Methods to Establish Input-Output Relationship for System Identification-Based Models

Feifei Bai[1,2], Yong Liu[1], Yilu Liu[1,3], Kai Sun[1], Xiaoru Wang[2]
[1]The University of Tennessee at Knoxville
[2]Southwest Jiaotong University
[3]Oak Ridge National Laboratory
fbai1@utk.edu
liu@utk.edu

Navin Bhatt[4], Alberto Del Rosso[4], Evangelos Farantatos[4]
[4]Electric Power Research Institute
nbhatt@epri.com

Abstract—**Both model-based and measurement-based methods are presented in this paper to describe the correlation between measurement locations. Transfer impedance (the model-based method) and correlation coefficient (the measurement-based method) are compared and applied to input location selection of power system dynamic modeling for dynamic response estimation. The comparison results show that both methods can describe the correlation between measurement locations effectively.**

Index Terms—**Autoregressive with exogenous (ARX), correlation coefficient map, dynamic response estimation, measurement-based, transfer impedance.**

I. INTRODUCTION

Increasing interconnection of power systems has made power system dynamic studies more complex. Therefore, it is necessary to reduce a large power network to a smaller equivalent system while keeping the dynamic characteristics of the full system model within the desired accuracy [1, 2]. The traditional and widely-used reduction approaches are the coherency method [3] and the model reduction method [4]. The first category is to use coherency and aggregation methods to obtain reduced models in the form of nonlinear power system models [5]. The second category treats the external system or the less relevant part of the system as a reduced lower order linear or nonlinear model. However, both methods are based on the power system detailed models, which are not accurate enough for two reasons: 1) the power system dynamic model could not include all the details in the power grid, especially on the load side; 2) the topology variations of power system happen occasionally, especially at the lower voltage level. Neither of the two aspects can be captured accurately by the existing models.

Synchrophasors provide system dynamics information in real time [6]. Such information can be used to build purely measurement-based system models for dynamics assessment and control. System identification is a commonly used approach to model the behaviors of a dynamic system based on measurement data [7]. However, a real large power system involves a large number of potential input locations. The input locations actually used by the model must be optimized to an acceptable level since it directly affects the model complexity [7].

The input location selection approaches can be divided into two main groups: 1) Measurement-based methods, which do not need a detailed power system model to find significant inputs. These approaches only use measurement data and are based on statistical tests [8, 9]; 2) Model-based methods, which rely on a specific power system model.

In this paper, both model-based and measurement-based methods are proposed to find the correlation between measurement locations to reduce the number of inputs of system dynamic response estimation model. This paper is organized as follows: Section II provides an explanation of dynamic response estimation and autoregressive with exogenous (ARX) model structure. Section III describes the model-based method. The measurement-based method is developed in Section IV. Section V presents a comparison between two methods. Section VI presents an application example to show the effectiveness of the proposed methods in the input selection for system dynamic response estimations. The conclusion and future work are provided in Section VII.

II. DYNAMIC RESPONSE ESTIMATION AND ARX MODEL

A. Dynamic Response Estimation

To achieve the goal of dynamic estimation in the power system, a concept of dynamic response estimation was proposed in [10]. It is to estimate the dynamic response of the power system during transient situations. The basic idea of dynamic response estimation is to identify the real-time dynamic model or transfer function of the power system and use the obtained model to estimate the dynamic response of the power system. ARX model structure and least squares

(LS) technique is employed to identify the dynamic model for the estimation.

B. ARX Model Structure

The mathematical structure expression of the ARX model is given below:

$$A_k(z)\,\hat{y}_k(t) = \sum_{j=1}^{n_j} B_{jk}(z)u_j(t) + e(t) \tag{1}$$

$$A_k(z) = 1 + a_{k1}z^{-1} + \cdots + a_{kn_{ak}}z^{-n_{ak}} \tag{2}$$

$$B_{jk}(z) = b_{jk0} + b_{jk1}z^{-1} + \cdots + b_{jk(n_{b_{jk}}-1)}z^{-(n_{b_{jk}}-1)} \tag{3}$$

where t is the sampled data index, $e(t)$ is the system noise, $A_k(z)$ and $B_{jk}(z)$ are the ARX denominator and numerator polynomials, respectively. u_j and $\bar{y}_k$ are the model input and output, respectively. The ARX structure and algorithm details are described in [7].

To evaluate the identified ARX model, a fitness criterion can be performed [11]:

$$F = (1 - \frac{\sqrt{(Y_i - \hat{Y}_i)^2}}{\sqrt{(Y_i - \bar{Y}_i)^2}}) \times 100 \tag{4}$$

where Y_i, $\hat{Y}_i$, $\bar{Y}_i$ are the measured response, estimated response, and the mean value of the measured response, respectively. This index is used to reflect the accuracy of the model in describing system's dynamic characteristics. A fitness of 100 means a perfect fit between the estimated response and the measured response, while a fitness of zero means the estimated response is no better than the mean value of measured response.

III. MODEL-BASED METHOD TO ESTABLISH MODEL INPUT-OUTPUT RELATIONSHIP

In power flow calculations, power transfer distribution factor is the relative change in power flow on a particular line due to a change in injection and corresponding withdrawal at a pair of buses. This factor is an index to describe the "connectivity" or "sensitivity" and it depends on the topology of the electric power system and the operating point. Likewise, transfer impedance (TI) could be defined in a similar way to describe the correlation between two measurement locations instead of two lines. Specifically, the TI between two measured locations (one is selected as the fixed model output location, others are selected as the selectable model input location) could be defined as the absolute value of voltage change at bus i if the current injection at bus j is varied by 1.0 pu. Since this index has the unit of impedance, the transfer impedance Z_{ij} could be used to denote it.

The definition of TI is:

$$TI = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1j} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & \cdots & Z_{2j} & \cdots & Z_{2n} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ Z_{i1} & Z_{2j} & \cdots & Z_{ij} & \cdots & Z_{in} \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{nj} & \cdots & Z_{nn} \end{bmatrix} \tag{5}$$

where Z_{ij} is the inverse of power system admittance matrix Y_{ij}, in other words, Z_{ij} is the transfer impedance between bus i and bus j. From the definition, it is apparent that larger TI means stronger correlation while smaller TI indicates weaker correlation. In other words, the larger one is better to be selected as the model identification inputs.

IV. MEASUREMENT-BASED METHOD TO ESTABLISH MODEL INPUT-OUTPUT RELATIONSHIP

A. Correlation Coefficient

Consider there are two power system measurement locations with output signals: $x(t)$ and $y(t+\tau)$, which are assumed as the wide-sense-stationary random process, where τ is the time delay between these two signals. Their mathematical expectations are $u_x = E\{x(t)\}$ and $u_y = E\{y(t)\}$, respectively. So the cross-correlation function and cross-covariance function can be defined as follows:

$$R_{xy}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_0^T x(t)y(t+\tau)dt \tag{6}$$

$$C_{xy}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \{x(t) - \mu_x\}\{y(t+\tau) - \mu_y\}dt \tag{7}$$
$$= R_{xy}(\tau) - \mu_x\mu_y$$

For the special case where $x(t) = y(t)$, the auto-correlation function $R_{xx}(\tau)$ and auto-covariance function $C_{xx}(\tau)$ of $x(t)$ become:

$$R_{xx}(0) = \lim_{T \to \infty} \frac{1}{T} \int_0^T x(t)x(t)dt \tag{8}$$

$$C_{xx}(0) = \lim_{T \to \infty} \frac{1}{T} \int_0^T x(t)x(t)dt \tag{9}$$

If $\mu_x = 0$, $\mu_y = 0$, it can be obtained:

$$C_{xy}(\tau) = R_{xy}(\tau) \tag{10}$$

$$R_{xy}(\tau) = \int_{-\infty}^{+\infty} S_{xy}(f)e^{j2\pi f\tau}df \tag{11}$$

$$R_{xx}(\tau) = \int_{-\infty}^{+\infty} S_{xx}(f)e^{j2\pi f\tau}df \tag{12}$$

Using (8), (11) and (12), correlation function in time domain can be obtained:

$$r_{xy}(\tau) = \int_{-\infty}^{+\infty} \gamma_{xy}(f) e^{j2\pi f\tau} df = \frac{R_{xy}(\tau)}{\sqrt{R_{xx}(0)R_{yy}(0)}} \quad (13)$$

Using (6), (7), (10), and (13) the correlation coefficient function can be obtained as follows:

$$r_{xy}(\tau) = \frac{R_{xy}(\tau) - \mu_x \mu_y}{\sqrt{\left[R_{xx}(0) - \mu_x^2\right]\left[R_{yy}(0) - \mu_y^2\right]}} \quad (14)$$
$$= \frac{C_{xy}(\tau)}{\sqrt{C_{xx}(0)C_{yy}(0)}}$$

where r_{xy} ranges from -1 (complete linear inverse correlation) to 1 (complete linear correlation). $r_{xy} = 0$ means lack of linear interdependence. The sign of r_{xy} indicates the direction of correlation: $r_{xy} < 0$ implies inverse correlation, i.e., the two power signals are out of phase or oscillate with each other. $r_{xy} > 0$ implies direct correlation, i.e. a tendency of both signals to have similar value with the same sign.

Correlation Coefficient (CC) matrix is defined as:

$$CC = \begin{bmatrix} r_{11}(\tau) & r_{12}(\tau) & \cdots & r_{1j}(\tau) & \cdots & r_{1n}(\tau) \\ r_{21}(\tau) & r_{22}(\tau) & \cdots & r_{2j}(\tau) & \cdots & r_{2n}(\tau) \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ r_{i1}(\tau) & r_{2j}(\tau) & \cdots & r_{ij}(\tau) & \cdots & r_{in}(\tau) \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ r_{n1}(\tau) & r_{n2}(\tau) & \cdots & r_{nj}(\tau) & \cdots & r_{nn}(\tau) \end{bmatrix} \quad (15)$$

where r_{ij} is the correlation coefficient between measurement locations i and j:

$$r(x_i(t), y_j(t)) = \frac{C(x_i(t), y_j(t))}{\sqrt{C(x_i(t), x_i(t))C(y_j(t), y_j(t))}} \quad (16)$$

B. Correlation Coefficient Map

Correlation Coefficient Map (CCM) is developed with Matlab® mapping toolbox to show the measurement signals correlation in a visual way.

It is assumed that 135 Phasor Measurement Units (PMUs) are distributed in the EI system and the simulation results from the 135 buses in the EI model are considered as the measurement signals. In CCM, different colors are used to indicate the different correlation coefficient values, while the bar on the left shows the correlation coefficient. In the map, one interested measurement location is chosen as the fixed location. The map shows the correlation coefficients between this fixed location and the others in geographically.

Assuming an event of 838 MW generation trip occurs in the East of Alabama, the CC are calculated then the CCM and CC for the voltage magnitude, phase angle, and frequency are shown in Figure 1. The correlation coefficients between the fixed location shown as purple triangle (selected model output) and other locations shown as orange circle (selectable model inputs) can be found directly from the color map. The area in the same color has the same correlation with the fixed location. If the fixed location is the model output, the locations having the similar color with the fixed location can be selected as the input locations for the model system identification model development.
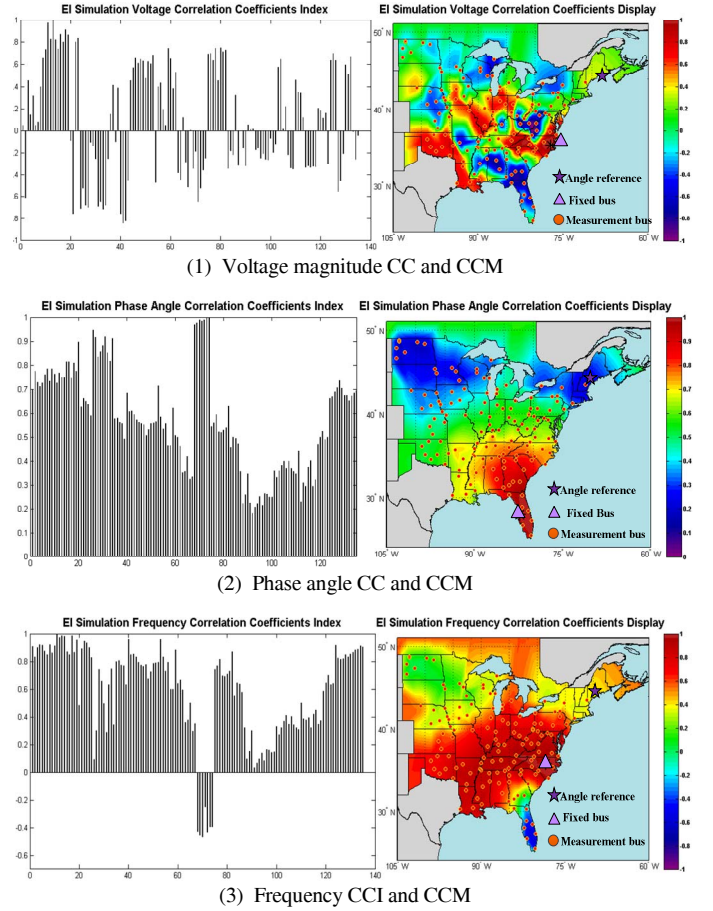

(1) Voltage magnitude CC and CCM

(2) Phase angle CC and CCM

(3) Frequency CCI and CCM

Figure 1. CC and CCM to show the correlation of the measurement locations

V. COMPARISON BETWEEN MODEL-BASED AND MEASUREMNT-BASED INPUT SELECTION METHODS

Correlation coefficients and transfer impedances are calculated between the fixed measurement location and other measurement locations in five different directions as shown in Figure 2, which are used to represent the correlation variation of the EI system. The comparison results are shown in Figure 3. The approach of comparison between model-based and measurement-based observability index is:

1) Find a series of buses in the five different directions shown by green arrows in Figure 2.

2) Calculate the correlation coefficients and transfer impedances between the fixed measurement location

(blue triangle) and other measurement locations (orange dots) in the five directions. As for the transfer impedance calculation, the EI system admittance matrix Y-bus is exported from PSS®E EI model while the inverse process of the transfer impedance is completed in Matlab®.

3) Draw the trend of the transfer impedances and correlation coefficients in the same direction from the fixed location to the farthest. The results are shown in Figure 3.
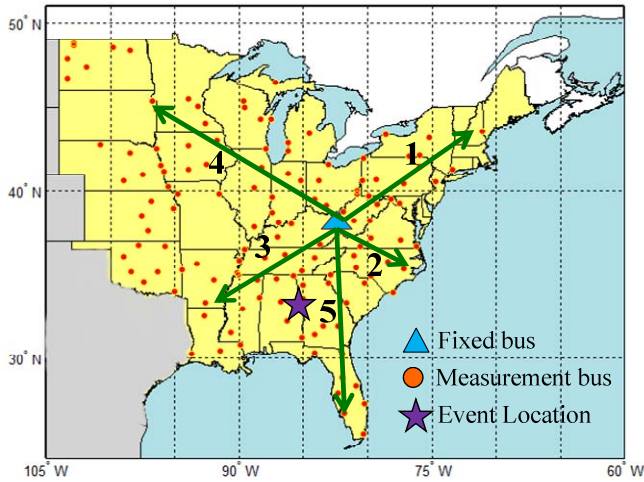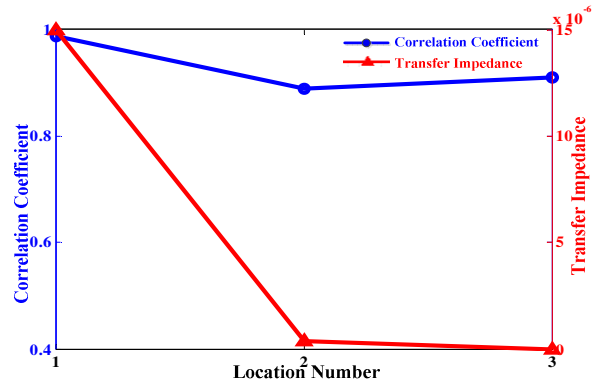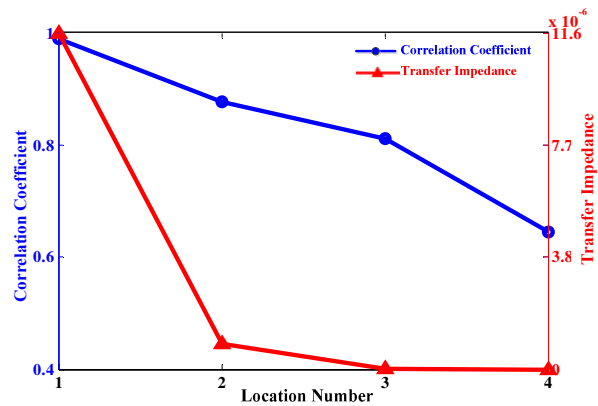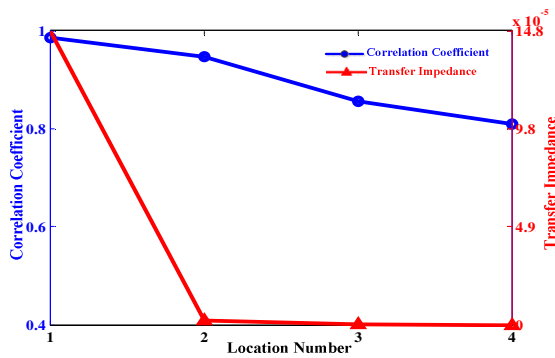

Figure 2. Five directions for the comparition of two methods

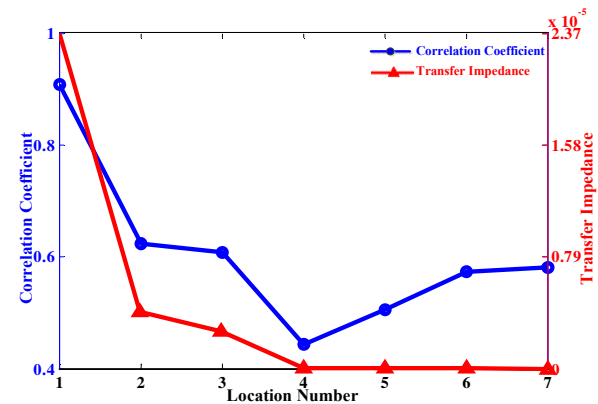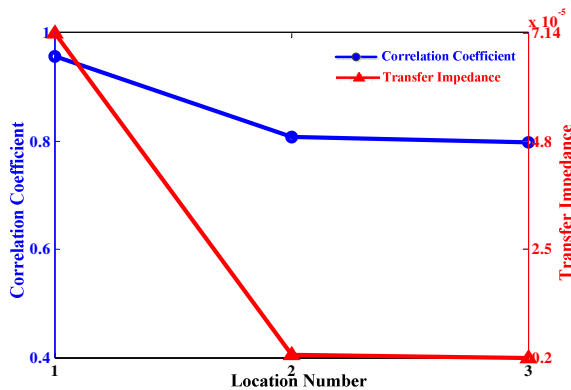(1) CC and TI variation trend in direction 1

(2) CC and TI variation trend in direction 2

(3) CC and TI variation trend in direction 3

(4) CC and TI variation trend in direction 4

(5) CC and TI variation trend in direction 5

Figure 3. Comparison between CC and TI variation trend in 5 directions

From the comparison in Figure 3, the variation trend of CC and TI is similar in the same direction, which suggests that the developed measurement-based and the model-based methods establish the input-output relationship work similarly. Furthermore, the relationship becomes weaker from the fixed bus to the remote bus. The transfer impedance calculation is simple; however, it is calculated in steady state so that it is not able to reflect the real-time operation conditions of the EI system. Comparatively, real-time data is used by measurement-based method (CCI), which contains the information of system dynamics.

VI. Application of the Proposed Methods

To illustrate the proposed method, a 16,000-bus EI model serves as the test bulk power system in all the following tests. About 3000 generators and 16,000 buses are included in the PSS/E EI dynamic model. The dynamic simulation time step is 0.001s and the time window for model development is 20s (from the beginning of one event). Assume that there are 135 buses with PMUs and two events are simulated. Event 1 is 814MW generation trip in Southeast Florida and event 2 is 838MW generation trip in East Alabama, which are shown in Figure 4. Event 1 is used to train the ARX model and event 2 is used to perform the verification of the proposed correlation method. Figure 4 shows CCM, which were obtained from the frequency response during Event 1. Based on these CCI and CCM information, various sets of input locations were selected, which were then used to estimate the response during Event 2 at the output bus. A comparison of actual versus estimated frequency response at the output bus for the selected set of input locations is provided in Figure 4.
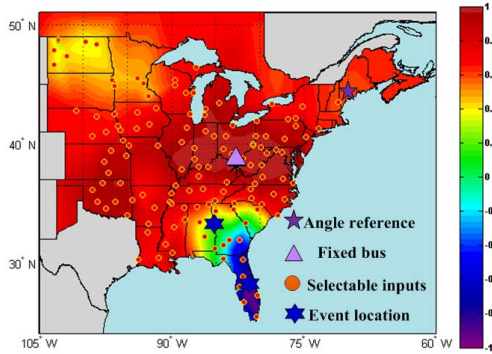

Figure 4. Frequency response correlation map for model input selection

(1) Top 4 and lowest 4 correlation signals as ARX model inputs

(2) Top 4 correlation signals as ARX model inputs

(3) Top 1 correlation signal as ARX model inputs
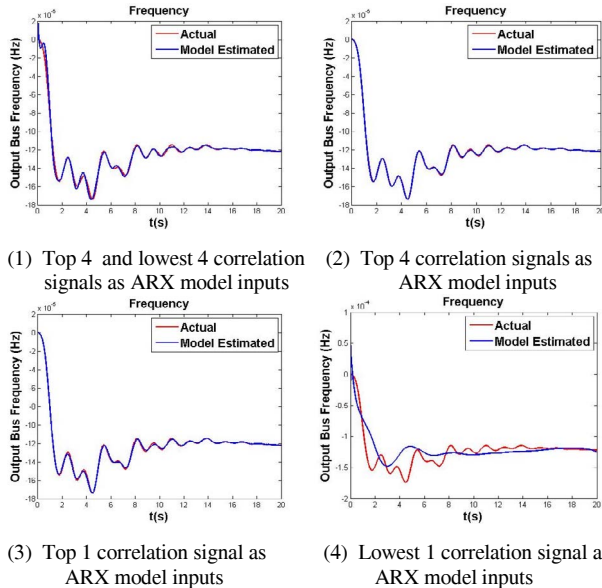
(4) Lowest 1 correlation signal as ARX model inputs

Figure 5. Application of proposed methods

The accuracy indices are 90, 98, 96 and 38 for figure 5 (1) to (4) respectively. From Figures 5, it is obvious that when input signals having high correlation coefficients with output signal location are selected, the model-estimated responses match the actual ones very well, which are shown in Figure 5 (1)-(3) for frequency. However, if only input signal locations with very low correlation coefficients are selected, the model-estimated responses do not match the actual response Therefore, only the signal locations with higher correlation coefficients should be selected as ARX input signal locations. Additionally, as shown in Figure 5 (3), if the input signal location with the highest correlation coefficient is selected, even one input signal location can provide a very good estimation, which can dramatically decrease the complexity of the ARX model.

VII. Conclusion

Both measurement-based and model-based methods are proposed to find the relationship of the measurement locations in power grid. The comparison results indicate that these two methods can be used to select the input location and reduce the input signals of ARX model. The application results demonstrate that the strong correlated input locations for dynamic response estimation can be quickly identified using the proposed methods. Therefore, the complexity of the model can be decreased. CCM has been developed for the Eastern Interconnection system to visually display the relationship of the measurement locations. The next step work is to take into account the on-line inertia calculation for the model-based method.

References

[1] Shankar Sastry, Pravin Varaiya, "Coherency for interconnected power systems," *IEEE Trans. Automatic Control*, vol. AC-27, pp. 218-226, Feb. 1981.

[2] Krishna K. Anaparthi, Balarko Chaudhuri, Bikash C. Pal, "Coherency identification in power systems through principal component analysis," *IEEE Trans.Power Systems,* vol. 20, pp.1658-1660, Aug. 2005.

[3] R. Podmore, "Identification of coherent generators for dynamic equivalents," *IEEE Trans. Power App. Syst.,* vol. PAS-97, pp. 1355-1353, 1978.

[4] J. E. Undrill and A. E. Turner, "Construction of power system electromechanical equivalent by model analysis," *IEEE Trans. Power App. Syst.,* vol. PAS-90, pp. 2049-2059, 1971.

[5] Chow, Joe H., editor, "Power system coherency and model reduction," Springer, 2013.

[6] J. Bertsch, C. Carnal, D. Karlsson, J. Mcdaniel, and K. Vu, "Wide-area protection and power system utilization," Proc. IEEE, vol. 93, no. 5,pp. 997–1003, May 2005.

[7] Lennart Ljung, System identification: theory for the user (2nd Edition), New Jersey: PTR Prentice, 1999.

[8] Sindelar, R., Babuska, R., "Input selection for nonlinear regression models," *IEEE Trans. Fuzzy Systems*, Vol. 12, No. 5, pp. 688-696, October, 2004.

[9] A. Poncet and G. S. Moschytz, "Optimal order for signal and system modeling," in Proc. IEEE Int. Symp, Circuits and Syatem, vol. 6, New York, 1994, pp.221-224.

[10] Yong Liu, Kai Sun, and Yilu Liu, "Measurement-based power system dynamic model for response estimation," in San Diego. 2012 IEEE PES Power and Energy Society General Meeting, pp. 1-6.

[11] N. Bhatt, "Synchrophasor-based power system dynamic modeling for stability estimation," EPRI, Palo Alto, CA, Tech. Rep. 2012.1026456.