Input Signals Selection for Measurement-based Power System ARX Dynamic Model Response Estimation

Feifei Bai
Dept. of Electrical Engineering
Southwest Jiaotong University
Dept. of Electrical Engineering and Computer Science
The University of Tennessee at Knoxville
fbai1@utk.edu

Yilu Liu
Kai Sun
Dept. of Electrical Engineering and Computer Science
The University of Tennessee at Knoxville
liu@utk.edu
kaisun@utk.edu

Navin Bhatt
Alberto Del Rosso
Evangelos Farantatos
Electric Power Research Institute
nbhatt@epri.com
adelrosso@epri.com
efarantatos@epri.com

Xiaoru Wang
Dept. of Electrical Engineering
Southwest Jiaotong University
xrwang@home.swjtu.edu.cn

Abstract—**This paper proposes a measurement-based approach to optimize the inputs of Auto-Regressive with eXogenous input (ARX) model identification in large power systems. Correlation Coefficient Index (CCI) is defined in this paper and Correlation Coefficient Map (CCM) is developed for the US Eastern Interconnection (EI) to show the correlation between any two power system output measurement signals visually. This approach is verified with EI system simulation data and applied to Frequency Disturbance Recorder (FDR) measurement data to estimate system dynamic response. The verification result shows that the number of ARX model inputs can be decreased and the estimation accuracy can be ensured by using the proposed approach.**

Index Terms—**Correlation coefficient index, correlation coefficient map, measurement-based, system identification, ARX model, dynamic response estimation.**

I. INTRODUCTION

Assessing power system dynamics is essential for safe and reliable operation and control of a power system. Synchrophasors provide rich information of system dynamic behavior in real time. Such information can be used to build purely measurement-based system models for dynamic assessment and control. System identification is a good approach to model the behavior of a dynamic system based on measurement data [1]. A linear power system dynamic model in the ARX structure is proposed in [2] to estimate the system dynamic response only based on measurement data. The model in [2] is simple and effective to estimate the system dynamic response, but it ignores one important aspect, namely: how to choose the input signal from hundreds of measurements for the developed model. Input signal design is a critical aspect for system identification since it directly affects the model accuracy [3]. Electrical distance is considered as an observability index to select the input for the ARX model in [4], but this method is based on the detailed EI system model. It is hard to calculate the electrical distance only using measurement data.

In this paper, CCI is defined and a new algorithm is developed to design the optimal input for the ARX dynamic model only using measurement data. Further, CCM has been developed to show the correlation between all the possible model inputs and outputs in a visual way for the EI system. Both CCI and CCM can be updated in real time. The algorithm is verified with EI simulation data and applied to estimate dynamic response by using FDR measurement data.

This paper is organized as follows. Section II gives a brief introduction of the ARX model. The ARX model input signal selection algorithm is developed in Section III. Section IV and V provide algorithm verification and application, respectively. Conclusion and future work are provided in Section VI.

II. MODEL IDENTIFICATION

A. ARX Model Structure

This ARX model provides a much simpler identification model of multi-variable system than the state-space model or other models [1]. 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 $\overline{y}_k$ are the model input and output, respectively. The model parameters of a multi-variable ARX model can be estimated by a linear least square technique. The least-squares estimation problem is solved by using QR factorization to optimize the ARX model parameters. The ARX structure and algorithm details are described in [1].

B. ARX Model Accuracy Index

To evaluate the identified ARX model, a fitness criterion can be performed [4]:

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

where Y_i, $\hat{Y}_i$, $\overline{Y}_i$ 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. ARX MODEL INPUT SIGNAL SELECTION ALGORITHM

Based on the synchrophasor-based power system, the coherency function is used to find the relationship among generators or bus measurement signals. Mode shape identification [5] and inter-area oscillation mode analysis [6] have already been performed using the frequency domain coherency function. In this paper we use the inverse fourier transform of coherency function to obtain the correlation coefficient function in time domain between any two power system output measurement signals.

A. Coherency Function

The measurement-based coherency function is defined [6]:

$$\gamma_{xy}^2(f) = \frac{\left|S_{xy}(f)\right|^2}{S_{xx}(f)\,S_{yy}(f)} \qquad \gamma_{xy}^2 \le 1 \tag{5}$$

where $S_{xy}(f)$ is the cross-spectral density (CSD) function between measured signals $x(t)$ and $y(t)$, $S_{yy}(f)$ is the power-spectral density (PSD) of signal $y(t)$. The coherence function gives the linear correlation between two power system output signals as a function of the frequency.

B. Correlation Coefficient Index

Consider there are two power system output measurement signals: $x(t)$ and $y(t+\tau)$, 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:

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

$$\begin{aligned}
C_{xy}(\tau) &= E\left[\{x(t)-\mu_x\}\{y(t+\tau)-\mu_y\}\right] \\
&= \lim_{T\to\infty}\frac{1}{T}\int_0^T \{x(t)-\mu_x\}\{y(t+\tau)-\mu_y\}dt \\
&= R_{xy}(\tau) - \mu_x\mu_y
\end{aligned} \tag{7}$$

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

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

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

If $\mu_x = 0$, $\mu_y = 0$, we will obtain:

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

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

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

If we do the invers fourier transform to (5) and using (11) and (12), we can obtain the correlation function in time domain:

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

Using (6), (8), and (13), the correlation coefficient function can be obtained as follows:

$$\begin{aligned}
\rho_{xy} &= \frac{R_{xy}(\tau) - \mu_x\mu_y}{\sqrt{\left[R_{xx}(0)-\mu_x^2\right]\left[R_{yy}(0)-\mu_y^2\right]}} \qquad \left|\rho_{xy}\right| \le 1 \\
&= \frac{C_{xy}(\tau)}{\sqrt{C_{xx}(0)C_{yy}(0)}}
\end{aligned} \tag{14}$$

where ρ_{xy} ranges from -1 (complete linear inverse correlation) to +1 (complete linear direct correlation). $\rho_{xy} = 0$ means lack of linear interdependence. The sign of ρ_{xy} indicates the direction of correlation: $\rho_{xy} < 0$ implies inverse correlation, i.e., the two power signals are out of phase or oscillate with

each other. $\rho_{xy} > 0$ implies direct correlation, i.e. a tendency of both signals to have similar value with the same sign.

From the derivation of (6) to (14), we can see that if the mean value of the measurement is zero, (5) and (14) are equivalent. Therefore, the coefficient function is reasonable to indicate the relationship between any two measurements.

CCI matrix is defined:

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

where r_{ij} is the correlation coefficient between any two arbitrary measurement signals:

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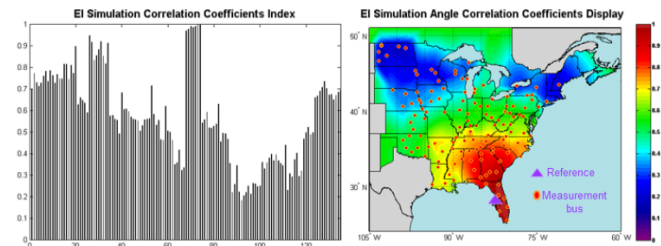
C. Correlation Coefficient Map

For a large power system, it is hard to read the correlation coefficient between any two measurement signals. CCM is developed with the Matlab mapping toolbox to show the correlation in a visual way. In CCM, we use different colors for the correlation coefficient values. The color bar on the right of the map shows the coefficient value, while the numerical value of the correlation coefficients can be found on the left part of the map. In the map, we choose one measurement bus as the reference bus and show the correlation coefficients between this reference bus and other measurement buses clearly at the same time.

Figure 1 shows the CCM and CCI for the voltage magnitude, voltage phase angle, and frequency of EI system with the simulation data. We assume that there are 135 Phasor Measurement Unites (PMUs) well-distributed in the EI system and the simulation results from the 135 buses are considered as measurement signals. The event is 838 MW Generation trip in the east of Alabama.
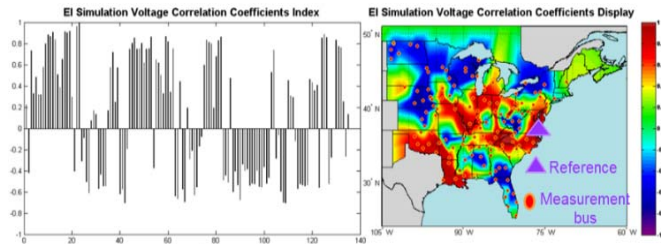
(a) Voltage magnitude

(b) Voltage phase angle
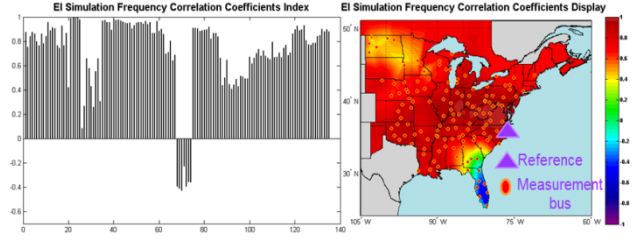
(c) Frequency

Figure 1. Correlation coefficient index and correlation coefficient map

D. ARX Input Signal Selection Algorithm Flowchart

In general terms, all the power system output signals measured by PMUs are selectable ARX model inputs. However, the key issue is how to identify a proper signal to obtain an ARX model that accurately represents dynamic system response. The algorithm developed to find the optimal inputs among hundreds of measurement signals for ARX model is described by the flowchart in Figure 2.
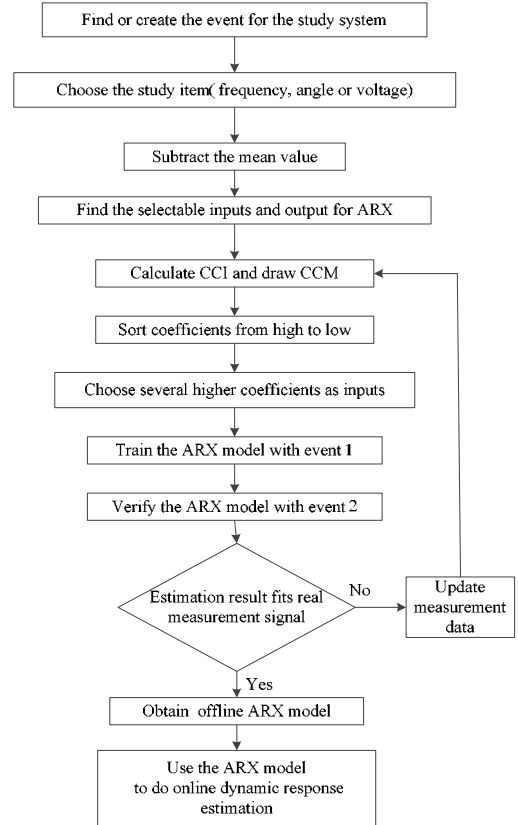

Figure 2. ARX model input signals selection flowchart

IV. ALGORITHM VERIFICATION WITH EI SIMULATION DATA IN DYNAMIC RESPONSE ESTIMATION

A. Concept of Dynamic Response Estimation

The Energy Management System (EMS) plays an important role in monitoring and control of power system, and state estimation forms its backbone. However, power system state estimation, in its current application status, is not capable

of capturing the system dynamic behaviors (except for some very slow system changes). It only provides monitoring information in the form of a sequence of steady states, or quasi-steady state [7], [8]. Dynamic state estimation is to predict the state vector one time step ahead and has the potential to foresee potential contingencies and security risks [9], [10]. Unlike the traditional state estimation and dynamic state estimation that focus on estimating relatively stationary state vectors, in this ARX dynamic response estimation, we detrend the DC component to identify the real-time dynamic model and use the obtained model to estimate the power system dynamic response [1].

B. Algorithm verification

The test system of EI used in this paper has about 3000 generators and 16000 buses. The frequency and voltage phase angle signals sampled at 0.001s for 20s are used as the measurement signals. We assume that there are 135 buses measured with PMUs and two events happened at different times and different locations: event 1 is 814MW generation trip in Southeast of Florida, as shown in figure 3; event 2 is 838MW generation trip in East of Alabama. Event 1 is used to train the ARX model and event 2 is used to perform the verification of the proposed algorithm. With the training frequency data we can obtain the CCI and CCM as shown in figure 3. Figure 4 shows the frequency dynamic response estimation results with this algorithm.
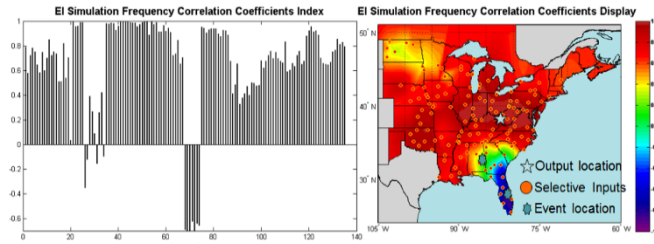

Figure 3. Frequency CCI and CCM

(a) Top 4 and lowest 4 coefficient signals as ARX model inputs

(b) Top 4 coefficient signals as ARX model inputs

(c) Top 1 coefficient signal as ARX model inputs

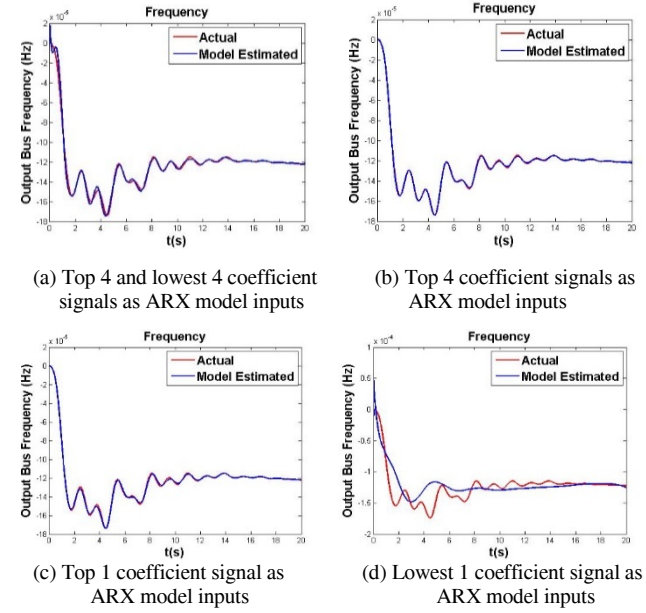(d) Lowest 1 coefficient signal as ARX model inputs

Figure 4. ARX input signals design algorithm verification for frequency dynamic response estimation

CCI and CCM are obtained with the training frequency and voltage phase angle data, which are shown in figure 5. Figure 6 shows the phase angle dynamic response estimation result with the proposed algorithm to select the ARX model inputs.
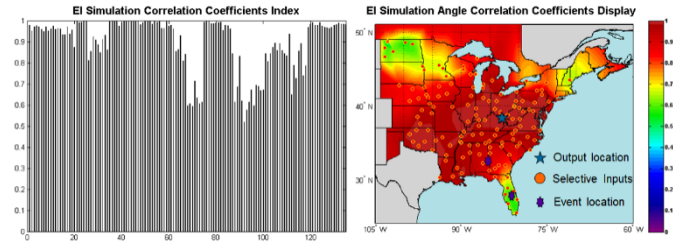

Figure 5. Voltage phase angle CCI and CCM

(a) Top 4 and lowest 4 coefficient signals as ARX model inputs

(b) Top 4 coefficient signals as ARX model inputs

(c) Top 1 coefficient signal as ARX model inputs

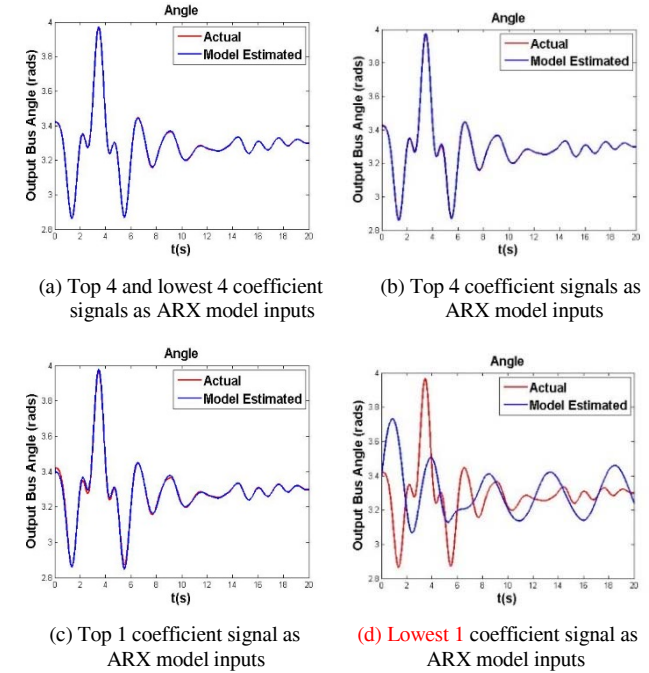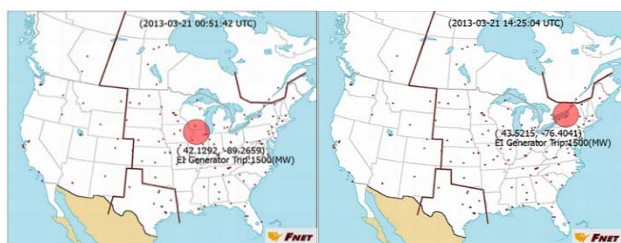(d) Lowest 1 coefficient signal as ARX model inputs

Figure 6. ARX input signals design algorithm verification for voltage phase angle dynamic response estimation

For the model identification, the more input signals are selected, the better to reflect the system dynamics of the identified model. Therefore, in Fig 4-a) and Fig 6-a), top 4 and lowest 4 correlation signals are roughly stand for all the selectable input signals, the estimation results are shown in Fig 4-a) and Fig 6-a). In Fig 4-b) and Fig 6-b), only top 4 correlation signals are selected as the input signals, the estimation results shows that although only top 4 correlation signals are selected as the input signals, the estimation result is as good as top 4 and lowest 4 correlation signals. Furthermore, the results also prove that only the strong correlation signals are the most important signals for the model identification and contain most of the system dynamic information. From the above comparisons between actual and model-estimated dynamic responses of the ARX output, it is obvious that when the input signals with high correlation coefficients are selected, the model-estimated responses fit the actual ones very well. However, if the input signals with very low correlation coefficients are selected, the model-estimated response does not fit the actual response. Therefore, only the

higher correlation coefficients are needed as the ARX input signals. If top 1 correlation coefficient of the input signals are found, only one input signal can obtain a very good estimation, which can decrease the complexity of ARX model.

## V. Algorithm Application in Real Dynamic Response Estimation

The north of American power Frequency Monitoring Network (FNET) is a wide-area measurement system that takes high accuracy, GPS synchronized measurements at standard end-user distribution voltages [11]. As a member of the PMU family, the FDRs used in the FNET system measure frequency, voltage phase angle at standard 120V outlets and transmit these measurement data through the Internet [12]. About the FDR measurement data, the frequency and phase angle signals sampled at 0.1s for 60s are used to calculate the CCI. There are two events monitored by FDR shown in figure 7 with 1500MW generation trip. We use event 1 to calculate the CCI and train the ARX model, then CCI is applied to the dynamic response estimation with event 2. The application results of the proposed approach are shown in figure 8 and figure 9.

(a) Event 1   (b) Event 2
Figure 7. Event location
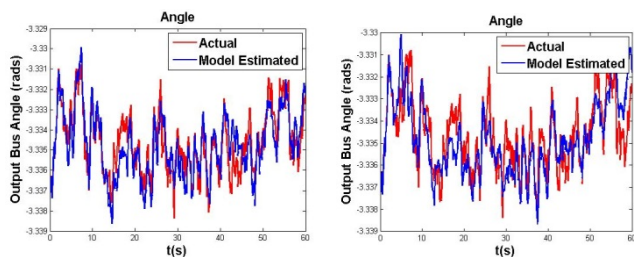
(a) Top 4 coefficient signals as ARX model inputs   (b) Top 1 coefficient signal as ARX model inputs
Figure 8. Voltage Phase Angle
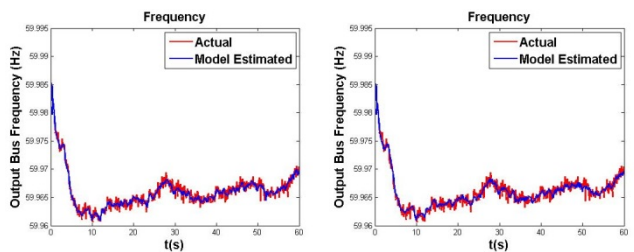
(a) Top 4 coefficient signals as ARX model inputs   (b) Top 1 coefficient signals as ARX model inputs
Figure 9. Frequency

The accuracy indices are 45 and 30 for figure 8 (a) and figure 8 (b), respectively. While the accuracy indices are 80 and 78 for figure 9 (a) and figure 9 (b), respectively. Which

means, for frequency, only one input signal is enough to get good estimation, however, for voltage phase angle, we need at least 4 input signals.

VI. Conclusion and Future Work

A measurement-based approach is proposed to select optimal input signals for ARX model. This approach works well with the real measurement data from FDRs in the EI system. With this approach, the strong correlated input signals for dynamic response estimation can be quickly identified. As a consequence, and the complexity of the ARX model can be decreased. A Correlation Coefficient Map has been developed for the Eastern Interconnection system to visually display the Correlation Coefficient Index. Next step of this work, includes comprehensive robustness test of the proposed approach with additional measurement data, and implementation in actural power system.

References

[1] Lennart Ljung, System identification: theory for the user (2nd Edition), New Jersey: PTR Prentice, 1999, pp1-296.

[2] 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.

[3] RAMAN K. MEHRA, "Optimal inputs for linear system identification," *IEEE Trans. Automatic Control*, vol. 19, pp. 2-7, Jun. 1974.

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

[5] D. J. Trudnowski, "Estimating electromechanical mode shape fromsynchrophasor measurements, " IEEE Trans. Power Syst., vol. 23, no. 3, pp. 1188-1195, August 2008

[6] Kai Sun,"Early warning of wide-area angular stability problems using synchrophasors,"in San Diego. 2012 IEEE PES Power and Energy Society General Meeting, pp. 1-6.

[7] Ding, Jun-ce, et al., "An Overview of State Estimation Based on WideareaMeasurement System," Automation of Electric Power Systems, Vol.30, No.7, Apr.10, 2006.

[8] R. Zivanovic, C. Cairns, "Implementation of PMU technology in stateestimation: an overview," 1996 IEEE Africon 4th africon conference inafrica, pp: 1006-1011.

[9] G. Durgaprasad, S.S. Thakur,"Robust dynamic state estimation of power systems based on M-Estimationand realistic modeling of system dynamics," Power Systems,IEEE Trans. on, Vol. 13 , pp. 1331-1336, 1998.

[10] E. Ghahremani, I. Kamwa, "Dynamic State Estimation in Power Systemby Applying the Extended Kalman Filter With Unknown Inputs toPhasor Measurements," Power Systems, IEEE Trans. on, Vol. 26, Iss. 4, pp. 2556–2566, Nov. 2001.

[11] Z. Zhong, C. Xu, B. J. Billian, L. Zhang, S.-J. S. Tsai, R. W. Conners,V.A. Centeno, A. G. Phadke, and Y. Liu, "Power System FrequencyMonitoring Network (FNET) implementation," IEEE Transactions onPower Systems, vol. 20, pp. 1914-1921, 2005.

[12] Yanzhu Ye, R. Matthew Gardner, and Yilu Liu, "Oscillation analysis in western interconnection using distribution-level phasor measurements," in Washington. 2012 IEEE PES ISGT,pp.1-8.

Dear Editors and Reviewers:

Thank you concerning our paper entitled "Input Signals Selection for Measurement-based Power System ARX Dynamic Model Response Estimation" (ID:14TD0513). The comments are all valuable and very helpful for revising and improving our paper. We have studied comments carefully and made corrections that we hope could meet with approval. Revised portion are marked in red in the paper. The main correlations in the paper and the responds to the reviewers' comments are as following:

Responds to the reviewer's comments:

Reviewer1:
Comments:
In general, this paper proposes a good tool (based on the index CCI and CCM) to select the inputs signals of ARX structure to estimate a linear power system dynamic model. In large power systems, this index can help to determine between different sampled signals obtained by PMUs, those that will result in accurate models. However, the paper presents some errors that must be corrected. Such mistakes are listed below:
1. In all the text, there are several typos in the spacing of words. Such mistakes should be carefully reviewed.

We are very sorry for our negligence of the typos in the spacing of words, the mistakes have been corrected in the revised version.

2. On Section IV-B:
In Fig 4-a), the correct legend is "Top 4 coefficient signals as ARX model inputs" since in the figure it is possible to see just the comparison between the estimated and the actual model. In Fig 6-a) First, the legend of Fig 6 is also wrong, I think the correct is just "Top 4 coefficient signals as ARX model inputs". Besides this, in this case, I would like to know if the same figure was used incorrectly in cases a) and b) or if the accurate model is the same with the use of Top 4 coefficient signal or Lowest 4 coefficient signals as ARX model inputs.

We are very sorry for our negligence of explaining the results in these Figures. The legends of Fig 4-a) and Fig 6-a) in the paper are correct. For the model identification, the more input signals are selected, the better to reflect the system dynamics of the identified model. Therefore, in Fig 4-a) and Fig 6-a), top 4 and lowest 4 correlation signals are roughly stand for all the selectable input signals, the estimation results are shown in Fig 4-a) and Fig 6-a). In Fig 4-b) and Fig 6-b), only top 4 correlation signals are selected as the input signals, the estimation results shows that even only top 4 correlation signals are selected as the input signals, the estimation result is as good as top 4 and lowest 4 correlation signals. Furthermore, the results also prove that only the strong correlation signals are the most important signals for the model identification and contain most of the system dynamic information.
We have added the explanation in Section IV-B with red words.

3. On Section V

The Fig. 8 and 9 were not referenced in the text.
The Figs. 8 and 9 are referenced in the last paragraph of Section V now.

Furthermore, in Fig 9, it is not possible to verify any difference in the precision of the model estimated in the case a) and case b). Are these figures correct? I believe that in these cases the ARX Model Accuracy Index should also be presented in the results.

We are very sorry for the incorrect description of the time on the x-axis and it has been corrected.
And the accuracy indices are added in the last paragraph of Section V according to the reviewer's suggestion.

4. The figures in the text should be bigger to facilitate the interpretation.

Thanks for your suggestion, but the pages of the paper are limited to 5, so I have to make the figures smaller.
Thanks for your good comments.

Reviewer 2:

Comments: The authors propose an approach to select optimal input signals for the identification of an ARX model.

The paper presents some contribution to the problem and the results seem to confirm the potential of the proposed approach.
It seems that in equations (6), (7) and (12) it should be Rxy(tau) and Cxy(tau).

We are really sorry for the mistakes in equations (6), (7) and (12). We have made correction according the reviewer's suggestion.
Thanks for your good comment and approval of the work in the paper.

Above all, according to reviewers' comments, the paper has been revised point-by-point. Thank you and the reviewers again!