Predictive Classification of Contaminants Encountered During the Distillation of Shochu, a Distilled Beverage Native to Japan.

Toshiyuki Yamashita[1]; Takeshi Serino[2]; David Peterson[2]; Anthony Gray[2] 1 Kinryo Electric Co. Ltd., Osaka, Japan; 2 Agilent Technologies, Santa Clara, CA

ASMS 2011
ThP 316

Agilent Technologies

Introduction

Shochu is a distilled alcoholic beverage native to Japan. It contains 25% alcohol by volume and is produced by distillation of rice, barley, potatoes or brown sugar. Mold, instead of yeast, is used to promote the fermentation processes. Shochu is typically consumed neat or mixed with water, tea, fruit juice or beer. In Japan, Shochu consumption exceeds that of Sake. Because of the social and economic importance of Shochu, robust and consistent quality control is critical to maintain shipments and customer satisfaction.

In recent years, the use of multivariate analysis to determine product quality has increased but today is mostly used for research and development (R&D) purposes. The multivariate analysis can produce a sample class prediction (SCP) model based on the statistical calculations from known samples. The SCP model can then be used for Quality Assurance (QA) and Quality Control (QC) in the manufacturing process, in addition to R&D. Utilizing the SCP model for QC will contribute to prevent any shipment of contaminated products. The SCP model can also be used for quick screening to predict the contaminants for quality assurance.

Experimental

Sample Information

In this study, we prepared several contaminated Shochu samples, as detailed below;

Table 1 Sample information

Sample name	Sample
IOSK	Shochu (Osaka), No contaminants
ITKO	Shochu (Tokyo), No contaminants
IUSA	Shochu (San Jose), No contaminants
DA	Shochu (Osaka) and Chlorine detergent
DB	Shochu (Osaka) and Sterile detergent
DC	Shochu (Osaka) and Anti-insect agent
GA	Shochu (Osaka) and Rubber Glove A
GB	Shochu (Osaka) and Rubber Glove B
OB	Shochu (Osaka) and Machine oil B

Shochu commercially available in Osaka, Tokyo (Japan) and San Jose (United States of America)
3 types of detergents commercially available in Osaka
1 type of machine oil commercially available in Osaka
2 types of rubber glove commercially available in Osaka

Experimental

Sample preparation

[Non-contaminated sample: IOSK, ITKO and IUSA]
1.0g of Shochu is inserted into the 20mL vial and sealed by the screw cap.

[Detergent contaminated sample: DA and DC]
1.0g of Shochu and 10~15mg of detergents are inserted into the 20mL vial and sealed by the screw cap.

[Detergent contaminated sample: DB]
1.0g of Shochu inserted into the 20mL vial spiked the detergent by a spray, then sealed by the screw cap.

[Rubber glove contaminated sample : GA and GB]
1.0g of Shochu and 20~23mg of a piece of rubber gloves are inserted into the 20mL vial and sealed by the screw cap.

[Machine oil contaminated sample : OB]
1.0g of Shochu and 10~15mg of machine oil are inserted into the 20mL vial and sealed by the screw cap.

Sample Introduction (SPME)
Incubation Temp : 40°C
Incubation Time : 40min
Agitation : None
Extraction Time : 20min
Desorption Time : 1.5min
SPME fiber : 100um PDMS (Supelco 57341-U), 1cm

GC/MS analysis

Instrument: GERSTEL MPS2 Multiple Purpose Sampler Agilent 7890/5975C TAD GC/MSD system, Agilent MSD Productivity ChemStation (E.02.02), Agilent Mass Profiler Professional software (B.02.02), Agilent Sample Class Predictor (B.02.00).

GC/MS Acquisition Parameters:
Injection: Splitless, 1.52 min @ 240°C
Column: J&W 121-7012LTMDB-Wax 10m x 0.18mm x 0.18um
Oven program: 35°C (2min)-30°C/min-240°C (3min)
GC Run time: 11.83 min
Column flow: 1.1 ml/min (Constant Flow)
MS Transfer line temp: 240°C
Ionization: EI, 70eV
Ion source Temp: 230°C
Acquisition mode: Scan (35-450u)
Trace Ion Detection : On
Tuning: atune.u

Peak detection and data process

Component extraction from the GC/MS data was done using NIST AMDIS (Automated Mass Spectral Deconvolution and Identification Software), version 2.69.

Results and Discussion

Total ion Chromatogram of samples

Figure 1 shows total ion chromatogram of all Shochu samples. The TIC contains many compound peaks such as ethers. Due to the complexity of the chromatogram, it is difficult to identify the contaminating compounds in the TIC.

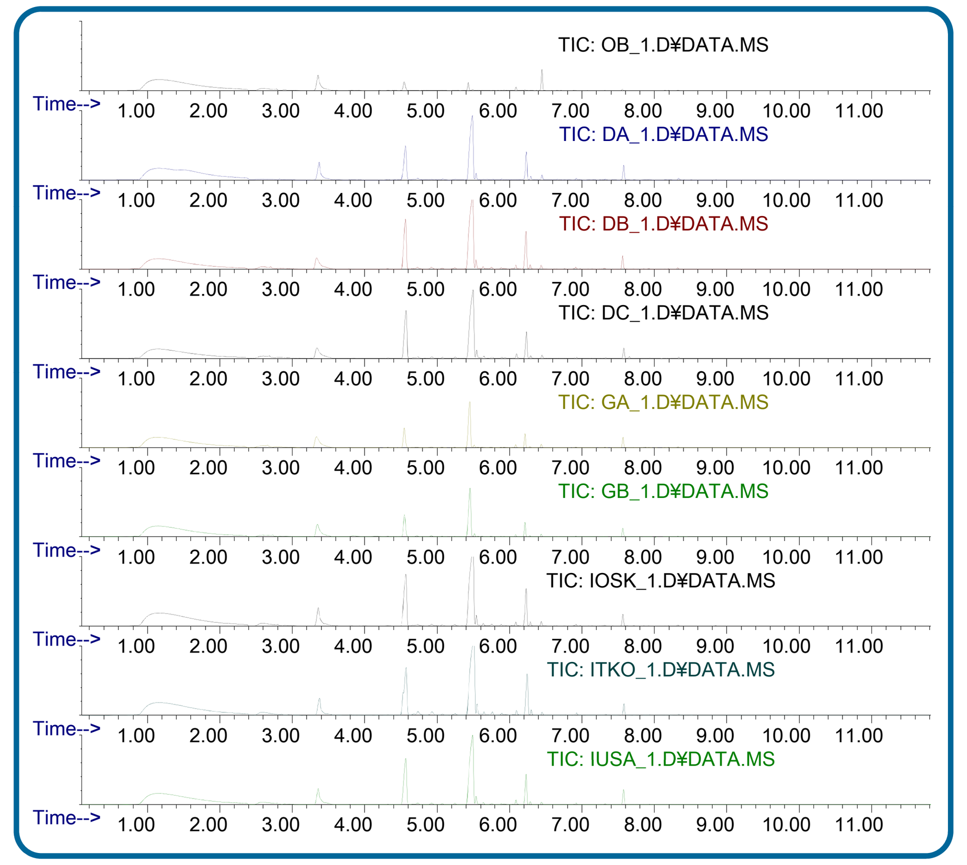

Figure 1 Total Ion Chromatograms of samples

Peak detection and filtering for multivariate analysis

AMDIS was used to extract components from the GC/MS data. The data consisted of 4 replicates each of Rubber, Detergent and Machine Oil contaminated sample, 23 replicates of Shochu sold in Osaka, 18 replicates of Shochu sold in Tokyo and 13 replicates of Shochu sold in San Jose, CA. AMDIS detected between 330 and 380 components per data file.

As the data set is imported into Mass Profiler Professional software (MPP) the unidentified components are aligned by spectral similarity and retention time window to form an "entity" list of 2,376 components. The MPP Frequency Table reveals that many of the entities are unique to one sample (i.e frequency = 1). Three entity filters were evaluated to eliminate these non-reproducible artifacts of the component extraction process and form a representative entity list of the sample set for creating the Class Prediction model.

Filter 1 *:Filter by Frequency, 100% at least one condition*
Entities which have 100% frequency in at least one sample are kept for the next step. 226 entities passed this filter.

Filter 2: *One way ANOVA*
Entities which have the statistic significance less than 5% in any of samples are kept. This filter is useful to filter out the entities that has NO statistic significance. 1,080 entities passed this filter.

Filter 3: *CV < 45% for all conditions after One way ANOVA*
This is the strongest filter in this experiment. Entities which passed one way ANOVA and CV <45% for all samples are kept. The objective of using these conditions is to intentionally create a strong filter and investigate the effect of the filter on the SCP model accuracy.

Principal Components Analysis (PCA)

Principal Components Analysis (PCA) was done for each entity list. Figure 2 to Figure 4 show the Score plot of each filtering conditions for Component 1 and Component 2.

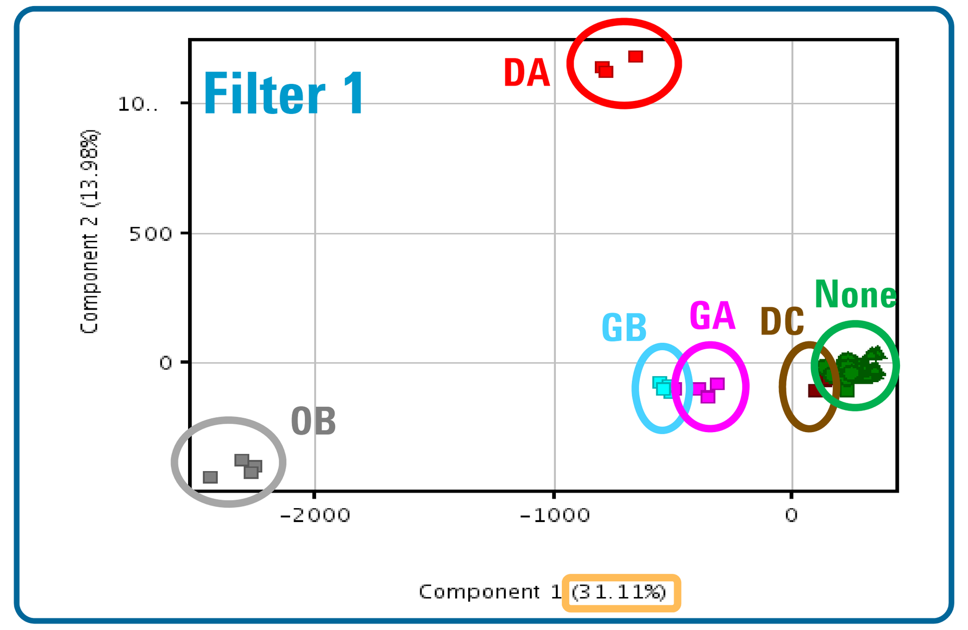

Figure 2 PCA Score Plot (Filter 1) Component 1 and 2

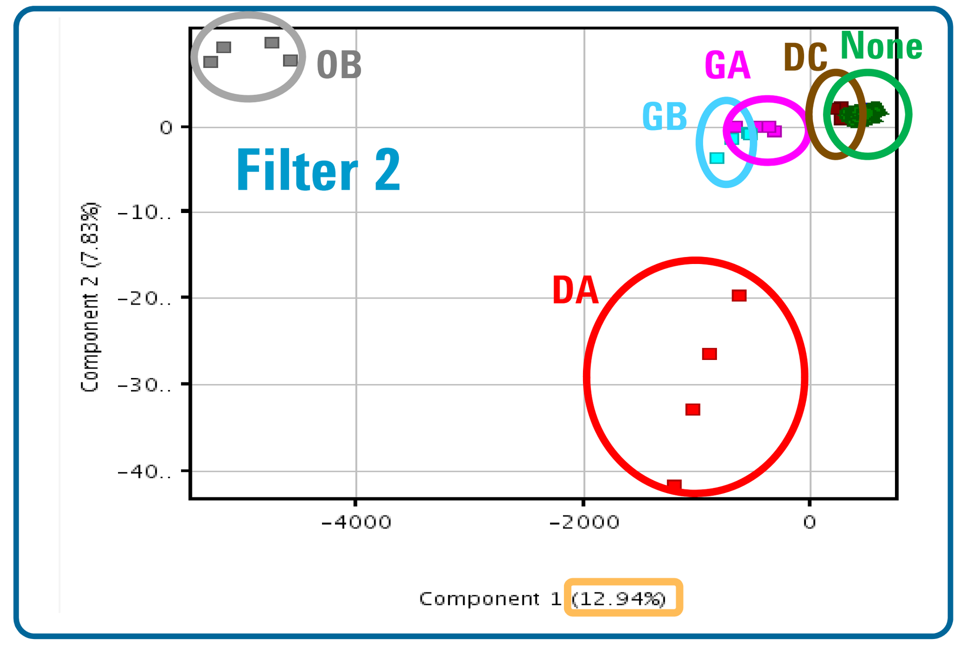

Figure 3 PCA Score Plot (Filter 2) Component 1 and 2

Results and Discussion

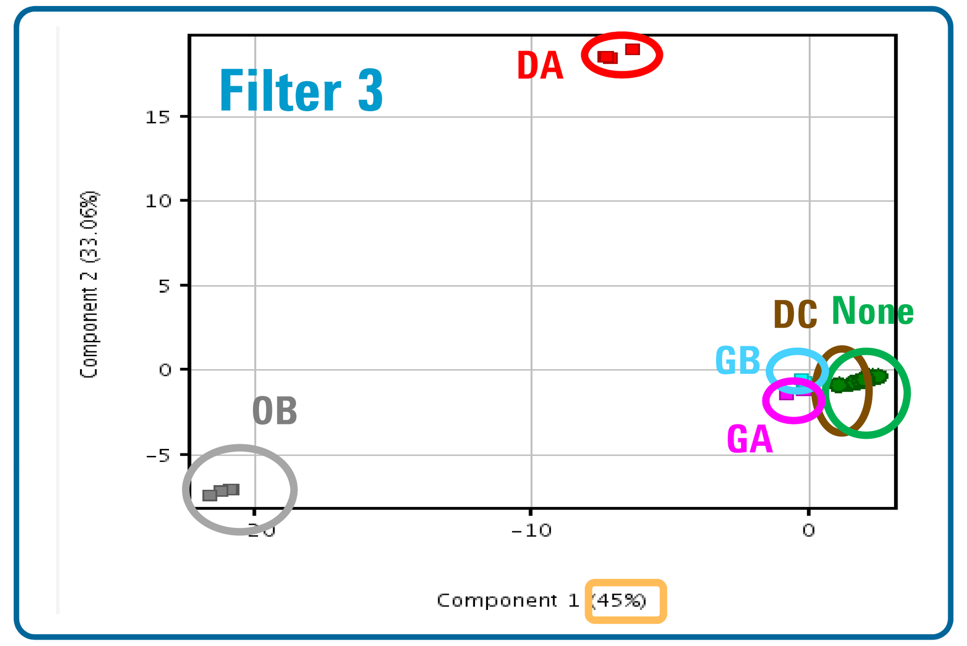

Figure 4 PCA Score Plot (Filter 3) Component 1 and 2

The PCA score plot with Filter 3 separates each sample type i.e no contaminants, detergents, rubber gloves and machine oil, better than other filters with 4 principal components.

Sample Class Prediction (SCP) Model

SCP models of 5 different algorithms (Decision Tree (DT), Naïve Bayes (NB), Neural Network (NN), Partial Least Square Discriminant (PLSD) and Support Vector Machine (SVM)) with the entities of 3 filters (Filter 1, Filter 2 and Filter 3) are created from all three entity lists described above.

During the SCP model creation, each sample in the set is evaluated against the model and an overall prediction accuracy is determined. An accuracy of 100% means that all samples were correctly predicted by the model.

Table 2 SCP model accuracy after training

Algorithm	Filter 1	Filter 2	Filter 3
DT	100%	100%	100%
NB	99%	100%	99%
NN	69%	73%	69%
PLSD	90%	90%	90%
SVM	100%	100%	100%

Applying Sample Class Prediction to unknown samples

In order to test the capability to detect for contaminants SCP was applied to an additional 12 samples that were not used to create the SCP model.
The additional, test samples consisted of Shochu contaminated with detergent (DA, DB, DC), rubber glove (GA, GB), machine oil (OB)), and 2 replicates of non-contaminated Shochu (IOSK, ITKO, IUSA).
The result of SCP model for unknown samples are shown in the Table 3.

Table 3 SCP model accuracy for unknown samples

Algorithm	Filter 1	Filter 2	Filter 3
DT	100%	100%	100%
NB	92%	100%	92%
NN	50%	50%	50%
PLSD	50%	42%	33%
SVM	83%	83%	100%

In table 2, you can see SVM and DT were able to accurately validate the models using all three filters. However, with these filters the DT model is more robust when predicting unknown contamination (Table3). This implies that there is a limited set of entities that heavily influence the prediction model. PLSD is better in classifying subtle differences in classes, however it is dependent on robust feature finding and entity filtering – i.e targeted analysis. In this experiment, the entities that influence the PLSD model may not have been present in the unknown sample entity lists, either in feature finding or with entity filtering.

Conclusions

SCP provides a robust way to determine Shochu quality that can be used in a production QC environment. By using multivariate analysis of GC/MSD data small differences between samples can be clearly visualized.

In order to get the SCP model with highest accuracy of prediction, the data quality is crucial. This allows you to create the right filtering and prediction model for your samples. We determined that SCP will provide the best results when the sample data is properly filtered and an appropriate prediction algorithm is used. Multiple prediction models allow you to evaluate and customize different prediction models to the analysis.

Better entity lists enable the development of better SCP prediction models. With better SCP model, improvement of the workflow of QA and QC of food analysis can be expected.