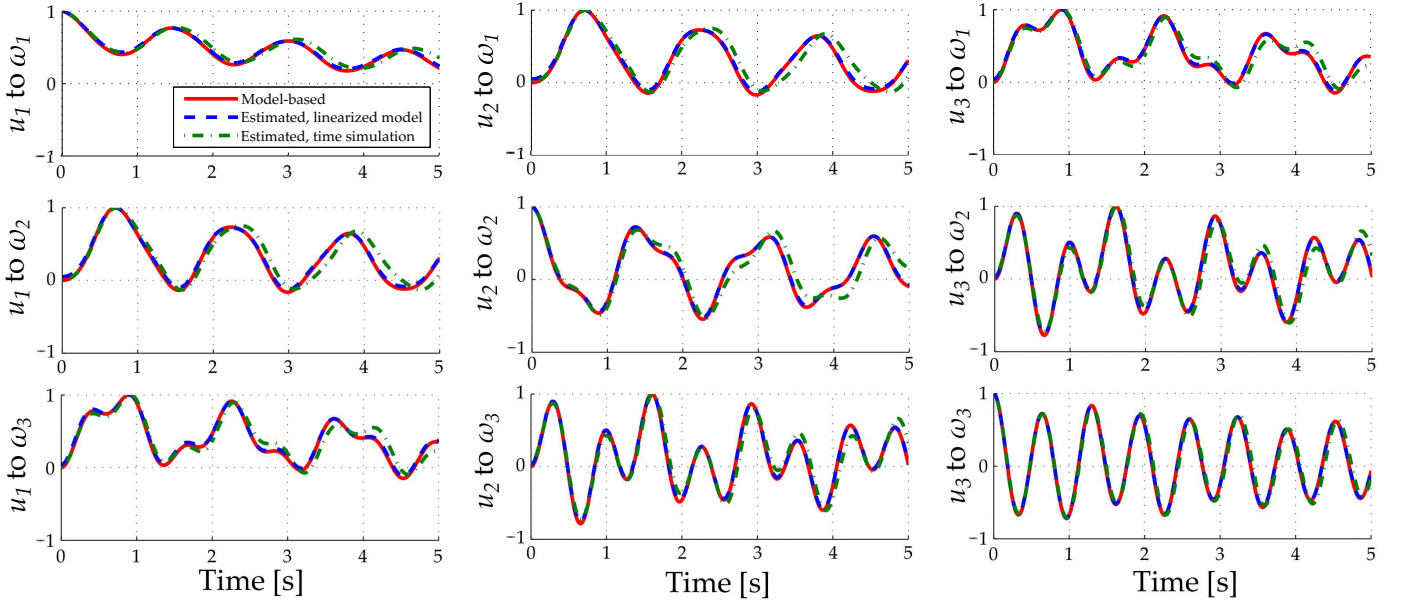


Fig. 3: Comparison of model-based and estimated frequency response for the WSCC test case with classical generator model and under uniform damping condition.

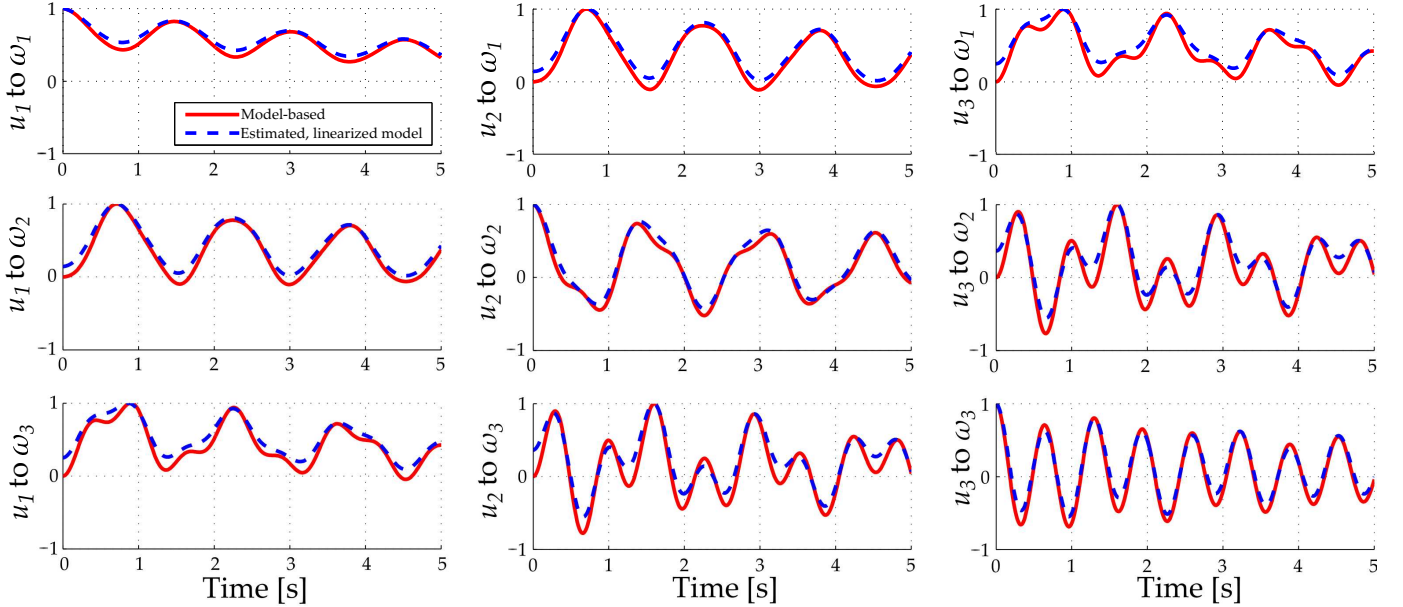


Fig. 4: Comparison of model-based and estimated frequency response for the WSCC test case with classical generator model and under non-uniform damping condition.

B. WSCC Case with Higher-order Generator Model

Moreover, we have used time-domain simulation based on the original non-linear model (1) to generate ambient frequency samples. This data generation method can more accurately represent real system responses by accounting for the non-linearity of power flow equations. However, it is also much more time consuming compared to the earlier method because of the numerical integration computations. We have slightly modified the integration module routine *pert.m* in PSAT [19] to incorporate white-noise perturbation as the generator power input. The noise covariance matrix

is set to be $\Sigma = 0.001\mathbf{M}$. Note that this method does not produce the generator speed deviation in the synchronous reference frame, but instead the actual values at around 60Hz. Hence, the output speed data need to be detrended to obtain the ambient deviation from nominal value. Subtracting the output from its time-averaged value serves this purpose well, because the system is slightly perturbed at a single operating point. By cross-correlating the detrended outputs, we plot the estimated frequency response in Fig. 3 using dash-dot green lines, along with the updated dissimilar scores included in Table I. In general, the effectiveness of the cross-correlation approach has been validated by the time-domain simulation

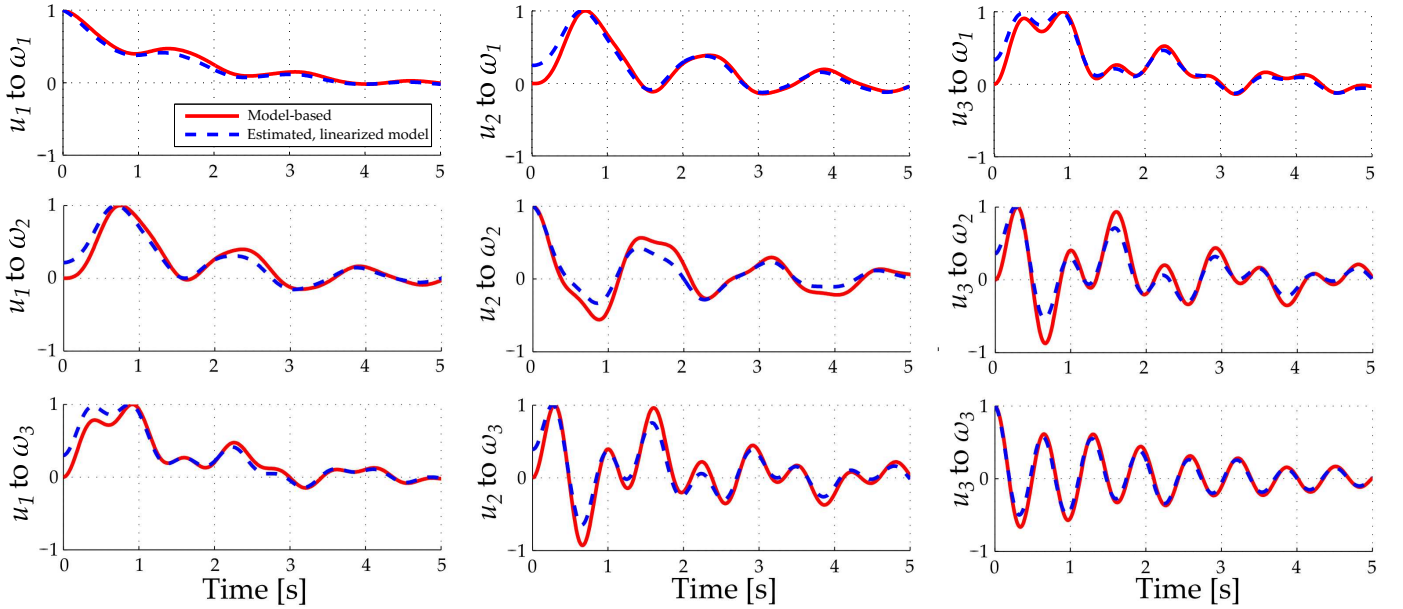


Fig. 5: Comparison of model-based and estimated frequency response for the WSCC test case with sixth-order generator model and under uniform damping condition.

output, with slightly increased dissimilarity scores compared to those obtained by the linearized model outputs. But still, the estimated frequency response can well approximate the model-based one, especially at the first few seconds when higher level of amplitude would occur as shown by Fig. 3. Similar to the validation results using linearized model outputs, relatively higher level of dissimilarity has been observed for the input-output pairs associated with generator 1, particularly for the pairs (u_1, ω_2) and (u_2, ω_1) . This observation can be again attributed to the transmission line (5,7) at high R/X ratio. Since the time-domain simulation method has shown similar estimation/validation results as the linearized model one, we would only use the latter method which allows to very quickly generate ambient speed data.

We further test the applicability of the proposed approach by relaxing the uniform damping assumption in (as2). All the models and parameters stay the same, except that the damping coefficient for each generator is changed as follows: $D_1 = 0.1M_1$, $D_2 = 0.2M_2$ and $D_3 = 0.3M_3$; see [3, pp. 165]. Under this setting, the damping coefficient for each oscillation mode is 0.279, 0.186, or 0.269, respectively, compared to the 0.2 value for all modes in the uniformly damped case. Fig. 4 updates the comparison of frequency response, while the corresponding dissimilarity scores are listed in Table II. Compared to the uniformly damped results using linearized model outputs, the dissimilarity score has increased, especially for the pairs (u_2, ω_3) and (ω_3, u_2) . A closer look at the corresponding curves in Fig. 4 shows a scaling difference in peak values. Under non-uniform damping, the M-orthogonality condition in (8) no longer holds, and thus the cross-correlation in (15) would involve inter-mode components and the coefficients would not match with the model-based analysis. Accordingly, the peak scaling cannot be accurately recovered. Nonetheless, Fig. 4 still shows that the peak locations are well aligned

between the estimated and actual ones. This observation sheds light on the applicability of the proposed approach in estimating disturbance propagation time under non-uniform damping scenarios.

To make the validation more realistic, we modify the WSCC case using the sixth-order generator model that includes controllers like exciter, governor, and power system stabilizer. This modified system is tested under similar settings as in Sec. IV-A. The updated plots for uniform and non-uniform damping conditions are shown in Figs. 5 and 6, respectively. The mismatch of the cross-correlation estimates has become further noticeable, especially in the peak scaling. This is again due to the violation of the ideal (as1)-(as2), in particular the orthogonality condition. It is also observed that the modes are better damped with the additional controllers. Interestingly, the estimated response has correctly reflected this improvement in damping as the match in the curvature shape is very perfect. As a result, the estimated response has an accurate peak timing alignment, which confirms the applicability of the proposed approach in estimating disturbance propagation time from real power system frequency measurements.

C. Real Data Test

We apply the cross-correlation method to real ambient frequency measurements at the Eastern Interconnection (EI) system to estimate its frequency response behavior. The data was collected by the frequency monitoring network (FNET) devices for 15min on January 20, 2017, at a 10Hz sampling rate. Measurement locations include places in the Midwest, South, and Northeast regions. The proposed cross-correlation result is compared with the recorded response during the 2008 Florida blackout. This event is among the most recent wide-area disturbances affecting an interconnected system, and thus

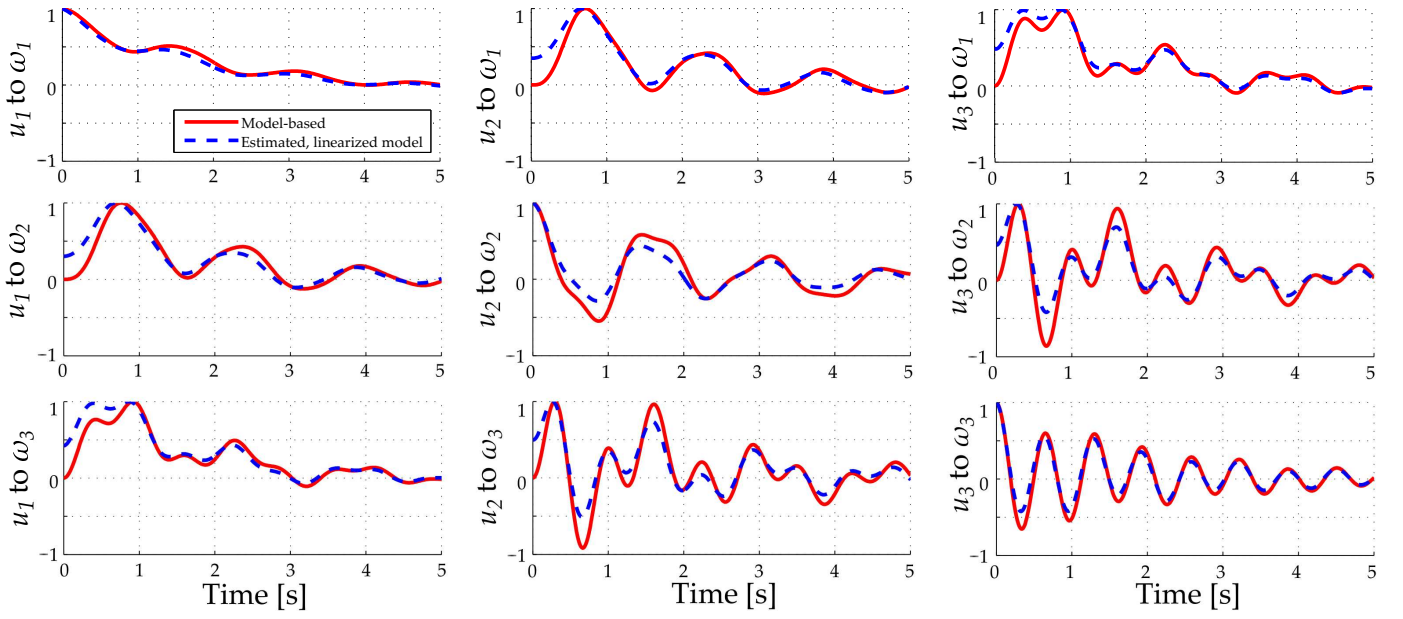


Fig. 6: Comparison of model-based and estimated frequency response for the WSCC test case with sixth-order generator model and under non-uniform damping condition.

can serve as a good reference for validating the effectiveness of our proposed data-driven approach.

The 2008 Florida blackout was caused by a load switching disturbance propagated across the whole EI footprint in the afternoon of February 26, 2008 [26]. This event can be approximated by an impulse input to a Florida substation at 18:09:8.2 UTC. Replay of the frequency response and its contour map using the recorded FNET data is available on Youtube¹. We have recovered the propagation time by observing how long after the impulse start time the frequency overshoot with the highest red color has arrived at each location equipped with FNET devices. For example, in 1.7 seconds, the frequency overshoot reached an intersected area among Arkansas, Tennessee, Kentucky and Missouri. In 2.7 seconds, the EM wave front reached the regions of North Dakota, Upstate New York and Massachusetts. Note that these recorded values may not be the ground-truth because of the observation error in the recovery process.

A sample of the ambient frequency data with its detrended output at a selected location is shown in Fig. 7. The bandpass filter in [S2] was implemented using the function *fir1* of Matlab[®] at an order of 300 samples with the lower and upper cutoff frequency at 0.075Hz and 1Hz, respectively. This high filter order can yield very low filter gain at low frequency regime and thus efficiently detrend the input data.

We cross-correlate the detrended frequency data at the Florida location with every other location to determine the corresponding frequency response. Fig. 8 plots selected response from the FNET meter 623, at Florida, to meters 726, 781 and 823, at North Dakota, Missouri and Arkansas, respectively. The peak of each response is located at 2.4, 2.0 and 1.6 seconds, respectively. This is used to estimate the arrival time of EM oscillations and compared with the recorded values,

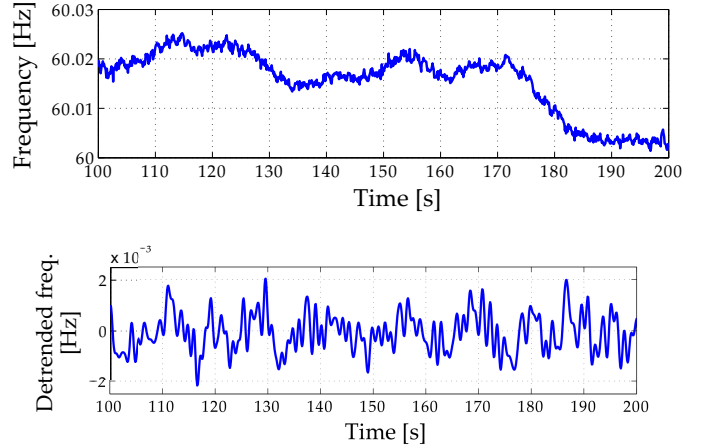


Fig. 7: A sample of the raw ambient frequency measurement (top) and its corresponding detrended component (down).

as listed in Table III for all locations. The recorded value is obtained from the replay video of this event. For the aforementioned three locations, the recorded value is 2.5, 2.1, and 1.5 seconds, respectively. Clearly, the cross-correlation approach has successfully estimated the propagation time from the source of disturbance to other location, with the average error around 0.2 second.

Various factors can attribute to the mismatch error between the recorded and estimated propagation time. First, the recorded values themselves may be different from the ground-truth propagation time. They have been recovered from the Youtube video showing frequency contour map and would suffer from some observation error. Second, the location of disturbance input is slightly different between the 2008 blackout and the 2017 ambient data. Due to limited deployment of FNET devices, the only frequency measurement available in

¹<https://youtu.be/awvS4TtN77E>

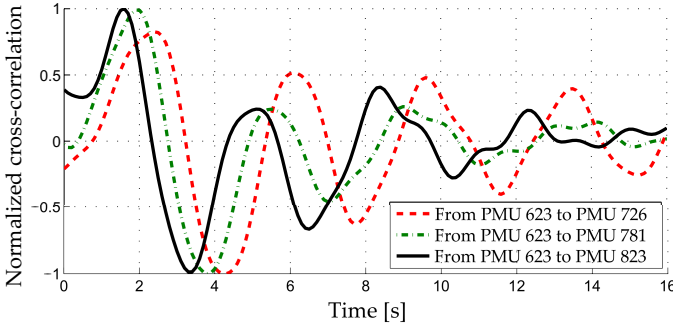


Fig. 8: Cross-correlation of ambient frequency deviation for the pairs 623–726 (Florida – North Dakota), 623–781 (Florida – Missouri) and 623–823 (Florida – Arkansas).

TABLE III: Recorded versus estimated propagation time

Location	Meter	Recorded	Estimated
Virginia	601	1.5	1.1
Alabama	671	0.7	0.4
Massachusetts	682	2.6	2.3
North Dakota	726	2.5	2.4
Missouri	756	1.6	1.9
Illinois	767	1.5	1.7
Missouri	781	2.1	2.0
Tennessee	787	1.6	1.5
Arkansas	823	1.5	1.6

Florida is located at Lakeland, about 250 miles north of the actual source of 2008 blackout, Turkey Point. Third, there is a fundamental limit in the estimated time resolution of 0.1 second as the ambient FNET data has a sampling frequency at 10 Hz. Last but not least, the blackout happened in 2008 while the ambient data was collected in 2017. Over the past decade, significant changes have taken place in the EI grid. Nonetheless, extensive real data based studies in [27] have demonstrated that the statistics of inter-area oscillation modes and EM waves has stayed almost the same for the EI grid. Hence, we believe that the 2008 Florida blackout event can still serve as a good reference for validating our real data test. Meanwhile, the mismatch error at 0.2 second is consistent with the time resolution of 0.1 second, which has verified the effectiveness of our proposed approach.

V. CONCLUDING REMARKS

This paper has presented the analytical conditions and numerical tests for a data-driven approach to estimate the dynamic frequency response from ambient synchrophasor data. To recovery the effects of EM oscillations propagated from any location to any other, it is advocated to cross-correlate the ambient frequency responses measured at the two locations during normal grid operations. We have explored a set of exact recovery conditions for the classical swing dynamics, including lossless systems and equally excited oscillation modes. Extensive numerical tests on the WSCC 3-generator case have demonstrated the cross-correlation technique can approximately recover the shape of frequency response for uniform damping conditions or even under higher-order generator dynamics. Moreover, the cross-correlation results of the EI grid frequency data have been shown to well match

with the actual disturbance propagation time during the 2008 Florida blackout, with the mismatch error consistent with the sampling resolution. Hence, it is true that the assumptions for exact recovery do not hold perfectly for all real power systems, nonetheless, the cross-correlation based algorithm is a powerful model-free tool to investigate the EM wave propagation effects in a wide-area interconnected grid.

Currently, we are exploring the possibility of relaxing the analytical conditions by going beyond pair-wise correlation to high-dimensional structure, as well as performing more real data based validations. Future research directions also include the development of data-driven instability monitoring or system design tools using ambient synchrophasor data.

REFERENCES

- [1] G. Rogers, *Power system oscillations*. Springer Science & Business Media, 2012.
- [2] C. W. Taylor, “Improving grid behaviour,” *IEEE Spectrum*, vol. 36, no. 6, pp. 40–45, Jun 1999.
- [3] P. W. Sauer and M. A. Pai, *Power System Dynamics and Stability*. Stipes Publishing L.L.C., 1997.
- [4] P. Kundur, N. J. Balu, and M. G. Lauby, *Power system stability and control*. McGraw-hill New York, 1994.
- [5] J. W. Pierre, D. J. Trudnowski, and M. K. Donnelly, “Initial results in electromechanical mode identification from ambient data,” *IEEE Trans. Power Syst.*, vol. 12, no. 3, pp. 1245–1251, Aug 1997.
- [6] N. Zhou, D. J. Trudnowski, J. W. Pierre, and W. A. Mittelstadt, “Electromechanical mode online estimation using regularized robust RLS methods,” *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1670–1680, Nov 2008.
- [7] N. Zhou, Z. Huang, L. Dosiek, D. Trudnowski, and J. W. Pierre, “Electromechanical mode shape estimation based on transfer function identification using PMU measurements,” in *Proc. IEEE PES General Meeting*, July 2009, pp. 1–7.
- [8] J. Ning, S. A. N. Sarmadi, and V. Venkatasubramanian, “Two-level ambient oscillation modal estimation from synchrophasor measurements,” *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 2913–2922, 2015.
- [9] T. Wu, V. M. Venkatasubramanian, and A. Pothan, “Fast parallel stochastic subspace algorithms for large-scale ambient oscillation monitoring,” *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1494–1503, 2017.
- [10] Y. C. Chen, J. Wang, A. D. Domínguez-García, and P. W. Sauer, “Measurement-based estimation of the power flow Jacobian matrix,” *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2507–2515, 2016.
- [11] X. Wang, J. Bialek, and K. Turitsyn, “PMU-based estimation of dynamic state Jacobian matrix and dynamic system state matrix in ambient conditions,” *IEEE Trans. Power Syst.*, 2017, (IEEE early access).
- [12] X. Wang and K. Turitsyn, “Data-driven diagnostics of mechanism and source of sustained oscillations,” *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 4036–4046, Sept 2016.
- [13] J. S. Thorp, C. E. Seyler, and A. G. Phadke, “Electromechanical wave propagation in large electric power systems,” *IEEE Trans. Circuits Syst. I*, vol. 45, no. 6, pp. 614–622, Jun 1998.
- [14] M. Parashar, J. S. Thorp, and C. E. Seyler, “Continuum modeling of electromechanical dynamics in large-scale power systems,” *IEEE Trans. Circuits Syst. I*, vol. 51, no. 9, pp. 1848–1858, Sept 2004.
- [15] O. I. Lobkis and R. L. Weaver, “On the emergence of the Green’s function in the correlations of a diffuse field,” *Journal of Acoustical Society of America*, vol. 110, no. 6, pp. 3011–3017, 2001.
- [16] R. Snieder, K. Wapenaar, and U. Wegler, “Unified Green’s function retrieval by cross-correlation,” *Physical Review E*, vol. 75, no. 3, p. 036103, 2007.
- [17] S. Backhaus and Y. Liu, “Electromechanical wave green’s function estimation from ambient electrical grid frequency noise,” in *Proc. 45th Hawaii Intl. Conf. System Sciences*, Jan 2012, pp. 2054–2061.
- [18] J. H. Chow and K. W. Cheung, “A toolbox for power system dynamics and control engineering education and research,” *IEEE Trans. Power Syst.*, vol. 7, no. 4, pp. 1559–1564, Nov 1992.
- [19] F. Milano, “An open source power system analysis toolbox,” *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1199–1206, Aug 2005.
- [20] B. N. Parlett, *The symmetric eigenvalue problem*. SIAM, 1998.

- [21] H. B. Callen and T. A. Welton, "Irreversibility and generalized noise," *Phys. Rev.*, vol. 83, pp. 34–40, Jul 1951. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRev.83.34>
- [22] R. Kubo, "The fluctuation-dissipation theorem," *Reports on Progress in Physics*, 1966.
- [23] E. Cotilla-Sanchez, P. D. Hines, C. Barrows, S. Blumsack, and M. Patel, "Multi-attribute partitioning of power networks based on electrical distance," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4979–4987, 2013.
- [24] "2016 frequency response annual analysis," North American Electric Reliability Corporation (NERC), Tech. Rep., 2016.
- [25] A. Esmailian and M. Kezunovic, "Fault location using sparse synchrophasor measurement of electromechanical-wave oscillations," *IEEE Transactions on Power Delivery*, vol. 31, no. 4, pp. 1787–1796, Aug 2016.
- [26] Y. Liu, L. Zhan, Y. Zhang, P. N. Markham, D. Zhou, J. Guo, Y. Lei, G. Kou, W. Yao, J. Chai, and Y. Liu, "Wide-area-measurement system development at the distribution level: An fnet/grideye example," *IEEE Transactions on Power Delivery*, vol. 31, no. 2, pp. 721–731, April 2016.
- [27] Y. Cui, L. Wu, W. Yu, Y. Liu, W. Yao, D. Zhou, and Y. Liu, "Inter-area oscillation statistical analysis of the u.s. eastern interconnection," *The Journal of Engineering*, vol. 2017, pp. 595–605(10), November 2017. [Online]. Available: <http://digital-library.theiet.org/content/journals/10.1049/joe.2017.0243>



Ahmed E. Elbanna Ahmed E. Elbanna holds a Ph.D. in civil engineering (2011) and an M.S. in applied mechanics (2006), both from the California Institute of Technology, and an M.S. in structural engineering (2005) and B.S. in civil engineering (2003) from Cairo University. He joined the faculty of Civil and Environmental Engineering at University of Illinois in 2013. He is a recipient of the National Science Foundation Faculty Early CAREER award (2018). His honors include the National Center for Supercomputing Applications Faculty Fellowship (2015), the George Housner Fellowship, California Institute of Technology, 2005, and the Certificate of Excellence, National Ceremony of Science, Egypt, 2004. Elbanna leads the Mechanics of Complex Systems laboratory at UIUC where his group conducts research on problems in theoretical and applied mechanics as arise in fields like geophysics, biomechanics and material design with special emphasis on fracture, wave propagation, and nonlinear dynamics.

BIOGRAPHIES



Phuc Huynh Phuc Huynh comes from Vietnam. He is currently a Ph.D. candidate at the University of Illinois at Urbana-Champaign. He received his B.S. from the University of Minnesota, Twin Cities in 2013 and M.Sc. from the University of Illinois at Urbana-Champaign in 2016 both in Electrical and Computer Engineering. His research interest includes power system dynamics & stability and power conversion system design.



Hao Zhu (M'12) is currently an Assistant Professor of ECE at the University of Illinois, Urbana-Champaign (UIUC). She received her B.S. from Tsinghua University, Beijing, China in 2006, and the M.Sc. and Ph.D. degrees from the University of Minnesota, Minneapolis in 2009 and 2012, respectively. She has been working as a postdoctoral research associate on power system validation with the Information Trust Institute at UIUC from 2012 to 2013. Her current research interests include power system monitoring and operations, dynamics and stability, distribution systems, and energy data analytics. Dr. Zhu is currently the IEEE Signal Processing Society's Representative on the *IEEE Trans. on Smart Grid*, and a steering committee member for *IEEE Smart Grid*.



Qianli Chen Qianli Chen is currently a PhD candidate in the Department of Civil and Environmental Engineering in University of Illinois at Urbana Champaign, USA. He received his B.S. in 2013 from Tongji University and M.S. in 2015 from UIUC. His main research interest includes wave propagation and metamaterials.