Reducing sensor complexity for monitoring wind turbine performance using principal component analysis

Yifei Wang, Xiandong Ma*, Malcolm J. Joyce
Engineering Department, Lancaster University, Lancaster, UK LA1 4YW
* Corresponding author. Tel.: +44 1524 593700; fax: +44 1524 381707
Email addresses: y.wang42@lancaster.ac.uk, xiandong.ma@lancaster.ac.uk, m.joyce@lancaster.ac.uk

ABSTRACT

Availability and reliability are among the priority concerns for deployment of distributed generation (DG) systems, particularly when operating in a harsh environment. Condition monitoring (CM) can meet the requirement but has been challenged by large amounts of data needing to be processed in real time due to the large number of sensors being deployed. This paper proposes an optimal sensor selection method based on principal component analysis (PCA) for condition monitoring of a DG system oriented to wind turbines. The research was motivated by the fact that salient patterns in multivariable datasets can be extracted by PCA in order to identify monitoring parameters that contribute the most to the system variation. The proposed method is able to correlate the particular principal component to the corresponding monitoring variable, and hence facilitate the right sensor selection for the first time for the condition monitoring of wind turbines. The algorithms are examined with simulation data from PSCAD/EMTDC and SCADA data from an operational wind farm in the time, frequency, and instantaneous frequency domains. The results have shown that the proposed technique can reduce the number of monitoring variables whilst still maintaining sufficient information to detect the faults and hence assess the system's conditions.

Keywords: principal component analysis (PCA), feature extraction, condition monitoring, wind turbine, distributed generation

Nomenclature

Acronyms

DG	Distributed generation
CM	Condition monitoring
CoE	Cost of electricity
O&M	Operation and maintenance
PCC	Point of common coupling
HHT	Hilbert-Huang transform
EMD	Empirical model decomposition
IMF	Intrinsic mode function
PCA	Principal component analysis
cppv	Cumulative percentage partial covariance
PMSG	Permanent magnet synchronous generator
SCADA	Supervisory control and data acquisition

Roman symbols

a_i	Instantaneous amplitude at level *i*
c_i	i^{th} intrinsic mode function
$\boldsymbol{L}$	Characteristic root matrix
m_{ik}	i^{th} envelope of a signal at *k* iteration
r	Pearson's correlation coefficient
r_i	i^{th} residual signal in EMD
r_z	Fisher's correlation coefficient
R	Resistance, Ω
$\boldsymbol{S}$	Covariance matrix
$\boldsymbol{S}_{rr}$	Covariance matrix of retained dataset
$\boldsymbol{S}_{dd}$	Covariance matrix of discarded dataset
$\boldsymbol{S}_{rr.d}$	Partial covariance matrix of retained dataset
$\boldsymbol{U}$	Characteristic vector matrix
V	Grid voltage, V
V_{dc}	DC-link voltage, V
V_w	Wind speed, m/s
$x(t)$	Real part signal in Hilbert transform
$\boldsymbol{X}$	Input dataset matrix
$y(t)$	Imaginary part signal in Hilbert transform
$\boldsymbol{Z}$	Principal component matrix

C	Capacitance, F	*Greek symbols*	
C_p	Wind turbine power coefficient	β	Pitch angle, °
$E(X)$	Information entropy of variable X, bit	η_e	Percentage entropy, %
h	Sum of the squared correlations	λ	Tip speed ratio
h_{ik}	i^{th} temporary IMF at k iteration	ω	Angular frequency, rads/s
$H(X)$	Normalised information entropy of variable X	ω_i	Instantaneous frequency, rads/s
I	Grid current, A	ϕ	Phase angle, °
I_{dc}	DC-link current, A	θ_i	Instantaneous phase angle at level i, °
L	Inductance, H		

1. Introduction

Distributed generation (DG) systems comprising of renewable energy generation technologies will play a significantly increasing role in future power systems [1, 2]. A distributed generation system normally consists of hybrid renewable energy generation units embedded in the system. An example of wind-turbine-based DG system is shown in Fig. 1, where turbines are interfaced with the grid at a point of common coupling (PCC). Two of the major challenges for deployment of a DG system are its reliability and maintainability, which can be overcome by condition monitoring. The condition monitoring process can be divided into several components including data acquisition, signal processing and diagnosis and prognosis [3]. To achieve effective condition monitoring, accurate and reliable measurements are crucial. Fig. 2 shows the architecture of a distributed condition monitoring system that was originally developed for conventional power plants but has been used for wind farm condition monitoring for some time. In this system, a large amount of condition monitoring data and SCADA (supervisory control and data acquisition) data need to be transferred to a local CM server for processing and storing or, alternatively, to a remote support centre for further fault analysis.

A condition monitoring system can incorporate present and past data monitored by the sensors to diagnose and predict potential failures. By doing so, the performance, availability and reliability of wind turbines can be improved. Studies have shown that operation and maintenance (O&M) cost plays a significant role in calculating the cost of energy (CoE); a higher-quality O&M regime can achieve higher availability, lower through-life costs and hence a lower CoE [4]. Moreover, the deployment of condition-based maintenance has been proven to be far superior to the conventional preventive and periodic maintenance strategies [5, 6]. However, handling, processing and transmitting a huge amount of data will lead to more complex CM systems being built up and hence result in a negative impact on the performance, maintainability and cost of the CM systems [7]. For a typical wind turbine, there can be more than 250 sensors required to monitor most subsystems; it is envisaged the number of sensors will be significantly increased for a wind farm [8, 9]. Therefore, if the number of sensors or measurements installed can be reduced whilst still maintaining a necessary number to assess the system's condition, the data acquisition system can be simplified and the performance, maintainability and cost benefit of CM systems to be developed can be enhanced.

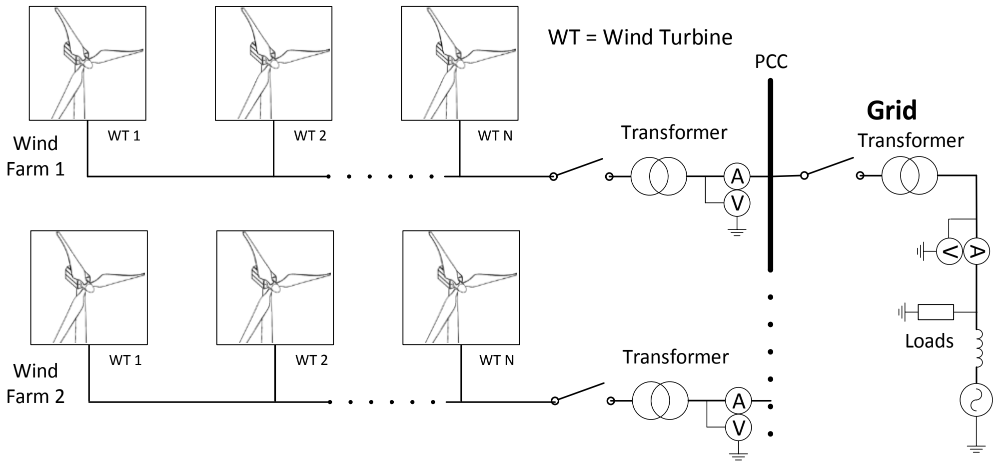

Fig. 1. An example of distributed generation (DG) network, taking the wind turbines as DG units.

Currently, data acquisition for condition monitoring systems is implemented mainly based on information maximisation principle, which means sensors are installed to obtain as much data as possible. Due to relationships existing among sensors, there is redundancy within the data collected. Thus, an appropriate sensor selection technique is desirable in order to identify and remove these unnecessary redundancies due to there being too many sensors carrying out similar functions. In the meantime, the method should be able to retain the provision of vital information, which is critical for fault diagnosis, prognosis and maintenance scheduling.

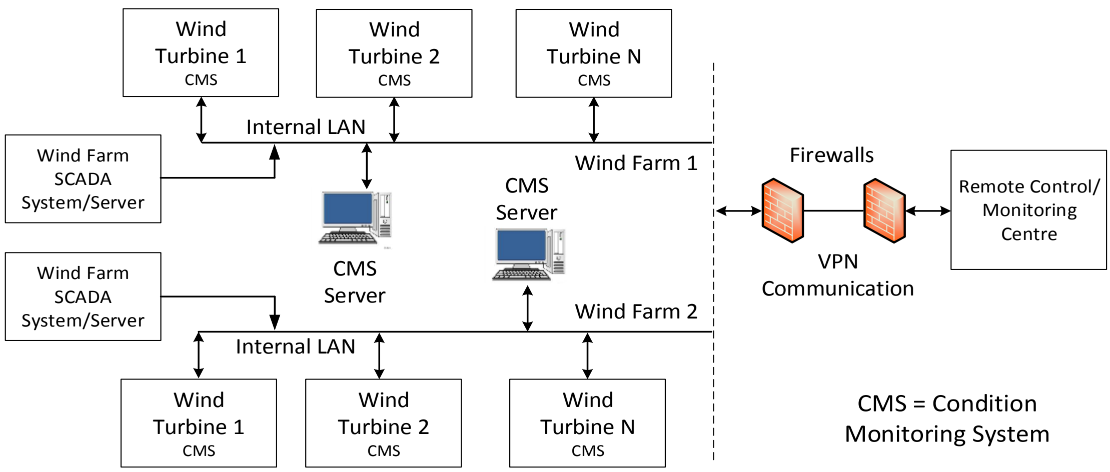

Fig. 2. A distributed condition monitoring system for wind farms.

There are a number of researches that have been carried out regarding sensor selection in complex sensor network systems. Information-based techniques are commonly adopted such as mutual information, information entropy, and fisher information. An entropy based sensor-selection approach has been proposed in [10] for an aerospace propulsion health monitoring system based on quantification of particular fault conditions and diagnostics. Sensor selection schemes were also proposed for tasks like target tracking and mission assignments in order to minimise the number of active sensors in a sensor network and hence reduce the energy use and prolong the lifetime of the sensor network [11]. A stochastic dynamic programming method was proposed to solve the sensor selection problem of robotic systems in real time [12]. Furthermore, filtering and estimation methods using Cramer-Rao bound criteria are also widely used in sensor selection for non-linear tracking problems [13]. It has been proven that there are fewer outputs from the filter or estimator than the input measurements, and the estimated parameters have better accuracy than from the direct measurement. However, all measurements are still required for prediction and update of the improved estimated outputs.

PCA has been used widely in dimension reduction and feature extraction applications because the transformed signals are orthogonal and found with a cost function of maximising variance. As with the PCA, other techniques like Linear Discriminant Analysis (LDA) and Locally Linear Embedding (LLE) are also commonly used for dimension reduction. The PCA performs dimension reduction while preserving as much of the data variance in the high-dimension space as possible, whereas the LDA performs dimension reduction while preserving as much of the class discriminatory information in the high-dimension space as possible. The LLE attempts to discover nonlinear structure in high dimensional data by exploiting its local properties. The objective of this nonlinear method is to maintain and reconstruct the local properties of the data manifold by writing the high-dimensional data points as a linear combination of their nearest neighbours [14].

However, PCA has the advantage of parametric mapping capability from the extracted features to the monitoring variables through estimation of the eigenvectors and principal eigenvalues. Each principal component corresponds to a particular feature of the data, and because these components are uncorrelated, there is no redundancy present. In this paper, PCA analysis incorporating the optimal variable selection based on data variability is investigated in order to optimise set of sensors for wind turbine condition monitoring systems. The optimal variable selection in this context is taken to mean that the variables are selected through maximising variability and minimising degrees of correlation among the retained variables. Moreover, one major contribution of the proposed method is that the actual number of physical sensors can be potentially reduced through estimation of the least significant variables. For wind turbines, the method can be used to reduce the complexity in developing a condition monitoring system. Furthermore, de-noising of data is not required prior to the analysis as the proposed method essentially assesses and selects the variables based on their variation. In our study, the analysis of measurement data focuses on transient characteristics not only in the time domain but also in terms of the frequency and instantaneous frequency domains (note: instantaneous frequency domain means frequency components as a function of time, referred to as the instantaneous frequency data later in this paper). The paper is organised as follows. The proposed sensor selection techniques are described in Section II. CM data used to test the proposed method are presented in Section III. The results are shown and discussed in Section IV, followed by the conclusions and a description of future work.

2. Methodology

The block diagram of the selection process is shown in Fig. 3, which comprises i) transformation of data into frequency and instantaneous frequency domain, ii) application of PCA to obtain the ranked principal components, and iii) use of different selection methods to retain the desirable variables.

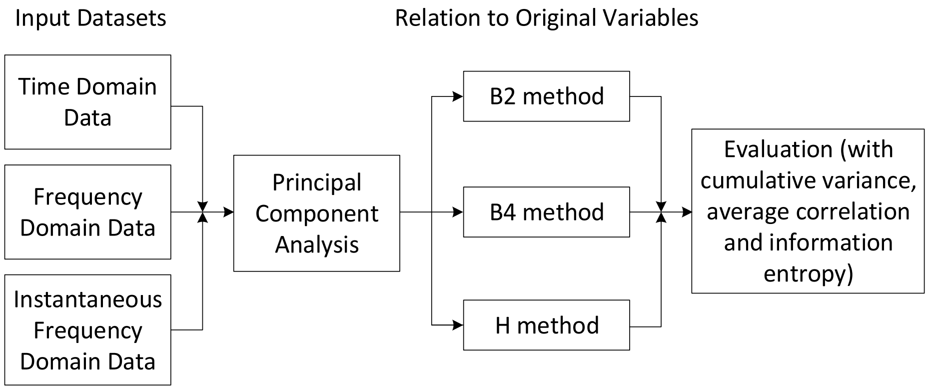

Fig. 3. Block diagram of the selection process using PCA.

Conventionally, time series data with large magnitude variations are retained and those with small magnitude variations are removed. This may not be ideal for sensor selection, simply because the selection process only captures features of the data in the time domain, whereas frequency components in measurement data are ignored. The wind sources are generally intermittent and stochastic, and hence are abundant in frequency components; so are the grid fluctuations. Frequency domain signals may contain more salient information than time domain, especially under fault conditions [15]. Therefore, it is worth to examine PCA with signals in the form of time-series, frequency-series and instantaneous-frequency data.

2.1 Instantaneous frequency transformation

We use the fast Fourier transform (FFT) and the Hilbert-Huang transform (HHT) to transform the time series data into frequency and instantaneous frequency domain data. The HHT is a combination of empirical mode decomposition (EMD) proposed by Huang [16] and the Hilbert spectral analysis. Zhang has applied the HHT in earthquake motion recordings [17], where it was proved that HHT outperforms the conventional methods such as FFT to analyse non-stationary dynamic earthquake motion recordings. Besides, EMD can decompose the signal into a series of intrinsic mode functions (IMF), which may contain critical physical information. Furthermore, the signal reconstructed from certain levels of IMF can be useful for capturing important frequency features contained in the original signal. It has also been shown that under certain conditions, the HHT is superior to the short time Fourier transform (STFT) and wavelet analysis to analyse vibration signals for machine health monitoring and to diagnose localised defects in roller bearings [18].

EMD decomposes the original signals $x(t)$ into a set of IMFs, each of which represents the intrinsic oscillatory modes of the signal. The IMF is found by first identifying the local extrema and then by fitting cubic spline line through all the maxima and minima to obtain the upper envelope $x_{up}(t)$ and lower envelope $x_{low}(t)$. Their mean is defined as $m_{ik}(t)$ and the difference between the original signal and the envelope mean is $h_{ik}(t)$.

$$m_{ik}(t) = [m_{up}(t) + m_{low}(t)]/2 \quad (1)$$

$$h_{ik}(t) = h_{i(k-1)}(t) - m_{ik}(t) \quad (2)$$

The process repeats k times until the $h_{ik}(t)$ satisfies the criteria defined for the IMF, where $h_{i(k-1)}$ is the original signal when k = 1.

Once a IMF is found, it is then subtracted from the original signal and a residual signal $r_i(t)$ is obtained. The process repeats i iterations until the final residual is a constant or a monotonic function.

$$c_i(t) = h_{ik}(t) \quad (3)$$

$$r_i(t) = x(t) - c_i(t) \quad (4)$$

The original signal can be reconstructed by summing all the n IMFs and the residual using the formula below.

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \quad (5)$$

The Hilbert transform calculates the instantaneous frequency of the IMFs obtained through EMD. The original signal can be expressed as the real part $\Re$ of the form $z(t)=x(t)+jy(t)$:

$$x(t) = \Re\left(\sum_{i=1}^{n} a_i(t) e^{j\int \omega_i(t)dt}\right) \quad (6)$$

where $y(t)$ is the complex conjugate of $x(t)$; n is the total number of IMFs; a_i is the amplitude of the signal of IMF at level i; $\omega_i(t)$ is the frequency of the signal at level i and j^2=-1.

At level i, the corresponding amplitude $a_i(t)$ and phase $\vartheta_i(t)$ can be found by,

$$a_i(t) = \sqrt{c_i(t)^2 + y_i(t)^2} \quad (7)$$

where $c_i(t)$ is the IMF at level i, and

$$\theta_i(t) = \tan^{-1}\left(\frac{y_i(t)}{c_i(t)}\right) \quad (8)$$

Finally, the instantaneous frequency $\omega_i(t)$ at level i can be found by,

$$\omega_i(t) = \frac{d(\theta_i(t))}{dt} \quad (9)$$

Consequently, the HHT transforms the original time series signal $x(t)$ into a new set of instantaneous frequency signals $f(t)$, i.e., the frequency changing with respect to time of $x(t)$. In our study, the signal reconstructed from all IMFs except the residual is used to produce the instantaneous frequency data for sensor selection in order to avoid feature losses due to data transformation.

2.2 Overview of PCA

Essentially, PCA is a variant of multivariate analysis relying on the data-analytic technique and tries to reveal the multivariate structure of the data. PCA transforms a set of data into a set of uncorrelated principal components (PCs). The uncorrelated PCs are calculated by maximising variance and then ranking them in terms of their magnitude [19]. PCA is initially used as a dimension reduction technique in different fields [20]. Researchers have shown that by retaining first few components, the dimension of the data can be reduced dramatically, while little information is sacrificed. These properties of PCA make it ideal to be used as a feature selection technique incorporated into the artificial neural network (ANN) to predict turbine performance and detect faults [21].

PCA is also known as the Karhumen-Loeve transforms. It is an orthogonal transformation that converts the original dataset **X** ($p \times n$ dimensions) with p variables and n samples into a set of principal components **Z** ($q \times n$ dimensions) with q PCs and n samples. The transformation of the dataset is completed by Single Value Decomposition (SVD) of the covariance matrix $\boldsymbol{S}$ ($\boldsymbol{S} = \boldsymbol{XX}^T$) of the dataset $\boldsymbol{X}$ by optimising the variance. This means that the first principal component has the highest variance. Therefore, each PC is uncorrelated and ranked with a descending order.

Finding the principal components involves eigenanalysis of the covariance matrix $\boldsymbol{S}$. The eigenvalues of $\boldsymbol{S}$ are solutions $\boldsymbol{L}$ ($l_1, l_2, \ldots, l_p$) to the characteristic equation $|\mathrm{S} - L\boldsymbol{I}| = 0$, where $\boldsymbol{I}$ is the identity matrix. The eigenvalues $l_1, l_2, \ldots, l_p$ are the variances of each principal component and the sum of all p eigenvalues equals the sum of the variances of the original variables. Hence, PCs are obtained by satisfying the relationship in (10) using SVD of the covariance matrix $\boldsymbol{S}$,

$$\boldsymbol{U'SU} = \boldsymbol{L} \quad (10)$$

As the diagonal matrix $\boldsymbol{L}$ ($l_1, l_2, \ldots, l_p$) is already known, the corresponding characteristic vectors or eigenvectors $\boldsymbol{U}$ ($u_1, u_2, \ldots, u_i, \ldots, u_p$, u_i are the columns of $\boldsymbol{U}$) are therefore calculated. $\boldsymbol{U}$ ($u_1, u_2, \ldots, u_p$) are also called as loadings representing correlations between variables and principal components.

The relationship between the PCs, $\boldsymbol{Z} = (z_1, z_2, \ldots, z_q)$, and the original dataset $\boldsymbol{X}$ is mathematically expressed below,

$$\begin{aligned} z_1 &= u_{11}x_1 + u_{12}x_2 + \cdots + u_{1p}x_p \\ z_2 &= u_{21}x_1 + u_{22}x_2 + \cdots + u_{2p}x_p \\ &\vdots \\ z_q &= u_{q1}x_1 + u_{q2}x_2 + \cdots + u_{qp}x_p \end{aligned} \quad (11)$$

Equation (11) represents the maximum possible proportion of variance in the original variables can be displayed in the first q principal components.

2.3 Selection methods

To link the ranked PCs back to the original variables, three different selection methods, *i.e.*, B2 method, B4 method and H method, are used.

The B2 selection process starts by selecting principal components which have a variance that is less than l_0 ($l_i<l_0$, $1\leq i\leq q$). The number of variables retained is highly dependent on the predefined threshold l_0. It was suggested by Jolliffe [22-23] that l_0=0.7 is a reasonable choice. For the k selected PCs, each component i ($1\leq i\leq k$) is related to the original variables as described in eq. (11). The original variable x_i, which has the largest absolute coefficient u_{ij} in the row vector $\boldsymbol{u}_i$ is eliminated. The process ends when all the selected PCs are examined. The rest of the original variables are then retained.

In contrast to B2, the B4 method starts with PCs, whose variance is larger than the predefined value. Original variable x_i with largest absolute eigenvector value u_{ij} is retained. It is also suggested by Jolliffe that, for the B4 method, the value of l_0 is reasonable if selected in a range of $0.66 \leq l_0 \leq 0.74$.

As a new method, the H method is performed based on one of the selection criteria for principal variables proposed in [24, 25]. The selection relies on the optimisation of minimising the squared norm of the original variables. The H method examines the H values, h_1, h_2, ..., h_p, which are known as the sum of the squared correlations between variable x_i as described in eq. (12). H values are ranked in a decreasing order after H values are calculated for all original variables. Variables that have the highest H value h_i are retained. The process stops when the sum of the H value of the k retained variables exceeds the predetermined threshold. The h_i is obtained by

$$h_i = \sum_{j=1}^{p}\left(l_j u_{ij}\right)^2 \tag{12}$$

where l and u are the eigenvalue and eigenvector, respectively, as described earlier.

2.4 Validation measures

Cumulative variance, average correlation and information entropy are used to validate the results from proposed selection algorithms. Each of these measures has its own purpose in examining the performance of retained variables.

Cumulative variance is a measure of percentage variability of the retained variables with regards to the whole dataset, where the multivariate structure of the dataset is considered [24]. For a dataset $\boldsymbol{X}$ ($p \times n$) with q variables ($q < p$) being retained and m ($m=p-q$) variables being discarded, the covariance matrix of the dataset $\boldsymbol{X}$ can be divided into $\boldsymbol{S_{rr}}$ ($q \times q$), $\boldsymbol{S_{rd}}$ ($q \times m$), $\boldsymbol{S_{dr}}$ ($m \times q$), $\boldsymbol{S_{dd}}$ ($m \times m$) as shown in (13). The subscripts r and d represent the retained set with q number of variables and the discarded set with p-q number of variables, respectively.

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S_{rr}} & \boldsymbol{S_{rd}} \\ \boldsymbol{S_{dr}} & \boldsymbol{S_{dd}} \end{bmatrix} \tag{13}$$

The partial covariance matrix $\mathbf{S}_{rr.d}$ for retained variables is:

$$\boldsymbol{S_{rr.d}} = \boldsymbol{S_{rr}} - \boldsymbol{S_{rd}}\boldsymbol{S_{dd}^{-1}}\boldsymbol{S_{dr}} \tag{14}$$

The cumulated percentage variance can then be obtained by the equation below, where tr is the trace of the partial covariance matrix, *i.e.*, sum of the elements on the main diagonal.

$$cppv = tr(\boldsymbol{S_{rr.d}})/tr(\boldsymbol{S}) \tag{15}$$

However, *cppv* does not explain the repetition of features among variables, for example, between power, current and voltage; this measure only calculates fluctuation of the magnitude of the signal. Thus, an average correlation coefficient is introduced to measure the degree of associations between variables in the dataset. Due to the fact that Pearson's correlation coefficients are not additive, the average correlation coefficient cannot be calculated using a simple arithmetic mean method. To be able to calculate the average correlation coefficient, Pearson's correlation coefficient is first transformed using Fisher's transformation, and then the arithmetic average of the transformed value is converted back. The Fisher's transform and its corresponding inverse transform are given below:

$$r_z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \tag{16}$$

$$r = \frac{e^{2r_z-1}}{e^{2r_z+1}} \tag{17}$$

where r_z is the transformed correlation coefficient and r is the Pearson correlation coefficient. This measure considers the multi-collinearity behaviour of the dataset, and a higher value indicates high degrees of correlation among dataset, and a low value indicates less dependency between variables.

As a measure of information discrepancy, entropy has been used extensively in communication, data compression and data encoding [26], and also in feature selection and classification for ANN and fault detection [27]. After application of PCA, the information entropy of original dataset and those retained variables from each selection method are calculated individually.

With a given variable X and the probability mass function of the variable $p(x)=Pr\{X=x\}$, $x \in \Re$, the information contained or the uncertainty in the variable X can be quantified by the information entropy $E(X)$,

$$E(X) = -\sum_{i=1}^{n} P(x_i)\log_b P(x_i) \tag{18}$$

where $P(x_i)$ is the probability $p(X=x_i)$ and b is the base for each different entropy unit. In this paper, the Shannon's entropy is used, where $b=2$; hence the unit of the entropy is the bit. Moreover, normalised entropy is also introduced in order to compare different variables, as the normalised entropy is bounded between 0 and 1, which is obtained by

$$H(X) = -\sum_{i=1}^{n} \frac{p(x_i)\log_b(p(x_i))}{\log_b(n)} \tag{19}$$

where n is the length of the signal and $\log_b(n)$ is the maximum entropy of the signal.

3. Condition monitoring data

3.1 Simulation data

The purpose of the simulations presented in this work is to investigate and therefore obtain useful data under various operation conditions. A DG network with wind turbines as the DG units is given in Fig. 1. A 2.1 MW wind turbine connected to the grid is modelled and shown in Fig. 4. The model has been simulated using PSCAD/EMTDC, a general-purpose time domain simulation program with a graphical interface for studying transient behaviour of complex electrical networks. The software allows a flexible time step ranging from nanoseconds to seconds to simulate electromagnetic transients in the electrical network; the time step chosen for the simulation is 100µs.

The input to the turbine is wind speed V_w. The turbine model is responsible for simulation of the mechanical energy generation including mechanical torque and power to drive the connected PMSG (permanent magnet synchronous generator), which converts mechanical power into electrical power. The aerodynamic torque and power are especially related to the effective wind speed and the pitch angle of the rotor blades β adjusted in a nonlinear relationship. The aerodynamic power coefficient, C_p, used for calculation of the aerodynamic power and torque is given by:

$$C_p = 0.5(\gamma - 0.022\beta^2 - 5.6)e^{-0.17\gamma} \tag{20}$$

where $\gamma = 2.237 V_w/\omega_t$, ω_t is the wind turbine shaft rotation speed. Essentially, γ is a function of the tip-speed ratio λ, i.e., $\gamma = 2.237 R_t/\lambda$, where $\lambda = \omega_t R_t / V_w$ and the turbine radius R_t = 46.2m in this study. The coefficients of C_p in equation (20) are obtained through nonlinear function fitting from experimental data in order to describe properly the aerodynamic behaviour of the blades under different operational conditions [28]. At low wind speed, the pitch angle β is forced to zero to maximise the power coefficient C_p. As the wind speed increases above the rated value (14 m/s in our simulation), dynamic pitch control is adopted to regulate the output power to its rated value. More information about the dynamic pitch control can be referenced in [29].

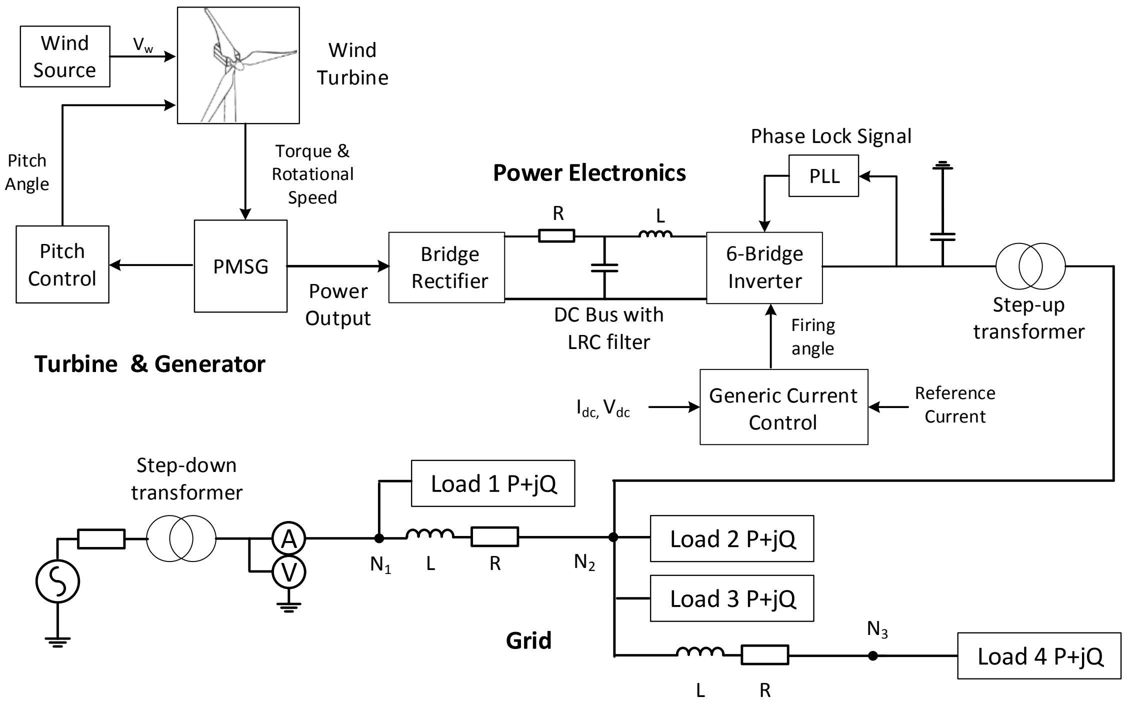

Fig 4. PMSG wind turbine model with grid connection simulated in PSCAD/EMTDC.

The turbine is coupled to a PMSG and the grid connection is made through an AC-DC-AC converter and a step-up transformer. The AC-DC-AC converter is necessary in order to connect the variable voltage and frequency output from the generator to the fixed grid voltage and fixed 50 Hz grid frequency. The converter is composed of a diode rectifier, a DC bus with a storage capacitance voltage and a six-pulse bridge thyristor inverter. The AC output from PMSG is rectified into DC voltage and a RLC circuit is then used to filter out noise and stabilize the electrical voltage input for a 6-bridge inverter. The inverter has two main purposes: control the active power flow from DC-link to grid and voltage stabilisation of the DC-link. A generic current controller is incorporated to maintain the voltage dependent current in the DC bus, and produce the firing pulses for the inverter based on the DC bus current I_{dc} and voltage V_{dc}. The phase angle of the converted AC voltage is synchronised through the phase locked loop (PLL). The transformer is required to step-up the voltage from 1.7kV to 12.5kV. A filtering capacitor is added to smooth output voltages and compensate for output reactive power.

The network is simulated by a three-phase 34.5 kV/300 MVA network with an ideal voltage source and equivalent system impedance. A transformer is used to step-down the voltage to 12.5kV. Grid faults between phases or between one or more phases and ground can be incorporated. A simplified radial distribution system is considered in this paper, where the loads are modelled with 2.133 MW and 1.6 MVar. Loads and loss in the transmission line are represented by resistive and inductive load R and L. The voltage drop ΔV due to losses from the loads in the grid is described by,

$$\Delta V = RI\cos\varphi + LI\omega\sin\varphi \quad (21)$$

where I is the current, ω is the angular velocity of power frequency and φ is a leading or lagging phase angle.

Wind speeds can be simulated as constant speed, constant speed superimposed with ramps and gusts representing wind speed fluctuations, or on-site wind speed measurements. In our study, real wind speeds collected on Hazelrigg site at Lancaster University are used, where a 64-metre wind turbine of 2.1 MW is erected and operational. As an example, Fig. 5 shows simulation results of the turbine mechanical torque and the active power under the actual wind speeds. The turbine torque is strongly related to the wind speed when a fixed power coefficient C_p is used, as shown in the Fig. 5. It is necessary to keep the rotor speed at an optimum value of the tip-speed ratio λ_{opt} when the wind speed varies. For the wind speed below the rated value, the generator produces maximum power at any wind speed within the allowable range following the adjusted λ_{opt}. For the wind speed above the rated value, the wind turbine energy capture is limited by applying the pitch control, as described above. Consequently the active power remains relatively constant under the given wind speeds (most wind speeds exceed the rated value of 14 m/s in our simulation).

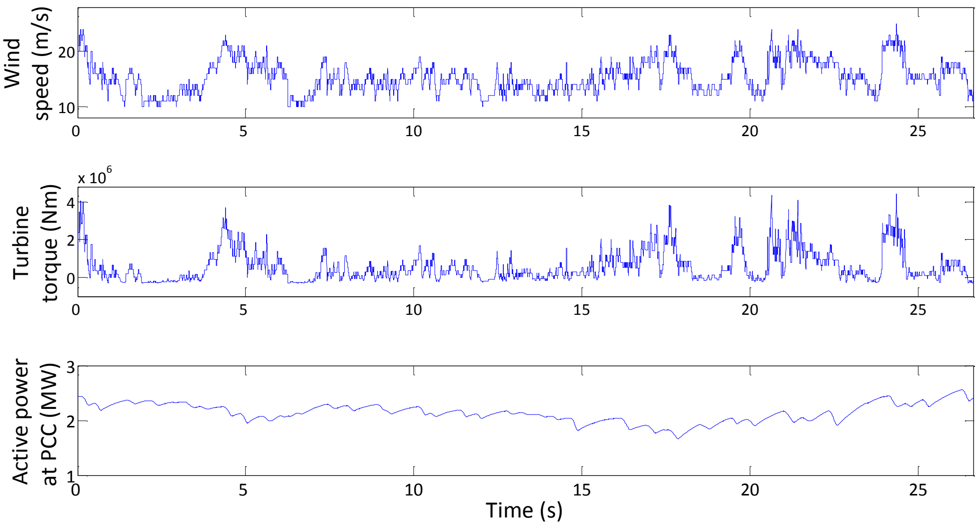

Fig. 5. Examples of simulation results showing wind speed, turbine torque and active power with actual wind speed as inputs.

3.2 SCADA data

SCADA data used in this paper are obtained from an operational wind farm, with time duration of 15 months. It is essential to use actual operational data of wind turbines to validate the proposed algorithms. SCADA data are usually sampled at 10-minute intervals in order to significantly reduce the amount of data that need to be processed while still reflecting normal and faulty status of wind turbine operations. The SCADA data for each turbine consist of approximately 128 readings for various temperatures, pressures, vibrations, power outputs, wind speed and digital control signals. Pre-processing of the data is carried out to eliminate those digital and constant signals, which are ineffective to the PCA analysis. Gaps in SCADA data exist due to occasions when a wind turbine is inactive during periods of low and high wind speeds, and due to the occurrence of maintenance periods. It is necessary to remove these gaps when no power is generated prior to PCA analysis. In order to obtain generic models applicable to the entire wind farm, SCADA data from a wind turbine selected at random have been used for validation of the general variable selection technique.

As an example, Fig. 6 shows wind speed, generator winding temperature and active power from one of the turbines in the wind farm for a time period of one month. Thermal aging is one of the most common stator insulation deterioration processes that might be caused by localised defects during operation; thus generator winding temperature monitoring has been widely used on multi-megawatt (multi-MW) wind turbines. The generator winding temperature depends not only upon the wind speed, but also the power output of the turbines [9]. In this case, at low wind speeds, the generator winding temperature fluctuates between 50-60 °C. When the wind speed increases and the turbine is operating at the rated value, the winding temperature can reach a maximum of approximately 80 °C.

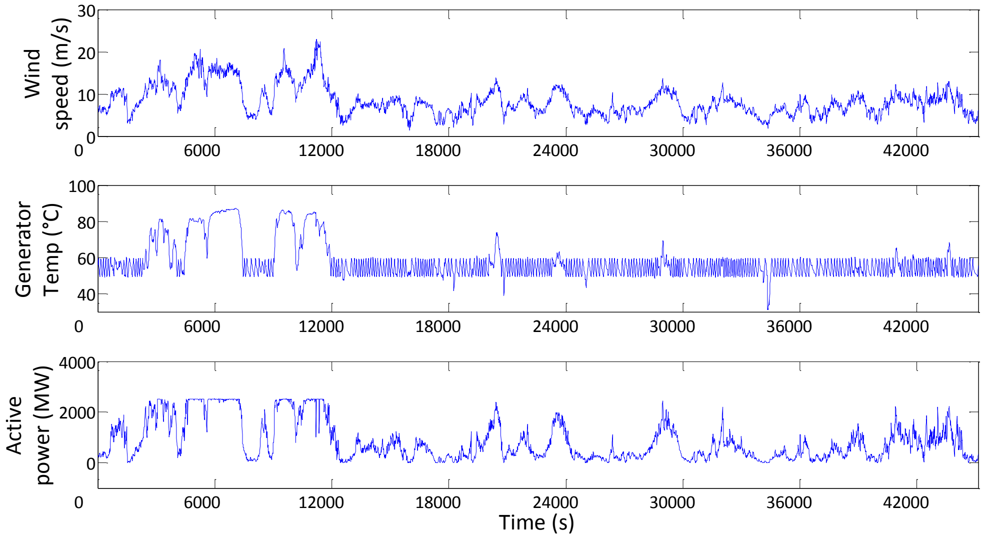

Fig. 6. Example of SCADA data showing wind speed, generator winding temperature and active power.

4. Results and discussions

4.1 General variable selection

The proposed selection algorithms are validated against both simulation and SCADA data. After pre-processing the data, there are a total of 29 and 77 variables for simulation and SCADA data, respectively. One crucial step of PCA concerns the choice of the number of principal components to be retained. In this paper, it is determined by a threshold value based on the cumulative variance, i.e., the ratio between the sum of the eigenvalues of the kept q principal components and the sum of all p eigenvalues of the original variables. In order to accommodate the dominant percentage of variances, the threshold value is set to be 99.7% in the paper. This means that 15 and 35 variables are sufficient to be used for the simulation and SCADA data, respectively. Measures described in previous sections are used to verify these retained variables and have shown that the retained variables have minimal information loss. Table 1 shows the results of three measures using each selection method in time, frequency and instantaneous frequency domains.

By looking at each measure individually, the cumulative variances from the time and frequency domains have similar values and are considerably higher than those from instantaneous frequency data. This might be because the instantaneous frequency data are reconstructed from IMFs in HHT. Moreover, among the three selection methods in the time and frequency domain, although H method has the lowest performance, they are still all above 83%; and the B2 method has the highest variance for both data cases. For the average correlation coefficients, the original datasets have a value of 0.34 and 0.11 for simulation and SCADA data. Out of these results, the B2 method in time domain for simulation data and the B2 method in frequency domain for SCADA data have the lowest average correlation coefficients. The results indicate that there is a lower interdependency among retained variables using the B2 method, which is desirable. In addition, it should be noted that the average correlation coefficient of the selected variables can be higher than the original dataset, implying that the presence of a higher degree of redundancy might be possible within the retained dataset, such as the result from using H method, which is undesirable.

Figures 7 and 8 show the combination of cumulative percentage variance and average correlation coefficient, taking the simulation and SCADA data in the time domain as the examples. The blue crosses are the variables of the original dataset and the red circles are the variables retained with the respective selection algorithm. The scatter plots show a relationship between the *cppv* and average correlation coefficient, where the variable of higher cumulative variance will give a higher correlation coefficient. Moreover, it can be seen that variables retained with the B2 and B4 methods are those variables selected across the entire area. On the contrary, the H method always retains variables with the highest cumulated variances, which also correspond to the highest average correlations. This indicates there still exists a considerable amount of information redundancy in the variables retained using the H method. This finding infers that critical information will also exist in variables with low variances, which is consistent with the result found by Hawkins in his research [30].

Table 1

Results from selection methods B2, B4 and H in the time, frequency and instantaneous frequency domains using simulation and SCADA data.

		Time			Frequency			Instantaneous frequency		
	Original dataset	**B2**	**B4**	**H**	**B2**	**B4**	**H**	**B2**	**B4**	**H**
SCADA data										
Cumulative variance	100%	99.00%	97.90%	83.62%	99.27%	98.79%	83.39%	72.24%	70.63%	51.19%
Average correlation	0.3418	0.162	0.1344	0.7945	0.1178	0.1313	0.7824	0.2634	0.2187	0.7175
Total entropy	59.47	48.36	47.25	22.45	45.55	48.06	22.93	41.68	43.38	16.18
Simulation data										
Cumulative variance	100%	99.81%	99.47%	91.43%	99.81%	99.55%	96.77%	79.81%	80.19%	55.79%
Average correlation	0.1107	0.0082	0.093	0.3702	0.1058	0.1329	0.0998	0.0113	0.0091	0.422
Total entropy	21.24	18.88	18.99	8.47	18.61	18.3	10	17.66	17.76	9.11

It seems that the selected variables are relatively random with the B2 and B4 method, as shown in Figures 7 and 8; however, those variables are discarded because they have high correlations with the retained ones. Moreover, the variance and correlation coefficients of the signals are dependent on the sample size, which may lead to a biased result. Therefore, information entropy is used to further validate the results. Suppose E_t is sum of information entropy of all variables, E_r is sum of entropy of the retained variables (Table 1) and the percentage of entropy η_e is the ratio of E_r/E_t. Thus, η_e can be used as a measure for comparison among selection methods. Figures 9 and 10 shows the percentage entropy of the selection methods (B2, B4 and H) in the time, frequency and instantaneous frequency domains, as represented by t, f and ft respectively, using simulation and SCADA data. It can be seen that for both cases, the H method has the lowest performance and the t_b4 and t_b2 have the highest percentage of entropy. This again agrees with the results obtained based on the cumulative variance and average correlation measures.

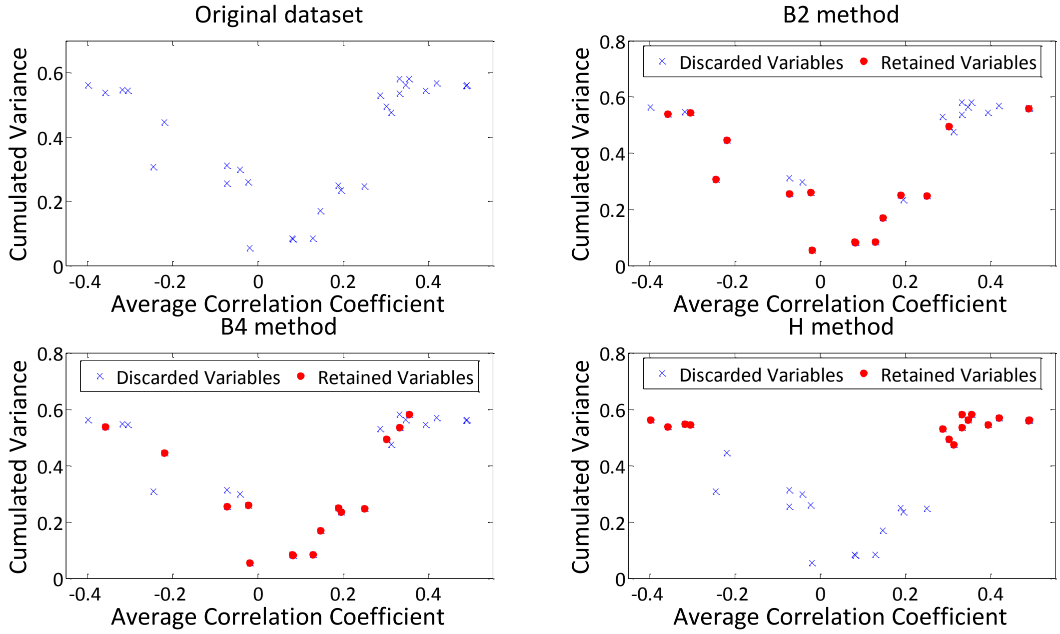

Fig. 7. Average correlation coefficient vs. cumulative variance with three selection methods in the time domain using simulation data.

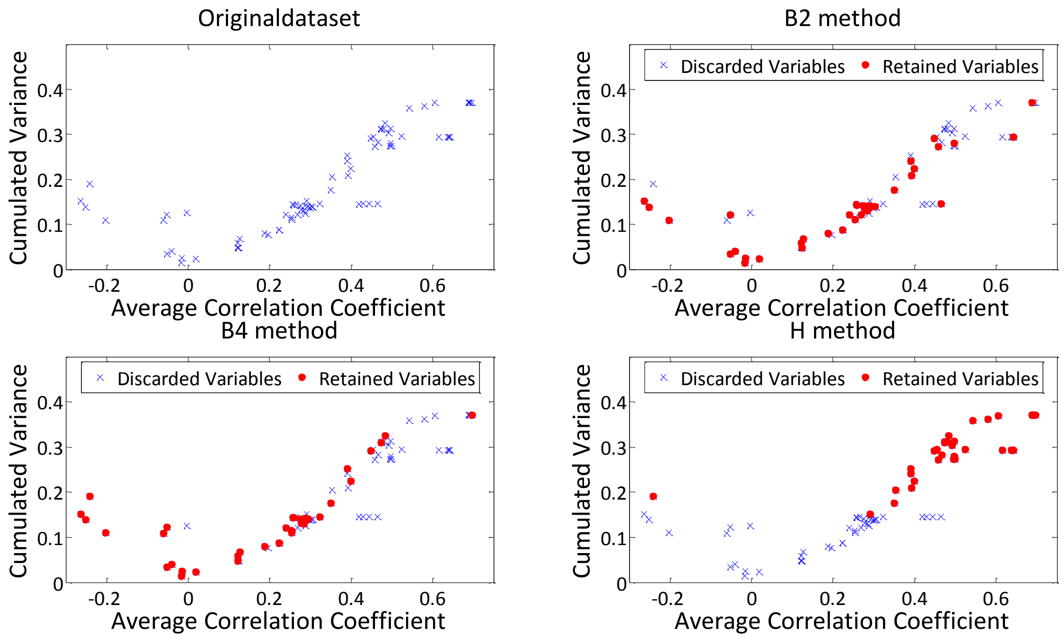

Fig. 8. Average correlation coefficient vs. cumulative variance with three selection methods in the time domain using SCADA data.

In order to further evaluate the optimality of the proposed methods, measures are also calculated using randomly selected variables in time domain for comparison. Table 2 gives the measures of the original dataset, B2 selection method and the mean measures of the 10 random trials. As with the B2 method, 15 and 35 variables are randomly selected for each trial using the simulation and SCADA data, respectively. It can be seen that the random trials have a lower performance across all three measures when compared to the t_b2 method. Consequently, based on these measures and the results of using them in combination, the B2 selection method in time domain demonstrates the best performance.

Table 2

Performance comparison between the B2 selection method and randomly selected variable set in time domain

	Original dataset	**t_b2**	**Random dataset (mean values of 10 trials)**
SCADA data			
Cumulative variance	100%	99.00%	92.62%
Average correlation	0.3418	0.162	0.3562
Total entropy	59.47	48.36	35.61
Simulation data			
Cumulative variance	100%	99.81%	95.75%
Average correlation	0.1107	0.0082	0.1144
Total entropy	21.24	18.88	14.42

For the simulation data, variables such as firing angle, pitch angle and active/reactive powers are almost always selected; variables with high dependency between them such as bus voltages and currents are not all selected. On the contrary, SCADA data have a more complex data structure than simulation data, as SCADA data consist of more signal variability, including variables like various temperatures and environmental conditions. Apart from these general parameters (*e.g.*, temperatures, oscillations and vibrations), most variables retained are related to the generator and grid. Variables related to blades (*e.g.* pitch angle, maximum pitch speed.) or environmental conditions (*e.g.* air pressure, relative humidity) are less likely to be selected because they may be highly dependent on wind speeds. It is worth noting that the wind speed and speed related variables are almost always selected. The B2 method is also applied to the SCADA data from a different turbine on the same wind farm; the variables selected are consistent with the result presented here.

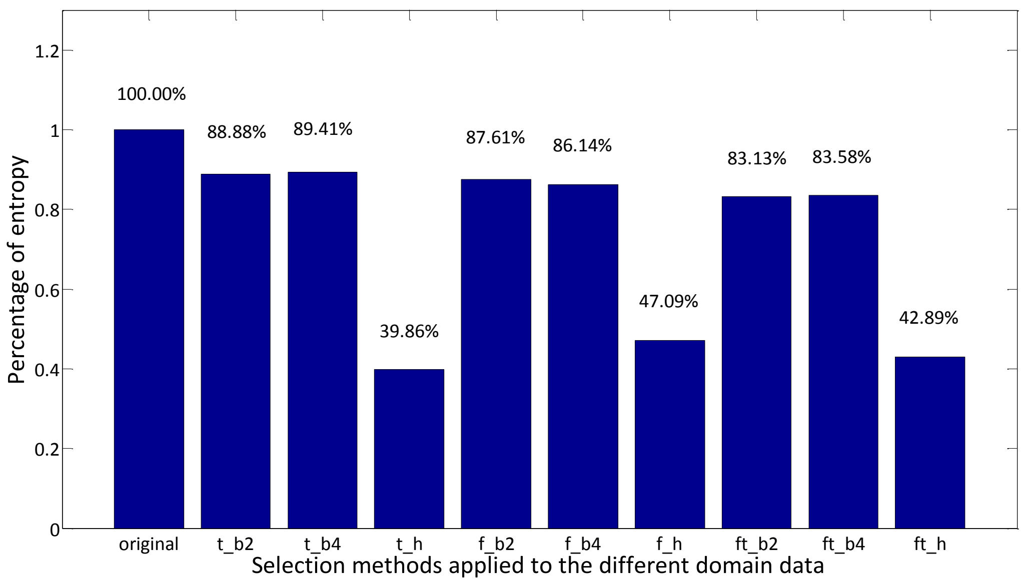

Fig. 9. Percentage of entropy obtained from different selection methods using simulation data in three domains.

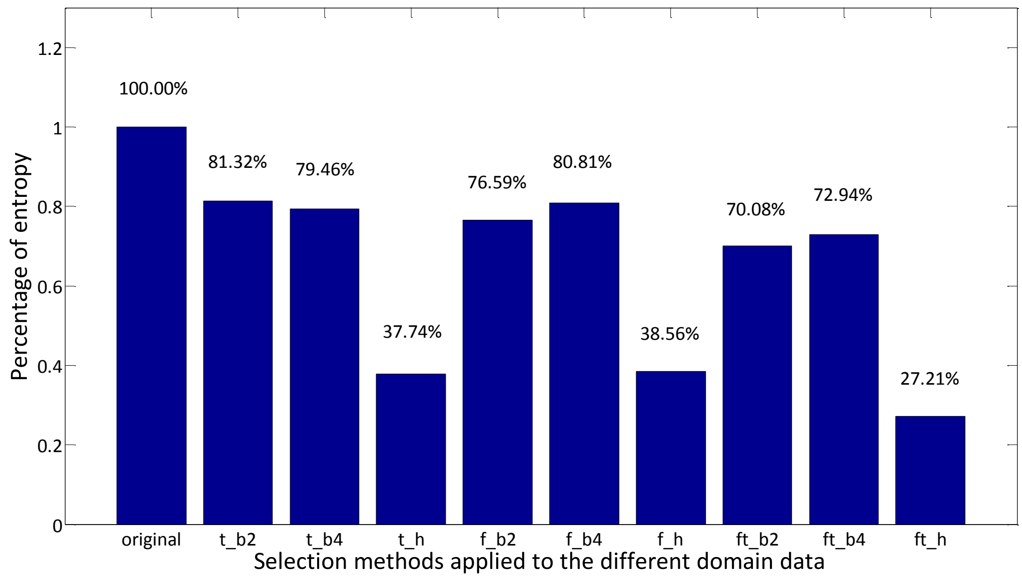

Fig. 10. Percentage of entropy obtained from different selection methods using SCADA data in three domains.

4.2 Selection under fault condition

To further evaluate whether the vital information relating to the particular fault is not being removed with the above selection methods, a DC-link capacitor fault, as an example, is simulated in PSCAD/EMTDC and data are collected. For multi-MW turbines, DC link capacitors are required to endure high ripple currents leading to self-heating, which, in addition to high ambient operating temperatures, can result in the deterioration of the electrolyte material and the loss of electrolyte by vapour diffusion. When the capacitor is operating at higher temperatures than the rated temperature, the DC voltage will be de-rated. The working life of a capacitor is also dependent upon operating voltage, current, and frequency. Consequently, DC link capacitors, although well designed, are considered one of the weakest components used in multi-MW power converters in the wind turbine.

Following the PCA of the original data, the PC, which is revealing the DC-link capacitor ageing fault in the original dataset, is first identified. The PCA then applies to the retained variables to obtain the new PCs, which are then compared to the original PC indentified. If the fault feature can be identified from the relevant new PC, it is confident to say that critical information associated with the fault is kept. In order to achieve this, the capacitor ageing fault is simulated several times to emulate the occurrence and severity of DC capacitor fault. The collected time-series data are then transformed using PCA to obtain the featured PCs. Having observed all the PCs, it is found that the DC capacitor fault is featured dominantly in the 7th principal components, *i.e.*, PC 7. Fig. 11 shows PC 7 transformed with data of the capacitor fault in increasing order of severity from no-fault occurred through 4% and 8% to 16% of capacitance loss. It can be seen that the peak amplitude during the fault increases rapidly when the fault severity increases. In order to quantify this change, the normalised entropy *H(X),* as given in eq. (19), is used as the measure, allowing comparison of entropy contained in different signals. Fig. 12 shows the normalised entropy of the capacitor fault at different fault levels. The fault level is simulated from the no-fault case to a highest level (16% in our study) with a constant increment of 1% capacitance loss. The blue dots are the actual calculated normalised entropy and the red line represents the fitted curve. The result clearly shows a decreasing trend of the normalised entropy. Consequently, a more severe fault will result in a larger change of waveform during the fault, which in turn leads to a larger change in the normalised entropy.

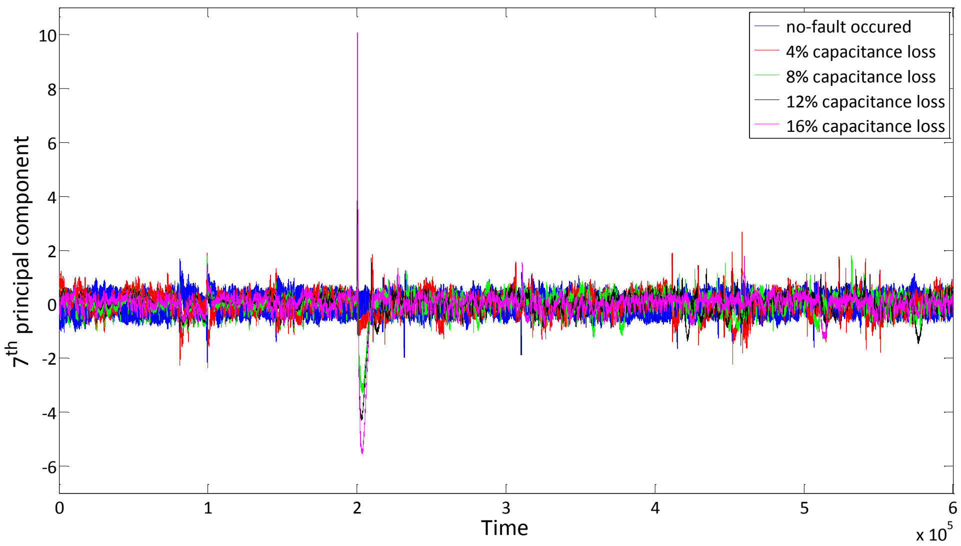

Fig. 11. The featured principal components of the dc-link capacitor fault at different ageing levels.

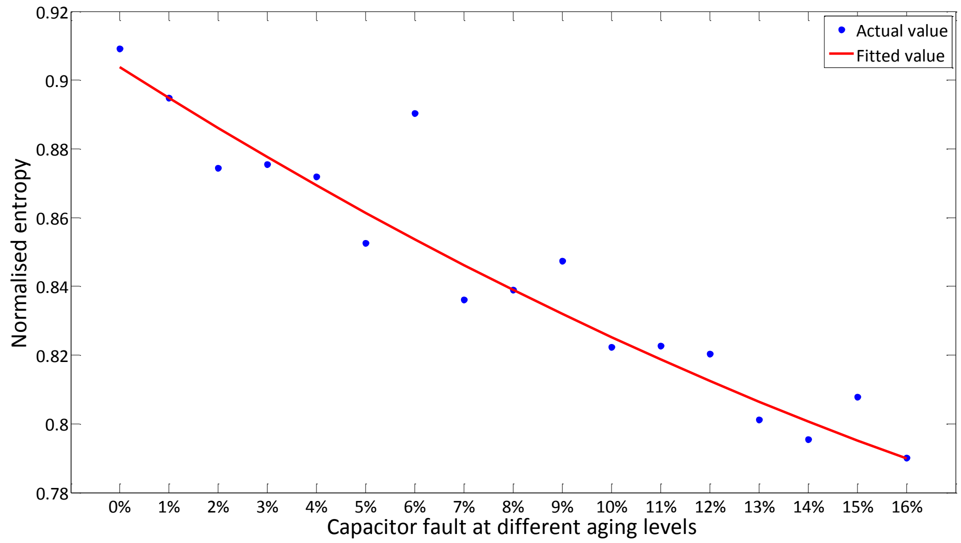

Fig. 12. Normalised entropy of the 7th principal components of the DC-link capacitor fault at different ageing levels.

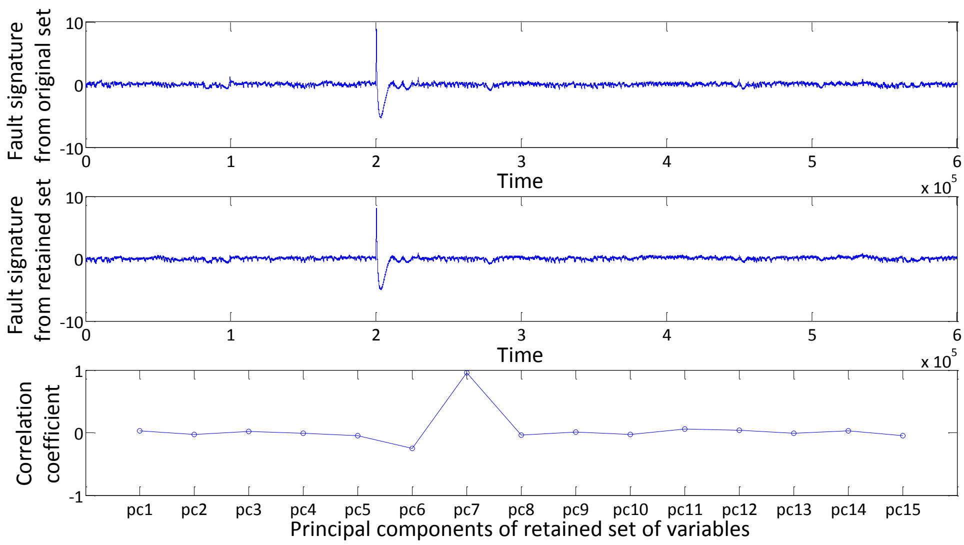

Fig. 13. Comparison of DC capacitor fault between original dataset and retained variables.

Fig. 13 shows an example of comparison of the signals for DC capacitor fault found in the original dataset and in the retained set of variables using the B2 method in the time domain. The first two plots are the featured 7th principal components transformed from the original dataset and from the reduced dataset; the Pearson's correlation coefficients between the first 15 principal complements from both datasets are also shown in the figure. This result clearly demonstrates that only the 7th principal component has a dominant correlation coefficient of 0.9682 and the rest are all close to 0. This again proves that the proposed selection algorithm has kept vital information of the fault, which can be used for further fault diagnosis.

Moreover, a nonlinear autoregressive exogenous artificial neural network (ANN) model with three layers and 10 neurons in the hidden layer [31] is used to further validate if fault feature is present in the retained variables based on the model prediction using different input datasets. We take SCADA data as an example. The ANN model is trained using SCADA data obtained from a fault-free turbine and then employed to predict the gearbox oil sump temperature of a faulty turbine on the same farm. The actual temperature and temperatures predicted using the original dataset and the B2 retained variables in time domain are shown in Fig. 14 (top). It can be seen that both predictions match the actual measurement precisely. The rise of temperature between 166.67 and 258.33 hours is due to a gearbox fault as indicated in the alarm log and from investigation of the data. The residuals, that is, the discrepancies between the model output and the actual output, using the original dataset and the B2 retained variables are shown in Fig 14 (bottom), where a zero line is also plotted as a reference. The coefficient of determination R^2 is employed here as a measure of how well the models explain the actual output data. The R^2 values for the models with all data variables and B2 selected variables are 0.9968 and 0.9934. This indicates that both models provide a precise fit, thus proving that the fault feature is present in the retained dataset. Consequently, results show that the proposed selection algorithm is able to reduce the dimension of the dataset while maintaining vital information of the fault.

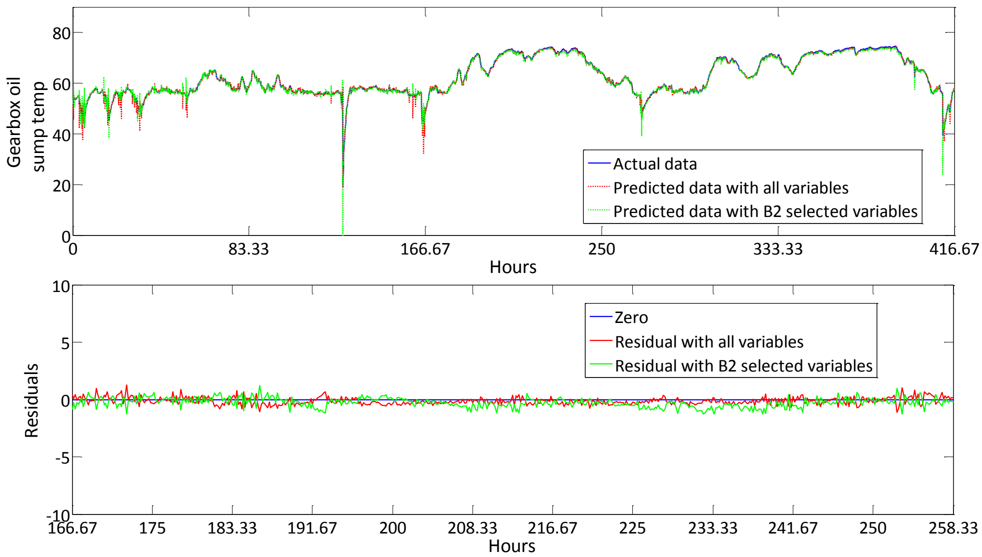

Fig 14. ANN model validation of SCADA data with gearbox fault. Upper: actual and predicted gearbox oil sump temperatures; bottom: residuals between the actual output and the model outputs.

5. Conclusions

In this paper, a new sensor selection technique is proposed, which uses PCA for condition monitoring of the distributed generation system oriented to wind turbines. The proposed method aims to identify a set of variables from huge amount of measurement data which can potentially reduce the number of physical sensors installed for condition monitoring whilst still maintaining sufficient information to assess the system's conditions. The selection process is examined not only with time series data but also with frequency series data and instantaneous frequency data in order to optimise sensor selection. The proposed technique is able to reduce the data dimension to 51.7% (15 out 29 variables) and 45.4% (35 out of 77 variables) for simulation and SCADA data, respectively. Findings from all three measures (cumulative variance, average correlation and information entropy) coincide with each other. It is found that the B2 method using both simulation and SCADA data in the time domain outperforms others, where the retained dataset has a cumulated percentage variance, average correlation and information entropy of 99.81%, 0.0082 and 81.32% for simulation data, and 99%, 0.162 and 88.88% for SCADA data, respectively. The results demonstrate that sufficient information is maintained in the retained dataset, while low degrees of correlation are ensured among the retained variables.

Furthermore, the selection methods are evaluated using simulation data of DC-link capacitor ageing fault to reveal whether the fault feature in the original dataset is still kept by comparing the featured principal components produced by the original dataset and retained dataset, respectively. Results have shown that under a fault condition, the selection algorithm not only reduces the dataset dimension but also keeps the vital features associated with the fault in the retained dataset with a high accuracy. The vital information of the fault present in the retained variables has been further validated by the ANN models. Consequently, the work has demonstrated the feasibility of the proposed selection methodology. Future work will be focussed on the study of the time-frequency domain data for the selection algorithms, as time-frequency data in 2D may reveal more abundant information. A more sophisticated selection criterion such as setting of multiple target objectives needs to be investigated for a more precise sensor selection process. Future work will also use simulation and practical data under different operational conditions of wind turbines to further validate the proposed algorithm and for further fault detection.

Acknowledgement

The authors would like to thank the UK Engineering and Physical Sciences Research Council (EPSRC) for their support under Grant EP/I037326/1. The permission of use SCADA data from Wind Prospect Ltd is also gratefully acknowledged.

References

[1] Ma X, Wang Y, Qin J. Generic model of a community-based microgrid integrating wind turbines, photovoltaics and CHP generations. Applied Energy 2013; 112:1475-1482.

[2] Krishna SK, Kumar SK. A review on hybrid renewable energy systems. Renewable and Sustainable Energy Reviews 2015; 52:907-916.

[3] Jardine AKS, Lin D, Bajevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing 2006; 20:1483-1510.

[4] Tavner PJ. Offshore wind turbines: reliability, availability and maintenance. Stevenage: Institution of Engineering and Technology 2012.

[5] Tian Z, Jin T, Wu B, Ding F. Condition based maintenance optimization for wind power generation systems under continuous monitoring. Renewable Energy 2011; 36:1502-1509.

[6] Nilsson J, Bertling L. Maintenance management of wind power systems using condition monitoring systems - life cycle cost analysis for two case studies. IEEE Transactions on Energy Conversion 2007; 22: 223-229.

[7] Marquez FPG, Tobias AM, Perez JMP, Papaelias M. Condition monitoring of wind turbines: Techniques and methods. Renewable Energy 2012; 46:169-178.

[8] Ma X. Novel early waning fault detection for wind turbine-based DG systems. Proceedings of 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies (ISGT Europe), 2011.

[9] Cross P, Ma X. Nonlinear system identification for model-based condition monitoring of wind turbines. Renewable Energy 2014; 71:166-175.

[10] Liu L, Wang S, Liu D, Zhang Y, Peng Y. Entropy-based sensor selection for condition monitoring and prognostics of aircraft engine. Microelectronics Reliability 2015; 55(9-10):2092-2096.

[11] Shen X, Liu S, Varshney P. Sensor selection for nonlinear systems in large sensor networks. IEEE Transaction on Aerospace and Electronics Systems 2014; 50(4):2664-2678.

[12] Hovland G, McCarragher B. Control of Sensory Perception in Discrete Event Systems Using Stochastic Dynamic Programming. Journal of Dynamic Systems, Measurement and Control 1999; 121(2):200-205.

[13] Mohammadi A, Asif A. Consensus-based distributed dynamic sensor selection in decentralised sensor networks using the posterior Cramer-Rao lower bound. Signal Processing 2015; 108:558-575.

[14] Van der Maaten LJP, Postma EO, Van den Herik HJ. Dimensionality Reduction: A Comparative Review. Tilburg University Technical Report, TiCC-TR 2009-005, 2009.

[15] Feng Z, Liang M. Fault diagnosis of wind turbine planetary gearbox under nonstationary conditions via adaptive optimal kernel time-frequency analysis. Renewable Energy 2014; 66:468-477.

[16] Huang NE, Shen SSP. Hilbert-Huang transform and its applications. Singapore: World Scientific 2005. ISBN: 978-981-4480-06-2.

[17] Zhang RR, Ma S, Safak E, Hartzell E. Hilbert-Huang transform analysis of dynamic and earthquake motion recordings. Journal of Engineering Mechanics 2003; 129:861–875.

[18] Yan R, Gao RX. Hilbert-Huang transform-based vibration signal analysis for machine health monitoring. IEEE Transactions on Instrumentation and Measurement 2006; 55: 2320-2329.

[19] Jackson J. Users guide to principal components. A Wiley-Interscience Publication, 1999. ISBN 0-471-62267-2.

[20] Skittides C, Fruh WG. Wind forecasting using Principal Component Analysis. Renewable Energy 2014; 69:365-374.

[21] Ata R. Artificial neural networks applications in wind energy systems: a review. Renewable and Sustainable Energy Reviews 2015; 49:534-562.

[22] Jolliffe IT. Principal component analysis. New York: Springer. 2002. ISBN: 978-0387954424.

[23] Al-Kandari N, Jolliffe IT. Variable selection and interpretation of covariance principal components. Communications in Statistics - Simulation and Computation 2001; 30(2):339-354.

[24] Cumming JA, Wooff DA. Dimension reduction via principal variables. Computational statistics & data analysis 2007; 52:550-565.

[25] McCabe P. Principal variables. American Statistical Association 1984; 26:137-144.

[26] Grunwald P, Vitanyi P. Shannon information and kolmogorov complexity. IEEE Trans. Information Theory 2004.

[27] Tang B, Song T, Li F, Deng L. Fault diagnosis for a wind turbine transmission system based on manifold learning and Shannon wavelet support vector machine. Renewable Energy 2014; 62:1-9.

[28] PSCAD power system simulation: wind turbine application technical paper, 2006. Available at http://www.cedrat-technologies.com/fileadmin/user_upload/cedrat_groupe/Publications/Publications/2006/06/2006_Wind_turbine_PSCAD_V42_ref.pdf, accessed on 29th Jan 2016.

[29] Haque ME, Muttaqi KM, Negnevitsky M. A control strategy for output maximisation of a PMSG-based variable-speedwind turbine. Australian Journal of Electrical and Electronics Engineering 2009; 5: 263-270.

[30] Hawkins DM. Exploring multivariate data using the minor principal components. Statistician 1984; 33:325-338.

[31] Cross P, Ma X. Model-based and fuzzy logic approaches to condition monitoring of operational wind turbines. International Journal of Automation and Computing 2015; 12: 25-34.