Preprints of the 10th IFAC International Symposium on Dynamics and Control of Process Systems
The International Federation of Automatic Control
December 18-20, 2013. Mumbai, India

Why Risk-Based Multivariate Fault Detection and Diagnosis?

O. Zadakbar[1], S. Imtiaz[2], F. Khan[3]

Process Engineering, Faculty of Engineering and Applied Science,
Memorial University of Newfoundland, St. John's, Newfoundland, Canada, A1B 3X5
email:o.zadakbar@mun.ca[1]; simtiaz@mun.ca[2]; fikhan@mun.ca[3]

Abstract: A novel risk-based fault detection method has been developed. The proposed method provides a dynamic process risk indication based on the probability of happening a fault and its consequence. In this method instead of generating an alarm based on residuals or signals an alarm is activated only when the calculated risk of operation exceeds the acceptable threshold. This is an important concept as it can funnel the attention and effort of operators to the faults which poses the most operational or safety risk. Application of this new risk-based approach provides early warning of the fault as well as the associated risk with the fault. Methodologies were developed to apply the concept with model based fault detection algorithm as well as multivariate history based fault detection techniques. In this paper we show the model based approach by combining Kalman filter with the risk based approach. The history-based approach was demonstrated using principal component analysis (PCA). This method has more power in discerning between operational changes and abnormal conditions which have potential to cause accidents.

1. INTRODUCTION

1.1. Why risk based multivariate fault detection?

In general fault detection techniques are aimed to detect operational faults that affect the control objectives of the process. However, in the context of process safety these methods are inadequate as none of the methods take into account the potential impact of the fault on the process and the environment. In order to address this issue risk based fault detection method was earlier proposed by Bao et al. (2010) and later developed by Zadakbar et al. (2013). Instead of generating an alarm based on residuals or signals crossing the threshold, the risk based fault detection method issues an alarm only when the risk of a process exceeds the acceptable threshold. The risk of a process is defined as a combination of probability of fault and severity of the fault. This is an important concept as it eliminates faults which are operational and not process safety concerns and also gives a dynamic indication of the operational risk. Bao et al. (2010) used univariate charting method to calculate the probability of fault. This potentially limits the effectiveness of the method due to inherent limitations of univariate fault detection and diagnosis approach. In this paper we propose multivariate risk-based fault detection and diagnosis technique. This technique can be implemented with both model-based and history-based approaches.

1.2. Model-based and Model-free approach

In this paper we describe the methodology for calculating process risk in combination with both model-based FDD and model-free FDD. For model-based approach, a Kalman filter based multivariable residual generation technique has been combined with the risk calculation procedure. The proposed method takes advantage of the known relationship between process input and output and therefore has more power in fault detection and precise risk calculation. Also, the use of Kalman filter makes the method robust to false alarms, which is an important element as our objective is to detect the most serious faults which are safety concern for the process.

In model-free approach, Principal Component Analysis (PCA) has been combined with the risk assessment procedure proposed earlier. PCA is a dimensionality reduction technique that takes advantage of the correlation between variables. PCA has got wide spread acceptance in process industries as a monitoring tool (Russell, 2000). In a chemical process typically thousands of variables are monitored. PCA offers an alternative way to represent process data, it compresses large set of variables into few important variables called Principal Components (PC)s along the direction of the most dominant variance of the process (Imtiaz, 2007). These principal components give early indication of faults and are used for process monitoring purpose. We developed the methodology to calculate the risk associated with each PC and an alarm is generated when the risk exceeds the allowable limit.

1.3. Dynamic Risk Assessment

Risk is defined as a measure of likely harm or economic loss caused by the fault if corrective action is not taken. It depends on two factors: the probability of occurrence of a fault leading to an unwanted event and severity of the loss caused by the event (CCPS, 2000). Bao et al. (2010) proposed the following formula for a univariate deterministic system:

$$Risk = p \times s \qquad (1)$$

where p is probability of the fault leading to a catastrophic event and s is severity of a fault. The probability of the fault is calculated using Equations 2 and 3. The probability of an event increases as the process moves away further from the normal operation. In this methodology, $\mu - 3\sigma$ and $\mu + 3\sigma$ are used as the low and high threshold for the normal operation. All data points between these two thresholds are considered as normal. If any predicted point goes outside this region; it could lead to a fault leading to a catastrophic event. When the monitored signals are at the threshold $\mu \pm 3\sigma$ the probability of fault is 0.5 as it can either go back to normal or may keep growing and ultimately lead to a catastrophic event. Based on this intuition, a cumulative normal distribution was developed by Bao et al. (2010) for fault probability, which is also used in the present study. In the original formulation, Bao et al. (2010) used the process variable as the monitored signal, however in a multivariable fault detection and diagnosis process variables are not directly monitored, rather residuals generated from the process or variables transformed to a different space (i.e., PCs) are monitored. In this paper we develop modified equations for risk calculation based on multivariate FDD. For positive fault signals when the signal approach the upper threshold the probability of the fault is calculated by $\varphi\left(\frac{x-(\mu+3\sigma)}{\sigma}\right)$. On the other hand, for negative fault signal the signal approach the lower threshold and the probability of the fault is calculated by the complement, $1-\varphi\left(\frac{x-(\mu+3\sigma)}{\sigma}\right)$, when

$$x > \mu \rightarrow p = \varphi\left(\frac{x-(\mu+3\sigma)}{\sigma}\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-(\mu+3\sigma))^2}{\sigma^2}} dx \tag{2}$$

$$x < \mu \rightarrow p = 1-\varphi\left(\frac{x-(\mu-3\sigma)}{\sigma}\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-(\mu-3\sigma))^2}{\sigma^2}} dx \tag{3}$$

Figure 1 gives a visual depiction of a fault probability. Probability of the fault for a point on the centerline or at the average value μ, is 0 and for a given point on the thresholds is equal to 0.5.

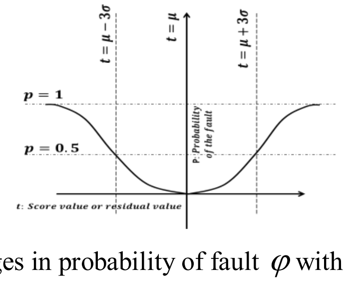

Fig. 1: Changes in probability of fault φ with deviation of score or residuals from the mean

The severity of fault is calculated using Equations 3-9. These equations are modified versions of the original equation proposed by Bao et al. (2010). The original equation needed modification since it was developed for univariate methods and did not take into account the various degree of severity caused by different types of fault. In the modified equation the severity of the fault associated with different process variables is considered. Further, it also takes into account that the increasing and decreasing rate of fault often do not have the same intensity of severity. Therefore, these cases should be treated differently. The modified severity equation is given below:

$$\text{For } x > \mu \rightarrow s = (a.100)^{\frac{(x-(\mu+3\sigma))}{x-\mu}} \tag{4}$$

$$x < \mu \rightarrow s = (a.100)^{\frac{((\mu-3\sigma)-x)}{\mu-x}} \tag{5}$$

$$a' = ab \tag{6}$$

Coefficient a in the above equations is called the intensity coefficient that indicates the intensity of the severity of the fault associated with each process variable. For instance, in a simple chemical reactor containing non-hazardous chemical compounds, the severity of damage caused by uncontrolled temperature is much higher than an uncontrolled concentration of a given component. Thus, a associated with temperature is larger than a associated with concentration.

Coefficient b in the above equations is called the moderation coefficient. Since the severity of the fault in case of a decreasing rate may not be equal to the severity of the fault while it has an increasing rate, coefficient b is used to consider this effect and moderate the severity. For example, the severity of abnormally increasing temperature in a given process can have much more damaging effects than an unusually decreasing temperature. On the other hand, a decreasing cooling water flow in a reactor can cause severe damage and needs more immediate attention than the increasing cooling water flow. Coefficient b gives the flexibility to treat increasing and decreasing faults differently. Both coefficient a and b are selected based on process and operational considerations e.g. process nature, number of people at risk, chemical and physical components, environment and costs in a given process system.

1.4. Univariate vs. Multivariate

Historically, univariate methods, such as limit or trend checking of measured output variables, are used for fault detection. However, the applicability of univariate method is limited as it is unable to distinguish between normal operational changes and real abnormal faulty conditions therefore univariate methods are prone to false alarms. More success in fault detection has been observed when multivariable fault detection techniques are applied, as these methods take advantage of the dependence among the process variables (i.e., between input and output), and flags a fault when any deviation from this process model is required or it can be purely data based where correlation between the variables are captured from the process data history (Isermann, 2005).

2. RISK-BASED FAULT DETECTION AND DIAGNOSIS

2.1 Risk Calculation in Model-based FDD

In this paper we propose a multivariate risk-based fault detection and diagnosis technique using a model. We use the model along with Kalman filter to generate the residual of the process. The residuals are used in Equations 2 & 3 to calculate the probability of fault. Severity of each fault is calculated using Equations 4 & 5 and subsequently these two quantities are combined according to Equation 1 to calculate process risk at any instant. The proposed method takes advantage of the relationship between process input and output and therefore has more power in fault detection and precise risk calculation. Also, the use of Kalman filter makes the method more robust to false alarms, which is an important element to any fault detection algorithm that is targeting the safety issues of an operation.

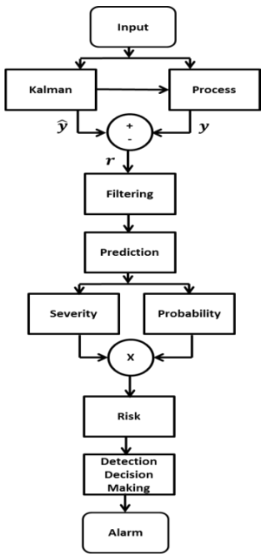

Fig. 2: Risk-based fault detection methodology using KF

The methodology for the risk-based fault detection is shown in Figure 2. Kalman filter estimates all process states and is also used for residual generation for measured process states. Subsequently based on the slopes of three consecutive real time data points the next three residual points are predicted. In the next step, severity of fault and probability of fault are calculated and the multiplication of these two quantities gives process risk. Finally, the risk profile is used for fault detection and the safety system will activate if the calculated risk exceeds threshold.

2.2 Risk Calculation in History-based FDD

We implemented the history based FDD in combination with a Principal Component Analysis (PCA). PCA is widely used in process industries for fault detection. The details of the theory and application of PCA can be found in (Imtiaz, 2007). Calculated variables, such as T^2, Q-statistics or Principal Components (PCs) are used to detect fault. Here we show the methodology to calculate risk associated with each Principal Component. The methodology for the risk-based fault detection using PCA is shown in Figure 3. The first step of the methodology is Principal Component Analysis. The output would be PCs, T^2 and Q-statistics.

Each PC is a linear combination of the original variables. PCs are uncorrelated as such can be monitored individually. Violation of threshold by these PCs, indicate serious faults in the process. In order to calculate the probability of a catastrophic fault we use the deviation of the PCs from the threshold in Equations 2 & 3. However, in order to calculate the severity of the fault modified forms of the severity equations (Eqns. 7 to 8) are used. Severity of a PC is not solely dependent on one variable rather depends on all variables which constitute the principal component. Therefore, pre-exponential term in the severity equation is calculated as a weighted average of the severity of individual variables given in Equation 9. The weights are the absolute loadings of each PC.

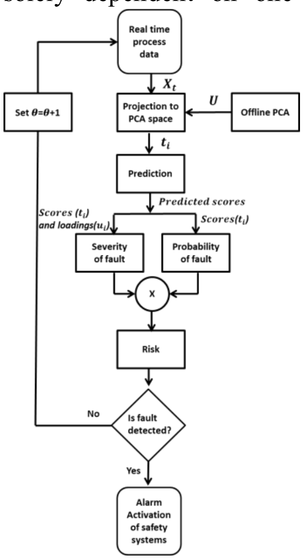

Fig. 3: Risk-based fault detection methodology using PCA

For $t > \mu \rightarrow s = (a'.100)^{\frac{(t-(\mu+3\sigma))}{t-\mu}}$ (7)

$$t < \mu \rightarrow s = (a'.100)^{\frac{((\mu-3\sigma)-t)}{\mu-t}} \quad (8)$$

$$a' = \sum_{i=1}^{n} w_i a_i \quad (9)$$

w in Equation 9 is absolute value of the loadings, a' is also called weighted intensity coefficient. Finally, similar to the model based approach the risk profile can be used for fault detection for activating the alarm or emergency shutdown system.

3. CASE STUDIES

We demonstrate the application of the proposed Risk based FDD technique on a simulated distillation column. The binary distillation column correlation structure is detected. These methods have more power in discerning between operational changes and abnormal conditions. Multivariable methods can be model based, where a priori and detailed description of the model can be found in (Skogestad, 1982 & 1997). Briefly, the binary distillation column separate a mixture with relative volatility of 1.5 into products of 96% purity (Figure 4). The assumptions considered to model the distillation unit are: binary mixture, equilibrium on all stages, constant pressure and relative volatility, total condenser, no vapour holdup and linearized liquid dynamics. The dynamic nonlinear model has 82 states, 6 inputs and 4 output variables. The first 41 states are compositions of lighter component with reboiler composition as x(1) and condenser as x(41). State x(42) is holdup in reboiler and x(82) is hold-up in condenser. Inputs are reflux flow rate L boilup flow rate V, top or distillate product flow D, bottom product flow B, feed rate F, and feed composition z_F. There are four sensors to measure top composition x_D, bottom composition x_B, condenser holdup M_D and reboiler holdup M_B.

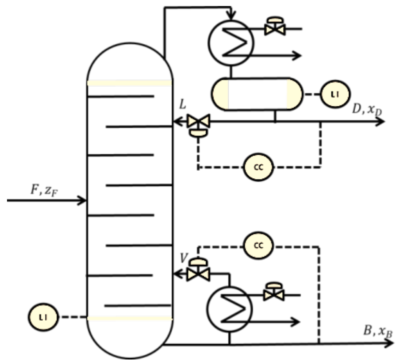

Fig. 4: Binary distillation column

3.1 Dynamic Risk calculation for Model-based FDD

In this study a sudden increase of reboiler heat flow is considered as a fault. This fault may cause increase in vapour flow rate with time which would affect other process states, e.g. top and bottom concentrations. It is assumed that vapour flow of the reboiler starts to increase at t=2000 min. There was no change in the feed or the reboiler steam setpoint. The only change made during this period was in the reflux setpoint. Figure 5 shows fluctuation of top concentration in normal and faulty conditions.

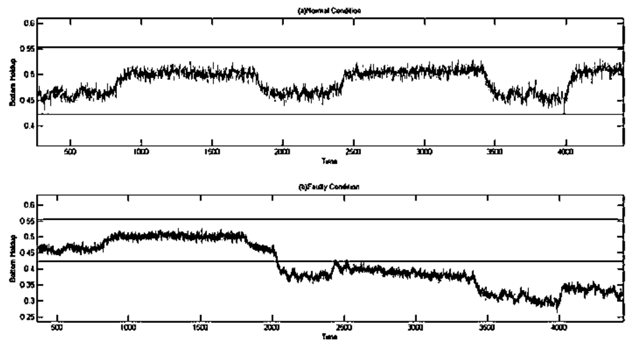

Fig. 5: Bottom holdup fluctuation in (a) normal and (b) faulty conditions

Fault in reboiler would affect top product concentration which is directly monitored. Figure 6 shows the top concentration residuals generated from Kalman filter.

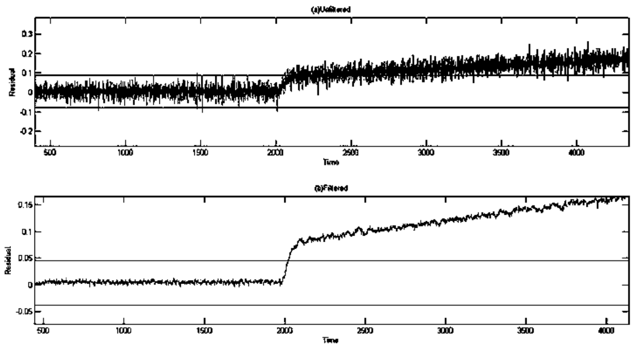

Fig. 6: Top concentration residuals in faulty conditions (a) unfiltered (b) filtered

Risk associated with the fault is calculated using these residues. Figure 7 shows the risk profile of bottom liquid holdup as well as top concentration. Since the severity of hazard associated with liquid holdup is generally higher than the severity of hazard associated with concentration a higher intensity coefficient is assigned for liquid holdup. In this case study it is assumed that $a_{Mb=2}$ and a_{xd}=1.

The first risk threshold is equal to 1 may be used for fault detection in parallel with the warning for the operators to take corrective action. The second threshold placed at 10 that would activate safety systems (emergency shutdown systems).

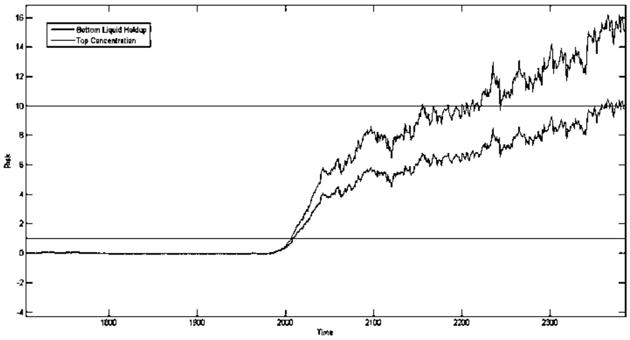

Figure 7: Risk profile of bottom liquid holdup (up) and top concentration (down) residuals

Increasing liquid holdup in each stage of the distillation column would be as hazardous as decreasing the liquid holdup. Therefore, for both increasing and decreasing fault moderation coefficient is considered one. As discussed in earlier, the risk threshold may be defined based on the acceptable risk of each process system.

In the risk based approach we use a cumulative distribution function to calculate the probability of fault. The risk is the accumulated risk up to that point and it goes up quickly if there is an increasing trend. The proposed method detects the fault early compared to the conventional methods. In this particular case based on risk alarm was activated at, t=2009 min while a residual based alarm will activate at t=2030min. On the other hand both these multivariate FDD methods detect the fault much earlier compared to the univariate method where the signals cross the threshold at t=2062min.

3.2 Dynamic Risk Calculation in History Based FDD

In this study a gradual decrease of reboiler heat flow is considered as a fault. This fault may cause increase in bottom flow rate with time which would affect other process states, e.g. top and bottom concentrations. In this simulated case the reboiler fault was introduced at t=2000 min. Set points of all other inputs except the reflux flow remain same. Figure 8 shows the bottoms flowrate, B under faulty conditions.

For applying PCA we consider the following process variables: feed composition, x_F; feed flowrate, F; distillate composition, x_D; distillate flowrate, D; bottoms composition, x_B and bottoms flowrate, B.

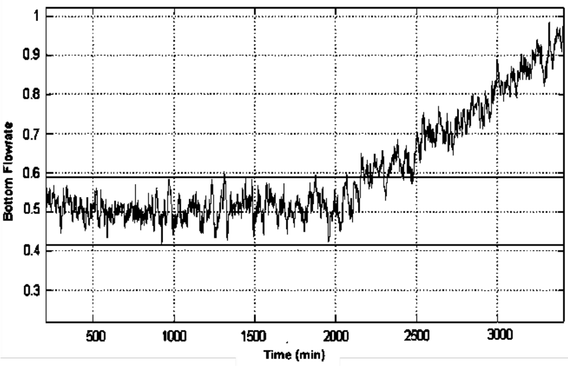

Fig.8: Bottom flow rate in faulty conditions

$$X = \begin{bmatrix} x_F \\ F \\ x_D \\ D \\ x_B \\ B \end{bmatrix} \tag{10}$$

The percentage variances explained by the PCs are shown in Table 1. First three PCs capture 89% of the total variance; therefore, we select the first 3 PCs in order to capture the correlation between the variables.

Table 1: Principal component analysis for the first case study

PC Number	%Variance captured this PC	%Variance captured total
1	48.99	48.99
2	23.7	72.69
3	16.7	89.39
4	8.16	97.56
5	2.05	99.6
6	0.4	100

The T-square and Q-statistics are shown in Figures 9 and 10. Both plots show early indication of fault. Also the scores of PC1 as shown in Fig 11 give a significant early indication of the fault. The noisy scores of PC1 are subsequently filtered and based on the slope of four real time data points in time series, the next five points of the score are predicted. Next we calculate the probability of a fault based on the score of the predicted scores.

A PC is a combination of several variables. The risk associated with a particular PC crossing the threshold is therefore a combination of the risk associated with each variable. We assign the following severity coefficient to each variable using a relative ranking method.

$$a = \begin{bmatrix} a_1 = a_{x_F} \\ a_2 = a_F \\ a_3 = a_{x_D} \\ a_4 = a_D \\ a_5 = a_{x_B} \\ a_6 = a_B \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 3 \\ 1 \\ 3 \end{bmatrix} \tag{11}$$

Loadings of each PC calculated by principal component analysis are used as weights. For the first PC loadings are:

$$w = \begin{bmatrix} w_1 = w_{x_F} \\ w_2 = w_F \\ w_3 = w_{x_D} \\ w_4 = w_D \\ w_5 = w_{x_B} \\ w_6 = w_B \end{bmatrix} = \begin{bmatrix} 0.27 \\ 0.27 \\ 0.04 \\ 0.03 \\ 0.12 \\ 0.27 \end{bmatrix}$$

$$a' = \sum_{i=1}^{6} |w_i a_i| = 1.87$$

Figures 9 through12 show PCA outputs.

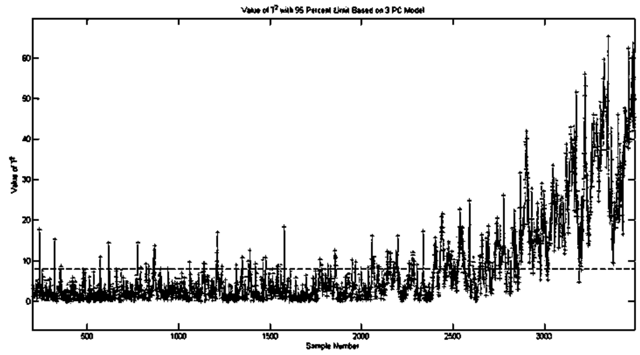

Fig.9: T^2 for the distillation unit data

The exponential term of severity is calculated using scores of PC1. The risk of the fault is obtained by combining the effect of severity with the probability. Figure 12 shows the risk profile for the PC. The threshold for the risk signal is based on the acceptable risk criteria of the process system. In Figure 11 guiding principles of acceptable risk in process operation are shown. The first risk threshold is at 1. It can be used to activate warning system for the operators to take action in order to bring the process back to normal. If no corrective action is taken or the corrective action fails to bring down the risk, and risk exceeds the second threshold then the automatic safety system is activated (emergency shutdown system).

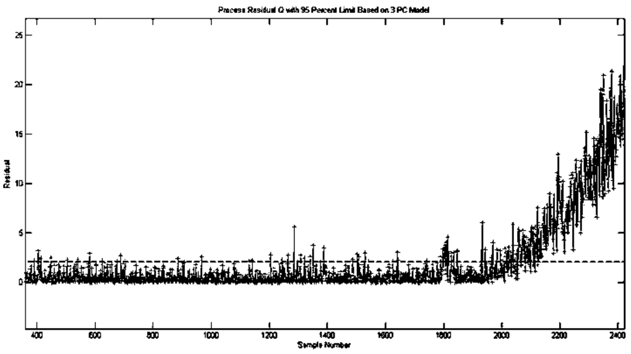

Fig.10: Q statistics for the distillation unit data

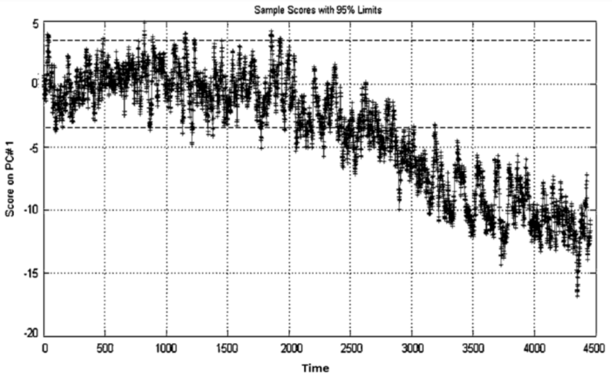

Fig.11: First PC for the distillation unit data

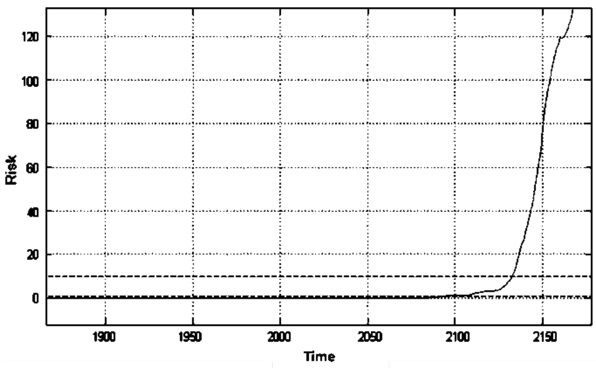

Fig.12: Risk profile of the first PC of the distillation unit data

Application of the intensity coefficient enhances the ability of the methodology to incorporate the impact of the fault on product quality, economics and potential accident.

These figures also show how application of the risk-based fault detection provides early warnings prior compared to other conventional methods. An alarm is activated at t=2090 min when the risk of operation exceeds the acceptable threshold while an alarm will be issued at t=2400s if an alarm is issued based on univariate method when the Bottoms flowrate exceeds the threshold. This improvement is coming from two sources, first from application of the multivariate method and second, since risk is a cumulative term it accumulates the past information. Therefore, in case a particular fault is sustained for some time risk will quickly start to build up and it will be reflected in the risk profile.

4. CONCLUSION

A multivariate risk-based fault detection and diagnosis technique targeting the safety issues of a process has been proposed in this paper. In this method instead of generating an alarm based on residuals or signals crossing the threshold an alarm is activated only when the risk of operation exceeds the acceptable threshold. This method has more power in discerning between operational changes and abnormal conditions which have the potential to cause catastrophic events. In addition, modifications have been made in the severity equations to adapt the method for multivariate fault detection method. The proposed risk-based fault detection technique is demonstrated on a binary distillation unit. The case studies show that the method can successfully generate early warning for different types of process faults. Further, due to risk-based approach, warning and recovery options can be prioritized.

5. ACKNOWLEDGMENT

Authors gratefully acknowledge the financial support provided by Vale Inc., Research & Development Corporation (RDC) and Atlantic Canada Opportunities Agency (ACOA) under the Vale Research Chair Program.

REFERENCES

Bao, H. Khan, F. Iqbal, T. Chang, Y. (2011). Risk-based fault diagnosis and safety management for process systems. Process Safety Progress, 30(1), pp 6–17.

Center for Chemical Process Safety of The American Institute of Chemical Engineers, CCPS, (2000). Guidelines For Chemical Process Quantitative Risk Analysis, (2), Wiley Publication.

Imtiaz, S. A. Shah, S. L. Patwardhan, R. H. Palizban, A. Ruppenstein, J. (2007). Detection, Diagnosis and Root Cause Analysis of Sheet-Break in a Pulp and Paper Mill with Economic Impact Analysis, The Canadian Journal of Chemical Engineering, 85, pp 512-525.

Isermann, R. (2005). Model-Based Fault-Detection and Diagnosis – Status and Applications, Annual Reviews In Control, 29, pp 71–85.

O. Zadakbar, S. Imtiaz, F. Khan, "Dynamic Risk Assessment and Fault Detection Using Multivariate Technique", Proc. Safety Prog., doi: 10.1002/prs.11609.

Russell, E. L. Chiang, L. H. Braat, R. D. (2000). Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis, Chemometrics and Intelligent Laboratory Systems, 51, 81-93.

Skogestad, S. (1997). Dynamics and control of distillation columns - A tutorial introduction, Trans IChemE (UK), 75(A), pp 539-562.

Skogestad, S. Morari, M. (1982). Understanding the Dynamic Behaviour of Distillation Columns, Industrial and Engineering Chemistry Research, 27(10), pp 1848-1862.