Method of Relationships Exploration for Online Small Signal Stability Analysis

Youping Fan*, Songkai Liu, Cheng Tian

School of Electrical Engineering, Wuhan University, Wuhan 430072, China

Abstract

This paper describes an online small signal stability (SSS) analysis method based on relationships exploration (RE) in a large data set for power systems. The proposed method can explore the relationships between operation variables and the SSS and find the optimal variables as input parameters for SSS analysis. Every relationship will be given score by maximal information coefficient (MIC) and Pearson correlation coefficient (PCC). The high ranked variables are selected to estimate the SSS. The SSS can be estimated online based on explored relationships if the selected variables can be obtained online from the wide area measurement system (WAMS). The method is examined on a 39-bus test system and a 1648-bus system provided by PSS/E, and the result is acceptable. Also the impacts of selected types of relationships, training data size, total number and ranks on estimation accuracy are analyzed. The robustness of the scheme to measurement errors and topology variation are studied.

Keywords: SMALL SIGNAL STABILITY, RELATIONSHIPS EXPLORATION, LARGE DATA SET, MAXIMAL INFORMATION COEFFICIENT, PEARSON CORRELATION COEFFICIENT

Nomenclature

V_i	Voltage amplitude of bus *i*.
θ_i	Voltage phase angle of bus *i*.
PL_i	Active power of loads at bus *i*.
QL_i	Reactive power of loads at bus *i*.
PG_i	Active power of generators at bus *i*.
QG_i	Reactive power of generators at bus *i*.
QS_i	Reactive power of shunts at bus *i*.
P_{i-j}	Active power from bus *i* to bus *j*.
Q_{i-j}	Reactive power from bus *i* to bus *j*.
S_{i-j}	Apparent power from bus *i* to bus *j*.
$Ploss_{i-j}$	Active power loss from bus *i* to bus *j*.
$Qloss_{i-j}$	reactive power loss from bus *i* to bus *j*.

1. Introduction

With the growth of interconnected power systems, modern power grids are operated near security limits and problems related to small signal stability occurred occasionally [1]. The SSS problem usually occurred when the damping ratio of power system is insufficient, which may result the system to be instable and experiences oscillation and even more serious consequences like splitting when small disturbance (such as load change) occurs. SSS focus on the capable of the power system to keep synchronized under small disturbances [2]. Many researches have been done in this area, and conventional time domain simulation based on system modeling has been used as the primary tool to analyze power system stability. Its analysis is normally based on the linearized system dynamic equations using modal (eigenvalue) analysis techniques, and usually its stable criterion is damping ratio (DR) [3], [4].

There are methods such as QR method [5], and more recently the BR method [6]. But they are not feasible for large-scale power system studies due to their excessively large CPU time and memory requirements. Other methods like improving small signal stability (ISSS) [7], probabilistic small signal stability (PSSS) [8], and so on, are not suitable for large power SSS analysis because of their excessive calculation. The above mentioned methods are relatively accurate, but infeasible for online SSS analysis.

This paper proposed a novel method that can estimate online SSS timely based on relationships explored between operation variables and DRs in large data set, including linear relationships and nonlinear functional relationships. Each relationship is given score by maximal information coefficient (MIC) [9] and the Pearson correlation coefficient (PCC) [10], and the high ranked variables are selected as input parameters for SSS estimation. The novel method includes three sections: offline training, model update, online estimation. In the RE process, large raw data set is created by PSS/E simulation, including all kinds of operation variables and the relative DRs. Then the relationships will be explored and given scores by MIC and PCC. After selecting the highly ranked variables as the optimal input parameters, curve fitting is applied to figure out the fitting equation between selected variables and the DRs. Finally, when coming to a new operation state, the DR can be quickly calculated through the equation with the data obtained from the WAMS, so that the SSS can be estimated online timely. The method is examined on a 39-bus test system and a 1648-bus system provided by PSS/E, and the result is accurate and acceptable.

The paper is organized as follows. Section 2 gives a detailed statement of small signal stability and mathematic methods. Section 3 introduces the integrated scheme of online SSS estimation process. The performance of the method is examined in Section 4. Conclusions are made in Section 5.

2. Problem Statement and Supporting Mathematical Methods

2.1. Small Signal Stability (SSS)

The linearized model of power system can be described as

$$\begin{cases} \dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}\tilde{y} \\ 0 = \tilde{C}\tilde{x} + \tilde{D}\tilde{y} \end{cases} \tag{1}$$

where $\tilde{x} \in R^2$ and $\tilde{y} \in R^2$ are system states and algebraic variable vectors, respectively. $\tilde{A} \in R^{n\times n}$, $\tilde{B} \in R^{n\times m}$, $\tilde{C} \in R^{m\times n}$ and $\tilde{D} \in R^{m\times m}$ are sparse matrices.

$$A = \tilde{A} - \tilde{B}\tilde{D}^{-1}\tilde{C} \tag{2}$$

A is the state matrix of power system. For the i th oscillation mode with the following conjugate pair of eigenvalues:

$$\lambda_i = \sigma_i \pm j\omega_i \tag{3}$$

The mode damping ratio (DR) is defined as

$$DR_i = \frac{-\sigma_i}{\sqrt{\sigma_i^2 + \omega_i^2}} \tag{4}$$

Every oscillation mode has its own DR, and the critical ones have greater influences on system. Only if the power system is stable under the most critical DR, the system is small signal stable [11]. So, in this paper the most critical DR value is selected as the indicator of SSS.

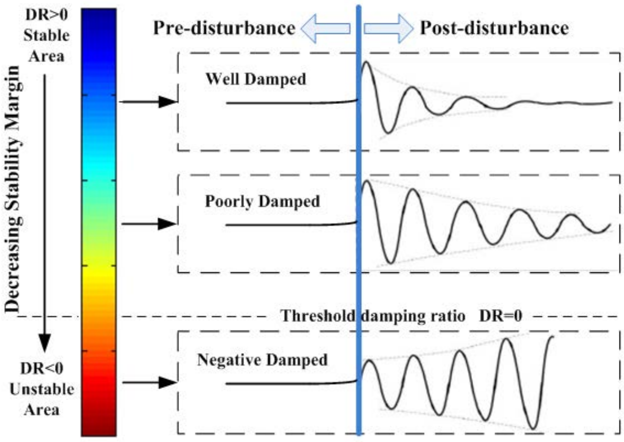

Figure 1. Damping ratio as SSS indicator

As shown in Fig. (1), after a small disturbance occurred, the system will finally be stable with positive DR value, while a negative one leads to the instability of the system. What's more, as the DR value decreases, the SSS becomes much more critical [12]. In this paper RE process is used to predict the online DR values.

2.2. MIC

MIC is an exploratory data analysis tool that can measure relationships between two dependent variables [9]. MIC is based on the idea that if a relationship exists between two variables, then a grid can be drawn on the scatterplot of the two variables that partitions the data to encapsulate that relationship. MIC captures a wide range of associations both functional and not, and for functional relationships provides a score that roughly equals the coefficient of determination of the data relative to the regression function. MIC belongs to a larger class of maximal information-based nonparametric exploration statistics for identifying and classifying relationships. In this paper MIC is applied to the study of power system SSS and to explore the relationships between DRs and power system operation variables.

Given a finite set D of ordered pairs, we can partition the x-values of D into x bins and the y-values of D into y bins, allowing empty bins. We call such a pair of partitions an x-by-y grid. Given a grid G, let $D|_G$ be the distribution induced by the points in D on the cells of G; that is, the distribution on the cells of G obtained by letting the probability mass in each cell be the fraction of points in D falling in that cell. For a fixed D, different grids G result in different distributions $D|_G$.

For a finite set $D \subset R^2$ and positive integers x, y

$$I^*(D,x,y) = \max I(D|_G) \tag{5}$$

where the maximum is over all grids G with x columns and y rows, and $I(D|_G)$ denotes the mutual information of $D|_G$.

The characteristic matrix *M(D)* of a set *D* of two-variable data is an infinite matrix with entries

$$M(D)_{x,y} = \frac{I^*(D,x,y)}{\log \min\{x, y\}} \tag{6}$$

The Maximal Information Coefficient (MIC) of a set *D* of two-variable data with sample size n and grid size less than *B(n)* is given by

$$MIC(D) = \max_{xy<B(n)}\{M(D)_{x,y}\} \tag{7}$$

where $B(n)=n^{0.6}$ is found to work well in practice, and all analyses carried out in this paper use this setting [9]. Some properties of MIC are as follows.

1. MIC assigns scores that tend to 1 to all never-constant noiseless functional relationships;
2. MIC assigns scores that tend to 1 for a larger class of noiseless relationships;
3. MIC assigns scores that tend to 0 to statistically independent variables.

2.3. PCC

PCC is a kind of linear correlation coefficient, which is used to define the linear correlation between two variables. It is widely used in statistical analysis, pattern recognition and image processing.

The PCC is defined by

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{S_X}\right)\left(\frac{Y_i-\bar{Y}}{S_Y}\right) \tag{8}$$

where n is the size of samples, $\bar{X}$ is the mean of X, S_X is the standard deviation of X; $\bar{Y}$ is the mean of Y, S_Y is the standard deviation of Y.

PCC falls between -1 and 1. Some properties are as follows.

1. $r > 0$: The two variables have positive correlation;
2. $r < 0$: The two variables have negative correlation;
3. $r = 0$: The two variables are not linear correlative, but there may exist other correlation (such as curve relationship);
4. the larger $|r|$ is, the stronger linear relationship between the two variables is.

3. Integrated Scheme of Online SSS Estimation Process

The integrated process of RE-based online SSS analysis is shown in Fig. (2). This method consists of three sections: offline training, model update, online estimation.

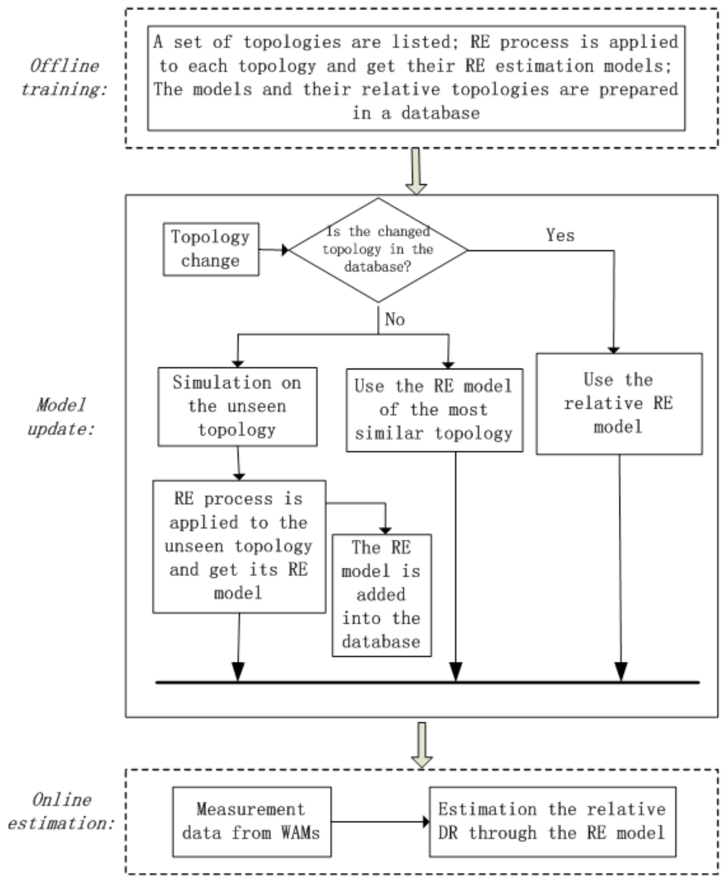

Figure 2. Integrated scheme of online SSS estimation method

1. Offline training: It takes a lot of time to analyze the SSS of a large scale power system. So ,in order to realizing online estimation timely, it's important to train the system offline beforehand and the model relationships can be prepared for online estimation. It is necessary to create a database by training a list of topologies that may emerge in real-time operation to improve the efficiency of online SSS analysis. RE process is applied to create the database and the database includes topologies and their relative RE estimation models.

2. Model update: The model needs to be updated when coming up to a new topology that is not included in the database. In online estimation, if the operating topology is in the database, then the relative relationships model is used to estimate the SSS. If the topology changes to an unseen one, first we can use the most similar topology model instead of the original one to see whether the estimation result is acceptable or not. If the result is not accurate, then the RE process will be used in the new topology to form the corresponding relationship models, and add the new topology and its model into the database. In this stage the database becomes larger and larger, thus the probability of encountering new topology is getting less and the efficiency of estimation higher.

3. Online estimation: With the first two steps to determine the relationship models, after obtaining the required input variable parameters by the WAMS Online, the DR can be quickly calculated through the relative fitting equation. Then the real-time SSS can be estimated.

The RE process includes 3 steps as shown in Fig. (3).

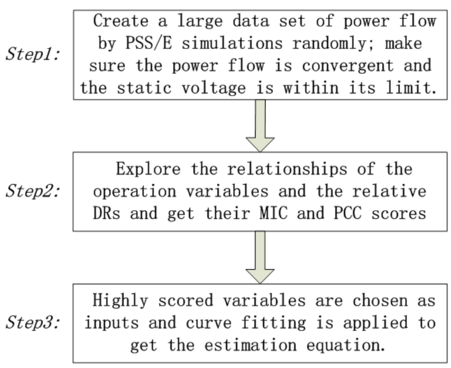

Figure 3 RE process

1. Step 1: PSS/E is used to simulate the power flow and the simulation results are used to create a large data set, including all kinds of operation variables and the relative DRs. Python and MATLAB programs are used to automate the PSS/E simulations, including initializing the operation data randomly in normal ranges, controlling the directions of loads, generators and shunts, and creating a large set of samples. In the simulation results, 80% of the samples are used for offline training and 20% of the samples are regarded as the test set.

2. Step 2: MIC and PCC are used to explore the relationships between every operation variable and DR, and give scores according to how strong the relationships they have. The stronger relationships they have, the higher scores are given. The MIC and PCC scores are shown in Fig. (4). In Fig. (5), the left column show some relationships of 39-bus system and the right column show some relationships of 1648-bus system. Fig. (5A) and Fig. (5B) show examples of inconspicuous relationships. Fig. (5C) and Fig. (5D) show examples of linear relationships. Fig. (5E) and Fig. (5F) show examples of nonlinear functional relationships.

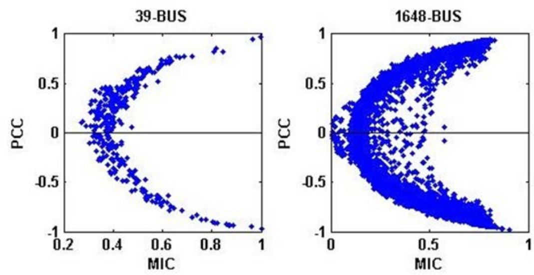

Figure 4. MIC and PCC scores

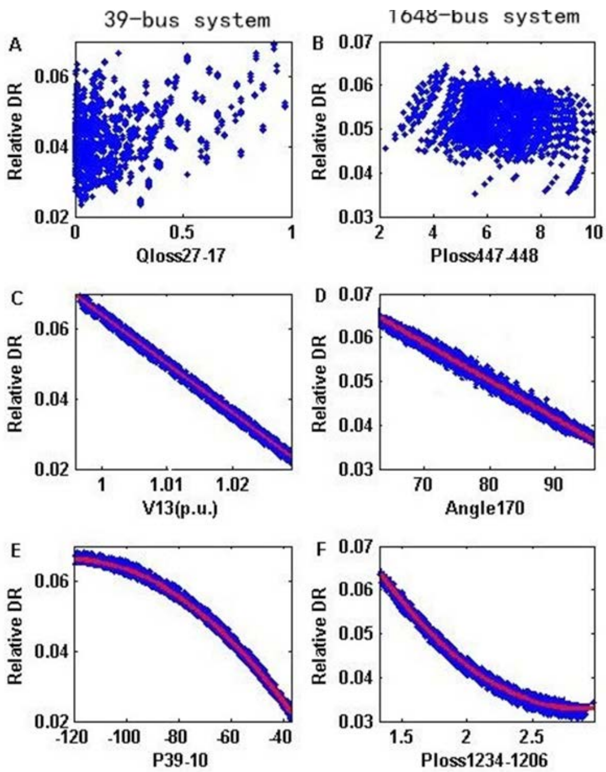

Figure 5. Some relationships between operation variables and DRs

3. Step 3: According to the scores given by MIC and PCC, the highly scored and top ranking variables are selected as optimal input parameters. The MATLAB curve fitting tool is applied to get the fit-

ting equation between the selected variables and DRs. For nonlinear functional relationships, the estimation result is given by Equation (9). For linear relationships, the estimation result is given by Equation (10). And the final result is given by Equation (11). The final equation is the RE estimation model.

$$E_{MIC} = \sum_{i=1}^{N_1} \frac{MIC_i}{\sum_{i=1}^{N_1} MIC_i} E_i \quad (9)$$

$$E_p = \sum_{j=1}^{N_2} \frac{|p_i|}{\sum_{j=1}^{N_2} |p_j|} E_j \quad (10)$$

$$E_F = \frac{N_1}{N_1 + N_2} E_{MIC} + \frac{N_2}{N_1 + N_2} E_p \quad (11)$$

where N_1 is the total number of selected nonlinear functional relationships, MIC_i is the MIC score of relationship i, E_i (or E_j) is the single estimation result by relationship i (or j), E_{MIC} is the comprehensive estimation result by the selected nonlinear functional relationships, N_2 is the total number of selected highly ranked linear relationships, p_j is the PCC score of relationship j, E_p is the comprehensive estimation result by the selected linear relationships. E_F is the final estimation result.

4. Performance Examination

The RE-based SSS estimation method is tes- ted in 39-bus and 1648-bus power system, and the test results are acceptable.

4.1. Relationships Exploration

Relationships between operation variables and DRs are explored by MIC and PCC. There are 447 relationships in 39-bus system and 26561 relationships in 1648-bus system, which are shown in Fig. (4). Some highly scored relationships are shown in Fig. (5), including linear relationships and nonlinear ones.

4.2. Estimation Test

The method is tested under $N_1 = 10, N_2 = 10$. The accuracy of the predicted DR is measured through two statistical indexes. One is called residuals squared error (R^2) [13], which is as follows:

$$R^2 = 1 - \frac{\sum[y_i - d(x_i)]}{\sum(y_i - \bar{y})^2} \quad (12)$$

where x_i is an input variable, y_i is the actual DR, $d(x_i)$ is the predictive value of DR, and $\bar{y}$ is the mean of y_i.

It is believed the closer the value of R^2 is to 1, the better the prediction is. But in practice, how good an R^2 is depends on the particular situation where it is used and the way it is measured [13]. In this work, $R^2 > 0.90$ shows acceptable results.

Another one is root-mean-square (RMS) [14]. It is utilized:

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[y_i - d(x_i)]} \quad (13)$$

where n is the number of test cases, y_i is the actual DR, and $d(x_i)$ is the predictive value of DR.

In 39-bus system, DR>3% is required to guarantee an sufficient SSS margin. In 1648-bus system, DR>4% is considered as secure operation condition [1].

The value of RMS depends on the base magnitude of the DR [15]. In 39-bus system RMS<0.025 and in 1648-bus system RMS<0.03 is considered as acceptable values.

Predictions of 300 new operation states is shown in Fig. (6), and its estimation accuracy is shown in Table 1.

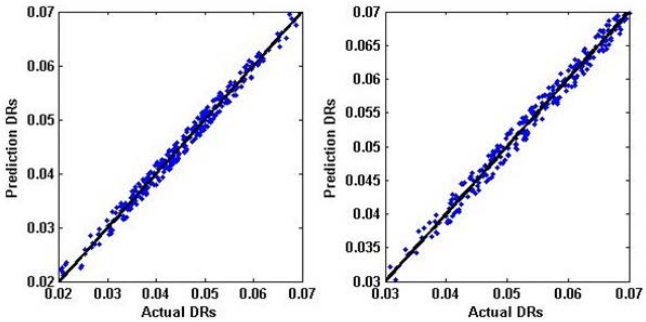

Figure 6. Predictions of 300 new operation states

Table 1. Estimation Accuracies of 300 New Operation States

System	R^2	RMS
39-bus	0.972	0.0174
1648-bus	0.985	0.0193

4.3. Impact of Training Set Size

Obviously, the more the training set size is, the better the estimation result will be. But at the same time, it will take more time and much more amount of calculation. So, it is important to find out an appropriate training set size. In this paper,39-bus system and 1648-bus system are tested to find out the suitable training set size of their own. 100%, 70%, 50%, 30%, 10% and 5% of the original training samples are tested under the same condition that $N_1 = 10, N_2 = 10$. The estimation results is shown in Fig. (7).

The results indicate that with the training set size increased, the accuracy of estimation will also be increased. Repeated tests show that in the 39-bus system, there is at least 25% of the original training samples (about 700 samples) that can get an accuracy of $R^2>0.90$,and it is at least 20% of original samples(about 1200 samples)in the 1648-bus system.

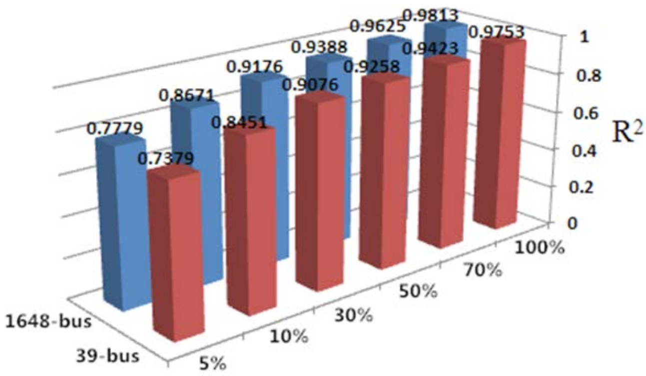

Figure 7. Estimation accuracies of different training set sizes

4.4. Impact of Selected Total Number and Types of Relationships

The impact of the selected relationships' total number and types are studied. In these tests, the input variables are divided into three groups. For each group,4, 8, 12, 16, 20, 24, 28 and 32 of the highly scored relationships are used respectively in each test. The three groups are as follows.

1. Group 1: Half of the relationships are highly scored linear ones by PCC and the other half are highly scored nonlinear ones by MIC;

2. Group 2: All of the relationships are highly scored nonlinear ones by MIC;

3. Group 3: All of the relationships are highly scored linear ones by PCC.

In the tests, 100% of its training samples are used. The accuracies of estimation are shown in Fig. (8) and Fig. (9). We can see from the chart that the more the total number of input variables are, the more accurate the estimation is. The combination of using linear relationships and using nonlinear functional ones has better performance than either of them. So it is better to use both linear and nonlinear relationships as inputs. At least 10 relationships in 39-bus system and 12 relationships in 1648-bus system is needed for guaranteeing an acceptable estimation accuracy of $R^2>0.90$. In practice, how many the inputs is needed depends on the specific system and can be easily figure out through tests. In this paper we choose $N_1=10, N_2=10$. Under this condition, it ensures high accurate without too much calculation.

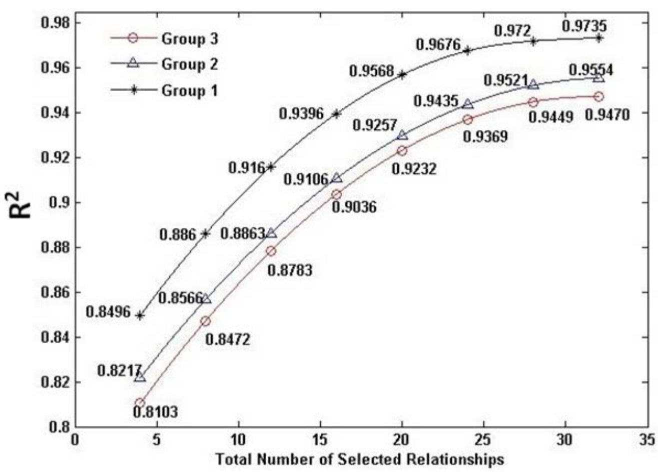

Figure 8. Estimation accuracies of 39-bus system

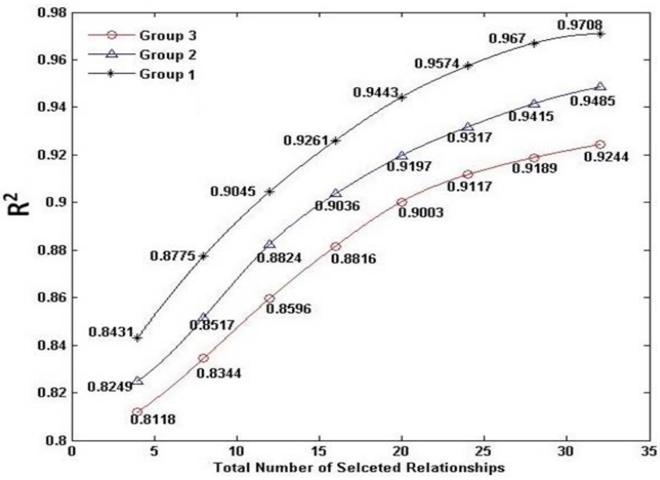

Figure 9. Estimation accuracies of 1648-bus system

4.5. Impact of the Ranks of selected Relationships

The impact of ranks of selected relationships on estimation accuracy is tested through 5 groups. There are 10 linear relationships and 10 nonlinear relationships in each group. The result is shown in Table 2. More tests have shown that if only the selected relationships are in the top percents (eg, first 290 about top 1.1% in 1648-bus system), in which we selected 20 inputs randomly, the estimation accuracy will be high and $R^2>0.90$ can be guaranteed at least. So there are various options of input relationships. The phasor measurement Units (PMUs) can be flexibly located due to the relationships. It is better to install a little more PMUs than the needed input relationships as back-ups. This property of the method is important, and it makes the estimation reliable and cost saving.

4.6. Impact of Measurement Errors

In reality, the data from PMUs in WAMS may contain some measurement errors. So the robustness of the proposed method is examined. Generally, PMUs that are level 1 compliant with the standard should provide a total vector error less than 1% [16]. The impact of measurement errors is examined according to this standard. The tests are divided into 2 groups as follows.

1. Group 1: Noise is added only to the test set;
2. Group 2: Noise is added to both training set and test set.

The result is shown in Table **3**. In group 2, it gets a better result.

Table 2. Estimation Accuracies of 1648-bus System

Group	Linear Relationships	Nonlinear Relationships	RMS
1	Top 1-10	Top 1-10	0.0175
2	Top 11-20	Top 11-20	0.0182
3	Top 21-30	Top 21-30	0.0189
4	Top 31-40	Top 31-40	0.0201
5	Top 41-50	Top 41-50	0.0213

Table 3. Estimation Accuracies Considering Measurement Errors

System	RMS	
	Group 1	Group 2
39-bus	0.0231	0.0197
1648-bus	0.0272	0.0212

Table 4. Computational Time

System	Training Time	Test Time
39-bus	about 24 s (447 relationships, 3045 samples)	about 2 s (1140 samples)
1648-bus	about 1 h 20 min (26561 relationships, 6678 samples)	about 4 s (1880 samples)

4.7. Computation time

The spending time of the estimation method is crucial to the online SSS analysis. In practice, the processing of PMU data should be less than 0.033s [12].

The RE process can give its MIC and PCC in about 0.02 min for one relationship of 3045 samples and about 0.06 min of 6678 samples. A computer can run about 20 programs at the same time. The computational time is shown in Table 4. The tests are executed on an Intel Core i-7 CPU 3.40-GHz CPU with 4 GB of RAM. We can see that a new operation state can be assessed in less than 0.002 s for both the 21-bus system and the 1648-bus system. According to the results, the RE-based method satisfies the speed requirement of online SSS analysis.

4.8. Impact of Topology Change

In fact, the topology of a power system hardly stays the same. In this paper, the robustness to topology changes is examined. The tests are divided into 2 groups. The results are shown in Table 5. In Table 5, the estimation accuracies that can't meet the requirement of R^2 >0.90 are marked.

1. Group 1: The estimation model is not updated;
2. Group 2: The estimation model is re-trained.

Table 5. Estimation Accuracies of Different Topologies

No.	Out of Work	System and Test Type	R^2	
			Group 1	Group 2
1	G30	39-bus N-1	0.9034	0.9564
2	G39	39-bus N-1	[0.8746]	0.9490
3	Line 23-22	39-bus N-1	0.9153	0.9637
4	Line 25-37	39-bus N-1	0.9097	0.9672
5	G28	1648-bus N-1	0.9254	0.9732
6	G274	1648-bus N-1	0.9237	0.9685
7	Line 45-25	1648-bus N-1	0.9338	0.9772
8	Line 1473-1469	1648-bus N-1	0.9376	0.9754
9	Shunt 76	1648-bus N-1	0.9273	0.9635
10	G204 and Line 233-181	1648-bus N-2	0.9044	0.9631
11	G421 and Shunt 460	1648-bus N-2	[0.8914]	0.9457
12	Shunt 498 and Line 291-290	1648-bus N-2	0.9157	0.9588
13	G561 and G602	1648-bus N-2	[0.8826]	0.9397
14	Line 331-392 and Line 383-410	1648-bus N-2	0.9174	0.9650
15	G1178, Shunt 76 and Line 1534-1536	1648-bus N-3	0.9089	0.9544
16	G1644, Shunt 1611 and Line 1630-1629	1648-bus N-3	[0.8743]	0.9431

From the table we can make some conclusions as follows.

First, in group 1, some of estimation accuracies are acceptable, while others are not. That means the intrinsic estimation model is somehow satisfy the topology changes. The intrinsic estimation model is efficient when the topology changes mildly.

Second, in group 2, all of the estimation accuracies meet the basic requirement. That means if we update the RE models for unseen system topologies, then we can get accurate estimation results.

Finally, whether the RE models should be updated or not depends on the combinations of the accuracy, estimation speed requirements and how dramatic the topology is changed. And we suggest it is better to update the RE models. If we update the unseen topology all the time, the database will become larger and larger, thus the probability of encountering new topology is getting less and the efficiency of estimation higher.

5. Conclusion

This paper proposes a novel method for online SSS analysis based on RE process. The scheme is applied to 39-bus and 1648-bus systems and the results are reliable and accurate. This method avoids to calculating the system eigenvalues and eigenvectors. Instead, it computes the DRs through the fitting equations prepared in the database. The scheme can meet the accurate and speeding requirements and has various options of input variables. The combination of linear relationships and nonlinear function relationships as input parameters has better performances. The robustness to the PMUs measurement errors is acceptable. When the system topology changes, the method can still hold its accuracy, if not, the RE process will update the estimation model, and thus get the acceptable results. So we believe that the RE-based method for online SSS analysis is a good choice.

Conflict of Interest

The authors confirm that this article content has no conflict of interest.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant (61074101, 50477018 and 51007093) and in part by the Specialized Research Fund for the Doctoral Program of Higher Education (20090141120062).

References

1. C. Y. Chung, L. Wang, F. Howell, and P. Kundur, “Generation rescheduling methods to improve power transfer capability constrained by small-signal stability,” *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 524–530, Feb. 2004.
2. C. Karawita and U. D. Annakkage, “A hybrid network model for small signal stability analysis of power systems,” *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 443–451, Feb. 2010.
3. M. M. Othman and S. Busan, “A novel approach of rescheduling the critical generators for a new available transfer capability determination,” *IEEE Trans. Power Syst.*, vol. PP, no. 99, pp. 1–15, Feb. 2015.
4. A. T. Saric and A. M. Stankovic, “Rapid small-signal stability assessment and enhancement following changes in topology,” *IEEE Trans. Power Syst.*, vol. 30, no. 3, pp. 1155–1163, May 2015.
5. S. H. Tsai, C. Y. Lee, and Y. K. Wu, “Efficient calculation of critical eigenvalues in large power systems using the real variant of the Jacobi-Davidson QR method,” *IET Gener., Transm., Distrib.*, vol. 4, no. 4, pp. 467–478, Apr. 2010.
6. J. Ma, Z. Y. Dong, and P. Zhang, “Comparison of BR and QR eigenvalue algorithms for power system small signal stability analysis,” *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1848–1855, Nov. 2006.
7. J. Ma, S. Wang, Z. Wang, and J. S. Thorp, “Power system small-signal stability region calculation method based on the guardian map theory,” *IET Gener., Transm., Distrib.*, vol. 8, no. 8, pp. 1479–1488, Aug. 2014.
8. J. L. Rueda, D. G. Colome, and I. Erlich, “Assessment and enhancement of small signal stability considering uncertainties,” *IET Gener., Transm., Distrib.*, vol. 24, no. 1, pp. 198–207, Feb. 2009.
9. N. R. David, A. R. Yakir, K. F. Hilary, R. G. Sharon, M. Gilean, J.T. Peter, S. L. Eric, M. Michael, and C. S. Pardis, “Detecting novel associations in large data sets,” *Science*, vol. 334, pp. 1518–1524, Dec.2011.
10. J. Benesty, J. Chen, and Y. Huang, “On the importance of the Pearson correlation coefficient in noise reduction,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 4, pp. 757–765, May 2008.
11. P. Kundur, et al., “Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions,” *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1387–1401, Aug. 2004.
12. C. Zheng, V. Malbasa, and M. Kezunovic, “Regression tree for stability margin prediction using synchrophasor measurements,” *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1978–1987, May 2013.
13. M. A. Pai, *Energy Function Analysis for Power System Stability*. Boston, MA: Kluwer, 1989.
14. R. G. Steel and J. H. Torrie, *Principles and*

Procedures of Statistics. New York: McGraw-Hill, 1960.
15. R. F. Nau, Forecasting, Feb. 2005. [Online]. Available: http://www.duke.edu/~rnau/rsquared.htm
16. *IEEE Standard for Synchrophasors for Power Systems*, IEEE Std. C37.118-2005, 2005.

Automatic Discovery and Application of Significant Relationships Between Steady-state Operation Data and Transient Stability Level in Electric Power Industry

Youping Fan*, **Songkai Liu, Cheng Tian**

School of Electrical Engineering, Wuhan University, Wuhan 430072, China

Abstract
The assessment of transient stability level is important for the automation of production processes in electric power industry. The connotative relationships of steady-state operation data and general critical clearing time $\overline{CCT}$ are explored in a large data set for power system. A novel online transient security assessment method is presented based on relationships exploration. Each relationship is given scores by the maximal information coefficient and Pearson correlation coefficient. Some highly ranked linear and nonlinear relationships are detected out and shown. Meanwhile, the generalized nonlinear relationships exploration coefficient is presented to discover connotative nonlinear relationships directly. Curve fitting is used for the explored linear relationships and functional nonlinear relationships to estimate $\overline{CCT}$ of new operation states. Weibull distribution and generalized extreme value distribution are adopted for distribution fitting of $\overline{CCT}$, and cumulative probability curve is used to determine the value range of $\overline{CCT}$ for each transient security level. The method is tested on a 21-bus system and various test results indicate it is accurate and effective. It can give accurate estimation results of $\overline{CCT}$, relative degree of transient stability and security level of transient stability. The applicability will not be influenced by the change of structure and scale since the selection of input features is based on data statistics and mining, and the way of selection is more intelligent than the current techniques. The automatic identification of transient stability level is meaningful for uninterrupted production in power industry.
Keywords: POWER SYSTEM AUTOMATION, TRANSIENT SECURITY ASSESSMENT, AUTOMATIC IDENTIFICATION, SIGNIFICANT RELATIONSHIPS, LARGE DATA SETS