Nicotine & Tobacco Research, 2019, 1–7
doi:10.1093/ntr/ntz066
Original investigation
Received January 10, 2019; Editorial Decision April 22, 2019; Accepted April 26, 2019
Advance Access publication May 6, 2019

Original investigation

Identification of Cigarette Brands by Soft Independent Modeling of Class Analogy of Volatile Substances

Zuzana Zelinkova PhD, Thomas Wenzl PhD

European Commission, Joint Research Centre (JRC), Geel, Belgium

Corresponding Author: Thomas Wenzl, European Commission, Joint Research Centre (JRC), Retieseweg 111, B-2440 Geel, Belgium. Telephone: 32-14-571 320; E-mail: Thomas.Wenzl@ec.europa.eu

Abstract

Introduction: This study aimed to develop a method for discriminating cigarette brands based on the profiles of volatile components extracted from the tobacco fraction of the finished cigarettes to authenticate branded cigarettes of unknown origin.

Methods: An analytical method comprising direct thermal desorption coupled with gas chromatography-quadrupole time-of-flight mass spectrometry was developed for acquiring volatile profiles of cigarettes. About 290 samples of commercially available cigarettes were analyzed. Within this batch, 123 samples represented four popular cigarette brands. They were selected for in-depth characterization. Multivariate analysis was used to investigate the interrelations among volatile compounds of cigarettes and to identify characteristic markers for the cigarette discrimination. Supervised pattern recognition techniques were used for designing classification models.

Results: Principal component analysis covering all detected volatiles allowed the differentiation of cigarettes based on the brand. A number of 56 volatile components were identified as markers with high discrimination power. These compounds were used for establishing classification models. A method of soft independent modeling of class analogy developed for the four studied cigarette brands proved to be efficient in the classification of unknown cigarettes, with accuracy between 95.9% and 100%.

Conclusions: The data evaluation by soft independent modeling of class analogy was highly accurate in classification of unknown cigarettes with a low rate of false positives and false negatives. The developed models can be used for discrimination of genuine from non-genuine products with high level of probability.

Implications: Profiling of volatiles, which is commonly used for authentication of different food commodities, was applied for the characterization of cigarette tobacco for the purpose of authentication a cigarette brand. Volatile components with a high discrimination power were identified by means of multivariate statistical methods and used for establishing of a classification model. The classification model was able to discriminate genuine from non-genuine cigarettes with a high level of prediction accuracy. This model could be a powerful tool for tobacco control to judge the authenticity of cigarettes.

Introduction

Tobacco is a very complex matrix consisting of a large number of different substances across multiple chemical classes and wide concentration ranges.[1,2] The physical and chemical properties of tobacco leaves are influenced by genetics, growing conditions, weather conditions, plant diseases, harvesting, post-harvesting procedures (such as curing, aging, and fermentation), and manufacturing.[3] An essential step of the manufacturing is addition of additives, particularly

© The Author(s) 2019. Published by Oxford University Press on behalf of the Society for Research on Nicotine and Tobacco.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/ntr/advance-article-abstract/doi/10.1093/ntr/ntz066/5486026 by guest on 22 May 2019

flavorings.[1] The flavor specialists have the task of improving and modifying the tobacco aroma and taste to fit the expectation of the consumer. Just as the blends and types of tobaccos used are determining factors in the design of a product, the flavorings, which are added, greatly influence the quality and acceptability of the finished product.[4,5] It is obvious that manufacturers aim for a high level of recognition of their products by the consumer and put, therefore, a lot of attention to maintaining their quality from batch to batch. It can be assumed that the composition of volatile compounds of a finished tobacco product is specific for the individual product and thus profiling of volatiles can be a useful tool for discrimination of different tobacco products.

Various analytical methods have been reported for the determination and quantification of volatile components in tobacco. These methods generally comprise sample preparation methods such as supercritical fluid extraction, simultaneous distillation and extraction, steam distillation, dynamic and static headspace extraction, solid phase microextraction and thermal desorption (TD) usually followed by gas chromatography–mass spectrometry (GC-MS) analysis.[6–12] TD is widely used for the analysis of volatile and semi-volatile organic compounds by GC-MS in a variety of sample matrices and over wide concentration ranges.[13,14] It offers the advantages of simplicity, speed, economy, and sensitivity of analysis of solid and liquid samples, compared to other extraction techniques TD achieves higher extraction yields and less discrimination between different classes of volatile compounds.[15]

Several studies applied profiling of volatiles for the differentiation of tobacco leaves according to their origin.[16,17,18] However, the application of profiling of volatiles for discriminating cigarette brands has not been published yet. The goal of this work was to develop a robust, simple, and automatic analytical method for profiling of tobacco volatiles and to establish a classification model for selected cigarette brands to judge with a high level of certainty whether or not cigarettes of these brands are genuine products. The work was divided into different phases. The first phase focused on the identification of volatile constituents with high discrimination power by applying multivariate statistical methods. The second phase concerned the development of a classification model, which accepts cigarette samples from the target class, while the probability of accepting nontarget samples is minimized. In the final phase, the classification model was validated by the analysis of independent samples belonging to both the target class and samples of other cigarette brands. The latter served as proxies of counterfeit products.

Methods

Samples

Samples of the four target cigarette brands (A, B, C, and D), manufactured by three producers (P01, P02, and P03), were collected at licensed tobacconists in Europe. Each package was sampled in a different location to avoid duplicity of samples and to assure that a representative sample set was obtained. Thirty-three cigarette packages of brand A manufactured by producer P01 were collected in 17 countries (Austria, Belgium, Croatia, Czech Republic, Estonia, Finland, France, Germany, Greece, Italy, Latvia, Portugal, Serbia, Slovakia, Slovenia, Spain, and Sweden). Twenty-five cigarette packages of brand B manufactured by producer P02 were purchased in 12 countries (Austria, Belgium, Czech Republic, Finland, Germany, Italy, the Netherlands, Portugal, Serbia, Slovakia, Spain, and Switzerland). Forty-three cigarette packages of brand C manufactured by producer P03 were collected in 16 countries (Austria, Belgium, Czech Republic, Estonia, France, Germany, Hungary, Italy, Latvia, Lithuania, Luxemburg, Portugal, Serbia, Slovakia, Spain, and Switzerland), and 22 cigarette packages of brand D manufactured by producer P02 were obtained from 10 countries (Austria, Belgium, Czech Republic, Germany, Latvia, the Netherlands, Spain, Sweden, Switzerland, and United Kingdom). In addition, 167 packs of cigarettes, comprising 114 other cigarette brands were collected in Europe for the purpose of validating the discrimination power of the developed model. The whole sample set consisted of 290 packs of cigarettes obtained in 2016 and 2017 from 28 European countries. Details on sample distribution based on country of purchase can be found in Supplementary data 1.

The research cigarette 3R4F, obtained from Tobacco-Health Research, University of Kentucky (Lexington, KY), was used as a quality control sample.

Reagents and Material

Isotopically labeled 2-ethylphenol-D_{10} was purchased from CDN Isotopes (Quebec, Canada). A spiking solution of the isotopically labeled standard of about 30 µg/mL was prepared gravimetrically in methanol. Methanol of LC-MS grade was obtained from VWR (Leuven, Belgium).

Equipment and Instrumentation

An automatic analytical syringe eVol (SGE Analytical Science Pty. Ltd., Ringwood, Austria) was used for spiking samples with the isotopically labeled standard.

The used direct TD gas chromatography-quadrupole time-of-flight mass spectrometry (GC/Q-TOF-MS) system consisted of a GC 7890A (Agilent Technologies, Santa Clara, CA) equipped with a cooled injection system, a programmable temperature vaporizing inlet (Gerstel, Mülheim an der Ruhr, Germany) and a TD unit (TDU, Gerstel) operating automatically in conjunction with a MultiPurpose Sampler (MPS, Gerstel). The GC was coupled to a 7200 Accurate Mass Q-TOF MS system (Agilent Technologies). Operation of the instrument was controlled by MassHunter GC/MS Acquisition software, version B.07.03.2129 (Agilent Technologies) and Maestro 1 software, version 1.4.31.10/3.5 (Gerstel).

Analytical Method

For the preparation of test samples, three randomly selected cigarettes were removed from each cigarette package. The tobacco contained in the cigarette sticks was separated from the cigarette paper and filter and ground and homogenized in a mortar under cooling by liquid nitrogen to prevent loss of volatile compounds. A volume of 5 µL of isotopically labeled standard solution was pipetted into a glass micro-vial insert and a portion of 30 mg tobacco sample was directly weighed over it. The micro-vial was inserted into an empty glass TD tube for analysis.

Splitless TD was realized by ramping the TDU from 20°C held for 0.1 minute to 100°C at 30°C/min held for 15 minutes with a helium purge flow of 100 mL/min. Volatile compounds were trapped in the programmable temperature vaporizing inlet on a Tenax TA packed liner at 15°C. The trapped compounds were transferred onto the HP-5MS GC column (30 m × 250 µm × 0.25 µm; Agilent Technologies) in split mode, with a split ratio of 15:1, while programming the programmable temperature vaporizing inlet from 15°C held for 0.8 minutes to 270°C at 12°C/s held for 30 minutes.

Downloaded from https://academic.oup.com/ntr/advance-article-abstract/doi/10.1093/ntr/nty066/5486026 by guest on 22 May 2019

The GC oven was programmed from 45°C (held for 2 minutes) to 210°C at 4°C/min further to 300°C at 10°C/min (held for 5 minutes). Helium was used as a carrier gas at 1.0 mL/min constant flow rate. The transfer line temperature was set at 300°C. The Q-TOF-MS was operated in electron ionization mode at 70 eV ionization energy, except for the time window from 23.25 to 23.95 minutes, when nicotine was eluted. During this period, the ionization energy was reduced to 25 eV. The data acquisition rate was 5 Hz in 2 GHz extended dynamic range mode for the mass range of *m/z* 45–450.

A quality control sample was included in each sample batch consisting of maximum 15 samples together with two blank samples.

Data Analysis

Chromatograms were first subjected to deconvolution, an automatic peak detection and component identification using Mass Hunter Unknown Analysis (Agilent Technologies). On average, 461 components were automatically detected by the software in each sample. Retention time shifts, which were for the whole period of the study below 0.1 minute, were compensated by mass spectra-based peak alignment. Hence, data of each sample were imported to Mass Profiler Professional (Agilent Technologies) for data alignment and data filtering. Data filtering was carried out via frequency-of-occurrence analysis (components found in only one or few samples might be caused by wrongly or inconsistently detected compounds), sample variability, and analysis of variance. Further reduction of the data was performed by plotting a cumulative sum of squares of all explained variables (R^2VX) and cumulative sum of squares of model prediction errors (Q^2VX) obtained by principal component analysis (PCA). Those components, which were poorly explained or predicted by the model, were removed.

Additional data analysis was carried out by using XCMS Online (Scripps Research Institute), a cloud-based informatics platform designed to process and visualize mass-spectrometry data. In this respect, data processing consisted of deconvolution, peak detection, alignment, and detection of features. A feature represents a peak in the chromatogram and is defined as a molecular entity with a unique mass and retention time. The XCMS Online platform was used to perform multivariate data processing, such as two-group pair comparisons to identify significant differences between two groups of samples, and meta-analysis to identify shared and different features between four groups of samples.[19,20]

Classification Method

Soft independent modeling of class analogy (SIMCA) is a classification method based on disjoint PCA modeling.[21] PCA is a widely used data analysis technique that allows reducing the dimensionality of the system while preserving information on the variable interactions. PCA transforms the original variables into a set of linear combinations, the principal components (PCs), which capture the data variability, are linearly independent and weighted in decreasing order of variance coverage.[22] In SIMCA, each of the classes (cigarette brands) is modeled separately by PCA. The concept of SIMCA is to build a confidence limit for each cigarette brand with the help of PCA and then to project an unclassified or unknown sample into each PCs space. On the basis of the residual variation of each class, the distance to the model of each observation can be computed. The final classification of an unknown sample is obtained comparing its residual variances to the residual variance within each class through an *F*-test. The critical limit for the distance to the model is based on the F-distribution using a 95% confidence interval. More details on the SIMCA modeling can be found in the literature.[21,23,24]

Before PCA, data were preprocessed by means of mean-centering, scaling to unit variance and logarithmic transformation.

Software

Mass Hunter Qualitative Analysis, version B.07.00 (Agilent Technologies); Mass Hunter Unknowns Analysis, version B.07.01 (Agilent Technologies); Mass Hunter Quantitative Analysis for QTOF, version B.07.01 (Agilent Technologies); Mass Profiler Professional, version 12.5 (Agilent Technologies); and XCMS, version v3.5.1 (Scripps Research Institute, La Jolla, CA) were used for data analysis. Multivariate statistics were carried out using SIMCA, version 14.1.0.2017 (MKS Umetrics, Malmo, Sweden). The mass spectra of measured peaks were compared with mass spectra in NIST spectral library, version 2.0 2011 (Gaithersburg, MD).

Results

Performance of the Analytical Method

The performance of the analytical method for the analysis of volatiles extracted from cigarette tobacco was validated. Method performance parameters, such as repeatability and intermediate precision, were evaluated from repeated analysis of quality control samples for 45 randomly selected compounds eluting between retention time of 5.1–44.1 minutes. Repeatability expressed as the relative standard deviation of absolute responses of nine consecutive measurements varied between 5.8% and 18.2%. Intermediate precision was calculated based on 18 measurements acquired within a 2-month period and expressed as the relative standard deviation of absolute responses. Intermediate precision ranged between 5.8% and 21.2%. Details can be found in Supplementary data 2.

Several quality control tools were used to monitor the variability of measurement. The absolute responses of the isotopically labeled standard added to each test sample were plotted in quality control charts to detect instrument problems and identify potential trends. In addition, quality control charts were kept for five selected compounds (2(5H)-furanone, benzyl alcohol, solanone, megastigmatrienone, and neophytadiene) measured in the quality control sample analyzed together with each batch of samples. Thresholds for quality control were set based on the relative standard deviation obtained for the particular compound during initial precision studies. The warning limit was equal to twice the standard deviation, whereas the action limit corresponded to three times the standard deviation. Examples of quality control charts can be found in Supplementary data 3.

All collected samples were analyzed in random order over a period of 8 months. Example chromatograms of the volatile fraction extracted from samples of the four cigarette brands by application of the described method are shown in Supplementary data 3.

Discrimination of Cigarettes Based on the Profile of Volatiles

PCA was performed to investigate any possible clustering of samples according to cigarette brand. The PCA model constructed from the entire preliminary data set was very poor, indicating a high amount of data with little discrimination power. However, the PCA score plot revealed a grouping of samples in clusters based on the cigarette brand and demonstrated that there are significant differences in the profiles of volatiles between different cigarettes. Data filtering was applied to reduce the data noise, to remove all non-relevant

Downloaded from https://academic.oup.com/ntr/advance-article-abstract/doi/10.1093/ntr/ntz066/5486026 by guest on 22 May 2019

information from the data matrix, and to keep only data with high discrimination power. The data were reduced to 75 variables. The PCA model consisted of nine PCs and captured 80.4% of total variance; the first and second components captured 41.1% of variance (Supplementary data 5). The total variation predicted by the model (Q^2 cumulative) was 60.4%. Characterizing compounds were identified for each cigarette brand from PCA score plots and loading plots.

In addition, two-group pair comparisons were carried out to investigate the difference between volatile profiles of two cigarette brands. This method identified features, whose relative intensities are significantly different between two brands. The differences in volatile profiles are visualized by cloud plots. Supplementary data 6 presents as an example the cloud plot for cigarette samples of brand A and brand B. Characterizing compounds were then identified by comparing the cloud plots with total ion chromatograms. The proper selection of discriminating components was corroborated by the additional second-order analysis called "meta" analysis. This method is useful for multiple group comparison. The outcome, illustrated in a Venn diagram (Supplementary data 7), enabled the selection of features that are unique for each of the cigarette brands (represented by the not overlapped areas in the Venn diagram) and to identify features common for two or more cigarette brands (represented by the overlapping areas in the Venn diagram).

A list of markers (characterizing compounds) was set up for each target cigarette brand by combining results from PCA, two-group pair analysis and meta-analysis. Cigarettes A were characterized by 20 compounds, cigarettes B by 17 compounds, cigarettes C by 14 compounds, and cigarettes D by 16 compounds. The majority of these compounds were tentatively identified by comparing mass spectra with the NIST library and based on the linear retention indexes of compounds (Table 1). The ratios of absolute responses of these compounds to the absolute response of the isotopically labeled standard were further used for developing classification models.

Classification of Cigarettes According to Brand

The sample set consisting of cigarettes A, B, C, D was split into two groups—a set of training and a set of prediction samples. The training sample set was used for model generation and consisted of 15–20 randomly selected samples of each brand. Four PCA disjoint models were built (Figure 1), each PCA represents one brand. In detail, three PCs modeled cigarettes of brand A and captured a total variance of 83.6%. Cigarettes brand B was characterized by the PCA model consisting of two PCs of a total variance of 72.1%. PCA models of three PCs characterized cigarette brand C (77.9% of total variance) and cigarette brand D (74.7% of total variance). One of the limitations of the experimental setup was the lack of confirmed genuine samples, which bears the risk of weakening the models with data of potentially counterfeit products. Hotelling's T^2 range plots (calculated for the range of components) were used to evaluate the homogeneity of sample sets and to detect potential outliers. Hotelling's T^2 plot did not reveal any extreme outliers in the training sample set, which could unnecessarily alter the classification results, and samples were found to be rather consistent (Supplementary data 8).

The SIMCA modeling scheme was applied to the prediction sample set to evaluate the accuracy of model predictions. The prediction sample set included independent (not included in the setup of the models) cigarette samples of the four targeted cigarette brands plus cigarettes of different other cigarette brands collected in Europe. The prediction sample set consisted in total of 220 samples (13 brand A, 10 brand B, 23 brand C, 7 brand D, and 167 other brands). Samples with a probability of membership less than 5% (less than 0.05) are considered to be outliers (outside 95% confidence level) and not belonging to the model (class). The performance of the model is expressed by the parameters sensitivity and specificity, which characterize the overall model accuracy. Sensitivity is defined as the percentage of samples correctly assigned by the model to the respective cigarette brand (true positive rate). Specificity is the percentage of samples correctly classified as not belonging to the respective cigarette brand (true negative rate). Accuracy represents the rate of correctly classified samples (sum of both true positive and true negative samples) among the performed classification tests. Sensitivity and specificity values of 100% were obtained for brand A. All cigarettes of brand A were correctly classified by the SIMCA model and were found similar to the rest of cigarettes A from the training sample set. All samples of other brands were determined as outliers and rejected by the model. Nine of 10 samples of brand B were correctly identified by SIMCA, whereas 7 of the 210 non-brand B samples were wrongly classified. Sensitivity and specificity for cigarette B were determined to 90% and 96.7%, respectively and an overall accuracy of 96.4%. Slightly lower sensitivity of 78.3% was observed for cigarettes C, where five of twenty-three samples were false negatives. However, specificity and total accuracy were 98.1% and 95.9%, respectively. Similarly, lower sensitivity of 85.7% was observed for cigarettes D. In this case, one of seven samples was misclassified. Specificity and total accuracy reached 97.7% and 97.3%, respectively.

Discussion

Tobacco is an agricultural product and as such subjected to geographical and seasonal influences on its composition. Manufacturing of fine cut tobacco adds additional variability to the product. Cigarette manufacturers aim to fabricate products with constant quality in terms of physical and sensorial properties. This is achieved by blending of tobacco and by addition of additives, which might compensate for deficiencies of the raw tobacco. The types and amounts of additives added to fine cut tobacco for maintaining a product-specific taste were considered rather stable. To account for the variability that can be expected in a finished branded product, sampling was performed in a way that assured the coverage of different production batches.

A limitation of the sampling of branded cigarettes from retail is the lack of certainty on the authenticity of the sampled products. However, the risk of unintendedly generating models based on counterfeit products was reduced by putting sampling on a broad geographical and temporal footage via sampling at licensed tobacconists distributed over a number of different European countries over a period spanning almost 2 years, which makes the inclusion of a large number of counterfeit products unlikely. In addition, statistical tests were performed for identifying outliers among the samples used for model generation.

The second limitation concerned the absence of possibilities to acquire counterfeits of branded cigarettes by lawful means. In assuming that counterfeiters will hardly be able to generate the same tobacco and additive blends as legal producers use for their branded products, cigarettes of other cigarette brands than the four target brands, comprising a multitude of tobacco and additive combinations, were used as a replacement for counterfeited target cigarettes.

The main goal of the developed analytical method was to obtain a comprehensive profile of volatile components of cigarette tobacco

Downloaded from https://academic.oup.com/ntr/advance-article-abstract/doi/10.1093/ntr/nty066/5486026 by guest on 22 May 2019

Table 1. Details of Characterizing Volatile Compounds (Markers) Used for SIMCA Classification Modeling for Cigarette Brand A, B, C, D

No.	Compound name	Cigarette brand	t_R [min]	CAS no.	LRI	LRI (lit)	*m/z*
1	Pyrazine, 2-methyl*	B; D	6.07	109-08-0	825	825	**94**; 64
2	Butanoic acid, 3-methyl-*	C; D	6.55	503-74-2	843	840	**60**; 87; 101
3	3-Furanmethanol	C	6.87	4412-91-3	855	835	**98**; 97; 81
4	Propionic acid, 3-methoxy	B	7.98	2544-06-1	897	851	**74**; 58; 45
5	2(5H)-Furanone	B	8.55	497-23-4	915	918	**55**; 84; 54
6	Pentanoic acid	C	8.83	109-52-4	924	924	**60**; 73
7	2-Furanmethanol, 5-methyl	B	9.88	3857-25-8	956	953	**112**; 111; 95
8	Benzaldehyde*	B; D	10.00	100-52-7	960	960	**105**; 77; 51
9	α-Methylstyrene	A	10.73	98-83-9	982	980	**118**; 117; 103
10	1-Hexanol, 2-ethyl-	B; D	12.37	104-76-7	1029	1020	**57**; 70; 83
11	2-Cyclopenten-1-one, 2-hydroxy-3-methyl-*	A	12.50	80-71-7	1033	1034	**112**; 69; 83
12	Ethanone, 1-(1H-pyrrol-2-yl)-*	D	13.57	1072-83-9	1063	1063	**94**; 109; 66
13	Acetophenone*	C	13.67	98-86-2	1066	1066	**105**; 77; 51
14	2-Pyrrolidinone	A	13.83	616-45-5	1070	1069	**85**; 84; 86
15	Pyrazine, tetramethyl-*	A	14.41	1124-11-4	1087	1087	**136**; 54; 137
16	Pyridine, 4-(1,1-dimethylethyl)-	B	14.44	3978-81-2	1088	1073	**120**; 135; 92
17	Phenol, 2-methoxy*	A	14.52	90-05-1	1090	1089	**124**; 81; 85
18	1,6-Octadien-3-ol, 3,7-dimethyl-*	B	14.89	78-70-6	1100	1100	**93**; 71; 121
19	Ehtanone, 1-(3-pyridinyl)-	B	15.22	350-03-8	1110	1105	**106**; 78; 121
20	Maltol*	A	15.43	118-71-8	1116	1114	**123**; 71; 55
21	Phenylethyl alcohol*	A	15.51	60-12-8	1118	1118	**91**; 92; 65
22	3-Pyridinemethanol	B	15.94	100-55-0	1130	1122	**109**; 108; 80
23	4H-Pyran-4-one, 2,3-dihydro-3,5-dihyroxy-6-methyl	A	16.74	28564-83-2	1152	1151	**144**; 101; 73
24	l-Menthone*	B	16.82	89-80-5	1154	1155	**139**; 154; 112
25	2(1H)-Pyridinone, 5,6-dihydro-	A	17.10	6052-73-9	1162	1160	**68**; 97; 69
26	D-Menthone*	B	17.19	1196-31-2	1165	1164	**139**; 112; 154
27	Acetic acid, phenylmethyl ester	A	17.20	140-11-4	1165	1160	**108**; 91; 79
28	Menthol*	B	17.50	89-78-1	1174	1172	**138**; 123; 109
29	1,3-Cyclohexadiene-1-carboxaldehyde, 2,6,6-trimethyl-	A	18.45	116-26-7	1200	1202	**107**; 121; 150
30	Benzeneacetic acid, ethyl ester*	A	20.02	101-97-3	1246	1247	**91**; 92; 164
31	Benzaldehyde, 4-methoxy*	D	20.33	123-11-5	1255	1252	**135**; 136; 107
32	Anethole*	D	21.40	104-46-1	1286	1286	**147**; 148; 117
33	Cyclohexanol, 5-methyl-2-(1-methylethyl)-, acetate*	B	21.70	16409-45-3	1295	1294	**95**; 123; 81
34	4-Acetylanisole*	A	22.60	100-06-1	1322	1325	**150**; 135; 77
35	Piperonal*	D	23.07	120-57-0	1337	1333	**149**; 150; 121
36	Triacetin*	C	23.68	102-76-1	1356	1350	**145**; 116; 115
37	Vanillin*	A	25.10	121-33-5	1399	1400	**151**; 152; 123
38	trans-Geranylacetone*	D	26.78	3796-70-1	1454	1453	**107**; 151; 136
39	2,6-Di-tert-butylbenzoquinone	C; D	27.37	719-22-2	1473	1472	**205**; 220; 165
40	1-Dodecanol	D	27.41	112-53-8	1474	1473	**55**; 69; 83
41	β-Ionone*	C	27.83	14901-07-6	1488	1488	**117**; 123; 178
42	2,6-Di-tert-butyl-4-methylphenol	D	28.63	128-37-0	1515	1515	**220**; 205; 177
43	2(4H)-Benzofuranone, 5,6,7,7a-tetrahydro-4,4,7a-trimethyl-	A	29.13	17092-92-1	1532	1538	**111**; 137; 109
44	Megastigmatrienone I	A; C	30.56	38818-55-2	1581	1588	**175**; 148; 190
45	Megastigmatrienone II	A; C; D	31.89	38818-55-2	1627	1623	**175**; 148; 190
46	6,10-Dodecadien-1-ol, 3,7,11-trimethyl-, (E)-(±)-	B	32.30	20576-54-9	1642	1654	**123**; 95; 69
47	Allyl α-ionone	C	33.02	79-78-7	1668	1664	**218**; 177; 175
48	Unknown (isoprenoid)	D	33.43		1683		**197**; 212; 155
49	1H-Indene, 2,3-dihydro-1,1,3-trimethyl-3-phenyl-	B; D	34.43	3910-35-8	1719	1716	**221**; 143; 128
50	Furan, 2-[(2-ethoxy-3,4-dimethyl-2-cyclohexen-1-ylidene)methyl]-	C	34.45	55162-49-7	1720	1723	**232**; 175; 121
51	Allyl ionone	C	34.94	79-78-7	1739	1734	**232**; 217; 135
52	Unknown (isoprenoid)	D	35.30		1752		**173**; 188; 201
53	Benzyl benzoate*	B	35.68	120-51-4	1766	1765	**105**; 194; 91
54	2-Pentadecanone, 6,10,14-trimethyl-	C	37.74	502-69-2	1846	1843	**58**; 59; 71
55	Benzeneacetic acid, 2-phenylethyl ester*	A	39.51	102-20-5	1917	1919	**104**; 91; 105
56	3-(4,8,12-Trimethyltridecyl) furan	A	40.77	54869-11-3	1969	1971	**82**; 81; 95

LRI = calculated linear retention index; LRI (lit) = linear retention index taken from databases[25–28]; *m/z* = *m/z* value selected from the compounds mass spectra used as quantifier (bold print) and qualifier ions; t_R = retention time.
*Compound used as a cigarette additive.

to investigate the possibility for discrimination and classification of selected cigarette brands (A, B, C, D) by the means of supervised pattern modeling techniques. Direct TD hyphenated to GC provides a fully automated solvent-less sample preparation technique consisting of desorption/extraction, pre-concentration, and GC injection. This method was found to be fit for the purpose providing a broad profile

Downloaded from https://academic.oup.com/ntr/advance-article-abstract/doi/10.1093/ntr/ntz066/5486026 by guest on 22 May 2019

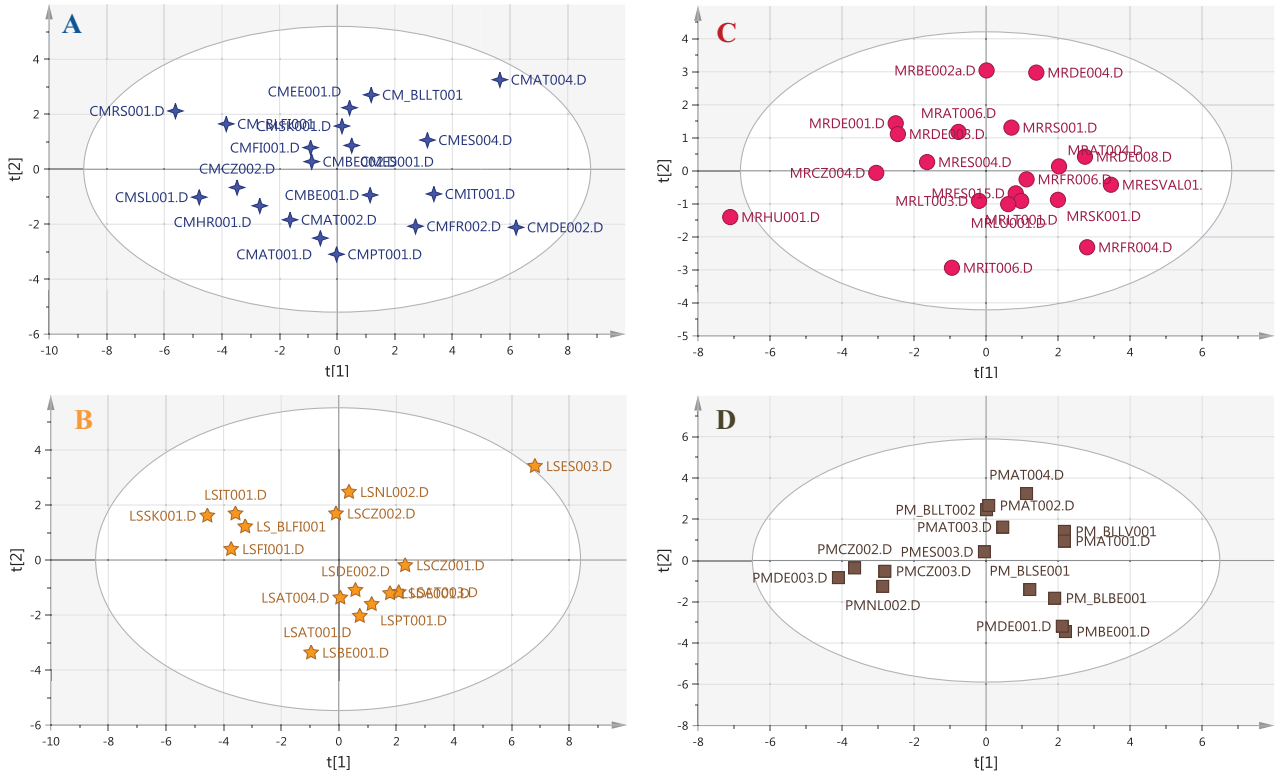

Figure 1. Disjoint principal component analysis (PCA) score plots of soft independent modeling of class analogy for cigarettes A (PC1 51.5%, PC2 18%), cigarettes B (PC1 50.3%, PC2 21.7%), cigarettes C (PC1 47.7%, PC2 18.2%), cigarettes D (PC1 32.1%, PC2 26.5%).

of volatile compounds contained in unburned tobacco. The stability of volatile compounds was evaluated when optimizing TD. A lack of stability was not observed under the selected conditions. However, it has to be stated that the assay did on purpose not focus on the most volatile compounds, as their content is likely altered during transport, storage, and manipulation of the cigarette sticks before instrumental analysis. The fine grinding of the cigarette tobacco provided not only a homogenous sample but also a high surface to volume ratio, which facilitated equilibration with laboratory climatic conditions. Conditioning of the sample in a climatic chamber was therefore not required. The optimized method was characterized by good repeatability and intermediate precision. It was also found that the measurement process does not add significant variability to the variability of the profile of volatiles of cigarette samples.

Several software platforms were applied for data treatment and statistical analysis. Multivariate analysis was used to investigate the interrelations among a set of variables to identify characterizing components for discrimination of cigarette brands and to depict important flavor-related compounds. PCA using preliminary selected volatile compounds clearly differentiated groups of cigarettes according to the cigarette brand. The first PC discriminated cigarette brand A from C (explaining 26.2% variation), whereas the second PC (14.9% variation) separated brand B and D from A and C (Supplementary data 4A). Discrimination of brand B and C was partly achieved by the third PC (Supplementary data 4B). Cigarettes brands B and D were made by the same producer and their volatile profiles were not as distinct as for other brands. However, a pair-analysis performed by XCMS Online revealed flavor components, which were unique for each brand (Supplementary data 9).

Twenty-six compounds of fifty-six selected markers were found in the cigarette ingredients list as a flavoring agent.[29] Compounds used as a cigarette additive are marked in Table 1. Flavorings are frequently used to provide a specific sensorial characteristic to a product and to improve the recognition of this product by consumers. Therefore, it could be expected that some of the flavoring additives would be among the markers for discrimination of cigarettes.

The developed SIMCA model was characterized by a high level of accuracy of class prediction of unknown cigarettes. Misclassifications occurred rarely, despite tests were performed on a large number of different cigarette samples obtained from the European market. However, an extension of model sensitivity assessment is planned for the future work. The established SIMCA model was considered efficient in identifying whether a tested cigarette is branded as A, B, C, or D, or neither of those brands. Consequently, the application of the developed model to an unknown, non-genuine cigarette sample, would have led with high probability to the conclusion that this sample is not authentic. In this way, the developed SIMCA model could be a powerful tool for tobacco control to judge the authenticity of cigarettes. If needed, the classification can be extended to other cigarette brands in a similar way as demonstrated in this article.

Supplementary Material

Supplementary data are available at *Nicotine and Tobacco Research* online

Funding

The Joint Research Centre is funded by the EU's Framework Programme for Research and Innovation.

Downloaded from https://academic.oup.com/ntr/advance-article-abstract/doi/10.1093/ntr/nlz066/5486026 by guest on 22 May 2019

Declaration of Interests

The Joint Research Centre is the European Commission's science and knowledge service, carrying out research and providing independent scientific advice in support of EU policies. The authors declared no conflict of interest in the subject matter or materials discussed in this article.

References

1. Rodgman A, Perfetti TA. *The Chemical Components of Tobacco and Tobacco Smoke*. Boca Raton, FL: CRC Press, Taylor & Francis Group. 2012:1221–1298, 1471-1475.
2. Chida M, Sone Y, Tamura H. Aroma characteristics of stored tobacco cut leaves analyzed by a high vacuum distillation and canister system. *J Agric Food Chem*. 2004;52(26):7918–7924.
3. Leffingwell JC. Basic chemical constituents of tobacco leaf and differences among tobacco types. In: Davis DL, Nielson MT, eds. *Tobacco: Production, Chemistry and Technology*. Malden, MA: Blackwell Science. 1999:265–284.
4. Leffingwell JC, Young HJ, Bernasek E. *Tobacco Flavouring for Smoking Products*. Winston-Salem, NC: R. J. Reynolds Tobacco Company. 1972:3–10.
5. Baker RR, Massey ED, Smith G. An overview of the effects of tobacco ingredients on smoke chemistry and toxicity. *Food Chem Tox*. 2004;42:53–83.
6. Leffingwell JC, Alford ED, Leffingwell D. Aroma constituents of a supercritical CO2 extract of Kentucky dark fire-cured tobacco. *Leffingwell Rep*. 2013;5(1):1–21.
7. Cai J, Liu B, Ling P, Su Q. Analysis of free and bound volatiles by gas chromatography and gas chromatography-mass spectrometry in uncased and cased tobaccos. *J Chromatogr A*. 2002;947(2):267–275.
8. Peng F, Sheng L, Liu B, Tong H, Liu S. Comparison of different extraction methods: steam distillation, simultaneous distillation and extraction and headspace co-distillation, used for the analysis of the volatile components in aged flue-cured tobacco leaves. *J Chromatogr A*. 2004;1040(1):1–17.
9. Leffingwell JC, Alford ED, Leffingwell D, et al. Identification of the volatile constituents of cyprian Latakia tobacco by dynamic and static headspace analysis. *Leffingwell Rep*. 2013;5(2):1–29.
10. Merckel C, Pragst F, Ratzinger A, Aebi B, Bernhard W, Sporkert F. Application of headspace solid phase microextraction to qualitative and quantitative analysis of tobacco additives in cigarettes. *J Chromatogr A*. 2006;1116(1–2):10–19.
11. Clark TJ, Bunch JE. Qualitative and quantitative analysis of flavor additives on tobacco products using SPME-GC-mass spectroscopy. *J Agric Food Chem*. 1977;45(3):844–849.
12. Ochiai N, Mitsui K, Sasamoto K, Yoshimura Y, David F, Sandra P. Multidimensional gas chromatography in combination with accurate mass, tandem mass spectrometry, and element-specific detection for identification of sulfur compounds in tobacco smoke. *J Chromatogr A*. 2014;1358:240–251.
13. Wilkes JG, Conte ED, Kim Y, et al. Sample preparation for the analysis of flavors and off-flavors in foods. *J Chrom A*. 2000;880(1–2):3–33.
14. Valero E, Sanz J, Martínez-Castro I. Direct thermal desorption in the analysis of cheese volatiles by gas chromatography and gas chromatography-mass spectrometry: comparison with simultaneous distillation-extraction and dynamic headspace. *J Chromatogr Sci*. 2001;39(6):222–228.
15. Huang LF, Zhong KJ, Sun XJ, et al. Comparative analysis of the volatile components in cut tobacco from different locations with gas chromatography-mass spectrometry (GC-MS) and combined chemometric methods. *Anal Chim Acta*. 2006;575(2):236–245.
16. Brokl M, Bishop L, Wright CG, Liu C, McAdam K, Focant JF. Multivariate analysis of mainstream tobacco smoke particulate phase by headspace solid-phase micro extraction coupled with comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry. *J Chromatogr A*. 2014;1370:216–229.
17. Li Y, Pang T, Li Y, et al. Gas chromatography-mass spectrometric method for metabolic profiling of tobacco leaves. *J Sep Sci*. 2011;34(12):1447–1454.
18. Xiang G, Yang H, Yang L, et al. Multivariate statistical analysis of tobacco of different origin, grade and variety according to polyphenols and organic acids. *Microchem J*. 2010;95(2):198–206.
19. Gowda H, Ivanisevic J, Johnson CH, et al. Interactive XCMS online: simplifying advanced metabolomic data processing and subsequent statistical analyses. *Anal Chem*. 2014;86(14):6931–6939.
20. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem*. 2006;78(3):779–787.
21. Wold S. Pattern recognition by means of disjoint principal components. *Pattern Recogn*. 1976;8(3):127–139.
22. Jolliffe IT. *Principal Component Analysis*. New York, NY: Springer. 1986:115–128.
23. Deming SN, Michotte Y, Massart DL, et al. *Chemometrics: A Textbook*. Amsterdam, the Netherlands: Elsevier Science. 1988.
24. Eriksson L, Johansson E, Kattaheh-Wold N, et al. *Multi- and Megavariate Data Analysis*. Umea, Sweden: Umetrics AB. 2006.
25. Flavournet and human odor space. http://www.flavornet.org/ Accessed May 8, 2019.
26. The Pherobase: Database of Pheromones and Semiochemicals. http://www.pherobase.com/ Accessed November 25, 2017.
27. NIST Chemistry WebBook. http://webbook.nist.gov/chemistry/ Accessed January 7, 2018.
28. PubChem Substance and Compound databases. https://pubchem.ncbi.nlm.nih.gov/ Accessed January 7, 2018.
29. Leffingwell & Associates. *Flavour-Base 10—Tobacco Version*. http://www.leffingwell.com/flavbase.htm/ Accessed May 08, 2019.

Downloaded from https://academic.oup.com/ntr/advance-article-abstract/doi/10.1093/ntr/ntz066/5486026 by guest on 22 May 2019