Comprehensive fault detection and isolation method applied to a recovery boiler

MOHAMED EL KOUJOK, MOULOUD AMAZOUZ, AND BRUNO POULIN

ABSTRACT: Early and accurate detection and isolation of industrial process faults are crucial to avoiding abnormal situations that cause productivity losses. Principal component analysis and reconstruction-based contribution (PCA-RBC) is a popular method used for such tasks. Unfortunately, this method does not guarantee correct fault isolation in cases where the faulty variables contribute little or do not contribute at all to the main principal components of the PCA model. This is the case, for example, of some pollutant emission levels that do not affect the global performance of a biomass boiler, but that should be maintained below certain thresholds. This paper proposes to adapt the PCA-RBC method to cope with such limitations. The idea is first to classify the data onto normal and abnormal conditions according to a selected parameter threshold, and then to build a PCA model using the normal dataset. The RBC approach is applied on the abnormal dataset to identify the variables that mostly contribute to the faulty situations. The proposed method is successfully demonstrated using real data from an industrial case. It is noted that an attempt to develop an accurate predictive model of the selected parameter using projection to latent structures (PLS) was unsuccessful.

Application: This paper presents a new development method to diagnosis total reduced sulfur (TRS) emission events by recovery boiler.

Fault detection and isolation (FDI) has been recognized as a vital tool to enable efficient and environmentally benign operation of industrial processes. The rapid progress and decreasing cost in sensor technology and data acquisition systems allows industry to collect a large amount of data. Data-driven methods utilize the massive amounts of available data to extract knowledge and useful information in order to generate cost-effective fault detection and isolation. This makes data-driven methods very attractive in many areas of application [1].

Principal component analysis (PCA) is one of the most popular data-driven methods. It has been proposed for the monitoring and FDI of complex systems [2]. PCA-based techniques typically involve the partitioning of data into principal and residual subspaces. Two fault detection indexes, namely the T^2 Hotteling and the squared prediction error (SPE) [3], are then derived by projecting the data onto each of these two subspaces. SPE measures the amount of faulty variations from the residual subspace, whereas T^2 captures the faulty variations from the principal subspace. A fault is detected when the values of the SPE and/or T^2 indexes are beyond certain threshold levels. The isolation of a fault is then performed by applying reconstruction-based contribution (RBC) to the detection indexes [3], by computing variable contribution to the fault. Unfortunately, this approach does not guarantee the correct detection and identification of all fault types. Such a situation was encountered in the industrial application of this conventional approach in the detection and diagnosis of sulfur emission (SE) events from a black liquor recovery boiler (BLRB). Because the SE (designated as a metric in this paper) does not appear as an important variable in the PCA model, the conventional PCA-RBC method does not allow for the detection or identification of faults related to this metric.

One alternative to this method is to develop a PLS (projection to latent structures) [4] with fault identification methods such as RBC or multivariate statistical process control [5] to detect and isolate the faulty metric. The PLS model developed for SE prediction shows very low R^2 and Q^2 values, and therefore cannot be used in this case. The low accuracy of the PLS model is due to the fact that SE events are sporadic and can be caused by contradictory complex combustion phenomena.

In this paper, a new approach is proposed to overcome the previously mentioned limitations of the standard PCA-PLS approach. The approach splits the data set into two subsets according to the targeted metric values. The first subset contains the data corresponding to periods where the values of the metric are below the high limit, and the second subset regroups the data where the values exceed the high limit. A PCA model is then built using the first subset (normal values), and the RBCs are calculated using the second subset (abnormal values).

Our work is devoted to the problem of multiple FDI using reconstruction-based contribution. The variable reconstruction approach estimates a set of variables using the PCA model from the remaining variables. If the faulty variables are reconstructed, the fault effect is removed, which is useful for fault isolation. To avoid the combinatorial multiplication of faulty scenarios related to multiple faults, an iterative approach is used. A short presentation of principal component analysis and its use for fault detection and isolation is presented first. Then, its successful application to multiple fault detection and isolation of recovery boilers (RBs) is explained.

PROCESS MONITORING AND FAULT DETECTION AND ISOLATION USING PCA-RBC

Process monitoring using the PCA model

Let $X \in \Re^{N\times m}$ denote the historical data with N observations and m variables defined by:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nm} \end{bmatrix} \tag{1}$$

The concept of the PCA is to project the data (Eq. [2]) onto a low-dimensional space that characterizes the comportment of the process ($\hat{X}$) and onto a second space that represents the residual ($\tilde{X}$) defined by:

$$X=\hat{X}+\tilde{X}=\hat{T}\hat{P}+\tilde{T}\tilde{P} \tag{2}$$

where the loading vectors ($P=[\hat{P} \quad \tilde{P}]$) contain the eigenvectors of the covariance matrix S. This matrix is given by:

$$S=\frac{1}{N-1}X^TX=P\Lambda P^T \tag{3}$$

where Λ is a diagonal matrix that contains the eigenvalues of S ranked in descending order. The columns of matrix $P(m\times m)$ are stocked according to the eigenvalues. The score matrix $T\ (N\times m)=[\hat{T} \quad \tilde{T}]$ is represented geometrically as the projection of X over the direction of P components. It is given by:

$$T=XP \tag{4}$$

Many approaches are used to find the dimension of the matrix $P\ (m\times r)$ [6]. The number 'r' represents the reduced dimension or principal components. The cross-validation technique [7] was used in this work to determine 'r'. Once the PCA model representing normal behavior is obtained, it can be used in real-time process monitoring to detect changes in operating point, sensor failure, process fault, and plant disturbance. This is done via statistical hypothesis tests on indexes in the principal and residual spaces.

Fault detection indexes

The most popular fault detection indexes are the squared prediction error (SPE) (known as the Q-statistic), the T^2 Hotelling, and a combination of the two indexes [8].

SPE index

The SPE index evaluates the residual data matrix ($\tilde{X}$) to show the variability of the data in the residual space:

$$SPE=x^T\tilde{P}\tilde{P}^T x=x^T\tilde{C}x \tag{5}$$

where $\tilde{C}=\tilde{P}\tilde{P}^T$, and the SPE control limit (δ^2) is defined as:

$$\delta^2=g^{SPE}\chi^2_a(h^{SPE}) \tag{6}$$

where $(1-a)\times 100\%$ is the confidence level,

$$g^{SPE}=\frac{\theta_2}{\theta_1} \tag{7}$$

and

$$h^{SPE}=\frac{\theta_1^2}{\theta_2} \tag{8}$$

where $\theta_1=\sum_{i=r+1}^{m}\lambda_i$, $\theta_2=\sum_{i=r+1}^{m}\lambda^2_i$ and λ_i is the i^{th} eigenvalue of the covariance S (Eq. [3]).

T^2 index

The T^2 index considers the variability of the projected data in the reduced principal space:

$$T^2=x^T\hat{P}\Lambda_r^{-1}\hat{P}^T x=x^TDx \tag{9}$$

where $D=\hat{P}\Lambda_r^{-1}\hat{P}^T$ and Λ_r is a diagonal matrix of the r first eigenvalues of the covariance matrix S. And the T^2 control limit (τ^2) is given by:

$$\tau^2=\chi^2_a(r) \tag{10}$$

with confidence level $(1-a)\times 100\%$.

Combined index

SPE and T^2 indexes behave in a complementary manner [9]. It therefore becomes possible to combine the two indexes into one index that can provide a complete measure of the variability in the entire space and that is convenient, in practice, for fault detection. The combined index (SPE and T^2) is given by:

$$\varphi=x^T\phi\, x \tag{11}$$

where

$$\phi=\frac{\tilde{C}}{\delta^2}+\frac{D}{\tau^2} \tag{12}$$

and the control limit ζ^2 is:

$$\zeta^2=g^{\varphi}\chi^2_a(h^{\varphi}) \tag{13}$$

where

$$g^{\varphi}=\left(\frac{r}{\tau^4}+\frac{\theta_2}{\delta^4}\right)/\left(\frac{r}{\tau^2}+\frac{\theta_1}{\delta^2}\right) \quad (14)$$

$$h^{\varphi}=\left(\frac{r}{\tau^2}+\frac{\theta_1}{\delta^2}\right)^2/\left(\frac{r}{\tau^4}+\frac{\theta_2}{\delta^4}\right) \quad (15)$$

with a $(1-a)\times 100\%$ confidence limit.

Since all of these fault detection indexes are in quadratic form, the notation can be simplified by considering just one general index as:

$$Index\ (x) = x^T M x \quad (16)$$

where M can be: $\tilde{C}$ (Eq. [5]), D (Eq. [9]), or ϕ (Eq. [12]). Once a fault is detected by one of the indexes, isolation methods must be used to find its cause.

Fault isolation methods

Fault isolation methods can be unified into three general isolation methods [10]:

- General decompositive contribution method, defined by:

$$GDC_i^M = x^T M^{1-\beta} \xi_i \xi_i^T M^{\beta} x \qquad 0 \le \beta < 1 \quad (17)$$

- Reconstruction-based contribution method, defined by:

$$RBC_i^M = x^T M \xi_i (\xi_i^T M \xi_i)^{-1} \xi_i^T M x \quad (18)$$

- Diagonal contribution method, defined by:

$$DC_i^M = x^T \xi_i \xi_i^T M \xi_i \xi_i^T x \quad (19)$$

where $i=1,\ldots,m$ and $M=\tilde{C}$, D or ϕ, and ζ_i is the i^{th} column of the identity matrix. These isolation methods were discussed in [3,11] and tested on simulation examples to compare the rates of correct isolation. The results showed that the $RBC_i^{M=\phi}$ (Eqs. [12] and [18]), i: variable) had the highest percent rates of correct isolation among the simulated faulty samples. Encouraged by this result, the $RBC_i^{M=\phi}$ was selected as useful support to isolate the inconsistent variables that cause the uncontrolled SE. It should be noted that this work does not suppose that the $RBC_i^{M=\phi}$ always has the best rate of isolation. The isolation can be enhanced by using another isolation method complementary to $RBC_i^{M=\phi}$. As fault isolation follows the detection step, it is important to ensure that a fault is correctly detected. Unfortunately, the previously-mentioned fault detection indexes are not able to detect all kinds of abnormal behavior in the process, as will be shown later in this paper.

PROPOSED PCA-RBC METHOD

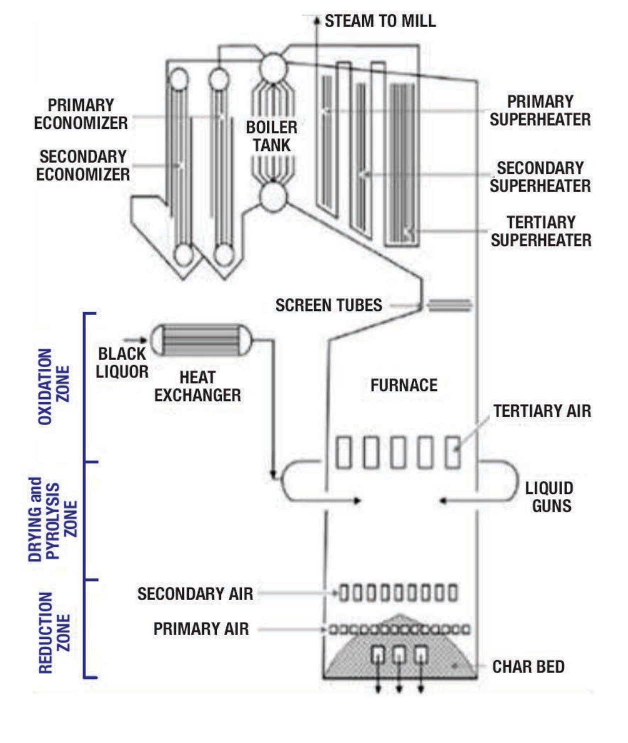
As stated earlier, the whole PCA-RBC approach does not guarantee the detection and isolation of all faults. Such is the case in the application presented in this paper, which concerns the isolation of SE events from a recovery boiler (see description in the following "Application to a black liquor recovery boiler" section). Applying the standard PCA-RBC method to this case did not work, because SE parameter loading is very low in the PCA model. Furthermore, an attempt to use PLS as an alternative multivariate statistical method to monitor SE failed due to the inaccuracy of the obtained model. Consequently, this paper aims to tackle the problem by adapting the standard PCA-RBC method.

The proposed approach consists of splitting the initial data set X into two subsets: normal X' and abnormal X'', in relation to the behavior of a selected parameter such as SE, instead of using principal and residual spaces of the standard approach. The X' subset is then used to build a PCA model that reflects the variability of the process while a selected parameter is normal or under a certain threshold (high limit). The second subset X'' is used to test the isolation method. Each observation from the X'' dataset is considered, and the $RBC_i^{M=\phi}$ (where $i=1,\ldots,m$) is calculated; the largest variable contribution given by the $RBC_i^{M=\phi}$ (where $i=1,\ldots,m$) is retained. The frequency of the occurrence of each large contributing variable on all abnormal observations is then computed. The variable with the largest frequency is considered the biggest contributor to the faulty SE. The proposed approach does not clarify the cause of the event; it simply extracts the set of process variables that are related to SE. Of course, expertise is required to find the cause of the SE events. The proposed method can be implemented online to monitor, detect, and isolate faults for selected metrics of equipment or processes. Fault detection is based on Shewhart-type control limits, rather than on the standard PCA-based indexes. When a metric value surpasses the threshold limit, an $RBC_i^{M=\phi}$ (where $i=1,\ldots,m$) of this observation can indicate the contribution of each variable to the fault. The magnitude of the $RBC_i^{M=\phi}$ (where $i=1,\ldots,m$) is a potential indicator of the fault's source.

APPLICATION TO A BLACK LIQUOR RECOVERY BOILER

Process description and problem statement

The objective of a black liquor recovery boiler (BLRB, **Fig. 1**) is to recover the inorganic chemicals as smelt (sodium carbonate [Na_2CO_3] and sodium sulfide [Na_2S]), burn the organic chemicals, and recover the heat of combustion in the form of steam. The latter is completed by a number of tubes (economizer, superheater, and boiler tank) filled with circulating water or steam to recover the heat from the flue gases. The furnace burns the concentrated black liquor by spraying it into the furnace through guns liquor. There are three zones: the upper section is the oxidizing zone where the sulfur can be oxidized (convert TRS to sulfur dioxide [SO_2]); the middle section (where the black liquor is injected) is the drying zone; and the bottom section is the reducing zone where, in a bottom bed of char, the sulfur compounds are converted to Na_2S.



1. Schematic representation of a black liquor recovery boiler (BLRB) [13].

The remaining sodium hydroxide (NaOH) and sodium salts of organic acids are converted to Na_2CO_3 [12].

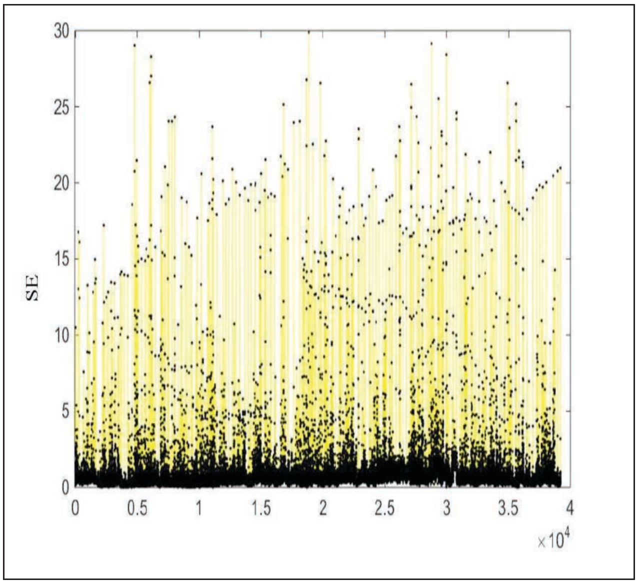
An optimized chemical recovery boiler must perform its functions safely, efficiently, and reliably, while minimizing both gaseous and particulate emissions to environmentally acceptable levels. This application focuses on an environmental issue, in which high SE in the BLRB flue gas must be kept at the lowest possible level: N_{SE} ($\mu_{SE}=0.95$ ppm, $\sigma_{SE}=0.2$). One year's worth of operations data was gathered from the plant database, using a 5-min sampling time. The data set contained 77000 observations and 60 variables representing temperatures; pressures; flow rates; and physical and chemical characteristics and contents of black liquor solids, water, steam, and flue gases.

Data preprocessing and PCA modelling

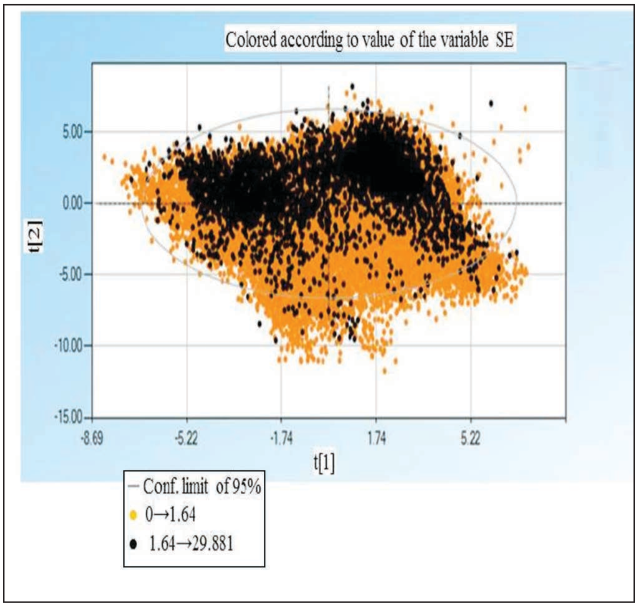
The EXPLORE software developed in-house was used to clean and filter the data before further analysis. Non-representative data corresponding to startup and shutdown periods were removed. A moving filter was also applied. The preprocessed data set, which contained 39000 observations and 60 variables including SE, was analyzed using PCA. The resulting model contains 7 principal components that account for 56% of the variability in the operation of the BLRB.

Sulfur emissions events detection and isolation using the standard PCA-RBC method

One of the objectives of this study was to monitor and isolate SE from the BLRB. The time series of the SE are shown in

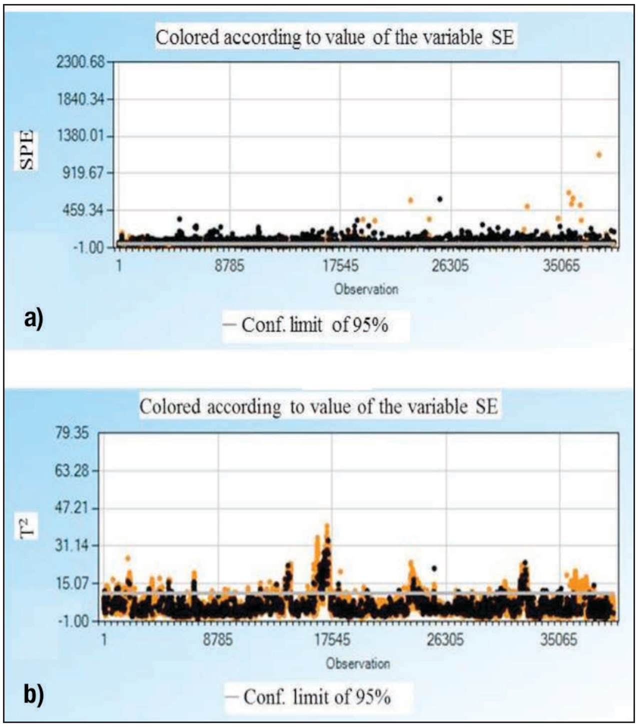


2. Time series of sulfur emissions (SE).

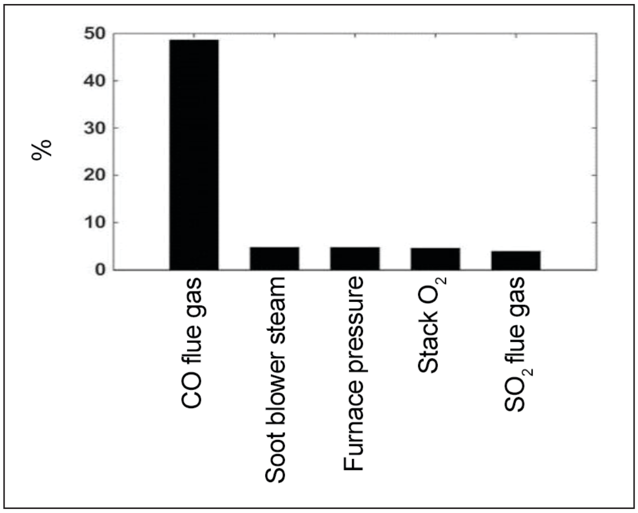


3. Scores plot for the first and second principal components (in black, SE higher than 1.64; in orange, SE lower than 1.64).

Fig. 2. The SE values vary from 0 to 30 ppm. Well-controlled BLRB operation should have the lowest possible SE values. Since zero SE is impossible in practice, the target is to keep SE below its average value, plus a few standard deviations. The scores plots of the obtained PCA model were verified, and there is no distinction between scores corresponding to low SE values and to high SE values. **Fig. 3** shows the scores plot of the first two components. Both low and high values of the SE are within the confidence limit and below the T^2 and SPE indexes limits (**Figs. 4a-b**). This clearly indicates that the standard PCA method cannot be used in this case to differentiate normal and abnormal values of SE. Thus, it cannot be used to detect and isolate high SE values.



4. a) Squared prediction error (SPE), and b) T^2 Hotteling. For both figures, the values are colored according to SE values.



5. Variable contributions to the SE fault.

Sulfur emissions events detection and isolation using the proposed method

The proposed method to detect and isolate high SE consists of splitting the original dataset into two subsets in accordance with the values of SE. The first subset, $X' \in \Re^{N_1 \times m}$, called normal data, corresponds to operation data that contain values of SE lower than the threshold $T_{SE} = \mu_{SE} + 3\sigma_{SE}$; and the second subset, $X'' \in \Re^{N_2 \times m}$, called abnormal data, corresponds to SE values higher than the threshold T_{SE}. A PCA model is then built using X'; the obtained PCA model is made of 5 principal components, explaining 50% of variability, and then $M=\phi$ (Eq. [12]) is determined in order to compute the RBC (Eq. [18]). Variable contribution for each observation belonging to X'' was obtained. It is then proposed to find the most common variable contribution to all observations. Carbon monoxide (CO) flue gas, soot blower steam, furnace pressure and stack oxygen (O_2) (**Table I**) are the variables that contributed most to SE faults (**Fig. 5**). This means that, in most cases, CO flue gas was the variable that contributed most to the SE. This does not mean that the cause of SE events is high CO, but that high CO is highly correlated to SE. In general, high CO means that part of the fuel is left unburned due to a lack of combustion air, which is one of the possible reasons that the sulfur compounds are not oxidized.

Variable	Description	Units
CO flue gas	Quantity of CO in the flue gas	ppm
Soot blower steam	Steam flow rate to lower soot	Kpph
Furnace pressure	Pressure in the BLRB furnace	kPa
Stack O_2	% of O_2 in the BLRB stack	%

I. Carbon monoxide (CO) flue gas, soot blower steam, furnace pressure, and stack oxygen (O_2) are the variables that contributed most to SE faults.

It is worth mentioning that Green and Hough [14], as well as Clement and Kagel [15] — recognized experts in BLRB design and operation — have found a strong correlation between SE and CO emissions in boilers and furnaces. As mentioned previously, a BLRB is a complex process, and a high SE level can be caused by multiple or combined factors. As shown in Fig. 5, 4% of the SE fault samples are related to soot blower steam, as is the case for furnace pressure and stack O_2. High stack O_2 means that excess air is present in the BLRB, which contradicts the effect of the lack of combustion air. In this case, it is possible that another phenomenon not captured by the data is present in the BLRB operation. Inadequate black liquor mixing due to fluctuating gun pressure might be one of these phenomena. Further analysis is therefore necessary to find the root cause of all faulty SE.

CONCLUSION

A modified application of the standard PCA-RBC approach was developed and demonstrated in a real industrial application. The motivation behind using this method is the limitation of the PCA-PLS to analyze and predict certain abnormal process behaviors, such as high SE levels. The proposed method is a useful decision-making tool in a context where accurate models are difficult to obtain and prior knowledge about the directions of the fault is unavailable.

This work suggests that the standard PCA-RBC method should be supplemented by sub-models that correspond to selected metrics that have low contribution and need to be monitored and diagnosed. **TJ**

ACKNOWLEDGEMENTS

The authors wish to thank the Program of Energy Research and Development (PERD) of Natural Resources Canada for its financial support.

LITERATURE CITED

1. Venkatasubramanian, V., Rengaswamy, R., Yin, K., et al., *Comput. Chem. Eng.* 27(3): 293(2003).
2. Chiang, L.H. and Colegrove L.F., *Chemom. Intell. Lab. Syst.* 88(2): 143(2007).
3. Alcala, C.F. and Qin, S.J., *J. Process Control* 21(3): 322(2011).
4. Wold, S., Sjöström, M., and Eriksson, L., *Chemom. Intell. Lab. Syst.* 58(2): 109(2001).
5. Li, G., Alcala, C.F., Qin, S.J., et al., *IEEE Trans. Control Syst. Technol.* 19(5): 1114(2011).
6. Weihua, L., Henry, H.Y., Valle-Cervantes, S., et al., *J. Process Control* 10(5): 471(2000).
7. Wold, S., *Technometrics* 20(4): 397(1978).
8. Qin, S.J., *J. Chemom.* 17(8-9): 480(2003).
9. Henry, H.Y. and Qin, S.J., *Ind. Eng. Chem. Res.* 40(20): 4403(2001).
10. Alcala, C.F. and Qin, S.J., "Unified analysis of diagnosis methods for process monitoring," *IFAC Proc. Vol.* 42(8): 1007(2009).
11. Pan, Y.C., Dong, Y., and Qin, S.J., "Fault diagnosis using concurrent projection to latent structures," *IFAC — PapersOnLine* 48(8): 1276(2015).
12. Biermann, J.C., *Handbook of Pulping and Papermaking*, Academic Press, San Diego, CA, USA, 1996, p. 107.
13. Cardoso, M., Avelar Costa, G.A., de Oliveira, E.D., et al., *Lat. Am. Appl. Res.* 42(2): 197(2012).
14. Green, R.P. and Hough, G., *Chemical Recovery in the Alkaline Pulping Processes*, TAPPI PRESS, Atlanta, GA, USA, 1998, p. 109.
15. Clement, R. and Kagel, R., *Emissions from Combustion Processes: Origin, Measurment, Control*, Lewis Publishers, Chelsie, MI, USA, 1990, p. 295.

ABOUT THE AUTHORS

Fault detection and isolation is an important cost effective technology that can help engineers and operators optimize process performance operation through automated fault diagnosis. The research here complements our previous research where the focus was on developing a predictive process model (soft sensor) of the metrics that need continuous monitoring. In this research, we worked on analysis and fault isolation using multivariate statistical process control (MSPC) charts. The most difficult aspect of this research was to overcome the limitations of the standard principal component analysis and reconstruction-based contribution to isolate faults that are not easily modelled, like TRS emissions. The important finding from this research is the powerfulness of the MSPC to isolate such faults.

Mills can use data from recovery boiler past operations to easily build a TRS events monitoring and diagnosis tool to timely react and avoid such events.

The next step in this research is to develop a multiple fault detection and diagnosis software tool that can assist pulp and paper engineers with continuously monitoring and rating the performance of the process, along with performing comprehensive online fault detection and diagnosis.

El koujok

Amazouz

Poulin

El koujok is research scientist, Amazouz is senior project manager, and Poulin is engineer, CEM, for CanmetENERGY, Varennes, QC, Canada.
Email El koujok at mohamed.elkoujok@canada.ca.