A Fault Detection Method Based on Dynamic Peak-valley Limit under the Non-Steady Conditions

TIAN-ZHEN WANG[1 2], MAN XU[1], TIAN-HAO TANG[1], CHRISTOPHE CLARAMUNT[2]
[1]Logistics Engineering Department
Shanghai Maritime University
NO.1550 Haigang Road, Pudong New Area, Shanghai, 200135,
CHINA
wtz0@sina.com, xumanaa_2006@126.com and thtang@shmtu.edu.cn

[2]Naval Academy Research Institute of France
BP 600, Lanveoc-Poulmic, Brest, 29240,
FRANCE
christophe.claramunt@ecole-navale.fr

Abstract: - The multivariate statistical methods are commonly used to fault detection through a straight limit line given by the HotellingT^2. However, the traditional straight limit line is difficult to detect the fault effectively under the non-steady conditions, and the rate of false alarm and missing alarm is high. For these problems above, a fault detection method based on dynamic peak-valley limit is proposed in this paper. The proposed method introduces relative principal component analysis (RPCA) to carry out data dimension reduction, extract principal component (PCs) and calculate T^2 statistics, then adopts moving least squares (MLS) to preprocess T^2 statistics to obtain the fitting curve which is called peak-valley curve, and finally connects peak and valley points in the curve to construct another control limit, by introducing a weight combined with the traditional straight limit line to construct the dynamic peak-valley limit. At the end, it is applied to wind power generation system, and the results could verify the effectiveness of the method.

Key-Words: - T^2-statistic, Dynamic Peak-valley Limit, RPCA, MLS, Peak-valley Curve, Fault Detect

1 Introduction

With the development of economic construction and improvement of electrification, modern industrial systems become increasingly complex. The reliability and safety of these systems are more and more concerned by people. If fault was discovered early, major failure and accidents could be avoided, so the fault detection and fault diagnosis become more important. Many scholars have done a lot of research, including data fusion, artificial immune, multivariate statistical methods and so on [1-3]. Fault diagnosis based on multivariate statistical is a kind of effective method. The two most commonly used statistical methods are principal component analysis (PCA) and partial least squares (PLS) [4,5], but they are generally suitable for steady working conditions, and the model is fixed. For the time-varying industrial process [6], recursive PCA and exponentially weighted PCA (EWPCA) can adaptively update control limit [7,8], thus have received widespread concern. But if the complex system is under the non-steady conditions (The non-steady conditions are referred to such motion process as starting, braking and other mutation conditions [9]), the above methods fail to detect fault effectively and computational load is large [10]. However, the actual industrial process is often unsteady, such as wind power generation system [11], because of the season, air pressure and topography influence, the speed of wind changes random and unsteady so significantly that the role of the wind vane is also random and unsteady. In addition, the failure probability of the system under the non-steady conditions is relatively high.

In this paper, we present a new method of fault detection based on the dynamic peak-valley limit to deal with the shortcomings of the traditional methods under the non-steady conditions, and the paper is organized as follows.

Section 2 will give a detailed review on data dimension reduction method based on RPCA and how to calculate the T^2 statistics. Section 3 introduces MLS method to preprocess T^2 statistic to get peak-valley curve. Section 4 will propose the fault detection method based on dynamic peak-valley limit, including the established procedure of the dynamic peak-valley limit and its fault detection process. In section 5, the model of wind power generation system will be built and its brief introduction will be given too. In section 6, we apply this method to the wind power generation system under the non-steady conditions to test the validity of the method and discuss its test results. Eventually, the conclusion is summarized in section 7.

2 RPCA and T² statistic

Normally the traditional principal component analysis method could detect and diagnose fault, but if the data is in "rotundity" scatter after dimensionless standardization, then it is difficult to extract the typical PCs and establish an effective fault detection model. For these problems, Tian-zhen Wang et al. proposed a relative principal component analysis method and the concepts of relative transform, relative principal component, "rotundity" scatter and so on [12,13]. The new concepts and new method could effectively overcome the shortcomings of the traditional PCA in data compression and fault detection, and have the following advantages:

1) If system variables distribute uniformly after dimensionless standardization, RPCA method not only could solve this problem, but could get the more representative relative principal component, which could achieve the purpose of data compression and fault detection.

2) To avoid the condition that large covariance variable would play a major role in selecting principal component, when the data is in different dimensional sense.

3) The energy of the system is not necessarily conserved after dimensionless standardization, but could keep conservation after the relative translation.

2.1 RPCA Algorithm

Given $X \in R^{N\times n}$ as a data matrix of system variable sequence, n is the number of variables, N is the number of samples.

Let

$$X^R = X^* \times M$$
$$= \begin{bmatrix} x_1^*(1) & x_2^*(1) & \cdots & x_n^*(1) \\ x_1^*(2) & x_2^*(2) & \cdots & x_2^*(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1^*(N) & x_2^*(2) & \cdots & x_n^*(N) \end{bmatrix} \times \begin{bmatrix} M_1 & 0 & \cdots & 0 \\ 0 & M_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & M_n \end{bmatrix} \quad (1)$$
$$= \begin{bmatrix} x_1^R(1) & x_2^R(1) & \cdots & x_n^R(1) \\ x_1^R(2) & x_2^R(2) & \cdots & x_n^R(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1^R(N) & x_2^R(N) & \cdots & x_n^R(N) \end{bmatrix}$$

Then Eq.(1) is called relative translation and could be written as the following matrix form.

$$X_i^R = M_i X_i^* \quad (2)$$

where

$$X_i^* = \frac{X_i - E(X_i)}{\zeta_i} \quad (3)$$

In Eq.(3), ζ_i is the standardization factor (Standardization factor has a variety of choices, such as $\zeta_i = \max_{1\le j\le N}|x_i(j)|$ or $\zeta_i = (\mathrm{var}(X_i))^{1/2}$), X^R is the relative matrix of matrix X, and M_i is the specific gravity coefficient of relative transform and its selection method is in [12].

The main operation process of RPCA could be summarized by the following steps:

1) Computing the covariance matrix R_{X^R} of X^R from Eq.(4).

$$R_{X^R} = E\{[X^R - E(X^R)]^T [X^R - E(X^R)]\} \quad (4)$$

2) Calculating relative eigenvalue λ and its eigenvector p respectively by

$$|\lambda I - R_{X^R}| = 0 \quad (5)$$

and

$$|\lambda_i I - R_{X^R}| p_i = 0 \quad i = 1, 2, \cdots, n \quad (6)$$

where

$$p_i = [p_i(1) \quad p_i(2) \quad \cdots \quad p_i(n)]^T \quad (7)$$

λ_i is the ith eigenvalue of R_{X^R}, and p_i is the corresponding eigenvector, suppose $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$.

3) Selecting the number of relative PCs according to the cumulative contribution rate

$$\delta\% = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \times 100\% \quad (8)$$

Where, m is the number of selected PCs, and the value of δ is determined by the user.

2.2 T^2 statistics

Calculate T^2 statistics by the above RPCA algorithm

$$T_i^2 = t_i \Lambda_m^{-1} t_i^T = X_i^R P_m \Lambda_m^{-1} P_m^T (X_i^R)^T \tag{9}$$

In Eq.(9), $\Lambda_m = diag(\lambda_1, \cdots, \lambda_m)$ is diagonal matrix, m is the number of selected PCs, $t_i = X_i^R P_m$ is the score vector, p_j is the load vector and also is the eigenvector of R_{X^R}, then $P_m = [p_1 \quad p_2 \quad \cdots \quad p_m]$

The T^2 control limit could be calculated by using F distribution, and its equation is as follows:

$$T_{ucl1} = \frac{m(N-1)}{N-m} F_\alpha (m, N-m) \tag{10}$$

Where, T_{ucl1} is the upper limit of T^2 statistic, N is the number of samples, and m is the number of reserved PCs. Confidence $1-\alpha$ could be determined by the user need, for example, if confidence is 95% or 99%, then $\alpha = 0.05$ or 0.01. $F_\alpha(m, N-m)$ is the critical value of F distribution corresponding to the test level α, the degree of freedom m and $N-m$.

3 Data Preprocessing

Compared with the traditional curve fitting method, moving least squares mainly has the following advantages:

1) There is no need to do piecewise fitting and smoothing, when the number of scattered data is large or its shape is complex.

2) Take different basis function for different accuracy and different weight function for changing the smoothness of the fitting curve, and other fitting methods could not do these.

Therefore, this paper chooses MLS method for data preprocessing. In this section, data preprocessing includes three aspects, that is, the introduction of MLS principle, the selection of parameters such as basic functions, weight functions and support radius and the specific process of curve fitting.

3.1 MLS Principle

MLS method is based on the least square method, but has a large improvement [14-17]. First of all, fitting function consists of basis function $q(x)$ and coefficient function $a(x)$, $a(x)$ is coordinate function. Secondly, introduce the concepts of weight function and compact support. Compact support means that the y value of point x is only influenced by nodes of the subdomain nearby point x, and the subdomain is called support domain of point x.

The fitting function $f(x)$ could be expressed below:

$$f(x) = \sum_{i=1}^{l} q_i(x) a_i(x) = q^T(x) a(x) \tag{11}$$

Where $a(x) = (a_1(x), a_2(x), \ldots, a_l(x))^T$ is a set of coefficients, $q(x) = (q_1(x), q_2(x), \ldots, q_l(x))^T$ is the basis function vector, and basis function usually need to select the complete polynomial base, for example 1D case

Linear basis: $q(x) = (1, x)^T \qquad l = 2$

Quadratic basis: $q(x) = (1, x, x^2)^T \qquad l = 3$

In order to get the more accurate local approximation, minimize a weighted square of discrete error of the local approximation $f(x_i)$ and y_i of node value expressed as

$$\begin{aligned} J &= \sum_{i=1}^{k} w(x - x_i)[f(x) - y_i]^2 \\ &= \sum_{i=1}^{k} w(x - x_i)[q^T(x) a(x) - y_i]^2 \end{aligned} \tag{12}$$

Where k is the number of nodes in the influence domain, $f(x)$ is the fitting function and $w(x - x_i)$ is weight function.

In matrix form, (12) could be rewritten as

$$J = (Qa(x) - Y)^T W(x) (Qa(x) - Y) \tag{13}$$

Where, $Y = (y_1, y_2, \cdots, y_k)^T$, $W(x) = diag(w_1(x), w_2(x), \cdots, w_k(x))$, $w_i(x) = w(x - x_i)$,

$$Q = \begin{bmatrix} q_1(x_1) & q_2(x_1) & \cdots & q_l(x_1) \\ q_1(x_2) & q_2(x_2) & \cdots & q_l(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ q_1(x_k) & q_2(x_k) & \cdots & q_l(x_k) \end{bmatrix}$$

Minimizing J with the respect to coefficients $a(x)$, then according to the least square principle, the equation (14) could be obtained.

$$a(x) = A^{-1}(x) B(x) Y \tag{14}$$

Where $A(x) = Q^T W(x) Q$, $B(x) = Q^T W(x)$. Substituting Eq.(14) into Eq.(11) to rearrange the equation, Eq.(11) could be re written as

$$f(x) = \sum_{i=1}^{k} \phi_i^c(x) y_i = \psi^c(x) Y \tag{15}$$

Where, $\psi^c(x)$ is known as the moving least square shape function, c is the order number of basis function.

$$\psi^c(x) = [\phi_1^c, \phi_2^c, \cdots, \phi_k^c] = q^T(x) A^{-1}(x) B(x) \quad (16)$$

3.2 Parameters Selection

By the MLS principle, it is known that the fitting function $f(x)$ is mainly determined by the basis function and weight function. Because of taking different basis function will get different accuracy and different weight function will change the smoothness of the fitting curve, this section will mainly explore how to select suitable parameters.

According to reference [18], basis function usually adopt the form of power function, and the number of the power function is commonly between once and three times. Increasing the number of basis function could improve the precision of the calculation, but it also brings so many other problems: not only significantly increase the amount of computation and reduce the calculation accuracy, but also the coefficient matrix obtained by the least square method generally appears morbid in a certain degree. Considering these problems, low time basis function could be chosen. Quadratic basis is selected as the basis function in this paper.

Weight function in the MLS method plays a very important role. First weight function should be positive and decrease monotonically with the increase of $\|x - x_i\|$. Weight function should also have compact support, namely, in support domain (the influence area of x) is not equal to zero, but is zero outside the domain. Generally choose round as the support domain, and support radius is r. Weight function has a variety of types, such as exponential weight function, spline weight function, trigonometric weight function, Gaussian weight function and so on. Because of the continuity of the spline function is good and the accuracy of the got best result is the highest, so spline function is commonly used. However, the number of spline function divides into three times, four times, five times, etc. But high-order spline function computes complex [19,20], and the result is not necessarily better, the cubic spline could generally meet the requirements of the curve fitting, so in this paper weight function choose cubic spline function. Set $s' = x - x_i, s = \frac{s'}{r}$, its expression is as follows

$$w(s) = \begin{cases} \frac{2}{3} - 4s^2 + 4s^3 & s \le \frac{1}{2} \\ \frac{4}{3} - 4s + 4s^2 - \frac{4}{3}s^3 & \frac{1}{2} < s \le 1 \\ 0 & s > 1 \end{cases} \quad (17)$$

Support radius r of the weight function is another factor to influence curve fitting, namely the range of compact support domain. Appropriate support radius should include the nodes in the influence “enough and as little as possible”, so that could highlight local fitting effect and improve the calculation efficiency. In order to maintain the approximate locality of MLS, the support radius of node x should be as small as possible, at the same time, in order to guarantee matrix $A(x)$ reversible, radius should be large enough to include sufficient points in the definition area of each node. Generally the support radius should make sure that support domain contains $l+1$ nodes at least (l is the item number of basis function). This paper selects quadratic basis, namely $l = 3$, so the support domain should contains four nodes at least, and the nodes are evenly distributed, then $r = 2$.

3.3 Curve Fitting Based on MLS

The basic idea of curve fitting based on MLS is to equally divide fitting region between adjacent nodes into d calculation points firstly, then calculate the value of the calculation points by using Eq.(15), finally connect all the calculation points to form a smooth fitting curve, namely peak-valley curve and its characteristics could be found in reference [9]. Fig.1 is the flow chart of curve fitting and the design process is as follows:

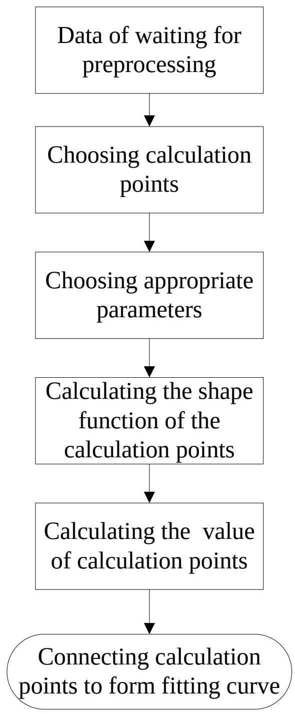

Fig.1 Flow chart of curve fitting

1) Divide the region between adjacent nodes into

d segments evenly, these segment points are called calculation points ($d=10$ in this paper).

2) Select the appropriate basis function $q(x)$, weight function $w(x)$, support radius r.

3) Calculate shape function $\psi^c(x)$ of each calculation point by Eq.(16).

4) Calculate the value of each calculation point by Eq.(15).

5) Connect all calculation points to form peak-valley curve $f(x)$.

4 A Fault Detection Method Based on Dynamic Peak-valley Limit under the Non-steady Conditions

The T^2 control limit of fault detection method based on RPCA is only related to the number of PCs and confidence $1-\alpha$, so it often remains unchanged or a constant in a sampling period and it is easy to occur the phenomena of false alarm or missing alarm when applied to monitor non-steady conditions.

So, dynamic peak-valley limit is proposed. According to curve function $f(x)$, peak points and valley points could be obtained firstly, then these points are connected one by one to construct a new control limit combined with the traditional T^2 control limit based on RPCA by introducing a weight to get the dynamic peak-valley limit. Finally, the process of the fault detection based on dynamic peak-valley limit is given.

4.1 The Establishment of Dynamic Peak-valley Limit

According to section 3, the function $f(x)$ of peak-valley curve could be obtained, $f'(x)$ is the derivative of function $f(x)$. Let

$$f'(x)=0 \tag{18}$$

According to Eq.(18), the abscissas of a set of peak points and valley points are obtained. Suppose $a_j\,(j=1,2,...,g)$, $b_j\,(j=1,2,...,h)$ ($g,h\in N$)are the abscissas of peak points and valley points in the curve respectively. Then according to the size of the abscissas, connect these points ordinal to form control limit T_{ucl2}.

T_{ucl1} is a straight line, there exists a lot of missing alarm and the detection sensitivity is not high. T_{ucl2} is close to T^2 statistic, so there exists a lot of missing alarm and the detection sensitivity is too high. Dynamic peak-valley limit combines the T^2 control limit T_{ucl1} and the control limit T_{ucl2} by introducing a weight ω, could effectively reduce the false alarm rate and missing alarm rate and change the detection sensitivity through adjusting the size of weight ω. Its mathematical expression is as follows:

$$T_{ucl1}=\omega * T_{ucl1}+(1-\omega)*T_{ucl2} \tag{19}$$

Where, ω is weight, and $0<\omega<1$.

The detection limit is no longer a straight line under the non-steady conditions, but could changes with condition changing, and is mainly composed of peak and valley points, so it is called dynamic peak-valley limit. As is shown in Fig.2.

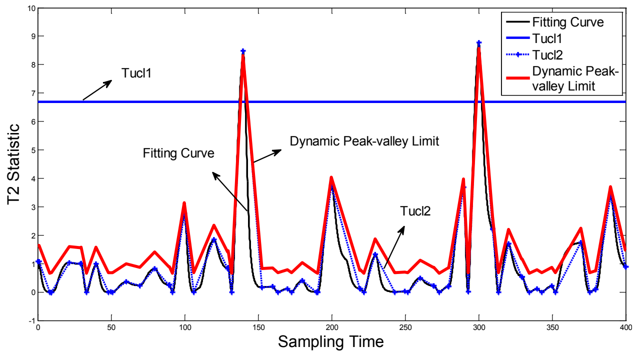

Fig.2 Construction of dynamic peak-valley limit

Dynamic peak-valley limit is constructed by the normal data, and the following steps will provide the detailed description.

STEP1. Sampling historic data

In the historical data set, collect a certain cycle length of historical data X_{ucl1}, the cycle length is N_t (the length of the sampling time).

STEP2. Dimension reduction process

1) Standardize data X_{ucl1}, then get X_{ucl}^R through relative transform and calculate covariance matrix $R_{X_{ucl}^R}$.

2) Calculate the eigenvalue λ and the corresponding eigenvector p of $R_{X_{ucl}^R}$, and the number m of PCs is decided by user need.

3) Calculate T^2 statistics according Eq.(9).

STEP3. Data preprocessing

1) Select the appropriate basis function $q(x)$, weight function $w(x)$ and support radius r.

2) Divide the region between adjacent nodes into d calculation points

STEP4. Constructing dynamic peak-valley limit

1) Calculate the value of T_{ucl1} according to Eq.(10).

2) Calculate the values of all peak points and valley points in the peak-valley curve, then connect all the points one by one to form the control limit T_{ucl2}.

3) Calculate dynamic peak-valley limit T_{ucl} according to Eq.(19).

4.2 Fault Detection Based on Dynamic Peak-valley Limit

We sample a group of measured data X_{test} online, then use the dynamic peak-valley limit constructed by the normal historical data X_{ucl} to process fault detection for the measured data. If the system is fault, output fault time, otherwise, the system is normal. Fig.3 is the flow chart of fault monitoring.

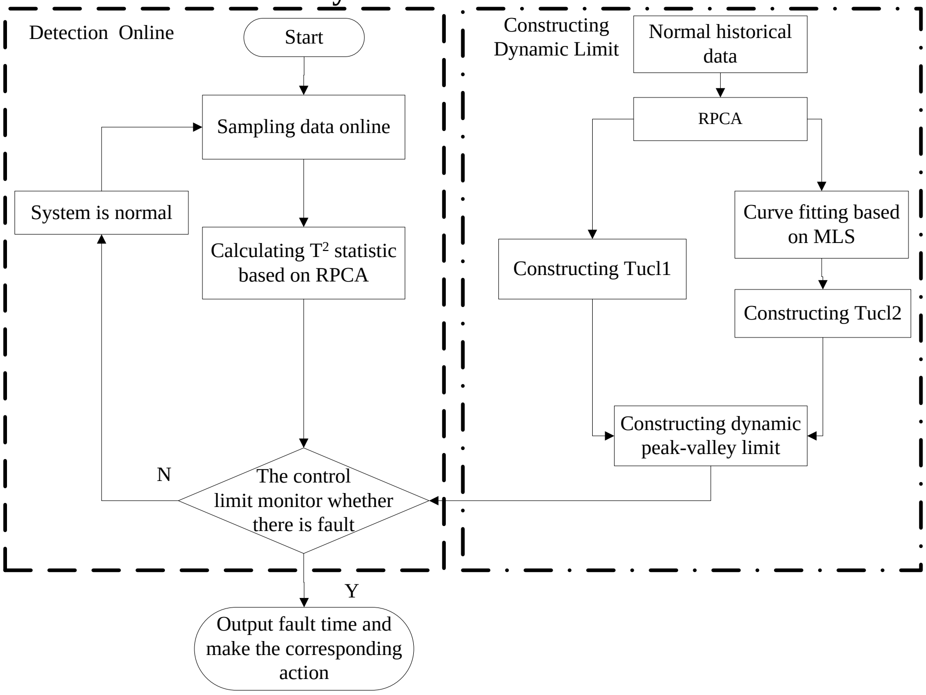

Fig.3 Flow chart of fault monitoring

The following steps will provide the detailed description of fault detection process.

STEP1. Sampling real-time data online

According the length of the sampling time N_t in Section 4.1 STEP1, sample a group of measured data X_{test} online.

STEP2. Dimension reduction process

Steps are same with STEP2 in section 4.1. Calculate T^2 statistics according to Eq.(9)

STEP3. Fault detection of system:

1) Detect whether T^2 statistics of X_{test} is over the control limit T_{ucl}, if it is true, then output the fault time and make the corresponding action according to the system requirements.

2) Otherwise, return to STEP1 and go on fault detection of the next process.

5 the Model of Wind Power Generation System

Wind power is a pollution-free renewable energy, which is exhaustless and wide distribution. With the requirement of ecological environment and energy, the development of wind energy is taken seriously increasingly. The wind power generation will be large-scale development in the 21st century [21,22].

Direct-drive permanent-magnet synchronous generation (DDPM) is one of a main direction of wind power generation [23,24]. The advantages of DDPM are as follows:

1) High operation efficiency;

2) Less control circuit and simple control;

3) Maintain the voltage of dc bus basically constant, and control electromagnetic torque of generator to adjust rotor speed;

4) Output constant frequency and the three-phase alternating current (AC) of the voltage, good adaptability to the fluctuation of power grid.

But the performance stability of permanent magnet material is high, if the weight of the motor increases, then the capacity of the inverter becomes larger. Those shortages lead that the cost of the generator is high. So generator fault could cause great economic loss, so it is very important to detect fault of DDPM. So we choose DDPM as fault detection model in this paper.

5.1 The Model of DDPM

DDPM is directly driven by variable-pitch wind turbines without gearbox. The output alternating current of generator converts to direct current by the high power electronic converter rectifier. Then, it becomes alternating current which frequency and voltage is constant. The output frequency and voltage of generator is along with the wind speed changes. So it Make the best use of the wind energy. The Fig.4 shows the structure of DDPM. The system is composed of wind turbine, rectifier, inverter and maximum power point tracking (MPPT) [25,26].

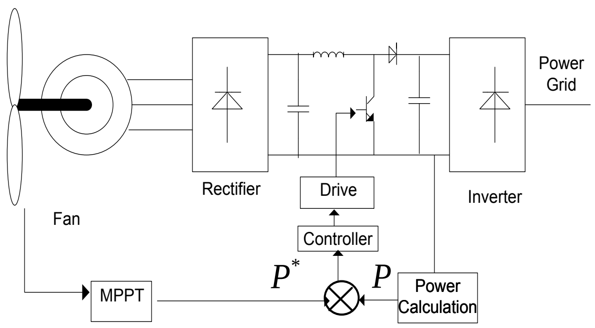

Fig.4 The model of DDPM

The paper use MATLAB to simulate the model of DDPM by for experiment. The simulation system is composed of wind turbine, MPPT, boost circuit, AC - DC conversion circuit. The Fig.5 shows the simulation model of DDPM.

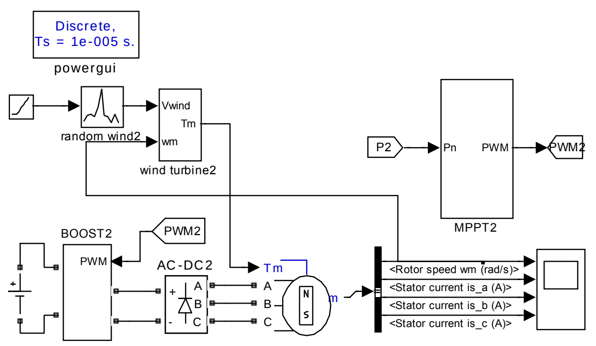

Fig.5 The simulation model of DDPM

According to the theory of the wind generator, the output of its mechanical power is:

$$P_{wind} = \frac{1}{2}\rho\pi R^2 C_p(\varepsilon,\gamma)V^3 \tag{20}$$

$C_p(\varepsilon,\gamma)$ is the conversion efficiency of wind energy and also is the function of tip speed ratio ε and pitch Angle γ, ρ is the air density, R is radius of the rotor, V is the wind speed, $\varepsilon = \frac{R\varpi}{V}$ is the ratio of Rotor tip speed $R\varpi$ and the wind speed V, ϖ is the speed of wind turbine rotor. The theoretical maximum limit of $C_p(\varepsilon,\gamma)$ is 59.3%. Considering the influence of wind speed fluctuation and wind direction fluctuation, $C_{p\max}$ is roughly 0.4 and difficultly exceeds 0.5. The expression is below:

$$C_p = 0.22(\frac{116}{\varepsilon_1} - 0.4\gamma - 5)e^{\frac{-12.5}{\varepsilon_1}} \tag{21}$$

$$\varepsilon_1 = 1/(\frac{1}{\varepsilon + 0.08\gamma} - \frac{0.035}{\gamma^3 + 1}) \tag{22}$$

$$T_{mwind} = \frac{P_{wind}}{\varpi} \tag{23}$$

According to Eqs.20-23 can get the simulation model of wind turbine as shown is Fig.6.

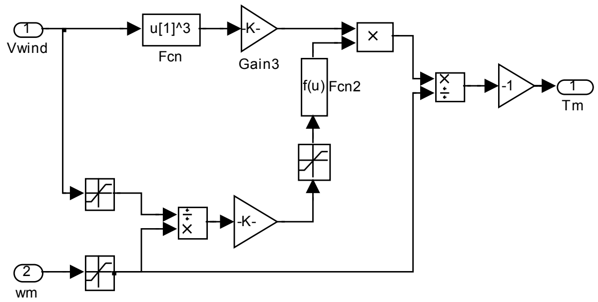

Fig.6 The simulation model of wind turbine

In order to absorb wind energy furthest, wind turbines always run in the maximum power point and the output power must match the captured mechanical power strictly. So MPPT must be added to the system, so as to obtain the conversion efficiency of the maximum wind energy. The Fig.7 is the simulation model of MPPT based on PSF.

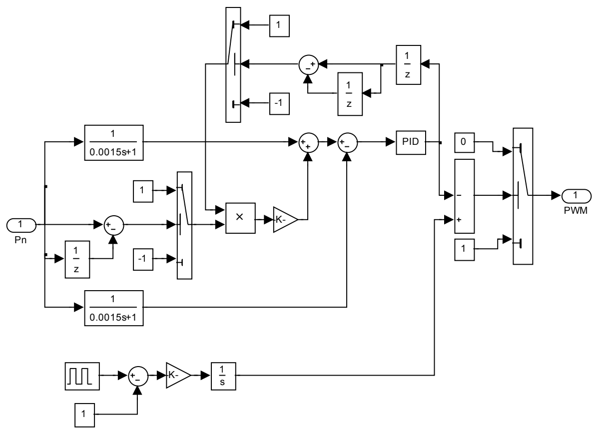

Fig.7 The simulation model of MPPT

Boost circuit is to boost dc power by using MPPT system to produce PWM wave. The boost converter will transform pulsating dc of the diode rectifier output to constant voltage. By adjusting PWM pulse, change the dutyfactor of switch tube and the function of load to match the load impedance, and then capture the maximum wind energy. The simulation model of boost circuit is shown in fig.8.

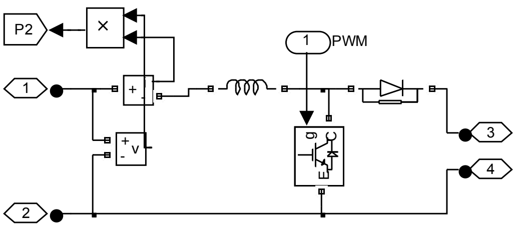

Fig.8 The simulation model of boost circuit

5.2 Main Parameters of DDPM

The simulation model of Permanent magnet synchronous generator is one of the MATLAB owns and the main parameters of DDPM in the paper is in table 1.

Table 1 The main parameters of the wind generator power feedback model

Mechanical (W)	39900	Rotor flux (Wb)	0.192
Generator power (VA)	44333	Friction coefficient ($N.m.s$)	0.001889
Pitch angle($^\circ$)	0	The optimal tip speed ratio	8.1
Stator resistance (Ω)	0.05	Fan radius R (m)	15
Inductance (H)	0.000635	Wind speed range (m/s)	6-11
Pole logarithmic (P)	36	The biggest wind power utilization coefficient	0.48

6 Application of the Fault Detection Method Based on Dynamic Peak-valley Limit in Wind Power Generation

According to the model of section 5, suppose the normal wind speed is at the range $6m/s-11m/s$. Get several groups of normal fan parameters (speed, voltage, power, three-phase rotor current, etc.) data to construct dynamic peak-valley limit through running the simulation model many times, and the cycle size is 951.

In sampling time 350-600, the wind speed of the simulation model take about $14m/s$ randomly. Wind speed of the fan in other times is random, but in the normal wind speed range. When the wind speed exceeds $11m/s$, the fan withstand mechanical stress is greater than the rated maximum stress, and the long time operation will damage the wind turbine, as a group of fault data. Next, we will use these data to do some experiments.

6.1 Fault Detection Results and Discussion

False alarm rate and missing alarm rate are important parameters to measure whether a fault detection model is reliable or not, and are also the mainstay to verify the feasibility of a method. In this paper, traditional PCA algorithm, dynamic peak-valley limit based on PCA algorithm, dynamic data window and dynamic peak-valley limit based on RPCA are adopted to monitor the working process of the fan respectively. Fig.9~Fig.12, respectively, are testing results graph of the four methods.

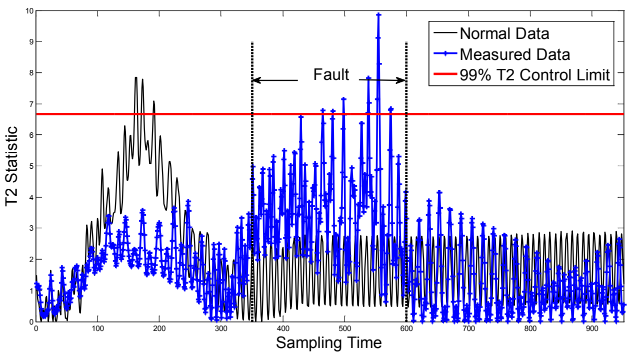

Fig.9 Fault detection result of wind power generation system based on PCA

Fig.9 is fault detection result of wind power generation system based on PCA. For the normal data, the T^2 control limit also has the phenomenon of false alarm rate, so the performance of the whole system is poor and this method is not suitable for fault detection under the non-steady conditions.

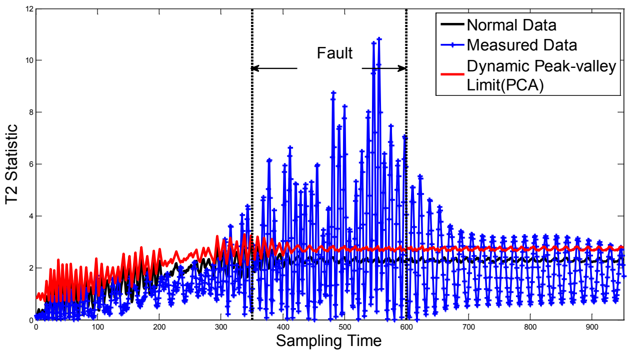

Fig.10 Dynamic peak-valley limit fault detection result of wind power generation system based on PCA

Fig.10 is the fault detection result of dynamic peak-valley limit based on PCA. From the figure we can see that there exists a great number of false alarm in the time 600-951.

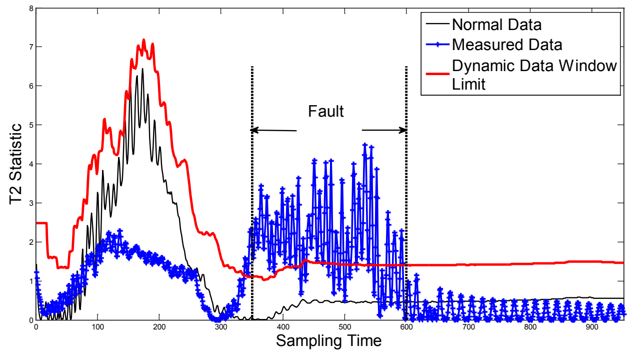

Fig.11 Dynamic data window fault detection result of wind power generation system based on RPCA

Fig.11 shows the dynamic data window fault detection result. According to observation, the phenomena of false alarm or missing alarm are rare and this method can detect fault effectively.

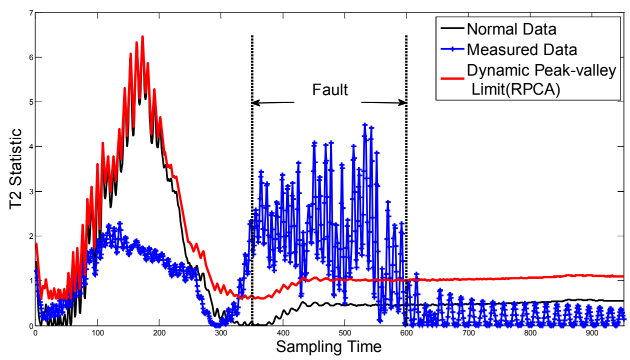

Fig.12 Dynamic peak-valley limit fault detection result of wind power generation system based on RPCA

Fig.12 shows the result of fault detection based on dynamic peak-valley limit (RPCA). This method maintains a lower false alarm rate and missing alarm rate compared with other methods from the above figures and can detect fault more effectively.

6.2 Result Comparison and Discussion

According to the analysis of the above examples, we can get the comparison results of the false alarm rate and the missing alarm rate of the four methods and the results are shown in Table 2.

Table 2 Four kind of fault detection methods performance comparison

Multivariate statistical fault detection method	False alarm rate (%)	Missing alarm rate (%)
PCA	0	26.13
Dynamic peak-valley limit based on PCA	14.51	10.25
Dynamic data window based on RPCA	2.21	6.53
Dynamic peak-valley limit based on RPCA ($\omega=0.1$, $(1-\alpha)=99\%$)	2.05	2.67

According to table 2 and the analysis of Fig.9, 10, 11, 12, we can see that: PCA algorithm is difficult to extract principal components effectively after dimension standardization, so it could not detect fault effectively and exists seriously missing alarm phenomenon and the missing alarm rate is up to 26.13%, detection sensitivity is low. Dynamic peak-valley limit based on PCA, compared with PCA, the missing alarm rate reduces greatly and is about 10.25%, but the false alarm rate becomes higher, so the reliability of fault detection model is relatively poor. In this paper, the proposed dynamic peak-valley limit based on RPCA could effectively extract principal components and get the needed information through according to the wind power generation system actual requirements. The method combines two control limits by a weight, could not only detect the fault effectively, but also reduce the false alarm greatly. Known from Fig.12 and table 2, the method makes the system keep lower missing alarm rate and false alarm rate, is 2.05% and 2.67% respectively, and could enhance the effectiveness of the system monitor greatly. At the same time dynamic data window method of Fig.11 could also maintain a low false alarm rate and missing alarm rate, is 2.21% and 6.53% respectively. But the dynamic peak-valley limit is superior to the dynamic data window no matter from false alarm rate or missing alarm rate.

Table 3 Performance comparison of different parameters of fault detection method based on dynamic peak-valley limit

the Fault Detection Method Based on Dynamic Peak-valley Limit		False alarm rate(%)	Missing alarm rate(%)
$\omega = 0.05$	confidence $(1-\alpha) = 99\%$	6.36	3.84
$\omega = 0.1$		2.05	2.67
$\omega = 0.2$		1.05	8.26
$\omega = 0.1$	$(1-\alpha) = 80\%$	11.25	1.15
	$(1-\alpha) = 95\%$	6.25	2.45
	$(1-\alpha) = 99\%$	2.05	2.67

Table 3 lists the value of false alarm rate and missing alarm rate when the system selects different confidence and weight. From the table we know that the fault detection method based on dynamic peak-valley limit could adjust the value of confidence and weight according to the users demand to achieve the expected accuracy and effectiveness of fault detection.

7 Conclusion

The detection limit of the traditional multivariate statistical method is a straight line, when applied to the non-steady conditions, there will be a large number of false alarm and missing alarm, so it could not effectively detect fault. For these problems, this paper introduces RPCA algorithm to get T^2 statistics firstly, then uses MLS method to preprocess T^2 statistics for obtaining the peak-valley curve, and connects the peak points and valley points in the curve to construct a new control limit combined with the traditional T^2 control limit based on RPCA by introducing a weight to get the dynamic peak-valley limit. Finally, a fault detection method based on dynamic peak-valley limit was put forward. This method could effectively reduce the false alarm rate and the missing alarm rate, and the monitoring sensitivity could be adjusted through changing the value of weight to maintain a balance between false alarm and missing alarm. Applied to the fault detection of the wind power generation system, the experimental results verified the effectiveness of the proposed method.

8 Acknowledgement

This paper was supported by the project of NSFC (NO.61203089). We were grateful to the anonymous reviewers of a draft of this paper for their insightful remarks and useful comments.

References:

[1] W.Y. Chen, Application of Multi-scale Principal Component Analysis and SVM to the Motor Fault Diagnosis, *International Forum on Information Technology and Application*, Vol.3, 2009, pp.131-134.

[2] X. Wang, U. Kruger, B. Lennox, Recursive Partial Least Squares Algorithms for Monitoring Complex Industrial Process, *Control Engineering Practice*, Vol.27, 2 003, pp.141-149.

[3] Z.M. Zhuang, W.Q. Li, F.L. Li, A Novel Method for the Fault Diagnosis of Asynchronous Motor Based on Multivariate Statistical Analysis, *Journal of Test and Measurement Technology*, Vol.27, No.2, 2007, pp.112-116.

[4] X.M. Tian, X.G. Deng, A Fault Detection Method Using Multi-Scale Kernel Principal Component Analysis, *Proceedings of the 27th Chinese Control Conference*, 2008, pp.25-29.

[5] S. Wold, Exponentially Weighted Moving Principal Component Analysis and Projection to Latent Structures, *Chemometrics and Intelligent Laboratory Systems*, Vol.23, 1994, pp. 78-83.

[6] M.J. Fuente, D. Garcia-Alvarez, G.I. Sainz-Palmero, Fault Detection and identification method based on multivariate statistical techniques, *Emerging Technologies & Factory Automation*, 2009, pp.1-6.

[7] W.H. Li, H.Henry Yue, Sergio Valle-Cervantes, S.J. Qin, Recursive PCA for Adaptive Process Monitoring, *Journal of Process monitoring, Journal of Press Control*, Vol.10, No.5, 2000, pp.471-486.

[8] S. Lane, E.B. Martin, A.J. Morrios, P. Gower, Application of Exponentially Weighted Principal Component Analysis for the Monitoring of a Polymer Film Manufacturing Process, *Transactions of the Institute of Measurement and Control*, Vol.25, No.1, 2003, pp. 17-35.

[9] T.Z. Wang, D.J. Gao, P. Liu , T.H. Tang, A Fault Detection Model Based on Dynamic Limit under Non-periodic Non-steady Conditions, *J. Shanghai Jiaotong Univ.(Sci.)*, Vol.46, No.4, 2012, pp. 607-612.

[10] N.A.A. Majid, B. Young, M. Taylor, J. Chen, Real-time for Monitoring Aluminium Reduction Cells using Multi-way PCA (MPCA) and Dynamic Euclidean Distances, *IEEE International*

Conference on Control and Automation, 2009, pp.454-458.

[11] H. Polinder, F.F.A. van der Pijl, G.-J de Vilder, P. Tavner, Comparison of Direct-Drive and Geared Generator Concepts for Wind Turbines, *IEEE International Conference on Electric Machines and Drives*, 2005, pp.543 - 550.

[12] C.L. Wen, J. Hu, T.Z. Wang, Z.G. Chen, Relative Principal Component Analysis with Application of Data Compression and Fault Diagnosis, ACTA AUTOMATICA SINICA, Vol.34, No.9, 2008, pp.1128-1139.

[13] T.Z. Wang, T.H. Tang, Relative Principal Component Analysis Algorithm and Its Application in Fault Detection, *Journal of System Simulation*, Vol.19, No.13, 2007, pp.2889-2894.

[14] Q.H. Zeng, D.T. Lu, Curve and Surface Fitting Based on Moving Least-Squares Methods, *Journal of Engineering Graphics*, Vol.25, No.1, 2004, pp.84-89.

[15] E.H.R. Coppoli, R.S. Silva, R.C. Mesquita, Treatment of Material Discontinuity in Meshless Methods for EM Problems Using Interpolating Moving Least Squares, *IET 7th International Conference on Computation in Electromagnetics*, 2008, pp.154 - 155.

[16]Lancaster P, Salksuskas K, Surfaces Generated by Moving Least Squares Methods, *Mathema-tics of Computation*, 1981, pp. 141-158.

[17] T. Belytschko, Y. Krongauz, D. Organ, M. Fleming, P. Krysl, Meshless Method: An Overview and Recent Developments, *Computer Methods in Applied Mechanics and Engineering*, 1996, pp. 3-47.

[18] Y.X. Zhang, Y. Li, B. Chen, A Nonlinear Approximation for Size Moisture Regain Using Moving Least Squares Method, *Control and Decision Conference(CCDC)*, 2012, pp.913-916.

[19] H. Ni, Research on Some Problems of Moving Least Square Method in the Data Fitting (Master thesis), *Zhejiang SCI-TECH University*, 2011.

[20] H. Ni, Z. Li, H.X. Song, Moving Least Square Curve and Surface Fitting with Interpolation Conditions, *Computer Application and System Modeling (ICCASM)*, Vol.13, 2010, pp:300-304.

[21] R. Esmaili, L. Xu, D.K. Nichols, A New Control Method of Permanent Magnet Generator for Maximum Power Tracking in Wind Turbine Application, *Proc.2005 IEEE Power Engineering Society General Meeting*, vol.3, 2005, pp.2090-2095.

[22] K. Tan, S. Islam, Optimum Control Strategies in Energy Conversion of PMSG Wind Turbine System without Mechanical Sensors, *IEEE Transactions On Energy Conversion*, Vol.19, No. 2, 2004, pp. 392–399.

[23] X.Y. Zhang, S. He, X.B. Zhang, Study of Fault Diagnosis Wind Turbine Generator System, *Journal of Xinjiang University*, Vol.2, 2009.

[24] T.H. Tang, X.H. Chen, D.K. Peng, J.G. Han, A Steady State Model of Direct-drive Permanent Magnet Synchronous Wind Generator and its Maximum Power Point Tracking Control, *Journal of Power Supply*, No.3, 2011, pp.13-20.

[25] Eftichios Koutroulis, Kostas Kalaitzakis, Design of a Maximum Power Tracking System for Wind-Energy-Conversion Applications, *IEEE Transactions on Industrial Electronics*, Vol.53, No.2, pp. 485-494.

[26] K.Y. Lo, Y.R. Chang, Y.M. Chen, Battery Charger with MPPT Function for Stand-Alone Wind Turbines, *Proc. of Power Electronics Conf., Sapporo, Japan,* 2010, pp. 932 - 937.