International Journal of Performability Engineering Vol. 4, No. 2, April 2008, pp. 183-192.

Printed in India

Fault Detection using DPCA-based Control Charts

JIANMOU WU AND VILIAM MAKIS*

Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, Canada M5S 3G8

(Received on May 24, 2007)

Abstract: In this paper, Hotelling's T^2 chart and the DPCA(Dynamic PCA)-based T_A^2 and Q charts are applied to real oil data histories obtained from condition monitoring of heavy hauler truck transmissions. The objective is to evaluate the impending failure detection capability and the false alarm rate of the control charts in a CBM (condition-based maintenance) application. When applied to failure histories, both T^2 and Q charts were able to correctly identify the impending failures in more than 50% cases, but the number of false alarms was considerably higher for the T^2 chart.

Key Words: *Practice of OR, Maintenance, Replacement Policy, Statistics, Time Series*

1. Introduction

This paper considers the application of multivariate SPC charts for failure prevention and condition-based maintenance (CBM) decision-making which has not been done previously in the literature. A CBM policy is based on monitoring equipment condition (e.g. machine vibration monitoring or spectrometric analysis of oil samples) and making maintenance decisions based on the partial information obtained from the observation process. Several CBM models have appeared in maintenance literature, such as a proportional hazards model [1], a counting-process model [2], a state-space model [3], and a hidden Markov model [4], among others.

The advantage of the approach in this paper when compared with the previously developed CBM models is relative simplicity of multivariate charting methods and easy implementation in industrial practice. The PCA-based (T_A^2, Q) control charts are very useful for monitoring multivariable industrial processes, but they cannot be directly applied in the maintenance area because PCA assumes independence of successive samples, whereas maintenance data typically exhibit both cross and auto-correlation. Dynamic PCA, an extension of PCA, can be successfully applied to such data and therefore a DPCA version of the combined T_A^2 and Q charts would be appropriate for a maintenance application. Recent applications of DPCA to oil data modeling and development of a proportional hazards based CBM model can be found in [5].

*Corresponding author Email: makis@mie.utoronto.ca

2. Hotelling's T^2 chart and PCA-based (T_A^2, Q) charts

The best-known control charts for monitoring multivariate processes are Hotelling's χ^2 and T^2 charts [6]. Assume that when the process is in control, a *k*-dimensional vector Y_t of measurements has a multivariate normal distribution with mean vector μ and covariance matrix Σ. The following statistic

$$\chi_t^{\ 2} = (Y_t - \mu)'\Sigma^{-1}(Y_t - \mu) \quad , \tag{1}$$

is plotted on the chart for which the upper control limit (UCL) is $\chi^2_{\alpha,k}$, where α is the selected probability of false alarm.

When the in-control mean vector μ and covariance matrix Σ are unknown and must be estimated from a sample of data, the Hotelling's T_t^2 statistic is used:

$$T_t^{\ 2} = (Y_t - \hat{\mu})' S^{-1}(Y_t - \hat{\mu}) \quad , \tag{2}$$

where $\hat{\mu}$ and S are the estimates of the process mean μ and the covariance matrix Σ, respectively. The upper control limit T^2_{UCL} is obtained from the *F* distribution with k and $N-k$ degrees of freedom, where N is the size of the sample used to estimate μ and Σ,

$$T^2_{UCL} = \frac{(N^2-1)k}{N(N-k)} F_\alpha(k, N-k) \quad , \tag{3}$$

Principal component analysis (PCA) is a linear transformation method that can obtain a set of uncorrelated variables, termed principal components (PCs), from the original set of variables. When the PCA is used to characterize the multivariable observation process, Hotelling's T^2 can also be expressed in terms of PCs:

$$T_t^2 = \sum_{a=1}^{k} \frac{z_{t,a}^2}{l_a} \quad , \tag{4}$$

where $z_{t,a}$ are the PC scores from the principal component transformation and l_a, $a = 1,2,\cdots,k$, are the eigenvalues of the correlation matrix estimate of the original data.

Furthermore, the PC score vector Z_t can be expressed as

$$Z_t = (z_{t,1}, z_{t,2}, \cdots z_{t,k})' = U' O_t = (u_1, u_2, \cdots u_k)' \cdot O_t \quad , \tag{5}$$

where $u_1, u_2, \cdots, u_k$ are the eigenvectors of the correlation matrix estimate of the original data set Y_t and the vector O_t is the standardized vector obtained from Y_t, $O_t = (\frac{Y_{t,1} - \overline{Y}_1}{s_1}, \frac{Y_{t,2} - \overline{Y}_2}{s_2}, \cdots, \frac{Y_{t,k} - \overline{Y}_k}{s_k})'$, with $\overline{Y}_i$ and s_i denoting the sample mean and standard deviation for variable $i, (i = 1, \cdots, k)$.

When PCA is used to reduce dimensionality, the T^2 chart based on the first A selected principal components is constructed as follows:

$$T_{A,t}^2 = \sum_{a=1}^{A} \frac{z_{t,a}^2}{l_a} = \sum_{a=1}^{A} \frac{(u_a' \cdot O_t)^2}{l_a} , \tag{6}$$

Then, based on the above formula, we can rewrite the T_t^2 statistic as

$$T_t^2 = \sum_{a=1}^{A} \frac{z_{t,a}^2}{l_a} + \sum_{a=A+1}^{k} \frac{z_{t,a}^2}{l_a^2} = T_{A,t}^2 + \sum_{a=A+1}^{k} \frac{z_{t,a}^2}{l_a^2} \quad (7)$$

The T_A^2 statistic based on the first *A* uncorrelated PCs provides a test for deviations in the condition monitoring variables that contribute most to the variability in the original data set Y_t. The upper control limit of T_A^2 can be calculated from formula (3) with k replaced by A.

However, only monitoring the process via T_A^2 is not sufficient. This method can only detect whether the variation in the condition monitoring variables in the space defined by the first *A* principal components exceeds UCL or not. In case a totally new type of special event occurs, which can cause the machine failure and was not presented when developing the in-control PCA model, then new PC's will appear and the new observation will move off the space defined by the in-control PCA model. Such new events can be detected by monitoring the squared prediction error (SPE) of the observation residuals, which gives a measure of how close an observation is to the *A*-dimensional space defined by the in-control PCA model:

$$SPE_t = \sum_{i=1}^{k} (O_{t,i} - \hat{O}_{t,i})^2 \quad (8)$$

where $\hat{O}_t = \sum_{a=1}^{A} Z_{t,a} \cdot u_a$ is computed from the in-control PCA model.

The SPE statistic is also referred to as a Q statistic [7]. The upper control limit for the Q statistic can be computed using approximate results for the distribution of quadratic forms [7]. With significance level α the UCL can be computed from the following formula:

$$Q_\alpha = \theta_1 [1 - \theta_2 h_0 (1 - h_0) / \theta_1^2 + z_\alpha (2\theta_2 h_0^2)^{1/2} / \theta_1]^{1/h_0}, \quad (9)$$

where z_α is the $100(1-\alpha)$ normal percentile, $\theta_i = \sum_{j=A+1}^{k} l_j^i, i = 1,2,3,$ and $h_0 = 1 - 2\theta_1\theta_3 / 3\theta_2^2$.

A T_A^2 chart based on the *A* dominant orthogonal PCs plus a Q chart is an effective set of multivariate SPC charts. On one hand, because the T_A^2 statistic is not affected by the smaller eigenvalues of the correlation matrix, it provides a more robust fault detection measure. T_A^2 can be interpreted as measuring the systematic variation in the PCA subspace, and a large value of T_A^2 exceeding the threshold would indicate that the systematic variation is out of control. On the other hand, the T^2 chart is overly sensitive in the PCA space because it includes the scores corresponding to small eigenvalues representing noise which may contribute significantly to the value of the T^2 statistic. The *Q* (SPE) statistic represents the squared perpendicular distance of a new multivariate observation from the PCA subspace and it is capable of detecting a change in the correlation structure when a special event occurs.

3. Selection of the in-control portion of the oil data

The data used in this paper are histories of the diagnostic oil data obtained from the 240-ton heavy hauler truck transmissions. Oil samples are taken roughly every 600 hours and the results of the spectrometric analysis consisting of the measurements of 20 metal elements in *ppm* are recorded. The total number of oil data histories considered is 51, 20 of them ended with a failure and the remaining 31 were suspended. A preliminary analysis by using EXAKT software (http://www.mie.utoronto.ca/labs/cbm) and also the results obtained previously indicate that it is sufficient to consider only 6 out of the total of 20 metal elements, namely potassium, iron, aluminum, magnesium, molybdenum and vanadium for maintenance decision-making.

Since we consider a set of 6-dimensional data and the in-control covariance matrix is unknown, we decided to calculate and plot Hotelling's T^2 statistic in order to determine the portions of the histories when transmission was in a healthy state. In a later section of this paper, the fault detection capability of the T^2 chart is compared with the Q chart. Before plotting the T^2 statistic and selecting the in-control portion of the oil data, it was necessary to pre-process the raw oil data. The Hotelling's T^2 statistic was calculated and plotted and the in-control portion of the data was identified.

The sample mean and covariance matrix were estimated from the data and the T_t^2 values were calculated using formula (2). For the upper control limit T_{UCL}^2, we set *k*=6 and *N*=527 in formula (3) because our oil data is 6-dimensional and we have totally 527 observations in the remaining 36 histories. For α =0.025, we found $F_\alpha(k, N-k) = 2.4325$ and $T_{UCL}^2 = 14.763$.

Values of T_t^2 were then plotted on the T^2 chart and the portions of histories when transmission was in a healthy state were determined. The following three working states of transmissions were considered: initial state (new transmission), healthy state and deteriorating state. Also, we note that the transmission oil was changed every 1200 hours, which dramatically affected the cumulative increasing trend of the T_t^2 series. We have decided to apply the following criteria to find out the in-control portion of the oil data. When the T_t^2 value exceeded the upper control limit T_{UCL}^2 during the first three observations of a history, we assumed, after consulting with maintenance engineers, that the transmission was in the initial run-in state before reaching the normal operating state. We excluded all the initial records in the histories up to and including the ones that exceeded the upper control limit. On the other hand, when an abnormal value of T_t^2 appeared later in the history ($t \geq 6$), it was assumed that the transmission reached a deteriorating state and all the following observations were excluded. The remaining observations are the portions of the histories when transmission was in a healthy state. Following this procedure, we selected 409 out of 527 records as the healthy portion of the oil data histories.

4. Time series modeling of the oil data in the healthy state

Using the in-control portion of the oil data, we can build a stationary AR model to describe the observation process.

By extending the Yule-Walker estimator formulae and the Wald statistic formula to a multi-history case and applying them to the in-control portion of oil data, we have the following modeling results. For *m=2,3,4,5* the order test Wald statistic values are $W_2 = 59.6080$, $W_3 = 29.8101$, $W_4 = 21.8461$ and $W_5 = 13.6612$. We can see that there is a clear drop in these values between W_2 and W_3. By setting the significance level $\alpha = 0.025$, we find the critical value $C = X^2_{36,0.025} = 54.437$ from the chi-square distribution with $k^2 = 36$ degree of freedom. Since $W_2 > C$ and $W_3 < C$, we reject $H_0 : \Phi_{22} = 0$ and fail to reject $H_0 : \Phi_{33} = 0$. Thus, we conclude that an VAR*(2)* model is adequate to model the in-control portion of the oil data. The order of the VAR model is also used in the Dynamic Principle Component Analysis (DPCA) later in the paper as the time lag value.

For the fitted VAR(*2*) model $(Y_t - \mu) = \Phi_{12}(Y_{t-1} - \mu) + \Phi_{22}(Y_{t-2} - \mu) + \varepsilon_t$, the estimates of the parameters are as follows,

$$\hat{\mu}_6 = (2.3899 \quad 8.9342 \quad 1.3215 \quad 8.2506 \quad 0.5114 \quad 0.1038)'$$

$$\hat{\Phi}_{12} = \begin{bmatrix} 0.7150 & -0.0237 & 0.1045 & 0.0008 & -0.1016 & -0.0433 \\ 0.2618 & 0.4857 & -0.8588 & -0.0211 & 0.4640 & -0.3494 \\ 0.0145 & 0.0082 & 0.2323 & 0.0003 & -0.0942 & 0.0196 \\ -0.2546 & -0.0100 & 0.9894 & 0.4606 & 0.7465 & -1.2370 \\ 0.0142 & 0.0012 & -0.0081 & -0.0025 & 0.0996 & -0.0034 \\ -0.0149 & 0.0023 & 0.0546 & -0.0012 & -0.0493 & 0.0625 \end{bmatrix}$$

$$\hat{\Phi}_{22} = \begin{bmatrix} -0.1385 & 0.0003 & -0.2326 & 0.0014 & 0.0591 & 0.2485 \\ -0.1378 & 0.1302 & 0.7855 & 0.0099 & 0.0037 & 0.7663 \\ -0.0300 & -0.0039 & 0.2115 & -0.0031 & -0.0626 & -0.0538 \\ 0.2238 & -0.0209 & 0.1360 & -0.0063 & 0.3450 & -0.9185 \\ -0.0119 & -0.0062 & -0.0377 & -0.0001 & 0.1411 & -0.0023 \\ 0.0143 & -0.0057 & -0.0188 & 0.0006 & 0.0899 & 0.0384 \end{bmatrix}$$

$$\hat{\Sigma}_2 = \begin{bmatrix} 6.6044 & 1.1622 & 0.0802 & -2.8131 & 0.0168 & 0.1088 \\ 1.1622 & 25.3816 & 0.2222 & 1.9658 & 0.4041 & 0.0739 \\ 0.0802 & 0.2222 & 0.2333 & -0.3565 & 0.0974 & 0.0245 \\ -2.8131 & 1.9658 & -0.3565 & 75.2260 & 0.0304 & -0.1669 \\ 0.0168 & 0.4041 & 0.0974 & 0.0304 & 0.2686 & 0.0211 \\ 0.1088 & 0.0739 & 0.0245 & -0.1669 & 0.0211 & 0.2239 \end{bmatrix}$$

The eigenvalues of Φ_2 are (-0.4261, -0.2512 + 0.1343i, -0.2512 - 0.1343i, -0.1792, -0.0000, 0.2915 + 0.2170i, 0.2915 - 0.2170i, 0.6724, 0.6135, 0.4358 + 0.0667i, 0.4358 - 0.0667i, 0.4230), where Φ_2 is described as $\Phi_2 = \begin{bmatrix} \Phi_{12} & \Phi_{22} \\ I & 0 \end{bmatrix}$. Since the eigenvalues are all less than one in absolute value, the fitted VAR(2) model is stationary. This result indicates that choosing the relatively simple Yule-Walker method for the VAR model parameters estimation and Wald statistic for the model order selection is appropriate.

5. Time series modeling of the oil data in the healthy state

The readings of the six metal elements representing oil data are both cross and auto-correlated. Since the oil data doesn't satisfy the assumption of independence of samples collected at different time epochs, the original PCA method is not suitable here. Directly applying PCA to oil data will not reveal the exact relations between the variables of the process. Dynamic PCA (DPCA), an extension of the PCA method, is used to process the oil data. The correlation relationship in oil data is represented by the cross-covariance and the auto-covariance matrices. The correlation relationship can be represented by the covariance matrices $\Gamma(0), \Gamma(1)$ and $\Gamma(2)$. $\Gamma(0)$ is the cross-covariance matrix and $\Gamma(i)$ is the auto-covariance matrix of time lag $i, i = 1,2$. These covariance matrices are used when applying DPCA to oil data. Unlike the PCA method, when DPCA is applied, the data matrix is composed of the time-shifted data vectors.

DPCA is based on conducting singular value decomposition on an augmented data matrix containing time lagged process variables. It is essentially the same as PCA except that the data vectors of the variables consist of the current data vector Y_t and the time-shifted vectors $Y_{t-1}, Y_{t-2}, \cdots$. For example, in the case of our oil data, the process dynamics is described by a vector VAR(*2*) model so that the data vector considered in the DPCA is $(Y'_t, Y'_{t-1}, Y'_{t-2})'$ instead of Y_t which would be considered in the PCA.

The starting point for DPCA is to obtain the sample covariance matrix $\hat{\Gamma}$. In our vector VAR(*2*) model, the covariance matrix Γ consists of $3 \times 3 = 9$ blocks, each of dimension 6×6, where the $(i, j)th$ block matrix is $\Gamma(i-j)$, $\Gamma(i-j) = \Gamma(j-i)'$ if

$i-j<0$, for $i,j=1,2,3$. In this section, $\Gamma(i)$ denotes the sample covariance matrix of time lag i. When using covariance matrix for the analysis, it is assumed that the variables do not have grossly different variance. If they do, the first few principal components tend to be determined by the variables with larger variances. In such cases the data should be standardized and correlation matrix should be used. For our oil data, since the means and the variances of the original variables vary widely, it is more appropriate to use the correlation matrix rather than the covariance matrix. The lag i correlation matrix can be obtained from the covariance matrix directly by $R(i)=D^{-1}\Gamma(i)D^{-1}, i=0,1,2$, where D is the diagonal matrix of standard deviations of the original variables. The structure of the sample correlation matrix used in DPCA is the same as the structure of the sample covariance matrix Γ, i.e., the $(i,j)th$ block matrix is $R(i-j)$, where $R(i-j)=R(j-i)'$ if $i-j<0$. Replacing O_t by $D_t=(O'_t, O'_{t-1}, O'_{t-2})'$ in formula (5), we can obtain the PC scores from the original oil data.

The eigenvalues $\{l_i\}$ of the sample correlation matrix R ordered from the largest to the smallest are the sample variances of the (sample) principal components. The purpose of DPCA is to generate a reduced set of variables (PCs) that accounts for most of the variability in the original oil data. A number of procedures for determining how many components to retain have been suggested in the literature. Here we apply the scree test, an approach first proposed by Cattell [8]. If we want to be sure that the principal components are properly selected, we may need to do the selection based on two or more criteria and all of them should give the same result. We use the successive difference of eigenvalues, l_i - l_{i+1}, as another criterion and look for the point at which l_i - l_{i+1} becomes fairly constant for the subsequent values. We also plot the eigenvalue difference in order to confirm the scree test PC selection result. The result of the scree test for the in-control oil data and the plot of eigenvalue differences are shown in Fig. 1. The eigenvalues obtained after applying DPCA are listed in Table 1.

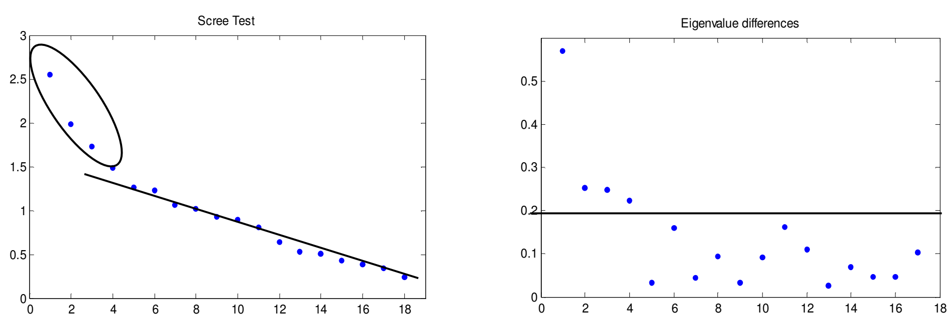

Fig. 1: Content of Grey Relation Analysis

Table 1: Successive eigenvalues l_i and eigenvalues differenece l_i-l_{i+1}

	i=1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
l_i	*2.5504*	*1.9820*	*1.7287*	*1.4809*	*1.2588*	*1.2259*	*1.0655*	*1.0203*	*0.9268*
l_i-l_{i+1}	0.5684	0.2533	0.2478	0.2221	0.0329	0.1604	0.0452	0.0935	0.0339
	10	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*
l_i	*0.8929*	*0.8020*	*0.6396*	*0.5292*	*0.5025*	*0.4326*	*0.3852*	*0.3398*	*0.2369*
l_i-l_{i+1}	0.0909	0.1624	0.1104	0.0267	0.0699	0.0474	0.0454	0.1029	--

The Scree Test plot in Figure 1 clearly shows that there is a break (an elbow) between the first four and the remaining fourteen eigenvalues that form approximately a straight line. The eigenvalue difference plot in Figure 1 and Table 1 also show that from *i=4*, l_i - l_{i+1} becomes relatively small. The above results indicate that we should retain the first four PCs to construct our DPCA model.

6. The DPCA-based $(T_{4,t}^2, Q_t)$ charts for the oil data

After applying the DPCA and selecting the principle components accounting for most of the process variation, we obtained a 4 dimensional PC scores series. Putting $A = 4$ in the formulas (6) and (8), we calculated the $T_{4,t}^2$ and Q_t values. Setting the significance level as $\alpha = 0.025$, we obtained: $T_{4,UCL}^2 = 11.4305$, $Q_{UCL} = 23.2147$ from formulae (3) and (9). We then plotted the $T_{4,t}^2$ and the Q_t values on the corresponding charts. The T_4^2 chart shows too many false alarms and misses most of the impending failures. The reason is that the 4-dimensional PC score series is still highly serially correlated, which makes the T_4^2 chart ineffective. On the other hand, the Q chart is based on the residuals of the DPCA model and serial correlation is not significant when the process is in control. In the Q chart, false alarms are significantly reduced and the true alarms coincide with the impending transmission failures.

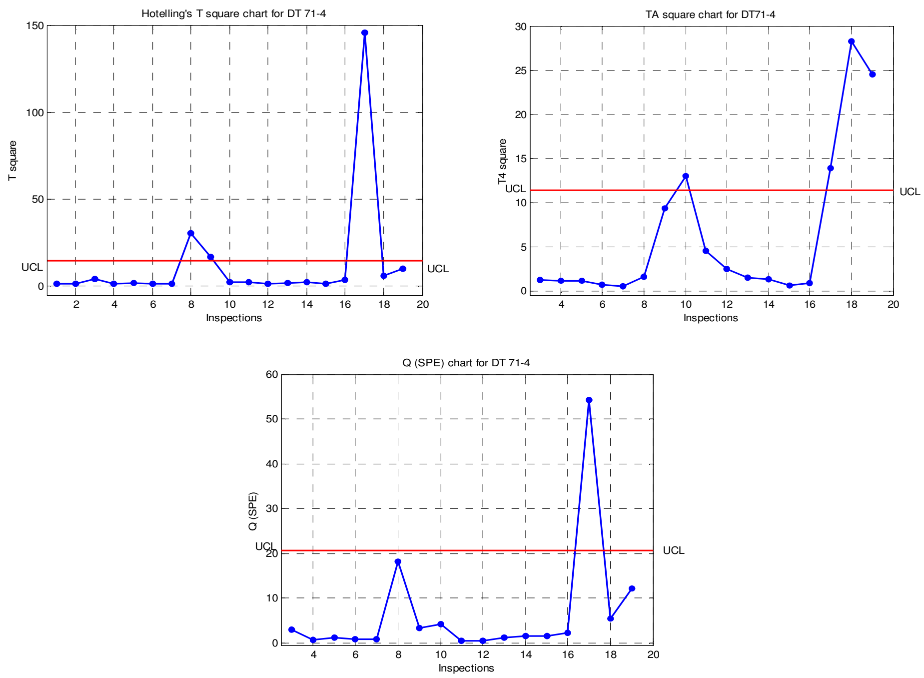

Fig. 2: Comparison of the T^2 chart, T_4^2 chart and Q chart

For example, Fig. 2 compares the traditional T^2 chart with the DPCA-based (T_4^2, Q) charts applied to history DT 71-4 which ended with suspension. On the T^2 chart, there are out-of-control alarms at the 8th, 9th and 17th sampling points. On the T_4^2 chart, out-of-control alarms appear at the 10th, 17th, 18th and 19th sampling points. It is clear that the 17th sampling point alarm in the T_4^2 chart has an effect on the next two $T_{4,t}^2$ values because of the high serial correlation of the PC scores and two subsequent alarms appeared that should not be there. On the Q chart, only one alarm appears at the 17th sampling point. Because this history ended with suspension, which means the transmission did not break before being suspended, all these alarms are false alarms although the alarm generated by the Q chart is much closer to the suspension time than the alarms on the T^2 and T_4^2 charts. Therefore, it is obvious that the T_4^2 chart and the traditional T^2 chart are not appropriate for transmission condition monitoring due to the serial correlation of the oil data, and the Q chart shows a better performance. Note that the T_4^2 chart and the Q

chart begin with the 3rd sample because the time shift in DPCA makes the first 2 $T_{4,t}^2$ and Q_t values unavailable.

7. Comparison of the fault detection schemes

From the previous analysis, we found that the DPCA-based T_4^2 chart is not an appropriate condition-monitoring tool due to the PC scores' serial correlation. In this section, we compare the fault detection capability and perform a maintenance cost analysis for the T^2 chart and the DPCA-based Q chart, denoted as Q_{DPCA} chart. When we applied the two charts to the histories that ended with suspensions, we only focused on the false alarms in the two multivariate SPC charts. In these cases, the histories ended with suspensions and the transmissions were replaced applying the age-based policy after about 12,000 working hours regardless of their actual conditions. By setting the significance level $\alpha = 0.025$, we first applied the T^2 chart and the Q_{DPCA} chart to the 23 histories that ended with suspension. The results of the two charts are summarized in Table 2.

Table 2: Comparison of the T^2 and Q_{DPCA} charts applied to suspended histories

Histories	Number of inspections	Alarms in T^2 chart	Alarms in Q_{DPCA} chart
DT 65-1	--	--	--
DT 65-2	20	13th	--
DT 65-3	--	--	--
DT 66-2	--	--	--
DT 66-3	17	4th, 6th	--
DT 67-2	20	13th, 14th, 15th	**15th**
DT 67-3	--	--	--
DT 68-1	--	--	--
DT 69-3	--	--	--
DT 70-1	--	--	--
DT 70-2	20	10th, 11th, 17th	**10th, 17th**
DT 70-3	13	2nd	--
DT 71-4	19	8th, 9th, 17th	**17th**
DT 71-5	--	--	--
DT 72-1	20	**1st, 2nd, 3rd**, 4th	3rd
DT 72-3	--	--	--
DT 74-1	20	18th	**18th**
DT 74-3	13	3rd	--
DT 75-1	--	--	--
DT 76-2	17	**1st**, 11th, 16th, 17th	--
DT 77-3	17	4th, 5th	**5th**
DT 78-1	15	3rd	--
DT 79-3	--	--	--

We can see that out-of-control signals occur more frequently in the T^2 chart than in the Q_{DPCA} chart, which indicates that the T^2 chart is more sensitive than the Q_{DPCA} chart. One can assume that the out-of-control signals occurring at the first three samplings are run-in period alarms. After the run-in period any subsequent signals can be considered to be false alarms. The out-of-control signals and the false alarms (in bold font) are summarized in Table 2. Totally, there are 19 false alarms for the T^2 chart while only 6 for the Q_{DPCA} chart. Thus, we can conclude that the Q_{DPCA} chart is superior to the traditional T^2 chart which, due to the serial correlation of the data, generates too many false alarms.

For failure histories, our main goal is to compare the failure detection capability of the two charting methods. Consider again the significance level $\alpha = 0.025$. A comparison of the charts applied to failure histories is given in Table 3.

Table 3: Comparison of the T^2 and Q_{DPCA} charts applied to the failure histories

Histories	Number of inspections	Alarms in T^2 chart	Alarms in Q_{DPCA} chart
DT 66-1	18	18^{th}	18^{th}
DT 67-1	13	**5^{th},** 13^{th}	13^{th} (Q increases)
DT 68-2	10	1^{st}, **5^{th},** 10^{th}	10^{th}
DT 68-3	--	--	--
DT 68-4	--	--	--
DT 69-2	16	16^{th}	16^{th}
DT 71-1	9	2^{nd}, 3^{rd}, **4^{th}, 5^{th}, 6^{th}**	3^{rd}
DT 71-2	9	1^{st}, **5^{th},** 9^{th}	9^{th}
DT 72-2	14	14^{th}	14^{th} (Q increases)
DT 73-1	--	--	--
DT 74-2	--	--	--
DT 77-1	7	7^{th}	7^{th}
DT 79-2	--	--	--

We can see from the Table 3 that the out-of-control signal occurs more frequently in the T^2 chart than in the Q_{DPCA} chart. The out-of-control alarm occurring at the last sampling point indicates that the chart correctly detects the impending failure and triggers a preventive replacement which avoids failure. Furthermore, since the Q_{DPCA} chart is less sensitive than the traditional T^2 chart, based on the same significance level, in case the Q_{DPCA} statistic value dramatically increases, even if it does not exceed the control limit, this can also be considered as an alarm indicating the impending failure. In the 13 failure histories, the traditional T^2 chart detects 7 impending failures giving 6 additional false alarms (in bold font), while the Q_{DPCA} chart triggers 7 preventive replacements without giving any false alarm. From the above comparison, we can see that the Q_{DPCA} chart is superior to the T^2 chart since it detects the same number of impending failures without giving any false alarm. So, from the results in Table 2 and Table 3, we can conclude that the Q_{DPCA} chart is an effective SPC charting method which can prevent considerable number of failures in a timely manner without generating an excessive number of false alarms.

With increasing false alarm cost, the fault detection capability of the T^2 chart may be further decreased due to the higher false alarm rate, which may also lead to the loss of confidence in this condition-monitoring tool.

8. Conclusion

In this paper, we have considered the DPCA-based (T_A^2, Q) charts for fault detection. After applying the DPCA transformation, serial correlation still exists in the PC scores, which considerably reduces the effectiveness of the DPCA-based T_A^2 chart. On the other hand, for the DPCA-based Q chart, no such autocorrelation effect exists. Considering the fact that regular transmission oil changes eliminate a cumulative increase of metal particles in oil samples, which makes it difficult to predict impending failures in condition monitoring, the fault detection ability of the Q_{DPCA} chart in this study is excellent.

Compared with the currently used age-based maintenance policy, both the T^2 chart-based and Q_{DPCA} chart-based maintenance policies result in significant cost savings. They both accurately indicated impending failures in more than 50% of the real failure data histories obtained under the age-based policy. The low false alarm rate of the Q_{DPCA} chart makes the policy based on this chart more cost effective and reliable than the T^2 chart-based policy. Both charts are based on sound statistical principles of multivariate SPC theory, they are easy to implement, and can prevent considerably large number of failures than the traditional age-based policy. We hope that the results presented in this study will motivate further research in this area.

Acknowledgement: Authors would like to thank the Natural Sciences and Engineering Research Council of Canada and the Imperial Oil Foundation for their financial support.

References

[1]. Makis V. and A. K. S. Jardine, *Optimal replacement in proportional hazard model,* INFOR, Vol.30, pp.172-183, 1992.

[2]. Aven T., *Condition-based replacement policies – a counting process approach,* Reliability Engineering and System Safety, Vol.51, pp.275-281, 1996.

[3]. Christer A. H., W. Wang and J. M. Sharp, *Case study a state space condition monitoring model for furnace erosion prediction and replacement,* European Journal of Operations Research, Vol.101, pp.1-14, 1997.

[4]. Makis V. and X. Jiang, *Optimal replacement under partial observations*, Mathematics in Operations Research, Vol.28, pp.382-294, 2003.

[5]. Makis V., J. Wu and Y. Gao, *An application of DPCA to oil data for CBM modeling,* European Journal of Operations Research, Vol.174, pp.112-123, 2006.

[6]. Hotelling H., Multivariate Quality Control, *Techniques of Statistical Analysis.* Edited by C. Eisenhart, M. W. Hastay, and W. A. Wallis. McGraw-Hill: New York, 1947.

[7]. Jackson J. E., *A Users' Guide of Principal Components*, John Wiley: New York, 1991.

[8]. Cattell R. B., *The scree test for the number of factors,* Journal of Multivariate Behavior Research, Vol. 1, pp. 245-276, 1966.

Jianmou Wu is a recent Ph.D. graduate from the Mechanical and Industrial Engineering Department of University of Toronto, Canada. His research interests include condition-based monitoring modeling and optimization, development of condition monitoring and fault detection schemes for deteriorating equipment, statistical data analysis, multivariate time series modeling and stochastic OR modeling.

Viliam Makis is a Professor in the Mechanical and Industrial Engineering Department of University of Toronto, Canada. His research and teaching interests include quality assurance, stochastic OR modeling, maintenance, reliability, and production control. His recent contributions have been in the area of modeling and optimization of partially observable processes with applications in CBM and multivariate quality control. He has contributed also to the development of EMQ and other production models with inspections and random machine failures, joint SPC and APC for deteriorating production processes, scheduling of operations in FMS, reliability assessment of systems operating under varying conditions, and modeling and control of queuing systems. He was a founding member of CBM Consortium at the University of Toronto in 1995.