

# Non-parametric statistics-based predictor enabling online transient stability assessment

ISSN 1751-8687

Received on 5th June 2018

Revised 18th September 2018

Accepted on 25th September 2018

E-First on 5th November 2018

doi: 10.1049/iet-gtd.2018.5802

www.ietdl.org

Tingjian Liu<sup>1</sup>, Youbo Liu<sup>1</sup>, Lixiong Xu<sup>1</sup> ✉, Junyong Liu<sup>1</sup>, Joydeep Mitra<sup>2</sup>, Yuting Tian<sup>2</sup>

<sup>1</sup>College of Electrical Engineering and Information Technology, Sichuan University, Chengdu, People's Republic of China

<sup>2</sup>Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan, USA

✉ E-mail: xulixiong@163.com

**Abstract:** Online transient stability assessment (TSA) is of great necessity for fast awareness of transient instability caused by fault contingencies. In this paper, a non-parametric statistics based scheme is proposed for response-based online TSA. A critical clearing time-based stability margin index is defined as the predictive output and 14 kinds of severity indicators are proposed as input features for the TSA predictor. With no prior knowledge of the correlation structure, the non-parametric additive model is used as the basis of the predictor. To screen out the weakly correlated indicators and reduce the dimensionality of the input space, two-stage feature selection is fulfilled by non-parametric independence screening and group Lasso penalised regression successively. The predictor is then learnt by least-squares regression in the reduced multi-feature space. With phasor measurement unit measurements at generator buses, severity indicators can be computed in the real-time and fast evaluation of post-fault stability margin can be made by the offline-trained predictor. The effectiveness of the proposed non-parametric statistics based scheme is demonstrated in a modified New England 39-bus system and a practical 756-bus transmission system in China.

## 1 Introduction

Transient stability refers to the ability of power systems to maintain synchronism when subjected to severe disturbances such as faults [1]. In large-scale interconnected power systems, a fault that occurs in a subsystem may cause unexpected out-of-step of some synchronous generators, network splitting from critical transmission corridors and ultimately blackouts. Online transient stability assessment (TSA) is necessary for system operation since it predicts the post-fault stability status in a real-time manner, providing the opportunity for timely execution of remedial action schemes to prevent system collapse.

Time-domain simulation (TDS) is the classical method for TSA. However, TDS is not practical for online prediction of transient stability due to the extensive computational burden of solving the differential algebraic equations. Currently, prospective methods for online TSA are the transient energy function (TEF) methods, the hybrid methods, the Lyapunov exponents (LEs) methods and the data-mining methods:

(1) The TEF methods construct the Lyapunov functions to evaluate the transient energy and predict the post-fault stability by comparing the energy at fault clearing time with the maximum potential energy that can be absorbed by the post-fault system. TEF models that incorporate with high-voltage DC transmission [2] and stochastic renewable generation [3] have been proposed, respectively, to accommodate the integration of these novel components and to provide a more accurate assessment. However, the difficulty of TEF-based real-time TSA is mostly the online computation of the controlling unstable equilibrium point (CUEP) and correspondingly the maximum potential energy. In [4], a look-up table of potential modes of disturbance is built offline to assist real-time identification of CUEP. However, the computation of CUEP is only quickened but not displaced.

(2) Hybrid methods are referred to the combination of TDS and TEF methods initially. However, with the deployment of phasor measurement units (PMUs), trajectories from TDS are replaced by online measurements, so that hybrid methods are adaptive to real-time application. The concept of potential energy in the critical corridors is used to detect loss of synchronism in [5]. The pair-wise

relative energy function is proposed in [6] for fast identification of the critical generators and then the single machine equivalents (SIME) and the equal-area criterion are employed to qualify the transient stability. Both the above-mentioned hybrid methods rely on the prediction of the SIME's unstable equilibrium point since they make use of the TEF concepts.

(3) On the basis of the ergodic theory of dynamic systems, LEs can indicate the divergence or convergence of the generator angle trajectories and thus the synchronism of generators can be detected by tracking the sign of the maximal LE (MLE) [7]. A data-driven approach for LE computation from time-series PMU measurements is proposed in [8], so that the model-dependent problem in [7] is overcome. Nevertheless, the post-fault stability status is still difficult to determine because the MLE may oscillate from positive and negative values before the post-fault system settles down. To address this problem, a recursive least-squares-based method is proposed in [9] for fast estimating the MLE, and therefore the monitoring window is shortened. So far, the connection between the change sign of MLE and the system's passing through the CUEP has not been studied. The optimal time for remedial control is thus not clear and future research is needed.

(4) As for data-mining-based TSA, post-fault stability can be assessed promptly by feeding the online measurements into the offline-learned predictor. With the advent of data-rich and processor-rich smart grid environment, data-mining techniques can promote response-based TSA, enabling real-time awareness of transient stability and also response-based remedial control schemes [10].

Some promising results of response-based TSA by data mining have been reported in [10–22]. Decision trees (DTs) have the superiority of transparency [11] and do not rely on the back-propagation (BP) training process [12] in comparison with neural networks (NNs). To reduce the input dimension without compromising the predictor's accuracy, a novel DT-based TSA scheme is proposed by introducing the characteristic ellipsoid (CELL) theory to extract the key features from limited PMU measurements in [13]. Another DT-based framework is proposed in [14] to predict the unstable generator grouping pattern in power systems with renewable generation. On the basis of stability

prediction, the response-based one-shot control scheme is also developed by using PMU measurements and DTs in [15]. Support vector machine (SVM) and its derivatives such as core vector machine (CVM) have also been applied to develop TSA classifiers. In [16], the SVM classifier takes proximity of the actual voltage variations to the pre-identified templates as inputs and satisfactorily prediction can be made within the six-cycles observation window. Another SVM classifier using TEF-based features as inputs is also investigated in [17] and accurate estimations can be provided for multiple contingencies with the maximum load/generation deviation to be  $\pm 20\%$ . In [18], case studies in two practical power systems in the USA and China have also validated the effectiveness of CVM classifiers. Apart from training a single conventional predictor, ensemble methods have been proposed as well. Catastrophe predictors based on the random forest (RF) are built on the basis of wide-area severity indices in [10] and these RF-based predictors have shown the robustness to different network dynamics when compared with a single DT. An intelligent system (IS) is developed for post-disturbance TSA by using an ensemble of extreme learning machines in [19]. Owing to the fast learning speed of extreme learning machines (ELMs), the ensemble classifiers can be updated by online pre-disturbance TSA results, thus improving the self-adaptiveness of the IS. The emerging deep learning techniques have also been proposed in recent years. Following the framework of [19], the long short-term memory network is used to develop a temporal self-adaptive TSA system in [20]. To develop a scalable TSA framework for large-scale power grids, MapReduce-based parallelised NNs are used for instability prediction and critical unstable generators identification in [21]. All the above-mentioned predictors only classify the post-

stability status, but none of them provide any quantitative assessment. Multi-variate adaptive regression splines are used in [22] to assess the transient stability margin based on real-time measurements. However, the literature on developing quantitative TSA predictor as [22] is very limited.

In this paper, a non-parametric statistics based scheme is proposed for response-based online assessment of transient stability margin. The stability margin based on critical clearing time (CCT) is first defined as the predictive response, and 14 kinds of severity indicators are proposed as input features. With no prior knowledge of the correlation structure between severity indicators and the corresponding stability margin, the non-parametric additive model is used as the basis of the predictor. After forming a knowledge base, weakly correlated severity indicators are screened out by a novel two-stage non-parametric analytics based feature selection. Predictor for online TSA is then learned by least-squares regression in the reduced multi-feature space. Since the severity indicators can be computed by post-fault PMU measurements, the stability margin can be estimated by the predictor in real-time operation. An illustrative case study on a modified New England 39-bus system and application of the proposed scheme on a practical 756-bus transmission system in China are provided.

The rest of this paper is organised as follows. The CCT-based stability margin index and the severity indicators are defined in Section 2. The technical details of the proposed scheme and the fundamentals of non-parametric statistics are introduced in Section 3. A case study on the New England 39-bus system is presented to illustrate the effectiveness of the proposed scheme in Section 4. Application to a practical transmission system is provided in Section 5. Finally, conclusions are drawn in Section 6.

**Table 1** Response-based severity indicators

Number	Severity indicators	Equations
1	separation of generator rotor speeds under COI reference at $t_0$	$\rho_1 =  \omega_g^{\text{COI}}(t_0) $
2	accelerations of generators under COI reference at $t_0$	$\rho_2 =  f_g^{\text{COI}}(t_0) $
3	voltage magnitudes of generator buses at $t_0$	$\rho_3 = V_g(t_0)$
4	separation of generator power angles under COI reference at $t_{\text{cl}}$	$\rho_4 =  \delta_g^{\text{COI}}(t_{\text{cl}}) $
5	coherency of generator power angles under COI reference [25]	$\rho_5 = \int_{t_{\text{cl}}}^{t_{\text{end}}} [\delta_g^{\text{COI}}(t)]^2 dt$
6	separation of generator rotor speeds under COI reference at $t_{\text{cl}}$	$\rho_6 =  \omega_g^{\text{COI}}(t_{\text{cl}}) $
7	coherency of generator rotor speeds under COI reference	$\rho_7 = \int_{t_{\text{cl}}}^{t_{\text{end}}} [\omega_g^{\text{COI}}(t)]^2 dt$
8	accelerations of generators under COI reference at $t_{\text{cl}}$	$\rho_8 =  f_g^{\text{COI}}(t_{\text{cl}}) $
9	maximum deviation of voltage magnitude of generator buses [26]	$\rho_9 =  V_g(t) - V_{gN} , t \in [t_{\text{cl}}, t_{\text{end}}]$
10	accumulated deviation of voltage magnitude of generator buses [26]	$\rho_{10} = \int_{t_{\text{cl}}}^{t_{\text{end}}}  V_g(t) - V_{gN}  dt$
11	recovering rate of voltage magnitude of generator buses at $t_{\text{cl}}$	$\rho_{11} = \frac{dV_g}{dt} \approx \frac{V_g(t_{\text{cl}} + \Delta t) - V_g(t_{\text{cl}})}{\Delta t}$
12–14	magnitude of dot-product trajectories [24]	$\rho_{12} = \text{dot}1^{\text{max}}(t) - \text{dot}1^{\text{min}}(t)$ $\rho_{13} = \text{dot}2^{\text{max}}(t) - \text{dot}2^{\text{min}}(t)$ $\rho_{14} = \text{dot}3^{\text{max}}(t) - \text{dot}3^{\text{min}}(t)$

## 2 Stability margin and severity indicators

Data mining is performed to establish a projection between the response and the features. As for data-mining-based TSA in this paper, a stability margin index based on CCT is defined as the predictive response, while 14 kinds of response-based severity indicators are proposed as the inputs of the predictor.

### 2.1 CCT-based stability margin index

CCT represents the boundary of transient stability from the perspective of fault clearing time. The margin of fault clearing time as is shown in (1) can be used to evaluate the post-fault stability of power systems

$$\eta = \frac{t_{\text{cr}} - t_{\text{cl}}}{t_{\text{cr}}} \quad (1)$$

where  $t_{\text{cr}}$  and  $t_{\text{cl}}$  are the CCT and the fault clearing time of a fault contingency, respectively.

### 2.2 Response-based severity indicators

Post-fault responses of power angles, rotor speeds and accelerating powers of generators, and voltage magnitudes of buses contain key information relating to transient stability. Numerous severity indicators have been proposed for stability assessment [18, 23–26]. In this paper, 14 kinds of indicators are proposed to form the multi-dimensional space of input features. Definitions of these indicators are shown in Table 1.

The nomenclature is given as follows. Here,  $t_0$  and  $t_{\text{cl}}$  are the faults occurring time and the fault clearing time, respectively, while  $t_{\text{end}}$  is the moment at which observation window ends.  $g$  denotes the serial number of generators while  $G$  represents the set of generators.  $\delta$ ,  $\omega$ ,  $P_m$  and  $P_e$  are power angle, rotor speed, mechanical power input and electrical power output of a generator.  $V$  is the voltage magnitude of the generator bus and  $V_N$  is its rated value.

Besides

$$M = \sum_{g \in G} M_g \quad (2)$$

$$\delta_{\text{COI}}(t) = \frac{1}{M} \sum_{g \in G} M_g \delta_g(t) \quad (3)$$

$$\omega_{\text{COI}}(t) = \frac{1}{M} \sum_{g \in G} M_g \omega_g(t) \quad (4)$$

$$P_{\text{COI}}(t) = \sum_{g \in G} [P_{mg}(t) - P_{eg}(t)] \quad (5)$$

$M_g$  ( $g \in G$ ) is the inertia coefficient of a generator while  $M$  is the aggregated inertia coefficient of a power system.  $\delta_{\text{COI}}$ ,  $\omega_{\text{COI}}$  and  $P_{\text{COI}}$  are the phase angle, the angular frequency and the accelerating power of centre of inertia (COI). The superscript of COI represents that the parameters are under COI reference. Thus, we have

$$\delta_g^{\text{COI}}(t) = \delta_g(t) - \delta_{\text{COI}}(t) \quad (6)$$

$$\omega_g^{\text{COI}}(t) = \omega_g(t) - \omega_{\text{COI}}(t) \quad (7)$$

$$\begin{aligned} f_g^{\text{COI}}(t) &= \frac{d\omega_g^{\text{COI}}(t)}{dt} \\ &= \frac{P_{mg}(t) - P_{eg}(t) - M_g P_{\text{COI}}(t)/M}{M_g} \end{aligned} \quad (8)$$

These indicators capture the post-fault response of a single generator. To evaluate the impact of a fault contingency on the system, five kinds of statistic indices, which are the maximum, the minimum, the maximum separation, the average and the standard deviation, are utilised. Therefore, an input space of  $14 \times 5 = 70$  indicators is formed. These statistic indexes will be presented by  $\rho_i^{\max}$ ,  $\rho_i^{\min}$ ,  $\rho_i^{\text{diff}}$ ,  $\rho_i^{\text{mean}}$  and  $\rho_i^{\text{std}}$  in this paper.

### 3 Proposed scheme

The proposed scheme for online TSA consists of four stages, which are data preparation, feature reduction, predictor training and online application. The technical details of these four stages will be introduced along with the fundamentals of non-parametric statistics in this section.

#### 3.1 Data preparation

The performance of a predictor relies heavily on the comprehensiveness of the knowledge base. As for TSA, the uncertainties that have impacts on post-fault stability are a pre-fault operating condition (OC), fault location and fault clearing time. A database of OC should be first produced by historical or forecasted OC data. To enrich this database, more stochastic OCs can be generated by random variation of loads and generations. On each OC, an arbitrary transmission line is selected as the faulted device according to its average failure rate and the fault clearing time is sampled by its probability distribution. For each fault contingency, TDS is performed to compute the post-fault responses, the severity indicators and the CCT-based stability margin. Each instance of fault contingency is then represented by a row vector of severity indicators and the corresponding stability margin as is shown in the equation below:

$$s = [\{\rho_1\}, \dots, \{\rho_i\}, \dots, \{\rho_{14}\} : \eta] \quad (9)$$

where

$$\{\rho_i\} = [\rho_i^{\max}, \rho_i^{\min}, \rho_i^{\text{diff}}, \rho_i^{\text{mean}}, \rho_i^{\text{std}}], \quad 1 \leq i \leq 14$$

Simulation and recording of fault contingencies are repeated until a knowledge base of a given scale is generated.

#### 3.2 Feature reduction by two-stage non-parametric analytics

During the stage of offline learning, feature reduction and predictor training are two major issues. Feature reduction is to screen out some weakly correlated severity indicators as they may not have a significant contribution on the performance improvement of the trained TSA predictor, but instead leads to the increased computation burden of predictor training. On the other hand, predictor training is to learn the marginal correlation structure and the relating parameter.

There is little prior knowledge that indicates the correlation takes a linear form or belongs to any other finite-dimensional parametric family. To address this difficulty, the non-parametric additive model, as is shown in (10), can be used for better flexibility and approximation accuracy

$$\eta = M(X) + \varepsilon = \mu + \sum_{i=1}^N f_i(x_i) + \varepsilon \quad (10)$$

where

$$x_i \in X = [\{\rho_1\}, \dots, \{\rho_{14}\}]$$

For simplicity, the severity indicators  $\rho$  are rewritten as  $x$  and the saying of 'feature  $x$ ' is invariably referred to the severity indicators in the rest of this paper.  $M(\cdot)$  denotes the correlation between the stability margin  $\eta$  and the severity indicators  $X$ .  $\mu$  is the intercept,  $f_i(\cdot)$  is the non-parametric component relating to the  $i$ th indicator  $x_i$  and  $\varepsilon$  is an unobserved regression error.

For feature reduction in the high-dimensional non-parametric additive model, a two-stage feature selection technique is proposed by combining the non-parametric independence screening (NIS) and the group Lasso penalised regression.

**3.2.1 Feature reduction based on NIS:** The principle of NIS [27] is to identify and screen out the weakly correlated features by fitting the univariate correlations and thresholding the regression error of the fitting functions.

Supposing there is a correlation between the stability margin  $\eta$  and the severity indicator  $x_i$ , this correlation can be estimated by a fitting function as is shown in the equation below:

$$\eta = f_i(x_i) + \varepsilon_i \quad (11)$$

where  $f_i(x_i)$  denotes the fitting function and  $\varepsilon_i$  is the regression error that obeys normal distribution.

The goodness of each fitting function can be evaluated by the root mean square error (RMSE) as is shown in the equation below:

$$\text{RMSE} = \sqrt{\frac{1}{K} \sum_{k=1}^K [\eta_k - f_i(x_{ik})]^2} \quad (12)$$

where  $k$  denotes the serial number of instances and  $K$  is the number of instances.

Without any prior knowledge of the parametric model for correlation fitting, basis-spline function, namely B-spline function, is often used for non-parametric regression because of its ability to fit both linear or non-linear correlations. The B-spline function can be represented by a linear combination of basis functions as is shown in the equation below:

$$f_i(x_i) = \sum_{j=1}^d \beta_j \phi_{j,\gamma}(x_i) \quad (13)$$

where basis function  $\phi_{j,\gamma}(x_i)$  is a piecewise polynomial function of  $\gamma$  order,  $d$  denotes the degree of freedom and  $\beta_j$  represents the contribution coefficient of the basis functions  $\phi_{j,\gamma}(x_i)$ . Three-order B-spline is used in this paper, thus order  $\gamma$  is chosen as 3. To fit non-linear correlations smoothly, the marginal degree of freedom has to be searched by repeatedly performing univariate regression

---

```

1: Input: Data matrix  $S = \{X_1, X_2, \dots, X_n, Y\}$ 
   Selection threshold  $\tau_s$ 
2: for  $i = 1$  to  $N$ 
3:   Extract the  $i$ th column  $X_i$  and the last column  $Y$  from
    $S$ 
4:   for  $d=3$  to  $13$ 
5:     Fit  $Y$  against  $X_i$  with (6) of  $d$  degree of freedom
6:     Evaluate the regression error  $RMSE_{i,d}$  by (5)
7:   end  $d$  Loop
8:   Determine the marginal function by smallest  $RMSE_{i,d}$ 
9: end  $i$  Loop
10: Rank all the features by  $RMSE_{i,marginal}$ 
11: Eliminated features with  $RMSE_{i,marginal} > \tau_s$ 
12: end algorithm

```

---

**Fig. 1** Algorithm 1: Feature selection based on NIS

under a different degree of freedom. In this paper, the degree of freedom takes a value within the set of  $\{d|3 \leq d \leq 5 \text{ and } d \in \mathbb{Z}\}$ .

All the severity indicators are ranked by RMSE and features with RMSE higher than a given threshold are then screened out. The procedure of feature reduction based on NIS is summarised by Algorithm 1 (see Fig. 1). Each row of data matrix  $s$  represents an instance. The last column of  $s$  is the stability margin  $\eta$  while the rest denote the severity indicators.

**3.2.2 Feature reduction based on group Lasso:** Group Lasso [28] is an enhanced penalised method for feature selection and non-parametric regression. For traditional Lasso, components  $f_i$  in the additive model is considered to take the universal form of linear parameters, which makes it incapable of accurate estimation of non-linear correlations. In contrast, feature reduction based on group Lasso is achieved by penalising the non-parametric components in the additive model. Assuming that the non-parametric components share the same structures as the marginal B-spline univariate functions that are determined in the stage of NIS, the non-parametric additive model can be rewritten as is shown in the equation below:

$$\begin{aligned} \eta &= \mu + \sum_{i=1}^N f_i(x_i) + \varepsilon \\ &= \mu + \sum_{i=1}^N \sum_{j=1}^{d_i} \beta_{ij} \phi_{ij}(x_i) + \varepsilon \end{aligned} \quad (14)$$

The estimation of (14) by group Lasso can be expressed by an optimisation problem as shown in the equation below: (see (15)) where

$$\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_N]^T, \quad \boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{i,d_i}]^T$$

$\lambda$  is the penalty factor and  $w_i$  is the weighting factor for the  $i$ th severity indicator. As  $\lambda$  increases, coefficients  $\beta_i$  of some features may decrease to 0. This characteristic of group Lasso ensures that ineffective indicators can be removed from the non-parametric additive model.

Usually, the weighting factors cannot be determined in advance. To avoid the extra computation burden of tuning weighting factors, the adaptive group Lasso is proposed. The procedure of the adaptive group Lasso is demonstrated as follows:

(1) Initialise all the weighting factors such as such as 1 and set the iterator  $l$  as 0.

(2) Compute the  $\mu$  and  $\beta$  by solving (15).

(3) Assign the new value to the weighting factors  $w$  by the equation below:

$$w_i = \begin{cases} \|\beta_i\|_2^{-1}, & \text{if } \|\beta_i\|_2 > 0 \\ \infty, & \text{if } \|\beta_i\|_2 = 0 \end{cases} \quad (16)$$

(4) Compute the maximum difference  $\Delta$  of the weighting factors at the current stage and the previous stage by the equation below:

$$\Delta = \max |w_i^l - w_i^{l-1}|, \quad 1 \leq i \leq N \quad (17)$$

(5) If the maximum difference  $\Delta$  is less than a given threshold, stop the loop and the values of  $w$ ,  $\mu$  and  $\beta$  at the current stage are then the final result of predictor training; otherwise, return to step (2) to proceed the computation.

In summary, the two-stage feature reduction is fulfilled by successively performing the NIS and the group Lasso. NIS helps determine the marginal structure of the non-parametric components and screen out weakly correlated features in the high-dimensional non-parametric additive model. Unlike that the NIS evaluates the features independently, the group Lasso realises feature reduction in a joint multi-variate space, providing a comprehensive evaluation of multiple features.

### 3.3 Predictor training

After two-stage feature reduction, the indicators that are preserved in the non-parametric additive model are selected to form a reduced multi-feature space. With this reduced input space, TSA predictor can be trained by the component-wise least-squares regression. The non-parametric additive model is rewritten as is shown in the equation below:

$$\begin{aligned} \eta &= \mu + \sum_{i=1}^{N_R} f_i(x_i) + \varepsilon \\ &= \mu + \sum_{i=1}^{N_S} \sum_{j=1}^{d_i} \alpha_{ij} \phi_{ij}(x_i) + \varepsilon \end{aligned} \quad (18)$$

In (18), the coefficients are represented by  $\alpha$  for the purpose of distinguishing them from those solved by group Lasso.  $N_R$  denotes the number of reduced severity indicators.

Given the reduced data matrix  $\mathbf{S}_R$

$$\mathbf{S}_R = [X_1, X_2, \dots, X_{N_R} : \eta] \quad (19)$$

B-spline basis matrix  $\boldsymbol{\Psi}$  can be formed by basis functions  $\phi_{ij}$

$$\boldsymbol{\Psi} = [1, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \dots, \boldsymbol{\Phi}_{N_R} : Y] \quad (20)$$

where

$$\boldsymbol{\Phi}_i = [\phi_{i1}(X_i), \phi_{i2}(X_i), \dots, \phi_{i,d_i}(X_i)]$$

Then the coefficients  $\alpha$  can be solved by the equation below:

$$\begin{bmatrix} \mu \\ \alpha \end{bmatrix} = (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T Y \quad (21)$$

---


$$\begin{aligned} \min L(\mu, \beta) &= \min \left\{ \frac{1}{K} \sum_{k=1}^K [\eta_k - \mu - \sum_{i=1}^N \sum_{j=1}^{d_i} \beta_{ij} \phi_{ij}(x_{ik})]^2 + \lambda \sum_{i=1}^N w_i \|\beta_i\|_2 \right\} \\ \text{s.t.} \quad & \sum_{k=1}^K \sum_{j=1}^{d_i} \beta_{ij} \phi_{ij}(x_{ik}) = 0 \end{aligned} \quad (15)$$

where

$$\alpha = [\alpha_1, \alpha_2, \dots, \alpha_{N_R}]^T, \quad \alpha_i = [\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{i,d_i}]^T$$

Once the coefficients are solved, the non-parametric additive model in (14) can be used for fast prediction of transient stability margin.

### 3.4 Online application

Although the predictor for TSA is trained, online application cannot be realised unless the severity indicators can be computed with real-time measurements. Provided that PMUs are installed at the HV buses of power plants, the power angle  $\delta$  and the rotor speed  $\omega$  of generators can be approximated by the phase angle and angular frequency of these buses. The electrical power output of generators  $P_e$  can be approximated by the active power flow of step-up transformer. Considering the delay of speed governors to adjust the mechanical power input,  $P_m$  is assumed to be the same as its pre-fault value, thus  $P_m \approx P_m^0 = P_e^0$ . Moreover, the voltage magnitude of generators' terminal buses  $V$  can be directly measured by PMUs. On the basis of the assumption, severity indicators can be computed with real-time PMU measurements.

With the above-mentioned assumption, once a fault is detected in real-time operation, post-fault PMU measurements within the observation window are used to compute all the severity indicators. After that, post-fault stability margin can be estimated by feeding the severity indicators to the predictor.

## 4 Illustrative case study

The proposed scheme is first illustrated by a case study on a modified New England 39-bus system with centralised wind power integration. The synchronous generator at Bus #37 is replaced by  $200 \times 3.6$  MW double-fed induction generators (DFIGs). All the DFIGs operate at constant power factor mode and the power factor is set to be 1.0.

### 4.1 Data generation

Operating conditions of nine different loading scenarios (varying from 80 to 120% of base condition with the increment of 5%) are used to form the database of OCs. To simulate the uncertainty of real-time operation and enrich the knowledge base, stochastic OCs are also generated by random sampling. Wind power generation at Bus #37 is sampled according to the Weibull distribution while the active power outputs of synchronous generators vary from 50 to 150% of their base condition. The unbalanced power consumption is met by the synchronous generator at Bus #39.

Fault location is randomly chosen at 0, 50 and 100% of the length of an arbitrary transmission line. The fault clearing time is modelled as a normal distribution with a mean value of 0.2 s and a standard deviation of 0.02 s. TDS is performed to simulate the post-fault responses, where the simulation time is 10 s.

About 2000 instances are generated. About 80% of them are randomly chosen to form the training set while the others serve as testing data. With PMU installed at synchronous generator buses, severity indicators are computed with the relevant post-fault trajectories and the observation window is set to be 0.1 s after fault clearing. The CCT-based stability margin index is searched by recursive simulations.

### 4.2 NIS

After data generation, NIS is performed to determine the structure of the additive model and identify the severity indicators that are weakly correlated to the stability margin. The goodness of the univariate fitting functions is evaluated by RMSE and the results are shown in Fig. 2.

As can be seen from Fig. 2, the 7th and the 14th severity indicators, which are the coherency indices of generator rotor speeds ( $\rho_7$ ) and the dot-product index of generator power angle and rotor speed ( $\rho_{14}$ ), are ranked as the top two, making them the most

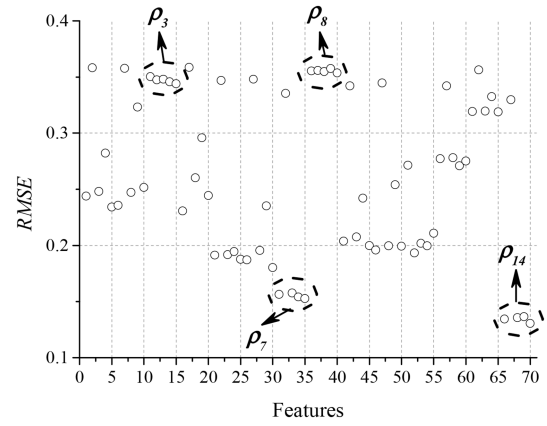


Fig. 2 Goodness of the univariate fitting functions evaluated by RMSE

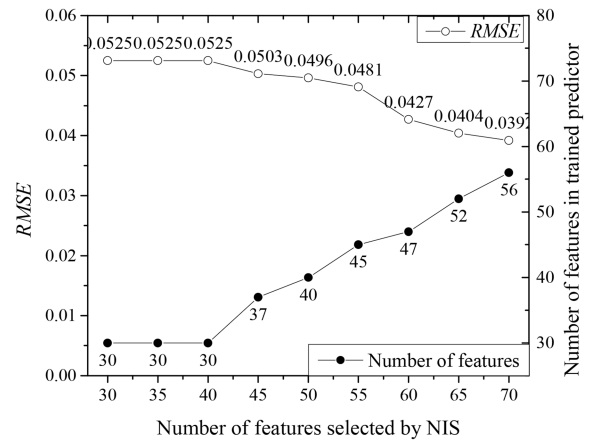


Fig. 3 Tuning of the threshold for NIS

correlated features to the stability margin  $\eta$ . Besides, generally for all the severity indicators, the minimum index is less correlated when compared with the other four indexes.

### 4.3 Tuning of the parameters

The threshold for NIS and the penalty factor for group Lasso are two key parameters to be tuned. The marginal threshold for NIS is first studied while the penalty factor is set to be 0.5. The goodness of the non-parametric predictor and the number of input features in this predictor are shown in Fig. 3. For both simplicity and accuracy of the TSA predictor, the threshold for NIS is chosen as 0.333 so the first 50 indicators are selected.

After determining the threshold for NIS, the marginal penalty factor for group Lasso regression is then studied. The goodness of the non-parametric predictor and the number of input features in this predictor are shown in Fig. 4. As can be seen from Fig. 4, the number of severity indicators in the trained predictors decreases with the increase of penalty factor, meanwhile resulting in the decrease of the predicting accuracy.

### 4.4 Testing data validation

With the NIS threshold and the penalty factor set to be 0.333 and 0.5, a non-parametric predictor is trained under the reduced set of severity indicators. The TSA predictor takes the non-parametric additive form as (17). To demonstrate the non-parametric components of this functional predictor, the B-spline component relating to the 68th input feature ( $\rho_{14}^{\text{diff}}$ ), is shown in (22) as an example. To understand the non-linear correlation between  $\rho_{14}^{\text{diff}}$  and the stability margin  $\eta$ , this B-spline component is also plotted in Fig. 5

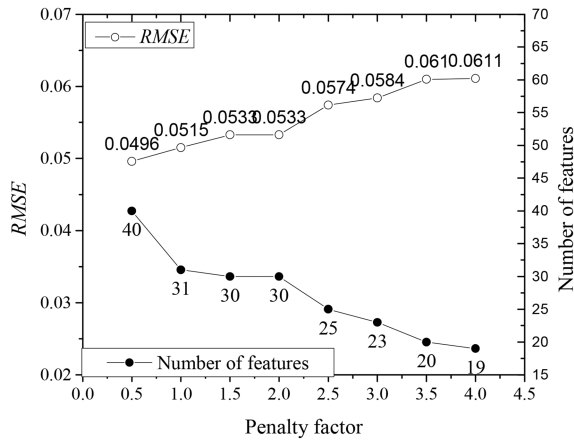


Fig. 4 Tuning of the penalty factor of group Lasso regression

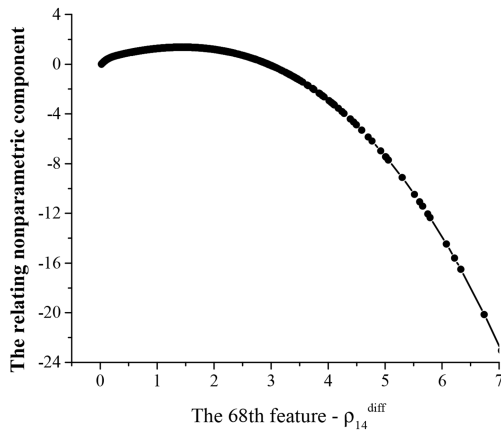


Fig. 5 Non-parametric component relating to  $\rho_{14}^{\text{diff}}$

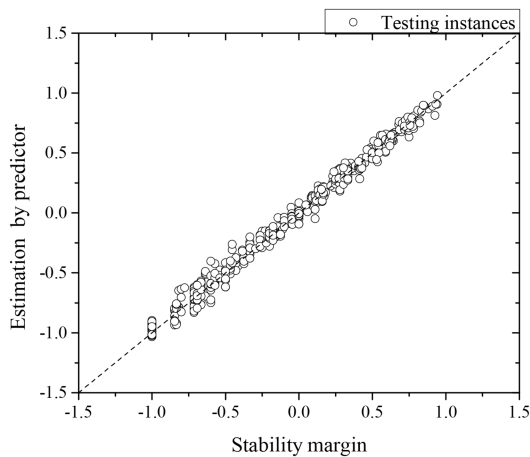


Fig. 6 Regression between the stability margin and its estimation by predictor for the testing data

$$f_{68}(x_{68}) = \begin{cases} 14.64x_{68}^3 - 13.98x_{68}^2 + 5.614x_{68} - 0.116, & x_{68} < 0.3027; \\ 0.211x_{68}^3 - 0.875x_{68}^2 + 1.646x_{68} + 0.284, & 0.3027 \leq x_{68} < 0.6833; \\ -0.043x_{68}^3 - 0.354x_{68}^2 + 1.290x_{68} + 0.365, & x_{68} \geq 0.6833. \end{cases} \quad (22)$$

After predictor training, the testing data is used to assess the goodness of the trained predictor. The stability margin is estimated by feeding the selected severity indicators into the non-parametric predictor. The regression between the stability margin  $\eta$  and the estimated value  $\hat{\eta}$  is plotted in Fig. 6. Each scatter in Fig. 6

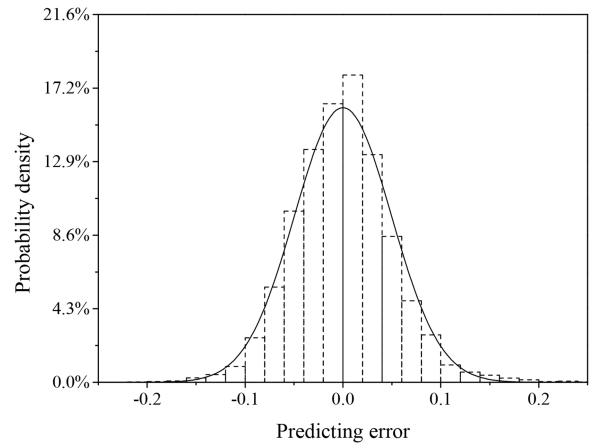


Fig. 7 Histogram of the predicting error and the fitted curve of the normal probability density function

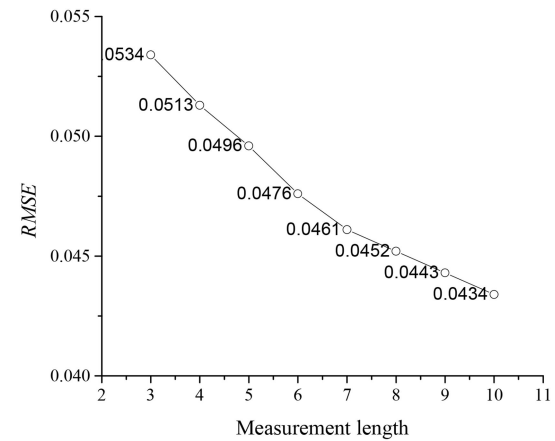


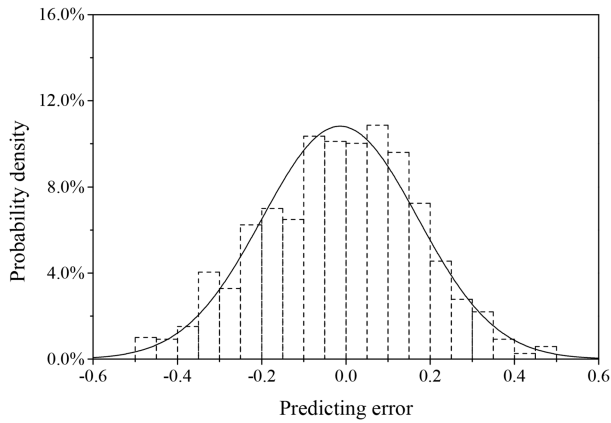
Fig. 8 Impact of measurement length on the predicting accuracy

represents an instance in the testing set. As can be seen from Fig. 6, all the scatters locate close to the line of  $\hat{\eta} = \eta$ , indicating the predictor has fitted the correlation between the stability margin and the severity indicators properly.

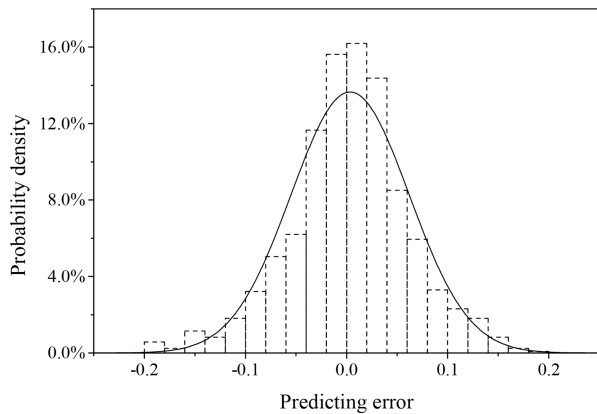
Statistical analysis is performed to investigate the probability distribution of predicting error. The histogram of predicting error is shown in Fig. 7. The probability distribution of predicting error follows a normal distribution. The mean  $\mu$  of predicting error is 0 while its standard deviation  $\sigma$  is 0.0496. According to the characteristics of normal distribution, the  $\mu \pm 2\sigma$  interval includes 95% of all the instances. Therefore, when applying the predictor to unseen instances, the predicting error is expected to be  $< \pm 0.0992$  at the confidence level of 95%.

#### 4.5 Impact of measurement length

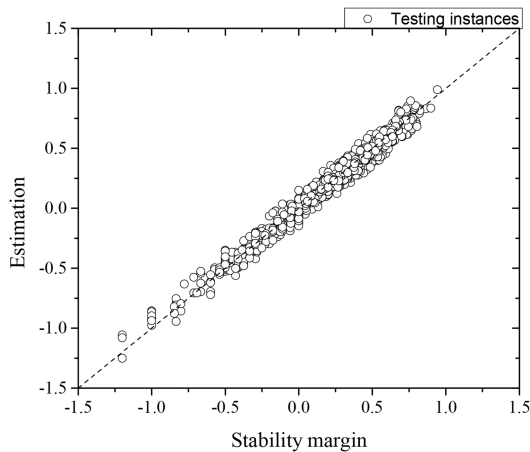
Either the 7th or the 14th severity indicator relies on the length of the observation window, so the impact of measurement length on predictor performance cannot be ignored. Predictors of different measurement lengths are trained under an identical set of input features. The performance of these predictors is evaluated by RMSE and the comparative result is shown in Fig. 8. With the observation window lengthened, the predicting error decreases and the performances of predictors are enhanced. For fast evaluation of stability margin, while maintaining the expected maximum predicting error should be  $< 0.1$ , the measurement length is chosen as five cycles. Notably, the measurement length in the existing literature varies from four cycles to ten cycles [16, 17, 19, 20]. Therefore, the observation window of the proposed TSA predictor is comparative with the existing predictors of other data-mining techniques.



**Fig. 9** Histogram of the predicting error for the noisy testing data (predictor trained by noise-free training data)



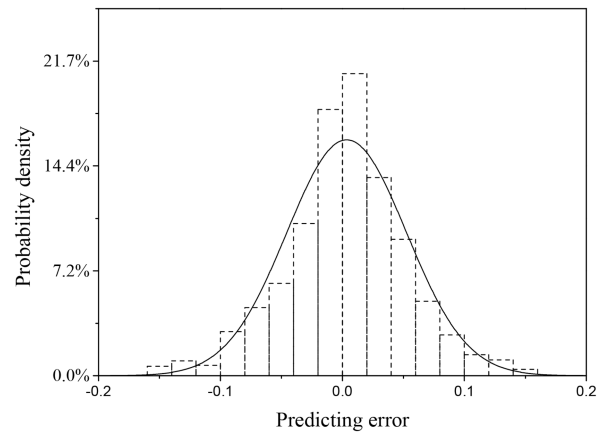
**Fig. 10** Histogram of the predicting error for the noisy testing data (predictor re-trained by noisy training data)



**Fig. 11** Regression between the stability margin and its estimation by predictor for the testing data of  $N-1$  scenarios

#### 4.6 Impact of measurement error

According to the ‘IEEE Standard for Synchrophasors for Power Systems’, PMUs with level 1 compliance should have a total vector error  $<1\%$  [29]. To study the impact of measurement error on predictor performance, a random error between  $-1$  and  $1\%$  is added to the simulated post-fault trajectories in the testing data. With the observation window to be five cycles, the indicators are then re-computed by these noisy trajectories. The histogram of the predicting error for the noisy testing data is shown in Fig. 9. The probability distribution of the predicting error follows a normal distribution. The mean and the standard deviation of the predicting error is  $-0.0138$  and  $0.1843$ , respectively. Clearly, the predicting performance on the noisy testing data is worst than that on the noise-free testing data.



**Fig. 12** Histogram of the predicting error and the fitted curve of the normal probability density function

**Table 2** Comparison of the RMSE between different TSA predictors

Methods	The proposed method	MLR	BPNN	RT	SVR
training data	0.0495	0.0693	0.0396	0.0272	0.1076
testing data	0.0499	0.0715	0.0480	0.0721	0.0824
overall	0.0496	0.0702	0.0431	0.0603	0.1060

However, if the TSA predictor is re-trained by using the noisy training data, the predicting performance is improved. In this case, the mean and the standard deviation of the predicting error are  $0.0035$  and  $0.0584$ , respectively. The histogram of the predicting error, in this case, is shown in Fig. 10. A similar conclusion about the impact of measurement errors can be found in [30].

#### 4.7 Impact of topology change

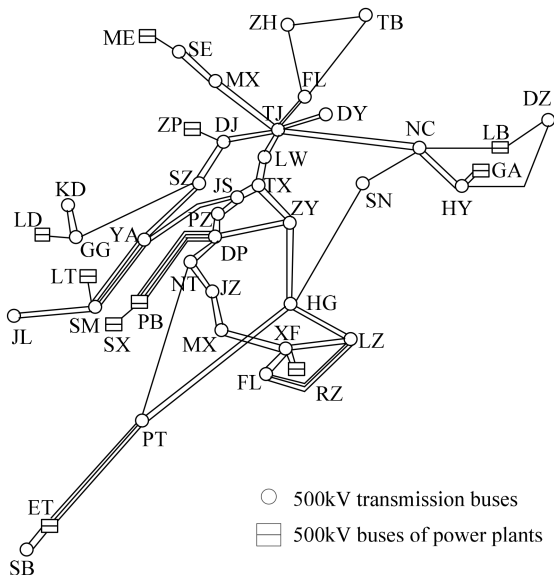
Topology change has a significant impact on the post-fault stability of power systems. Therefore, the robustness of TSA predictors should be tested under the scenarios of topology change. For this propose, 1000 instances are generated under stochastic  $N-1$  pre-fault OCs. The predictions of these instances and distribution of the prediction error are shown, respectively, in Figs. 11 and 12. The RMSE of these instances is  $0.0498$ , thus it can be concluded that the predictor has shown a consistent performance in  $N-1$  condition.

#### 4.8 Comparison of different predictors

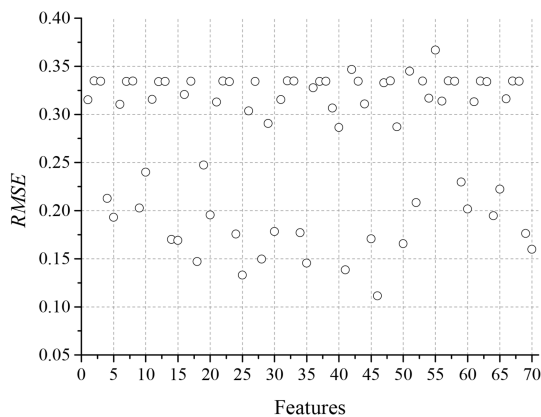
The comparison is performed among the proposed method, multi-linear regression (MLR), NN, regression tree (RT) and support vector regression (SVR). An identical set of instances is utilised for predictor training with the rest to be tested. The predicting error of these techniques is provided by Table 2. The NN-based predictor has the best overall performance among all the techniques. The RT-based predictor is an unpruned tree with 1471 children nodes and thus suffers from overfitting badly. There is a trade-off between the accuracy and the transparency of data-mining technologies [11]. The proposed non-parametric predictor has an explicit formulation rather than a ‘black-box’ projection, which provides better interpretation while maintaining comparative evaluation accuracy with NN.

### 5 Application in a practical system

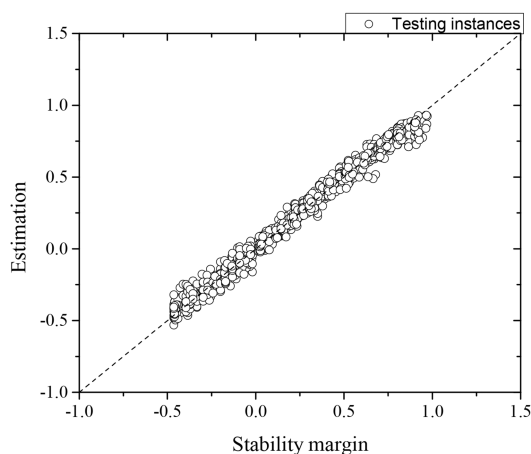
The proposed scheme is also applied to a practical 756-bus transmission system in China [14]. The 500 kV backbone network is demonstrated in Fig. 13. The detailed dynamic models for synchronous generators, turbines, speed governors, excitation systems and power system stabilisers are used for simulation.



**Fig. 13** 500 kV bulk grid of the practical transmission system



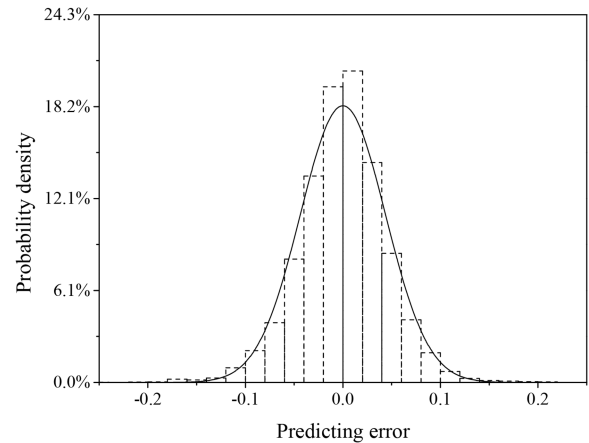
**Fig. 14** Correlation strength evaluated by NIS in the practical transmission system case



**Fig. 15** Regression between the stability margin and its estimation by predictor for the testing data in the practical transmission system case

### 5.1 Data generation

Following the proposed TSA scheme and the previous New England 39-bus system case, a database of OCs is first generated by the OCs of different loading scenarios and the stochastic OCs. Fault location is randomly selected among 500 kV transmission lines while the fault clearing time is also modelled as a normal distribution with a mean value of 0.2 s and a standard deviation of 0.02 s. The length of the observation window is set to be five cycles. TDS is performed to simulate the post-fault responses and



**Fig. 16** Histogram of the predicting error and the fitted curve of the normal probability density function in the practical transmission system case

compute the severity indicators and the transient stability margin. About 2500 TSA instances under normal OCs and 500 TSA instances under  $N-1$  OCs are generated.

### 5.2 Feature selection

About 2000 instances of normal OCs are used for predictor training while the others including the rest instances of normal OCs, and all the instances of  $N-1$  OCs serve as testing data. On the training data, NIS is performed to determine the structure of the additive model and identify the severity indicators that are weakly correlated to the stability margin. The result of correlation evaluation is shown in Fig. 14.

The threshold for NIS and the penalty factor for group Lasso are tuned to be 0.25 and 0.5, respectively. Group Lasso algorithm is employed to carry out the second stage of feature selection. Eventually, 27 indicators are selected to form the input space for predictor training.

### 5.3 Predictor training and application results

A TSA predictor is then trained with the selected severity indicators. The stability margin of all the testing instances are estimated by the trained predictor and the RMSE of these instances is 0.0437. The regression between the stability margin and the estimation for testing instances is shown in Fig. 15 and the distributions of the predicting error are given by Fig. 16. It can be concluded that the proposed non-parametric statistics-based scheme can also provide a promising assessment of post-fault stability margin in the larger-scale practical transmission system.

## 6 Conclusion

A non-parametric statistics based scheme is proposed for response-based online TSA in power systems. To train a TSA predictor, the CCT-based transient stability margin and 14 kinds of severity indicators are defined as the predictive response and the input features, respectively. To address the problem of lack of prior knowledge of correlation structure, the non-parametric additive model is used as the basis of the TSA predictor. To screen out the indicators that do not significantly help improve the performance of the predictor, two-stage feature selection is fulfilled by successively NIS and group Lasso penalised regression. After that, the predictor is learnt by component-wise least-squares regression in the reduced multi-feature space. With PMU measurements at generator buses, severity indicators can be computed in the real-time and fast evaluation of post-fault stability margin can be made by the offline-trained predictor.

Case studies on the modified New England 39-bus system and the practical 756-bus transmission system in China are provided to illustrate the effectiveness of the proposed scheme. Numerical results demonstrate that:



(1) with the observation window to be five cycles, the RMSEs for the well-trained predictors in these test systems are 0.0496 and 0.0437, which indicates the expected maximum error to be  $<0.1$  and satisfies the need of accurate estimation of post-fault stability margin;

(2) the proposed two-stage non-parametric statistics based feature selection scheme can identify the weakly correlated indicators and thus reduce the dimension of input features; and

(3) the proposed non-parametric TSA predictor has the ability to adapt to normal and possible  $N - 1$  pre-fault OCs.

Future research emphasis lies in two aspects. First, PMU measurements at all the generator buses are assumed to be available in this paper. For practical application, no PMU measurements at some generator buses or data missing problem should be addressed. So, the robustness of the TSA predictor under incomplete observation should be studied and improved. Second, remedial control schemes should be taken when the post-fault system is about to lose synchronism. Hence a data-mining-based scheme for response-based control should be studied to incorporate with the proposed TSA predictor.

## 7 Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 51437003).

## 8 References

- [1] Kundur, P., Paserba, J., Ajarapu, V., *et al.*: 'Definition and classification of power system stability', *IEEE Trans. Power Syst.*, 2004, **19**, (2), pp. 1387–1401
- [2] Jiang, N., Chiang, H.-D.: 'Energy function for power system with detailed DC model: construction and analysis', *IEEE Trans. Power Syst.*, 2013, **28**, (4), pp. 3756–3764
- [3] Odun-Ayo, T., Crow, M.L.: 'Structure-preserved power system transient stability using stochastic energy functions', *IEEE Trans. Power Syst.*, 2012, **27**, (3), pp. 1450–1458
- [4] Bhui, P., Senroy, N.: 'Real-time prediction and control of transient stability using transient energy function', *IEEE Trans. Power Syst.*, 2017, **32**, (2), pp. 923–934
- [5] Padiyar, K., Krishna, S.: 'Online detection of loss of synchronism using energy function criterion', *IEEE Trans. Power Deliv.*, 2006, **21**, (1), pp. 46–55
- [6] Gou, J., Liu, Y., Liu, J., *et al.*: 'Novel pair-wise relative energy function for transient stability analysis and real-time emergency control', *IET Gener. Transm. Distrib.*, 2017, **11**, (18), pp. 4565–4575
- [7] Yan, J., Liu, C.-C., Vaidya, U.: 'PMU-based monitoring of rotor angle dynamics', *IEEE Trans. Power Syst.*, 2011, **26**, (4), pp. 2125–2133
- [8] Dasgupta, S., Paramasivam, M., Vaidya, U., *et al.*: 'PMU-based model-free approach for real-time rotor angle monitoring', *IEEE Trans. Power Syst.*, 2015, **30**, (5), pp. 2818–2819
- [9] Wei, S., Yang, M., Qi, J., *et al.*: 'Model-free MLE estimation for online rotor angle stability assessment with PMU data', *IEEE Trans. Power Syst.*, 2018, **33**, (3), pp. 2463–2476
- [10] Kamwa, I., Samantaray, S.R., Joos, G.: 'Catastrophe predictors from ensemble decision-tree learning of wide-area severity indices', *IEEE Trans. Smart Grid*, 2010, **1**, (2), pp. 144–158
- [11] Kamwa, I., Samantaray, S.R., Joos, G.: 'On the accuracy versus transparency trade-off of data-mining models for fast-response PMU-based catastrophe predictors', *IEEE Trans. Smart Grid*, 2012, **3**, (1), pp. 152–161
- [12] Amraee, T., Ranjbar, S.: 'Transient instability prediction using decision tree technique', *IEEE Trans. Power Syst.*, 2013, **28**, (3), pp. 3028–3037
- [13] Yang, Y., Huang, Y., Liu, J., *et al.*: 'Measurements based cell-DT method for power system transient stability classification', *CSEE J. Power Energy Syst.*, 2017, **3**, (3), pp. 278–285
- [14] Papadopoulos, P.N., Guo, T., Milanović, J.V.: 'Probabilistic framework for online identification of dynamic behavior of power systems with renewable generation', *IEEE Trans. Power Syst.*, 2018, **33**, (1), pp. 45–54
- [15] Gao, Q., Rovnyak, S.M.: 'Decision trees using synchronized phasor measurements for wide-area response-based control', *IEEE Trans. Power Syst.*, 2011, **26**, (2), pp. 855–861
- [16] Rajapakse, A.D., Gomez, F., Nanayakkara, K., *et al.*: 'Rotor angle instability prediction using post-disturbance voltage trajectories', *IEEE Trans. Power Syst.*, 2010, **25**, (2), pp. 947–956
- [17] Geeganage, J., Annakkage, U.D., Weekes, T., *et al.*: 'Application of energy-based power system features for dynamic security assessment', *IEEE Trans. Power Syst.*, 2015, **30**, (4), pp. 1957–1965
- [18] Wang, B., Fang, B., Wang, Y., *et al.*: 'Power system transient stability assessment based on big data and the core vector machine', *IEEE Trans. Smart Grid*, 2016, **7**, (5), pp. 2561–2570
- [19] Zhang, R., Xu, Y., Dong, Z.Y., *et al.*: 'Post-disturbance transient stability assessment of power systems by a self-adaptive intelligent system', *IET Gener. Transm. Distrib.*, 2015, **9**, (3), pp. 296–305
- [20] Yu, J.J.Q., Hill, D.J., Lam, A.Y.S., *et al.*: 'Intelligent time adaptive transient stability assessment system', *IEEE Trans. Power Syst.*, 2018, **33**, (1), pp. 1049–1058
- [21] Liu, Y., Liu, Y., Liu, J., *et al.*: 'High-performance predictor for critical unstable generators based on scalable parallelized neural networks', *J. Mod. Power Syst. Clean Energy*, 2016, **4**, (3), pp. 414–426
- [22] Rahmatian, M., Chen, Y.C., Palizban, A., *et al.*: 'Transient stability assessment via decision trees and multivariate adaptive regression splines', *Electr. Power Syst. Res.*, 2017, **142**, pp. 320–328
- [23] Liu, T., Liu, Y., Liu, J., *et al.*: 'Multi-indicator inference scheme for fuzzy assessment of power system transient stability', *CSEE J. Power Energy Syst.*, 2016, **2**, (3), pp. 1–9
- [24] Fu, C.J., Bose, A.: 'Contingency ranking based on severity indices in dynamic security analysis', *IEEE Trans. Power Syst.*, 1999, **14**, (3), pp. 980–986
- [25] Li, G., Rovnyak, S.M.: 'Integral square generator angle index for stability ranking and control', *IEEE Trans. Power Syst.*, 2005, **20**, (2), pp. 926–934
- [26] Alvarez, J.M., Mercado, P.E.: 'Online inference of the dynamic security level of power systems using fuzzy techniques', *IEEE Trans. Power Syst.*, 2007, **22**, (2), pp. 717–726
- [27] Fan, J., Feng, Y., Song, R.: 'Nonparametric independence screening in sparse ultra-high-dimensional additive models', *J. Am. Stat. Assoc.*, 2011, **106**, (494), pp. 544–556
- [28] Huang, J., Horowitz, J.L., Wei, F.: 'Variable selection in nonparametric additive models', *Ann. Stat.*, 2010, **38**, (4), pp. 2282–2312
- [29] Martin, K.E., Benmouyal, G., Adamiak, M.G., *et al.*: 'IEEE standard for synchrophasor for power systems', *IEEE Trans. Power Deliv.*, 1998, **13**, (1), pp. 73–77
- [30] Gomez, F.R., Rajapakse, A.D., Annakkage, U.D., *et al.*: 'Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements', *IEEE Trans. Power Syst.*, 2011, **26**, (3), pp. 1474–1483