

Review Article

Chemometrics in Fingerprinting by Means of Thin Layer Chromatography

Łukasz Komsta

Department of Medicinal Chemistry, Medical University of Lublin, Jaczewskiego 4, 20-090 Lublin, Poland

Correspondence should be addressed to Łukasz Komsta, lukasz.komsta@umlub.pl

Received 29 July 2011; Revised 29 September 2011; Accepted 30 September 2011

Academic Editor: Monika Waksmundzka-Hajnos

Copyright © 2012 Łukasz Komsta. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The paper is written as an introductory review, presenting summary of current knowledge about chemometric fingerprinting in the context of TLC, due to a rather small interest in the literature about joining TLC and chemometrics. The paper shortly covers the most important aspects of the chemometric fingerprinting in general, creating the TLC fingerprints, denoising, baseline removal, warping/registering, and chemometric processing itself. References being good candidates as a starting point are given for each topic and processing step.

1. Introduction

Exactly 40 years ago, in 1971, Swante Wold used the term “chemometrics” for the first time in his grant application. These 40 years of continuous development of computer abilities and analytical chemistry methods made chemometrics accessible to any interested scientist. The chemometric techniques allow the researcher to compare and explore complex and multivariate data by a proper projection methods [1].

The chemometric exploration of chromatographic data is a well-established topic in the case of HPLC equipped with different detection techniques [2–4]. The application of this approach is very wide, from ion chromatography of mineral waters [5], through oil spills [6], metabolic profiling [7], different kinds of food [8], petroleum oils [9], and ending in pharmaceutical and herbal investigations [10–16]. Although such approach needs some chemometric (mathematical) knowledge, it is more and more often preferred because of the lack of the need of identifying every peak or, at least significant peaks. The chromatogram is treated not as the set of any peaks, but as an unique multivariate signal. Therefore, there is an increasing interest in chemometric chromatographical fingerprinting.

Although thin-layer chromatography is a well-established technique and its pharmaceutical and herbal applications are very wide (see e.g., books of Macek [17] and

Waksmundzka-Hajnos et al. [18]), the fingerprinting is done here mainly in a nonchemometric way. The most often used approach relies on identifying peaks and comparing their height, area, or intensity. There is still lack of comprehensive fingerprinting TLC approaches based on the full chemometric processing of collected data. The reason for this cannot be definitely justified; it is most probably caused by narrow specializing of thin-layer chromatographers, some fear of chemometrics, and lack of cooperation between TLC chromatographers and chemometricians.

Therefore, the purpose of this paper is to review and summarize all possibilities in TLC fingerprinting and settle some starting point for any TLC chromatographer, which would be interesting in introducing chemometrics to his/her research.

2. Fingerprinting: Classical and by Means of Chemometrics

The classical fingerprinting by TLC is done visually. The chromatographic system is optimized to separate and visualize (e.g., by different colors) all interesting components from particular herbal sample. Next, the samples belonging to different herbal species are compared and the differences in particular peaks are manually notified. Several peaks can be then proposed as markers, that is, substances responsible for differences between discriminated classes.

This allows the chromatographer to analyze an unknown sample and—based on absolute or relative intensity of marker peaks—to distinguish between particular species or chemotaxonomic groups. The fingerprints can be done with several TLC systems simultaneously, each system allowing to see and compare the markers found in another group of phytochemicals found in the sample.

The main disadvantage of classical fingerprinting is its subjectivity. The comparison of chromatograms is mainly based on a subjective (not objective) method, mainly chromatographer perception. The chemometric approach, although more difficult, is based only on objective mathematical methods.

The chemometric fingerprinting treats the chromatogram as a unique signal, without a need to identify and interpret the peaks. The chromatogram is represented as a vector—a series of numbers with a particular sampling rate. Therefore, each chromatogram can be treated as one point in a multidimensional space, where the number of dimensions is defined by the signal length. As the chromatograms consist of tens, hundreds, or thousands of points, the high multidimensionality does not allow the chemometrician to imagine such point; however, the spaces of higher dimensions have analogous to 3D space metric properties (angles, distances, hyperplanes, etc.).

Considering the chromatograms as multivariate points, one could measure a similarity/dissimilarity between any two of them in different manner. The most often used measure is Euclidean distance (which is analogous to distance between points in 3D space), and angular distance (which is the angle between vectors pointed from origin to the two points). The first one is sensitive to absolute values of the peaks, the second one works in a correlation-like manner and compares only shape of the chromatograms. Based on the similarity/dissimilarity or correlation matrix one can perform a number of unsupervised chemometric techniques, which will be explained below.

3. Creating Fingerprints

In the case of HPLC, the chromatography equipment allows a researcher to export a chromatogram as a serie of numbers, because the chromatogram is already digitized during its development. In the case of TLC, the plate must be scanned by densitometer or photographed after development. In a general way, three approaches can be considered.

- (1) Scanning a plate with a densitometer.
- (2) Taking a photo of the plate and extracting univariate, densitogram-like signal along a particular track.
- (3) Taking a photo of whole plate, when 2D TLC is performed.

The densitometric scanning can be performed in ultraviolet wavelength or visual wavelength, before or after applying a developing agent. The scanning can be performed with one wavelength, several wavelengths, or in multiwavelength mode. The last one results in the data analogous to DAD chromatogram, but the main disadvantage is a time needed

to scan the track. Only the newest densitometers have DAD built in, and can collect many wavelengths during one pass of the scanner. The plate can be scanned two times, before and then after development by visual agent, densitograms can be then combined together with one matrix with more channels.

In the case of videoscanning, one can take a color photo at 254 nm light, 366 nm light, or visible light. Each photo has three channels (RGB), so there is a theoretical possibility to collect 9 channels of information for the same plate. However, one must not move the plate under the camera during taking subsequent pictures to preserve compatibility of spatial information. For this reason, it is almost impossible to move the plate out of videoscanner, develop it by some agent, and then scan it again: obtained photo will not exactly match the previous one and the spatial difference can make the chemometric processing almost impossible (or at least the results will be incorrect). It could be possible with a special holder, always holding the plate at an exact position, but to the best knowledge of the author, no commercially available videoscanning camera is equipped with it.

The images of the plate can be collected not only by special videoscanning equipment—in many cases a popular digital camera is also an appropriate hardware. However, it must be done in repetitive conditions, in the same place, with the same ambient light, and with the camera fixed in one position with tripod. The univariate tracks can be then obtained and exported from the photos by an open source software, such as ImageJ [19].

4. Chemometric Pretreatment

Before applying the chemometric techniques, proper pretreatment of the signals is a necessity. The typical order of preprocessing is as follows: normalization, denoising, and baseline removal, followed by warping/registering.

The normalization step is not mandatory and there is no consensus when it is obligatory—sometimes it makes the results better, sometimes even worse. Therefore it is recommended to compare the results with and without normalization. The most often method of signal normalization is the standard normal variate (SNV) method [20]. It is a centering to zero mean and scaling to unit variance of each signal itself (independent of scaling or centering used in chemometric methods themselves).

4.1. Denoising. The denoising step is also not mandatory and it should be performed only in the case of visible noise. However, the densitograms and videoscans have quite a high noise level in general. The signal processing science knows many methods of denoising and smoothing of the signals. The simplest one is Savitzky-Golay smoothing filter (which can be used also with simultaneous differentiation) [21]. The filter parameters should be optimized against the correlation of filtered noise [22]. The other most important method is the wavelet shrinkage [23, 24]. The other ones, such as digital filters, Fourier denoising, or Whittaker smoother are not so often used.

The optimum denoising method is different in the case of different noise properties—noise spectrum (and autocorrelation) and homo/heteroscedasticity. The densitometer noise is in most cases pink and heteroscedastic whereas the CCD camera noise is white and homoscedastic. Therefore the recommended methods are different. The reader can be referred to two papers with comprehensive comparison of denoising techniques in the case of densitograms [25] and univariate videoscans [26]. In the first case, denoising is more difficult and the Savitzky-Golay filter or Haar wavelet shrinkage with soft thresholding is a good choice. In the second case, the Haar mother wavelet with soft denoising and any decomposition level larger than 1 was the best choice.

There is also another report [27] on image processing techniques on whole images (when the 2D chromatogram is processed and one cannot extract a univariate path). Several other kind of filters (averaging, circular, Gaussian, median, and Wiener) can be used here regardless of 2D version of previous techniques. This study recommends use of median filter with the window width optimized against filtered noise autocorrelation. The wavelet shrinkage at high decomposition level performs also very well.

4.2. Baseline Removal. The baseline removal process is almost always mandatory. If this step is not applied, a baseline variation causes a random variance greater than the one caused by chromatogram variability. The real trends in data are then completely obscured by baseline trends; therefore one cannot make any real conclusion without baseline removal.

If the univariate signals are considered (densitograms, univariate videoscans), the reader is referred to a comprehensive study and review of current baseline removal methods [28]. Many approaches exist in the literature and the methods of choice could be asymmetric penalized least squares (e.g., in airPLS implementation) or quantile regression.

In the case of 2D videoscans (images), the simple algorithms, such as rollerball method [29] are sufficient to filter the smooth baseline trends caused by inhomogeneous illumination or spraying. The more advanced methods, used mainly in the case of 2D gel electropherograms [30] are not necessary in the case of TLC [29].

4.3. Warping. Another significant problem is the random shift of the chromatographic peaks. This phenomenon cannot be fully avoided and the variability caused by these shifts can also obscure the variability between the peaks themselves. This problem is extensively discussed on univariate chromatograms in [31]. The most often used approach is called correlation optimized warping (COW) whereas dynamic time warping (DTW), Fuzzy Warping or several other ideas are also present in the literature [32–34].

Regardless of the method used, there is a need that all aligned fingerprints contain a set of common peaks to be warped correctly and the target chromatogram (the reference one to be aligned) is also carefully selected [35].

The only one way to bypass the alignment/warping procedure is to use so-called Gram-matrix in the case of multichannel chromatograms [36]. In TLC this would mean multiwavelength DAD-like signals. However, the time is the main drawback here in the case of classical densitometer.

In the case of image processing, a warping step (called also in the image context “registering”) is a little bit more difficult. The example approach [29] requires a registering done by special curves called splines, against the correlation. The random variation of the spot location in 2D TLC is very significant and correct processing of such fingerprints are almost impossible without this step.

5. Chemometrics Itself: Unsupervised and Supervised Techniques

The chemometric techniques applied to TLC fingerprints can be unsupervised or supervised. The mainly used unsupervised techniques [1, 37] are principal component analysis [38] with its robust variant [39] independent component analysis [40], and cluster analysis, including dendrograms [41]. In the case of multimode data, such as multichannel fingerprints, the PARAFAC method can be also considered [42].

The unsupervised approaches perform analysis taking into the account some objective criteria, such as dissimilarity or variance in the data matrix. PCA and ICA find directions in the datasets, therefore after applying this technique on the “calibration set” one can store the directions (expressed as loading vectors) and make a further projections of any unknown future sample. In the case of dendrograms, one can add the further samples to dendrogram, because adding another object does not change shape of the dendrogram. In all the cases, one should point a significant attention for outlying objects, as they can disturb data analysis completely [43].

The supervised techniques are used as multivariate calibration methods [44], or as discriminative methods [45], when data do not cluster against any objective criterion. Although there is a possibility to construct a regression model with a number of first latent variables of PCA and such approach is called principal component regression (PCR), the analysis of data matrix is then based only on variance and, therefore, is unsupervised. Due to this limitation, the most often used linear supervised method is the partial least squares (PLS), known also as projection to latent structures [46, 47]. It finds the directions in the data matrix, which maximize the covariance between matrix itself and predicted variable, and then constructs a regression model with several first such directions. There are also some extensions-generalizations of the method, such as cyclic subspace regression [48] or continuum regression [49]. In the case of outlying object, there may be a need to introduce their robust variants [50]. The linear methods are sufficient in many cases; however, sometimes there is a need to use nonlinear ones. The PLS method can be nonlinearized by applying the radial basis function concept [51]. The neural networks are also often used [52].

6. Conclusion

Although the chemometric approach in TLC fingerprinting requires some prior mathematical and informatical knowledge, there are many publications which can be treated as starting points for interested TLC chromatographers. Tracing the citations inside these papers, one can gain all the knowledge for chemometric treatment for TLC data. However, it is always a nice idea to consult an idea with some experienced chemometrician. Due to increasing interest in chemometrics, we can expect an increasing interest in chemometric evaluation of all TLC data.

References

- [1] M. Daszykowski, B. Walczak, and D. L. Massart, "Projection methods in chemistry," *Chemometrics and Intelligent Laboratory Systems*, vol. 65, no. 1, pp. 97–112, 2003.
- [2] J. H. Christensen, J. Mortensen, A. B. Hansen, and O. Andersen, "Chromatographic preprocessing of GC-MS data for analysis of complex chemical mixtures," *Journal of Chromatography A*, vol. 1062, no. 1, pp. 113–123, 2005.
- [3] Y. Ni, L. Zhang, J. Churchill, and S. Kokot, "Application of high performance liquid chromatography for the profiling of complex chemical mixtures with the aid of chemometrics," *Talanta*, vol. 72, no. 4, pp. 1533–1539, 2007.
- [4] Y. Ni, Y. Peng, and S. Kokot, "Fingerprinting of complex mixtures with the use of high performance liquid chromatography, inductively coupled plasma atomic emission spectroscopy and chemometrics," *Analytica Chimica Acta*, vol. 616, no. 1, pp. 19–27, 2008.
- [5] E. Blicharska, Ł. Komsta, R. Kocjan, A. Gumieniczek, and A. Wiśniewska, "Chemometric processing of ion chromatograms application to comparative analysis of polish bottled mineral and spring waters," *Polish Journal of Environmental Studies*, vol. 19, no. 5, pp. 1071–1075, 2010.
- [6] J. H. Christensen and G. Tomasi, "Practical aspects of chemometrics for oil spill fingerprinting," *Journal of Chromatography A*, vol. 1169, no. 1-2, pp. 1–22, 2007.
- [7] T. Gröger and R. Zimmermann, "Application of parallel computing to speed up chemometrics for GC × GC-TOFMS based metabolic fingerprinting," *Talanta*, vol. 83, no. 4, pp. 1289–1294, 2011.
- [8] K. A. Aliferis, P. A. Tarantilis, P. C. Harizanis, and E. Alisandrakis, "Botanical discrimination and classification of honey samples applying gas chromatography/mass spectrometry fingerprinting of headspace volatile compounds," *Food Chemistry*, vol. 121, no. 3, pp. 856–862, 2010.
- [9] I. Eide and K. Zahlsen, "A novel method for chemical fingerprinting of oil and petroleum products based on electrospray mass spectrometry and chemometrics," *Energy and Fuels*, vol. 19, no. 3, pp. 964–967, 2005.
- [10] X. H. Fan, Y. Y. Cheng, Z. L. Ye, R. C. Lin, and Z. Z. Qian, "Multiple chromatographic fingerprinting and its application to the quality control of herbal medicines," *Analytica Chimica Acta*, vol. 555, no. 2, pp. 217–224, 2006.
- [11] Y. Hu, Y. Z. Liang, B. Y. Li, X. N. Li, and Y. P. Du, "Multicomponent spectral correlative chromatography applied to complex herbal medicines," *Journal of Agricultural and Food Chemistry*, vol. 52, no. 26, pp. 7771–7776, 2004.
- [12] F. Gong, Y. Z. Liang, Y. S. Fung, and F. T. Chau, "Correction of retention time shifts for chromatographic fingerprints of herbal medicines," *Journal of Chromatography A*, vol. 1029, no. 1-2, pp. 173–183, 2004.
- [13] F. Gong, B. T. Wang, F. T. Chau, and Y. Z. Liang, "Data preprocessing for chromatographic fingerprint of herbal medicine with chemometric approaches," *Analytical Letters*, vol. 38, no. 14, pp. 2475–2492, 2005.
- [14] Y. Liang, P. Xie, and F. Chau, "Chromatographic fingerprinting and related chemometric techniques for quality control of traditional Chinese medicines," *Journal of Separation Science*, vol. 33, no. 3, pp. 410–421, 2010.
- [15] Y.-Z. Liang, P.-S. Xie, and K. Chan, "Chromatographic fingerprinting and metabolomics for quality control of TCM," *Combinatorial Chemistry and High Throughput Screening*, vol. 13, no. 10, pp. 943–953, 2010.
- [16] Y. Z. Liang, P. S. Xie, and K. Chan, "Perspective of chemical fingerprinting of Chinese herbs," *Planta Medica*, vol. 76, no. 17, pp. 1997–2003, 2010.
- [17] K. Macek, *Pharmaceutical Applications of Thin-Layer and Paper Chromatography*, Elsevier, 1972.
- [18] M. Waksmundzka-Hajnos, J. Sherma, and T. Kowalska, *Thin Layer Chromatography in Phytochemistry*, Taylor-Francis, 2010.
- [19] <http://rsbweb.nih.gov/ij/>.
- [20] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," *Applied Spectroscopy*, vol. 43, no. 5, pp. 772–777, 1989.
- [21] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
- [22] G. Vivó-Truyols and P. J. Schoenmakers, "Automatic selection of optimal Savitzky-Golay smoothing," *Analytical Chemistry*, vol. 78, no. 13, pp. 4598–4608, 2006.
- [23] B. K. Alsberg, A. M. Woodward, and D. B. Kell, "An introduction to wavelet transforms for chemometricians: a time-frequency approach," *Chemometrics and Intelligent Laboratory Systems*, vol. 37, no. 2, pp. 215–239, 1997.
- [24] B. Walczak and D. L. Massart, "Wavelets—something for analytical chemistry?" *Trends in Analytical Chemistry*, vol. 16, no. 8, pp. 451–462, 1997.
- [25] Ł. Komsta, "A comparative study on several algorithms for denoising of thin layer densitograms," *Analytica Chimica Acta*, vol. 641, no. 1-2, pp. 52–58, 2009.
- [26] Ł. Komsta, "Suppressing the charged coupled device noise in univariate thin-layer videoscans: a comparison of several algorithms," *Journal of Chromatography A*, vol. 1216, no. 12, pp. 2548–2553, 2009.
- [27] Ł. Komsta, "Dealing with charged-coupled device noise in thin-layer videodensitometry. Optimization of several image-denoising techniques," *Acta Chromatographica*, vol. 21, no. 3, pp. 355–367, 2009.
- [28] Ł. Komsta, "Comparison of several methods of chromatographic baseline removal with a new approach based on quantile regression," *Chromatographia*, vol. 73, no. 7-8, pp. 721–731, 2011.
- [29] Ł. Komsta, T. Cieśla, A. Bogucka-Kocka, A. Józefczyk, J. Kryszewski, and M. Waksmundzka-Hajnos, "The start-to-end chemometric image processing of 2D thin-layer videoscans," *Journal of Chromatography A*, vol. 1218, no. 19, pp. 2820–2825, 2011.
- [30] M. Daszykowski, I. Stanimirova, A. Bodzon-Kulakowska, J. Silberring, G. Lubec, and B. Walczak, "Start-to-end processing

- of two-dimensional gel electrophoretic images,” *Journal of Chromatography A*, vol. 1158, no. 1-2, pp. 306–317, 2007.
- [31] M. Daszykowski, Y. Vander Heyden, C. Boucon, and B. Walczak, “Automated alignment of one-dimensional chromatographic fingerprints,” *Journal of Chromatography A*, vol. 1217, no. 40, pp. 6127–6133, 2010.
- [32] B. Walczak and W. Wu, “Fuzzy warping of chromatograms,” *Chemometrics and Intelligent Laboratory Systems*, vol. 77, no. 1-2, pp. 173–180, 2005.
- [33] V. Pravdova, B. Walczak, and D. L. Massart, “A comparison of two algorithms for warping of analytical signals,” *Analytica Chimica Acta*, vol. 456, no. 1, pp. 77–92, 2002.
- [34] A. M. van Nederkassel, M. Daszykowski, P. H. C. Eilers, and Y. V. Heyden, “A comparison of three algorithms for chromatograms alignment,” *Journal of Chromatography A*, vol. 1118, no. 2, pp. 199–210, 2006.
- [35] M. Daszykowski and B. Walczak, “Target selection for alignment of chromatographic signals obtained using monochannel detectors,” *Journal of Chromatography A*, vol. 1176, no. 1-2, pp. 1–11, 2007.
- [36] M. Daszykowski, R. Danielsson, and B. Walczak, “No-alignment-strategies for exploring a set of two-way data tables obtained from capillary electrophoresis-mass spectrometry,” *Journal of Chromatography A*, vol. 1192, no. 1, pp. 157–165, 2008.
- [37] M. Daszykowski, “From projection pursuit to other unsupervised chemometric techniques,” *Journal of Chemometrics*, vol. 21, no. 7–9, pp. 270–279, 2007.
- [38] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1–3, pp. 37–52, 1987.
- [39] C. Croux and A. Ruiz-Gazen, “A fast algorithm for robust principal components based on projection pursuit,” in *Proceedings of the Computational Statistics (COMPSTAT '96)*, Physica, Heidelberg, Germany, 1996.
- [40] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley and Sons, New York, NY, USA, 2001.
- [41] N. Bratchell, “Cluster analysis,” *Chemometrics and Intelligent Laboratory Systems*, vol. 6, pp. 105–125, 1987.
- [42] R. Bro, “PARAFAC. Tutorial and applications,” *Chemometrics and Intelligent Laboratory Systems*, vol. 38, no. 2, pp. 149–171, 1997.
- [43] M. Daszykowski, K. Kaczmarek, Y. Vander Heyden, and B. Walczak, “Robust statistics in data analysis—a review. Basic concepts,” *Chemometrics and Intelligent Laboratory Systems*, vol. 85, no. 2, pp. 203–219, 2007.
- [44] H. Martens and T. Naes, *Multivariate Calibration*, John Wiley and Sons, 1989.
- [45] M. Barker and W. Rayens, “Partial least squares for discrimination,” *Journal of Chemometrics*, vol. 17, no. 3, pp. 166–173, 2003.
- [46] S. Wold, M. Sjostrom, and L. Eriksson, “PLS—a basic tool of chemometrics,” *Chemometrics and Intelligent Laboratory Systems*, vol. 58, pp. 109–130, 2001.
- [47] J. Hinkle and W. Rayens, “Partial least squares and compositional data: problems and alternatives,” *Chemometrics and Intelligent Laboratory Systems*, vol. 30, no. 1, pp. 159–172, 1995.
- [48] J. H. Kalivas, “Cyclic subspace regression with analysis of the hat matrix,” *Chemometrics and Intelligent Laboratory Systems*, vol. 45, no. 1-2, pp. 215–224, 1999.
- [49] S. De Jong and R. W. Farebrother, “Extending the relationship between ridge regression and continuum regression,” *Chemometrics and Intelligent Laboratory Systems*, vol. 25, no. 2, pp. 179–181, 1994.
- [50] M. Daszykowski, Y. Vander Heyden, and B. Walczak, “Robust partial least squares model for prediction of green tea antioxidant capacity from chromatograms,” *Journal of Chromatography A*, vol. 1176, no. 1-2, pp. 12–18, 2007.
- [51] B. Walczak and D. L. Massart, “The radial basis functions—partial least squares approach as a flexible non-linear regression technique,” *Analytica Chimica Acta*, vol. 331, no. 3, pp. 177–185, 1996.
- [52] B. Walczak, “Neural networks with robust backpropagation learning algorithm,” *Analytica Chimica Acta*, vol. 322, no. 2, pp. 21–29, 1996.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

