Chemometrics and Intelligent Laboratory Systems 150 (2016) 41–50

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Common components and specific weights analysis: A tool for metabolomic data pre-processing

Elodie Dubin [a], Marc Spiteri [a], Anne-Sophie Dumas [a], Jérôme Ginet [a], Michèle Lees [a], Douglas N. Rutledge [b,*]

[a] Eurofins Analytics France, 9 Rue Pierre Adolphe Bobierre, B.P. 42301, F-44323 Nantes Cedex 3, France
[b] UMR Ingénierie Procédés Aliments, AgroParisTech, Inra, Université Paris-Saclay, 91300 Massy, France

ARTICLE INFO

Article history:
Received 4 August 2015
Received in revised form 9 November 2015
Accepted 13 November 2015
Available online 17 November 2015

Keywords:
LC-TOF-MS
Metabolomics
Pre-processing
Intensity normalisation
Common Components and Specific Weights Analysis

ABSTRACT

The metabolomic approach using LC-MS analyses suffers from substantial intensity variability which must be corrected before extracting useful biological information. In this paper, Common Components and Specific Weights Analysis (CCSWA) is proposed as a novel method for the correction of this analytical bias. This method was compared to LOESS normalisation for within-batch correction and to the median of the quality controls for between-batch correction. In the first case, the correction of a non-continuous effect in the batch was investigated using both LOESS signal correction and CCSWA on fish samples. In the second case, four batches were analysed and combined to create a larger cohort of honey samples. CCSWA was successfully applied to correct both within- and between-batch effects observed in the LC-MS signals.

1. Introduction

Metabolomics was first developed in the clinical and pharmaceutical areas primarily to compare metabolomes of different classes of patients and identify markers specific to each class (e.g. control vs. treated). This metabolomic approach has recently been extended to other fields such as food control where it is used to complement targeted analyses, as the untargeted nature of metabolomics makes it possible to access information about both endogenous and exogenous compounds in samples.

Such untargeted analyses are carried out on different analytical platforms. In the case of LC-MS (Liquid Chromatography-Mass Spectrometry) analyses, pre-processing steps are necessary to correct for analytical biases before applying any statistical tools to extract interpretable information [1]. The idea of pre-processing is to remove systematic errors which are not due to the biological variations in the samples, such as dilution, detector sensitivity, ionisation processes, contamination of the source and retention time drifts, so as to get comparable data within batches and between batches [2]. Some normalisation strategies are based on applying a correction using a marker considered as unvarying (creatinine in urine), or on total ion intensity, or on "total useful signal" (compounds common to all samples) [3]. However, each compound may react differently during the analysis (because of detector saturation, matrix effects, dilution response), consequently a unique correction factor does not always make it possible to obtain an appropriate signal correction [2]. Probabilistic Quotient Normalisation (PQN) had been proposed to correct ^{1}H Nuclear Magnetic Resonance (NMR) spectra by scaling all the intensities in a spectrum using the most probable multiplicative factor calculated as the median of the quotients of the amplitudes of each point in a spectrum and a reference spectrum [4]. This corrections is not often applied to LC-MS data, since not all variables will react in the same way to the within- and between-batch variations. Variance Stabilisation Normalisation (VSN) has also been tested for correction of LC-MS data [5]. VSN adjusts sample-to-sample variations through shifting and scaling intensities and then transforming the intensities to a scale where the variance is constant over the entire data set. Although VSN applies a different correction factor to each variable in a signal, it is based on adjusting their variances. It is preferable to have a method that makes corrections based on the influence of variable variations that generate within- and between-batch drift.

Moreover, in large-scale studies, a considerable number of samples may have to be analysed, which implies more than one analytical

Abbreviations: CC, Common Components; CCSWA, Common Components and Specific Weights Analysis; ESI, Electrospray Ionisation; FWHM, Full Width at Half Maximum; LC-MS, Liquid Chromatography-Mass Spectrometry; LOESS, Locally Estimated Scatterplot Smoothing; NMR, Nuclear Magnetic Resonance; OSC, Orthogonal Signal Correction; PC, Principal Component; PCA, Principal Components Analysis; PQN, Probabilistic Quotient Normalisation; QC, Quality Control; QuEChERS, Quick Easy Cheap Efficient Rugged and Safe; RT, Retention Time; UPLC, Ultra High Performance Liquid Chromatography; SVD, Singular Value Decomposition; TOF-MS, Time of Flight Mass Spectrometry; VSN, Variance Stabilisation Normalisation.

* Corresponding author. Tel.: +33 1 44 08 16 48; fax: +33 1 44 08 16 53.
E-mail address: rutledge@agroparistech.fr (D.N. Rutledge).

http://dx.doi.org/10.1016/j.chemolab.2015.11.005
0169-7439/

batch. In this case, the principal difficulty with LC-MS is the global differences in response between batches for a given variable, mainly due to the detector evolution and cleaning processes [6].

Different methods have been proposed to correct for analytical bias in a variable by variable manner. For the within-batch correction, Dunn et al. [6,7] proposed a quality control (QC) based LOESS signal correction (Locally Estimated Scatterplot Smoothing) to correct intensity within a batch. The QC is a sample representative of samples analysed during the batch and is injected all along the analytical run. It is used to evaluate the analytical platform stability and to correct intensity deviation. For that, each variable in each sample is individually corrected according to the evolution of its value in the neighbouring QCs. Concerning the between-batch correction, Van der Kloet et al. [8] proposed in 2009 a correction based on the average or on the median of the QC replicates analysed in all batches.

Common Components and Specific Weights Analysis (CCSWA) is a multiblock analysis originally developed by Qannari et al. [9] in order to solve a problem regarding sensory analyses. In the case of free-choice sensory analyses, the judges may use different descriptors. For a set of test-samples, each judge may produce a data matrix with the same number of rows (observations), but not the same number of columns (variables or descriptors). CCSWA makes it possible to analyse simultaneously these matrices, considering them as blocks, and extracts the information that is common among them. Moreover, each block receives a weight called “salience” that quantifies its contribution to each extracted component. CCSWA has since been used to study samples analysed on different instruments, to find relations between tables and to discriminate samples using the global information included in all tables. For example, Mazerolles et al. [10] applied it to the characterisation of food products using different analytical methods to obtain a global picture of the product. This technique has also been successfully applied for the combination of data obtained on the same samples by different laboratories and instruments [11].

In the present paper, CCSWA is proposed as a novel method for correction of analytical bias. This method has been successfully applied to correct both within- and between-batch effects observed in LC-MS analyses. The CCSWA results are compared to those obtained using the LOESS normalisation method for within-batch correction [6] and to the median of the QC for between-batch correction [8]. Within-batch correction with CCSWA is illustrated using signals acquired on fish samples and between-batch correction with honey samples. In the first case, the correction of a non-continuous effect in the batch was investigated using both LOESS signal correction and CCSWA. In the second case, four batches were analysed and combined to create a larger cohort. However, between-batch effects masked the relevant information and needed to be corrected. Interestingly, within-batch effect was not strong in any of those four batches and so no correction was necessary at this level.

2. Material and methods

2.1. Samples

2.1.1. Honey samples

57 commercial honeys of 6 different botanical origins (polyfloral, acacia, orange blossom, mountain, lavender and eucalyptus) and from various geographical sources (both from European countries such as France and Spain, and elsewhere in the world, such as Chile and Australia) were collected at different retail outlets: discount stores, supermarkets, luxury shops, or local retailers.

The samples were tested for authenticity using pollen analysis, Isotope Ratio Mass Spectrometry [12], and a recently published NMR-profiling approach [13]. All results were in agreement with the labelling description, showing no evidence of adulteration.

2.1.2. Fish samples

Samples of Alaska pollock consisted in 14 blocks of minced and compressed fish caught between January and August 2014. Samples were ground prior to extraction.

2.2. Chemicals and reagents

Reagents used were all of high analytical purity. Methanol (Fluka, ultra chromasolv), acetonitrile (Fluka, ultra chromasolv), ammonium acetate (Fluka, purity ≥ 99.0%), anhydrous magnesium sulphate (purity ≥ 99.5%), sodium acetate (purity ≥ 99.0%) and leucine enkephalin acetate salt hydrate (purity >95%) were purchased from Sigma-Aldrich and formic acid (Optima LC/MS grade) was from Fisher Chemical. 18 MΩ deionised water was provided by a purification system (Veolia Environment).

2.3. Sample extraction

2.3.1. Honey extraction

Samples were prepared using a modified QuEChERS method (Quick Easy Cheap Efficient Rugged and Safe) [14]. First, they were homogenized and if the honey was crystallised, pots were heated in a water bath at 40 °C for 30 min (as 40 °C is the standard hive temperature, no degradation of the honey is to be expected at this temperature).

For the extraction, 5 g of sample was weighed in 50 mL polypropylene tubes. After the addition of 4 mL of 18.2 mΩ water, samples were mixed for at least 10 min until complete dissolution. Then 10 mL of acetonitrile were added to the tubes, which were agitated again for 5 min. 2 g of magnesium sulphate and 2.5 g of sodium acetate were added to each sample and they were mixed again for 5 min. Samples were centrifuged at 4500 rpm for 5 min. 4 mL of supernatant was dried under nitrogen at 40 °C, then 80 μL of methanol, followed by 720 μL of 18.2 mΩ water, was added and samples were agitated briefly after each solvent addition. Finally, extracts were filtered through PVDF filter at 0.22 μm. A blank was prepared using the same method but without any honey. Each sample was prepared twice to give two technical replicates.

One within-batch quality control sample per batch was prepared by pooling an aliquot of all the samples in its batch. A between-batch quality control sample was prepared by pooling the acetonitrile extract of 18 honeys selected based on their botanical and geographical origins. Aliquots of this mix were frozen. One replicate of this between-batch control was analysed at each new batch.

2.3.2. Fish extraction

Samples were also prepared using a modified QuEChERS method. 10 g of ground fish sample was weighed in 50 mL polypropylene tubes. Then 10 mL of acetonitrile was added to the tubes, and they were agitated again for 10 min and sonicated for 10 min. 1 g of magnesium sulphate, 1 g sodium chloride, 1 g trisodium citrate dihydrate and 0.5 g sodium hydrogen citrate sesquihydrate were added to each sample and they were mixed again for 5 min. Samples were centrifuged at 4500 rpm for 10 min at −18 °C. 2 mL of supernatant was frozen at −20 °C for 1 h and samples were again centrifuged for 2 min at 4 °C to eliminate residual fat. 1 mL of supernatant was dried under nitrogen at 40 °C, then 100 μL of methanol, followed by 900 μL of 18.2 mΩ water, was added and samples were agitated briefly after each solvent addition. Finally, extracts were filtered through PVDF filter at 0.22 μm. A blank was prepared using the same method but without incorporating any of the fish sample. Each sample was prepared twice to give two technical replicates.

2.4. Sample analysis

Samples were analysed by UPLC-TOF-MS on a Xevo G2-S TOF (Waters) equipped with an Acquity I Class UPLC (Waters). Chromatographic separation was achieved with a 1.7 μm, 2.1 mm × 100 mm,