

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,400

Open access books available

174,000

International authors and editors

190M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Introductory Chapter: Novel Aspects in Gas Chromatography and Chemometrics

Vu Dang Hoang, Victor David and Serban C. Moldoveanu

1. Introduction

Gas chromatography and chemometrics are important topics of analytical chemistry. They are both mature areas of research, and for this reason, the more recent progress made in these fields is not necessarily revolutionary. Nevertheless, the progress continues. A wide range of applications associated with continuous demands to improve analytical techniques is also reflected in the progress seen in gas chromatography and chemometrics.

2. An overview of progress in gas chromatography

The first gas chromatographic separations were performed more than 75 years ago [1, 2]. These separations used hydrogen as a carrier gas, a chromatographic column containing silica gel on activated carbon, and a thermal conductivity detector [3]. Important developments followed, such as the invention of flame ionization and electron capture detectors, the introduction of temperature gradient for the GC separation, the connection of a gas chromatograph with a mass spectrometer, the introduction of open-tubular (capillary) column, the introduction of capillary columns made from fused silica, etc. From the beginning of GC and up to this day, the progress in the nature of the stationary phase was also made. The first stationary phases in the packed column were made of solid porous support coated with a high-boiling fluid or porous plastic, and those were followed by the capillary columns with a bonded, cross-linked coating [4]. The minicomputer revolution allowed the introduction of computer control of the gas chromatographic instrumentation and the data processing in GC. Throughout its history, numerous other important improvements were made in gas chromatography. Among these can be mentioned the invention of other types of detectors, development of various injection procedures allowing large volume injections or cold on-column injection, development of solid-phase micro-extraction (SPME), introduction of autosamplers, development of comprehensive two-dimensional GC, introduction of fast gas-chromatography, etc.

Modern gas chromatography is strongly associated with the use of gas chromatography-mass spectrometry, which has a mass spectrometer as a detector for a GC system. As a result, the progress in mass spectrometry has been very important for the utilization of GC, and GC-MS became the most utilized and powerful technique for

compound identification in mixtures, and for the detection and quantitation of trace components when they are volatile. For this reason, the progress in GC is strongly associated with the developments in MS. Besides, the important improvements in mass spectrometric sensitivity, stability, and mass range, other examples of the progress include the development of large mass spectral libraries, introduction of GC-MS/MS systems, hyphenation of GC with high-resolution MS instruments, etc.

Over the years of utilization, significant effort has been made to adjust by chemical derivatization of the analytes in order to make them volatile and stable at a higher temperature such that to extend the range of compounds capable to be analyzed by GC and GC/MS [5]. Another extension of GC and in particular of GC-MS utilization is that on polymeric materials analysis. Polymeric compounds cannot be analyzed directly by GC due to their lack of volatility, but can be analyzed after thermal decomposition using a pyrolyzer.

The previously indicated developments are far from covering many developments in gas chromatography, mainly related to incremental developments. These developments are not seen as “milestones” although they can be very useful for practical purposes. For example, in the development of chromatographic columns for GC, among such incremental improvements is the decrease in column bleeding. As the temperature of the GC oven increases, some compounds from the stationary phase start to be generated and in the mass spectrometric detection produce an undesirable background. An important progress in making stationary phases aimed for the reduction of column bleeding allowing a better detection and also extending toward higher values the range of temperatures at which the columns can be utilized. Instrumentation in gas chromatography also has experienced numerous incremental improvements. Among these can be indicated the better control of oven temperature, better pneumatic control of pressures in the instrument, replacement of gas cylinders with gas generators, introduction of gas switches for helium conservation, development of portable GC (and GC-MS) systems, etc.

Another aspect related to the progress in gas chromatography refers to the extension of a range of applications for this technique. From the initial applications of separating gases and highly volatile compounds, gas chromatography continually expanded its range of applications. This included numerous applications for oil industry as well as for the analysis of fragrances and flavors, environmental pollutants (air, water, soil, etc.), pharmaceutical drugs, compounds in food, beverages and agricultural products, etc. A special role of GC (and GC-MS) is in the analysis of biological samples such as breath condensate, volatiles emitted from skin or bodily fluids, various xenobiotics, etc. A large part of biological samples is, however, non-volatile or highly polar molecules that cannot be directly analyzed by GC. Because of the importance of biological sample analysis, a significant effort has been made to process this type of sample by transforming them into volatile/semi-volatile compounds amenable for GC analysis.

Comprehensive two-dimensional gas chromatography coupled to mass spectrometry (GC × GC-MS) is now a common analytical technique used for the study of various complex samples, and the chemometrics based approaches are designated to decode the large amount of analytical information produced by this process (e.g. [6]).

A very large body of publications including papers in peer-reviewed journals, books, manufacturer catalogs, and information on the internet cover novel aspects in gas chromatography. Several novel items from a long list of improvements in the field are discussed in the present book.

3. Trends in chemometrics-based GC analysis

Historically, the application of chemometrics to GC analysis probably commenced in the 1990s with the works of Mayfield et al. employing a chemometrics software package to classify orange essence oil varieties analyzed by GLC (with FID and MS detectors) [7] and Jurášek et al. on the use of a chemometric detector (i.e., a computer method to selectively detect isotope cluster patterns in a time series of mass spectra of GC-MS analyses) [8].

Over the past decades, the proliferation of perceptibly sophisticated analytical systems coupled with more powerful detection techniques applied to the separation of complex samples has indeed requested highly efficient data analysis and optimization strategies. Chemometrics methods play a vital role in revealing the chemical information/knowledge hidden in high-dimensional datasets acquired from multidimensional separations. The idea of dimensionality of a mixture of compounds was defined by Giddings as “the number of independent variables that must be specified to identify the components of the sample” [9]. To improve the resolution and separation power of an analytical method, Giddings also suggested that a sample is subjected to a number of different separation mechanisms (i.e., the dimensionality of a separation method) [10]. With reference to gas chromatography, heart-cutting (i.e., conventional) techniques were initially proposed as multidimensional separations, in which the effluent from the first column was fractionally injected (i.e., a fraction or several consecutive fractions) onto a secondary column coated with a different type of stationary phase [11]. Although these conventional techniques could be useful in some cases when allowing additional separation of a modest number of the target analytes in specific regions of a single GC chromatogram, such analysis is not popular for complex matrices nowadays. The dimensionality of GC analysis has thus been mainly demonstrated by the application of GC-MS and comprehensive two-dimensional gas chromatography ($GC \times GC$). For the former, although GC-MS is commonly employed in the analytical sciences [12] many readers may not realize that it is, in effect, a multidimensional technique as a mass spectrometer as the detector adds a dimension to that of the chromatographic separation. For the latter, $GC \times GC$ is, in fact, a separation in which many sequential heart-cuts are further separated in a second column i.e., the effluent existed from the first column is periodically sampled in such a manner to preserve the separation in the first dimension and to subject all the compounds in a sample to both separation dimensions. To do so, the system must contain two orthogonal GC columns integrated by a special interface (modulator) [13].

It is worth mentioning that multidimensional GC data can have the second-order advantage for quantification using calibration standards containing only the analytes of interest without prior knowledge of possible matrix interference [14]. This can also considerably shorten analysis time by eliminating the need to resolve chromatographic overlapping signals. In practice, however, the full advantage of a combination of complementary techniques has not been efficiently exploited. For instance, GC-MS data are usually analyzed using the mass spectrometer as a filter to generate an entire chromatogram for single ions or using the GC chromatograph as a filter to identify particular peaks based on mass spectra matching. The fact that either filter is completely selective, necessitates the requirement for chemometric tools to extract all valuable information from huge amounts of multidimensional data sometimes referred to as “a tsunami of data” or more generally “Big data”.

Basically, the application of chemometrics to GC analysis can be categorized into two main groups: data preprocessing and data analysis. The aim of data processing

Chemometric tools	Overall features	Ref.
Data Preprocessing		
Baseline correction		
Savitzky-Golay	Using a polynomial fit of m^{th} order to $(2n + 1)$ neighboring points (inclusive of the point to be smoothed) with $n \geq m$	[15]
Asymmetric least squares	Using a smoother with deviations asymmetrically weighted to estimate a baseline	[16]
Polynomial fitting	Using a polynomial of n^{th} order that is a best fit (in a least-squares sense) for the data	[17]
Penalized least squares	Balancing the fit of a model to the data generated by the sum of squares against its roughness by altering a smoothing parameter, provided that the location of peaks in a chromatogram is established	[18]
Moving window minimum value	Sliding a window of length k across neighboring points (inclusive of the point to be smoothed) to give an array of local k -point centered minimum values.	[18]
Local minimum values coupled with robust statistical analysis	Using a linear interpolation to estimate the local minimum values in a chromatogram as a new baseline vector, with the help of a robust statistical strategy to detect outlier data points (corresponding to the unseparated peaks)	[18]
Retention-time-alignment		
Correlation-optimized warping	Dividing chromatograms into several local regions to be iteratively stretched and compressed until the Pearson correlation coefficient between the test chromatogram and the reference chromatogram is maximized	[19]
Local minimum value	Using a linear interpolation to predict the baseline after finding local minimum values in a chromatogram and eliminating outlier data points by an iterative optimization	[20]
Automatic peak detection and background drift correction	Accepting a signal as a true peak if (i) the absolute value of its first-order derivative is five times larger than a noise threshold and (ii) its second-order derivative crosses the zero-line fewer than eight times; correcting background drift by replacing regions containing peaks by linear baselines and using three-point moving-window averaging for denoising	[21]
Data Analysis		
Classification		
Unsupervised Pattern Recognition		
Cluster analysis (CA)	Grouping objects into clusters according to their similarity (proximity); monitoring the correctness of clusters and detecting deviation points	[22]
Hierarchical cluster analysis (HCA)	Creating a classification hierarchy that starts with each object in a single cluster and puts together clusters until only one is left	[23]
Principal component analysis (PCA)	Reducing the dimensionality of an original data set and creating new dimensions of data by conversion of strongly correlated input variables into uncorrelated values, called principal components; the first components can represent the maximum variance direction in the data and the omission of the remaining components does not result in a significant loss of information.	[23]
Supervised Pattern Recognition		
Linear discriminant analysis (LDA)	Data are projected from a D dimensional feature space down to a D' ($D > D'$) dimensional space to maximize inter-class variability and reduce intra-class variability	[24]

Chemometric tools	Overall features	Ref.
Quadratic discriminant analysis (QDA)	An LDA closely related algorithm that omits the assumption of equal covariance for all classes, but maintains the assumption of normality (unsuitable for very small sample sizes)	[25]
K-nearest neighbor (KNN)	A non-parametric algorithm that uses proximity to make classifications about grouping an individual data point. The classification of an object is based on a plurality vote of its neighbors when assigning it to the class most common among its k nearest neighbors (k is a positive integer, selected by cross-validation procedures to ensure the lowest classification error)	[26]
Random Forest (RF)	An ensemble learning algorithm that constructs multiple decision trees to find the best split to subset the data	[26]
Soft independent modeling of class analogy (SIMCA)	Samples are analyzed by PCA, with only the significant components retained. They can be identified as belonging to multiple classes, not necessarily classified into non-overlapping classes.	[27]
Support vector machine (SVM)	A linear classifier based on the kernel function can be used for non-linearly separable data by implicitly mapping them into higher dimensional feature spaces.	[28]
Partial least squares discriminant analysis (PLS-DA)	A discriminatory variant of Partial Least Squares regression	[29]
Quantification		
Partial least squares regression (PLSR)	A multivariate calibration method that reduces the predictors to a smaller set of uncorrelated components for least squares regression performance. It is very useful when the predictors are highly collinear and can be measured with error.	[30]
Artificial Neural Networks (ANN)	It is a subset of machine learning, constructed by using a set of algorithms that mimic the behavior of the human brain for pattern recognition. ANN are comprised of node layers; each node (aka. Artificial neuron) connects to another to form an extensive network for exchanging messages. It can be used as a prognostic model.	[31]
Multivariate curve resolution-alternating least squares (MCR-ALS)	Multivariate curve resolution, similar to PCA, seeks solutions accounting for the most variation possible with non-negativity constraints. Using a constrained Alternating Least Squares algorithm, MCR-ALS solves the MCR basic bilinear model.	[32]
Parallel factor analysis (PARAFAC)	It decomposes multidimensional arrays into component matrices and commonly uses ALS to calculate the decomposition.	[33]
Generalized rank annihilation (GRAM)	It is for solving an eigenvalue problem by using two data matrices simultaneously (unknown and calibration). The introduction of factor analysis is to project a target bilinear matrix onto another PC bilinear matrix space.	[34]

Table 1.
Chemometric tools for preprocessing and analyzing GC data.

strategies is to render GC data ready for accurate identification and quantification. They include (i) background correction (i.e., denoising and smoothing for the removal of low-amplitude signals irrespective of their frequency and high-frequency signals irrespective of their amplitude; drift correction for subtracting the baseline shape from a measurement) and (ii) retention time alignment (i.e., correcting inter-run variation in retention time for similar samples). After preprocessing data,

the translation of complex data for a sample into useful information covers a series of steps such as (i) peak detection (i.e., locating true signals in a chromatogram), (ii) information extraction (i.e., applying data dimension reduction), (iii) classification (i.e., discriminating between sample classes with different chemical characteristics). **Table 1** displays commonly used chemometric tools for and analyzing GC data with some typically applied studies for illustration purpose.

4. Conclusions and future outlook

In the literature, GC has been undoubtedly proven to be one of the most sensitive and popularly applied techniques for the separation and analysis of volatile and semi-volatile organic compounds. Its application can be adopted in a wide range of analytical studies in the biomedical, pharmaceutical, forensic, environmental and food sciences. To enhance the power for maximum sensitivity and selectivity, the use of multidimensional GC (especially hyphenated with mass spectrometry) is a helpful hint for sure. The analysis that relies on such modern GC techniques can generate a very powerful data platform, e.g., the use of profile data leads to more accurate information than centroid spectra. Thus, it necessitates robust data analysis strategies to extract relevant information from such GC data. Although there have been many interesting developments in the field of chemometrics-based GC analysis, it is still difficult to judge which algorithms can give the best results in general. This is because most chemometric methods were reported when addressing a specific challenge in a data set and comparisons with other approaches were infrequently sighted. It is suggested that more comprehensive studies should be done with different types of data and algorithms to shed light on the pros and cons of each chemometric tool.

In the 21st century, Artificial Intelligence (AI) is a fast-augmenting sector that has dramatically changed many aspects of daily life worldwide. It is unsurprising that AI, if exploited correctly as demonstrated e.g., [35–37], can help scientists achieve unimaginable breakthroughs and solutions regarding GC data analysis in the future.

IntechOpen

Author details

Vu Dang Hoang^{1*}, Victor David² and Serban C. Moldoveanu^{3*}


1 Faculty of Analytical Chemistry and Drug Testing, Hanoi University of Pharmacy, Hanoi, Vietnam

2 University of Bucharest, Romania

3 R.J. Reynolds Tobacco Co., United States of America

*Address all correspondence to: hoangvd@hup.edu.vn; moldovs@rjrt.com

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Prior F. Determination of Adsorption Heats of Gases and Vapors by Application of the Chromatographic Method in the Gas Phase, Doctoral thesis. Austria: University of Innsbruck; 1947 (in German)
- [2] Ettre L. The beginnings of gas adsorption chromatography 60 years ago. *LC-GC North America*. 2008;**26**(1):48-60
- [3] Bobleter O. Exhibition of the first gas chromatographic work of Erika Cremer and Fritz Prior. *Chromatographia*. 1996;**43**(7):444-446
- [4] Jennings WG, Poole CF. Milestones in the development of gas chromatography. In: *Gas Chromatography*. Second ed. Amsterdam: Elsevier; 2021. pp. 1-17
- [5] Moldoveanu SC, David V. Derivatization methods in GC and GC/MS. In: Kusch P, editor. *Gas Chromatography: Derivatization, Sample Preparation, Application*. London, UK, London: IntechOpen; 2018. DOI: 10.5772/intechopen.81954
- [6] Stefanuto P-H, Smolinska A, Focant J-F. Advanced chemometric and data handling tools for GC×GC-TOF-MS: Application of chemometrics and related advanced data handling in chemical separations. *TrAC Trends in Analytical Chemistry*. 2021;**139**:116251
- [7] Mayfield HT, Bertsch W, Mar T, Staroscik JA. Application of chemometrics to the classification of orange essence oil varieties by GLC. *Journal of High Resolution Chromatography*. 1986;**9**(2):78-83
- [8] Jurášek P, Slimák M, Košík M. Determination of isotope cluster patterns in mass spectra of GC-MS analyses by a chemometric detector. *Microchimica Acta*. 1993;**110**(4):133-142
- [9] Giddings JC. Sample dimensionality: A predictor of order-disorder in component peak distribution in multidimensional separation. *Journal of Chromatography A*. 1995;**703**(1):3-15
- [10] Giddings JC. Two-dimensional separations: Concept and promise. *Analytical Chemistry*. 1984;**56**(12):1258A-1270A
- [11] Marriott PJ, Chin S-T, Maikhunthod B, Schmarr H-G, Bieri S. Multidimensional gas chromatography. *TrAC Trends in Analytical Chemistry*. 2012;**34**:1-21
- [12] Matheson A, Botcherby L. Trends and developments in GC and GC-MS: A panel discussion on the latest advances and future developments in gas chromatography mass spectrometry (GC-MS). *The Column*. 2020;**16**(10):27-32
- [13] Liu Z, Phillips JB. Comprehensive two-dimensional gas chromatography using an on-column thermal modulator Interface. *Journal of Chromatographic Science*. 1991;**29**(6):227-231
- [14] Booksh KS, Kowalski BR. Theory of analytical chemistry. *Analytical Chemistry*. 1994;**66**(15):782A-791A
- [15] Mikaliunaite L, Sudol PE, Cain CN, Synovec RE. Baseline correction method for dynamic pressure gradient modulated comprehensive two-dimensional gas chromatography with flame ionization detection. *Journal of Chromatography A*. 2021;**1652**:462358
- [16] Samanipour S, Dimitriou-Christidis P, Gros J, Grange A,

- Samuel AJ. Analyte quantification with comprehensive two-dimensional gas chromatography: Assessment of methods for baseline correction, peak delineation, and matrix effect elimination for real samples. *Journal of Chromatography A*. 2015;**1375**:123-139
- [17] Mecozzi M. A polynomial curve fitting method for baseline drift correction in the chromatographic analysis of hydrocarbons in environmental samples. *APCBEE Procedia*. 2014;**10**:2-6
- [18] Fu H-Y, Li H-D, Yu Y-J, Wang B, Lu P, Cui H-P, et al. Simple automatic strategy for background drift correction in chromatographic data analysis. *Journal of Chromatography A*. 2016;**1449**:89-99
- [19] Zushi Y, Gros J, Tao Q, Reichenbach SE, Hashimoto S, Arey JS. Pixel-by-pixel correction of retention time shifts in chromatograms from comprehensive two-dimensional gas chromatography coupled to high resolution time-of-flight mass spectrometry. *Journal of Chromatography A*. 2017;**1508**:121-129
- [20] Fu H-Y, Hu O, Zhang Y-M, Zhang L, Song J-J, Lu P, et al. Mass-spectra-based peak alignment for automatic nontargeted metabolic profiling analysis for biomarker screening in plant samples. *Journal of Chromatography A*. 2017;**1513**:201-209
- [21] Yu Y-J, Fu H-Y, Zhang L, Wang X-Y, Sun P-J, Zhang X-B, et al. A chemometric-assisted method based on gas chromatography–mass spectrometry for metabolic profiling analysis. *Journal of Chromatography A*. 2015;**1399**:65-73
- [22] Passarella S, Guerriero E, Quici L, Ianiri G, Cerasa M, Notardonato I, et al. Dataset of PAHs determined in home-made honey samples collected in Central Italy by means of DLLME-GC-MS and cluster analysis for studying the source apportionment. *Data in Brief*. 2022;**42**:108136
- [23] Gilbert N, Mewis RE, Sutcliffe OB. Classification of fentanyl analogues through principal component analysis (PCA) and hierarchical clustering of GC–MS data. *Forensic Chemistry*. 2020;**21**:100287
- [24] Zhou X, Li X, Zhao B, Chen X, Zhang Q. Discriminant analysis of vegetable oils by thermogravimetric-gas chromatography/mass spectrometry combined with data fusion and chemometrics without sample pretreatment. *LWT*. 2022;**161**:113403
- [25] Aghili NS, Rasekh M, Karami H, Azizi V, Gancarz M. Detection of fraud in sesame oil with the help of artificial intelligence combined with chemometrics methods and chemical compounds characterization by gas chromatography–mass spectrometry. *LWT*. 2022;**167**:113863
- [26] Yun J, Cui C, Zhang S, Zhu J, Peng C, Cai H, et al. Use of headspace GC/MS combined with chemometric analysis to identify the geographic origins of black tea. *Food Chemistry*. 2021;**360**:130033
- [27] Becerra V, Odermatt J, Nopens M. Identification and classification of glucose-based polysaccharides by applying Py-GC/MS and SIMCA. *Journal of Analytical and Applied Pyrolysis*. 2013;**103**:42-51
- [28] Gerhardt N, Schwolow S, Rohn S, Pérez-Cacho PR, Galán-Soldevilla H, Arce L, et al. Quality assessment of olive oils based on temperature-ramped HS-GC-IMS and sensory evaluation: Comparison of different processing approaches by LDA, kNN, and SVM. *Food Chemistry*. 2019;**278**:720-728

- [29] Toraman HE, Abrahamsson V, Vanholme R, Van Acker R, Ronsse F, Pilate G, et al. Application of Py-GC/MS coupled with PARAFAC2 and PLS-DA to study fast pyrolysis of genetically engineered poplars. *Journal of Analytical and Applied Pyrolysis*. 2018;**129**:101-111
- [30] Aishima T. Comparing predictability of GC-MS and e-nose for aroma attributes in soy sauce using PLS regression analysis. In: WLP B, Petersen MA, editors. *Developments in Food Science*. Vol. 43. Amsterdam: Elsevier; 2006. pp. 525-528
- [31] Vyviurska O, Koljančić N, Gomes AA, Špánik I. Optimization of enantiomer separation in flow-modulated comprehensive two-dimensional gas chromatography by response surface methodology coupled to artificial neural networks: Wine analysis case study. *Journal of Chromatography A*. 2022;**1675**:463189
- [32] Izadmanesh Y, Garreta-Lara E, Ghasemi JB, Lacorte S, Matamoros V, Tauler R. Chemometric analysis of comprehensive two dimensional gas chromatography–mass spectrometry metabolomics data. *Journal of Chromatography A*. 2017;**1488**:113-125
- [33] Valverde-Som L, Reguera C, Herrero A, Sarabia LA, Ortiz MC. Determination of polymer additive residues that migrate from coffee capsules by means of stir bar sorptive extraction-gas chromatography-mass spectrometry and PARAFAC decomposition. *Food Packaging and Shelf Life*. 2021;**28**:100664
- [34] Prazen BJ, Bruckner CA, Synovec RE, Kowalski BR. Second-order chemometric standardization for high-speed hyphenated gas chromatography: Analysis of GC/MS and comprehensive GC×GC data. *Journal of Microcolumn Separations*. 1999;**11**(2):97-107
- [35] Baccolo G, Quintanilla-Casas B, Vichi S, Augustijn D, Bro R. From untargeted chemical profiling to peak tables – A fully automated AI driven approach to untargeted GC-MS. *TrAC Trends in Analytical Chemistry*. 2021;**145**:116451
- [36] Bi K, Zhang D, Qiu T, Huang Y. GC-MS fingerprints profiling using machine learning models for food flavor prediction. *Processes*. 2020;**8**:23. DOI: 10.3390/pr8010023
- [37] Matyushin DD, Sholokhova AY, Buryak AK. Deep learning driven GC-MS library search and its application for metabolomics. *Analytical Chemistry*. 2020;**92**(17):11818-11825