Journal of Process Control 50 (2017) 56–65

Contents lists available at ScienceDirect

Journal of Process Control

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

An integral approach to inferential quality control with self-validating soft-sensors

José L. Godoy[a,b,*], Jacinto L. Marchetti[a], Jorge R. Vega[a,c]

[a] INTEC (CONICET and Universidad Nacional del Litoral), Güemes 3450, 3000, Santa Fe, Argentina
[b] FRP-UTN (Facultad Regional Paraná – Universidad Tecnológica Nacional), Av. Almafuerte 1033, 3100, Paraná, Argentina
[c] FRSF-UTN (Facultad Regional Santa Fe – Universidad Tecnológica Nacional), Lavaisse 610, 3000, Santa Fe, Argentina

ARTICLE INFO

Article history:
Received 16 July 2015
Received in revised form 20 July 2016
Accepted 2 December 2016

Keywords:
Integral design space
Multivariate quality control
Partial least squares
Self-validating soft-sensor
Fault detection and diagnosis
Styrene-butadiene rubber (SBR)

ABSTRACT

This paper presents an integral technique for designing an inferential quality control applicable to multivariate processes. The technique includes a self-validating soft-sensor and a multivariate quality control index that depends on the specifications. Based on a partial least squares (PLS) decomposition of the online process measurements, a fault detection and diagnosis technique is used to develop an improved self-validation strategy that is able to confirm, correct or reject the soft-sensor predictions. Model extrapolations, disturbances or sensor faults are first detected through a combined statistic (that considers the calibration region); then, a diagnosis is made by combining statistics pattern recognition, contribution analysis, and disturbance isolation based on historical fault patterns. An off-spec alarm is produced when the proposed index detects that an operating point lies outside the integral design space driven by the specifications. The effectiveness of the proposed technique is evaluated by means of two numerical examples. First, a synthetic example is used to interpret the fundamentals of the method. Then, the technique is applied to the industrial Styrene-Butadiene rubber process, which is emulated through an available numerical simulator.

1. Introduction

In most industrial processes, the automatic control systems (or even the operators) adjust the manipulated variables with the aim of fulfilling several goals, e.g. to maintain the product quality within specifications, to limit the waste or contaminating emissions according to government regulations, and to keep the process close to an optimal operation condition that is normally established from both economical and technical point of views. An adequate driving of the manipulated variables is clearly conditioned to the availability of accurate and fast measurements. Typically, the process measurements are available from either online analyzers or offline laboratory equipments. Unfortunately, many important process variables cannot accurately be measured in relatively short times, thus negatively affecting real-time control strategies. In such cases, soft-sensors can help to overcome the problem. A soft-sensor is a mathematical code able to infer some unmeasured variables from a set of online measured variables. In the last years, soft-sensor applications have brought significant attention in the process industry [1–5]. In particular, several techniques based on partial least squares (PLS) have been used for monitoring complex industrial processes where the quality variables are important [6,7].

The efficiency of a soft-sensor as a predictive tool depends on several factors, such as: i) the accuracy of the process model that was used to derive the soft-sensor; ii) the proper adjustment of the soft-sensor to the actual process operating point; and iii) the availability of adequate online measurements. In general, process disturbances and sensor faults can strongly modify the correlated measurements, and therefore the quality predictions become unreliable. To overcome these difficulties, several strategies have been reported in the literature [8,9]. For example, Liu et al. [8] proposed a PLS-based soft-sensor with self-validation and reconstruction of faulty readings that improve the reliability of the predictions. The strategy consisted in an initial validation of the measurements prior to predicting the quality variables through the soft-sensor. The detection of a faulty sensor was followed by a reconstruction of the corresponding faulty readings; however, the identification of the faulty sensors was not fully reliable [8,10].

A fault detection and diagnosis strategy applicable to multivariate processes typically includes three main tasks: 1) the detection of an anomaly or an out-of-control condition; 2) the classification

* Corresponding author at: INTEC (CONICET and Universidad Nacional del Litoral), Güemes 3450, Santa Fe, 3000, Argentina.
E-mail address: jlgodoy@santafe-conicet.gov.ar (J.L. Godoy).

http://dx.doi.org/10.1016/j.jprocont.2016.12.001
0959-1524/

of the fault that generated the abnormal behavior, and 3) the isolation of the disturbed variables and ideally of the disturbing sources. Godoy et al. [11] proposed a PLS-based fault detection and diagnosis technique for multivariate processes that assumes available online measurements of the quality variables. According to such approach, the projections of the process measurements onto the latent space induce a PLS-decomposition of such measurements into four non-overlapped subspaces, and then, a combined index can be used to detect the process anomalies. The technique allows for an efficient anomaly classification as well as the identification of the disturbed variables. The pattern of the statistics that compose the detection index can be used to classify the anomaly type. However, no specification-dependent control limits were included for those statistics, as it is normally required for establishing quality control strategies.

Contribution plots are tools typically used for the identification of sources of faults without requiring any prior fault information [7,10]. On the basis of a principal component analysis (PCA) model, Alcala and Qin [12] have proposed a reconstruction-based contribution (RBC) technique to diagnose process faults. This technique inherits the merit of traditional contribution plots and has a solid theoretical foundation for detecting faulty sensors without smearing problems. However, RBC is only useful to isolate those failed sensors that do not cause the fault propagation to other variables, while it is ineffective for complex faults such as process faults and disturbances [10]. On the other hand, the angle between the vectors corresponding to the measurements and the fault signatures has been used to isolate complex faults [2,13]. Generalized RBCs based on fault subspaces have also been used to isolate know process faults [14], but need several faulty samples for extracting each fault subspace. In contrast, the angular measures have the advantage of only requiring a single faulty sample of each know fault to implement the diagnosis stage. It has been proven that the diagnosis results are the same for both approaches, because the ratio "generalized RBCs/detection index" is equal to the angular measure given by the squared cosine of the angle between the current PCA-projection and the faulty PCA-projection [10].

The role of design spaces (or multivariate specification regions) is undoubtedly important in several processes, and aims at reducing statistical variations in the final product quality by design rather than by inspection. The International Conference on Harmonization Q8 (ICH-Q8) document [15] defines a design space as "the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality." The current trend is towards defining multivariate specifications in the low dimensional subspace defined by a PLS model [16,17]. An integral design space should include a quality driven specification for the raw materials to be used in the process under certain operating conditions. Such specification will have to account for the inherent variability in the process and the combined effect of incoming materials with process conditions onto product quality [18].

Control of industrial polymerization processes is difficult due to the lack of sensor devices capable of providing with accurate online measurement of most quality variables [19]. Currently, there are several applications of soft-sensors in polymerization processes (e.g., Gonzaga et al. [20]). In particular, Godoy et al. [21] have developed a PLS soft-sensor capable of monitoring the production of Styrene-Butadiene rubber (SBR) in an industrial train of 7 continuously-stirred tank reactors. However, the presence of disturbances or sensor faults can turn unreliable the soft-sensor predictions.

In this work, an integral technique for inferential quality control is presented. To this effect, a PLS model is used to define an integral design space driven by quality specifications that accounts for the relationships between incoming materials, process conditions and product quality. Based on this integral design space, a statistical index is proposed for quality control. Additionally, this work presents a self-validation strategy for the inferences provided by a soft-sensor that estimates quality variables in a multivariate process (see Fig. 1). Such strategy is based on a fault detection and diagnosis method previously designed for processes with online measurable outputs [11], and includes: i) an extrapolation control limit, ii) pattern analysis of the statistics that compose a fault detection index, iii) RBC analysis for the identification of the contributing variables or faulty sensors, and iv) a disturbance isolation method based on the angle differences between current measurements and the historical disturbances. The proposed self-validation strategy contributes to improve the soft-sensor inferences by adding a procedure able to alarm the presence of extrapolations, disturbances or sensor faults, and to eventually correct faulty readings. The effectiveness of the proposed integral technique for inferential quality control is first demonstrated through a numerical example. Then, a self-validation strategy is developed for the simulator of an industrial SBR process; and in such a sense, the present work extends the applicability of the soft-sensor developed by Godoy et al. [21].

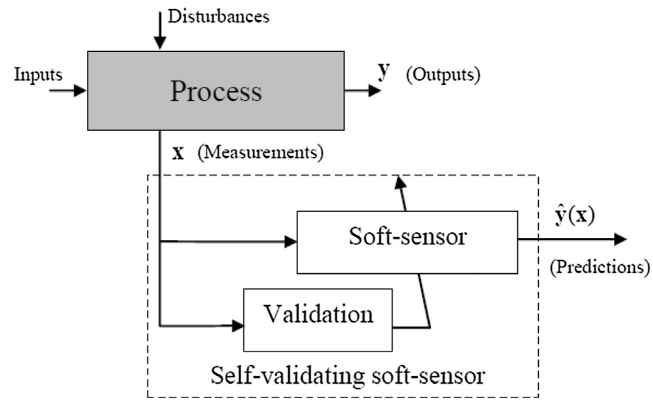

Fig 1. Inferential quality control with a self-validating soft-sensor. The soft-sensor predictions ($\hat{\mathbf{y}}$) are confirmed, corrected, or rejected by the proposed validation strategy.

2. Inferential quality control with self-validating soft-sensor

For a given multivariate process, call $\mathbf{x} = [x_1 \ldots x_m]' \in \Re^m$ the vector of online measurements and $\mathbf{y} = [y_1 \ldots y_p]' \in \Re^p$ the vector of quality variables. Both $\mathbf{x}$ and $\mathbf{y}$ are standardized vectors (i.e., mean-centered and scaled). Assume that N offline measurements of each variable were collected while the process was operating under normal conditions. Then, the following extended PLS regression model can be derived [11]:

$$\mathbf{x} = \mathbf{P}\mathbf{t} + \tilde{\mathbf{x}}, \mathbf{y} = \mathbf{Q}\mathbf{u} + \tilde{\mathbf{y}}, \tag{1}$$

$$\mathbf{u} = \mathbf{B}\mathbf{t} + \tilde{\mathbf{u}}, \tag{2}$$

where $\{\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{u}}\}$ are the model residuals, and the latent vectors $\mathbf{t}$ and $\mathbf{u}$ are respectively calculated from $\mathbf{x}$ and $\mathbf{y}$, as follows:

$$\mathbf{t} = \mathbf{R}'\mathbf{x}, \mathbf{u} = \mathbf{S}'\mathbf{y}. \tag{3}$$

The matrices $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{S}$, and $\mathbf{B}$, are obtained through the PLS-NIPALS algorithm [11,22,23]. This technique implicitly assumes that both $\mathbf{x}$ and $\mathbf{y}$ are online measured, and projects the measurement vectors into low-dimension spaces defined by A latent variables which are then regressed.

In this work we assume that the quality variables, $\mathbf{y}$, are not online available. On the basis of Eqs. (1), (2) and (3), it is possible to