



Assessment of the orthogonality in two-dimensional separation systems using criteria defined by the maximal information coefficient



Ahmad Mani-Varnosfaderani*, Mostafa Ghaemmaghami

Department of Chemistry, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Article history:

Received 12 May 2015

Received in revised form 23 August 2015

Accepted 24 August 2015

Available online 28 August 2015

Keywords:

Comprehensive two-dimensional gas chromatography

Maximal information coefficient

Maximal information-based nonparametric exploration

Abraham solvation parameter model

Orthogonality of 2-D separation system

ABSTRACT

A novel method based on the maximal information coefficient (MIC) is developed to assess the orthogonality of comprehensive two-dimensional separation systems. The proposed method is based on a modification of Marriott's method, which was previously reported in 2013. Marriott's method relies on the calculation of two separate parameters. The first term is C_{pert} which defines the peak coverage percent in separation space, and the second is C_{peaks} which corresponds to 2-D distribution correlation of the peaks. Marriott's method for estimating the values of C_{peaks} is based on the calculation of the coefficient of determination (R^2) between the retention indices of the peaks. Herein, we present some examples where R^2 is an inefficient way to estimate the values of C_{peaks} . The results in this work illustrate that when there are either functional or non-functional local dependencies between the distributions of the peaks, R^2 values fail to thoroughly estimate the values of C_{peaks} . We proposed using the MIC instead of R^2 to estimate the values of C_{peaks} for orthogonality calculations. Simulations of comprehensive two-dimensional gas chromatograms were performed using the Abraham solvation parameter model in order to generate examples for orthogonality assessment. The results indicate that the suggested modifications in this work correct the shortcomings of Marriott's model, and the proposed equation accurately measures the column dependencies in 2-D separation systems.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Due to the complicated nature of biological and environmental samples, there is a huge demand in analytical chemistry for performing separations in two or three dimensions [1–4]. In recent years, gas chromatography setups have been designed in multi-dimensions in order to create capabilities for resolving complicated mixtures, such as proteins and petroleum samples [5,6]. The basic logic for applying comprehensive two-dimensional separations is to obtain more information than from traditional one-dimensional techniques. In a simple view, if the efficiency in two separation spaces increases separately, the overall separation efficiency should also increase. Mathematically explained, the peak capacity in 2-D separation (P_{2-D}) is described by $P_{2-D} = P_1 P_2$ and can be calculated as the multiplication of peak capacities of both dimensions (P_1 , P_2). This definition is valid only when the selectivity of separations is completely independent and the entire 2-D separation space is randomly populated by the peaks. The orthogonality of the chromatographic columns is an important issue when trying to resolve

complicated samples in 2-D systems. Selection of proper columns as supporting media for separation spaces is critical to yield maximum capacities and should be investigated before analysis [7–9].

Through a comprehensive study on different LC systems, Gilar et al. [10] introduced a simple geometrical model to assess the orthogonality of different 2-D separation modes. They tried to bin the whole separation space and calculate the number of bins occupied by data points ($\Sigma bins$). If the total number of bins in a separation space is shown by P_{max} , then the orthogonality of the 2-D separation system is defined by the following equation:

$$O = \frac{\Sigma bins - \sqrt{P_{\text{max}}}}{0.63P_{\text{max}}} \quad (1)$$

Implementation of this equation on some simulated and real 2-D separation protocols revealed that the calculated values of orthogonality depend only on space coverage and not on the distribution patterns of the peaks [11]. In order to overcome this problem, Marriott et al. [11] introduced a method that uses information of peak distributions in a 2-D separation system. In their method, orthogonality was theoretically defined as the multiplication of two independent parameters:

$$O = C_{\text{pert}} \times C_{\text{peaks}} \quad (2)$$

* Corresponding author.

E-mail address: a.mani@modares.ac.ir (A. Mani-Varnosfaderani).

C_{pert} and C_{peaks} correspond to the bin coverage percent and the correlation distribution pattern of 2-D peaks, respectively. Orthogonal separation is achieved if C_{pert} and C_{peaks} attain one simultaneously, but complete non-orthogonality is achieved if either of them attains zero. In Marriott's formula, C_{pert} and C_{peaks} are defined by Eqs. (3) and (4), respectively:

$$C_{\text{pert}} = \frac{\Sigma \text{bins}}{0.63P_{\text{max}}} \quad (3)$$

$$C_{\text{peaks}} = 1 - R^2 \quad (4)$$

Following the above expressions, the orthogonality of a 2-D system is defined by Eq. (5):

$$O = \frac{\Sigma \text{bins}}{0.63P_{\text{max}}} \times (1 - R^2) \quad (5)$$

where R^2 is the regression coefficient for the peaks propagated in 2-D separation space. Marriott's work revealed that linear regression analysis is effective for discovering quantitative relationships of retention times in two separation dimensions. The evident advantage of Marriott's work is the reasonable employment of occupation of bins and further refinement of the distribution of the peaks with original retention data information. They compared their method with previously reported techniques and in most cases, their algorithm had superior performance for calculating orthogonality of 2-D separation systems [11].

In the present study, we detect some shortcomings in Eq. (5) for calculating the 2-D orthogonality. We found that the R^2 parameter itself is not enough to consider the shape of peak distributions in 2-D space. Instead, we propose a more powerful and general parameter, the maximal information coefficient (MIC), to assess the distribution patterns of peaks in separation space [12]. While R^2 is a useful parameter for measuring the linear relationships, the MIC captures all types of dependencies between peaks in 2-D space. In reality, the overall distribution patterns of the peaks are not linear in most cases and therefore, the R^2 parameter is not enough to explain the dependency in such distributions. In this paper, we present some examples which show in detail the inadequacy of R^2 for calculating the orthogonality. Then, we show how the MIC solves this problem. Briefly, we compare two methods for calculating the orthogonality in 2-D separation protocols: (1) Eq. (5), and (2) using MIC values instead of R^2 in Eq. (5). The results in this work reveal the superiority of the latter method.

Finally, we simulate some GC \times GC chromatograms using the Abraham solvation parameter model (ASPM) [13,14] and present the advantages of the newly suggested formula for measuring the degrees of orthogonality in 2-D systems.

2. Theoretical basis

Marriott's method captured the effects of both bin coverage and peak distributions for estimating the orthogonality. Inspection of Eq. (5) reveals that C_{pert} measures the effect of peak coverage and C_{peaks} corrects the formula for the linear dependencies between the retention times of the peaks in two separation spaces. For example, for a 2-D chromatogram when peaks are partially propagated on a linear line, R^2 value is high and C_{peaks} is on its minimum values and therefore, the orthogonality of the separation system is rather low. Generally, for cases where the peaks are linearly dependent or even independent, Eq. (5) works well. However, the problem arises when complicated dependencies are seen between the retention times of the peaks in two separation spaces. In these situations, we suggest the use of the MIC instead of R^2 for estimating the orthogonality. The MIC is able to capture all types of functional and even non-functional relationships between retention indices of peaks in 2-D space.

2.1. Maximal information coefficient

The MIC belongs to a larger class of maximal information-based nonparametric exploration (MINE) [12,15] techniques for discovering and classifying complicated associations. This parameter was introduced in 2011 by Reshef et al. [12] for identifying interesting relationships between pairs of variables in large data sets. MIC captures a wide range of associations and provides an index that thoroughly determines different types of dependencies between two variables. In order to calculate the MIC for two x and y variables, the algorithm grids the space of the two studied variables. The degree of gridding is explored up to the highest grid resolution based on the sample size in x and y . In the next step, the highest possible mutual information achieved by any x -by- y grid is computed. The calculated mutual information for different grid sizes is normalized to ensure a fair comparison between grids of different sizes and eventually, the MIC will be the maximum calculated normalized mutual information. More detailed descriptions about the theory of MIC can be found in literature [12]. For completely independent variables, the value of the MIC is near zero and for functional or non-functional related variables, it is equal to one. When there are some partial dependencies between two variables, the value of the MIC lies between zero and one. Previous investigations on the use of the MIC revealed that there is a good correlation between the MIC and R^2 values for linear distributions [12]. However, when analyzing real datasets, researchers found particular distributions with high MIC and low R^2 values [12]. We believe these types of distributions make mistakes when considering R^2 for orthogonality calculations in GC \times GC setups. Following, we present some examples which show the superiority of the MIC over R^2 . Some simulated distributions are shown in Fig. 1. For all of the shapes, the values of R^2 are small or even near zero. However, the MIC captures the sinusoidal, ellipsoidal, and circular dependencies. As can be seen in this figure, R^2 is not sensitive to these types of dependencies and therefore, the calculation of C_{peaks} by Eq. (5) has some detectable shortcomings. The differences between the calculated values of R^2 and the MIC in Fig. 1 clearly show that the R^2 values fail to capture complicated peak distributions. It is worth mentioning that some 2-D profiles, such as D, E, F, G and L in Fig. 1 can be seen in different real GC \times GC chromatograms in literature [16–19]. Therefore, implementation of the MIC instead of R^2 would be a good suggestion. We implemented this modification in Eq. (5) to address this problem:

$$O = \frac{\Sigma \text{bins}}{0.63P_{\text{max}}} \times (1 - \text{MIC}) \quad (6)$$

Briefly, we believe Eq. (6) is superior to Eq. (5), and this hypothesis is explored and tested in the next sections using simulated and reconstructed 2-D chromatograms.

3. Experimental

3.1. Abraham solvation parameter model

In order to simulate the GC \times GC chromatograms in this work, the Abraham solvation parameter model (ASPM) has been used. This method proposes a linear model to predict the retention indices of solutes in different chromatographic columns. According to the ASPM, the retention indices are described using the following equation:

$$I = l'L' + s'S' + a'A' + b'B' + e'E' + c' \quad (7)$$

where L' , S' , A' , B' and E' are the solute parameters and represent the size, dipolarity/polarizability, hydrogen bond acidity, hydrogen bond basicity, and excess polarizability of the solute, respectively. The parameters l' , s' , a' , b' , e' , and c' are system constants that

Download English Version:

<https://daneshyari.com/en/article/7611370>

Download Persian Version:

<https://daneshyari.com/article/7611370>

[Daneshyari.com](https://daneshyari.com)