Case Study

Quality and Reliability Engineering International

(www.interscience.wiley.com) DOI: 10.1002/qre.1070

Published online 9 October 2009 in Wiley InterScience

Multivariate Process Monitoring of an Experimental Blast Furnace

Erik Vanhatalo*†

Process monitoring by use of multivariate projection methods has received increasing attention as it can reduce the monitoring problem for richly instrumented industrial processes with many correlated variables. This article discusses the monitoring and control of a continuously operating experimental blast furnace (EBF). A case study outlines the need for monitoring and control of the EBF and the use of principal components (PCs) to monitor the thermal state of the process. The case study addresses design, testing and online application of PC models for process monitoring. The results show how the monitoring problem can be reduced to following just a few PCs instead of many original variables. The case study highlights the problem of multivariate monitoring of a process with frequently shifting operating modes and process drifts and stresses the choice of a good reference data set of 'normal' process behavior. Possible solutions for adaptations of the multivariate models to process changes are also discussed. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: multivariate statistical methods; process monitoring; principal components; blast furnace; experiments, continuous process

1. Introduction

Statistical monitoring and control of industrial processes are vital to improve the product quality and to attain and maintain an efficient manufacturing process. Today, control charts such as the Shewhart, cumulative sum (CUSUM), exponentially weighted moving average (EWMA) and multivariate control charts form the basis for statistical process control (SPC). Multivariate process monitoring has received increasing attention, not least in the area of chemometrics, due to the fast development of computers and software during the last decades.

Sometimes, univariate (one-variable-at-a-time) control charts provide sufficient information for the engineer to make correct control decisions. However, MacGregor[1] argues that when multiple variables need to be monitored, a univariate approach is normally neither effective nor efficient.

Many examples of multivariate process monitoring in the literature come from the process industry. The production in process industry is typically continuous or batch-wise and the products often exhibit characteristics such as liquids, powders, slurries and other nondiscrete states, see Dennis and Meredith[2] and Fransoo and Rutten[3]. Hence, measuring the phenomena of interest in process industries is often difficult. Instead, industries measure what they can and often end up with a multitude of process variables used as proxies for process phenomena. Process industries are often richly instrumented with sensors routinely collecting measurements on many process variables, such as temperatures, pressures and physical properties, see Wise and Gallagher[4]. The multiple measurements are typically not independent as a few underlying events often drive the process at any given time. Many of the measured variables are therefore just different reflections of the same underlying event, see Kourti and MacGregor[5]. Because of many, often highly cross and auto-correlated process variables, Hild *et al.*[6] argue that shifts in the processing state that are visible in multivariate representations may, due to normal variation, not deviate significantly in univariate plots. Hild *et al.*[6] further point out that although measurement data often are dominated by frequent online logging of process variables, measurements of product characteristics are usually made less frequently by off-line analyses. Hence, the cause and effect relations between process conditions and product characteristics, needed to make relevant control decisions, can be hard to establish in

Quality Technology, Luleå University of Technology, SE-97187, Luleå, Sweden
Correspondence to: Erik Vanhatalo, Quality Technology, Luleå University of Technology, SE-97187, Luleå, Sweden.
†*E-mail: erik.vanhatalo@ltu.se*

Contract/grant sponsor: LKAB
Contract/grant sponsor: European Union
Contract/grant sponsor: European Regional Development Fund
Contract/grant sponsor: Produktion Botnia

continuous processes. In addition, continuous processes are often dynamic and nonstationary in nature, which further complicates monitoring and control.

The practical problem for the engineer to simultaneously keep track of many univariate control charts is another strong argument in favor of a multivariate approach to process monitoring and control. One of the key benefits of multivariate monitoring is that the performance of a process can be monitored by the operator looking at only a few multivariate charts as indicators of the process performance.

Most examples in the literature illustrate multivariate monitoring and control of production processes. Another situation, not often referred to, and the focus of this work is the need for process monitoring and control during ongoing experimentation. Consider the following example: an experiment is conducted in a process plant where a number of experimental factors, *X*s, are investigated by analyzing one (or probably many) response variables, *Y*s, from the experiment. Assume further that the process needs to be controlled (this is normally the case) during the experiment by the manipulation of other variables to keep the process in control. Moreover, suppose that the manipulation of these 'control variables' can affect the experimental results substantially. Indeed, this presents a tricky experimental situation that requires careful planning, see Vanhatalo and Vännman[7] and Vanhatalo and Bergquist[8]. In such situations, unbiased and objective process control decisions are prerequisites to be able to trust the conclusions from the experiment.

Indeed, the use of replication, randomization and blocking in experimental design make it possible to run experiments in a process subjected to simultaneous control, or more formally, a process that is not in statistical control, see Bisgaard[9]. However, due to practical and economic reasons, it is often hard to completely randomize experiments in the industry. In fact, split-plot experimental designs are frequently used, either intentionally or unintentionally, in the industry, see, for example, Kowalski *et al.*[10]. Furthermore, it is difficult to correctly block incorrect control actions before the experiment as predicting them is nearly impossible. Hence, poorly chosen control actions during an industrial experiment may in a best-case scenario merely inflate the experimental error making it harder to find significant effects of experimental factors. In the worst case, incorrect control actions can come to bias the estimated effect of one or many experimental factors or invalidate the entire experiment, especially if the randomization of the experiment is restricted. Therefore, it is argued here that an unbiased and objective process monitoring and control is essential for successful experimentation in industrial processes that require simultaneous process control.

The purpose of this article is to describe how a multivariate approach to process monitoring can be used to improve information as the basis for control decisions for a continuous process during experimentation and to discuss the application benefits and problems. A case study at an experimental blast furnace (EBF) operation illustrates the approach where principal components (PCs) are used to monitor the thermal state of the process.

2. Multivariate process monitoring

Process monitoring and control by use of multivariate statistical methods has been an active area during the last decades. Excellent and comprehensive overviews of developments of methods within the area are given by, for example, Bersimis *et al.*[11], Kourti[12] and Qin[13].

Multivariate extensions of the univariate Shewhart, CUSUM and EWMA control charts are examples of methods that can be of help in a multivariate situation, see MacGregor and Kourti[14]. The two latent variable techniques, principal component analysis (PCA) and projection to latent structures by use of partial least squares (PLS) are of special importance for multivariate process monitoring today, which is pointed out by, for example: Kourti[12], Qin[13] and Wise and Gallagher[4]. Mastrangelo *et al.*[15] discuss how PCs can be used for statistical process monitoring. PLS is used to model the covariance between two blocks of data, for example, an *X*-block and a *Y*-block from a process, see Höskuldsson[16]. In this article only one block of process data is modeled, for which the PCA technique is better suited.

Multivariate monitoring by use of latent variable methods has been used in a variety of industries and processes, to name a few examples: a continuous steel caster process[17], a titanium dioxide production process[18], a biological wastewater treatment process[19] and for blast furnace processes[20–22]. The following section explains multivariate process monitoring using PCs.

2.1. Multivariate monitoring using principal components

When the variables in the data matrix are many and correlated, PCA can reduce the dimensionality of the data matrix by extracting a few latent, uncorrelated variables called PCs (linear combinations of the original variables) that together explain the main variability in the data, see, for example, Jackson[23] and Johnson and Wichern[24]. The PCA technique and the associated monitoring tools are briefly explained below.

Let $\mathbf{x}' = [x_1, x_2 \ldots x_m]$ represent a random vector describing an *m*-dimensional variable with covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{\Sigma}$ have the eigenvalue–eigenvector pairs $(\lambda_1, \mathbf{p}_1), (\lambda_2, \mathbf{p}_2), \ldots, (\lambda_m, \mathbf{p}_m)$. The *m* PCs are formed as linear combinations of the original variables:

$$\begin{aligned} PC_1 &= \mathbf{p}_1'\mathbf{x} = p_{11}x_1 + p_{12}x_2 + \cdots + p_{1m}x_m \\ PC_2 &= \mathbf{p}_2'\mathbf{x} = p_{21}x_1 + p_{22}x_2 + \cdots + p_{2m}x_m \\ &\vdots \\ PC_m &= \mathbf{p}_m'\mathbf{x} = p_{m1}x_1 + p_{m2}x_2 + \cdots + p_{mm}x_m \end{aligned} \tag{1}$$

The PCs are orthogonal to one another and ordered with respect to their variances. The first PC has the largest variance, the second PC the second-largest variance and so on, where the eigenvalues, λ_a, $a=1,2,\dots,m$, are the variances of the PCs. The eigenvectors, $\mathbf{p}_a$, $a=1,2,\dots,m$, have unit length, $\mathbf{p}_a\mathbf{p}_a'=1$, and are called the PC loading vectors. PCA is scale-dependent and the variables are often scaled to unit variance before PCA is conducted. With standardized variables, the correlation matrix of $\mathbf{x}$ is used to derive the eigenvector–eigenvalue pairs instead of the covariance matrix, see, for example, Johnson and Wichern[24].

In practice the covariance (correlation) matrix is unknown and estimated by the sample covariance (correlation) matrix calculated from an observed $\mathbf{X}$ matrix with n observations of each of the m variables. The values of the PCs for each observation are here called PC scores, and the score vectors, $\mathbf{t}_a$, $a=1,2,\dots,m$, represent the n observed values of the PCs based on the observed $\mathbf{X}$ matrix.

The loading vectors, $\mathbf{p}_a$, $a=1,2,\dots,A$, where $A<m$, define the reduced dimension space (A) with respect to the original variables and the score vectors, $\mathbf{t}_a$, $a=1,2,\dots,A$ are the projection of the original observations onto the A-dimensional reduced space.

The number of retained PCs (A) can be derived by several methods, see, for example, Jackson[23]. One way is to extract the number of components that is needed to reproduce a specific fraction of the variance of the original data. Cross-validation, see Wold[25], is recommended when the PCA model is used to evaluate future data.

If all PCs are retained the original $\mathbf{X}$ matrix is given by:

$$\mathbf{X}=\mathbf{T}\mathbf{P}'=\sum_{a=1}^{m}\mathbf{t}_a\mathbf{p}_a' \tag{2}$$

Often, the variability in matrix $\mathbf{X}$ can be well approximated using a limited number of the PCs. Throughout this article, the term 'PCA model' is used to denote the approximation of the variability in $\mathbf{X}$ by the use of the A first PCs. We can then write:

$$\mathbf{X}=\mathbf{T}\mathbf{P}'=\sum_{a=1}^{A}\mathbf{t}_a\mathbf{p}_a'+\mathbf{E} \tag{3}$$

where the remaining PCs typically explain noise, summed up in a matrix of residuals $\mathbf{E}$.

The developed PCA model can be combined with tools and techniques from univariate SPC to form multivariate statistical process monitoring tools. There are many examples of control charts that can be used for process monitoring using projection methods. According to Kourti[12], a common and useful monitoring chart is the Hotelling's T^2 chart based on the A first PCs. The Hotelling T^2 chart checks if a new observation of the variables from the process projects onto the PC hyperplane within the limits determined by the reference data set. Let

$$T_A^2=\sum_{a=1}^{A}\frac{t_a^2}{s_{t_a}^2} \tag{4}$$

where $s_{t_a}^2$ is the variance of the elements in the ath score vector, $\mathbf{t}_a$, from the reference data set, and t_a^2 is the squared PC score of the ath PC for the new observation. An upper control limit based on the A first PCs and significance level α is given by:

$$T_{A,UCL}^2=\frac{(n^2-1)A}{n(n-A)}\cdot F_{\alpha(A,n-A)} \tag{5}$$

where n is the number of observations in the reference data set and $F_{\alpha(A,n-A)}$ is the upper αth percentile of the F distribution with A and $n-A$ degrees of freedom.

However, if a new type of special event occurs, which was not present in the observations used to build the reference PCA model, 'new' PCs can appear and the observations may move away from the initial hyperplane of PCs. Therefore, it is often recommended to combine the Hotelling's T^2 chart with a chart monitoring the squared prediction error (SPE) of the residuals of new observations, see Kourti[12], MacGregor and Kourti[14] and Kresta *et al.* [26]. Wikström *et al.*[27] use a similar measure of the a row-wise summary of the residual matrix $\mathbf{E}$ that they call the 'distance to the model' (DModX). The DModX measure is also applied by the online multivariate monitoring software used during the case study, see Umetrics[28]. Below, the DModX measure is explained according to the description by Eriksson *et al.* [29]. The DModX, s_i, for the ith observation is the residual standard deviation (RSD) of the observation computed as:

$$s_i=\sqrt{\frac{\sum_{j=1}^{m}e_{i,j}^2}{(m-A)}} \tag{6}$$

where $e_{i,j}^2$ is the squared residual with respect to the jth variable for the ith observation.

The DModX can be expressed in normalized units as the DModX of the observation divided by the pooled RSD of the reference model, s_0:

$$s_0=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}e_{i,m}^2}{(n-A-1)\cdot(m-A)}} \tag{7}$$

Eriksson *et al.*[29] claim (without proof) that the ratio s_i/s_0 is approximately F-distributed with $(m-A)$ and $(n-A-1)\cdot(m-A)$ degrees of freedom. This result is then used in the software to calculate a critical distance to the model for the normalized

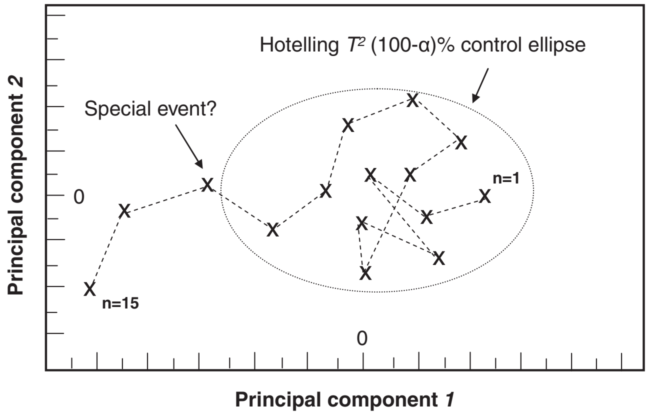

Figure 1. Example of a two-dimensional score plot (time-series plot) of observations projected onto the plane of principal components 1 and 2. The Hotelling T^2 control ellipse represents the limits of 'normal' process behavior

DModX at a specified significance level. The normalized DModX can then be plotted in a control chart together with the critical distance, see Wikström *et al.*[27].

The use of projection methods also makes it possible to monitor the process by, for example, following a time-series plot of PC scores in two dimensions, see Figure 1. A Hotelling T^2 'control ellipse' for future observations in the two-dimensional score plot can be used to determine if they come from a process in control. An observation outside the ellipse represents a possible out-of-control observation, see Johnson and Wichern [24, p. 248]. The axes of the ellipse in a score plot of, for example, the two first PCs are given by:

$$\left(s_{t_1}^2 \cdot F_{\alpha(2,n-2)} \cdot \frac{2(n^2-1)}{n(n-2)}\right)^{1/2} \quad \text{and} \quad \left(s_{t_2}^2 \cdot F_{\alpha(2,n-2)} \cdot \frac{2(n^2-1)}{n(n-2)}\right)^{1/2} \tag{8}$$

Assignable causes that are detected using latent variable methods such as PCA can be analyzed by contribution plots. Plots of the original process variables' contribution to, for example, the Hotelling T^2, DModX, or PC scores are helpful to analyze what has caused a specific event, see Kourti[12]. Hence, analyzing the changes in the original variables is important to be able to diagnose an event detected in a multivariate control chart.

3. The case: The EBF

To illustrate how experimental campaigns can benefit from multivariate monitoring, a case study was performed at the EBF plant in Luleå, Sweden. The blast furnace process can be characterized as a high temperature counter current reactor for reduction and smelting of iron ore into hot metal, see Geerdes *et al.*[30]. Coke and coal are used to reduce iron oxide, normally in the form of sinter and/or pellets, into liquid iron. The production capacity of the EBF is approximately 35 metric tons of hot metal per day (compared with up to 10 000 metric tons per day for the largest full-scale furnaces).

The blast furnace process is a dynamic continuous process with a multitude of measured process variables as well as product variables and hence a good candidate to study the benefits of multivariate monitoring. The research engineers at the EBF plant (hereafter the EBF engineers) looked to improve their decision-making abilities during experiments at the EBF and have collaborated with the author during the study.

The EBF was inaugurated in 1997 by Luossavaara-Kiirunavaara AB (LKAB), a Swedish producer of highly developed iron ore products (pellets in particular). The EBF is a pilot scale blast furnace, specifically designed for experimental purposes and intended mainly for product development, but also to improve knowledge about LKAB's customers' process—the blast furnace process. More information about experiments run in the EBF and special considerations that are needed when planning, conducting and analyzing experiments in the EBF can be found in Vanhatalo and Vännman[7] and Vanhatalo and Bergquist[8].

The pilot scale enables a realistic, controlled and safe way to conduct experiments and it is possible to create reactions and progress that can be expected of full-scale blast furnaces, but with much shorter times than those of commercial furnaces. The experimental costs and risks associated with performing experiments are substantially lower than in full-scale operation, but still large in the pilot scale. An outline of the EBF and examples of measurement possibilities are presented in Figure 2.

3.1. The control problem: The thermal state of the EBF process

An important problem during experiments in the EBF is to control the thermal state of the process and control actions during experimentation are unavoidable, see Vanhatalo and Bergquist[8]. Maintaining a good thermal state in the furnace is especially intricate as frequent changes of the experimental setups are made, for example, when running a factorial design. Without controlling the thermal state during experiments in the EBF, there is a risk for the melt to freeze or overheat, which can jeopardize

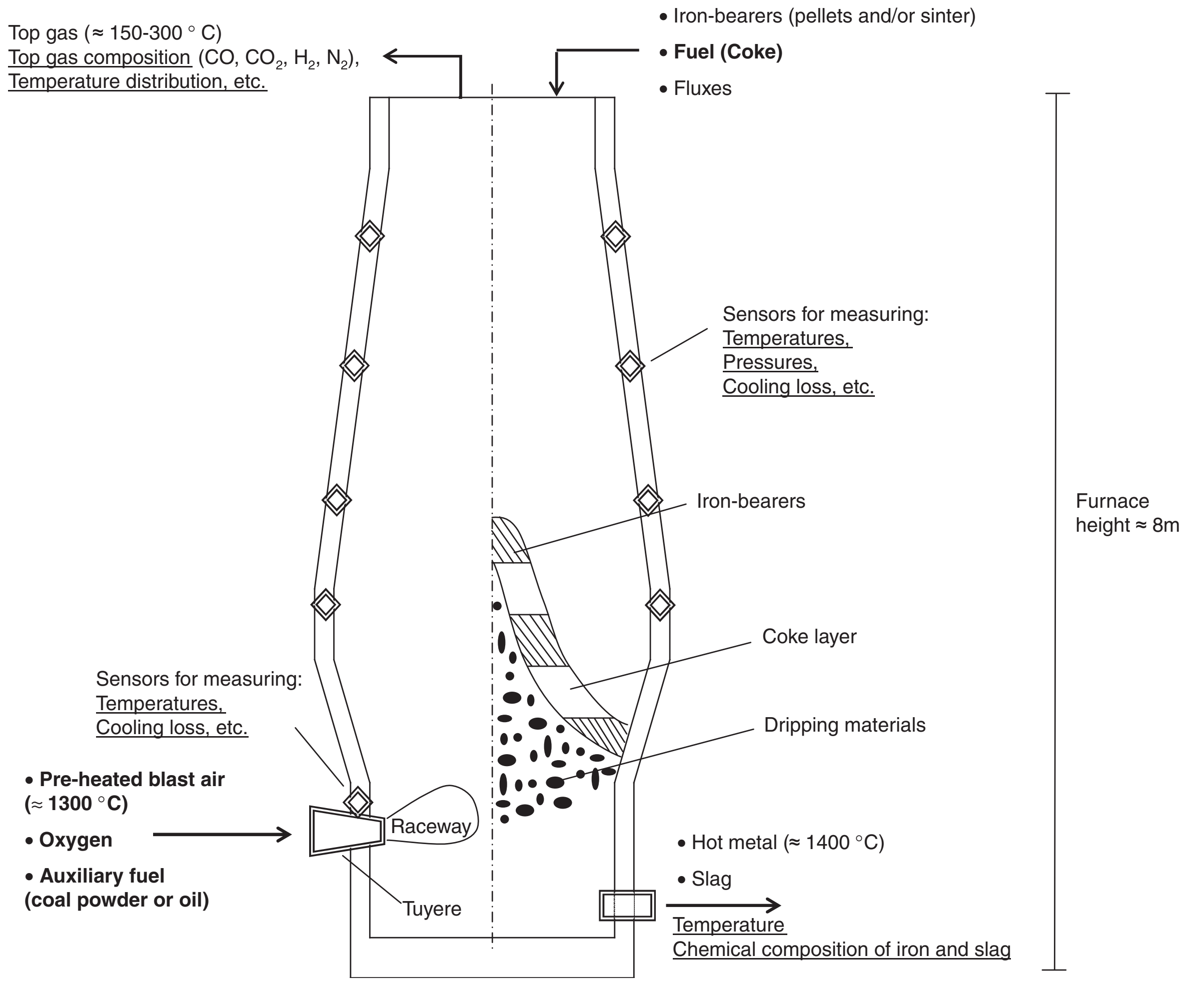

Figure 2. Outline of the EBF process. In the figure, examples of measured process and product variables are underlined. Variables possible to use for control of the thermal state of the process are indicated by bold font

the experimental campaign, the plant, as well as personal safety. The thermal state of the process will affect most response variables used to assess, for example, the performance of a new product in the blast furnace, which makes it even more critical to control.

Often, the thermal state of the blast furnace is depicted by the hot metal temperature, which is normally measured manually during tapping of the furnace. However, the manual measurements of hot metal temperature are subjected to significant measurement error. Hence, chemical analyses of the hot metal and slag are normally used to asses the thermal state. The levels of, for example, carbon, silicon and sulfur in the hot metal are usually good indicators of the thermal state in the process, see[31]. Furthermore, signals of changing thermal state in the furnace are often visible in process variables such as: gas utilization; cooling loss at the tuyeres; shaft temperatures and top gas temperatures.

The control of the thermal state in the EBF is difficult as it is manual and requires human intervention at certain processing states and the engineers need to take many variables into account before making a control decision. The control also includes a large but often unknown time lag due to process dynamics. The thermal state is normally controlled by increasing or reducing the coke rate in the burden mixture but can also be controlled by changing the auxiliary fuel rate, oxygen content of the blast air or the blast temperature, see Figure 2. The fuel rate during the experiment can also be an important response variable by itself although it is subjected to manipulation. This further stresses the need of an unbiased control.

Furthermore, many variables used as basis for control decisions are generated by off-line analysis of the chemical quality of the produced pig iron and slag. New observations of the chemical quality are normally available about once every hour. Hence, there is a built-in information lag for the control decision. Many variables further complicate the monitoring and control due to practical problems of simultaneously monitoring many univariate plots and weighing them together to form an idea of the thermal state in the process. Although larger shifts of the thermal state become evident in many univariate plots, early detection of minor shifts and trends in univariate plots is harder. Therefore, occurrence of 'cool furnace operation' is not uncommon during experimental campaigns in the EBF. Though the process can be brought back to a normal thermal state after a cooler period, a substantial portion of the experimental time usually needs to be disregarded, which is costly. During postexperimental analysis another problem is to try to resolve if the cooler periods depended on, for example, the tested pellets' performance in the furnace or on delayed thermal control decisions. Clearly, finding an unbiased, effective and time-efficient method to monitor the thermal state of the EBF process and a good control strategy is important to secure the quality of the experimental results. This is what motivated the multivariate approach described in the next section.

4. Multivariate monitoring of the thermal state in the blast furnace

This section describes how multivariate monitoring of the thermal state in the EBF was developed. Five main steps are described: choosing original variables representing the thermal state; building PCA models based on reference data sets; testing model performance; online monitoring and troubleshooting.

4.1. Variables used as indicators of the thermal state

The chemical quality of the pig iron and slag together with the hot metal temperature are normally the strongest signals of the thermal state in the lower parts of the furnace. Furthermore, there are many other process variables, such as those mentioned in Section 3.1 that can provide early signals of changing thermal state in the furnace. The process variables are measured online, often each second, while the chemical analyses of iron and slag are conducted by off-line lab tests available approximately once every hour. Because of the differing logging frequency of variables it was decided that two complementing PCA models were to be developed, see Figure 3. The 'tap data model', model 1, provides the strongest correlation to the thermal state of the furnace and the 'process data model', model 2, can provide early signals of unusual process behavior due to shifting thermal state or other disturbances.

This article will focus on PCA model 1 as it provides better signals of the thermal state in the process than PCA model 2. Table I presents the 10 product variables that were chosen, together with EBF engineers, to represent the thermal state of the process in the model. Each variable was judged to be correlated with the thermal state of the process.

4.2. PCA modeling of the thermal state

The next step is to model the thermal state of the process by building a PCA model based on the 10 variables in Table I. The PCA model building approach follows traditional SPC philosophy, see Kourti[12] and Mastrangelo *et al.*[15]. That is, an appropriate reference data set is selected to define normal operating conditions for the EBF process. The limits for normal operation on

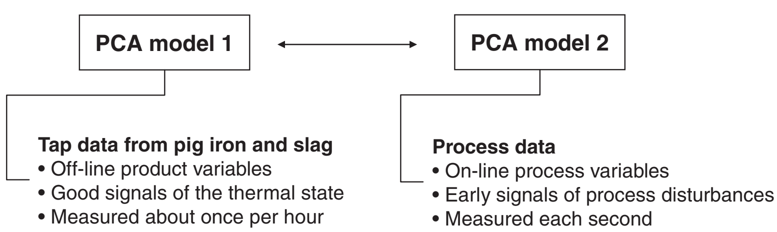

Figure 3. A summary of the two PCA models that were developed to monitor the EBF process

Table I. The ten variables used in the PCA model based on the chemical analysis of iron and slag

Variable	Units	Remarks	Thermal state (+)	Abbreviation
Hot metal temperature	°C	Measured manually	+	Temp
Chemical composition of pig iron	weight %	Off-line lab tests		
Carbon (C)			+	C
Silicon (Si)			+	Si
Sulfur (S)			−	S
Manganese (Mn)			+	Mn
Titanium (Ti)			+	Ti
Chemical composition of slag	weight %	Off-line lab tests		
Silicon dioxide (SiO_2)			−	SiO_2
Manganese oxide (MnO)			−	MnO
Sulfur (S)			+	S in slag
Titanium dioxide (TiO_2)			−	TiO_2

The (+) and (−) signs in the 'thermal state' column indicate how each variable normally is correlated to the thermal state. For example, increasing silicon content normally indicates a warmer process state.

Table II. Details of the PCA model built on 245 h of normal operation in the blast furnace			
Principal component	Explained variance	Cum. explained variance	Eigenvalue
1	0.447	0.447	4.47
2	0.264	0.711	2.64

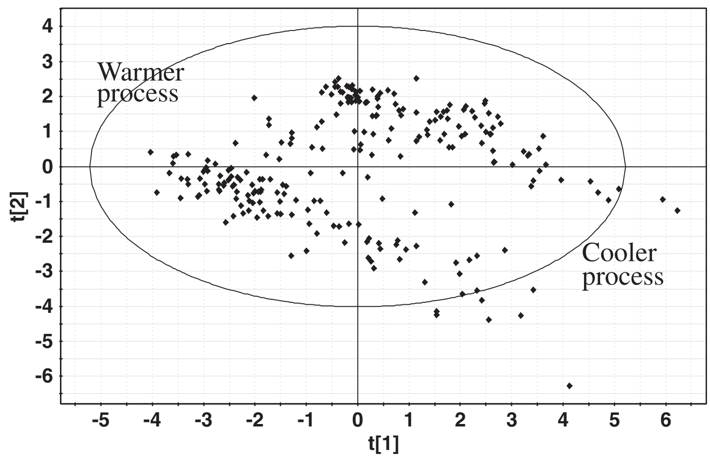

Figure 4. PC score scatter plot, $\mathbf{t}_1$ vs $\mathbf{t}_2$

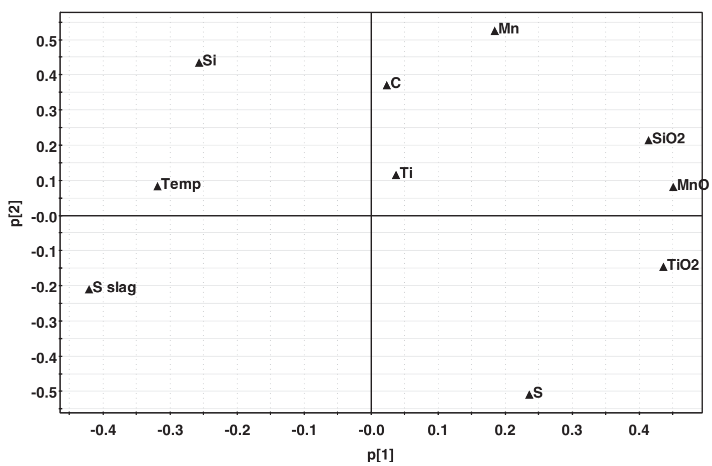

Figure 5. PC loading plot, $\mathbf{p}_1$ vs $\mathbf{p}_2$. See Table I for explanations of the variable abbreviations

the multivariate control charts are defined based on this reference set and future observations are compared with these limits. To secure good monitoring performance, any periods of disturbed furnace operation arising from special events (such as cool furnace operation) that one would like to detect in the future are removed from the reference data set.

Consider the first PCA model that was built using data from a previous experimental campaign in the EBF. Normal furnace operation during the first part of the campaign was used to build the PCA model, and the model's ability to monitor the thermal state during the rest of the campaign was studied. After excluding periods in the reference data representing cool furnace operation, 245 h of operation were left to represent normal process operation, see Figure 6. That is, approximately 22 per cent of data from the campaign were used to build the model, which was judged to be enough to capture the normal process variation and the correlation structure among the variables.

The 10 variables in Table I were first mean-centered and scaled to unit variance and then PCA was performed. Table II presents some results from the PCA model built on the 245 h of normal operation in the furnace. By cross-validation two PCs are derived, together explaining 71 per cent of the variation in the original data. Figure 4 presents a PC score scatter plot for the observations in the reference data set and Figure 5 presents the loadings for the first two PCs.

The loadings in Figure 5 describe the correlation between the original variables and the first two PCs. Figures 4 and 5 show that the variability of the thermal state in the reference data is visible in the reduced two-dimensional space of the first two PCs. Higher values on PC_1 together with lower values on PC_2 indicate a cooler process. Note, that the loadings in Figure 5 do not completely depict the correlation among the variables indicated in Table I as periods of cool furnace operation were removed to form the reference data set. Furthermore, the number of retained PCs as well as the 'warm' and 'cool' areas in the score plots can change somewhat depending on choice of reference data set. In Figure 4 two groups of observations may be observed in the score plot, which can be explained by decreased manganese content in the burden mixture during the reference period.

By using PCA, the dimensionality of the monitoring problem is reduced from 10 original variables to two PCs where the first describes the main variability of the thermal state (in this case).

4.3. Assessing the PCA model's monitoring performance

The PCA model was tested for the remainder of the experimental campaign to evaluate its performance; see Figure 6 for an overview of the campaign. The model performance was evaluated together with EBF engineers for five of the cool periods of the campaign, see Figure 6. The PCA model provided clear signals of cool furnace operation for all five periods. Logbooks were

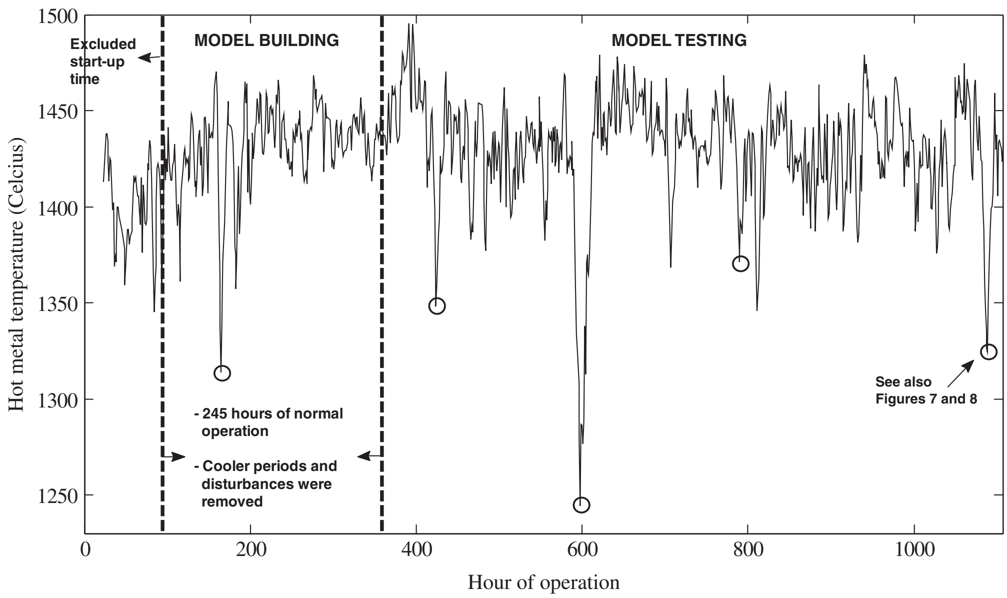

Figure 6. A simplified representation of the thermal state (only hot metal temperature) during the campaign. It shows the excluded 'start-up period' of the blast furnace, the 245 h of 'normal' operation used for model building, and the part used for model testing. Cooler periods of process operation where the PCA model performance was studied more closely are marked by circles

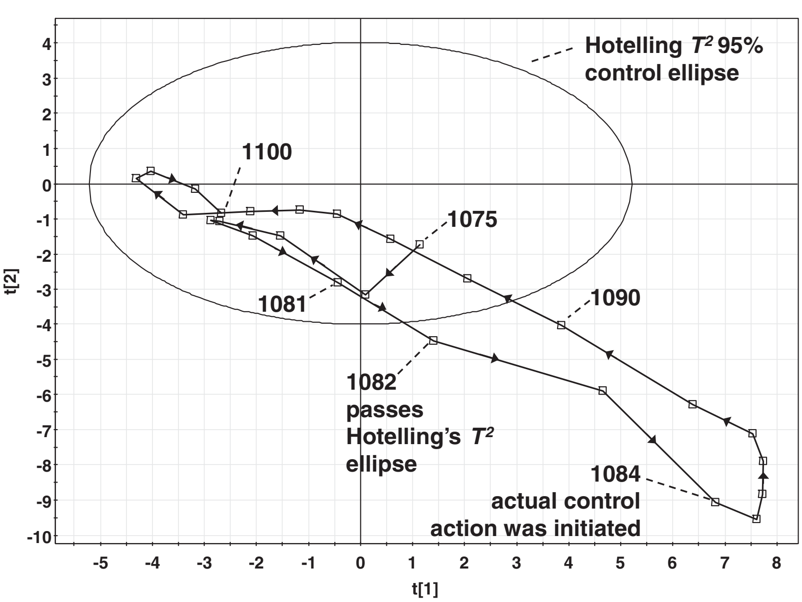

Figure 7. Score plot of PC_1 vs PC_2 for observations 1075–1100

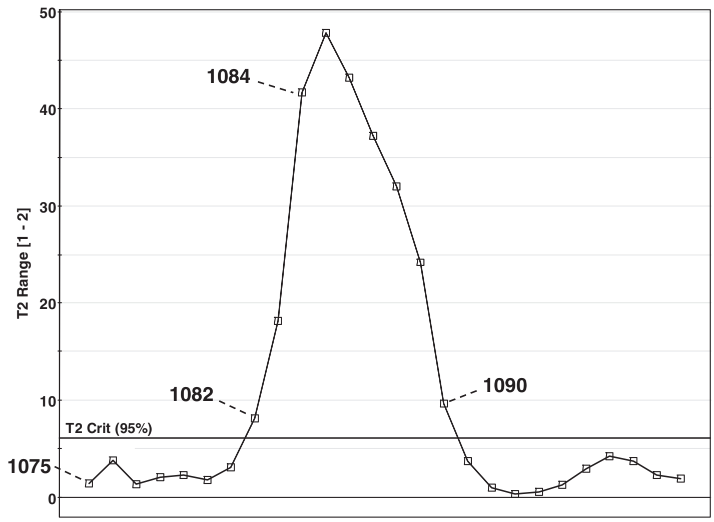

Figure 8. Hotelling's T^2 chart for the first two PCs for observations 1075–1100

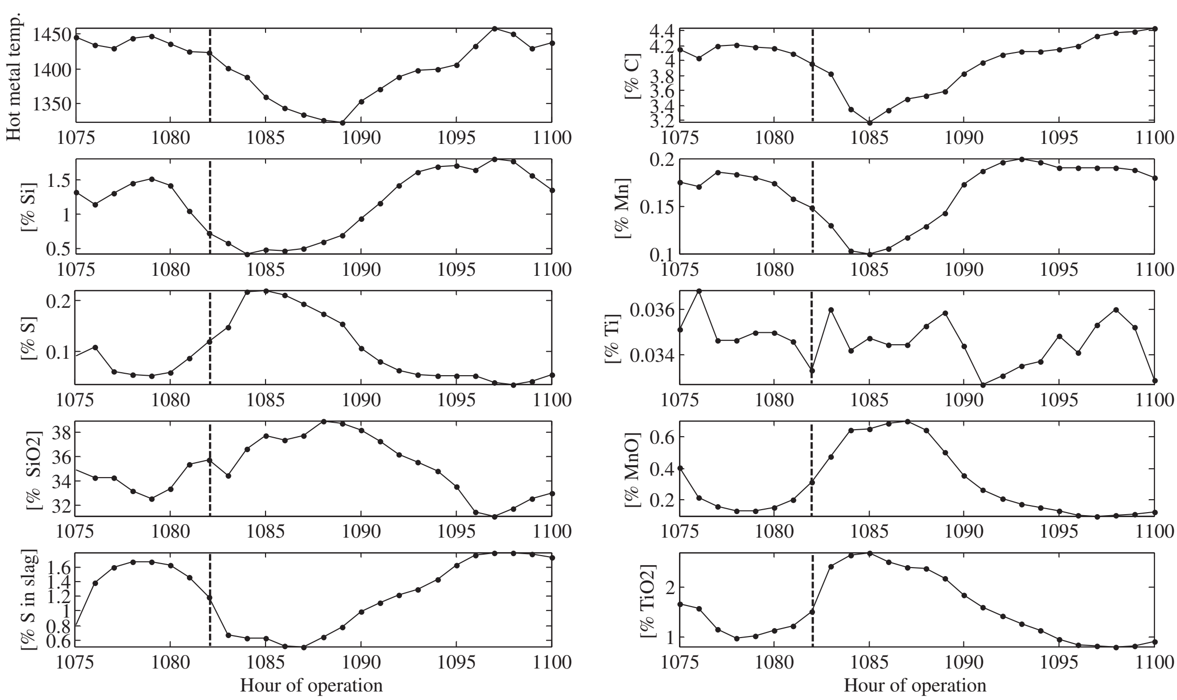

Figure 9. Ten univariate plots of the variables in the PCA model for observations 1075–1100. Observations 1082 is marked by a dashed line for comparison with Figures 7–8

studied to compare the actual control actions with the multivariate control charts' signals. On average actual control actions (based on univariate signals) were made about 1–2 h after unusual process behavior could be detected in, for example, the Hotelling's T^2 chart. As an illustration of a cool period in the campaign, the hours 1075–1100 in the campaign have been chosen to illustrate the PCA model's performance at the point 'far away' from the reference data set. Figures 7–9 show the corresponding score plot, Hotelling's T^2 plot, and univariate plots of the original variables. Note that the actual control action is made about two hours after the control charts in Figures 7–8 signal unusual behavior.

In Figures 7 and 8 we see that signals of non-normal process behavior become evident at about operation hour 1082. The point of unusual process behavior is not as easily determined when studying the 10 univariate plots of the original variables in Figure 9, which can explain that actual process control actions were decided on as late as observation 1084.

4.4. Online monitoring of the thermal state

A test with the PCA models for online monitoring of the EBF was made during an experimental campaign in the fall of 2008. Both PCA models (see Figure 3) were rebuilt using new data from the ongoing campaign and were then used to monitor the process. The models were updated as more campaign data became available. The PCA models were executed online using the Simca-4000 software[28]. For monitoring purposes, three charts, of those available in the software, were used:

- a time-series plot of PC scores in two dimensions,
- a Hotelling's T^2 chart based on all PCs in the model, and
- a control chart of the normalized DModX for the observations.

As an illustration, Figure 10 presents the online monitoring charts for one of the cooler periods in the experimental campaign using the tap data model (PCA model 1). Note that this is not the same model as described in Sections 4.2–4.3. Contribution plots were then used to diagnose deviating behavior in the multivariate charts. For example, Figure 11 displays the differences, in scaled units, for all the terms in the model, between the latest observation in Figure 10 and the average of the reference data set (model average) weighted by the loadings of components 1 and 2. Clearly, from Figure 11, the deviation in the multivariate charts is caused by: an increased magnesium oxide content in the slag; decreased sulfur content in the slag; increased sulfur content in the pig iron and so on. These are all signals of cool furnace operation.

During the trial, information from the multivariate monitoring was compared with information from prevailing monitoring charts such as univariate plots of important variables. Benefits from the multivariate monitoring mentioned by the EBF engineers during and after the trial are:

- that the multivariate charts provide a quick overview of the thermal state in the process or as they often say: *'a thermal index'*,
- that an index which takes many variables into consideration can provide a sense of comfort during process monitoring. It is hard to keep track of many univariate plots and different people tend to have their own favorite plots,
- and the multivariate charts can be a way of standardizing the way the thermal state is assessed, which is important to achieve unbiased control of the blast furnace during the experiments.

Another benefit of the multivariate monitoring connected to experiments run in the EBF process is that, for example, the multivariate control charts can be used to develop formal criteria for when to perform control actions. Hence, decisions of *when* to perform control actions can be made less subjective even though human deliberations still are needed to determine *what* the appropriate action is.

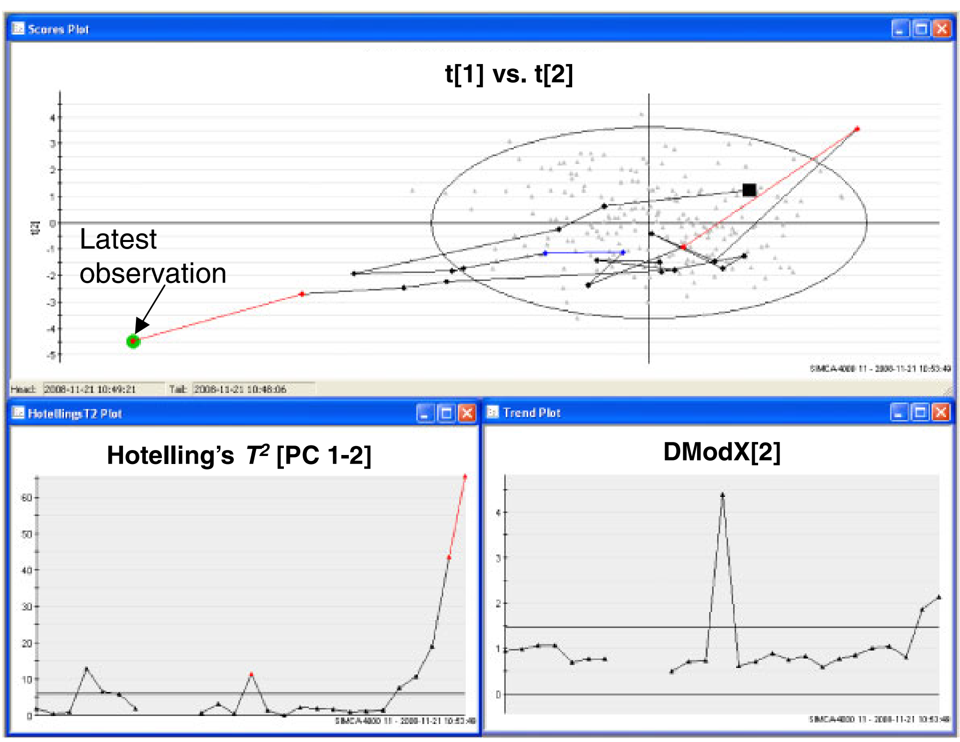

Figure 10. The process monitoring charts used in Simca-4000. The time period viewed in the charts represents one of the cooler periods in the process during experimental campaign. Note that the model used in these plots is not the same as in, for example, Figure 4. This figure is available in colour online at www.interscience.wiley.com/journal/qre

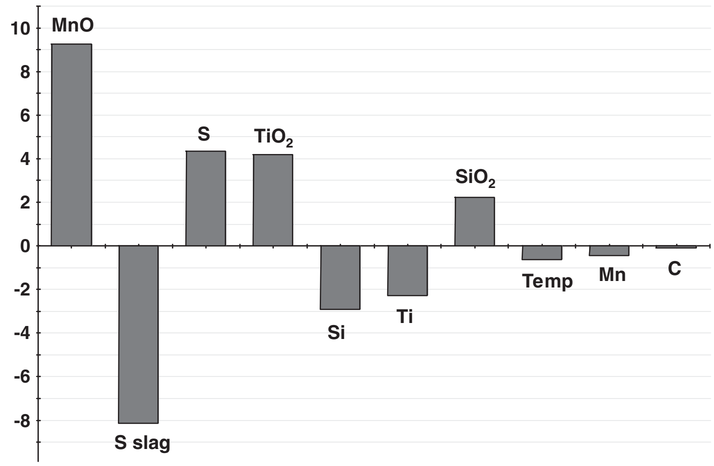

Figure 11. Principal component score contribution plot in scaled units, for all the terms in the model, between the latest observation in Figure 10 and the model average weighted by the loadings $\mathbf{p}_1$ and $\mathbf{p}_2$

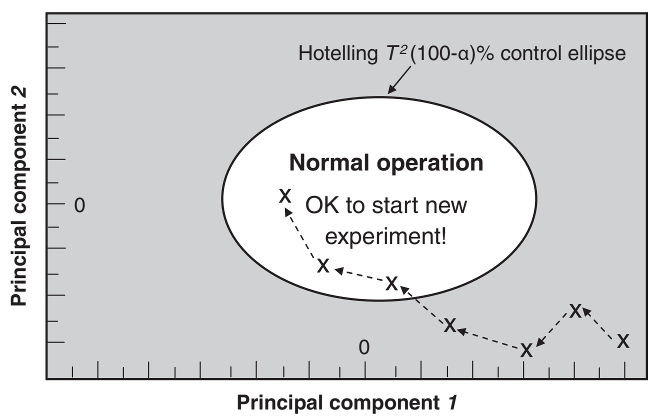

Figure 12. The idea of using principal component score plots to determine whether the EBF is operating normally or not. A new experiment should be started during normal operation

In addition, multivariate charts provide the engineer with a formal way of determining whether the process is operating normally, which is important during experiments and for postexperimental analysis. For example, the location of the process with respect to the Hotelling's T^2 ellipse or in the T^2 chart can be used as decision criteria whether or not to initiate a new experiment. A new experiment should be started when the process is performing normally to give it a fair chance and not bias the results due to prior disturbances in the EBF process, see Figure 12 for the general idea. Formal decision criteria of normal process operation are also important when choosing samples of the EBF process' performance to analyze after the experiment.

Drawbacks of the multivariate monitoring approach observed by the author and mentioned by the EBF engineers are:

- that development, maintenance and interpretation of the models require knowledge, training and time,
- that multivariate representations can be considered unnecessarily abstract compared with monitoring of variables that have a direct or indirect physical or chemical meaning to the engineer, and
- multivariate monitoring of a regularly changing process due to different experimental setups requires the multivariate models to be updated several times during an experimental campaign.

4.5. Troubleshooting—varying operating modes

A limitation of PCA-based monitoring of the EBF process is that once the PCA model is built it is time-invariant while the EBF process, as many other industrial processes, is time-varying. Changes in means; changes in variance; and changes in correlation structure often occur over time in industrial processes, see Li *et al.*[32]. The changes occur due to, for example, equipment aging, sensor drifts, process drifts and maintenance activities. Consider, for example, the process drift of increasing carbon content during one of the experimental campaigns in the EBF in Figure 13. If the PCA model is not updated to reflect such changes, it will soon cause frequent false alarms and poor monitoring performance.

In addition, the EBF is stopped, dug out and dismantled after each campaign and hence in a way a 'new process' is started at the beginning of each campaign. A typical experimental campaign is about 7–10 weeks long, which limits the time available to

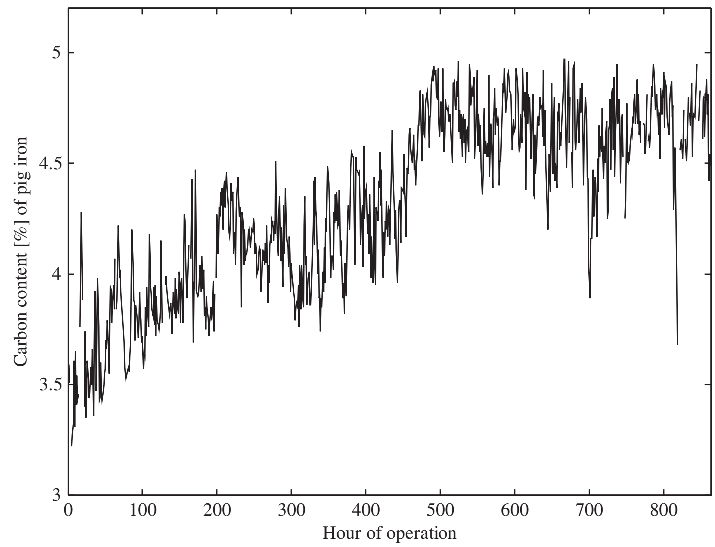

Figure 13. The carbon content during one of the experimental campaigns in the EBF

build the reference model. Furthermore, several experiments are typically run in each campaign with often quite different setups and operating modes. Hence, the most difficult part of getting the online model to perform well in the EBF is to maintain a relevant reference data set of normal process operation. Clearly, the PCA models need frequent recalibration due to the varying operating modes of the EBF.

Hwang and Han[21] discuss the problem of monitoring a process with multiple operating modes and propose using hierarchical clustering and a 'super PCA model' to monitor, for example, a blast furnace. However, their solution is not recommended for processes operating in a wide range of operating modes (like the EBF). Furthermore, the limited up-time of the EBF process (7–10 weeks) makes it hard to get good estimations of the many operating modes. Other feasible solutions to this problem are, according to Hwang and Han[21], to either develop a set of local monitoring models, each valid for the corresponding operating mode, or to adaptively update the model to the changes. As the EBF is an experimental pilot plant, the operating modes change from experiment to experiment and from campaign to campaign due to differing process setups depending on the experimental purpose. Hence, developing a limited number of local monitoring models is not realistic. Instead, the PCA models need to be adaptively updated. In the EBF case, two ways to update the PCA models were used as they were possible to combine with the monitoring software used:

1. Adaptation of PCA model averages online (model centering)
2. Manual calibration of the observations in the reference data set

For online adaptation to the slow-moving drifts in the EBF process, EWMA were used to continuously update the averages of the reference model:

$$\bar{\mathbf{x}}_{i+1}^{adj} = \phi \cdot \mathbf{x}_i + (1-\phi) \cdot \bar{\mathbf{x}}_i^{adj} \quad (9)$$

where $\bar{\mathbf{x}}_{i+1}^{adj}$ is a vector with the adjusted averages of the variables at time $i+1$, $\mathbf{x}_i$ a vector with the observed values at time i, $\bar{\mathbf{x}}_i^{adj}$ are the adjusted averages at time i, and the coefficient ϕ determines the rate of adaptation. Letting $\phi=0$ means no adaptation while $\phi=1$ means complete adaptation. According to Umetrics[28], reasonable values on ϕ usually lie between 0.01 and 0.05.

Sometimes more dramatic and faster changes of variable averages occur in the process such as when a new experimental setup takes effect in the EBF process or due to calibration of sensors. When these types of faster and more dramatic changes occur, manual calibration of the averages in the reference PCA model can be faster and more appropriate than the EWMA adjustment. Calibration of the averages can be made by either adjusting the averages in the reference data set or by calibrating the new observations:

$$\mathbf{x}_i^{cal} = \mathbf{x}_i + \boldsymbol{\delta} \quad (10)$$

where $\mathbf{x}_i^{cal}$ is the calibrated vector of the new observations, $\mathbf{x}_i$ is the new observation vector, and $\boldsymbol{\delta}$ is a calibration vector to adjust the new observations to the changed average of the process. $\boldsymbol{\delta}$ is calculated as:

$$\boldsymbol{\delta} = \bar{\mathbf{x}}_{ref} - \bar{\mathbf{x}}_{new} \quad (11)$$

where $\bar{\mathbf{x}}_{ref}$ is a vector with the averages for each variable from the reference data set and $\bar{\mathbf{x}}_{new}$ is a vector with new averages from, for example, observations from the new operating mode.

Indeed, adaptations of variable averages in the models are not always enough to cope with process changes that affect the variance or the correlation structure among the variables, such as significant changes of the operating mode of the EBF. Instead, the PCA models may need to be recalibrated by assigning new observations to the reference data set and removing older observations that are no longer relevant. A persistent trend of high and increasing DModX values and Hotelling's T^2 even though the process seems to be operating normally is usually a signal of an irrelevant reference model.

More sophisticated methods for adapting multivariate models to process drifts have been proposed in the literature. Different algorithms to recursively update the PCA model, each time a new vector of observations becomes available, have been proposed by, for example, Lane *et al.*[33], Li *et al.*[32] and Wold[34]. Essentially, these algorithms introduce an exponentially weighted moving window for the observations in the reference data set, where the influences of the most recent observations are the greatest. Such algorithms would be interesting to study in the future in the EBF case. However, these methods cannot be straightforwardly applied without answering important questions such as: how large should the exponential 'forgetting factor' be, can the exponential forgetting factor be increased for periods when experimental setups are changed to speed up model adaptation, and how are the recursive models prevented from adapting to faults and unusual process behavior while at the same time effectively adapted to slow-moving process drifts?

5. Conclusions and discussion

This article shows how a multivariate approach based on PCs can be applied to monitor the thermal state of an EBF process. Using the multivariate approach, the monitoring problem is reduced to monitoring plots of a few latent variables instead of many one-variable-at-a-time plots. The latent variables give a good overview of the thermal state in the process and thereby facilitate control of the furnace during ongoing experimentation. The multivariate model does not, however, eliminate the need for human deliberation about the proper control action to counteract unwanted process states but provides means of detection rules and analysis tools of the thermal state of the process. The multivariate charts provide the possibility to develop formal decision rules of when to perform control actions as well as a standardized way of evaluating the process state among different people, which is the starting point for an unbiased control action and in extension an unbiased experiment.

A multivariate model such as PCA can be a good alternative to monitor a process like the blast furnace process. The models do not require complete theoretical knowledge about what is going on inside the blast furnace at any given time, but they require good process knowledge. With access to empirical data in the form of periods of good and stable operation in the blast furnace, models can be set up to identify deviating operation and present which of the many original variables that carry important information. From a monitoring perspective this is an interesting alternative as knowledge of what exactly is going on inside the blast furnace can be something of a mystery sometimes. Critical success factors for the multivariate approach are to choose relevant variables that measure the phenomenon that you want to monitor, and to identify and assign a proper reference data set for normal process behavior.

Choosing relevant variables requires good process knowledge and the choice of reference data set is important for good model performance. A reference data set that underestimates the variation in the process compared with 'normal' operation will produce a too sensitive model while overestimation of the normal variability gives you a model that reacts slowly to significant changes of the process state. Assigning the proper reference data set for the PCA models is complicated in the EBF and for any continuous process that run short production campaigns or makes frequent changes to production setup conditions with frequently shifting operating modes as a result. Hence, frequent recalibration of the PCA models may be required, for example, by adapting the models to process drifts by continuously updating the averages of the original variables in the models or by updating the observations in the reference data set. From a blast furnace perspective, this is probably less of an issue in large full-scale furnaces where production is run under more stable process setups for longer times.

Thus, the multivariate approach to monitoring of the EBF has showed a promising potential to provide unbiased information and overview of the thermal state in the EBF. Ongoing efforts are made to implement the models for all experimental campaigns in the EBF and to develop more effective and efficient methods to update the models to frequent process shifts.

Acknowledgements

The financial support from the Swedish mining company LKAB and the European Union, European Regional Development Fund, Produktion Botnia is gratefully acknowledged. The author thanks all members of the LKAB 'EBF methodology development project' for their important contribution to the results presented here. Special thanks to Gunilla Hyllander at LKAB for valuable support. Furthermore, thanks to Bjarne Bergquist and Kerstin Vännman at Luleå University of Technology for valuable comments on earlier drafts and Conny Wikström at Umetrics AB for help with software issues.

References

1. MacGregor JF. Using On-line process data to improve quality: Challenges for statisticians. *International Statistical Review* 1997; **65**(3):309–323. DOI: 10.1111/j.1751-5823.1997.tb00311.x.
2. Dennis D, Meredith J. An empirical analysis of process industry transformation systems. *Management Science* 2000; **46**(8):1085–1099. DOI: 10.1287/mnsc.46.8.1085.12031.

3. Fransoo JC, Rutten WGMM. A typology of production control situations in process industries. *International Journal of Operations and Production Management* 1994; **14**(12):47–57.
4. Wise BM, Gallagher NB. The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control* 1996; **6**(6):329–348. DOI: 10.1016/0959-1524(96)00009-1.
5. Kourti T, MacGregor JF. Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemometrics and Intelligent Laboratory Systems* 1995; **28**:3–21. DOI: 10.1016/0169-7439(95)80036-9.
6. Hild C, Sanders D, Cooper T. Six sigma* on continuous processes: How and why it differs. *Quality Engineering* 2000; **13**(1):1–9. DOI: 10.1080/08982110108918618.
7. Vanhatalo E, Vännman K. Using factorial design and multivariate analysis when experimenting in a continuous process. *Quality and Reliability Engineering International* 2008; **24**(8):983–995. DOI: 10.1002/qre.935.
8. Vanhatalo E, Bergquist B. Special considerations when planning experiments in a continuous process. *Quality Engineering* 2007; **19**(3):155–169. DOI: 10.1080/08982110701474100.
9. Bisgaard S. Must a process be in statistical control before conducting designed experiments. *Quality Engineering* 2008; **20**(2):143–150. DOI: 10.1080/08982110701826721.
10. Kowalski SM, Parker P, Vining GG. Tutorial: Industrial split-plot experiments. *Quality Engineering* 2007; **19**(1):1–15. DOI: 10.1080/08982110601057179.
11. Bersimis S, Psarakis S, Panaretos J. Multivariate statistical process control charts: An overview. *Quality and Reliability Engineering International* 2007; **23**(5):517–543. DOI: 10.1002/qre.829.
12. Kourti T. Application of latent variable methods to process control and multivariate statistical process control in industry. *International Journal of Adaptive Control and Signal Processing* 2005; **19**:213–246. DOI: 10.1002/acs.859.
13. Qin SJ. Statistical process monitoring: Basics and beyond. *Journal of Chemometrics* 2003; **17**:480–502. DOI: 10.1002/cem.800.
14. MacGregor JF, Kourti T. Statistical process control of multivariate processes. *Control Engineering Practice* 1995; **3**(3):403–414. DOI: 10.1016/0967-0661(95)00014-L.
15. Mastrangelo CM, Runger GC, Montgomery DC. Statistical process monitoring with principal components. *Quality and Reliability Engineering International* 1996; **12**(3):203–210.
16. Höskuldsson A. PLS regression methods. *Journal of Chemometrics* 1988; **2**:211–228. DOI: 10.1002/cem.1180020306.
17. Miletic I, Quinn S, Dudzic M, Vaculik V, Champagne M. An industrial perspective on implementing on-line applications of multivariate statistics. *Journal of Process Control* 2004; **14**:821–836. DOI: 10.1016/j.jprocont.2004.02.001.
18. Morud TE. Multivariate statistical process control; example from the chemical process industry. *Journal of Chemometrics* 1996; **10**:669–675. DOI: 10.1002/(SICI)1099-128X(199609)10:5/6<669::AID-CEM467>3.0.CO;2-Q.
19. Miettinen T, Hurse TJ, Connor MA, Reinikainen S-P, Minkkinen P. Multivariate monitoring of a biological wastewater treatment process: A case study at Melbourne Water's Western Treatment Plant. *Chemometrics and Intelligent Laboratory Systems* 2004; **73**(1):131–138. DOI: 10.1016/j.chemolab.2003.11.012.
20. Warren P, Harvey S. Development and implementation of a generic blast-furnace expert system. *Transactions of the Institution of Mining and Metallurgy Section C: Mineral Processing and Extractive Metallurgy* 2001; **110**:43–49.
21. Hwang D-H, Han C. Real-time monitoring for a process with multiple operating modes. *Control Engineering Practice* 1999; **7**:891–902. DOI: 10.1016/S0967-0661(99)00038-6.
22. Gamero FI, Colomer J, Meléndez J, Warren P. Predicting aerodynamic instabilities in a blast furnace. *Engineering Applications of Artificial Intelligence* 2006; **19**:103–111. DOI: 10.1016/j.engappai.2005.05.006.
23. Jackson JE. *A User's Guide to Principal Components*. Wiley: Hoboken, NJ, 2003.
24. Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis*. Prentice-Hall: Upper Saddle River, NJ, 2002.
25. Wold S. Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* 1978; **20**(4):397–405.
26. Kresta JV, Macgregor JF, Marlin TE. Multivariate statistical monitoring of process operating performance. *The Canadian Journal of Chemical Engineering* 1991; **69**:35–47.
27. Wikström C, Albano C, Eriksson L, Fridén H, Johansson E, Nordahl Å, Rännar S, Sandberg M, Kettaneh-Wold N, Wold S. Multivariate process and quality monitoring applied to an electrolysis process. Part I. Process supervision with multivariate control charts. *Chemometrics and Intelligent Laboratory Systems* 1998; **42**:221–231. DOI: 10.1016/S0169-7439(98)00014-8.
28. Umetrics. *Users Guide to SIMCA-4000: Version 11, Online Multivariate Modeling*. Umetrics AB: Umeå, Sweden, 2006.
29. Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C, Wold S. *Multi- and Megavariate Data Analysis: Part I Basic Principles and Applications*. Umetrics AB: Umeå, Sweden, 2006.
30. Geerdes M, Toxopeus H, van der Vliet C. *Modern Blast Furnace Ironmaking: An Introduction*. Verlag Stahleisen GmbH: Düsseldorf, 2004.
31. Zuo G. *Improving the Performance of the Blast Furnace Ironmaking Process*. Department of Chemical and Metallurgical Engineering, Luleå University of Technology: Luleå, 2000.
32. Li W, Yue HH, Valle-Cervantes S, Qin SJ. Recursive PCA for adaptive process monitoring. *Journal of Process Control* 2000; **10**(5):471–486. DOI: 10.1016/S0959-1524(00)00022-6.
33. Lane S, Martin EB, Morris AJ, Gower P. Application of exponentially weighted principal component analysis for the monitoring of a polymer film manufacturing process. *Transactions of the Institute of Measurement and Control* 2003; **25**(1):17–35. DOI: 10.1191/0142331203tm071oa.
34. Wold S. Exponentially weighted moving principal component analysis and projections to latent structures. *Chemometrics and Intelligent Laboratory Systems* 1994; **23**:149–161.

Author's biography

Erik Vanhatalo is a PhD student in the subject of Quality Technology and Management at the Luleå University of Technology, Luleå, Sweden. He holds a Licentiate degree of Engineering in the subject of Quality Technology and Management and a Master's degree in Industrial and Management Engineering from Luleå University of Technology. His research is focused on combining experimental design and multivariate statistical methods in continuous processes. He is a member of the European Network for Business and Industrial Statistics (ENBIS).