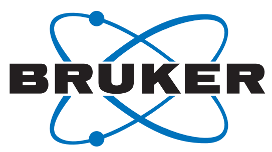

Application Note LCMS-86

Food Authenticity – Classification of Coffee Types Based on LC-MS

Introduction

The increasing global trade in food and beverages has created a commensurate demand for authentication and adulteration testing of products [1]. In the past, food scares have brought these procedures to the attention of a wide public. Traditionally, standard assays for food testing are based on measurements of a relatively small number of target compounds or characteristic features. Recent advances in LC-MS technology enable simultaneous measurement of many compounds in parallel and without prior knowledge of sample constituents (that is, a non-targeted approach). In the case of small molecule analysis, this is typically referred to as metabolomics or metabolic profiling. MS and NMR are the most widely used complementary instruments for detecting possible authenticity-defining compounds and adulterants.

An example of how technology helps to determine authenticity and species purity, detect false labeling, and monitor production process control and sample similarity is the NMR-based JuiceScreener™ [2]. Combined with SGF Profiling* the JuiceScreener™ delivers the required information from a single NMR experiment instead of multiple individual analytical steps.

Authors

Verena Tellström, Sven Brand, Michael Becker, Klaus Meyer, Gabriela Zurek, Aiko Barsch
Bruker Daltonik GmbH, Bremen, Germany

Keywords	Instrumentation and Software
compact	compact
Metabolomics	ProfileAnalysis
Classification	
Coffee	
PCA	
PLS	

Combining LC with high-resolution MS enables non-targeted analysis of hundreds or even thousands of small molecules. In a recent study, the **compact** QTOF System was used to analyze extracts from 13 different types of coffee capsule [3]. A non-targeted metabolomics workflow enabled differentiation of coffee types based on their flavor intensity and readily identified the compounds responsible for the differentiation. In this study, which serves as an example for food authenticity assessment in general, the established statistical PCA and PLS models were used to classify different coffee samples into similar or distinct groups, and to predict the relative intensity for each group.

Experimental

The experimental setup for the LC-MS measurements was the same as described in Bruker Application Note LCMS-79. Coffee extracts were diluted with water before three 5 µL replicates were analyzed by UHPLC-MS. Chromatographic separation was carried out using an RSLC system (Dionex) with a 50 x 2.1 mm BEH C18, 1.7 µm column (Waters) at a flow rate of 0.45 mL/min (Solvent A: Water + 0.1% HCOOH and Solvent B: methanol + 0.1% HCOOH). An LC gradient of a linear increase from 2% B to 98% B (over 5 min) and constant 98%B (for 1 min) was used. MS detection was performed using a **compact** QTOF mass spectrometer (Bruker Daltonics). The instrument was operated in ESI positive mode acquiring MS full scan data at an acquisition speed of 3 Hz.

After measuring samples for the statistical models [3], around 150 further samples – unrelated to the current study and including proteomics and small molecule samples of different complexity – were measured over the space of five weeks on the same **compact** QTOF instrument. Following these five weeks of "non-coffee-related" measurements, the measurements for this Application Note were performed. The samples evaluated in this study included one of the coffee samples used for creating the original model. This sample was stored at –20°C before re-injection. For this study, 8 coffee capsules were extracted by a different operator using a different espresso machine and 25 mL of water (instead of 35 mL as in the original study). These samples were diluted with water before injection to match the concentration of the other samples. The identity of the test samples was revealed before the LC-MS analysis ("non-blinded" experiment). As a "blinded" experiment, 8 capsules were extracted by yet another operator on a third espresso machine using 35 mL of water. These coffee samples originated from the same coffee vendor but their identity was not revealed until after the prediction of the coffee was performed using established PCA and PLS models.
A further sample – originating from a different vendor and with an assigned roasting degree of 6 (indicating a medium intensity) – was also extracted. All samples were injected as two technical replicates.

Data derived from the coffee extracts was classified using PCA and PLS models generated from the original 13 different types of coffee using the **profileanalysis** 2.1 software (Bruker Daltonik). The coffee intensity assigned by the manufacturer was used as Y Matrix for calculating the PLS model. The PLS model also allows prediction of the assigned intensity of the new extracts. It should be noted that the caffeine peak was not included for building

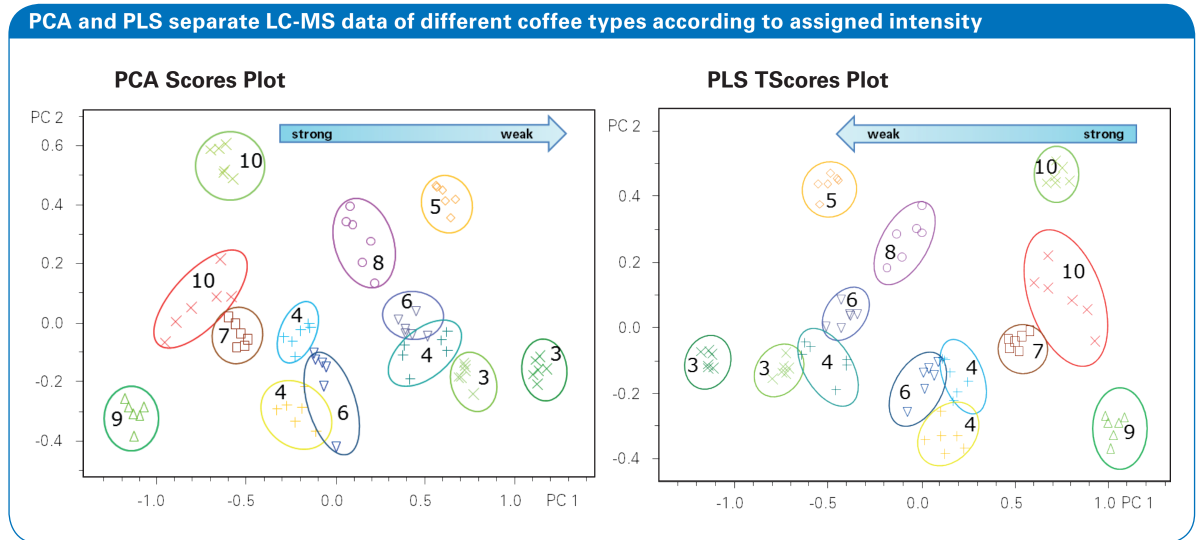

Figure 1: Scores plots for 13 coffee metabolic profiles measured on a compact LC-MS. A: PCA Scores plot reveals a separation of samples on PC1 according to coffee intensity assigned by the coffee manufacturer (numbers from 3 (weak) – 10 (strong)) B: PLS TScores plot using the same original data reveals a similar separation with an inverted clustering on PC1 compared to the PCA. The assigned coffee intensity was used as Y matrix for PLS calculation. Note that the caffeine peak was removed before generating both models.

PCA and PLS models, and therefore the caffeine amount was not taken into account for coffee intensity prediction. For sample classification, **profileanalysis** automatically performed data recalibration and applied the same parameters used for calculating the original models.

Results

PCA and PLS models can separate coffee types on the basis of the assigned coffee intensity

In a previous study [3] 13 different coffee types were analyzed by LC-QTOF MS. Figure 1 A and B show the PCA and PLS TScore plots generated from the original data. The caffeine peak was not included for building these pareto-scaled models. Both models could separate the coffee types on the basis of the intensity assigned by the coffee vendor.

Sample classification: Coffee sample stored at –20°C for 5 weeks

PCA and PLS models can not only help to determine the grouping of samples to find similarities or differences but can also be used for classifying new samples using existing models. Several validation experiments were performed to determine the validity of the PCA model established using the 13 different coffee types. First, a coffee sample used in building the original model was re-injected into the same **compact** LC-QTOF several weeks after the first samples were measured. During this five week interlude, around 150 proteomics and small molecule samples of different complexity were analyzed on the same MS, reflecting the normal workload of an analytical laboratory. For this study, this sample was injected twice using the same LC-MS method. The data was classified using the existing PCA model by automatically applying the same peak detection and bucketing parameters. As shown in Figure 2, the two samples cluster closely to a group of samples highlighted in green which correspond to the same coffee type as the archived sample. This indicates that the model is able to correctly classify the archived sample after the MS was in use for several weeks and the sample was frozen at –20°C.

Sample classification: different operator, extraction instrument, and extraction volume (“non-blinded” experiment)

Eight coffee samples were extracted by a different operator using a different espresso machine and extraction volume (25 mL instead of 35 mL). Two of the samples belonged to the same capsule type. The obtained data was classified using the existing PCA model. Figure 3 shows that all test samples clustered closely to the original samples. Knowing the identity of the test samples facilitated the assignment of some samples that clustered between two similar coffee blends. This experiment showed that even if standard operation procedures (SOPs) are not applied for the sample extraction, a reasonable classification is possible. However, we strongly recommend that SOPs are established for sample preparation and analysis in routine sample classifications.

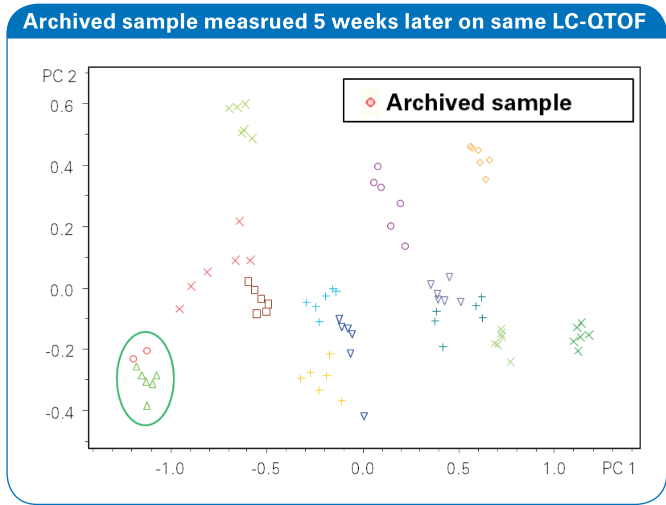

Figure 2: A sample used for building the original PCA model was stored at –20°C and re-injected five weeks after the first batch of samples was measured. The classification of these samples (red circles) using the established PCA model shows a close clustering to the previously measured injections of this sample (green triangles).

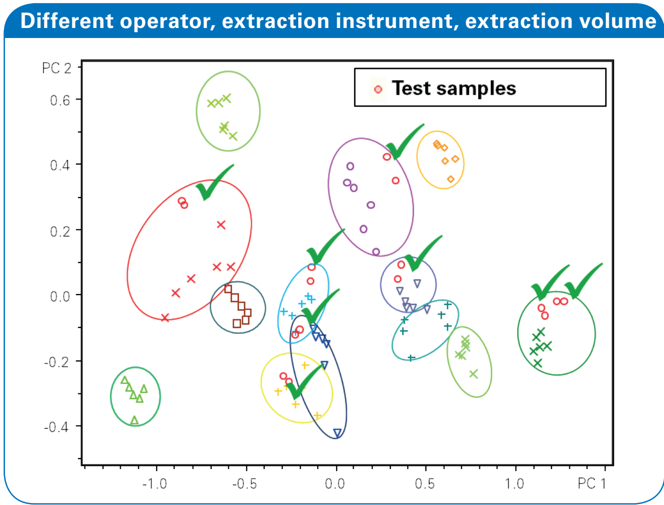

Figure 3: “Non-blinded” experiment: Eight samples were extracted by a different operator, on a different espresso machine, using a different extraction volume. Classification of two technical replicates using the original PCA model enabled a correct assignment of all samples. The sample ID was known before samples were classified.

Sample classification: different operator and extraction instrument in a "blinded" experiment

Subsequently, a further 8 coffee samples were extracted by a third operator using yet another espresso machine. This time, the same extraction volume (35 mL) was used but the identity of the samples was not revealed to the MS operator. Classification using the existing PCA model enabled assignment of the most likely sample type. Seven of eight samples could be correctly assigned. As shown in Figure 4, the incorrectly assigned coffee clusters closely to several different coffee types. The original model in this proof-of-concept study was built using two extracts per coffee type, each injected three times. The results of this "blinded" classification indicate that this model can be used for sample classification, but should ideally be extended with more data to make it more robust.

PLS-based coffee intensity prediction

Projection to latent structures by means of partial least squares (PLS) can be considered as a supervised regression extension of PCA [4]. One advantage of PLS models is the fact that they can be used not only for sample classification but also for prediction. PLS relates input data (X data; for example, LC-MS metabolic profiles) to output data (Y data; for example, coffee intensity). The assigned coffee intensity was used as Y matrix for calculating the PLS model from the original 13 coffee types. This enables prediction of the coffee intensity (Y) for new samples using the existing model. Table 1 shows the predicted coffee intensity for the 8 samples measured in the blinded experiment. The predicted coffee intensity closely matches the correct intensity assigned by the manufacturer. As in PCA, the medium-intensity coffee types show a slightly higher deviation from the assigned values, but the trend for strong and weak coffees is clearly revealed.

Classification and intensity prediction for a medium roast coffee from a different vendor

Finally, an extract from a coffee capsule from a different vendor was analyzed using the established method and classified using the existing models. These capsules (ZB) were described by the vendor as having a roasting degree of 6 (on a scale of 1–10, weak to strong). Several colleagues who formed a "non-expert" tasting panel described this blend as much more intense compared to the coffees used to build the model. The "ZB" coffee was extracted two times and injected twice into the **compact** LC-MS. Figure 5 A presents the corresponding PLS classification into the existing model. The samples cluster very closely to the intense coffees from the original model. Interestingly, they did not cluster close to the medium intensity samples, as would be expected from the vendor's assignment. Table 2 lists the predicted intensity for "ZB" coffee using the original PLS model.

The predicted intensity of ~12.5 might indicate that the intensity scale of each coffee vendor is different. Nevertheless, the prediction is in accordance to the observation of the "non-expert" tasting panel.

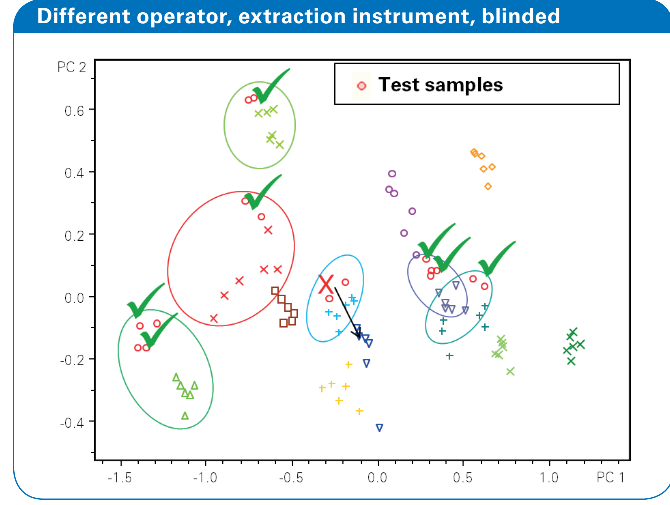

Figure 4: "Blinded" experiment: Eight samples were extracted by a different operator, on a different espresso machine. Classification of two technical replicates using the original PCA model enabled correct assignment of seven of eight samples in this blinded experiment (the sample ID was only revealed after the sample ID was predicted using the established PCA model).

Table 1: Predicted coffee intensities based on PLS model

Sample ID	Predicted Intensity	Stated Intensity
N1	9.74	10
N2	4.79	6
N3	8.84	9
N4	7.67	9
N5	4.12	6
N6	3.26	4
N7	5.12	6
N8	8.93	10

Table 2: Prediction of intensity for coffee from different vendor

Sample ID	Injection	Predicted Intensity
ZB1	1	12.35
ZB1	2	12.75
ZB2	1	13.00
ZB2	2	13.38

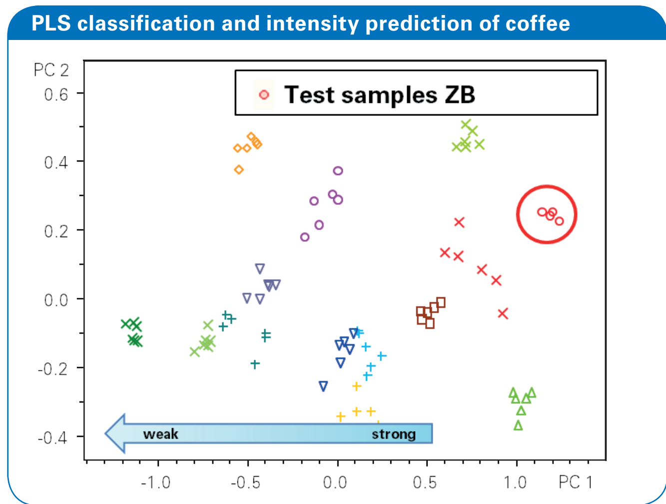

Figure 5: Classification of a coffee from a different vendor using the existing PLS model. The ZB coffee was assigned a medium degree of roasting; interestingly it clusters closely to the strong coffees of the original model.

Conclusion

In this proof-of-concept study, different coffee samples were classified using PCA and PLS models based on high-resolution LC-MS metabolic profiles. The established models enabled correct assignment of coffee types, even if they were extracted by different operators, on different espresso machines, and using different extraction volumes. Most coffees were also correctly assigned using this workflow in a blinded experiment. PLS models not only enabled sample classification, but also an intensity prediction for the analyzed coffees. A coffee extract from a different vendor could be classified and the intensity could be predicted using the existing PLS model. The strength assignment matched the intensity characterization of a (non-expert) tasting panel.

In summary, the presented workflow demonstrates that non-targeted, high-resolution LC-MS metabolic profiles combined with statistical data evaluation are perfectly suited to fulfilling the increasing public demand for testing food and beverages for authenticity and adulteration.

References

[1] Carcea ,M. et al. 2009: Food authenticity assessment: ensuring compliance with food legislation and traceability requirements. Quality Assurance and Safety of Crops & Foods 1(2) 93-100

[2] http://www.bruker.com/products/mr/nmr/food-screener/overview.html

[3] Application Note LCMS-79: Metabolic Profiling of Different Coffee Types on the Bruker compact™ QTOF System

[4] Eriksson, L. et al. 2006: Multi- and Megavariate Data Analysis: Part I: Basic Principles and Applications, Umetrics Academy, Umea

* A collaboration between Bruker BioSpin GmbH and SGF International e.V. (http://www.bruker.com/products/mr/nmr/food-screener/juicescreener/overview.html)

For research use only. Not for use in diagnostic procedures.

Bruker Daltonics is continually improving its products and reserves the right to change specifications without notice. © Bruker Daltonics 08-2013, LCMS-86, 1822451

Bruker Daltonik GmbH

Bremen · Germany
Phone +49 (0)421-2205-0
Fax +49 (0)421-2205-103
sales@bdal.de

www.bruker.com

Bruker Daltonics Inc.

Billerica, MA · USA
Phone +1 (978) 663-3660
Fax +1 (978) 667-5993
ms-sales@bdal.com

Fremont, CA · USA
Phone +1 (510) 683-4300
Fax +1 (510) 687-1217
cam-sales@bruker.com