Measurement-based Power System Frequency Dynamic Response Estimation Using Geometric Template Matching and Recurrent Artificial Neural Network

Feifei Bai, Student Member, IEEE, Xiaoru Wang, Senior Member, IEEE, Yilu Liu, Fellow, IEEE, Xinyu Liu, Yue Xiang, Student Member, IEEE, Yong Liu, Member, IEEE

Abstract—Understanding power system dynamics after an event is essential for online stability assessment and control applications. Wide area measurement systems (WAMS) based on synchrophasors 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 measurements can be inaccessible for various reasons, e.g., most notably communication failure. To reconstruct the inaccessible measurements, radial basis function artificial neural network (RBF-ANN) is used to estimate the system dynamics in this paper. In order to find the best input features of the RBF-ANN model, geometric template matching (GeTeM) and quality-threshold (QT) clustering are employed from the time series analysis to compute the similarity of frequency dynamic responses in different locations of the power system. The proposed method is tested and verified on the bulk system Eastern Interconnection (EI) transmission system in United States and the results obtained indicate that the proposed approach provides a compact and efficient RBF-ANN model that can accurately estimate the inaccessible frequency dynamic responses considering different operating conditions with fewer inputs.

Index Terms—Artificial neural network, clustering, dynamic response estimation, geometric template matching, radial basis function.

I. INTRODUCTION

Wide area measurement systems (WAMS) based on synchrophasors, e.g., phasor measurement units (PMUs) [1] and frequency disturbance recorders (FDRs) [2], can continuously provide high-sampling-rate measurements, which enable to make the real-time monitoring of wide-area power system dynamics visible [3]. These measurement data contain important pieces of information to help power system researchers and operators understand the dynamics of the power grid. Many studies have been carried out based on the real measurement data of the power grid, e.g., transient stability analysis [4], frequency prediction [5], and inter-area oscillation mode identification [6]. However, some issues still need to be solved: firstly, 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; secondly, some PMUs/FDRs may lose partial measurements due to the communication failure between PMUs and the central operator that could make the system unobservable; and third, some PMUs/FDRs may deliver bad data because of cyber attacks, thus causing wrong operating decisions [7-8]. The dynamic information in these conditions must be accurately estimated to enhance system stability and reliability.

Some research has been done to estimate the inaccessible measurements [9-10]. Gao et al. [9] formulated the missing PMU data reconstruction problem into a low-rank matrix completion problem, but the method only works for problems that lose some data.
points. Traditional state estimators usually generate pseudo-measurements from either historical data or context of the generation of pseudo-measurements to replace the missing data [10]. Static state estimation, which is the most widely studied does not capture the dynamics very well after disturbances [10-11]. Therefore, the dynamic state estimation is studied to predict or estimate the dynamic states whenever an operation suddenly changed [12-13]. However, these methods depend on the system circuit model, which may not be adequately accurate because of the power system operating condition and topology changing frequently (although usually not dramatically). To avoid these shortcomings, measurement-based approaches have been proposed: a linear model auto-regressive with exogenous input (ARX) was used to estimate the dynamic response in [14-16], which is a simple and efficient estimation approach during small disturbances. Considering the nonlinearity of the power system, a nonlinear model should be proposed to adapt to the nonlinearity of the transient signals.

The artificial neural network (ANN) has had a wide area of applications in solving difficult issues in power system analysis because of its following advantages: it can be fully close to the complex relationship between non-linear mappings; it can learn and adapt to the dynamic nature of an uncertainty system; and it has self-learning ability[17-22]. Three fast learning ANNs: the radial basis function (RBF) network, the progressive learning network (PLN), and the self-organizing map (SOM) are compared in fast voltage prediction in power system dynamic analysis and the most accurate prediction method being RBF-ANN[23]. Most previous studies focus on transient stability assessment [24-26] and dynamic state estimation (DSE) [27-28]. Some system dynamic response estimation work[23,29] use the measurement information along with the knowledge of the system physical models. When multi-layer feed forward ANN models were used to estimate the generator rotor speed in a two machine system, the estimation comparison results show that the measurement-based model is better than the classical circuit-based model[30].

On the other hand, most of the previous estimations [30-31] have focused on small systems, which did not involve lots of candidate input locations for the construction of the estimation model. A large power system indicates a great number of potential input locations for the estimation model construction. The number of input locations actually used by the model must be reduced to an acceptable level for the sake of the model updating speed. Therefore, input location selection is a critical aspect of system identification since it directly affects the accuracy and complexity of the model[32]. [15] proposed an linear index to measure the similarity of the transient signals to aid in choosing the input locations of the estimation model. This paper proposed a nonlinear input signal selection approach utilizing geometric template matching (GeTeM)[33] and quality-threshold (QT) clustering[35] in order to grasp the similar response by extracting the dynamic characteristics from measured transient signals in large power systems.

In this paper, a nonlinear clustering index based on GeTeM is proposed to select the input features for model constructions, which is proposed based on RBF-ANN by making use of measurements for frequency dynamic response estimation. The RBF-ANN performs fast training by exploiting the linear separability principle, which will benefit the online application.

The rest is organized as follows. The methodology of the input selection index is proposed in Section II. Section III introduces the RBF-ANN estimation framework. The case study and verification of the proposed approach are shown in Section IV. The application of the proposed estimation method with the actual FDR measurement data is carried out in Section V. The conclusion is provided in Section VI.

II. METHODOLOGY OF THE INPUT SELECTION INDEX

Time series analysis and a clustering algorithm has been applied in a power system for grouping of coherent generators, e.g., principal component analysis (PCA)[36], independent component analysis (ICA)[37], and dynamic time warping (DTW)[38] based on the feature extraction method to group similar perturbed trajectories. All methods mentioned above attempt to match shapes or motifs by looking at the raw data or some projection thereof. These kind of techniques can be called pattern matching[33]. An alternative viewpoint is to consider the raw data as an observation from the underlying dynamical system, and to compare properties of the systems that generate the different signals. In this paper, GeTeM based on Time-delay embedding is proposed to measure similarity of frequency dynamic responses after an event. QT is then used to cluster.

A. Geometric Template Matching

GeTeM[33] based on time-delay embeddings is a nonlinear dynamical system model. The idea is to reconstruct the state and dynamics of a dynamical system from measurements utilizing time delay coordinates.

The measured bus frequency dynamic response can be looked as a time series \(x\) consisting of \(T\) univariate measurements by the same time interval sampling:
Each \( x(i) \) can be reconstructed by the phase space reconstruction technology\[34\]:

\[
X(t) = \left[ x(t), x(t + \tau), \ldots, x(t + (m-1)\tau) \right]^T
\]

where \( t = 1, 2, \ldots, M; \ M = T(m-1)\tau, \ m \) is the embedding dimension, \( \tau \) is the time-delay, \( m \) and \( \tau \) were chosen using the nonlinear multiple autocorrelation functions method\[43\]. Each \( X(t) \) is:

\[
X(1) = \left[ x(1), x(1+\tau), \ldots, x(1+(m-1)\tau) \right]
\]

\[
X(2) = \left[ x(2), x(2+\tau), \ldots, x(2+(m-1)\tau) \right]
\]

\[ \vdots \]

\[
X(M) = \left[ x(M), x(M+\tau), \ldots, x(T) \right]
\]

Then the time-delay reconstruction of \( x(t) \) is given by the following sequence of points in \( R^n \):

\[
u = \{ X(1), X(2), \ldots, X(M) \} = \{ X(t) \}_{t=1}^M
\]

\( u \) is the reconstructed model for the sequence \( x(t) \).

Based on (1) to (3), GeTeM starts by building models \( u_1 \) and \( u_2 \) for measured bus frequency signals \( x_1(t) \) and \( x_2(t) \), respectively, using same parameters \( m \) and \( \tau \).

\[
u_1 = \{ X_1(1), X_1(2), \ldots, X_1(M) \} = \{ X_1(t) \}_{t=1}^M
\]

\[
u_2 = \{ X_2(1), X_2(2), \ldots, X_2(M) \} = \{ X_2(t) \}_{t=1}^M
\]

Let \( n_{1j}, \ldots, n_{ik} \) be the indices in \( X_1(t) \) of the \( k \) nearest neighbors \[42\] of \( X_1(t) \) in terms of Euclidean distance, that is,

\[
 n_{1j} = \arg \min_{j \neq n_{1j}} \| X_1(i) - X_1(j) \|
\]

\[
 \ldots
\]

\[
 n_{ik} = \arg \min_{j \neq n_{ik}} \| X_1(i) - X_1(j) \|
\]

Let \( X_1(i) = k^{-1} \sum_{j=n_{1j}}^{n_{1j}+1} X_1(n_{1j}) \) and \( X_1(i+1) = k^{-1} \sum_{j=n_{1j}+1}^{n_{1j}+1} X_1(n_{1j}+1) \) be the mean of the \( k \) nearest neighbors of \( u_1 \) in \( u_1 \) and the mean of the subsequent points in \( u_1 \), respectively.

The similarity between models \( u_1 \) and \( u_2 \) is defined as:

\[
S(X_1, X_2) = \frac{1}{M-1} \sum_{i=1}^{M-1} \frac{\left( \sum_{j=n_{1j}}^{n_{1j}+1} X_1(i+1) - X_1(i) \right) \left( \sum_{j=n_{1j}+1}^{n_{1j}+1} X_1(i+1) - X_1(i) \right)}{\max \left( \left( \sum_{j=n_{1j}}^{n_{1j}+1} X_1(i+1) - X_1(i) \right) \left( \sum_{j=n_{1j}+1}^{n_{1j}+1} X_1(i+1) - X_1(i) \right) \right)}
\]

Finally, the distance of each two measured signals are defined as:

\[
d(X_1, X_2) = \exp(-S(X_1, X_2))
\]

where the value of \( d \) is \((0,1)\).

B. QT Algorithm

The goal of QT clustering \[35\] is to form large clusters with similar patterns and ensure a quality guarantee for each cluster. It does not require the specified number of the class since cluster quality is defined by a pre-defined threshold radius specified in terms of a similarity measure. Moreover, the QT algorithm is a kind of deterministic algorithm. The basic process is as follows:

1. Give a radius \( d \) of clustering.
2. Creating a candidate cluster for each data set. The candidate cluster includes all data points within the radius.
3. The candidate cluster that has the most data will be retained as the first cluster and signals grouped in this cluster will be removed from the data set.
4. The procedure is iteratively repeated with the reduced set of signals until all signals are clustered.

The clustering radius is used to determine whether two data points will be regarded as in the same class or not. Given the distance \( d(x, y) \) between data point \( x \) and \( y \), and clustering radius \( r_0 \), if \( d(x, y) < r_0 \), then two signals will be in the same class. The signals in the same class are selected as the candidate input signals for the estimation model construction.

III. RBF-ANN FOR DYNAMICAL SYSTEM ESTIMATION

The nonlinear models can express and provide robust and accurate estimation/prediction results of the power system dynamics compared with the linear models, especially in the case of the dramatic change of the power system data and topology. In this paper, the Multi-Input-Single-Output (MISO) RBF-ANN model is constructed based on GeTeM and QT clustering to estimate system dynamic response under different events.

A. Radial Basis Function Artificial Neural Network

The main advantage of ANN is the ability to learn complex non-linear relationships and their modular structures, which allows parallel processing\[39\]. Thus, a RBF-ANN topology is employed to estimate the frequency dynamic response after events occur.

The generic three-layer structure of MISO RBF-ANN\[40\] is shown in Fig. 1, which performs fast training by exploiting the linear separability principle. The input vectors are transformed in vectors of an \( n \)-dimensional space by the \( n \) non-linear units (called bases) of the hidden layer. If the conditions of the linear separation are reached, the weights of the output layer are easily computable by linear regression. Therefore, the relationship between the input and output could be approximated by a linear combination of non-linear functions.

![Fig.1. MISO RBF artifical neural network structure](image)
In the RBF-ANN models, the Gaussian function is selected as the activation function:
\[
\phi_i(x_p-c_i) = \exp \left(-\frac{1}{2\sigma^2_i} \|x_p-c_i\|^2 \right)
\]  
(8)
where \(x_p\) is the \(p^{th}\) model input, that is the measured signal by PMU/FDR, \(c_i\) is the \(i^{th}\) hidden neuron center vector, \(\sigma_i\) is the \(i^{th}\) hidden neuron standard deviation (width), \(\phi(.)\) is the Gaussian function. In each hidden neuron, the unit can be described as shown in Fig. 2, where \(w_i\) the weight connected to the hidden layer. \(\|\text{dist}\|\) is the distance between the input vector and weight vector.

![RBF neuron](image)

The output of the MISO model is:
\[
y = \sum_{i=1}^{h} \omega_i \phi_i(x_p-c_i) \quad i=1\ldots h
\]  
(9)
where \(\omega_i\) is the weight connecting the output node to the \(i^{th}\) hidden neuron, \(h\) is number of hidden neurons. The Gaussian function centers \((c_i)\) are initialized using \(k\)-means clustering algorithm [41].

The Gaussian function widths \((\sigma_i)\) are initialized according to:
\[
\sigma_i = \frac{C_{\max}}{\sqrt{2h}} \quad i=1\ldots h
\]  
(10)
\[
c_{\max} = \text{MAX} \{\|c_i-c_k\|\}
\]  
(11)

The weight \((\omega_i)\) is initialized using the linear quadratic (LQ) method as:
\[
\omega_i = \exp \left(\frac{h}{C_{\max}} \|x_p-c_i\|^2 \right) \quad p=1\ldots P, i=1\ldots h
\]  
(12)

**B. RBF ANN Input Feature Selection**

The signals from interested locations can be chosen as RBF-ANN output signals during neural network training. The input features are selected by the approach described in Section II. The goal of the input signal selection is to reduce the complexity of the RBF-ANN complexity and improve the accuracy of the ANN.

**C. RBF-ANN Training**

In order to establish a good estimation network, RBF-ANN training is an important step. Different kinds of events are created to generate the input-output data pairs as ANN training and testing samples. In this paper, 10s samples after an event are used. During the training process, \(\sigma_i\) are adjusted according to minimum sum-squared error of the estimation.

**D. Signal Detrending**

Trend in a time series is a slow, gradual change in some property of the series over the whole study time window. For system identification, all signals should be detrended. Zero mean filter is used in this paper for practical application. It is defined as
\[
y(k) = y'(k) - E(y'(k))
\]  
(13)
where \(y(k)\) is the orginal signal, \(E(y'(k))\) is the mean value of \(y'(k)\), and \(y(k)\) is the detrended signal.

For dynamics study, we focus on the dynamics of original signal \(y'(k)\) instead of detrended signal \(y(k)\). To recover the original signal from detrended signal, the inverse form filter is
\[
y'(k) = y(k) + E(y'(k))
\]  
(14)

**E. Error Index**

The target of training a neural network is to determine the optimal estimation of frequency response that yields the minimum error over training input-output pairs. The estimation error could be measured as a function of deviation between the actual response and the estimated output. Mean square error (MSE) is considered as the performance criterion:
\[
\text{MSE} = \frac{1}{N} \sum_{k=1}^{N} (y(k) - y'(k))^2
\]  
(15)
where \(N\) is the number of training samples, \(y'(k)\) is the target actual response and \(y(k)\) is the corresponding estimated output.

**F. Frequency Dynamic Response Estimation**

The following are the major steps to realize the frequency dynamic response estimation scheme:

Step 1) Select 10s to 20s signals of bus frequency dynamic response after an event occurs.

Step 2) Set \(m\), \(\tau\) and \(d\) for GeTeM and QT clustering. Use the proposed approach described in Section II to cluster the bus frequency dynamic responses to some distinct groups.

Step 3) Define the interested location as the fixed location, where is the output of the RBF-ANN as well. Then locate the clustered group where the fixed location belong. Choose first 2 to 4 signals in the group as the input of the RBF-ANN model.

Step 4) Detrend the selected input and output signals in Step 3). In order to obtain a wide-sense stationary random process, all measurement data should be detrended to remove direct-current components. A simple zero-mean filter is used in this paper.

Step 5) Train the MISO RBF-ANN model with the selected signals in Step 2) and Step 3). And the threshold of the estimation error is \(1 \times 10^{-5}\). Estimation modes with a high error index will be rejected while only the model with the acceptable
accuracy can be selected as the final model for further estimation studies.

Step 6) Verify the obtained model in Step 5) by estimating the bus frequency response at the fixed location during other event scenarios. Please note that the estimated signals need to be recovered since all the signals are detrended in Step 4).

Step 7) Apply the model to estimate the inaccessible frequency dynamic response at the fixed location.

IV. CASE STUDY WITH SIMULATED DATA

To demonstrate the efficiency of the proposed frequency dynamic response estimation method in different scenarios, the Eastern Interconnection ( EI) transmission system[14] of the United States is used as the test system, as shown in Fig. 3. There are about 3000 generators and 16000 buses in the model. 135 buses are assumed to be installed and measured by PMUs and these 135 buses are well-distributed in the EI system as marked with a dot in Fig.3. The simulation results from the 135 buses are considered as the measured signals, which are all available as the RBF-ANN input features. The interested output location is fixed with a circle as shown in Fig. 3.

A. Bus Frequency Dynamic Response Cluster

The simulation time step is set to be 0.001s and the time window is 10s. The event to train the RBF-ANN is an 838-MW generation trip occurring in eastern Alabama. Therefore, the signals to cluster bus frequency dynamic responses are the same as the signals for training. Then the QT clustering is applied on the 135 signals using threshold diameter of 0.65 and the GeTeM distance using $m=4$, $r=17$ and $k=2$. These parameters were chosen using the nonlinear multiple autocorrelation functions method[43]. Finally, 45 distinct clusters are identified as shown in Table I.

The bus frequency dynamic responses are shown in Fig.4 and the first eight large groups are plotted out in Fig.5.

<table>
<thead>
<tr>
<th>No.</th>
<th>Number of signals</th>
<th>No.</th>
<th>Number of signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13</td>
<td>9-12</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>13-20</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>21-30</td>
<td>2</td>
</tr>
<tr>
<td>4,5</td>
<td>7</td>
<td>31-45</td>
<td>1</td>
</tr>
<tr>
<td>6,7,8</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 5 First 8 groups of the cluster results. (a) Cluster 1 (13 buses); (a) Cluster 2 (11 buses); (c) Cluster 3 (9 buses); (d) Cluster 4 (7 buses); (e) Cluster 5 (7 buses); (f) Cluster 6 (5 buses); (g) Cluster 7 (5 buses); (h) Cluster 8 (5 buses).

From Fig. 5, it can be seen that each clustered group has a similar dynamic response. Bus-41 in the largest Group-1 is chosen as the fixed bus and also the output of the RBF-ANN model, then the first three buses in the Group-1 are chosen as the inputs. The width ($\sigma$) of the RBF-ANN is 1.8. The estimation model is obtained and verified in Subsection B with different event scenarios.

**B. Frequency Dynamic Response Estimation**

Numerous case studies were carried out to test the effectiveness of the RBF-ANN estimation model obtained in Subsection A. Three typical scenarios are given as follow:

Scenario I: Generation trip events occur in the same location (marked with star in the Fig.3) with different amounts, which is shown in Table II.

<table>
<thead>
<tr>
<th>Event Number</th>
<th>Amount (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>328</td>
</tr>
<tr>
<td>2</td>
<td>510</td>
</tr>
<tr>
<td>3</td>
<td>717</td>
</tr>
<tr>
<td>4</td>
<td>1149</td>
</tr>
</tbody>
</table>

Scenario II: Generation trip events occur in different locations (marked with stars in Fig.3) with similar amount, which is shown in Table III.

<table>
<thead>
<tr>
<th>Event Number</th>
<th>Amount (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,000</td>
</tr>
<tr>
<td>2</td>
<td>1,017</td>
</tr>
<tr>
<td>3</td>
<td>1,150</td>
</tr>
<tr>
<td>4</td>
<td>990</td>
</tr>
</tbody>
</table>

Scenario III: Line trip events occur in different locations (marked with triangles in Fig.3).

It can be seen that the model estimated responses match the actual ones closely with one event occurring with different amounts, at different locations, or different types as shown in Fig.6, Fig.7 and Fig.8. The verification results demonstrate that the trained RBF-ANN model can accurately describe the dynamic relationship among the measured frequency dynamic responses at the input and output locations. On the other hand, the verification results demonstrate the effectiveness of the clustering in Subsection A to select the input signals for the ANN-RBF model as well.

**V. CASE STUDY WITH REAL MEASUREMENT DATA**

The power system frequency monitoring network (FNET)[44] is a wide-area measurement system that takes high accuracy, GPS-synchronized measurements at standard end-user distribution voltages, which is shown in Fig. 9(a). As a member of the PMU family, the FDRs used in the FNET system measure frequency, voltage and phase angle at standard 120V outlets and transmit these measurements through the Internet[45]. 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 visualization, and post event analysis[46]. The FDR measurements used here are the frequency signals sampled at 0.1s. As shown in Fig. 9 (b), one disturbance monitored by FDRs was selected for training. The circle is
the detected event location. The GeTeM and QT clustering are applied to the event data as shown in Fig.10 with \( d=0.68, m\hat{=}4, \tau=17 \) and \( k=2 \), and the clustering results are shown in Table IV and the first four large groups are plotted out in Fig.11.

\[
\begin{array}{c|c|c|c|c}
\hline
\text{No.} & \text{Number of signals} & \text{No.} & \text{Number of signals} \\
\hline
1 & 14 & 5-7 & 2 \\
2-4 & 3 & 8-13 & 1 \\
\hline
\end{array}
\]

VI. CONCLUSION

This paper presents a novel idea to estimate the inaccessible measurements using a data driven model RBF-ANN to describe the relationship among measurements in different locations. The input locations selection results demonstrate that GeTeM is an effective measurement-based approach to process abundant transient signals and the complexity of the RBF-ANN model was reduced consequently by the input preselection and the model can accurately estimate the inaccessible frequency dynamic responses in different operating conditions. Furthermore, the test using the real FDR measurement indicates that the proposed approach can be implemented in actual power systems for inaccessible data estimation. Besides, the dynamic estimation method proposed in this paper is purely based on measurement, indicating that the model can be updated online.

Acknowledgement

The authors gratefully acknowledge the FNET team at the University of Tennessee for providing the phasor measurement data from FNET/GridEye for the verification work.

REFERENCES


