Journal of Chromatography A, 1361 (2014) 178–190

Contents lists available at ScienceDirect

Journal of Chromatography A

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

A universal comparison study of chromatographic response functions

Eva Tyteca*, Gert Desmet

Department of Chemical Engineering (CHIS-IR), Vrije Universiteit Brussel, B-1050 Elsene, Belgium

ARTICLE INFO

Article history:
Received 4 July 2014
Received in revised form 4 August 2014
Accepted 5 August 2014
Available online 13 August 2014

Keywords:
Chromatographic response function (CRF)
Automated method development
Chromatogram simulation

ABSTRACT

We report on a large scale *in silico* comparison study of so-called chromatographic response functions (CRFs). These are single number descriptors of the separation quality that can be used to guide search-based optimizations for chromatographic separations. A comprehensive set of literature and new CRFs were compared for their ability to guide a search based on first order chromatographic data (i.e., no spectral information available) and for cases where the number of sample compounds is not known beforehand. The results are discussed based on the available separation power. It was found that CRFs increasing monotonically with the number of observed peaks perform significantly better than those that do not possess this property. CRFs based on the discrimination factor or the peak-to-valley ratio can better cope with peak asymmetry than CRFs based on Snyder resolution R_s. Unfortunately, the former lose their advantage as soon as the noise level becomes significant. Most CRFs perform best when the search is conducted on a column offering just, or, even better, a bit less than the required efficiency to baseline separate the sample. The best results over the entire range of possible efficiencies are obtained with a CRF giving preference to the number of observed compounds before further ranking the conditions based on the achieved separation resolution or the required analysis time. When the search is conducted on columns with an insufficient efficiency, even the best possible CRFs suffer from the incomplete information about the sample, and deviating searches cannot be avoided without resorting to spectral information of the sample.

1. Introduction

Chromatographic method development (MD) usually involves a lot of trial and error because of the many adjustable parameters (type of organic modifier and stationary phase, temperature, gradient profile, pH, ionic strength, etc.) and the high probability for peak overlap [1]. To facilitate this search, a large variety of different automated MD strategies have been proposed in the past. Automated MD strategies can be either retention model-based (Drylab, Chromsword), search-based (Simplex, Genetic Algorithms (GA)) or a combination of both (Design of Experiments (DoE), Predictive Elution Window Shifting and Stretching (PEWS^2)) [2–9].

In complex design problems, where the number of variables is too large to just simply scan all possible design parameters, a numerical criterion describing the overall separation quality in a single number is needed to guide these automated MD strategies as efficient as possible to the optimal solution. In the literature, this quality criterion is generally referred to as chromatographic response function (CRF), a concept introduced by Morgan and Deming in 1975 [10]. Since then, a panoply of different CRFs has been proposed in literature and the debate on which CRF is the most suitable is still ongoing [11–26]. Designing a good CRF is particularly difficult when the number of components in the sample is unknown, or when a secondary goal needs to be satisfied (e.g., when also the analysis time needs to be minimized).

In the present study, we have made a global comparison of the most widely used CRFs proposed in literature, as well as some newly proposed ones. To base the conclusions of the study on a sufficiently large number of samples with widely differing composition, the comparison is made via a numerical study, using *in silico* samples (=computer simulated, but with realistic input parameters). This approach also has the advantage that the best possible solution is always known, such that the ability to find this solution can be unambiguously determined.

Emphasis is put on the case where the number of components in the sample is not known, as this is a more general and difficult problem than the one wherein all components of the sample are known. For the same reason, we also focus on the case of 1st order chromatographic data [27,28], where only a single detector signal is available (e.g., single-wavelength UV/vis-trace), hence assuming

* Corresponding author. Tel.: +32 629 36 17.
E-mail address: eva.tyteca@vub.ac.be (E. Tyteca).

http://dx.doi.org/10.1016/j.chroma.2014.08.014
0021-9673/

that the occurrence of fully overlapping peaks is not detectable and that the resolution of partially overlapping peaks is not quantifiable. If all peaks are always visible (via the use of spectral detection or using modeling or any other chemometric technique [29,30]), the CRF-selection problem becomes trivial (and hence not interesting) since in this case the resolution or the peak purity of the most critical pair is always the most direct and hence correct criterion [24,31].

The adopted search method was a simple brute force search, scanning all possible points on a search grid. This approach was preferred over any of the many existing "smart" search methods (Simplex, Genetic Algorithms, PEWS^2) [2,8,9], as we did not want the outcome of the study to depend on any possible peculiar properties of the employed search algorithm.

An overview of the different considered CRFs is given in Table 1. Distinction is made between CRFs without and with a time optimization component (category I and II respectively). The subcategories A and B have been introduced to distinguish between CRFs whose value increases monotonically with the number of observed compounds (category B) or not (category A). Since most of the CRFs proposed in literature belong to category A, whereas it can be anticipated that a dependency on the number of components is essential in searches where the number of components present in the sample is unknown, an effort has been made to transform most of the literature CRFs belonging to category A into a CRF belonging to category B by normalizing the original CRF-expression and adding the number of observed peaks (see Section 3, leading to a number of newly proposed CRFs). This is done in Section 3.1.

2. Theory and adopted numerical procedures

2.1. Chromatogram simulation

For reasons of simplicity the linear solvent strength (LSS) model [32,33] was adopted to implement the retention properties of the sample components.

$$\ln(k) = \ln(k_w) - S \cdot \phi \tag{1}$$

where the k_w-value is the extrapolated value of k for $\phi = 0$ (i.e., pure water) and S is the solvent strength parameter which is a constant for a given column, compound and organic solvent. Chromatograms were generated for different sample categories, respectively containing 5, 10, and 15 different compounds ($n_c = 5$, 10 and 15). Each sample was created by randomly picking n_c different combinations of the k_w- and S-values of the LSS-model from a prescribed interval using the *rand* function of Matlab®. To cover a sufficient variability in the considered samples, 10 different sample sub-categories with different k_w- and S-ranges, as well as with different concentration ratios were considered (see Table 2 for the different intervals). For each sample sub-category, 10 different samples were selected.

The randomly picked k_w- and S-values were subsequently used, together with a given value for the column volume V_0, the column efficiency N, and the gradient parameters ϕ_0 (initial mobile phase concentration, i.e. fraction of organic modifier at the start of the gradient) and βt_0 (gradient steepness $\beta \equiv (\phi_e - \phi_0) \cdot t_0/t_G$ with ϕ_e the fraction of organic modifier at the end of the gradient and t_0 the column dead time) to simulate the expected peak profiles using a computer program written in Matlab®. To calculate the peak profiles, both perfectly Gaussian peaks as well as skewed peaks were considered. The purely Gaussian peaks were calculated using: [34]

$$c\left(\frac{t}{t_0}\right) = \frac{c_0 \cdot \sqrt{N}}{(1+k_e)\cdot G} \exp\left(\frac{-N\left((t/t_0) - 1 - k\right)^2}{2(1+k_e)^2 G^2}\right)$$

$$\text{with} \quad c_0 = \frac{M}{\sqrt{2\pi}\cdot V_0} \tag{2}$$

where t is the time (min), c_0 the injected concentration (g/mL), k_e the retention factor experienced by a compound at the end of the column, i.e. at elution, k the effective retention factor of a compound (k is given by $k = t_R - t_0/t_0$ wherein t_R is the retention time of the compound), G the gradient compression factor [35], M the injected mass (g), V_0 the column dead volume (mL). The values of k_e and k have been calculated for each compound as a function of ϕ_0 and βt_0 and the known values of k_w and S using the linear gradient expressions given by Schoenmakers et al. [36].

To model the case of skewed chromatographic peaks, a large number of different expressions have already been proposed in literature, with varying complexity and accuracy [37]. In the present study, the well-established exponentially modified Gaussian (EMG) function has been used. Although polynomially modified Gaussian have been shown to be more accurate (at the expense of an additional complexity), the EMG is more widespread known and therefore provides a better reference to most readers. Using the same nomenclature as used in Eq. (2), the EMG shaped peaks were simulated using [38]:

$$c\left(\frac{t}{t_0}\right) = \frac{c_0 \cdot t_0(1+k)}{\tau \cdot \sqrt{N}} \sqrt{\frac{\pi}{2}} \exp\left(\frac{t_0^2(1+k)^2}{2\cdot\tau^2 N} - \frac{(t/t_0) - 1 - k}{\tau}\right)$$

$$\times \left(1 - \text{erf}\left(\frac{1}{\sqrt{2}}\left(\frac{1+k-(t/t_0)}{t_0(1+k)/\sqrt{N}}\right) + \frac{t_0(1+k)/\sqrt{N}}{\tau}\right)\right) \tag{3}$$

wherein τ is the relaxation time parameter of the exponential function used to modify the Gaussian and wherein erf is the error function. Two different types of EMG-peaks have been considered, differing in the width of the range from which the τ-values were randomly picked for each different compound: EMG$_1$-type peaks (with $0.5\sigma \le \tau \le 1.5\sigma$) and EMG$_2$-type peaks (with $0.5\sigma \le \tau \le 2.5\sigma$).

2.2. Chromatogram read-out

The chromatograms were read-out using a self-written Matlab®-routine determining the horizontal and the vertical position of the start, end and maximum of each peak or peak shoulder of the simulated chromatograms.

Using this information, the separation resolution of each peak pair i (comprising peaks i and $i+1$) was calculated, as well as the discrimination factor $d_{0,i}$ introduced by El Fallah and Martin [39,40] and Kaiser's peak-to-valley ratio's (f_i/g_i) [41–43] (see Fig. 1):

$$R_{s,i} = \frac{t_{R,i+1} - t_{R,i}}{1/2(w_{p,i+1} + w_{p,i})} \tag{4}$$

$$d_{0,i} = 1 - \frac{h_{\text{valley}}}{h_{\text{peak,min}}} \quad \text{and} \quad f_i/g_i = 1 - \frac{h_{\text{valley}}}{1/2\cdot(h_{\text{peak,min}} + h_{\text{peak,max}})} \tag{5}$$

The definition of h_{valley}, $h_{\text{peak,min}}$ and $h_{\text{peak,max}}$ is given in Fig. 1a. Fig. 1a also shows that the determination of the peak width (at either the height corresponding to the 4σ- or the 5σ-width) was carried out in a very crude way in the case of a non-fully resolved peak pair. This was done on purpose, to select CRFs that require only a minimal degree of sophistication for the read-out of their quality measures. The main difference between R_s on the one hand and d_0 and f/g on the other hand is that the former is read-out in the horizontal direction, whereas the latter are read-out in the vertical