Analytica Chimica Acta 642 (2009) 206–211

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Uncertainty and periodic behavior of process derived from online NIR

Maaret Paakkunainen*, Jarno Kohonen, Satu-Pia Reinikainen

Lappeenranta University of Technology, Department of Chemical Technology, P.O. Box 20, FIN-53851 Lappeenranta, Finland

ARTICLE INFO

Article history:
Received 29 July 2008
Received in revised form 28 October 2008
Accepted 19 November 2008
Available online 25 November 2008

Keywords:
Uncertainty
Periodic behavior
NIR measurements
Variographic analysis
Autocorrelation
Fast Fourier transformation

ABSTRACT

Past years have shown that near infra-red (NIR) can be successfully applied in online process control. The NIR measurements are commonly utilized because they are fast, versatile and relatively cost-effective. The online instruments produce an enormous amount of data, which need to be analyzed for, e.g., reliability, like any other online data. Instrumental data containing huge amount of simultaneously determined variables is multivariate in nature, and it has to be taken into account when the data is analyzed. The aim of this study was to show that variographic analysis gives a novel insight to online NIR data and the total uncertainty including variation arising from process itself can be estimated. It will be shown, that variographic analysis can be utilized in monitoring the process dynamics, as well as, in optimization of sampling interval.

The periodic behavior was identified with autocorrelation and fast Fourier transformation (FFT) as well as with the variographic analysis. However, the variographic analysis gave a more detailed insight to the process dynamics and enabled estimation of uncertainty as a function of sampling interval. These approaches are illustrated with real industrial data originating from a petrochemical plant. Similar periodic behavior could be detected by applying any of the three mathematical methods to the online variable sets containing either NIR or other process control variables. The total uncertainty of the NIR data was estimated by applying variographic analysis with an assumption that the different principal components (PC) are individual "error sources" causing uncertainty.

1. Introduction

The use of near infra-red (NIR) technology has rapidly been spread across the industry as a continuous process control tool due to its cost-effectiveness and high capacity for accurate online process control and to the many other advantages it offers. Online installations are usually the most challenging ones of the NIR installations, and in practice there exists several possible sources of uncertainty. The instrument is typically placed close to the monitored process requiring that the instrument withstands the harsh conditions of the process environment. Frequently, these conditions include extreme temperature conditions, high humidity, high dust levels, etc. The sampling is done by a probe that is directly inserted into the process line, and the design of the probe should be such that the probe will have the ability to withstand the process conditions and provide reliable information during the cleaning and calibration cycles.

The online NIR measurements establish time series, which are most commonly utilized in prediction of a quality parameter related to sampled material or the process flow [1–2]. In several instruments there exist some inbuilt algorithms, which are utilized to assure robustness of the instrument. However, the data can be analyzed further to gain advantageous information about the process or instrumental behavior. The spectral data could be analyzed for, e.g., autocorrelation structure, drifting or periodic behavior by applying fast Fourier transformation (FFT), autocorrelation or the variographic analysis. These mathematical methods are applied widely in analyzing univariate data, but they can also be utilized in analyzing multivariate data, such as spectra. For instrumental data of this kind, the estimation of the periodic behavior might be difficult without multivariate methods, because of the noise and the number of collinear variables.

The instrumental data of this type is often collinear having low rank and to handle the data the multivariate methods, such as principal component analysis (PCA) and partial least squares (PLS), are found essential. In the approaches presented here PCA is utilized to extract the information from online NIR data for the investigation of periodicity and uncertainty. The data analyses of the behavior in time domain and the total uncertainty, i.e., combined standard uncertainty, are based on principal components instead of the original data. These score vectors reducing the noise of original data represents each a certain amount of the total variance, described by R^2 or eigenvalues of the PCs, and thus they are handled as independent error components or sources in the variographic analyses.

* Corresponding author. Tel.: +358 5 621 2263.
E-mail address: maaret.paakkunainen@lut.fi (M. Paakkunainen).

0003-2670/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2008.11.037

The online processes are always dynamic, and the data tend to have some inner correlation structure in time domain, as well. This autocorrelation structure should be taken into account when total uncertainty of the measurements is analyzed. Variographic analyses presented by Pierre Gy are applicable in autocorrelated data of this kind. From the measurement point of view, it is often argued, that one of the main advantages that the NIR method offers is the high precision, which can be higher than that of the primary analytical method, and the speed (several quality variables can be measured in the order of seconds). In that case the measure of precision is the standard errors reported for the NIR measurements typically based on the calibration samples, which are used during the calibration development. However, the online NIR technology is utilized to process control and the measurements reflects the changes in production. Changes in standard deviation or in other words, in process variation, can be monitored via the NIR instruments as well. With variographic analysis the total standard uncertainty including uncertainty originating with heterogeneity of the process flow and conditions is analyzed.

2. Methods

2.1. Variographic analysis

Variographic analysis is primarily applied when the uncertainty of a process variable or its optimal sampling interval is estimated. It is also commonly utilized to identify the variability of the process. Theory of variographic analysis is described in details in several publications introducing the Pierre Gy's sampling theory [3–5].

In the variographic analysis, heterogeneity, h_i, i.e. the relative fluctuation from the process mean, is defined as in Eq. (1).

$$h_i = \frac{a_i - a_L}{a_L}\frac{M_{s_i}}{\bar{M}_s}, \quad i = 1, 2, \ldots, N \tag{1}$$

where i is the index of a sample; a_i is the analytical result; a_L is the weighted mean of the lot; M_{s_i} is the size of sample i; $\bar{M}_s$ is the mean sample size (or flow rate); N is the total number of samples.

From the heterogeneity, an experimental variogram can be estimated. The variogram expresses correlation between any two heterogeneity values, i.e. autocorrelation of the function. The variogram is estimated based on the differences between the heterogeneities as in Eq. (2).

$$V_j = \frac{1}{2(N-j)}\sum_{i=1}^{N-j}(h_{i+j} - h_i)^2, \quad j = 1, 2, \ldots, N/2 \tag{2}$$

For the optimal frequency of sampling, a Y-intercept of the variogram is needed. In this case study, the intercept is estimated graphically. The possible periodic behavior, as well as random drifts in the data, can be observed from the variogram. When the variance of the sampling is estimated, the variogram has to be integrated as in Eq. (3).

$$W_j = \frac{1}{j}\int_0^j V_j dj = \frac{1}{j}S_j \tag{3}$$

The sample selection methods utilized in the variographic analysis are systematic sampling and stratified sampling. In the systematic sampling the lot to be sampled is first divided into equal sizes sub-lot and from each sub-lot one sample is taken systematically. In the stratified sampling strategy, the lot is also first divided into equal sizes sub-lots, but from each sub-lot, one random sample is taken. In general, systematic sampling leads to smallest uncertainty estimates. However, the stratified sampling strategy should be utilized, if systematic periodic behavior is found. Periodic behavior has also been taken into account, when an optimal sampling frequency is considered. The variance of the stratified sampling can be expressed as in Eq. (4) and the variance of the systematic sampling as in Eq. (5).

$$s_{st}^2 = \frac{2}{j^2}\int_0^j S_j \, dj \tag{4}$$

$$s_{sy}^2 = 2W_{j/2} - s_{st}^2 \tag{5}$$

2.2. Autocorrelation

Autocorrelation is the correlation of a signal with itself at different points in time. It is widely utilized in finding repeating patterns in a signal, i.e., in this particular case in the score vectors or in the original process variables. Autocorrelation is simply computed as correlation of the vector against its time-shifted values. Theory of autocorrelation is well-explained in the literature [6–7].

2.3. Fast Fourier transformation

The Fourier transformation is a powerful technique that is commonly utilized in detecting periodicity, patterns and tendem repeats in various data sequences. Its' ability to represent time domain data in the frequency domain and vice versa has also numerous other applications. FFT is an efficient method for computing the discrete Fourier transform (DFT), and was first developed by Cooley and Tukey [8]. The algorithm was later refined for even greater speed and for using with different data lengths through the "mixed-radix" algorithms. The theory of FFT is explained in detail in several mathematical handbooks, for example Ref. [9].

The Fourier transformation maps time domain functions into frequency domain representations and is defined as in Eq. (6).

$$X(f) = F\{x(t)\} = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt \tag{6}$$

where $x(t)$ is time domain signal; $X(f)$ is its Fourier Transform.

Similarly, the discrete Fourier transform maps discrete–time sequences into discrete–frequency representations and is given by Eq. (7).

$$X_k = \sum_{i=0}^{n-1} x_i e^{-j2\pi ik/n}, \quad \text{for } k = 0, 1, 2, \ldots, n-1, \tag{7}$$

where x is input sequence; X is its drift; n is the number of samples in both the discrete–time and the discrete–frequency domains.

In this study FFT with a periodogram and power spectrum has been utilized together with the other two mathematical methods revealing periods in the multivariate or in the original data vectors. In FFT windowing multiplies input data supplied to the FFT with a value that decreases to zero at each end of data. To investigate the cycles or the periodicity in the data periodograms and power spectrum has been computed with Matlab [10].

The power spectrum $S_{xx}(f)$ of time domain signal $x(t)$ is defined as in Eq. (8).

$$S_{xx}(f) = X(f)X*(f) = |X(f)|^2 \tag{8}$$

where $X(f)$ is $F\{x(t)\}$; $X^*(f)$ is complex conjugate of $X(f)$.

A complex magnitude squared of $X(f)$ is called the power. A plot of power versus frequency is called periodogram. However, since a scale cycles/(sample interval) is found inconvenient, the power is studied as (sampling interval)/cycle.