Journal of Chromatography A, 1213 (2008) 130–144

Contents lists available at ScienceDirect

Journal of Chromatography A

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

# Dynamic analysis of on-line high-performance liquid chromatography for multivariate statistical process control

Sila Kittiwachana[a], Diana L.S. Ferreira[a], Louise A. Fido[b], Duncan R. Thompson[b], Richard E.A. Escott[c], Richard G. Brereton[a,*]

[a] Centre for Chemometrics, School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, UK
[b] GlaxoSmithKline, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, UK
[c] GlaxoSmithKline, Old Powder Mills, Tonbridge, Kent TN11 9AN, UK

ARTICLE INFO

*Article history:*
Received 9 July 2008
Received in revised form
22 September 2008
Accepted 26 September 2008
Available online 1 October 2008

*Keywords:*
High-performance liquid chromatography
Multivariate statistical process control
Continuous processes
Reaction monitoring

ABSTRACT

A continuous process was studied over 83.32 h using on-line high-performance liquid chromatography, involving the acquisition of 252 chromatograms. A method for analysis of these data using multivariate statistical process control on peak tables, in real-time, is described. The normal operating condition (NOC) region of the process was identified using evolving principal components analysis to be between 5.77 and 8.13 h. 19 out of the 37 peaks detected throughout the process were found in the NOC region, the remainder representing undesirable contaminants found elsewhere in the process. A major challenge is to develop the peak table as the process evolves, which is dynamically updated as new peaks are detected after the NOC region: this approach involving an "unlocked" peak table is contrasted to an approach using a "locked" peak table where only peaks detected during the NOC region are included in the model. In addition, results are compared to those obtained using baseline corrected and aligned chromatograms, using a NOC region of 5.85–8.33 h. D- and Q-charts were obtained. It is shown that the "unlocked" peak table detects out of control samples best and provides good diagnostic insight into problems with the process.



## 1. Introduction

Multivariate statistical process control (MSPC) [1–6] has been used for monitoring of reactions for more than a decade. Most commonly, within chemometrics, probes such as near infrared [7] and ultraviolet visible [8] are employed to monitor reactions over time. These approaches are particularly useful for steady state reactions such as occur during continuous processes, where the reaction mixture should be of constant composition throughout. Usually a normal operating condition (NOC) region of the process is defined whereby the reaction is considered to be under control. Deviations from these conditions are considered to be a consequence of a reaction getting out-of-control, and a warning to the process operator or plant manager that there could be problems.

Special difficulties often lie with small impurity peaks, as the quality of a product depends on there being impurities only below a given level. Hence it is important to be able to detect, in real-time, if the levels of impurities are changing significantly from those in the NOC region. Whereas spectroscopy has classically been employed for this purpose, chromatography adds extra insight, because it can be used to directly relate the origins of the faults to changes in chemical composition, especially in the context of impurity or contaminant peaks. With suitable chromatographic conditions, deviations from the steady state can be detected often by looking at the intensities of small peaks. Whereas manual inspection of chromatograms could be feasible, this often requires time and expertise, not always available on site during the running of a process. For a multistage process running over many days if several chromatograms are to be acquired per hour, sometimes at different sampling points, it is nearly impossible to arrange for manual inspection of such chromatograms, and so there is a need for automated on-line analysis of these data. In a previous paper, we reported the preliminary analysis of a process [6] using HPLC for monitoring but the MSPC was performed off-line in retrospect and not designed for dynamic monitoring of the reaction.

In this paper, we report an approach to solve this problem. Rapid HPLC can be acquired on-line during a process, and chemometric approaches employed to analyse these data. A major challenge

* Corresponding author. Tel.: +44 117 9287658; fax: +44 1179251295.
*E-mail address:* r.g.brereton@bris.ac.uk (R.G. Brereton).

0021-9673/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2008.09.082

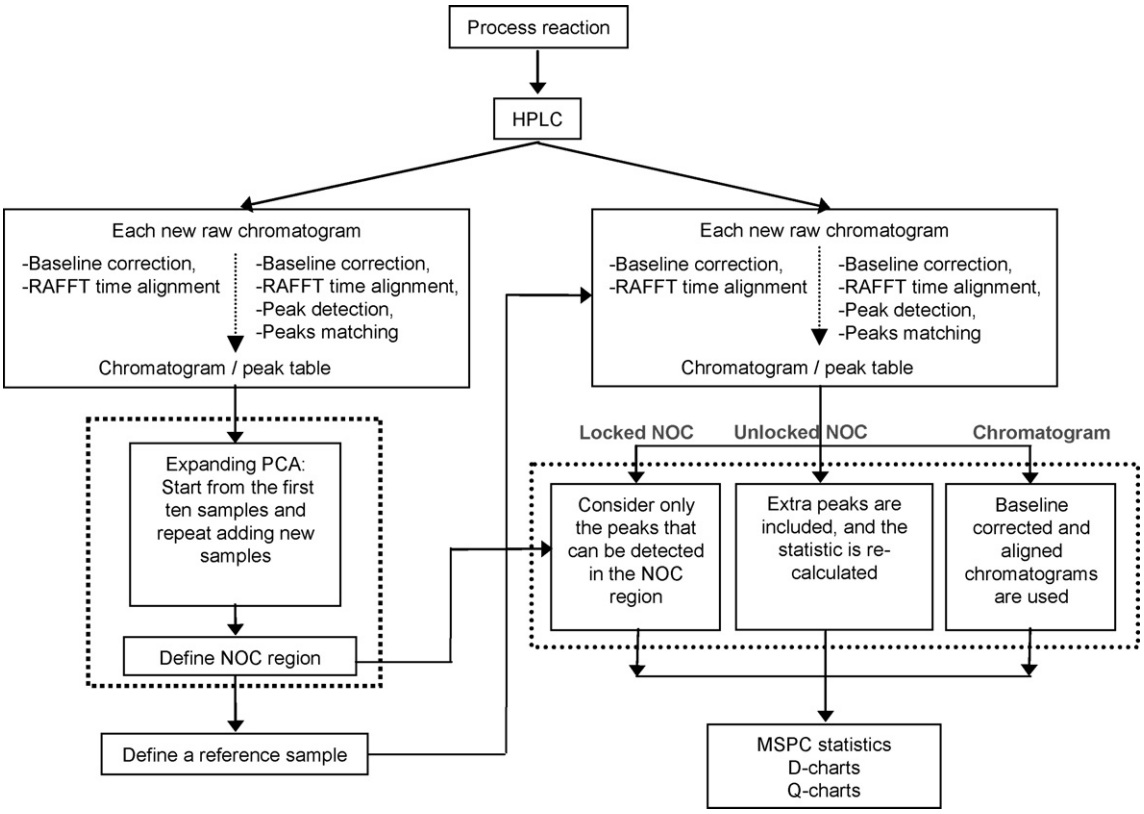


**Fig. 1.** Flowchart for on-line HPLC data analysis.

involves real-time data analysis: ability to form models and to obtain information as the process evolves. The most informative models take advantage of the information about the change in intensity of each peak in the chromatogram as these contain potential chemical origins of problems with the process. Sometimes process chemists are aware of specific impurities or reactants that provide information about why there may be difficulties, and so the use of HPLC has the potential to relate this back to the specific chemical problems. However, a specific difficulty is that peaks that are diagnostic of new faults may not have been detected in the NOC region. In a dynamic model formed during the evolution of the process, these peaks will be absent in the NOC model, and have to be added afterwards. This paper tackles this problem by introducing the idea of an "unlocked" peak table that is a table of detected compounds that changes as new peaks are detected after the NOC region: the intensity of these additional compounds within the NOC region is assumed to be 0. The models using an "unlocked" peak table are compared both to those using a "locked" peak table where only compounds detected within the NOC region are taken into account, and also using aligned baseline corrected chromatograms using raw data that has not been separated into individual peaks.

The approach described in this paper can be implemented dynamically in real-time as the process runs, with certain key features of the models evolving during the process.

This paper demonstrates that MSPC combined with on-line HPLC is a powerful approach for monitoring processes in real-time providing insight into the original of faults. It also shows that there is better insight if the intensities of individual peaks are used as input to the MSPC software, rather than using total chromatographic intensity, but in order to gain this additional information, attention has to be paid to the method for peak detection and alignment.

## 2. Experimental

### 2.1. Process

There are several stages in the process that has been studied. The results in this paper are from the last stage. The first sample was taken from the process at 3.49 h after the reaction started. Subsequent samples were taken from the process with 5–31 min intervals between each sampling time resulting in 252 samples overall; the sampling time was uneven and the reaction was monitored over 83.32 h. Due to a detector fault, there was a gap between 55.85 and 59.04 h when no samples were monitored. Each sample was analysed chromatographically using a single wavelength ultraviolet spectrometer detector at 245 nm, over an elution time of 2.5 min sampled at the rate of 13.72 scan $s^{-1}$ resulting in 2058 datapoints per chromatogram. Chromatographic data were exported to Matlab version 7 (Matworks, Natick, MA, USA) for further analysis. All software was written in Matlab. A data matrix of 252 rows corresponding to the samples and 2058 columns corresponding to the scans was obtained for further analysis.

### 2.2. Chromatography

An HPLC method was developed to facilitate fast analysis involving three steps: a sampling mechanism, a dilution device and a liquid chromatograph for on-line analysis. All systems were controlled by a personal computer. The sampling mechanism was connected to the chemical process. For each analysis, a sample volume of 365 μL was obtained and then pumped into the dilution device. In this study, 16.3 mL of ethyl acetate (Fisher Scientific, Loughborough, UK) was used for dilution and then pumped into an Agilent 1100 HPLC system controlled by Agilent Chemstation, v10.02 (Agilent Technologies, Stockport, UK). The HPLC