IET Generation, Transmission & Distribution

Research Article

IET Journals
The Institution of Engineering and Technology

Measurement-based correlation approach for power system dynamic response estimation

ISSN 1751-8687
Received on 30th March 2014
Revised on 22nd February 2015
Accepted on 22nd March 2015
doi: 10.1049/iet-gtd.2014.1013
www.ietdl.org

Feifei Bai[1,2] ✉, Yong Liu[2], Yilu Liu[2,3], Kai Sun[2], Navin Bhatt[4], Alberto Del Rosso[4], Evangelos Farantatos[4], Xiaoru Wang[1]

[1]*Department of Electrical Engineering, Southwest Jiaotong University, Chengdu, Sichuan, People's Republic of China*
[2]*Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA*
[3]*Oak Ridge National Laboratory, Oak Ridge, TN, USA*
[4]*Electric Power Research Institute, Columbus, OH, USA*
✉ *E-mail: feifei_bai@126.com*

Abstract: Understanding power system dynamics is essential for online stability assessment and control applications. Global positioning system-synchronised phasor measurement units and frequency disturbance recorders (FDRs) make power system dynamics visible and deliver an accurate picture of the overall operation condition to system operators. However, in the actual field implementations, some measurement data can be inaccessible for various reasons, for example, most notably failure of communication. In this study, a measurement-based approach is proposed to estimate the missing power system dynamics. Specifically, a correlation coefficient index is proposed to describe the correlation relationship between different measurements. Then, the auto-regressive with exogenous input identification model is employed to estimate the missing system dynamic response. The US Eastern Interconnection is utilised in this study as a case study. The robustness of the correlation approach is verified by a wide variety of case studies as well. Finally, the proposed correlation approach is applied to the real FDR data for power system dynamic response estimation. The results indicate that the correlation approach could help select better input locations and thus improve the response estimation accuracy.

1 Introduction

Presently, wide area measurement systems based on synchrophasors, for example, phasor measurement units (PMUs) [1] and frequency disturbance recorders (FDRs) [2], can continuously provide high-sampling-rate, global positioning system (GPS)-synchronised phasor measurements over wide areas, which enable the real-time monitoring of wide-area power system dynamics [3, 4]. These measurement data contain very important information for power system researchers and operators to understand the power grid dynamics. Therefore a lot of research has been carried out based on the real measurement data of the power grid, for example, frequency prediction [5], inter-area oscillation analysis and damping control [6, 7] and stability analysis [8]. However, there are still some issues, which need to be solved: first, PMUs/FDRs are usually installed in limited locations of the power grid, which makes it difficult to study the dynamic characteristics of certain locations of interest; second, some PMUs/FDRs may lose partial measurements because of the failure of communication between PMUs and the central operator, which may make the system unobservable; third, some PMUs/FDRs may deliver bad data because of cyber attacks, which may cause wrong operation decisions [9, 10]. Therefore the dynamic information in these conditions should be accurately estimated for enhancing system stability and reliability.

Traditionally, missing data reconstruction methods are used to help find the missing measurements [11, 12]. Gao *et al.* [11] formulated the missing PMU data reconstruction problem into a low-rank matrix completion problem, but the method only works for problems that are missing some data points. Traditional state estimators usually generate pseudo-measurements from historical data or context of the generation of pseudo-measurements to replace the missing data [12]. As for the most widely studied static state estimation, it could not capture the dynamics very well after disturbances [12, 13]. Therefore the dynamic state estimation is studied [14–17]. Aminifa *et al.* [14] proposed a dynamic state estimation approach to predict voltage magnitude and phase angle whenever an operation suddenly changed. Ghahremani and Kamwa [15] presented a dynamic state estimation process based on Kalman filtering techniques to simultaneously estimate synchronous generator rotor angle and speed. However, both of the proposed methods depend on the system circuit model, which may not be adequately accurate for two reasons: on the one hand, the power system circuit-based dynamic model in the simulation software could not include all the details of the power grid; on the other hand, the operation condition and topology change frequently (although usually not dramatically) in the power grid. Neither of these two aspects can be addressed by the existing circuit-based model. To avoid these shortcomings, measurement-based approaches had been proposed. Multi-layer feed-forward artificial neural network (ANN) models were trained and used to estimate the generator rotor speed in a two machine system, whose estimation comparison results show that the measurement-based model is better than the classical circuit-based model [16]. Angel *et al.* [17] proposed transient rotor angle estimation using ANN in a five-generator system, the results show that the measurement-based method could be used in real time for the ultra-fast estimation. Besides, most of the previous dynamic state estimations have focused on small-sized systems, which did not involve lots of candidate input locations for the estimation model construction.

This paper proposes a measurement-based dynamic response estimation method, which could estimate a period of dynamic response, for example, the entire frequency response after a generation trip. In addition, the validation of this approach is also carried out using FDR data from the US Eastern Interconnection (EI). In contrast with all the methods above, the main advantages of the proposed approach are: (i) it is purely based on

measurements, thus does not rely on the circuit model of the power grid; (ii) the missing measurements are estimated by other strong correlated accessible measurements, rather than their own historical data; (the correlation between system dynamics can be shown in a visual way on a power grid map, which could be updated in real time.

The rest of the paper is organised as follows. Section 2 introduces the model structure used to estimate the missing measurements. The methodology to select the best input locations for the model construction is described in Section 3. Section 4 provides a case study of the proposed approach in the EI system. The robustness of the algorithm is demonstrated and strong correlation areas are defined in Section 5. The algorithm is applied to real measurement data for dynamic response estimation in Section 6. Conclusions are offered in Section 7.

2 Model structure for dynamic response estimation

System identification is the approach to model behaviours of a dynamic system based on measurement data [18]. An identification model could be constructed using the PMU/FDR measurements as model input(s) and output(s). However, it is unrealistic to develop a model that can include all the details of power system dynamic characteristics. Generally, the authors had to utilise a model that correctly captures the specific dynamic phenomenon or interaction for a particular study. Depending on the purpose of the study, the appropriate model of a given power system component could vary significantly. Therefore a power system is usually modelled as a black box with finite measurement points, which is usually expressed in a transfer function form [18]. In this paper, the linear auto-regressive with exogenous input (ARX) model is used as a transfer function to estimate the system dynamics. The mathematical structure expression of the single-input single-output (SISO) ARX model structure is described by a linear difference equation [18]

$$\begin{aligned} &y(k) + a_1 y(k-1) + \cdots + a_{n_a} y(k-n_a) \\ &\quad = b_1 u(k-1) + \cdots + b_{n_b} u(k-n_b) + e(k) \end{aligned} \tag{1}$$

For simplification, (1) can be rewritten as

$$A(q)y(k) = B(q)u(k) + e(k) \tag{2}$$

where k is the data point index and $e(k)$ is a white noise. n_a and n_b are the orders of the signal $y(k)$ and $u(k)$, respectively. $u(k)$ is the input signal and $y(k)$ is the output signal, which is the currently inaccessible system response, respectively. $A(q)$ and $B(q)$ are polynomials in q^{-1} as shown in (3) and (4), respectively. q^{-1} is a backward shift operator, which is the conventional definition of the z-transform [18]

$$A(q) = 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a} \tag{3}$$

$$B(q) = b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b} \tag{4}$$

With the SISO ARX model structure in (2), the multi-input single-output ARX model structure can be derived

$$A(q)y(k) = \sum_{j=1}^{n_j} \boldsymbol{B}_j(q) u_j(k) + e(k) \tag{5}$$

where j is the number of input signal, n_j is the order of the jth input signal.

Because of the linear structure of the ARX model, the model parameters of a multi-variable ARX model can be estimated by a linear least-square (LS) estimation method which usually uses QR factorisation for overdetermined linear equations to optimise the ARX model parameters and to minimise the function V_{LS}. The details of ARX structure and algorithm are discussed in [18, 19]

$$\text{Min } J(V_{\mathrm{LS}}) = \sum_{k=n_s+1}^{N} \big(\varepsilon(k)\big)^2 \tag{6}$$

where N is the total number of the data points and n_s is order of the ARX model; the equation error $\varepsilon(k)$ is described by

$$\varepsilon(k) = \hat{y}(k) - y(k) \tag{7}$$

where $\hat{y}(k)$ and $y(k)$ are the measured response and estimated response, respectively.

To evaluate the identified ARX model, a model fitness criterion is employed as the model accuracy index

$$F = \left(1 - \frac{\sqrt{(\hat{Y} - Y)^2}}{\sqrt{(\hat{Y} - \bar{Y})^2}}\right) \times 100 \tag{8}$$

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

For easier interpretation, a normalisation process that converts the accuracy index from $(-\infty, 100]$ to $(0, 1]$ can be performed as

$$A = \mathrm{e}^{(F/100)-1} \tag{9}$$

After obtaining the model orders and parameters, the identified models can be evaluated by comparing their outputs with the actual system responses. The objective function for the model evaluation is

$$\mathrm{e}^{(F/100)-1} > \varepsilon \tag{10}$$

where ε is the model accuracy threshold. ARX models with low accuracy index will be rejected, whereas only the model with the acceptable accuracy can be selected as the final model for further estimation studies. The threshold of the accuracy index can be defined according to the study system and the event type [20].

3 Methodology of the proposed approach

A large power system indicates a great number of potential input locations. The number of input locations actually used by the model must be reduced to an acceptable level for the sake of model updating speed. Therefore input location selection is a critical aspect of system identification since it directly affects the model accuracy and complexity [21, 22].

In a power grid, the measurements in different locations have some underlying relationships, for example, all the bus frequencies change similarly after a generation trip. In this paper, a linear correlation concept in signal processing is employed to describe the relationship between arbitrary measurement locations in power grids. On the basis of the correlation ranked results, the inputs of the measurement-based model will be selected. In [23], ARX model is used to estimate the system dynamics, but it ignores one important aspect – how to choose the input locations from hundreds of measured locations. The methodology of how to define the correlation between measurements is described as follows.

3.1 Correlation coefficient index (CCI)

In power systems, a commonly used measurement-based coherency function [24, 25] is defined as

$$\gamma_{xy}(f) = \frac{\left|S_{xy}(f)\right|}{\sqrt{S_{xx}(f)S_{yy}(f)}} \quad \left|\gamma_{xy}\right| \leq 1 \tag{11}$$

where f is the frequency, γ_{xy} is the coherency between power system measured signals $\{x(t)\}$ and $\{y(t)\}$. $S_{xy}(f)$ is the cross-spectral density function between $\{x(t)\}$ and $\{y(t)\}$, $S_{xx}(f)$ and $S_{yy}(f)$ are the power-spectral density of $\{x(t)\}$ and $\{y(t)\}$, respectively. And these two signals are assumed as the wide-sense stationary random processes. This coherency function gives the linear correlation between two power system output signals as a function of the frequency. However, this frequency domain function is not convenient when dealing with a large number signals. On the other hand, the frequency domain function pays more attention to the coherency of the specific frequency point. The following equations will be used to derive the time domain correlation function for a wide frequency.

The cross-correlation function $R_{xy}(\tau)$, self-correlation functions $R_{xx}(\tau)$ and $R_{yy}(\tau)$ are given by the inverse Fourier transform of $S_{xy}(f)$, $S_{xx}(f)$ and $S_{yy}(f)$, respectively

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

$$R_{xx}(\tau) = \int_{-\infty}^{+\infty} S_{xx}(f)\, e^{j2\pi f\tau}\, df \tag{13}$$

$$R_{yy}(\tau) = \int_{-\infty}^{+\infty} S_{yy}(f)\, e^{j2\pi f\tau}\, df \tag{14}$$

where τ is the time delay.

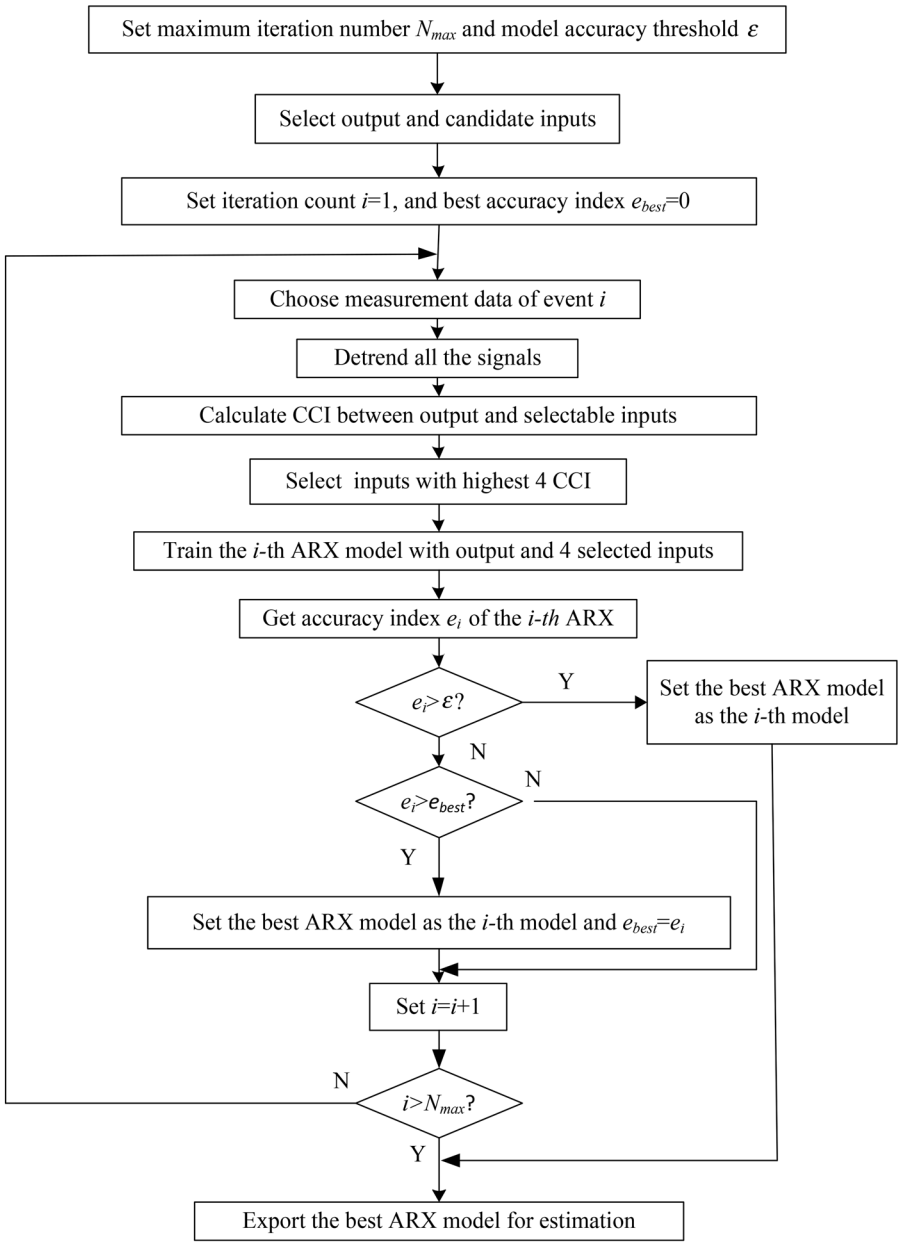

Fig. 1 *Flowchart of the input selection approach for model construction*

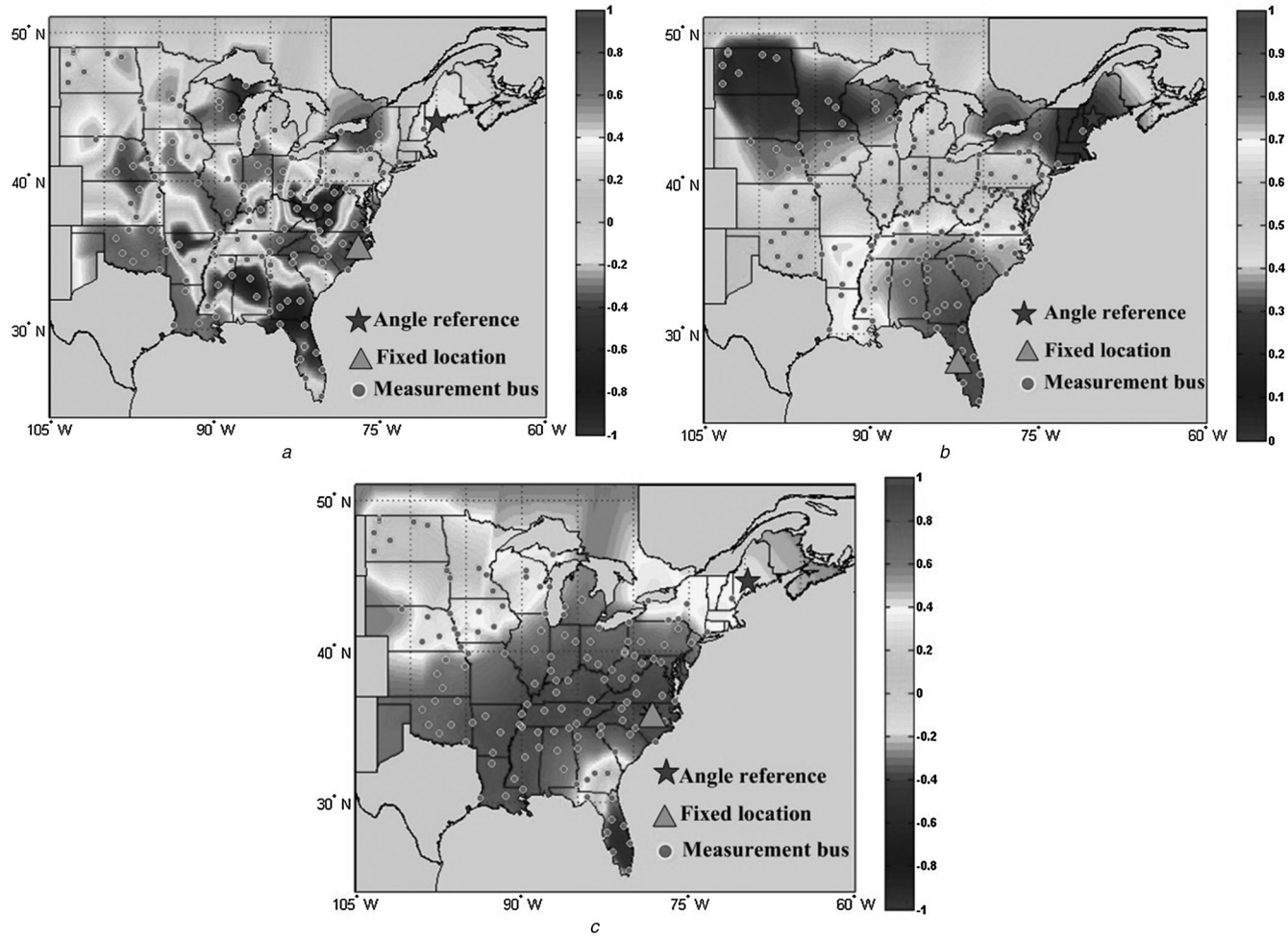

Fig. 2 *Correlation coefficients of the measured voltage magnitude, phase angle and frequency*

a Voltage magnitude CCM
b Phase angle CCM
c Frequency CCM

Applying the inverse Fourier transform to (11) and using (12), (13) and (14), the correlation function in the time domain is

$$r_{xy}(\tau) = \int_{-\infty}^{+\infty} \gamma_{xy}(f)\, \mathrm{e}^{\mathrm{j}2\pi f\tau}\, \mathrm{d}f = \frac{R_{xy}(\tau)}{\sqrt{R_{xx}(\tau)R_{yy}(\tau)}} \tag{15}$$

The mathematical expectations of these two signals are $u_x=E\{x(t)\}$ and $u_y=E\{y(t)\}$, respectively. Hence, the cross-correlation function $R_{xy}(\tau)$ and cross-covariance function $C_{xy}(\tau)$ are defined

$$R_{xy}(\tau) = E\big\{x(t)y(t+\tau)\big\} = \lim_{T\to\infty}\frac{1}{T}\int_0^T x(t)y(t+\tau)\,\mathrm{d}t \tag{16}$$

(see (7))

For the special case where $x(t)=y(t)$, the self-covariance function of $C_{xx}(\tau)$ and $C_{yy}(\tau)$ are

$$C_{xx}(\tau) = R_{xx}(\tau) - \mu_x^2, \quad C_{yy}(\tau) = R_{yy}(\tau) - \mu_y^2 \tag{18}$$

If μ_x=0, μ_y=0 in (15),we can obtain

$$C_{xy}(\tau) = R_{xy}(\tau), \quad C_{xx}(\tau) = R_{xx}(\tau), \quad C_{yy}(\tau) = R_{yy}(\tau) \tag{19}$$

Assuming the time delay $\tau=0$ and using (15) and (19), the correlation coefficient function can be obtained

$$r_{xy}(0) = \frac{C_{xy}(0)}{\sqrt{C_{xx}(0)C_{yy}(0)}} \tag{20}$$

where r_{xy} is the correlation coefficient between signals $\{x(t)\}$ and $\{y(t)\}$ in time domain, which ranges from -1 (complete linear inverse correlation) to 1 (complete linear correlation) with $r_{xy}=0$ meaning lack of linear interdependence. The sign of r_{xy} indicates the direction of correlation: $r_{xy}<0$ implies inverse correlation, that is, the two power signals are out of phase or oscillate with each other, whereas $r_{xy}>0$ implies direct correlation, that is, a tendency of both signals to have similar value with the same sign.

The CCI between the ith input signal x_i and the jth output signal y_j is defined as

$$r_{ij} = \frac{C(x_i, y_j)}{\sqrt{C(x_i, x_i)C(y_j, y_j)}} \tag{21}$$

where i and j are the input and output signal numbers, respectively.

The CCI will be a criterion to find the strong correlated measurements of different locations as inputs of the estimation model.

$$C_{xy}(\tau) = E\Big[\big\{x(t)-\mu_x\big\}\big\{y(t+\tau)-\mu_y\big\}\Big] = \lim_{T\to\infty}\frac{1}{T}\int_0^T \big\{x(t)-\mu_x\big\}\big\{y(t+\tau)-\mu_y\big\}\,\mathrm{d}t = R_{xy}(\tau) - \mu_x\mu_y \tag{17}$$

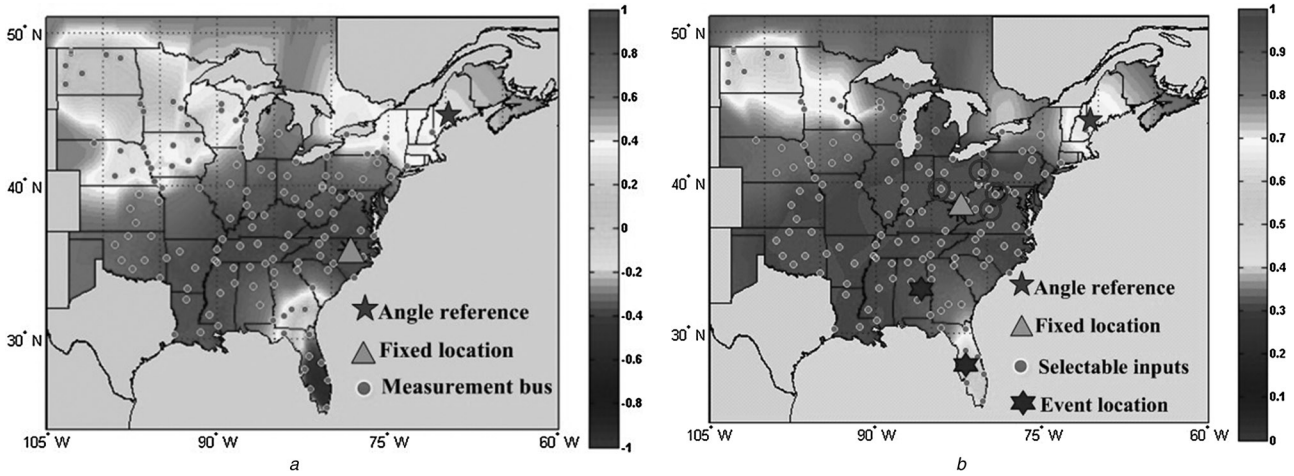

Fig. 3 *CCMs obtained from the frequency and phase angle response*

a Frequency CCM
b Phase angle CCM

3.2 Input location selection algorithm for ARX model construction

The ranking of input location correlations is obtained by calculating CCI using (21) with measured signals in different locations. The performance of the ARX model is the criterion to evaluate the selected input signals. Please note that the 'fixed location' is the output location of ARX model, where the measurement is currently inaccessible. The target of the algorithm is to obtain an accurate ARX model for the dynamic response estimation. The whole algorithm is described in Fig. 1. The basic steps outlined by this flowchart are:

Step 1: Detrend the candidate measured signals. To obtain a wide-sense stationary random process, all measurement data

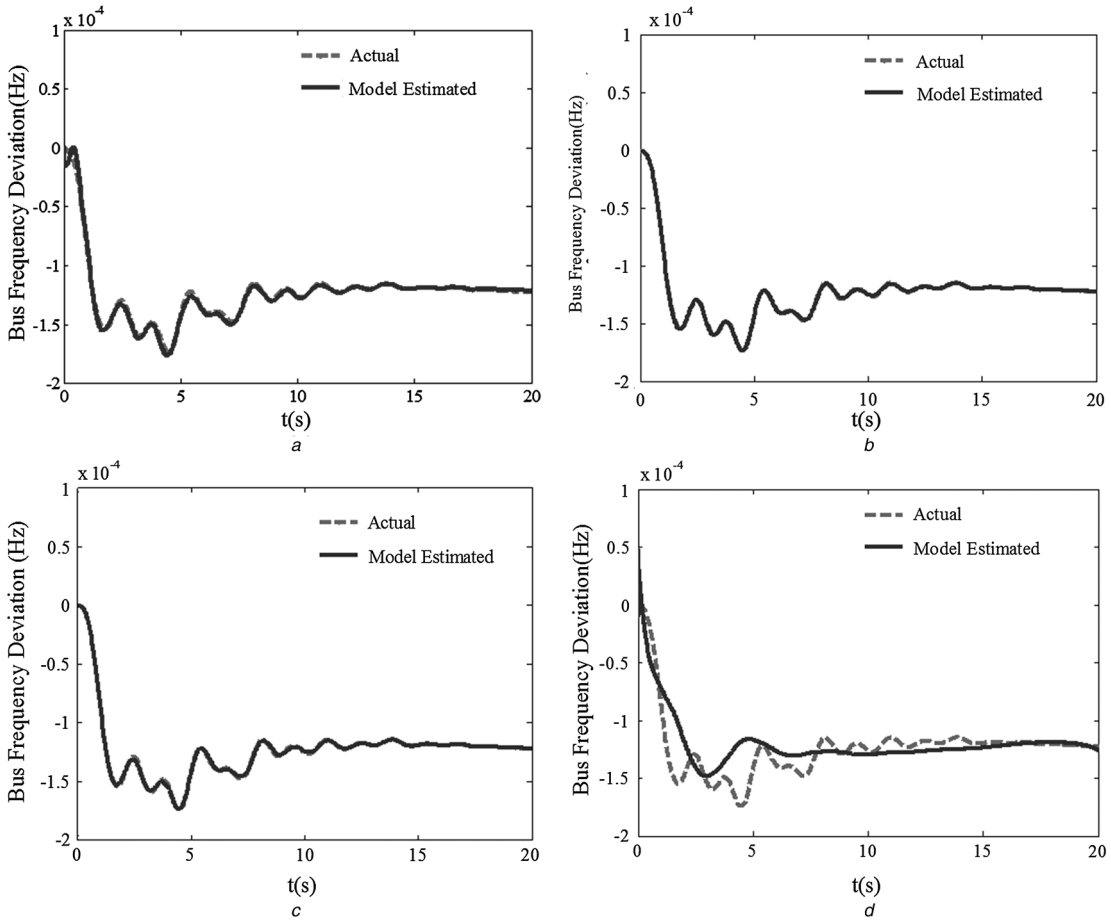

Fig. 4 *Input location selection algorithm verification using frequency dynamic response estimation*

a Highest four and lowest four coefficient locations as ARX model input locations
b Highest four coefficient locations as ARX model input locations
c Highest one coefficient location as ARX model input location
d Lowest one coefficient location as ARX model input location

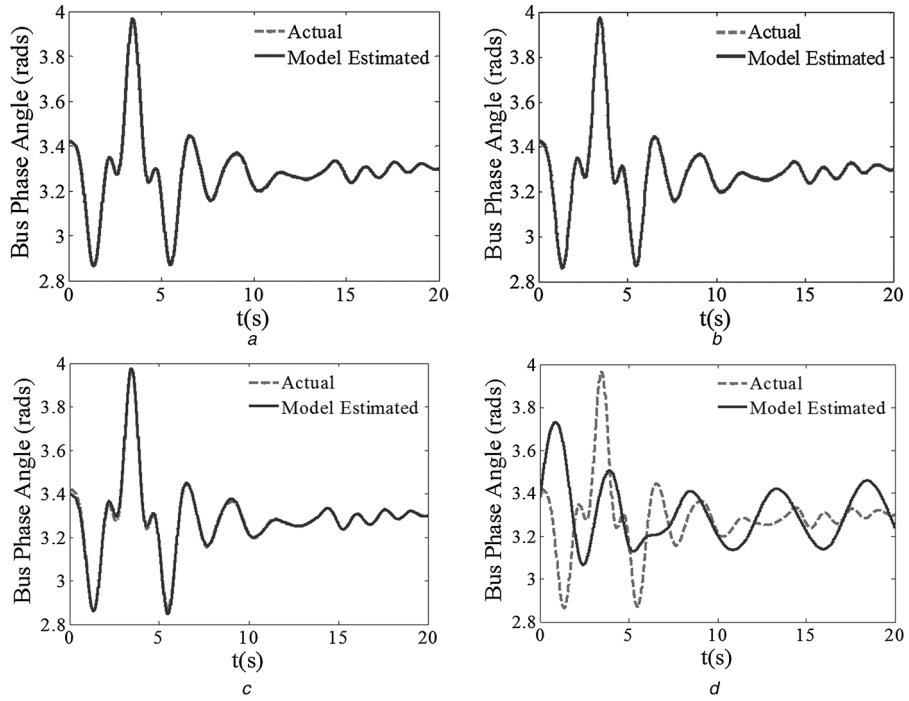

Fig. 5 *Input location selection algorithm verification using phase angle dynamic response estimation*

a Highest four and lowest four coefficient locations as ARX model input locations
b Highest four coefficient locations as ARX model input locations
c Highest one coefficient location as ARX model input location
d Lowest one coefficient location as ARX model input location

should be detrended to remove direct-current components. Here, a simple zero-mean filter should be used.

Step 2: Use the detrended signals to calculate the CCI between fixed location and the candidate input locations.

Step 3: Choose several inputs with the highest correlation coefficients as the ARX model inputs. For the EI system, only four inputs with the highest CCI are selected since the lower inputs the lower complexity of the ARX model. These four inputs are usually enough to obtain an acceptable model accuracy in the following cases. However, different power systems can require a higher number of inputs, which should be further checked.

Step 4: Train the ARX model with the selected input signals by CCI. The generation trip is considered as the excitation source to train the model.

Step 5: Test the ARX model accuracy. The model accuracy can be reflected by the accuracy index. ARX models with low accuracy index will be rejected while only the model with the acceptable accuracy can be selected as the final model for further estimation studies. The threshold of the accuracy index can be defined according to the study system and the event type [20]. The threshold or EI simulation of the case study is 0.9.

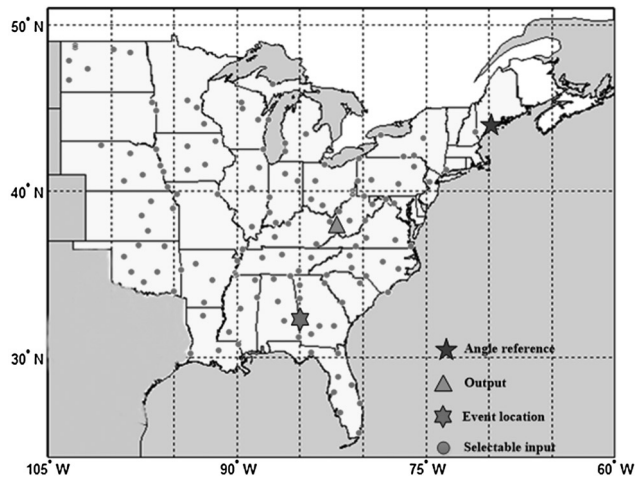

Fig. 6 *Event location and details of the test system*

Usually, an ARX model can be obtained with acceptable accuracy e, that is, model accuracy higher than preset accuracy threshold, ε, in one or several iterations. However, to prevent the proposed algorithm from infinite iteration, an outer loop is used to make sure the algorithm can successfully exit within limited iterations, that is, maximum iteration number of $N_{\max}$.

3.3 Correlation coefficient map (CCM)

A CCM can be developed for large power systems with the MATLAB® mapping toolbox to show the CCI in a visual way. The EI system can be taken as an example to explain how to obtain a CCM. In the example shown in Figs. 2*a*–*c*, the artificial measurement data were generated by simulating an 838 MW generation trip in eastern Alabama. In the CCM, the grey level is used to indicate the correlation coefficient values. The colour bar on the right of the map shows the coefficient value. In the map, one measurement location is chosen as the fixed location, which is the CCI calculation reference location. Then, the correlation coefficients between this fixed location and the other measurement locations can be shown vividly on the map. In Fig. 2, the correlation coefficients of the measured voltage magnitude, phase angle and frequency between the fixed location and other locations can be found from the map. If the fixed location is the ARX

Table 1 Generation trip events for the robustness test

Event number	1	2	3	4	5
amount, MW	328	510	717	863	1149

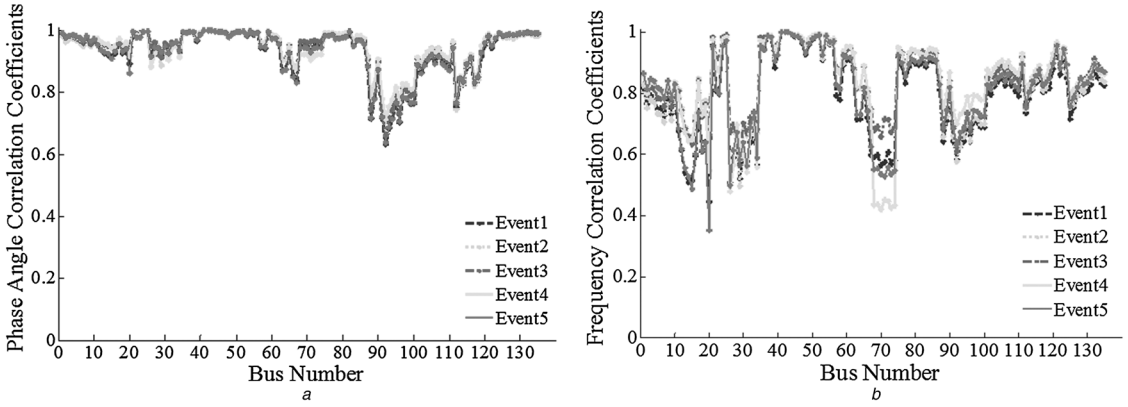

Fig. 7 *CCI variation tendency among five events*

a Phase angle CCI variation tendency among five events
b Frequency CCI variation tendency among five events

model output location, the measured locations sharing a similar grey level can be selected as the input locations. From the CCM in Fig. 2, the frequency and phase angle correlation are strong for the entire system, whereas the voltage magnitude correlation is shown as weak over wide areas. In this paper, only frequency and phase angle dynamic response estimations are studied.

4 Case study

To demonstrate the effectiveness of the proposed approach in dynamic response estimation, the simulation was used to simulate the artificial measurements. A case study was performed on the EI system, which contains 3000 generators and 16 000 buses in the power system simulation for engineering (PSS/E) EI model. One hundred and thirty-five buses are assumed to be measured by PMUs, and these 135 buses are evenly distributed throughout the EI system. The simulation results for the 135 buses are considered as the measured signals and 20 s of data after one disturbance is used. Note that the EI system serves as the test system in following cases. Two events were simulated to generate the artificial measurement data. Event 1 was an 814 MW generation trip in southeastern Florida, and Event 2 was an 838 MW generation trip in eastern Alabama, both of which are shown in Fig. 3 with stars. Event 1 was used to train the ARX model and Event 2 was used to perform the verification of the proposed algorithm. CCMs were obtained from the frequency and phase angle response during Event 1 and they are shown in Fig. 3. Based on the CCM information, various sets of inputs were selected, which were then used to estimate the dynamic response during Event 2 at the fixed location (shown in Figs. 4 and 5). A comparison of actual against estimated responses and the accuracy index are provided in Fig. 4.

From the comparison results between the actual and estimated frequency responses in Fig. 4, it is obvious that the estimated frequency responses match the actual ones very well when input locations have high correlation coefficients, as shown in Figs. 4*a*–*c* for frequency. If input locations with very low correlation coefficients are selected, the estimated responses do not match the actual response and the model estimation accuracy index is very low.

The comparison results between the actual and estimated phase angle responses are given in Fig. 5. It is obvious that the model estimated responses match the actual ones very well when input locations have high correlation coefficients, as shown in Figs. 5*a*–*c*. On the contrary, the estimation result may be completely wrong like the estimation result in Fig. 5*d*.

From these two dynamic response estimation results in Figs. 4 and 5, it can be seen that only the locations with higher correlation coefficients should be selected for the model estimation. Otherwise, the estimation results will be very bad. According to the input selection results, the selected highest four input locations are shown in Fig. 3 with circles.

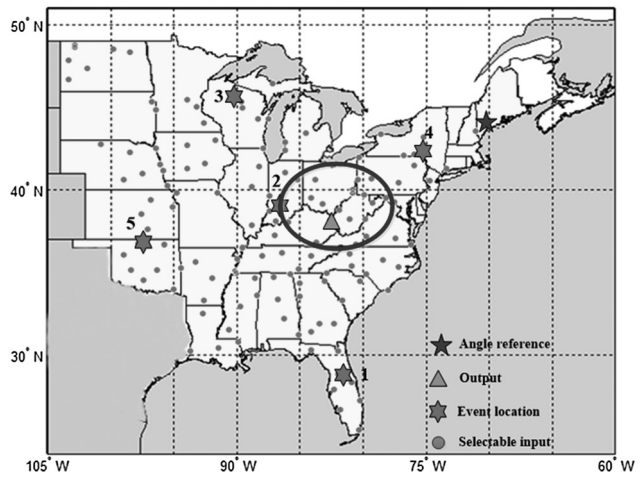

Fig. 8 *Event locations and details of the test system*

5 Correlation coefficient index robustness test

The results from the previous section show that the proposed input location selection approach can help achieve the accurate estimations of frequency and phase angle in the test cases. However, the proposed correlation calculation method is based on the disturbance data, which means that the CCI may be influenced by the disturbance location and magnitude. Therefore the robustness of this approach still needs to be verified in order to be applied to a real power system.

5.1 Disturbances in the same location with different magnitudes

The test system is the same as in Section 4. The event location is shown in Fig. 6 with a star. In the PSS/E dynamic simulation, the five event locations are the same while the generation trip

Table 2 Generation trip events for the robustness test

Event number	1	2	3	4	5
amount, MW	957	1000	1017	1150	990

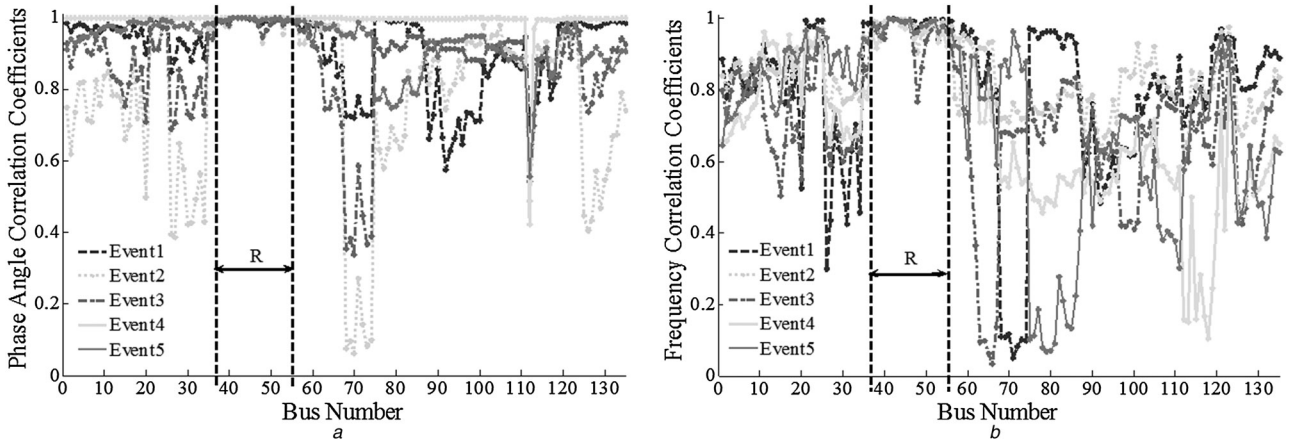

Fig. 9 *CCI variation tendency among five events*

a Phase angle CCI variation tendency among five events
b Frequency CCI variation tendency among five events

magnitudes are different. Table 1 shows the event details. The test results are shown in Figs. 7*a* and *b*, respectively.

As shown in Fig. 7, for both phase angle and frequency, the CCI variation tendencies remain similar for all the events, which indicate that the magnitudes do not influence the correlation relationship much and the proposed approach is robust to different disturbance magnitudes.

5.2 Disturbances in different locations with similar magnitudes

As shown in Fig. 8 by the star, five generation trip events, each of which is around 1000 MW, occur independently in the south, north, west, east and middle of the EI system (to show the universality of the event locations). The output location is labelled by the purple triangle in Fig. 8 and the event details are shown in Table 2. The phase angle and frequency CCI variation tendencies are shown in Fig. 9.

Fig. 9 indicates that most of the CCI values change when an event occurs in different locations. However, there are always some measurement locations which have very stable and high correlation coefficients no matter where the event occurs. These CCI variation trends are shown in Fig. 9 with the part marked 'R'. For this part, the measurements are all in the circle in Fig. 8. Additionally, it is apparent that all these measurement locations are close to the output location. Using the proposed input location selection approach, all of the four selected input locations are in this stable strong correlation area. That also means the ARX model input locations selected by the proposed approach always have strong and robust correlation with the ARX model output location.

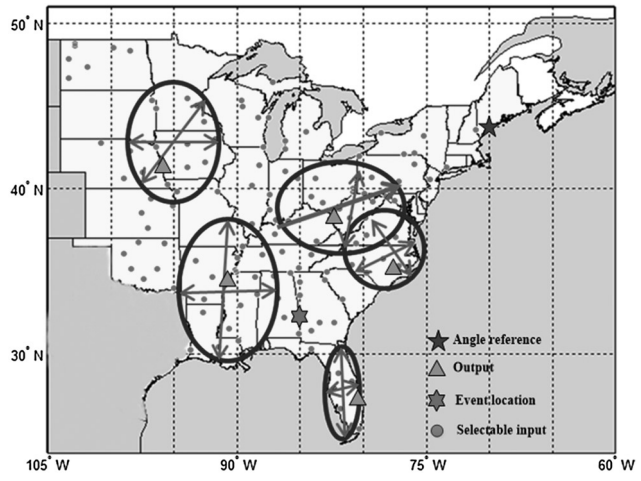

Fig. 10 *Strong correlation area for different outputs*

5.3 Strong correlation areas

Sections 5.1 and 5.2 prove that the strong correlation part of input locations has good robustness, and a strong correlation area exists around the output location. In this section, five typical output locations (triangles in Fig. 10) are chosen to find their strong correlation areas, which are shown in the circles in Fig. 10, and the area diameters are shown in Table 3.

The results in Fig. 10 give a geographic area of the input location selection for dynamic response estimation. In other words, all the measured signals in the blue circle can represent inputs of the measurement-based model to estimate the dynamics at the location of the triangle. Therefore the strong correlation areas will benefit the selection of input locations.

6 Algorithm test using real measurement data

The power system frequency monitoring network (FNET) is a wide-area measurement system that takes high accuracy, GPS-synchronised measurements at standard end-user distribution voltages [26], which is shown in Fig. 11*a*. As a member of the PMU family, the FDRs used in the FNET system measure frequency, voltage and phase angle at standard 120 V outlets and transmit these measurements through the internet [27]. It serves the entire North American power grid through advanced situational awareness techniques, such as real-time event alerts, accurate event location estimation, animated event visualisation and post-event analysis [28]. The FDR measurements used here are the frequency and phase angle signals sampled at 0.1 s. As shown in Figs. 11*b* and *c*, two disturbances monitored by FDRs were selected in this test. The circles are the detected event locations and the triangle is the ARX output location. Event 1 is used to calculate the CCI and train the ARX model using the

Table 3 Strong correlation area for different outputs

Interested location	Florida	North Carolina	East Iowa	West Virginia	East Mississippi
largest diameter, mile	320	240	390	520	600
smallest diameter, mile	110	230	320	400	400

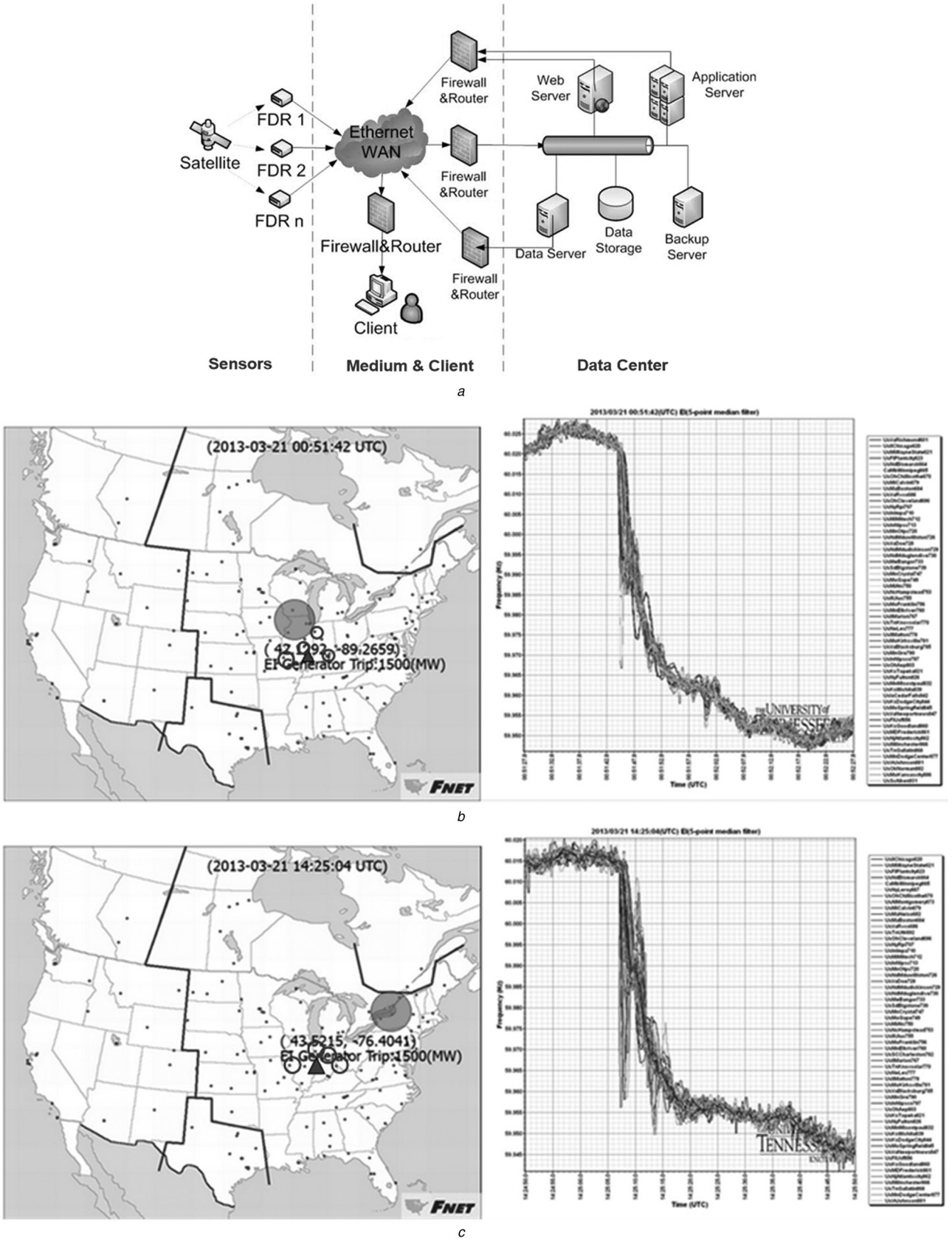

Fig. 11 *FNET and display of frequency response measured by FDR*

a Architecture of the FNET system at University of Tennessee, Knoxville, USA
b Event 1 and measured frequency responses by FDR
c Event 2 and measured frequency responses by FDR

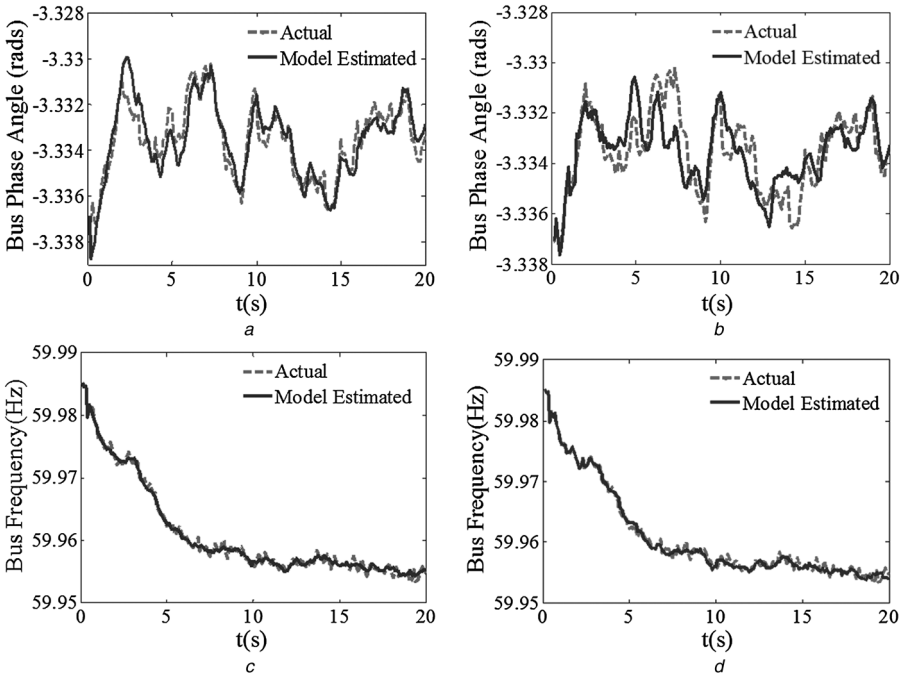

Fig. 12 *Bus frequency and phase angle dynamic response estimation*

a Highest four coefficient locations as ARX model input locations
b Highest one coefficient location as ARX model input location
c Highest four coefficient locations as ARX model input locations
d Highest one coefficient location as ARX model input location

proposed input location selection approach. The estimation results and the estimation accuracy index are shown in Fig. 12.

As shown in Fig. 12, it is demonstrated that even one input location may be enough for the frequency to obtain an accurate estimation in some cases, whereas four input locations are needed for phase angle estimation. The highest four FDR locations are marked as circles in Fig. 11. By comparing the accuracy index in Figs. 5 and 12, it can be shown that the estimation accuracy using real measurement data is lower than using simulation data. The main reason is that the real measurement data contains noise.

7 Conclusions

The proposed CCI is an effective measurement-based approach to calculate the correlation relationship between different measurements at different locations. Moreover, the strong correlated input locations for dynamic response estimation can be quickly identified using the proposed approach. In this paper, the proposed algorithm is verified using both artificial measurement data generated by simulation and real measurements from FDRs. The complexity of the ARX model can be reduced consequently and the estimated results match the real dynamics very well. Furthermore, the real FDR measurement test indicated that the proposed approach gave a stable performance and can be implemented in the actual power systems for the missing data estimation. Besides, the developed CCM could visually demonstrate the correlation between different measurement locations. In the future, ambient data will be considered to study the normal operation estimation.

8 Acknowledgments

This work is partially supported by the Electric Power Research Institute and also made use of Engineering Research Center Shared Facilities supported by the DOE under NSF Award Number EEC1041877. Additional support is provided by the CURENT Industry Partnership Program and China National Government Building High-level University Graduate Programs ([2012]3013). The authors also gratefully acknowledge the FNET team at the University of Tennessee for providing the phasor measurement data from FNET/Grideye for the verification work.

9 References

1 Phadke, G., Thorp, J.S.: 'Synchronized phasor measurements and their applications' (Springer-Verlag, New York, NY, USA, 2008)
2 Zhong, Z., Xu, C., Billian, B.J., *et al.*: 'Power system frequency monitoring network (FNET) implementation', *IEEE Trans. Power Syst.*, 2005, **20**, (4), pp. 1914–1921
3 Sun, K., Lee, S., Zhang, P.: 'An adaptive power system equivalent for real time estimation of stability margin using phase-plane trajectories', *IEEE Trans. Power Syst.*, 2011, **26**, (2), pp. 915–923
4 Phadke, A.G., de Moraes, R.M.: 'The wide world of wide-area measurement', *IEEE Power Energy*, 2008, **6**, (5), pp. 52–65
5 Dong, J., Ma, X., Djouadi, S.M., Li, H., Liu, Y.: 'Frequency prediction of the power systems in FNET based on state-space approach and uncertain basis functions', *IEEE Trans. Power Syst.*, 2014, **29**, (6), pp. 2602–2612
6 Sun, K., Zhou, Q., Liu, Y.: 'A phase locked loop-based approach to real-time modal analysis on synchrophasor measurements', *IEEE Trans. Smart Grid*, 2014, **5**, (1), pp. 260–269
7 Kamwa, I., Grondin, R., Hebert, Y.: 'Wide-area measurement based stabilizing control of large power systems – a decentralized/hierarchical approach', *IEEE Trans. Power Syst.*, 2001, **16**, (1), pp. 136–153
8 Diao, R., Sun, K., Vittal, V., *et al.*: 'Decision tree-based online voltage security assessment using PMU measurements', *IEEE Trans. Power Syst.*, 2009, **24**, (2), pp. 832–839
9 Huang, Y., Esmalifalak, M., Nguyen, H., *et al.*: ' Bad data injection in smart grid: attack and defense mechanisms', *IEEE Commun. Mag.*, 2013, **51**, (1), pp. 27–33
10 Kim, J., Tong, L.: 'On topology attack of a smart grid: undetectable attacks and countermeasures', *IEEE J. Sel. Area Commun.*, 2013, **31**, (7), pp. 1294–1305
11 Gao, P., Wang, M., Ghiocel, S.G., Chow, J.H.: 'Modeless reconstruction of missing synchrophasor measurements'. IEEE PES General Meeting, 2014
12 Miranda, V., Krstulovic, J., Keko, H., Moreira, C., Pereira, J.: 'Reconstructing missing data in state estimation with autoencoders', *IEEE Trans. Power Syst.*, 2012, **27**, (2), pp. 604–611
13 Qi, J., Sun, K., Kang, W.: 'Optimal PMU placement for power system dynamic state estimation by using empirical observability gramian', *IEEE Trans. Power Syst.*, **PP**, (99), pp. 1–14
14 Aminifar, F., Shahidehpour, M., Fotuhi-Firuzabad, M., Kamalinia, S.: 'Power system dynamic state estimation with synchronized phasor measurements', *IEEE Trans. Power Syst.*, 2014, **63**, (2), pp. 352–363

15 Ghahremani, E., Kamwa, I.: 'Dynamic state estimation in power system by applying the extended Kalman filter with unknown inputs to phasor measurements', *IEEE Trans. Power Syst.*, 2011, **26**, (4), pp. 2556–2566
16 Kamwa, I., Baraboi, B., Wamkeue, R.: 'Sensorless ANN-based speed estimation of synchronous generator: improved performance through physically motivated pre-filters'. IJCNN, Canada, 2006, pp. 1710–1718
17 Angel, A.D., Geurts, P., Ernst, D., Glavic, M., Wehenkel, L.: 'Estimation of rotor angles of synchronous machines using artificial neural networks and local PMU-based quantities', *Neurocomputing*, 2007, **70**, (16–18), pp. 2668–2678
18 Ljung, L.: 'System identification: theory for the user 2nd Ed' (PTR Prentice, New Jersey, USA, 1999), pp. 79–88
19 Ljung, L.: System identification toolbox™ user's guid, Matlab R2014b. pp. 292
20 Bhatt, N.: 'Synchrophasor-based power system dynamic modeling for stability estimation'. Technical Report 1026456, EPRI, Palo Alto, CA, 2012.
21 Sindelar, R., Babuska, R.: 'Input selection for nonlinear regression models', *IEEE Trans. Fuzzy Syst.*, 2004, **12**, (5), pp. 688–696
22 Poncet, A., Moschytz, G.S.: 'Optimal order for signal and system modeling'. IEEE Int. Symp. on Circuits and Systems, 1994, pp. 221–224
23 Liu, Y., Sun, K., Liu, Y.: 'A measurement-based power system model for dynamic response estimation and instability warning', *Electr. Power Syst. Res.*, 2008, **124**, pp. 1–9
24 Trudnowski, D.J.: 'Estimating electromechanical mode shape from synchrophasor measurements', *IEEE Trans. Power Syst.*, 2008, **23**, (3), pp. 1188–1195
25 Sun, K., Hur, K., Zhang, P.: 'A new unified scheme for controlled power system separation using synchronized phasor measurements', *IEEE Trans. Power Syst.*, 2011, **26**, (3), pp. 1544–1554
26 Dong, J., Xia, T., Zhang, Y., Weeks, T., Thorp, J.S., Liu, Y.: 'Monitoring power system disturbances at the distribution level'. IEEE PES General Meeting, Pittsburgh, Pennsylvania, USA, 2008
27 Ye, Y., Matthew Gardner, R., Liu, Y.: 'Oscillation analysis in western interconnection using distribution-level phasor measurements'. IEEE PES ISGT, Washington, DC, USA, 2012
28 Zhang, Y., Markham, P., Xia, T., *et al*.: 'Wide-area frequency monitoring network (FNET) architecture and applications', *IEEE Trans. Smart Grid*, 2010, **1**, (2), pp. 159–167