Postprint of: Kalinowska K., Bystrzanowska M., Tobiszewski M., Chemometrics approaches to green analytical chemistry procedure development, Current Opinion in Green and Sustainable Chemistry (2021), 100498, DOI: 10.1016/j.cogsc.2021.100498

Chemometrics approaches to green analytical chemistry procedure development

Kaja Kalinowska, Marta Bystrzanowska, Marek Tobiszewski*

Department of Analytical Chemistry, Faculty of Chemistry, Gdańsk University of Technology (GUT), 11/12 G. Narutowicza St., 80-233 Gdańsk, Poland.

* author for correspondence: marek.tobiszewski@pg.edu.pl

Abstract

Chemometric tools are widely used in analytical chemistry for the reduction of data dimensionality, grouping of variables and processing of analytical signals. They have also the potential to be applied in analytical procedure development with the aim of minimizing the procedure's environmental impact. The design of experiment gives the possibility to obtain much better information on the system response than in case of "changing one variable at a time" approach. This results in savings of materials and energy. Desirability functions applied together with the design of experiment, create a possibility to include in procedure development the variables that directly refer to the procedure's greenness. In this way analysis time, consumption of solvents or reagents and, mobile phase (in the case of liquid chromatography) can be minimized. Cluster analysis and principal component analysis are successfully applied to find greener solvent alternatives.

Keywords: design of experiment; desirability function; green solvents; green chemistry.

1. Introduction

Chemometric tools are widely used in analytical chemistry to group objects and variables, categorize them, predict characteristics, reduce the dimensionality of datasets for their easier evaluation. They are widely applied at the data treatment stage in the characterization of environmental media [1], food products (especially for their authentication) [2] or at the stage of signals processing [3, 4].

Green analytical chemistry is aimed at the reduction of environmental and human health impacts related to analytical processes [5]. These goals are reached by the application of different strategies, such as using direct analytical methodologies [6], solventless extraction techniques [7], microextraction techniques [8], the application of greener mobile phases and their lower volumes in liquid chromatography [9] among many others. In this context, the application of chemometric techniques and related numerical tools is rather overlooked in the scientific literature.

The aim of this paper is to present the chemometrics applications for greener analytical chemistry. For this purpose design of experiment, multiobjective procedure optimization and optimal process parameters selections are summarized.

2. Design of Experiment

Experimental design or design of experiment (DoE) is quite widely applied in the optimization of analytical procedures but it is rarely considered as useful in green analytical chemistry. Screening designs are useful as they allow to identify the factors that have the biggest influence on the response of the system. In this way, the material and energy savings are made by simply leaving the factors that do not influence analytical response without further interest [10]. Then, the number of experiments to be performed for 3 factors system is 8 when the factorial design is applied, 17 for application of Box-Behnken design and 20 for central composite design [11]. After the application of factorial design, the response function is investigated to find the values of variables that are close to optimal.

What is more, the application of DoE makes finding the optimal values more probable than changing one variable at a time approach. As it can be seen in figure 1 the second approach does not cover the entire surface of variable range [12]. The same number of measurements results in obtaining better information on the system for DoE approach or fewer measurement points are needed to obtain the same information than it is in the case of changing one variable at a time approach. Changing both variables (or all included ones) seems to be a better approach.

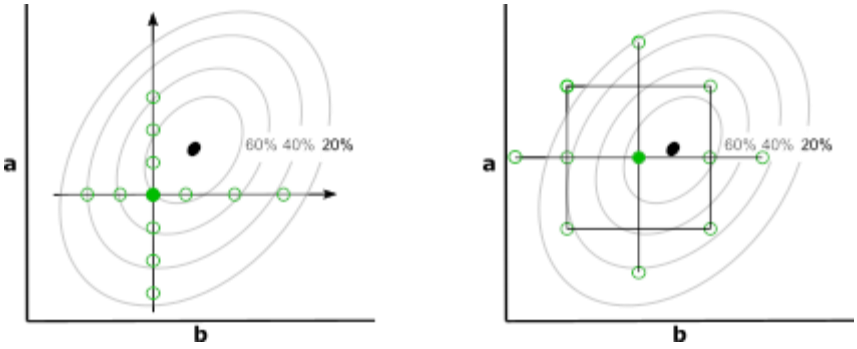

Figure 1. The response surface in case of change one at a time optimization (left) and design of experiment (right).

DoE is also applied for screening the variables relevant to the optimization of the procedure. It results in saving of time and materials as variables without impact (or with just small impact) on response are not considered. A good example can be the Plackett-Burman design that is applied for screening of variables to find significant ones that influence the performance of procedure for selenium species determination in water samples. After the reduction of variables by two, the remaining five are optimized with a central composite design [13]. Another example can be the

optimization of extraction procedure of phenols with deep eutectic solvents [14]. Only seventeen runs are needed to optimize sample to solvent ratio, extraction time and extraction temperature. Similarly, the reduction of initial seven variables to three (extraction solvent volume, agitation time and buffer volume) can be done for dispersive liquid-liquid microextraction of fluconazole [15]. Then the aim of optimization is to obtain good recovery with minimized extraction solvent consumption.

LC chromatographic conditions can be also optimized [16]. Two-level fractional factorial design is applied to screen for relevant parameters and Box-Behnken design to find optimal values of the separation process. Different responses are investigated, including resolution between peaks, tailings of both peaks, run time and number of theoretical plates, analytical eco-scale and HPLC-environmental impact assessment scores. The responses of the system are investigated with overlay plots, which seem to be appropriate to treat 2-3 responses together but for more responses, other multiobjective optimization tools could be more appropriate.

3. Multiobjective optimization

DoE even better fulfils the principles of green analytical chemistry if it is applied together with a multiobjective optimization method such as Derringer's desirability function [17]. In this function, the responses are transformed into individual desirabilities that are expressed in unitless zero – one scale [18]. Typically in analytical chemistry, the responses of the system should be minimized or maximized as presented in figure 2.

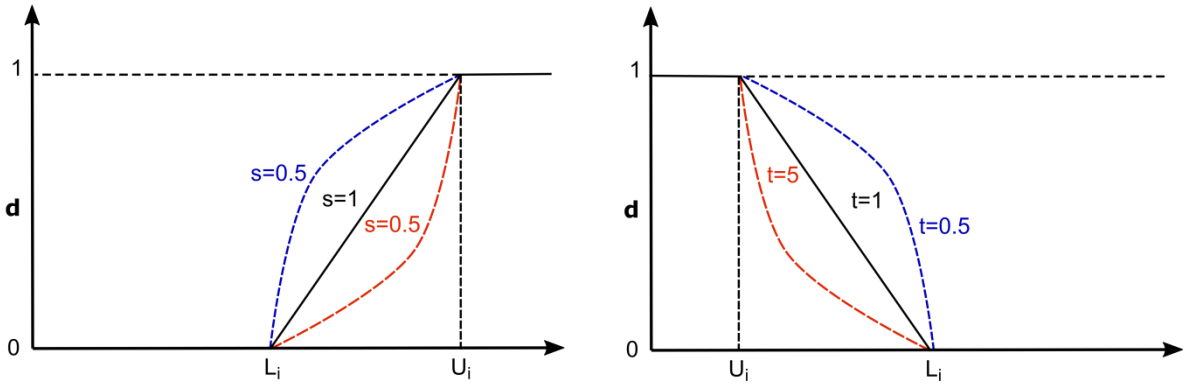

Figure 2. The transformation of values into individual desirabilities a) for the response that should be maximized; b) for the response to be minimized.

After the calculation of individual desirabilities d_n, the global desirability D is calculated according to the equation:

$$D = (d_1^{r1} \times d_2^{r2} \times \ldots \times d_n^{rn})^{1/n} \qquad (1)$$

Where r_1, r_2 (…) r_n are weights that express the relative importance of individual desirabilities. It should be noted, that if any of the individual desirabilities equal 0 then the global desirability D also gives a non-desirable response.

The multiple variables to be optimized are typically chromatographic peak areas, peak resolutions, precision of responses but also the time of chromatographic separation, which can be treated as green chemistry – related parameter [19, 20]. Optimization of peaks resolution and chromatographic separation is a very good example as general improvement of one variable can result in deterioration of performance of the second one. The resolution is variable to be maximized, while separation time is the variable to be minimized [21].

Six responses (LC separation time, tailing factors and resolutions) are simultaneously optimized for the separation of active compounds and their impurities [22]. As tailing factors and resolutions are directly referring to the quality of separation, proper optimization of separation time can result in decrease of mobile phase consumption, which is the green analytical chemistry parameter. Similarly, separation time and the resolutions for sixteen analytes, i.e. tryptophan derivatives are optimized with DoE and desirability function as well as more suitable chromatographic column is selected [23]. The optimization of LC separation with desirability function is also done for non-steroid anti-inflammatory drugs [24]. The responses considered are peak areas, resolutions, retention time of last eluting compound and environmental impact of mobile phase (the function of volume and hazards).

Desirability functions can be applied in the development of green extraction procedures [25], by defining the responses of the system that refer to procedure greenness. It is done for the extraction of organophosphorus pesticides with magnetic solid-phase extraction [26]. The optimized parameters are responses for four analytes, equilibration time to be minimized and pH value, salt content and amount of sorbent to be in pre-defined ranges.

4. Selection process conditions

One of the typical applications of chemometrics is a grouping of variables or objects [27], typically with cluster analysis or principal component analysis. Finding similar objects (of similar properties that are grouped) is a good clue in finding alternatives for a given process. Such an approach can be applied for the selection of greener alternatives.

Sustainable Solvent Selection and Substitution (SUSSOL) tool is based on Kohonen self-organizing maps clustering of ~500 solvents that is based on their physicochemical properties [28]. The grouping is presented in figure 3. Three groups are formed according to two dimensions. The dimension D1 is generally referred to as volatility with low boiling point solvents to the left, and D2 can be read as the measure of solvents polarity. After finding solvent substitutes with similar physicochemical properties, safety, health and environment criteria are considered to select green solvent.

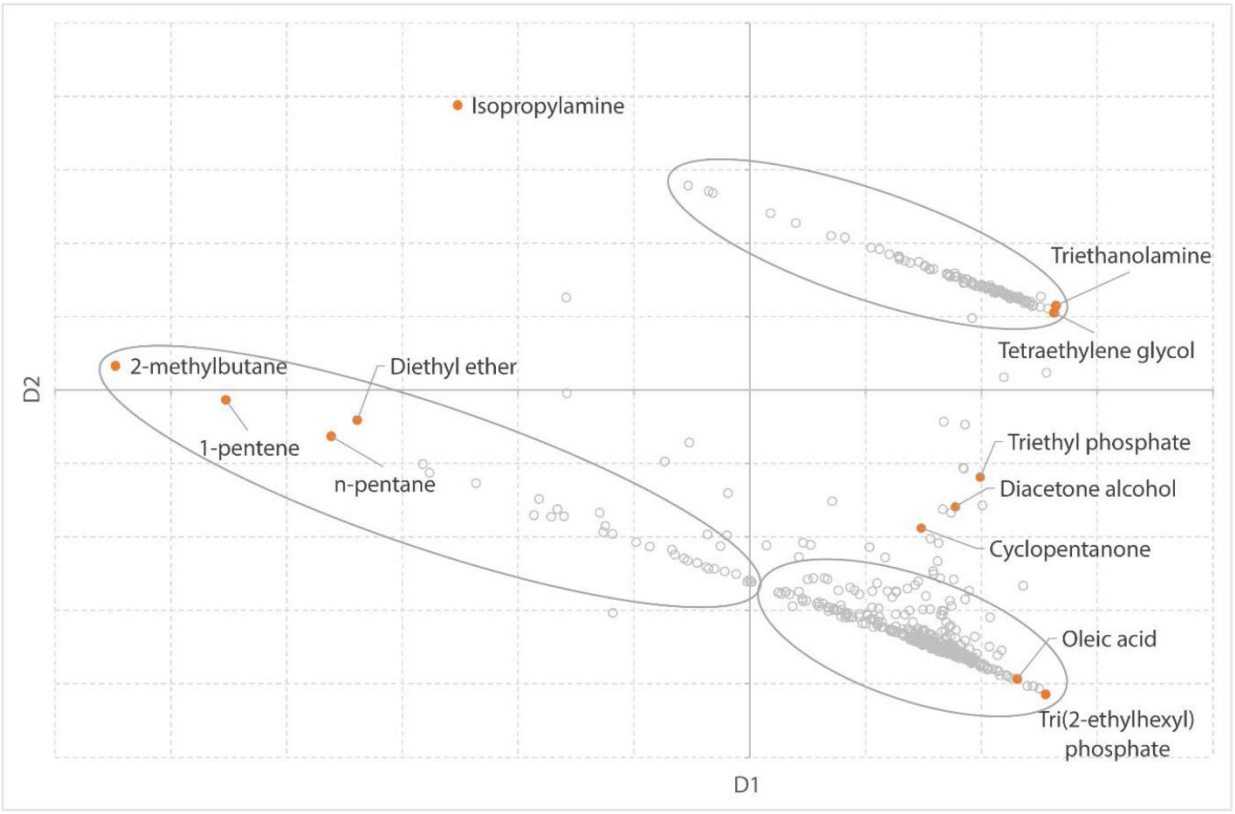

Figure 3. Grouping of solvents obtained with SUSSOL. The well-separated groups are formed. Reprinted from [28], licensed under a Creative Commons Attribution (CC BY) license.

Another chemometric approach is based on a grouping of solvents according to physicochemical properties with cluster analysis [29]. In this study, three clusters are formed – polar solvents, nonpolar and volatile solvents and the third cluster consists of nonpolar and rather nonvolatile solvents. Then for each cluster ranking of solvents is performed with multicriteria decision analysis with several criteria describing solvents toxicity, degradation potentials, safety of application. Cluster analysis is also used to calculate solvent similarity indexes [30]. With this tool, the substitutes for a given solvent or solvent mixture can be found on the basis of the distance between solvents on the dendrogram obtained by the classification of 261 solvents.

The third similar approach incorporates principal component analysis at the stage of a chemometric grouping of solvents [31]. The first six principal components (PCs) carry the explanation power of 87.9 % if the initial variability of the system PC1 represents the polarity variability of the solvents, PC2 the ability of the solvent to create a hydrogen bond. After finding the solvent on the PC1-PC2 plot (as the one explaining the biggest part of initial variability), the potential substitutes can be easily found as they are close neighbors. It is performed in user-friendly software.

All three above mentioned examples show that a big enough dataset of physicochemical properties describing organic solvents treated with unsupervised (the algorithm finds patterns in

the dataset without pre-defined assumptions) chemometric techniques nicely separate solvents according to volatility and polarity. As these two variables are typically the main descriptors during solvent selection for a given purpose, chemometric unsupervised classification can be very helpful to find greener alternative solvent of similar properties.

5. Conclusions

The application of chemometrics in green procedure development is still limited. The examples shown in this contribution prove that the greenness of analytical procedure and the optimization process itself can be improved with the adaptation of chemometric tools. This can be obtained with a multivariate selection of greener solvents, application of DoE and including greenness variables in multi-response optimization.

ACKNOWLEDGEMENTS

References:

[1] M. F. Dupont, A. Elbourne, D. Cozzolino, J. Chapman, V. K. Truong, R.J. Crawford, K. Latham, Chemometrics for environmental monitoring: a review, *Anal. Methods* 12(38) (2020) 4597-4620. https://doi.org/10.1039/d0ay01389g

[2] I. Reinholds, V. Bartkevics, I. C. Silvis, S. M. van Ruth, S. Esslinger, Analytical techniques combined with chemometrics for authentication and determination of contaminants in condiments: A review, *J. Food Compost. Anal.* 44 (2015) 56-72. https://doi.org/10.1016/j.jfca.2015.05.004

[3] N. P. Kalogiouri, R. Aalizadeh, M. E. Dasenaki, N. S. Thomaidis, Application of High Resolution Mass Spectrometric methods coupled with chemometric techniques in olive oil authenticity studies-A review, *Anal. Chim. Acta* 1134 (2020) 150-173. https://doi.org/10.1016/j.aca.2020.07.029

[4] D. Salazar-Rojas, A. Intilangelo, S. E. Vignaduzzo, R. M. Maggio, Development and validation of a green method for dissolution monitoring of pharmaceutical combinations. Meloxican and pridinol, *J. Pharm. Biomed. Ana.* 170 (2019) 228-233. https://doi.org/10.1016/j.jpba.2019.03.038

[5] M. Koel, Do we need Green Analytical Chemistry?, *Green Chem.* 18 (2016) 923-931. https://doi.org/10.1039/c5gc02156a

[6] S. Zivkovic, J. Savovic, M. Kuzmanovic, J. Petrovic, M. Momcilovic, Alternative analytical method for direct determination of Mn and Ba in peppermint tea based on laser induced breakdown spectroscopy, *Microchem. J.* 137 (2018) 410-417. https://doi.org/10.1016/j.microc.2017.11.020

[7] N. Lorenzo-Parodi, W. Kaziur, N. Stojanović, M. A. Jochmann, T. C. Schmidt, Solventless microextraction techniques for water analysis, *Trends Anal. Chem.* 113 (2019) 321-331. https://doi.org/10.1016/j.trac.2018.11.013

[8] J. M. Kokosa, Advances in solvent-microextraction techniques, *Trends Anal. Chem.* 43 (2013) 2-13. https://doi.org/10.1016/j.trac.2012.09.020

[9] C. J. Welch, N. Wu, M. Biba, R. Hartman, T. Brkovic, X. Gong, R. Helmy, W. Schafer, J. Cuff, Z. Pirzada, L. Zhou, Greening analytical chromatography. *Trends Anal.l Chem.* 29(7) (2010) 667-680. https://doi.org/10.1016/j.trac.2010.03.008

[10] B. Dejaegher, B., Y. V. Heyden, Experimental designs and their recent advances in set-up, data interpretation, and analytical applications, *J. Pharm. Biomed. Anal.*, 56(2) (2011) 141-158. https://doi.org/10.1016/j.jpba.2011.04.023

** The paper is good introduction to types of experimental designs.*

[11] S. T. Narenderan, S. N. Meyyanathan, V. V. S. R. Karri, Experimental design in pesticide extraction methods: A review, *Food Chem.* 289 (2019) 384-395. https://doi.org/10.1016/j.foodchem.2019.03.045

[12] D. B. Hibbert, Experimental design in chromatography: a tutorial review, *J. Chromatogr. B,* 910 (2012) 2-13. https://doi.org/10.1016/j.jchromb.2012.01.020

** The paper discussed good practices in reporting DoE studies, software applied and different designs.*

[13] B. Mostafavi, A. Feizbakhsh, E. Konoz, H. Faraji, Hydrophobic deep eutectic solvent based on centrifugation-free dispersive liquid–liquid microextraction for speciation of selenium in aqueous samples: one step closer to green analytical chemistry, *Microchem. J.* 148 (2019) 582-590. https://doi.org/10.1016/j.microc.2019.05.021

[14] R. Canales, M. Espino, S. Pasini, M. F. Silva, Chemometric and green metrics strategies for sustainable analytical methods: phenolic compounds in lettuce-NADES extracts, *Anal. Methods* (2021). https://doi.org/10.1039/d0ay02318c

[15] B. J. Moreira, L. A. Schiave, R. Martinez, S. G. Dias, C. M. de Gaitani, Dispersive liquid–liquid microextraction followed by green high-performance liquid chromatography for fluconazole determination in cerebrospinal fluid with the aid of chemometric tools, *Anal. Methods* 12(24) (2020) 3106-3114. https://doi.org/10.1039/d0ay00704h

[16] A. A. Habib, S. F. Hammad, S. M. Megahed, A. H. Kamal, Innovative Quality by Design Approach for Development of Green Micellar HPLC Method for Simultaneous Determination of Atorvastatin and Amlodipine, *Chromatographia* 83(10) (2020) 1221-1231. https://doi.org/10.1007/s10337-020-03937-5

*** The approach presented in this article includes green analytical chemistry metrics in procedure development. Environmental assessment tool and ecoscale scores are calculated for different HPLC conditions included in design of experiment.*

[17] L.. V. Candioti, M. M. De Zan, M. S. Cámara, H. C. Goicoechea, Experimental design and multiple response optimization. Using the desirability function in analytical methods development, *Talanta* 124 (2014) 123-138. https://doi.org/10.1016/j.talanta.2014.01.034

[18] M. A. Bezerra, S. L. C. Ferreira, C. G. Novaes, A. M. P. Dos Santos, G. S. Valasques, U. M. F. da Mata Cerqueira, J. P. dos Santos Alves, Simultaneous optimization of multiple responses and its application in Analytical Chemistry–A review, *Talanta*, 194 (2019) 941-959. https://doi.org/10.1016/j.talanta.2018.10.088

[19] M. R. Hadjmohammadi, S. S. S. Nazari, Optimization of separation of flavonoids in micellar liquid chromatography using experimental design and Derringer's desirability function, *J. Liq. Chromatogr. Relat. Technol.* 36(7) (2013) 943-957. https://doi.org/10.1080/10826076.2012.678458

[20] M. Hadjmohammadi, V. Sharifi, Simultaneous optimization of the resolution and analysis time of flavonoids in reverse phase liquid chromatography using Derringer's desirability function, *J. Chromatogr. B,* 880 (2012) 34-41. https://doi.org/10.1016/j.jchromb.2011.11.012

[21] F. Safa, M. R. Hadjmohammadi, Simultaneous optimization of the resolution and analysis time in micellar liquid chromatography of phenyl thiohydantoin amino acids using Derringer's desirability function, *J. Chromatogr. A,* 1078(1-2) (2005) 42-50. https://doi.org/10.1016/j.chroma.2005.04.081

[22] F. H. Edrees, A. S. Saad, M. T. Alsaadi, N.H. Amin, N. S. Abdelwahab, Experimentally designed chromatographic method for the simultaneous analysis of dimenhydrinate, cinnarizine and their toxic impurities, *RSC Adv.* 11(3) (2021) 1450-1460. https://doi.org/10.1039/d0ra09585k

[23] W. Setyaningsih, I. E. Saputro, C. A. Carrera, M. Palma, C. G. Barroso, Multiresponse optimization of a UPLC method for the simultaneous determination of tryptophan and 15 tryptophan-derived compounds using a Box-Behnken design with a desirability function, *Food Chem.* 225 (2017) 1-9. https://doi.org/10.1016/j.foodchem.2016.12.034

[24] A. Dogan, M. Tobiszewski, Optimization of liquid chromatographic separation of pharmaceuticals within green analytical chemistry framework, *Microchem. J.* 152 (2020) 104323. https://doi.org/10.1016/j.microc.2019.104323

[25] M. Bystrzanowska, M. Tobiszewski, Multi-objective optimization of microextraction procedures. *Trends Anal. Chem.* 116 (2019) 266-273. https://doi.org/10.1016/j.trac.2018.12.031

[26] H. Heidari, H. Razmi, Multi-response optimization of magnetic solid phase extraction based on carbon coated Fe3O4 nanoparticles using desirability function approach for the determination of the organophosphorus pesticides in aquatic samples by HPLC–UV, *Talanta* 99 (2012) 13-21. https://doi.org/10.1016/j.talanta.2012.04.023

[27] M. Bystrzanowska, M. Tobiszewski, Chemometrics for selection, prediction, and classification of sustainable solutions for green chemistry—A review, *Symmetry* 12(12) (2020) 2055. https://doi.org/10.3390/sym12122055

[28] H. Sels, H. De Smet, J. Geuens, SUSSOL—Using Artificial Intelligence for Greener Solvent Selection and Substitution, *Molecules*, 25(13) (2020) 3037. https://doi.org/10.3390/molecules25133037

*** It is good presentation how chemometric tool can be applied in practice. The procedure to find alternative for N-methyl pyrrolidone is an example.*

[29] M. Tobiszewski, S. Tsakovski, V. Simeonov, J. Namieśnik, F. Pena-Pereira, A solvent selection guide based on chemometrics and multicriteria decision analysis, *Green Chem.* 17(10) (2015) 4773-4785. https://doi.org/10.1039/c5gc01615k

[30] M. D. Driver, C. A. Hunter, Solvent similarity index, *Phys. Chem. Chem. Phys.* 22(21) (2020) 11967-11975. https://doi.org/10.1039/d0cp01570a

[31] L. J. Diorazio, D. R. Hose, N. K. Adlington, Toward a more holistic framework for solvent selection, *Org. Process Res. Dev.* 20(4) (2016) 760-773. https://doi.org/10.1021/acs.oprd.6b00015