



## Pre-selection and assessment of green organic solvents by clustering chemometric tools



Marek Tobiszewski<sup>a</sup>, Miroslava Nedyalkova<sup>b,\*</sup>, Sergio Madurga<sup>c</sup>, Francisco Pena-Pereira<sup>d</sup>, Jacek Namieśnik<sup>a</sup>, Vasil Simeonov<sup>b</sup>

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

<sup>b</sup> Faculty of Chemistry and Pharmacy, University of Sofia “St. Kl. Ohridski”, 1164 Sofia, J. Bourchier Blvd. 1, Bulgaria

<sup>c</sup> Materials Science and Physical Chemistry Department & Research Institute of Theoretical and Computational Chemistry (IQTUB) of Barcelona University (UB), C/ Martí i Franquès, 1, 08028 Barcelona, Catalonia, Spain

<sup>d</sup> Analytical and Food Chemistry Department, Faculty of Chemistry, University of Vigo, Campus As Lagoas-Marcosende s/n, 36310 Vigo, Spain

### ARTICLE INFO

#### Keywords:

Green chemistry  
Solvents  
Chemometrics  
Bioconcentration factor  
Greenness assessment

### ABSTRACT

The study presents the result of the application of chemometric tools for selection of physicochemical parameters of solvents for predicting missing variables – bioconcentration factors, water-octanol and octanol-air partitioning constants. EPI Suite software was successfully applied to predict missing values for solvents commonly considered as “green”. Values for logBCF, logK<sub>OW</sub> and logK<sub>OA</sub> were modelled for 43 rather nonpolar solvents and 69 polar ones. Application of multivariate statistics was also proved to be useful in the assessment of the obtained modelling results. The presented approach can be one of the first steps and support tools in the assessment of chemicals in terms of their greenness.

### 1. Introduction

Green chemistry is the concept introduced by Anastas and Warner (1998) with the publication of the twelve principles that are specific guidance on the introduction of sustainability to chemical science. Since then this concept has developed much and efforts were made to develop zero-waste technologies, design benign products that maintain their properties, find renewable and bio-based feedstock for chemicals or apply energy-efficient technologies. Also, solvents gained a lot of attention, the fifth principle of green chemistry states that solvents should not be applied if possible (Jessop, 2016) otherwise they should be as inert to the environment as possible.

As the application of solvents cannot be avoided in many technological processes, it is highly desired to use green solvents. The green solvent is characterised by preferential environmental, health and safety (EHS) parameters (Capello et al., 2007). The first, very basic information about solvent greenness can be obtained from its physicochemical parameters and phase distribution constants. For example, solvents with low boiling points are very volatile; therefore the exposure by inhalation is very likely. High values of octanol – water partitioning coefficients give initial information indicating that the compound can be accumulated in the animal tissues. More specific information on solvents greenness can be obtained from toxicological and

ecotoxicological data, such as oral (Sathish et al., 2016) or inhalation toxicities, toxicity towards aquatic organisms or carcinogenicity (Tobiszewski and Namieśnik, 2015). Similarly, environmental persistence data such as biodegradability or hydrolysis potential give information about environmental related hazards. Remarkably, the European REACH Regulation (EC No 1907/2006) (“Regulation (EC) No 1907/2006 – REACH – Safety and Health at Work – EU-OSHA”, 2017) has set as a priority the assessment of chemicals’ bioaccumulative potential, which is the potential of a substance to accumulate in biota and, eventually, to pass through the food chain. A parameter that is widely used to measure a chemical substance’s bioaccumulative potential is the bioconcentration factor (BCF). BCF is commonly defined as the ratio between the concentration of a chemical substance present in an aquatic organism and in the surrounding environment at thermodynamic equilibrium under controlled laboratory conditions (Arnot and Gobas, 2006).

One of the problems related to the assessment of solvents in terms of their greenness is the non-availability of data that are required (Alder et al., 2016). The missing data can be approximated with the average value for a given class of chemicals. However, usually, the obtained assessment estimation is characterised by high uncertainty. Therefore, it is desired to find reliable methods for the predictions of missing data values. It is especially important in the case of solvents that are

\* Corresponding author.

E-mail address: [mici345@yahoo.com](mailto:mici345@yahoo.com) (M. Nedyalkova).

<http://dx.doi.org/10.1016/j.ecoenv.2017.08.057>

Received 26 February 2017; Received in revised form 20 August 2017; Accepted 21 August 2017

Available online 14 September 2017

0147-6513/ © 2017 Elsevier Inc. All rights reserved.

relatively novel and still poorly characterised, like esters or ethers derived from renewable feedstocks (Pena-Pereira et al., 2015).

There are several methodologies for modelling the bioconcentration values of various chemicals (Pavan et al., 2008). Quantitative structure-activity relationship (QSAR) applied to a dataset of chemicals allowed to predict logBCF with the model with an  $r^2 = 0.491$  (Petoumenou et al., 2015). QSAR supported by partial least square modelling allowed to obtain fit the model with an  $r^2 = 0.868$  (Qin et al., 2009). Also, an extensive investigation was made of the physico-chemical properties of 152 solvents in a search for quantitative structure–property relationships (QSPR) (Gramatica et al., 1999). The artificial neural network followed by relatively simple generic model also allowed to establish the values of logBCF with acceptable accuracy (Fatemi et al., 2003). Application of linear and nonlinear models allowed to model logBCFs of 107 pesticides successfully (Yuan et al., 2016).

Although different logBCF prediction methods have been mainly applied to organic pollutants, efforts to create models for predictions of these values for solvents, according to authors' knowledge were not made.

The aim of this study is to interpret two sets of solvents described by several physicochemical and biological features by multivariate statistics. Also, we present the assessments of physicochemical properties such as logarithms of partitioning constants between octanol and air ( $\log K_{OA}$ ), octanol and water ( $\log K_{OW}$ ) and logBCF by using the Estimation Programs Interface (EPI) Suite. Results of the EPI Suite predictions of the physical properties of the solvents sets and comparisons with the available literature values are validated. The above-mentioned software is a valuable tool to be applied to experimental data for such kind of solvents properties when there is lack of sufficient data. In this way, it is possible to find the relationships that are the basis for modelling of solvents parameters that define their greenness. Easy but reliable prediction of hazards related to greener solvents introduction is highly desired. The paper presents the application of multivariate statistics tools in the prediction of unknown properties and assessment of their modelling results.

## 2. Materials and methods

### 2.1. Dataset

The dataset consists of 43 solvents – *non-polar and sparingly volatile solvents* and 69 solvents *a priori* categorised as *polar* solvents. In this study, the datasets are treated separately. This *a priori* classification is based on previous research (Tobiszewski et al., 2015), where large dataset of 151 solvents was analysed with cluster analysis using Melting and boiling point, density, water solubility, vapour pressure, Henry's law constant,  $\log K_{OW}$ ,  $\log K_{OA}$  and surface tension as physicochemical parameters for solvent classification. The identified third cluster – (conditionally determined as “*non-polar and volatile*” solvents) in Fatemi et al. (2003), is not investigated in the present study, as solvents grouped there (mainly chlorinated solvents or other causing EHS problems) are out of interest in the light of green chemistry. Not every process can be carried out with water miscible solvents. In fact, immiscible solvents are required in a wide number of processes (e.g., extraction and separation processes, synthesis, etc.), so their greenness evaluation is also of high relevance.

Each solvent in the present study was characterised by 10 parameters, namely melting point, boiling point, density, surface tension, water solubility, vapour pressure and Henry's law constant, as well as logarithms of partitioning constants between octanol and air ( $\log K_{OA}$ ) and octanol and water ( $\log K_{OW}$ ) and logarithm of bioconcentration factor (logBCF). The physicochemical parameters were extracted from material safety data sheets of chemicals and from Handbook On Physical-Chemical Properties And Environmental Fate For Organic Chemicals (Mackay et al., 2006). Three additional parameters were included for the chemometric analysis, namely,  $\log K_{OAc}$ ,  $\log K_{OWcalc}$ , and  $\log BCF_{calc}$ . These parameters were calculated with models described in Section 2.3.

$\log K_{OAc}$ , and  $\log BCF_{calc}$ . These parameters were calculated with models described in Section 2.3.

### 2.2. Multivariate statistics

Hierarchical clustering (HC) is a well-documented approach to the unsupervised pattern recognition (Massart et al., 1983; Massart and Kaufman, 1998). It aims to select groups of similar objects (clusters) within different data sets and to interpret the meaning of the clustering either between the objects of interest or between the parameters used for the description of the objects. Usually, the hierarchical cluster analysis requires several steps in performing the algorithm of clustering: standardisation of the raw data (in order to avoid the effect of the different dimensionality of parameters); determination of the distance between the objects for clustering (in order to introduce a similarity measure); procedure for linkage. The results are normally presented on a tree-like plot called dendrogram and in the final stage, a criterion for determination of the cluster significance is needed in order to improve the interpretation. The use of chemometrics for the treatment of different data sets provides a valuable tool for objective decision-making (Hristov et al., 2016; Nedyalkova et al., 2017).

Principal component analysis (PCA) is one of the several multivariate methods that allows us to explore patterns in complex data sets allowing to classify the information and detect structure in a diffuse data set. In general, PCA is a mathematical treatment of the input data matrix (objects described by many features or variables) where the goal is to represent the variation present in many variables by a small number of factors or latent variables. A new space of the features is formed which it makes possible to visualise and project the multivariate nature of the data set.

The central task in PCA is to reduce the original dimension of the input matrix  $X$  to two parts – factor loadings (part  $A$  matrix) and factor scores (part  $F$  matrix). The first one includes the weights of each feature (variable) in each identified factor (new latent variable). The higher the weights the higher is the contribution of the original variable. Thus, this procedure allows us to identify which variables influence the objects.

If the objects have to be presented in the space of the new latent variables, then the factor scores matrix must be used. The specific rules for performing and interpreting PCA are presented, for instance (Einax et al., 1997).

### 2.3. Modelling – EPI Suite™

In the current work the following subprograms of EPI Suite™ version 4.10 were used: KOAWIN™, KOWWIN™ and BCFBAF™. EPI Suite™ is available from the US Environmental Protection Agency – (US EPA et al., n.d.) (Computer Program Estimation Programs Interface Suite™ for Microsoft® Windows Version 4.10, available on <http://www.epa.gov/oppt/exposure/pubs/episuite.htm>)

This KOAWIN™ program estimates the logarithm of the octanol-air partition coefficient ( $K_{OA}$ ) of an organic compound using the compound's octanol-water partition coefficient ( $K_{OW}$ ) and Henry's Law constant (HLC). KOAWIN requires only a chemical structure to estimate  $K_{OA}$ . Structures are entered into KOAWIN through SMILES (Simplified Molecular Input Line Entry System) notations, which are also used by other estimation programs in EPA's EPI Suite. It is possible to estimate  $K_{OA}$  from the octanol-water partition coefficient ( $K_{OW}$ ) and Henry's law constant ( $H$ ) by the following equation:  $K_{OA} = K_{OW} (RT)/H$ , where  $R$  is the ideal gas constant and  $T$  is the absolute temperature.  $K_{OA}$  and  $K_{OW}$  are unitless values.  $H/RT$  is the unitless Henry's law constant, also known as the air-water partition coefficient ( $K_{AW}$ ) (Meylan and Howard, 1995). Therefore, the equation to estimate  $K_{OA}$  is:  $K_{OA} = K_{OW}/K_{AW}$ .

The KOWWIN™ program predicts the logarithm of the octanol-water partition coefficient. KOWWIN uses a “fragment constant” methodology to predict logP. In a “fragment constant” method, a structure is divided

Download English Version:

<https://daneshyari.com/en/article/5747862>

Download Persian Version:

<https://daneshyari.com/article/5747862>

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