Control Engineering Practice 72 (2018) 177–191

Contents lists available at ScienceDirect

Control Engineering Practice

journal homepage: www.elsevier.com/locate/conengprac

Canonical variable analysis and long short-term memory for fault diagnosis and performance estimation of a centrifugal compressor

Xiaochuan Li [a,*], Fang Duan [b], Panagiotis Loukopoulos [c], Ian Bennett [d], David Mba [a]

[a] Faculty of Technology, De Montfort University, Leicester, LE1 9BH, UK
[b] School of Engineering, London South Bank University, London, SE1 0AA, UK
[c] School of Power and Propulsion, Cranfield University, London, MK430 AL, UK
[d] Rotating Equipment Department, Royal Dutch Shell, Hague, 2501 AN, Netherlands

ARTICLE INFO

Keywords:
Condition monitoring
Long short-term memory
Canonical variable analysis
Fault identification
Performance estimation

ABSTRACT

Centrifugal compressors are widely used for gas lift, re-injection and transport in the oil and gas industry. Critical compressors that compress flammable gases and operate at high speeds are prioritized on maintenance lists to minimize safety risks and operational downtime hazards. Identifying incipient faults and predicting fault evolution for centrifugal compressors could improve plant safety and efficiency and reduce maintenance and operation costs. This study proposes a dynamic process monitoring method based on canonical variable analysis (CVA) and long short-term memory (LSTM). CVA was used to perform fault detection and identification based on the abnormalities in the canonical state and the residual space. In addition, CVA combined with LSTM was used to estimate the behavior of a system after the occurrence of a fault using data captured from the early stages of deterioration. The approach was evaluated using process data obtained from an operational industrial centrifugal compressor. The results show that the proposed method can effectively detect process abnormalities and perform multi-step-ahead prediction of the system's behavior after the appearance of a fault.

1. Introduction

Modern industrial natural gas processing plants are becoming increasingly complex due to the use of diverse equipment. Because of their complexity, developing an accurate first-principle failure model for such large-scale industrial facilities can be challenging (He, Li, & Bechhoefer, 2012). Thus, existing condition monitoring approaches for industrial processes are typically derived from routinely collected system operating data. Due to the rapid growth and advancement in data acquisition technology, long-term continuous measurements can be taken with sensors mounted on machinery systems. The monitored data are easily stored and analyzed to extract important process condition information.

A number of advanced multivariate statistical techniques have been developed based on condition monitoring data for diagnostic and prognostic health monitoring, such as filtering-based models (Guerra & Kolodziej, 2017), multivariate time-series models (Serdio, Lughofer, Pichler, Buchegger, & Pichler, 2014) and neural networks (Tran, Althobiani, & Ball, 2014). Key challenges in the implementation of these techniques include strongly correlated variables, high-dimensional data, changing operating conditions and inherent system uncertainty (Jiang, Huang, Zhu, Yang, & Braatz, 2015). Recent developments in dimensionality reduction techniques have shown improvements in identifying faults from highly correlated process variables. Conventional dimensionality reduction methods include principal component analysis (PCA) (Harrou, Nounou, Nounou, & Madakyaru, 2013), independent component analysis (ICA) (Hyvärinen, Karhunen, & Oja, 2004) and partial least-squares analysis (PLSA) (Kruger & Dimitriadis, 2008). These basic multivariate methods perform well under the assumption that process variables are time independent. However, this assumption might not hold true for real industrial processes (especially chemical and petrochemical processes) because sensory signals affected by noise and disturbances often show strong correlations between the past and future sampling points (Jiang et al., 2015). Therefore, variants of the standard multivariate approaches (Li & Qin, 2001; Stefatos & Hamza, 2010; Yin, Zhu, Member, & Kaynak, 2015) were developed to solve the time-independency problem, which makes these approaches more suitable for dynamic process monitoring. In addition to approaches derived from PCA, ICA and PLSA, canonical variable analysis (CVA)

* Corresponding author.
E-mail addresses: lix29@lsbu.ac.uk, xiaochuan.li0309@gmail.com (X. Li).

https://doi.org/10.1016/j.conengprac.2017.12.006
Received 5 August 2017; Received in revised form 16 December 2017; Accepted 16 December 2017
Available online 3 January 2018

is a multivariate analysis tool. CVA is a subspace method that takes serial correlations between different variables into account and hence is particularly suitable for dynamic process modeling (Stubbs, Zhang, & Morris, 2012). The effectiveness of CVA has been verified by extensive simulation studies (Huang, Cao, Tian, & Deng, 2015; Stubbs et al., 2012) and data captured from experimental test rigs (Cárcel, Cao, & Mba, 2007). However, the effectiveness of CVA in real complex industrial processes has not been fully studied. In the present study, condition monitoring data acquired from an operational industrial centrifugal compressor were used to prove the superior performance of CVA for fault detection and identification in industrial processes.

Once a fault is detected in industrial processes, a prognostic tool is required to predict how the system will behave under faulty operating conditions. Examples of successful applications of different methodologies for performance estimation in the presence of faults are available (Ai et al., 2009; Ai, Zheng, Wang, Jang, & Song, 2010; Zheng et al., 2010). In addition to the abovementioned approaches, CVA is a subspace identification method that can be used to build a dynamic model using measurements of a system's input and output signals. The obtained model can be utilized to predict system performance given expected future input conditions. System inputs used in subspace identification are typically manipulated or controllable variables such as inlet liquid and gas flow valve position. However, the performance of complex industrial systems, such as turbomachines, is not only associated with the system's input signals, which can be manipulated, but is also affected by variations in environmental conditions such as the ambient temperature (Campanari, 2000). The inlet gas temperature of the compressor in this study is a prime example of how environmental conditions can affect a system's performance. Specifically, for the centrifugal compressor, the temperature of the gas to be compressed is largely determined by the ambient temperature when the gas passes through long transmission pipelines to the compressor. As a result, the magnitude of the compressor's inlet gas temperature changes periodically, most commonly every 24 h. To account for the impact of ambient temperature on a system's performance, and thereby allow both the environmental factors and the human interventions to be factored in when predicting the system's future behavior, a time series prediction method is required to forecast the magnitude of the inlet gas temperature based on historical data.

Many data-driven methodologies are available for the prediction of time series, including the widely applied support vector machine (SVM) (Zhang, Wang, & Zhang, 2017), echo state network (ESN) (Chouikhi, Ammar, Rokbani, & Alimi, 2017) and nonlinear auto-regressive moving average (Wootton, Butcher, Kyriacou, Day, & Haycock, 2017) methods. One main challenge of sequence prediction tasks that involve temporal dependencies is handling long-range dependencies (Bengio, Simard, Frasconi, & Member, 1994). Long short-term memory (LSTM) is a powerful learning model that has been extraordinary capable in a wide range of machine learning tasks such as machine remaining useful life prediction (Wu, Yuan, Dong, Lin, & Liu, 2017), visual object recognition (Son, 2017) and speech recognition (Chen & Wang, 2016). LSTM networks use special units in hidden layers that allows inputs to be remembered for long periods; therefore they have great potential in constructing end-to-end systems (Lecun, Bengio, & Hinton, 2015). However, few studies have been conducted to predict sensory signals collected from industrial processes. In this investigation, we explore the ability of LSTM to model the compressor inlet temperature time series. The predicted future inlet gas temperature along with the manipulated system's input signals were fed into a CVA model to perform machine behavior estimation.

The major contributions of this paper are as follows:

- The use of CVA for fault detection using data captured from an operational industrial centrifugal compressor;
- The combination of the canonical state space and the residual space information for fault root-cause analysis;
- The application of LSTM to predict the inlet gas temperature of the compressor in the study;
- The combination of CVA and LSTM for multi-step-ahead prediction of the system's behavior after the occurrence of a fault.

2. Methodology

2.1. CVA for fault detection and identification

CVA is a dimensionality reduction technique used to monitor machine operation by transferring high-dimensional process data into one-dimensional health indicators. Condition monitoring data captured from the system operating under healthy conditions are used to calculate the threshold for normal operating limits. Abnormal operating conditions can be detected when the value of the health indicator exceeds pre-set limits.

The objective of CVA is to maximize the correlation between two sets of variables (Russell, Chiang, & Braatz, 2000). To generate two data matrices from the measured data $y_t \in \mathcal{R}^n$ (n indicates that there are n variables being recorded at each sampling time t), the data were expanded at each sampling time by including a, the number of previous samples, and b, the number of future samples, to construct the past and future sample vectors $y_{a,t} \in \mathcal{R}^{na}$ and $y_{b,t} \in \mathcal{R}^{nb}$.

$$y_{a,t} = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-a} \end{bmatrix} \in \mathcal{R}^{na} \tag{1}$$

$$y_{b,t} = \begin{bmatrix} y_t \\ y_{t+1} \\ \vdots \\ y_{t+b-1} \end{bmatrix} \in \mathcal{R}^{nb}. \tag{2}$$

To avoid excessive influence of variables with large absolute values, the past and future sample vectors were normalized to zero mean vectors $\hat{y}_{a,t}$ and $\hat{y}_{b,t}$, respectively. Then, the vectors $\hat{y}_{a,t}$ and $\hat{y}_{b,t}$ at different sampling times were rearranged according to Eqs. (3) and (4) to produce the reshaped matrices $\hat{Y}_a$ and $\hat{Y}_b$:

$$\hat{Y}_a = [\hat{y}_{a,t+1}, \hat{y}_{a,t+2}, \dots, \hat{y}_{a,t+N}] \in \mathcal{R}^{na \times N} \tag{3}$$

$$\hat{Y}_b = [\hat{y}_{b,t+1}, \hat{y}_{b,t+2}, \dots, \hat{y}_{b,t+N}] \in \mathcal{R}^{nb \times N} \tag{4}$$

where $N = l - a - b + 1$ and l represents the total number of samples for y_t. The Cholesky decomposition was then applied to the past and future matrices $\hat{Y}_a$ and $\hat{Y}_b$ to configure a Hankel matrix $\mathcal{H}$ (Samuel & Cao, 2015). The purpose of using the Cholesky decomposition here is to transfer $\hat{Y}_a$ and $\hat{Y}_b$ into a new correlation matrix with reduced dimensionality such that the subsequent calculations can be conducted in a stable and fast manner. To find the linear combination that maximizes the correlation between the two sets of variables, the truncated Hankel matrix $\mathcal{H}$ is then decomposed using the singular value decomposition (SVD):

$$\mathcal{H} = \sum_{a,a}^{-1/2} \sum_{a,b} \sum_{b,b}^{-1/2} = U \sum V^T \tag{5}$$

where $\Sigma_{a,a}$ and $\Sigma_{b,b}$ are the sample covariance matrices and $\Sigma_{a,b}$ denotes the cross-covariance matrix of $\hat{Y}_a$ and $\hat{Y}_b$.

If the order of the truncated Hankel matrix $\mathcal{H}$ is r, then U, V and $\sum$ have the following form:

$$U = [u_1, u_2, \dots, u_r] \in \mathcal{R}^{na \times r}$$

$$V = [v_1, v_2, \dots, v_r] \in \mathcal{R}^{nb \times r}$$

$$\sum = \begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_r \end{bmatrix} \in \mathcal{R}^{r \times r}.$$

The columns of $U = [u_1, u_2, \ldots, u_r]$ and the columns of $V = [v_1, v_2, \ldots, v_r]$ are called the left-singular and right-singular vectors of $\mathcal{H}$, respectively. $\sum$ is a diagonal matrix, and its diagonal elements are called singular values, which depict the degree of correlation between the corresponding left-singular and right-singular vectors. The right-singular vectors in V, which correspond to the largest q singular values, are retained in the truncated matrix $V_q = [v_1, v_2, \ldots, v_q] \in \mathcal{R}^{na\times q}$. This matrix will be used later to perform dimension reduction on the measured data.

With the truncated matrix V_q, the na dimensional past vector $\hat{Y}_a \in \mathcal{R}^{na\times N}$ can be further converted into a reduced q-dimensional matrix $\Phi \in \mathcal{R}^{q\times N}$ (the columns of Φ are z_t, which are called canonical state variates) by:

$$\Phi = [z_{t=1}, z_{t=2}, \ldots, z_{t=N}] = K \cdot \hat{Y}_a. \tag{6}$$

Similarly, the residual variates $\Psi \in \mathcal{R}^{na\times N}$ can be calculated according to Eq. (7):

$$\Psi = [\varepsilon_{t=1}, \varepsilon_{t=2}, \ldots, \varepsilon_{t=N}] = G \cdot \hat{Y}_a \tag{7}$$

where K and G are the projection matrices and can be computed as $K = V_q^T \sum_{a,a}^{-1/2} \in \mathcal{R}^{q\times na}$ and $G = (I - V_q V_q^T)\sum_{a,a}^{-1/2} \in \mathcal{R}^{na\times na}$, respectively. The canonical variates z_t and residual variates ε_t consist of valuable information that is needed to construct health indicators. To be specific, the state space matrix z_t represents the projection of the measurement matrix into the q-dimensional space, while the residual space matrix ε_t is associated with system variations not represented by the state space.

The health indicators adopted in this study were the Hotelling T^2 and Q (SPE) statistics, which were introduced by Hotelling in 1936 (Hotelling, 1936). The Hotelling health indicators at sampling time t can be calculated as follows:

$$T_t^2 = \sum_{j=1}^{q} z_{t,j}^2 \tag{8}$$

$$Q_t = \sum_{j=1}^{na} \varepsilon_{t,j}^2. \tag{9}$$

Since the Gaussian distribution does not hold true for non-linear processes, the actual probability density function of the health indicator in this study is calculated using a method called the kernel density estimation (KDE) (Odiowei & Yi, 2010). Then, machine faults will be considered every time the value of the health indicator exceeds the threshold.

In addition to fault detection, T^2 and Q statistics can also be used to calculate contributions of variables to the detected fault. For a new observation y_t, the CVA-based state space contributions at time instant t can be computed from the canonical state variates as follows:

$$\begin{aligned} c_t^{state} &= \left(K \cdot \hat{Y}_{a,t}\right)^T \left(K \cdot \hat{Y}_{a,t}\right) = \left(K \cdot \hat{Y}_{a,t}\right)^T \sum_{i=1}^{q} \left(\hat{Y}_{a,t} K_i^T\right)^T \\ &= \sum_{i=1}^{q} \left(\hat{Y}_{a,t} K_i^T\right)\left(\hat{Y}_{a,t} K_i^T\right)^T \end{aligned} \tag{10}$$

where $\hat{Y}_{a,t}$ denotes the column vector of $\hat{Y}_a$ at time instant t. K_i is the ith row of matrix K. Similarly, the CVA-based residual space contributions at time instant t can be computed as follows:

$$\begin{aligned} c_t^{residual} &= \left(G \cdot \hat{Y}_{a,t}\right)^T \left(G \cdot \hat{Y}_{a,t}\right) = \left(G \cdot \hat{Y}_{a,t}\right)^T \sum_{i=1}^{na} \left(\hat{Y}_{a,t} G_i^T\right)^T \\ &= \sum_{i=1}^{na} \left(\hat{Y}_{a,t} G_i^T\right)\left(\hat{Y}_{a,t} G_i^T\right)^T \end{aligned} \tag{11}$$

where G_i is the ith row of matrix G. According to the literature (Jiang et al., 2015), a limitation of the CVA model is that the calculated contributions can be overly sensitive due to the inversion procedure of $\sum_{a,a}^{-1/2}$, which would lead to incorrect identification of faulty variables. To alleviate this sensitivity, a combination of state and residual space contributions is adopted for the identification of faulty variables in this study.

2.2. CVA-based state space model for performance estimation

CVA can be used to build a state-space model that represents the dynamics of the system using condition monitoring data. Given system input time series u_t and output time series y_t, the linear state-space model can be built as follows Russell et al. (2000):

$$x_{t+1} = Bx_t + Cu_t + w_t \tag{12}$$

$$y_t = Dx_t + Eu_t + +Lw_t + v_t \tag{13}$$

where x_t is the state vector with an order q; B, C, D, E and L are model coefficient matrices; and w_t and v_t are independent white noise. According to the literature (Odiowei & Yi, 2010), if the number of retained states q is no less than the actual order of the system, we can substitute the matrix x_t with the canonical state variates z_q obtained from Eq. (6). L can be calculated from the covariance matrix of the error in the prediction (Russell et al., 2000). Moreover, the authors suggested the use of multivariate regression for the calculation of the unknown coefficient matrices B, C, D and E:

$$\begin{bmatrix} \hat{B} & \hat{C} \\ \hat{D} & \hat{E} \end{bmatrix} = \mathrm{Cov}\left[\begin{pmatrix} z_{t+1} \\ y_t \end{pmatrix}, \begin{pmatrix} z_t \\ u_t \end{pmatrix}\right] \cdot \mathrm{Cov}^{-1}\left[\begin{pmatrix} z_t \\ u_t \end{pmatrix}, \begin{pmatrix} z_t \\ u_t \end{pmatrix}\right]. \tag{14}$$

The procedures of performance estimation using the model described above for the centrifugal compressor in this study are summarized as follows:

- Determine the system inputs u_t (u_t consists of two variables: a manipulated variable (suction throttling valve position) and an unmanipulated variable (suction temperature)) and outputs y_t (measured performance variables).
- Obtain a training data set from the compressor during the early stages of deterioration. Construct a state space model based on the obtained training data. Model coefficient matrices can be calculated according to Eqs. (12)–(14). The constructed CVA-based state space model can be used to predict system outputs in the future $\hat{y}_{t(future)}$ for the future expected input conditions $u_{t(future)}$.
- Capture a validation data set from the compressor that covers the entire degradation process to test the capabilities of the CVA model to estimate performance deterioration of the compressor under faulty conditions. Then, feed the CVA model with the same suction throttling valve set points used during the total duration of the validation data set. Meanwhile, use a trained LSTM to predict the suction temperature for the total duration of the validation data set and feed the predicted suction temperature values into the CVA model together with the suction throttling valve set points. Predict the values of system outputs as per Eqs. (12) and (13). Compare the predicted outputs with the actual measured outputs in the validation data set to evaluate the predictive accuracy of the CVA model.

Section 2.3 details how an LSTM model can be built and trained to estimate a compressor's future suction temperature. In reality, when a fault occurs, site engineers can predict suction throttling valve set points by looking at the production plan and predict suction temperature values using a trained LSTM model. Then, a trained CVA model can be used to predict how the system will behave under faulty conditions for the expected (predicted) future input conditions.

2.3. LSTM for time series prediction

Section 2.2 describes how a CVA model can be used to estimate the behavior of a system under faulty operating conditions given expected future input conditions. This section describes how uncontrollable system input variables (i.e., temperature of the gas to be compressed) can be predicted by the LSTM based on historical data.

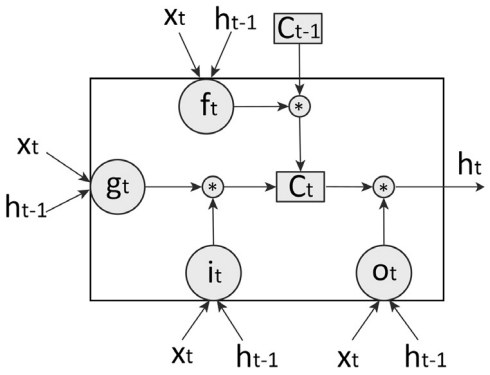

Fig. 1. Schematic diagram of an LSTM cell.

2.3.1. Long short-term memory

LSTM is a special type of recurrent neural network that eases the "vanishing gradient" problem (Hochreiter & Schmidhuber, 1997). The schematic diagram of an LSTM cell is shown in Fig. 1. The core idea behind LSTMs is a memory cell C_t that can maintain its state information over time, which allows gradients to flow over long sequences. The information flow into and out of the memory cell C_t is regulated by three gates: an input gate i_t, a forget gate f_t and an output gate o_t. At every time instant, the LSTM cell reads the input of the current time x_t and the hidden state h_{t-1} from the previous step. The combination of input x_t and hidden state h_{t-1} is processed by passing through a squashing tanh function:

$$g_t = \tanh(W_g x_t + U_g h_{t-1} + b_g) \tag{15}$$

where W_g, U_g and b_g are input weights, recurrent weights and bias, respectively. The forget gate f_t will determine which information should be removed from the memory cell C_t through an element-wise sigmoid function:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f). \tag{16}$$

Then, the input gate will determine which information will be stored in the memory cell C_t through an element-wise sigmoid function:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i). \tag{17}$$

Next, the information in the memory cell is updated through partial forgetting of the information stored in the previous memory cell C_{t-1} via the following:

$$C_t = f_t * C_{t-1} + i_t * g_t \tag{18}$$

where $*$ denotes the element-wise product function of two vectors. At this step, the learnable forget gate f_t determines the extent to which past information stored in C_{t-1} will be forgotten. The value of f_t is set between 0 and 1. If $f_t \to 0$, the old state will be completely forgotten, and if $f_t \to 1$, the past information will be maintained in the memory cell. Lastly, the output hidden state h_t is updated based on the computed cell state C_t as follows:

$$h_t = o_t * \tanh(C_t). \tag{19}$$

Network input weights $W_{g,i,f,o}$, recurrent weights $U_{g,i,f,o}$ and bias $b_{g,i,f,o}$ are calculated through the training process, and the trained network will be used to predict time series data in Section 3.

2.3.2. Data pre-processing

It is common practice to perform data pre-processing when using machine learning algorithms for time series forecasts (Mohammad & Nishida, 2016). Seasonality decomposition analysis (Cleveland, Cleveland, & Terpenning, 1990) was first performed on the original suction temperature signals, and the results revealed that the seasonality is multiplicative because the magnitude of the seasonal variation changes depending on the mean level of the time series (Koehler, Snyder, & Ord, 2001). Therefore, the pre-processing in this study starts with applying log-transformation to the suction temperature time series. Then, seasonal decomposition is conducted to decompose the log-transformed data into seasonal, trend and random components. In this way, the seasonal decomposition's normally additive split is changed into a multiplicative split, thereby transferring the suction temperature's seasonality into the multiplicative seasonality. After removal of the seasonality, two moving windows (an input window and an output window) are applied to the sum of the smooth trend and the noisy component to prepare training samples for LSTM. Fig. 2 illustrates how the moving windows are used to prepare training samples based on the decomposed time series. The last value of the trend data at the end of the input window (shown as a black dot in the figure) is subtracted from all samples covered by the input and output windows. The purpose of this step is to normalize the data to avoid excessive influence of extremely large values. Then, the input and output windows slide by a single increment each time until the rightmost point of the output window reaches the end of the training data set. With this process, constant size inputs and outputs are extracted and then fed into an LSTM model to train the model parameters.

2.3.3. Network architecture

The LSTM network used in this study consists of a sequence of 6 layers: an input layer, 3 LSTM hidden layers, a dense hidden layer with linear activation and an output layer. Fig. 3 illustrates the training

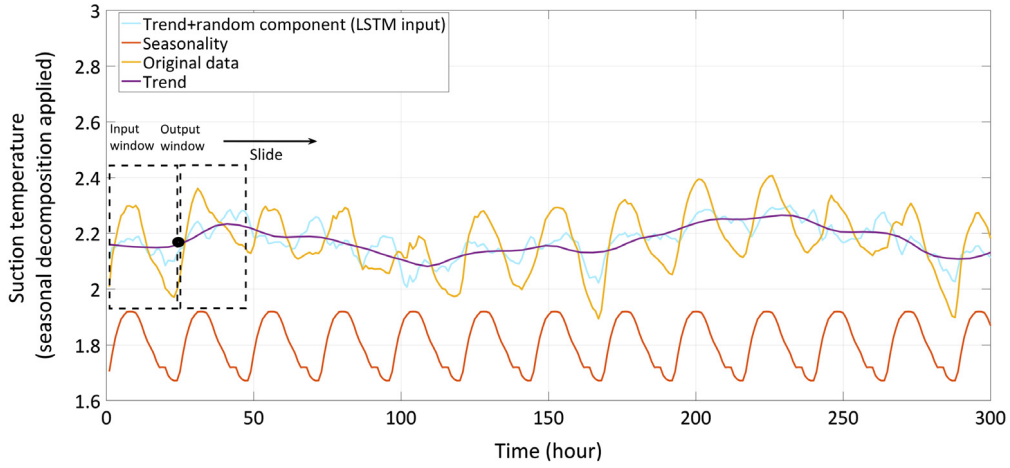

Fig. 2. Illustration of preparing the LSTM inputs.

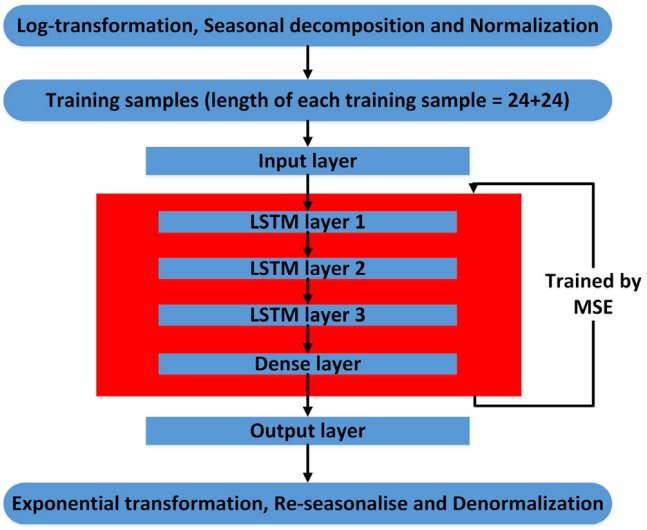

Fig. 3. Illustration of the use of the proposed LSTM network for time series prediction.

process of the proposed LSTM network. The trained LSTM network can be used to perform multi-step-ahead prediction of the compressor's suction temperature. The procedure used to determine the optimal LSTM architecture as well as how to choose the optimal network parameters will be detailed in Section 3.

2.3.4. Multi-step-ahead time series prediction

To perform multi-step-ahead predictions of the future suction temperature, the length of the input and output moving window is set to 24 time-steps in this study. In other words, each LSTM training sample contains a 24-long input vector and a 24-long output vector, and the prediction horizon is the same as the output window length, which is 24 future time steps. Compared to the one-step-ahead prediction, this approach provides the power to predict the entire daily (24 h) suction temperature ahead at once. In addition, setting the prediction horizon to 24 time steps could provide site engineers with the entire daily performance estimation of the compressor, which allows ample time for decision making and maintenance scheduling. Longer prediction horizons may provide site engineers with more reaction time for a severe fault predicted by the CVA model, but a longer prediction horizon will result in fewer available training samples for an LSTM network and therefore lower forecast accuracy given that we have limited historical data. Taking into consideration both aspects, the prediction horizon was set to 24 h in this study.

After each prediction, the output window was moved by 24 steps, in effect creating a sequence of predictions of 24 time steps (see Fig. 4).

2.4. Combining CVA and LSTM for performance estimation

The proposed CVA–LSTM system for performance estimation of the centrifugal compressor comprises four main steps as shown in Fig. 5: detecting and identifying an incipient fault, training the CVA-based faulty model, training-validating the LSTM model and predicting the performance. The role of each step is detailed as follows:

- Step 1. Detecting and identifying an incipient fault: Detect and identify an incipient fault using a CVA-based fault detection model.
- Step 2. Training the CVA-based faulty model: Build and train a CVA-based state space model using data acquired during early stages of deterioration.
- Step 3. Training-validating the LSTM model: Build and train an LSTM network using historical suction temperature data. Predict future suction temperature values (under faulty operating conditions) using the trained LSTM model.
- Step 4. Predicting the performance: A validation data set that covers the entire degradation process is first obtained from the compressor. Then, the trained CVA model is fed with the same suction throttling valve set points used during the total duration of the validation data set. Meanwhile, the predicted suction temperature values by the LSTM network are fed into the CVA model together with the suction throttling valve set points. (In reality, when a fault occurs, site engineers could predict suction throttling valve set points by looking at the production plan and predict suction temperature values using a trained LSTM model). The trained CVA model is used to predict system output variables for the given inputs. The performance of the model is evaluated by calculating the error between the predicted values and actual measured values in the validation data set.

3. Case study

3.1. Data acquisition

Centrifugal compressors are widely used in a large number of different compression applications in the oil and gas industry. These machines are equipped with a large variety of sensors to enable fully automated online supervision of various operating parameters. Moreover,

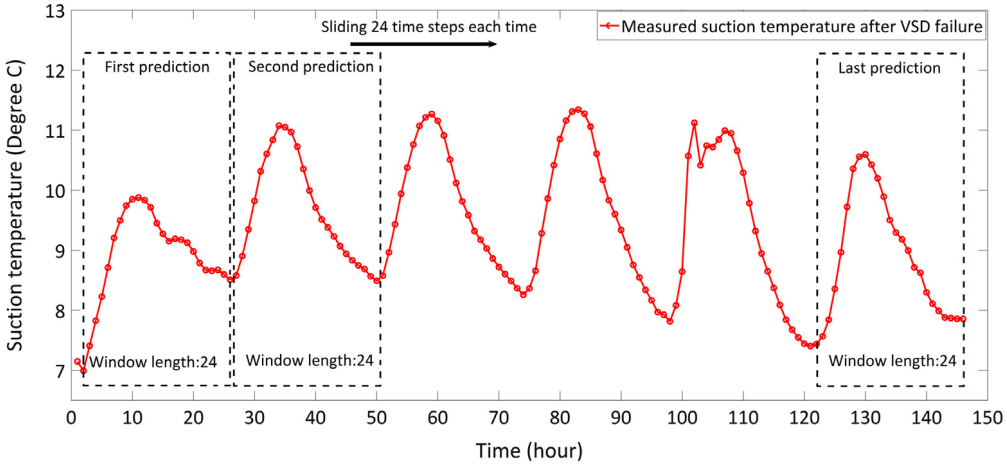

Fig. 4. Illustration of multi-step-ahead prediction.

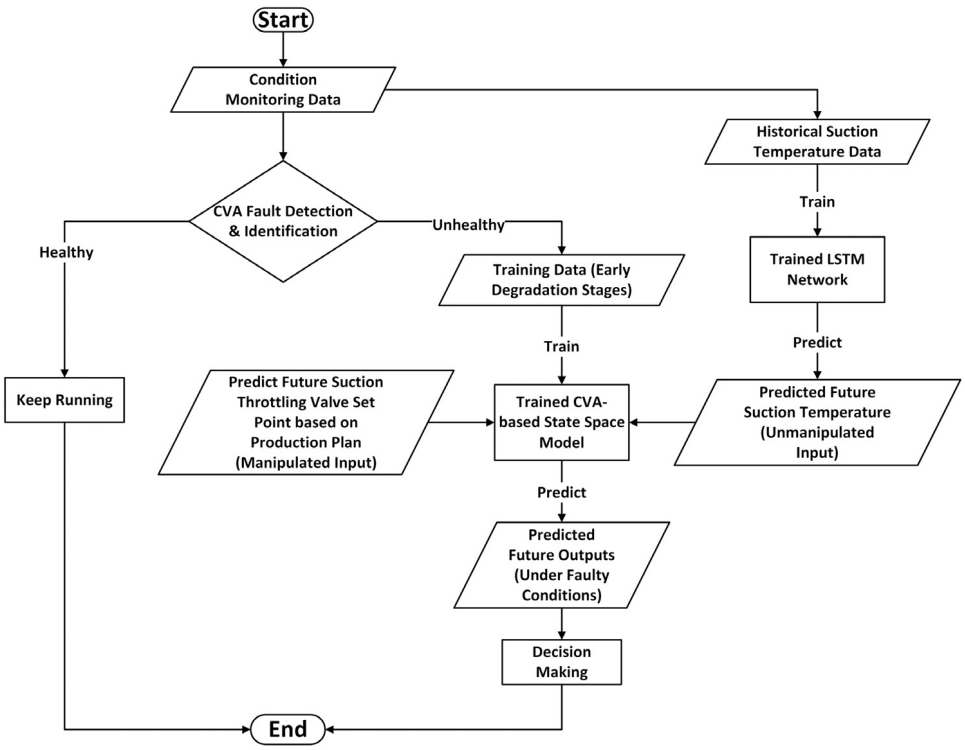

Fig. 5. Illustration of the proposed CVA–LSTM system for performance estimation of an industrial system under faulty operating conditions.

Table 1
Measured variables for compressor A.

ID	Variable name	ID	Variable name	ID	Variable name
1	Shaft speed	7	Suction pressure	13	Journal bearing 2 vibration
2	Flow meter	8	Discharge pressure	14	Journal bearing 3 vibration
3	Suction throttling valve	9	Actual flow	15	Journal bearing 4 vibration
4	Actual power	10	Total driver power	16	Thrust bearing vibration
5	Suction temperature	11	Actual molecular weight		
6	Discharge temperature	12	Journal bearing 1 vibration		

the measured signals from different sensors can be stored and accessed through an e-maintenance system and can be used for diagnostic and prognostic purposes.

In practice, fault data are typically scarce because very few safety-critical and expensive components are allowed to run to failure. If a fault occurs and the fault is not serious, it may be that machine operators choose to keep the machine running until repair facilities and spare parts are available at the plant. In such a case, the proposed CVA–LSTM method can be used to predict the impact of the fault on machine operation, plant safety and product quality before replacements and substitutes are available. The proposed method is applied to an operational industrial centrifugal compressor to predict the performance of the machine after a variable speed drive (VSD) fault occurs. This compressor is a one-cylinder, two-section, six-stage centrifugal compressor running at a large refinery in Europe (hereafter referred to as Compressor A). This compressor is driven by a variable-speed electric motor. For this study, all the data were captured at a sampling rate of one sample per hour by the machine's condition monitoring system. Table 1 summarizes all of the measured variables for this compressor.

As shown in Fig. 6, the compressor is in the healthy condition during the first 170 points of the time series. The SVD fault occurred at the 171th sampling point, and the fault affected many measured sensor signals and the most affected variables include the shaft speed and suction pressure (see Figs. 7 and 8). Since the fault severity was not very critical, this machine was kept running until the 316th sampling point. After that time, site engineers removed the malfunctioning VSD and replaced it with a newly arrived VSD.

3.2. Fault detection and identification using a CVA-based diagnostic model

To evaluate the performance of the proposed fault diagnostic method, the data captured during healthy operating conditions (hereafter referred to as T1), as shown in Fig. 6, were used as the training data for the CVA diagnostic model, and the data captured throughout the degradation process (hereafter referred to as T3, see Fig. 6) were used to validate the trained model. T1 includes 170 observations, and T3 contains 146 observations, which covers the entire fault degradation process. Fig. 6 depicts the trend of five specific performance variables in data sets T1 and T3 (all sensor measurements were normalized to provide a clear view of the trends for different variables in the training data set and the validation data set).

The numbers of time lags a and b were determined by calculating the autocorrelation function of the root summed squares of all variables in data set T1 against a confidence bound of ±5%. The autocorrelation function indicates how long the signal is correlated with itself and thus can be used to determine the maximum number of significant lags. As shown in Fig. 9, the sample autocorrelation analysis of the training data demonstrates that the maximum number of significant lags was 6. Therefore, the number of time lags a and b were set to 6 in this study.

According to the literature (Negiz & Çinar, 1998; Odiowei & Yi, 2010), different methods can be used to determine the number of states retained q, among which those based on the Akaike information criterion (AIC) and the dominant singular values (SVs) in the diagonal matrix $\sum$ are most commonly used. In this study, the optimum number of retained states was determined by considering the dominant SVs

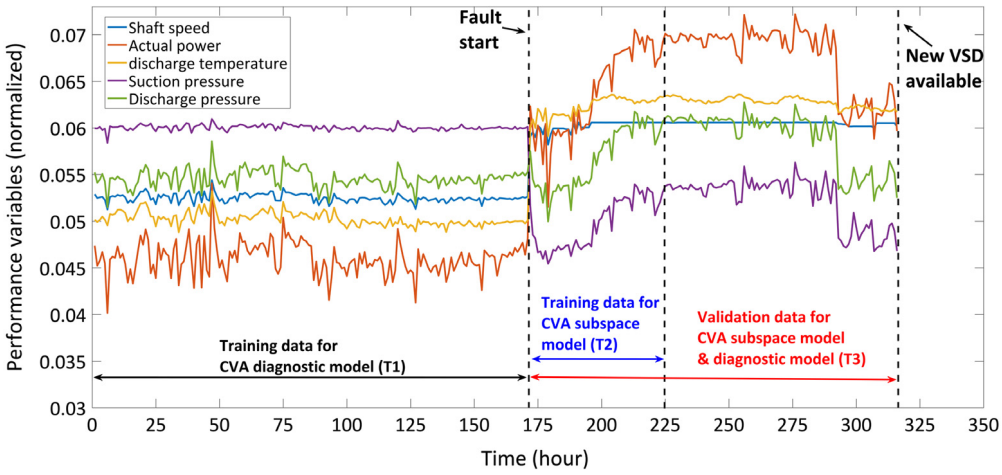

Fig. 6. Trend of five different performance variables (normalized) before and after the VSD failure.

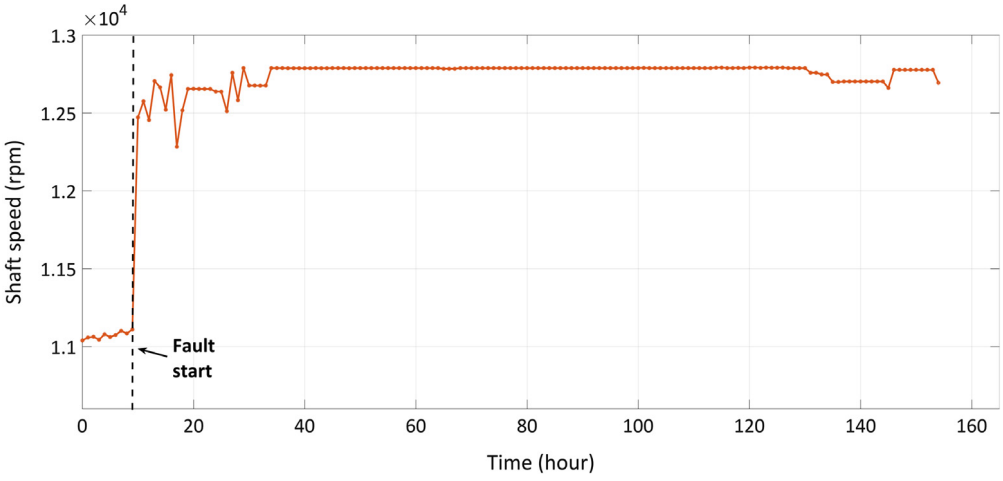

Fig. 7. Trend of the shaft speed after the VSD failure.

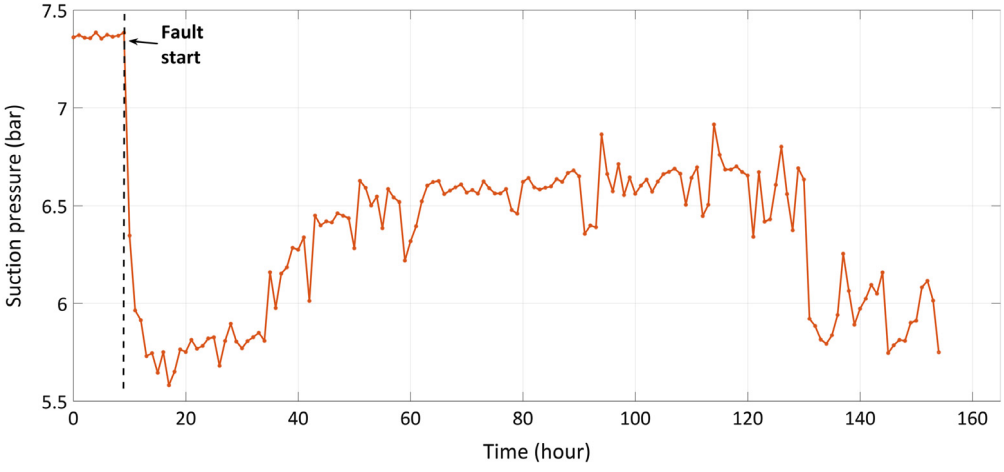

Fig. 8. Trend of the suction pressure after the VSD failure.

obtained from formula (5). As shown in Fig. 10, the SVs were placed in descending order with a gradually decreasing trend, and setting the number of q based on these SVs will lead to an unrealistic high-order system. Moreover, the performance of the fault detection in this study is not relevant to the value of q because both statistics (T^2 and Q) are used as detection metrics. In other words, a fault that fails to generate a significant deviation in the canonical state space will be captured by the residual space. To select the optimal number of retained state q that gives the lowest false alarm rate, a data set containing 887 observations under healthy operating conditions was used to test the trained CVA

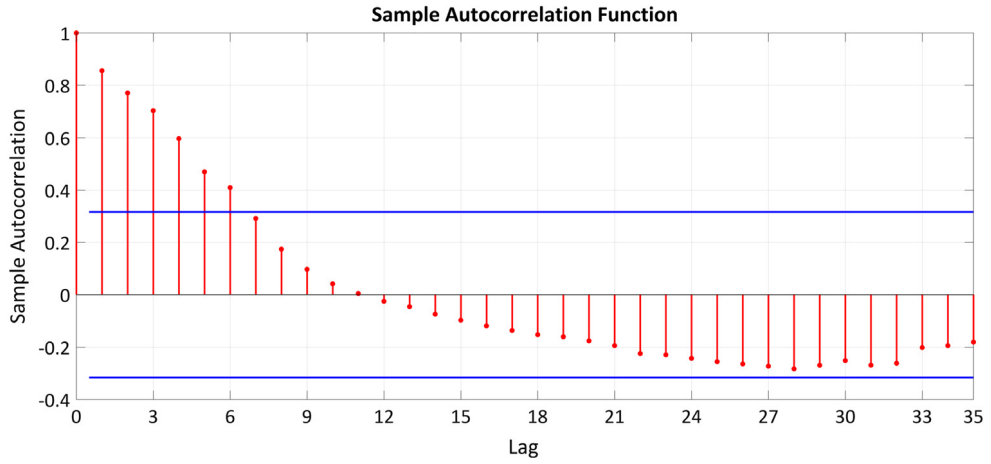

Fig. 9. Autocorrelation analysis of data set T1.

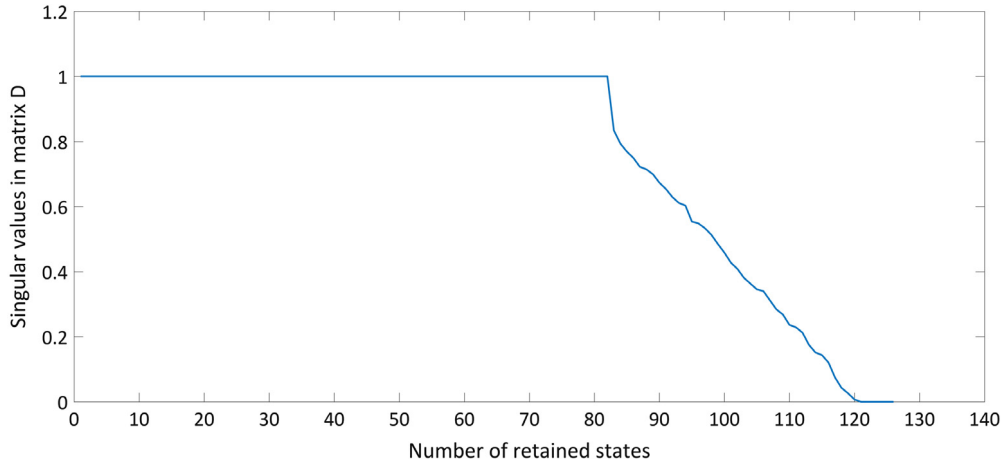

Fig. 10. Singular values in matrix D.

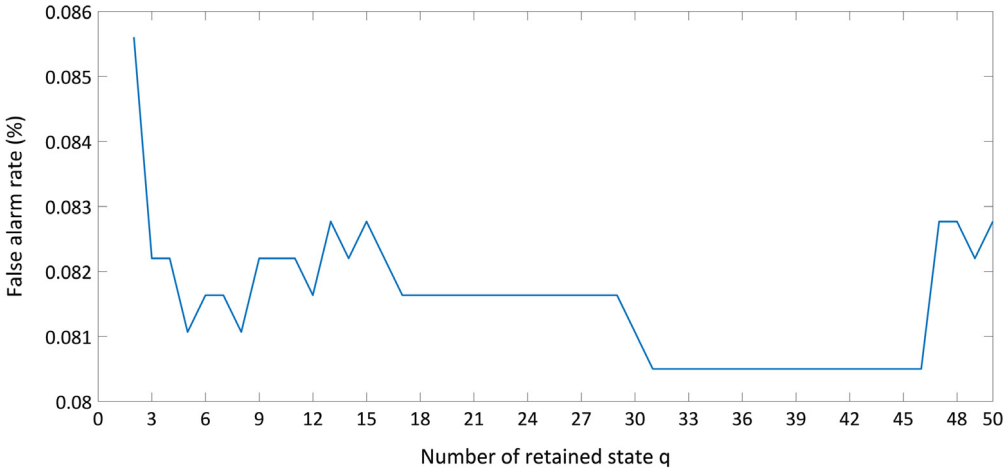

Fig. 11. False alarm rate for different values of the retained state q.

diagnostic model. A confidence bound of 99% was adopted for the calculation of the normal operating threshold during the testing process. The false alarm rate was calculated for different values of q. The false alarm rate in this study was calculated by dividing the number of false detections by the length of the testing data. Fig. 11 shows the false alarm rate against different values of q for the testing data set. For low values of system order q, the false alarm rate is high because the information in the retained space is not able to fully represent the system dynamics, which results in a large number of T^2 threshold violations. Meanwhile, to avoid the CVA model overfitting the training data, the value of q cannot be set too large, and $q = 31$ was finally adopted in this study to perform fault detection. Fig. 12 shows the T^2 and Q statistics of the

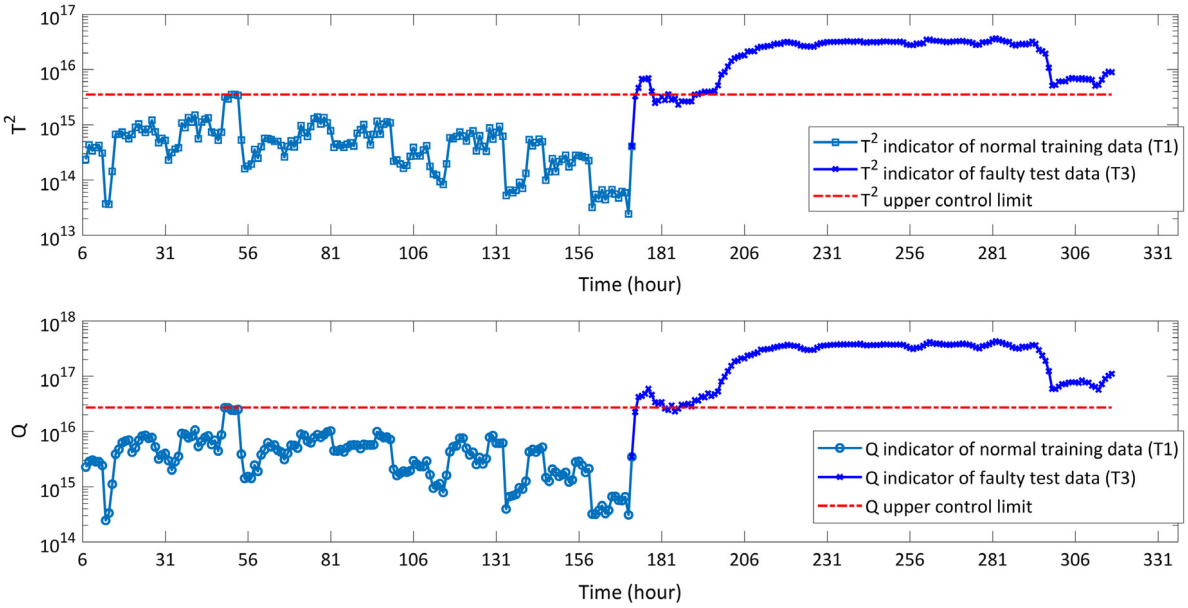

Fig. 12. The T^2 (upper) and Q (lower) statistics for training data set T1 and validation data set T3. Red dashed: thresholds for health indicators. Fault is detected after the 172th sampling point (one hour after the VSD failure has occurred).

training data set T1 and the validation data set T3 (all of the measured variables were used to perform fault detection in this study). The upper control limit for healthy operating conditions was calculated at the 99% confidence level. Both indicators detect the VSD fault at approximately the 172th sampling point, which is one hour after the fault has occurred. The false alarm rate for the T^2 and Q statistics is 9.59% and 4.11%, respectively, indicating that the Q statistic is more sensitive to the fault than the T^2 statistic.

Once a fault occurs in an industrial heavy-duty compressor, it is valuable to identify which components are most likely associated with the root cause of the malfunction. To identify the most fault-related variables for compressor A, a combined CVA-based 2-D contribution plot is displayed in Fig. 13, in which the variable name is the vertical axis and the sampling time is the horizontal axis. The contributions for the canonical (c_t^{state}) and the residual space ($c_t^{residual}$) were calculated as per Eqs. (10) and (11), and the combined contributions $c_t^{combined}$ used equal weights for c_t^{state} and $c_t^{residual}$ ($c_t^{combined} = 0.5 * c_t^{state} + 0.5 * c_t^{residual}$).

After the VSD failure, the malfunctioned VSD could not maintain relatively constant suction pressure in the compressor as it did before the failure occurred. Similarly, the malfunctioned VSD could not control the speed of the compressor as it did before the failure occurred. In other words, shaft speed and suction pressure will show large variations after a VSD failure; therefore, they can be used as two key indicators of a VSD failure for this compressor. As shown in Fig. 8, the suction pressure decreased from 7.384 bar to 5.582 bar and remained unstable until the faulty VSD was replaced; meanwhile, the shaft increased from 11,110 rpm to 12,790 rpm and remained at a higher speed than that of normal operating conditions (see Fig. 7). In Fig. 13, both the shaft speed and suction pressure are reported by the combined contribution plot in which they show consistently strong bands of contribution during the degradation process. This information can be used by experienced machine operators to find the root cause of the fault. It can be observed that aside from the shaft speed and suction pressure, the actual flow shows large contributions to the detected fault because the actual flow is calculated based on the suction pressure.

3.3. LSTM for suction temperature prediction

In addition to fault diagnostics and identification, site operators may be more interested in how the system will behave under faulty operating conditions given future compressor inputs and how the detected fault will affect the safety of plant operation, quality of product and power efficiency of the system. The future values of the manipulated system input variable (i.e., suction throttling valve position in this study) can be obtained by looking at the production schedule. However, the inlet gas temperature of the machine is largely determined by the ambient temperature when the gas passes through long transmission pipelines to the compressor and therefore is not a manipulated variable. To account for the impact of the ambient temperature on a system's performance under faulty operating conditions, LSTM was used in this study to forecast the magnitude of the future inlet gas temperature based on historical data.

With the aim of forecasting the values of suction temperature during the entire degradation process, the measured suction temperature signals before the failure time were used to train the proposed LSTM network. The data set for training contains 449 sampling points, as shown in Fig. 14. The predicted suction temperature during the entire deterioration process was plotted against the actual measured temperature in Fig. 14 as well. The predictive deviation at each time instance was plotted as vertical blue lines to show consistent over- and under-prediction, as well as a sense of the variance in the predictions. The procedures used to determine the LSTM network architecture as well as the optimal network parameters are detailed in the following subsections.

3.3.1. Determination of the LSTM network architecture

The proposed LSTM network consists of a sequence of 6 layers: an input layer, 3 LSTM hidden layers, a dense layer and an output layer. The LSTM hidden layers are used to model the relationships between the past and future time series signals. The layers also allow the network to represent more complex models than possible without the hidden layers. The dense layer is used to change the dimensions of the output vectors from the previous LSTM layer and map the outputs into a final predicted time sequence. To determine the number of LSTM hidden layers l_r, 10-fold cross-validation was performed on the training data (i.e., 449 suction temperature points measured before the VSD failure) and the mean square error (MSE) on the cross-validation set was calculated for different numbers of l_r. The results are summarized in Table 2, and $l_r = 3$ resulted in the lowest prediction error on the cross-validation set. The results indicate that the deeper network architecture helps the LSTM understand complicated relationships between the time series, which generates more accurate predictions. Meanwhile, the predictive

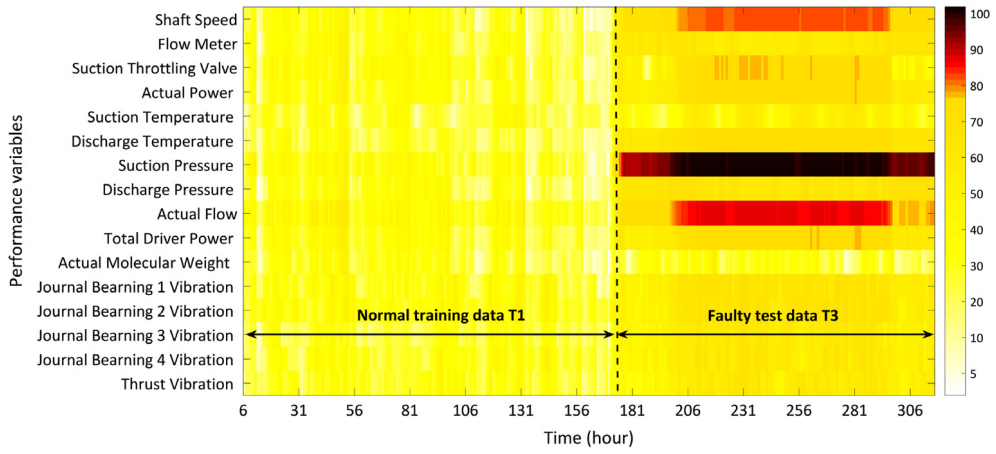

Fig. 13. Combined CVA-based 2-D contributions for identifying the VSD failure.

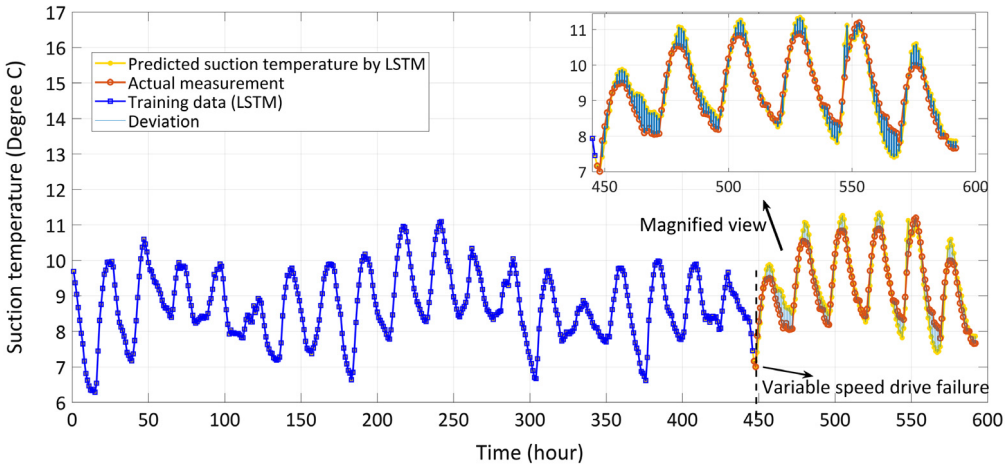

Fig. 14. Predicted suction temperature (yellow) against actual measure suction temperature (red) during the entire degradation process (length of the predicted time sequence is equal to the length of data set T3). The training data set for the proposed LSTM network (blue) is also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Cross-validation set performance (MSE, %) based on different numbers of LSTM hidden layers l_r.

Number of neurons	$l_r = 3$			$l_r = 2$			$l_r = 1$		
	Test1	Test2	Test3		Test1	Test2	Test3		Test1
128	0.461	0.482	0.563	128	0.461	0.482	0.563	128	0.461
256	0.385	0.477	0.473	256	0.385	0.477	0.473	256	0.385
384	0.475	0.504	0.494	384	0.475	0.504	0.494	384	0.475
512	0.538	0.435	0.495	512	0.538	0.435	0.495	512	0.538
640	0.440	0.577	0.481	640	0.440	0.577	0.481	640	0.440
768	0.598	0.511	0.504	768	0.598	0.511	0.504	768	0.598
896	0.540	0.539	0.548	896	0.540	0.539	0.548	896	0.540
1024	0.479	0.547	0.533	1024	0.479	0.547	0.533	1024	0.479
Average		**0.503**			0.507			0.617	

Best result is in bold. The number of epochs was set to 5, and the number of units (neurons) in each LSTM layer was set to 256 during the experiments. Cross-validation was repeated 3 times for each number of LSTM hidden layers, and the averaged MSE of 3 replicates was used to determine the network's final structure. The number of LSTM layers was chosen in the range of $\{1, 2, 3\}$ because the computational cost of a larger number of l_r will become huge.

accuracy was not obviously improved when the number of hidden layers was increased from 2 to 3. In order to achieve the lowest prediction error, $l_r = 3$ was adopted in this study for suction temperature prediction.

3.3.2. Determination of the LSTM network parameters

In this section, the number of neurons in the LSTM layers $N_{neurons}$ and batch size B were obtained by 10-fold cross-validation to obtain the optimal LSTM model. In this investigation, $N_{neurons}$ was studied in the range of $N_{neurons} = \{128, 256, 384, \ldots, 1024\}$ and B was studied in the range of $B = \{128, 256, 384, \ldots, 1024\}$. The results of cross-validation are shown in Table 3. The best testing performance was $MSE = 0.424\%$ with $N_{neurons} = 384$ and $B = 896$. Therefore, $l_r = 3$, $N_{neurons} = 384$ and $B = 896$ were adopted to predict the suction temperature of the compressor for the entire degradation process. To avoid overfitting, the number of epochs was set to 50 for the training process. The predicted suction temperature is shown in Fig. 14.

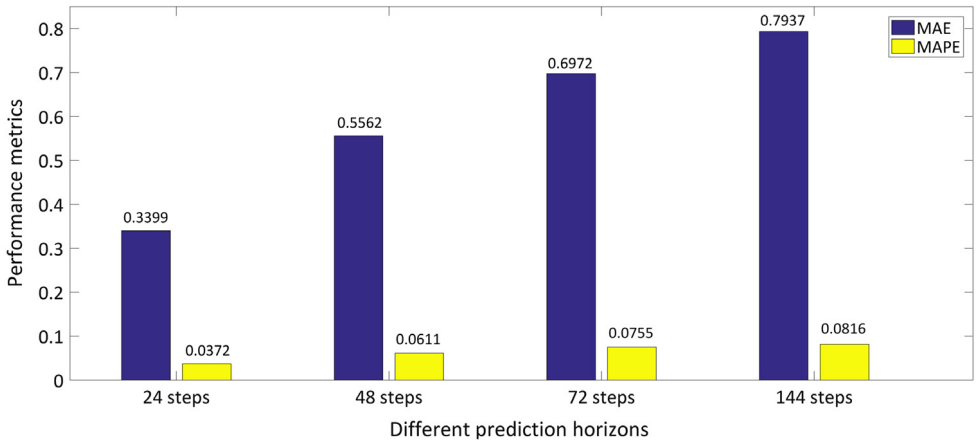

Fig. 15. Comparison of different forecasting horizons.

Table 3
Cross-validation set performance (MSE, %) based on different values of batch size B and number of neurons $N_{neurons}$ in the hidden layers.

Number of neurons	Batch size							
	128	256	384	512	640	768	896	1024
128	0.483	0.486	0.469	0.516	0.538	0.526	0.519	0.467
256	0.460	0.494	0.487	0.492	0.548	0.506	0.499	0.517
384	0.512	0.503	0.434	0.471	0.494	0.437	**0.424**	0.520
512	0.461	0.425	0.494	0.459	0.494	0.516	0.468	0.472
640	0.468	0.494	0.496	0.434	0.486	0.500	0.511	0.579
768	0.501	0.482	0.472	0.548	0.570	0.547	0.557	0.526
896	0.475	0.443	0.489	0.537	0.457	0.539	0.527	0.481
1024	0.435	0.485	0.466	0.445	0.523	0.528	0.511	0.590

Best result is in bold. The results are based on the configuration of 3 LSTM hidden layers and 1 fully connected dense layer. The number of epochs was set to 5, and each value represents the mean of three replicates.

3.3.3. Comparison of different forecasting horizons

Following the steps described in the previous subsections, the LSTM network was used to predict the suction temperature throughout the entire degradation process with different prediction horizons (24, 48, 72 and 144 time-steps). The mean absolute error (MAE) and mean absolute percentage error (MAPE) were computed for different forecasting horizons (see Fig. 15). The results show that the predictive accuracy decreases as the forecasting horizon increases because a longer prediction horizon will result in fewer available training samples for an LSTM network and therefore lower forecast accuracy given the limited historical data. To achieve the highest predictive accuracy, the prediction horizon was set to 24 h in this study.

3.4. Combining CVA and LSTM for performance estimation

Following the steps described in Section 2.4, the training data set T2, which was obtained during the early stages of deterioration, was used to build and train a CVA-based state space model as per Eqs. (12) and (14). Then, the trained model was fed with two input sequences: 1. suction temperature predicted by the LSTM network (i.e., the predicted time series shown in Fig. 14); and 2. Actual suction throttling valve position recorded during the entire degradation process (i.e., the length of this sequence is equal to that of the validation data set T3) to provide estimations of the system outputs for data set T3. Finally, the predicted outputs were compared with the actual system outputs measured in data set T3 to evaluate the performance of the model.

3.4.1. Determination of the CVA-based state space model parameters

Similar to the procedure described in Section 3.2, the number of time lags a and b were calculated based on the autocorrelation analysis of the training data set T2, and a and b were set to 6. Then, T2 was used to build and train a CVA subspace model. To determine the optimal number of trained retained states q, the trained CVA model was first used to predict system outputs for data set T2. The purpose of this validation process was to find the optimal value of q that minimizes the predictive error for T2. Two performance metrics—1) mean absolute error (MAE) and 2) mean absolute percentage error (MAPE)—were used in this study to evaluate the performance of the predictive model. Interested readers are referred to Saxena et al. (2008) for further information about the two metrics. The summed MAE and MAPE of all the output variables in T2 were plotted against different numbers of retained states q in Fig. 16. The errors increase as the system order becomes larger, and q was finally set to 1 to obtain the optimal model that provides the highest predictive accuracy.

3.4.2. Results and discussion

Following the steps described in Section 2.4, data set T2 was first used to build and train a CVA-based state space model as per Eqs. (12) and (14). Then, the trained model was fed with two input sequences: 1. suction temperature predicted by the LSTM network; and 2. actual suction throttling valve set points used throughout the degradation process to provide estimations of the system outputs for data set T3. Figs. 17–21 show the prediction results for some of the most significant variables of T3. A large deviation from the actual measurements at the beginning of each prediction was observed. This error was caused by an inaccurate initial state estimation, for which the model can rapidly compensate. Other than the inaccurate initial predictions at the very beginning of each time sequence, the estimated values are close to the actual measurements, indicating that the developed model is able to accurately predict the system behavior under faulty operating conditions based on the expected future suction valve set points and predicted future suction temperature. Table 4 shows the precision analysis for the prediction error of the proposed CVA–LSTM model. The results demonstrated the superior performance of the proposed

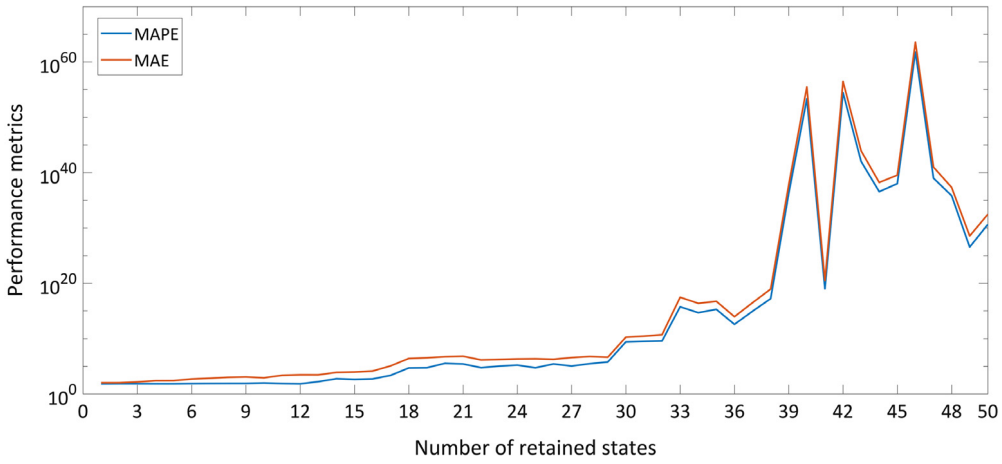

Fig. 16. Summed prediction error (MAE and MAPE) for all output variables for different values of q.

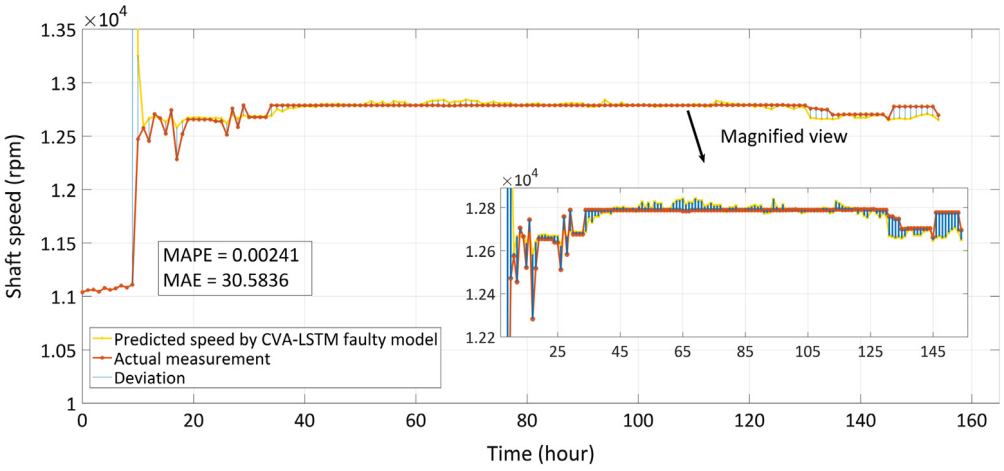

Fig. 17. Shaft speed under faulty operating conditions predicted by the CVA–LSTM model. The predicted shaft speed can be used to estimate the extra power cost caused by the VSD failure.

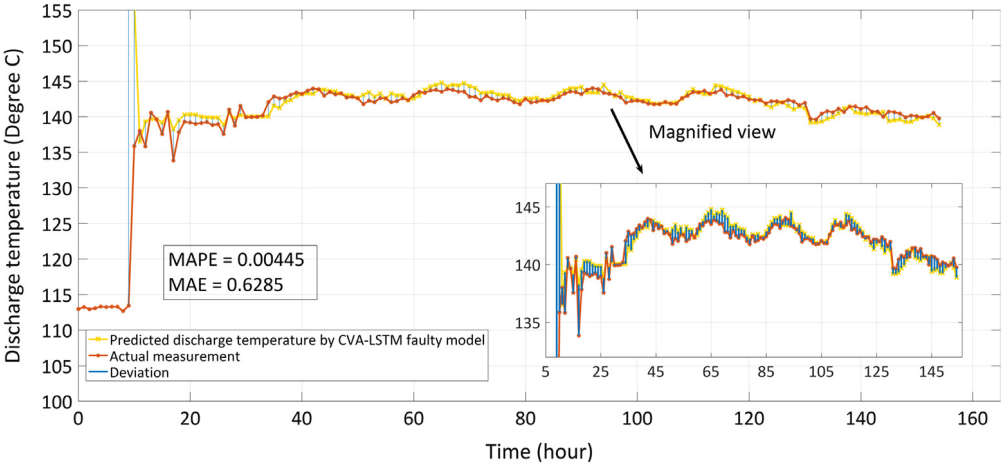

Fig. 18. Discharge temperature under faulty operating conditions predicted by the CVA–LSTM model. The predicted discharge temperature can be used to estimate the impact of the VSD failure on product quality.

model and proved that the model can be used to accurately predict the system behavior under slowly evolving faulty conditions. Site engineers could use the predicted key performance variables to plan maintenance with the aim of minimizing plant power consumption and unexpected breakdown costs as well as maximizing product quality and the safety and reliability of the plant.

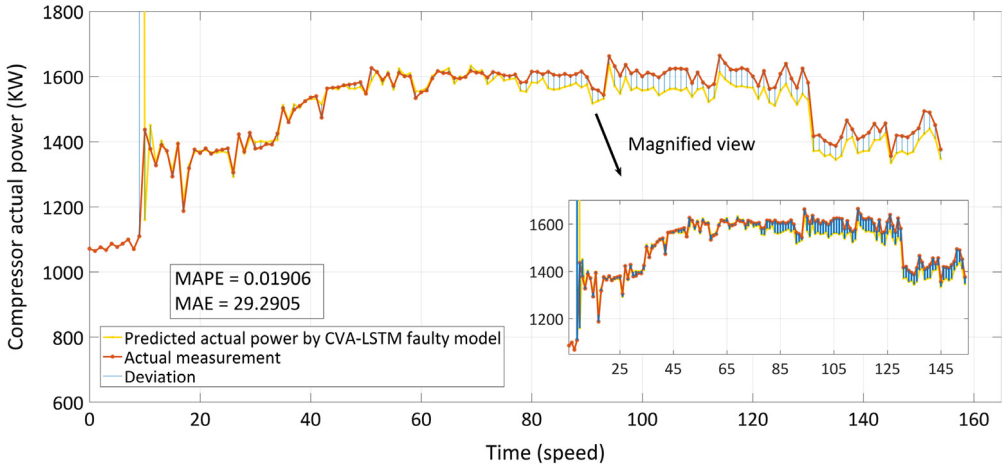

Fig. 19. Actual power under faulty operating conditions predicted by the CVA–LSTM model. The predicted actual power can be used to estimate the extra power cost caused by the VSD failure.

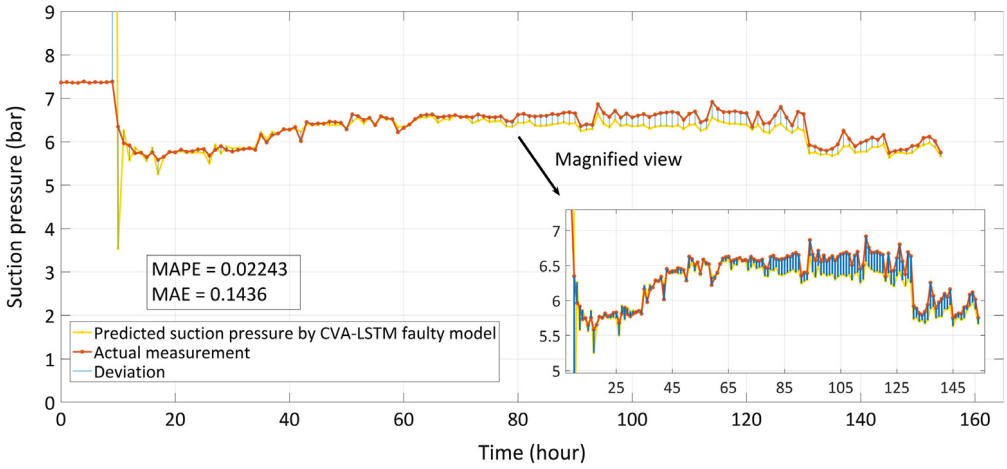

Fig. 20. Suction pressure under faulty operating conditions predicted by the CVA–LSTM model. The predicted suction pressure is related to plant safety since a low-pressure alarm will be triggered when the suction pressure is lower than 6 bar for this compressor.

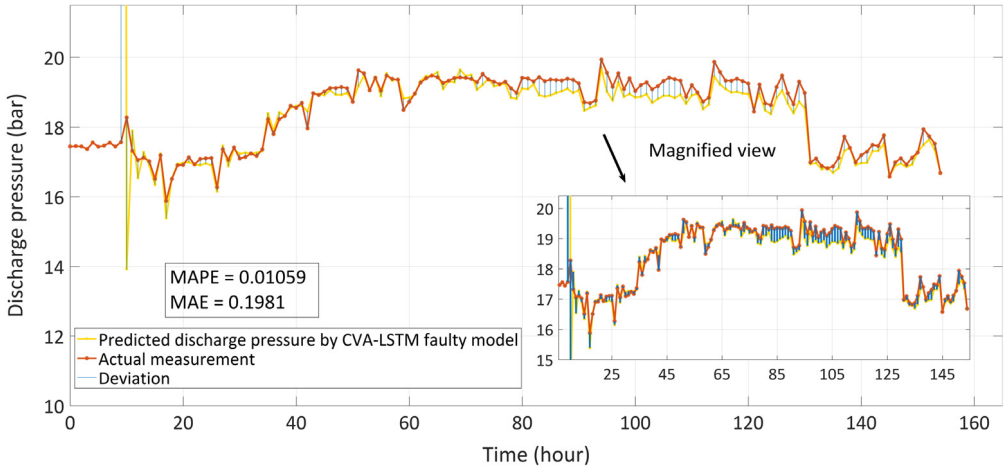

Fig. 21. Discharge pressure under faulty operating conditions predicted by the CVA–LSTM model. The predicted discharge pressure can be used to estimate the impact of the VSD failure on product quality.

Table 4
Prediction error for different performance variables.

Variable ID	1	2	4	6	7	8	9
MAPE	0.00241	0.01168	0.01906	0.00445	0.02243	0.01059	0.01202
MAE	30.5836	0.0068	29.2905	0.6285	0.1436	0.1981	37.6962
Variable ID	10	11	12	13	14	15	16
MAPE	0.01888	0.00311	30.22971	4.56105	0.33150	0.29680	0.40970
MAE	29.0465	0.5454	4.0882	4.2302	4.2503	4.5429	4.1582

4. Conclusion

In this study, condition monitoring data acquired from an operational industrial centrifugal compressor were used to test the capabilities of CVA for fault detection and identification. In addition, CVA combined with LSTM were applied for the first time to predict the behavior of the system under faulty operating conditions for the expected future suction throttling valve set points. The VSD failure in data set T3 was successfully detected by T^2 and Q health indicators within a short detection time. Contribution plots based on both the canonical state space and the residual space information were utilized to identify the root cause of the failure. Condition monitoring data obtained from an incipient fault were used to train a CVA subspace model. The trained CVA model was then used to provide estimations of the system's performance after the fault had occurred for the specified future suction throttling valve set points. Although large oscillations were observed in the initial estimations, the average prediction error was low, which proves that the model is able to represent the system dynamics under faulty conditions. The predicted key performance variables can be used to plan maintenance with the aim of minimizing plant power consumption and maximizing product quality and the safety and reliability of the plant. The proposed model is based on the assumption that the faults evolve slowly enough and the correlation between system inputs and outputs will not change rapidly during the entire degradation process so that the estimation routine can track them properly. In addition, this study focuses on the future prediction of a single manifestation of the input time series, which can create difficulties when managing prediction uncertainties. The use of a probabilistic model can overcome this limitation, thereby allowing uncertainty to be incorporated into the results (the reader is invited to refer to Olivier and Craig, 2017). Efforts will be made in future research to quantify the influence of uncertainty in the disturbance prediction on the performance estimations.

The combined method takes into account the impact of the ambient temperature on a system's performance under faulty operating conditions and thereby allows both the environmental factors and the human interventions to be factored in when predicting the system's future behavior. The information provided by the proposed method can be used by site engineers to estimate the impact of a fault on the operational process and to develop appropriate production plans and optimal maintenance strategies, which makes the plant operations more safe, productive and profitable.

References

Ai, B., Zheng, Y., Jang, S. S., Wang, Y., Ye, L., & Zhou, C. (2009). The optimal drift-compensatory and fault tolerant approach for mixed-product run-to-run control. *Journal of Process Control*, *19*(8), 1401–1412. http://dx.doi.org/10.1016/j.jprocont.2009.04.008.

Ai, B., Zheng, Y., Wang, Y., Jang, S. S., & Song, T. (2010). Cycle forecasting EWMA (CF-EWMA) approach for drift and fault in mixed-product run-to-run process. *Journal of Process Control*, *20*(5), 689–708. http://dx.doi.org/10.1016/j.jprocont.2010.03.004.

Bengio, Y., Simard, P., Frasconi, P., & Member, S. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157–166.

Campanari, S. (2000). Full load and part-load performance prediction for integrated SOFC and microturbine systems. *Journal of Engineering for Gas Turbines and Power*, *122*, 239–246.

Cárcel, C. R., Cao, Y., & Mba, D. (2007). A benchmark of canonical variate analysis for fault detection and diagnosis. In *2014 UKACC international conference on control 9th–11th July 2014* (pp. 329–356). Loughborough U.K. https://doi.org/10.1016/S0167-4137(07)80012-6.

Chen, J., & Wang, D. (2016). Long short-term memory for speaker generalization in supervised speech separation. In *INTERSPEECH 2016* (pp. 3314–3318). San Francisco.

Chouikhi, N., Ammar, B., Rokbani, N., & Alimi, A. M. (2017). PSO-based analysis of Echo State Network parameters for time series forecasting. *Applied Soft Computing Journal*, *55*, 211–225. http://dx.doi.org/10.1016/j.asoc.2017.01.049.

Cleveland, R. B., Cleveland, W. S., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, *6*(1), 3–73.

Guerra, C. J., & Kolodziej, J. R. (2017). A data-driven approach for condition monitoring of reciprocating compressor valves. *Journal of Engineering for Gas Turbines and Power*, *136*(1), 1–13. http://dx.doi.org/10.1115/1.4025944.

Harrou, F., Nounou, M. N., Nounou, H. N., & Madakyaru, M. (2013). Statistical fault detection using PCA-based GLR hypothesis testing. *Journal of Loss Prevention in the Process Industries*, *26*. http://dx.doi.org/10.1016/j.jlp.2012.10.003.

He, D., Li, R., & Bechhoefer, E. (2012). Stochastic modeling of damage physics for mechanical component prognostics using condition indicators. *Journal of Intelligent Manufacturing*, *23*, 221–226. http://dx.doi.org/10.1007/s10845-009-0348-9.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, *2*(3/4), 321–377.

Huang, L., Cao, Y., Tian, X., & Deng, X. (2015). A nonlinear quality-relevant process monitoring method with Kernel input–output canonical variate analysis. In *IFAC proceedings volumes (IFAC-PapersOnline), Vol. 48* (pp. 611–616). Elsevier Ltd. http://dx.doi.org/10.1016/j.ifacol.20150.9.035.

Hyvärinen, A., Karhunen, J., & Oja, E. (2004). *Independent component analysis*. John Wiley & Sons.

Jiang, B., Huang, D., Zhu, X., Yang, F., & Braatz, R. D. (2015). Canonical variate analysis-based contributions for fault identification. *Journal of Process Control*, *26*, 17–25. http://dx.doi.org/10.1016/j.jprocont.2014.12.001.

Koehler, A. B., Snyder, R. D., & Ord, J. K. (2001). Forecasting models and prediction intervals for the multiplicative Holt–Winters method. *International Journal of Forecasting*, *17*, 269–286.

Kruger, U., & Dimitriadis, G. (2008). Diagnosis of process faults in chemical systems using a local partial least squares approach. *AIChE Journal*, *54*(10), 2581–2596.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444. http://dx.doi.org/10.1038/nature14539.

Li, W., & Qin, S. J. (2001). Consistent dynamic PCA based on errors-in-variables subspace identification. *Journal of Process Control*, *11*(6), 661–678.

Mohammad, Y., & Nishida, T. (2016). *Data mining for social robotics: Toward autonomously social robots*. Springer. http://dx.doi.org/10.1007/978-3-319-25232-2.

Negiz, A., & Çinar, A. (1998). Monitoring of multivariable dynamic processes and sensor auditing. *Journal of Process Control*, *8*(5–6), 375–380. http://dx.doi.org/10.1016/S0959-1524(98)00006-7.

Odiowei, P. E. P., & Yi, C. (2010). Nonlinear dynamic process monitoring using canonical variate analysis and Kernel density estimations. *IEEE Transactions on Industrial Informatics*, *6*(1), 36–45. http://dx.doi.org/10.1109/TII.2009.2032654.

Olivier, L. E., & Craig, I. K. (2017). Should I shut down my processing plant? An analysis in the presence of faults. *Journal of Process Control*, *56*, 35–47. http://dx.doi.org/10.1016/j.jprocont.2017.05.005.

Russell, E. L., Chiang, L. H., & Braatz, R. D. (2000). Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *51*(1), 81–93. http://dx.doi.org/10.1016/S0169-7439(00)00058-7.

Samuel, R. T., & Cao, Y. (2015). Kernel canonical variate analysis for nonlinear dynamic process monitoring. In *9th international symposium on advanced control of chemical processes, Vol. 48* (pp. 605–610). https://doi.org/10.1016/j.ifacol.2015.09.034.

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., & Saha, S. et al., (2008). Metrics for evaluating performance of prognostic techniques. In *International conference on prognostics and health management, 2008. PHM 2008* (pp. 1–17). https://doi.org/10.1109/PHM.2008.4711436.

Serdio, F., Lughofer, E., Pichler, K., Buchegger, T., & Pichler, M. (2014). Fault detection in multi-sensor networks based on multivariate time-series models and orthogonal transformations. *Information Fusion*, *20*, 272–291. http://dx.doi.org/10.1016/j.inffus.2014.03.006.

Son, H. (2017). Toward a proposed framework for mood recognition using LSTM Recurrent Neuron Network. In *Procedia computer science, Vol. 109* (pp. 1028–1034). Elsevier B.V. http://dx.doi.org/10.1016/j.procs.2017.05.378.

Stefatos, G., & Hamza, A. Ben. (2010). Dynamic independent component analysis approach for fault detection and diagnosis. *Expert Systems with Applications*, *37*(12), 8606–8617. http://dx.doi.org/10.1016/j.eswa.2010.06.101.

Stubbs, S., Zhang, J., & Morris, J. (2012). Fault detection in dynamic processes using a simplified monitoring-specific CVA state space modelling approach. *Computers and Chemical Engineering*, *41*, 77–87. http://dx.doi.org/10.1016/j.compchemeng.2012.02.009.

Tran, V. T., Althobiani, F., & Ball, A. (2014). An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks. *Expert Systems with Applications*, *41*(9), 4113–4122. http://dx.doi.org/10.1016/j.eswa.2013.12.026.

Wootton, A. J., Butcher, J. B., Kyriacou, T., Day, C. R., & Haycock, P. W. (2017). Engineering Applications of Artificial Intelligence Structural health monitoring of a footbridge using Echo State Networks and NARMAX. *Engineering Applications of Artificial Intelligence*, *64*(November 2016), 152–163. http://dx.doi.org/10.1016/j.engappai.2017.05.014.

Wu, Y., Yuan, M., Dong, S., Lin, L., & Liu, Y. (2017). Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing*, *275*, 167–179. http://dx.doi.org/10.1016/j.neucom.2017.05.063.

Yin, S., Zhu, X., Member, S., & Kaynak, O. (2015). Improved PLS focused on key-performance- indicator-related fault diagnosis. *IEEE Transactions on Industrial Electronics*, *62*(3), 1651–1658.

Zhang, X., Wang, J., & Zhang, K. (2017). Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. *Electric Power Systems Research*, *146*, 270–285. http://dx.doi.org/10.1016/j.epsr.2017.01.035.

Zheng, Y., Ai, B., Wong, D. S. H., Jang, S. S., Wang, Y., & Zhang, J. (2010). An EWMA algorithm with a cycled resetting (CR) discount factor for drift and fault of high-mix run-to-run control. *IEEE Transactions on Industrial Informatics*, *6*(2), 229–242. http://dx.doi.org/10.1109/TII.2009.2039904.