Research Article

PTR-TOF-MS and data-mining methods for rapid characterisation of agro-industrial samples: influence of milk storage conditions on the volatile compounds profile of Trentingrana cheese[†]

Alessandra Fabris,[a] Franco Biasioli,[a]* Pablo M. Granitto,[b] Eugenio Aprea,[a] Luca Cappellin,[a,c] Erna Schuhfried,[c] Christos Soukoulis,[a] Tilmann D. Märk,[c] Flavia Gasperi[a] and Isabella Endrizzi[a]

Proton transfer reaction-mass spectrometry (PTR-MS), a direct injection mass spectrometric technique based on an efficient implementation of chemical ionisation, allows for fast and high-sensitivity monitoring of volatile organic compounds (VOCs). The first implementations of PTR-MS, based on quadrupole mass analyzers (PTR-Quad-MS), provided only the nominal mass of the ions measured and thus little chemical information. To partially overcome these limitations and improve the analytical capability of this technique, the coupling of proton transfer reaction ionisation with a time-of-flight mass analyser has been recently realised and commercialised (PTR-TOF-MS). Here we discuss the very first application of this new instrument to agro-industrial problems and dairy science in particular. As a case study, we show here that the rapid PTR-TOF-MS fingerprinting coupled with data-mining methods can quickly verify whether the storage condition of the milk affects the final quality of cheese and we provide relevant examples of better compound identification in comparison with the previous PTR-MS implementations. In particular, 'Trentingrana' cheese produced by four different procedures for milk storage are compared both in the case of winter and summer production. It is indeed possible to set classification models with low prediction errors and to identify the chemical formula of the ion peaks used for classification, providing evidence of the role that this novel spectrometric technique can play for fundamental and applied agro-industrial themes. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: proton transfer reaction mass spectrometry; time of flight; Trentingrana; volatile compounds; cheese

Introduction

Proton transfer reaction-mass spectrometry (PTR-MS),[1] a direct injection mass spectrometric technique based on an efficient implementation of chemical ionisation by proton transfer from hydronium ions, allows for fast and high-sensitivity monitoring of volatile organic compounds (VOCs). Applications of PTR-MS in food science and technology, as well as in environmental and medical areas, have been proposed since its invention. Several groups actively contributed to exploring the potential of PTR-MS in food science investigating the possibility to use the quadrupole-based version of PTR-MS (PTR-Quad-MS) for rapid product characterisation,[2,3] on-line process monitoring[4] and real-time in vivo VOCs detection during food consumption.[5]

PTR-MS, in the conventional configuration, does not use separation before mass spectrometric analysis and this limits the analytical information available: usually only the nominal mass of the ions. To partially overcome this issue and to increase the analytical information provided, the coupling of the ionisation method of PTR-MS with different mass analysers has been proposed[6] and, in particular, that with a time-of-flight (TOF) spectrometer has been recently realised[7,8] and commercialised.[9]

The coupling of different modern data-mining methods to rapidly and efficiently exploit the information entangled in PTR-Quad-MS spectra has been proposed and implemented in several case studies.[3,10] It provides an efficient method to build models for sample identification (classification models like random forest[11]), for correlation with other analytical determinations (calibration models like partial least squares[12]) and for identifying relevant features of the measured samples (feature selection[13]). However,

the application of these methods to PTR-TOF-MS data is not straightforward as, compared with quadrupole mass analysers, TOF analysers have the drawback of producing much larger data sets (i.e. 200 data points for quadrupole-generated spectra compared to 400.000 in the example discussed below with the TOF analyser). Thus, an appropriate peak extraction procedure is necessary to provide a data set with a reasonable number of variables.

Grana Padano, as other cheese of the 'grana' family (with a distinctive granular texture) as Parmigiano Reggiano, is a semi-fat, hard, cooked cheese that undergoes a slow ripening period of up to 2 years. It is economically relevant in the northern part of Italy. 'Trentigrana' cheese is produced in the Alpine province of Trento and is a member of the 'Grana Padano' consortium enjoying PDO (protected designation of origin).[14] Trentigrana is produced from partly skimmed raw cow's milk acidified with a mixture of Lactobacillus and coagulated by the addition of bovine rennet but, having a more restrictive protocol as compared with Grana cheese, the use of lysozyme and silage for the cow's feeding are not allowed.[15]

At the 18 cheese factories belonging to the Trentingrana consortium, the milk used comes from a double milk collection: the full fat milk of the evening milking is delivered to the cheese factory in churns (traditional method) or in tanks at controlled temperature (18 °C) and undergoes a natural creaming process (or gravity separation[16]) in large vats. After skimming the evening milk is added to the milk of the morning milking and used to produce cheese according to the standard cheese-making procedure of the Trentigrana. A single milk collection has been proposed as more economic: the milk of the morning milking is stored at the dairy farm under controlled temperature and moderate stirring, added to the evening milk and transported to the cheese factory where a partial skimming takes place overnight. The effects of the different dairy maturation of milk on physico-chemical characteristics, on rennet-coagulation aptitude and on rheological properties have been thoroughly studied for the case of Parmigiano Reggiano by Malacarne *et al.*[17] Despite the number of studies about the influence of storage on the milk quality, little is known about the relationship between milk storage modalities and the quality of the cheese produced. Tavoria and Malcata[18] showed that cheese manufactured with refrigerated or non-refrigerated milk are different in terms of micro-structural characteristics, but no studies have been published on the effect on the final volatile compounds profile of ripened cheese.

Volatile organic compounds (VOCs) in cheese originate to a great part from the activity of microorganisms and enzymes on carbohydrates, fats and proteins naturally present in milk and rennet through biochemical pathways related to glycolysis, lipolysis and proteolysis.[19–22] The VOCs in cheese belonging to the 'grana' family, Grana Padano, Parmigiano Reggiano and Trentingrana, have been extensively investigated mostly by gas-chromatographic methods.[23–26] They include several classes of compounds: acids, esters (in particular ethyl esters originated by the corresponding fatty acids), aldehydes, ketones, alcohols, lactones and sulphur compounds.[27]

Because volatile compounds play a fundamental role in the development of cheese flavour, it seems relevant, both from the technological and fundamental point of view, to have a rapid method to evaluate the final effect on cheese flavour profile and quality.

In this work we propose the application, to the best of our knowledge for the first time, of the newly introduced PTR-TOF-MS instrument to agro-industrial questions. In particular the possibility to rapidly characterise the volatile compounds profile of Trentingrana cheese produced with milk stored in different conditions was studied, i.e. on one side, the possibility of sample classification by rapid PTR-TOF-MS characterisation and data-mining methods was investigated and, on the other, the chemical information provided by the new PTR-TOF-MS in comparison with the quadrupole-based version is presented and discussed.

Materials and Methods

Cheese samples

Cheese loaves (three for each case considered below) were produced in one of the cheese factories of the Trentingrana consortium (Primiero, Italy) and aged for 20 months. Half of the samples were produced in winter 2007 (labelled in the following with a 'W') and half during summer time (labelled with a 'S').

In the production of these cheese loaves, milk subjected to 4 different storage conditions was employed: on the one hand double milking collected in churns (B as 'Bidoni' in Italian) and kept at room temperature (samples labelled with S_B and W_B) or at 18 °C (samples labelled with S_18 or W_18); on the other, single milk collection (morning and evening milk together) with day-time storage at 12 °C (indicated with S_12 and W_12) and 8 °C (indicated with S_8 and W_8).

PTR-TOF-MS analysis

All measurements were performed with a commercial PTR-TOF 8000 instrument[9] supplied by Ionicon Analytik GmbH, Innsbruck (Austria). The inlet system, the production of hydronium ions and the ionisation in the drift tube proceed in the same way as discussed for quadrupole-based instruments.[1] The proton transfer reaction was controlled by drift voltage (600 V), drift temperature (110 °C) and drift pressure (2.11 mbar). Inlet flux (about 100 sccm) was adjusted to stabilise the drift tube pressure. The TOF was operated in V mode. The sampling time of the TOF spectra was 0.1 ns.

From a piece of cheese of about 1 kg, after removal of the rind (5 cm from the external side, 1 cm from the tip), slices of about $12 \times 6 \times 1.5$ cm^3 were grated and well mixed to reduce the variability of the sample due to the non-uniformity of the different parts of the cheese. About 3 g of this grated cheese was placed in glass vials (120 ml, Supelco, Bellefonte, USA) with bidistilled water (4 ml) and capped by PTFE/Silicone septa (Supelco). Each sample was prepared and measured in triplicate. Samples were equilibrated at 40 °C for 30 min in a water bath before the analysis; they were then measured by direct injection of the head space mixture into the PTR-TOF-MS drift tube via a heated (110 °C) peek inlet for 20 s, allowing to take 20 average spectra. We choose relatively high temperatures both for the drift tube and for the inlet to reduce memory effects and to reduce condensation. The measurement order has been randomised to avoid possible systematic memory effects. The absence of pre-treatments or pre-concentration and the equilibration at 40 °C provide a means to measure VOC concentrations as close as possible, for a static measurement, to the one released during cheese consumption. Reference blank vials containing only 4 ml of bidistilled water have been measured as well.

Peak intensity in ppb$_v$ has been estimated by the formula described in Lindinger *et al.*[28] using a constant value for the

reaction rate constant ($k = 2.10^{-9}$ cm^3 s^{-1}). This introduces a systematic error for the absolute concentration for each compound that is in most cases below 30% and can be accounted for if the actual rate constant is available.[29]

Data analysis

Spectra have been acquired using the data acquisition software TOF-DAQ (Tofwerk AG, Switzerland) with a mass range of 10–400 Th. A total of 28 860 spectra has been added before storage and only the resulting sum spectra, one each second, have been stored in HDF5 format[30] for efficient data storage and direct access to data structure and considered for data analysis.

Calibration and peak extraction has been performed as described in[31] and the 20 spectra referring to the same sample have been averaged after alignment of single spectra. Internal calibration was based on three peaks always present in the PTR-MS spectra at $m/z = 18.0338$ (NH$_4^+$), 21.0202 (H$_3^{18}$O$^+$) and 29.9974 (NO$^+$) and on three esters-related peaks that have always high concentrations and do not show any interfering structure in all measurements [$m/z = 89.0597$ (C$_4$H$_9$O$_2^+$), 117.0910 (C$_6$H$_{13}$O$_2^+$) and 145.1223 (C$_8$H$_{17}$O$_2^+$)]. Throughout the article, we use three decimal figures for estimated m/z values and four for the expected exact ones.

We observed, in the PTR-TOF-MS spectra, more than 800 peaks but we considered for further data analysis only the areas of the 399 most intense peaks (estimated peak area greater than 1 ppb). The data set thus consisted of a matrix with 72 lines (the samples: 2 seasons × 4 milk ripening × 3 loaves × 3 replicates) and 399 columns (the peaks). Every cell contained the area of the corresponding peak for the corresponding sample.

Multivariate data analysis (principal component analysis) and non-parametric ANOVA (Kruskal–Wallis test) have been performed by Statistica 8.0 (StatSoft, USA). Classification has been performed by four different data-mining methods: (1) random forest (RF),[11] (2) penalised discriminant analysis (PDA),[12] (3) support vector machine (SVM)[32] and (4) discriminant partial least squares (dPLS).[33] The four classification methods have been previously described and applied to PTR-Quad-MS analysis with good results.[10,13] In all cases, we used implementations available as free packages for the R statistical environment software.[34] To evaluate the results of the classification methods we used a leave-group-out (LGO) method: we iterated the process of leaving a group out as test set and using the remaining of the data set to fit the models. The free parameters of each classifier, such as the C constant of SVM or the number of dimensions considered in dPLS, were selected at this step by internal cross validation using only the training data set. After that, those models were used to individually classify the samples of the independent test batch. Each individual group in this LGO procedure consisted of the three replicates of the same loaf, in order to evaluate really independent test sets (the good reproducibility of PTR-MS evaluations results in a high correlation among replicates of the same sample, which bias the result of the discriminant analysis if not taken into account). We analysed the classification results using confusion matrices, in which rows correspond to the true classes and columns to the predicted ones. The results are given in percentage of samples of each product that the classifier assigns to the product given by the column title.

RF can also be used to analyse the data set in a graphical way, complementing the more typical PCA analysis. RF graphical outputs are multidimensional scaling projections[35] of the data set that utilise a particular measure of distance among samples based on the internal assignation of classes in the RF ensemble. Granitto et al.[36] discussed the use of this tool in the analysis of food data. The authors showed that RF visualisations can be very informative for discrimination tasks, as they use information about the real classes (opposite to PCA) and also can be less biased than other supervised visualisations (like LDA or PLS) because RF plots are based on unseen samples.

We used random forest - recursive feature elimination (RF-RFE) introduced by Granitto et al.[13] in order to select a few very discriminant peaks for each case. The procedure has two steps. First, RF-RFE is applied separately to each one of the several g partitions in training and test sets produced by the LGO procedure described earlier. The method produces an average error curve relating the classification error with the number of peaks used in the model. We use that curve to select a number p of peaks that is as low as possible but still yields good discriminant models.[13] In the second step, we select the top p peaks from each of the g runs of RF-RFE. We compute the average number of times that each of the original 399 extracted peaks is selected in these reduced lists of p discriminant peaks, and keep only the peaks that were selected more than 10% on average. It is important to note that the output of the process is a list of peaks that are highly relevant to the problem, not the subset that produces the lowest classification error.

Results and Discussion

PTR-TOF-MS spectra of cheese samples

While the quadrupole-based version of PTR-MS has proven to be highly sensitive and quantitative in many applications, it provided only the nominal mass of the observed spectrometric peaks and thus little chemical information in the case of the complex volatile compound mixtures often found in food science and technology. The mass resolution and mass accuracy (better than 10 ppm) of PTR-TOF-MS[31] provides, without any loss in sensitivity and speed, the possibility to have, for most peaks, an unambiguous determination of the chemical formula. Figure 1a exemplifies the output of PTR-TOF-MS analysis of a cheese sample before any data pre-processing such as calibration or baseline correction. The actual spectra extend up to $m/z = 400$ but for the sake of clarity only the mass range 20–100 is shown.

At this scale of visualisation PTR-TOF-MS spectra resemble those of the quadrupole versions with signals at integer nominal mass[37]. However, now every single peak provides more information. For instance, the estimated exact mass of a single peak at nominal mass 35 (Fig. 1b) is 34.9952 ± 0.0004 and thus corresponds unambiguously to protonated H$_2$S, an utmost important compound because of its low detection threshold (10 ppb in water[38] or 0.18 ppb in air[39]) and its characteristic odour. There is no other combination of elements that can explain the presence of this peak. It is important to notice that for compounds such as H$_2$S or CO$_2$ (detected at mass 45 and separated by the TOF from acetaldehyde) with a proton affinity close to that of water, back-reaction from the protonated ion to water is also possible. Thus, the concentrations provided for H$_2$S in the present article, on the basis of the above-mentioned formula,[28] may be values that underestimate the real concentration. Moreover, in these cases, a strong dependence on humidity is expected. This is, however, not an issue in the measurement presented here that are performed at constant relative humidity.
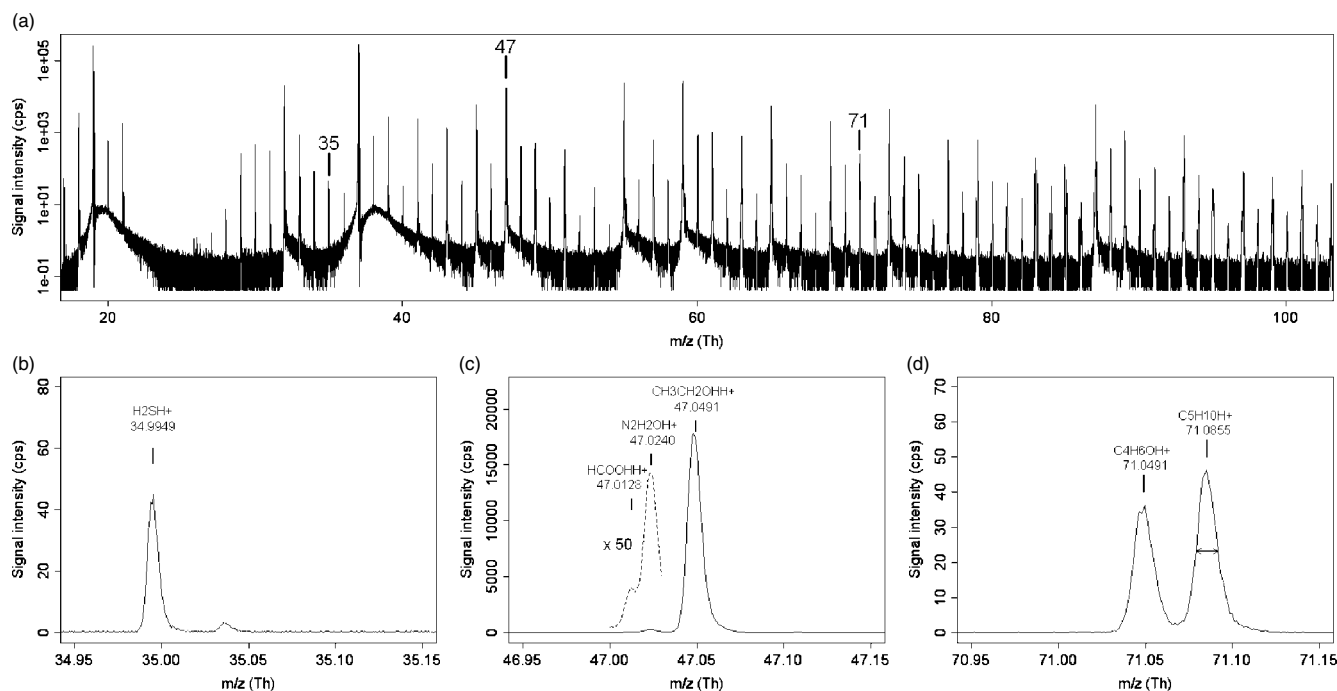
Figure 1. Example of PTR-TOF-MS spectrum of cheese samples headspace. The figure (a) represents a sample spectrum in the low mass region (20–100 Th) while figures (b), (c) and (d) enlarge the regions around three selected peaks indicated by ticks and nominal mass in (a). Refer the text for discussion.

The graph in Fig. 1c exemplifies the possibility of the simultaneous quantification of isobaric compounds, even at very different concentrations, as is the case for ethanol and formic acid (both at nominal mass of 47 Th) and demonstrates also the possibility to account for a contribution of an instrumental background at $m/z = 47.024$. Methanol is also detected at nominal mass 33 and the TOF allows the separation of the contribution of isobaric $^{18}O^{17}O^+$.

A further example, used here to quantify the instrument resolution, is the structure of the peaks at nominal mass 71 (Fig. 1d). Here, it is easy to separate the contribution of $C_4H_7O^+$ (m/z 71.050) from $C_5H_{11}^+$ (m/z 71.087) and the peak at $m/z = 71.087$ allows to estimate a value for the mass resolution of about 5000 [m/Δm (FWHM)].

Other interesting peaks are, for instance, (data not shown) the peak at nominal $m/z = 42$ where it is possible to distinguish the contribution of the isotope of the generic fragment $C_3H_6^+$ (m/z 42.046) from protonated acetonitrile at $m/z = 42.035$, a possible contaminant. At nominal mass 63, we have three well-separated peaks: m/z 63.008 (the cluster of CO_2 and water), m/z 63.028 (dimethylsulphide) and m/z 63.045 (acetaldehyde/water cluster). The possibility to measure sulphur volatile compounds by fast, non-invasive head space analysis is of utmost relevance for applications (discussed later).

Esters produce a series of intense peaks observed at nominal mass 75 +n·14 with a common fragment at $m/z = 61$ interfering with acetic acid; ketones and aldehydes at 59 + n·14. In general, acetate esters show a common fragment at $m/z = 61$ while ethyl esters show a common fragment at $m/z = 89$.[40] We observe that even-chained fatty acid and esters are more intense than the odd-chained ones. This is expected because the formation of fatty acids begins with acetyl-CoA and proceeds via the addition of C2 units. The chemical formula of many other peaks can be determined. We limit the analysis here to these relevant examples and discuss later on the case of peaks that turned out to be important for the classification models.

Sample classification

Principal component analysis on the data matrix extracted from PTR-TOF-MS spectra (Fig. 2) indicates the presence of differences among the samples and that part of the variance is related to the different milk storage conditions, at least for summer samples. Only three components describe 69.6% of the total variance for summer samples and 64.05% for the winter samples. There are, however, other effects related to the variability of the cheese production that contribute to the total variance as indicated, for example, by the fact that the second PCA (not shown) component is used to separate summer samples according to some other causes of variation. With only two components it is, however, possible to provide a relatively good separation of the investigated classes. For summer samples, the separation is clear: the first principal component separates the B samples from the others, while the third one allows the separation of the other group samples. Winter samples show a similar but less evident separation: two components allow the separation of cheese produced with the milk stored at the lowest temperature. It is not possible to distinguish W_B (room temperature) from W_18. However, due to the partial superposition of the seasonal effect and of the technological process and due to the presence of unavoidable variability associated with the production process, PCA, a non-supervised technique, is not the best way to set up an optimum classification model for sample discrimination. In particular, PCA is not able to separate the variation associated with the season from that associated to the process.

A more efficient way to investigate the possibility to differentiate the samples on the basis of the PTR-TOF-MS fingerprint is to use modern data-mining methods.
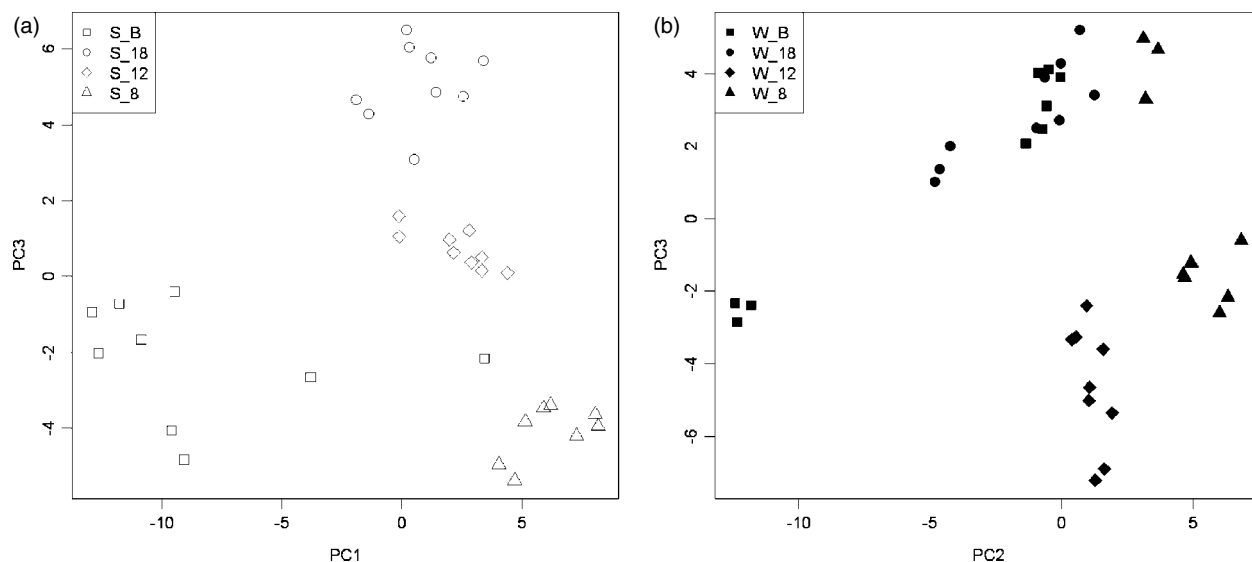
Figure 2. PCA analysis of the PTR-TOF-MS data of summer samples (a) and winter samples (b).
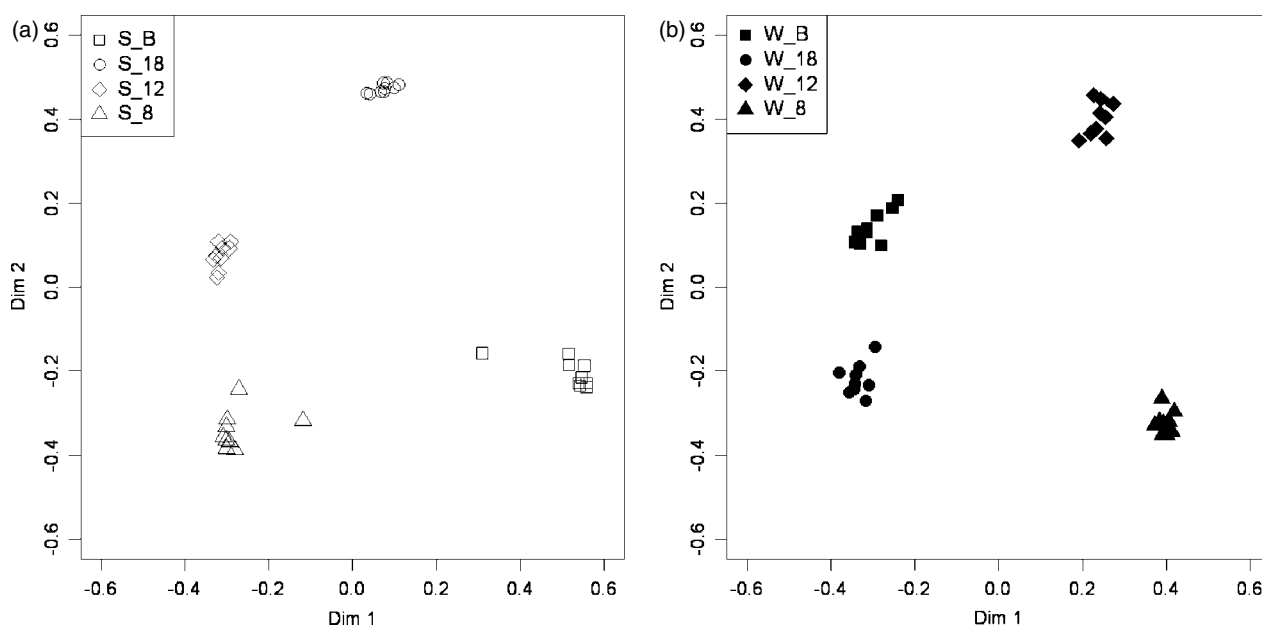

Figure 3. Random Forest graphical output for the classification analysis of summer samples (a) and winter samples (b).

As we illustrated earlier, RF provides a useful graphical output (Fig. 3) which clearly suggests the possibility of better discrimination and allows to emphasise the relative difference among samples. For instance, Fig. 3a shows how S_B samples are separated from S_8, S_12 and S_18, which are ordered along the second dimension according to the milk storing temperature and this description corresponds closely to the one obtained by PCA (Fig. 2a).

We built three different classification models: (1) for seasons (discrimination of winter and summer samples); (2) for milk storage conditions, for summer and winter samples separately; and (3) for all classes at a time. In all cases, we tested the four methods mentioned above (RF, PDA, SVM and dPLS).

The separation of summer samples from winter samples by PDA gives a high success rate: only 4 samples on 72 were wrongly classified.

Tables 1 and 2, (winter and summer data modelled separately) and (all classes together), show the cross-validated confusion matrices of the classification models obtained by PDA. They confirm that samples can be efficiently separated on the basis of rapid PTR-TOF-MS fingerprinting. In particular, summer samples are always well classified while the winter samples produced with milk stored at 8 or 12 °C are not distinguishable. We found slightly worse classification performances for the model that tries to classify all cases (Table 2). The classification errors are however among samples that are expected to be similar: for instance, coldest milk winter samples are confused with summer samples produced with milk stored at the same temperature.

We report the confusion matrices only for PDA. Because the other classification methods (RF, SVM and dPLS) have similar performances and provide similar indications, we report only the comparison of their average classification error (Table 3). PDA per-

Table 1. Confusion matrix for the classification by penalised discriminant analysis of winter samples (left) and summer samples (right) separately

	W_B	W_18	W_12	W_8		S_B	S_18	S_12	S_8
W_B	9	0	0	0	S_B	9	0	0	0
W_18	1	8	0	0	S_18	0	9	0	0
W_12	0	0	5	4	S_12	0	0	9	0
W_8	0	0	3	6	S_8	0	0	1	8

Table 2. Confusion matrix for the classification by penalised discriminant analysis of all samples

	W_B	W_18	W_12	W_8	S_B	S_18	S_12	S_8
W_B	5	0	0	0	0	4	0	0
W_18	1	8	0	0	0	0	0	0
W_12	0	0	6	3	0	0	0	0
W_8	0	0	1	3	0	0	2	3
S_B	0	0	0	0	9	0	0	0
S_18	0	0	0	0	0	9	0	0
S_12	0	0	0	0	0	0	9	0
S_8	0	0	1	0	0	0	0	8

Table 3. Comparison of the performances (mean classification errors) of the four classification methods tested

	Winter vs Summer	Winter samples	Summer samples	All classes
RF	0.11	0.31	0.06	0.21
PDA	0.06	0.22	0.03	0.21
SVM	0.10	0.25	0.03	0.21
dPLS	0.10	0.25	0.11	0.11
Mean	0.09	0.26	0.06	0.18

RF, random forest; PDA, penalised discriminant analysis; SVM, support vector machines; dPLS, discriminant partial least squares.

forms better in season classification, winter samples classification and together with SVM in summer samples classification. dPLS has the smallest classification error in the most complex case of the classification of all classes at a time.

Discriminant peaks

Table 4 represents the list of the most informative peaks for the models that classify winter samples against summer samples and for the models that classify the four considered cases in winter and in summer, respectively.

It is interesting to note that the peaks used in the setting of the model for the discrimination of summer samples are different from the ones used in the case of winter samples. In particular, all peaks related to the ester series, even carbon chains, (117.091, 118.092, 145.121 and 173.151) are used only for summer samples while several peaks related to aldehydes and ketones (59.049, 60.053, 73.065, 101.060, 101.097, 115.112 and 143.143) are used only for winter samples or to discriminate winter from summer samples. In this latter case, they have

higher concentrations for winter samples. Figure 4 shows the actual concentration data for selected peaks of the ester series. There is an evident effect on the concentration by the milk storage temperature, i.e. all peaks increasing with increasing milk temperature. In the winter samples, we observe a similar effect but the variability of the data is much larger and does not allow a clear separation. Although beyond the scope of this article, we propose a possible explanation of the observed effect based on evidence from literature[19,41,42] on the activity of psychrotrophic bacteria as Pseudomonas fragi in cheese: they show both hydrolytic and esterifying activity, the first reaction having an optimum efficiency at 25–27 °C, the latter at 12–15 °C. Only in the case of high temperature storage (more evident in summer S_B and S_18) we have efficient hydrolysis and formation of fatty acids that are then esterified during the skimming that occurs for all samples at 15 °C.

An important advantage of PTR-TOF-MS is that its mass resolution and accuracy allow to separate and quantify the contribution to the mass spectrometric signal of important classes of compounds as nitrogen and sulphur compounds.

As an example, let us consider a spectrum of a cheese sample in the region around $m/z = 95$ in Fig. 5b: the estimated mass of the highest peak is 94.999 that is in good agreement with the exact mass of dimethyl disulphide (94.9984). Other possible formulas are not chemically sound and can be excluded considering the isotope ratio of the peaks at $M + 1$ and $M + 2$. In fact, for protonated dimethyl disulphide, the expected isotopic pattern is 3.95% at $M + 1$ and 8.94% at $M + 2$ that is in good agreement with the observed one (Fig. 5). Figure 5a shows the peak at nominal mass of 79. Similar considerations can be carried out: again, estimated mass and isotopic patterns allow us to define the chemical formula of the ion (CH_3SSH^+) that can be assumed to be a fragment of dimethyl disulphide also supported by the high correlation between the two signals ($r^2 = 0.99$, $p < 0.001$) and in good agreement with fragmentation reported by Aprea et al.[43]

Dimethyl disulphide decreases in summer samples with increasing milk storage temperature. Other very interesting peaks related to sulphur compounds are present even if not indicated as discriminating by RF-RFE analysis as, for instance $m/z = 49$ (methanthiol), 63 (dimethyl sulphide, well separated from the isobaric water/acetaldyde cluster), 105 (methional) and 35 (hydrogen sulphide, Fig. 1). Volatile sulphur compounds are considered to be important flavour contributors to various cheese even if present at low concentrations.[26,44] Sulphur-containing compounds such as hydrogen sulphide and methanethiol are mainly produced by the degradation of methionine that result from cleavage of bonds between carbon and sulphur by a methionine–demethiolase, during cheese ripening and methanethiol can be converted to dimethyl disulphide and dimethyl trisulphide by oxidative reactions.[45]

Another interesting discriminating peak is the one found at $m/z = 109.075$. Both its m/z value and the isotopic pattern (Fig. 6) indicate that the formula of this ion is $C_6H_8N_2H^+$ (dimethylpyrazine). It is used in the discriminant models because its concentration in the winter samples produced with milk stored at the coldest temperature is lower. Alkyl pyrazines have been recognised as important trace flavour components of a large number of heated foods and are believed to form as a result of the Maillard non-enzymatic browning reaction and Strecker degradation reactions[46,47] and their presence in cheese is well documented.[23,48–51] Heterocyclic compounds may also originate

Table 4. List of the more relevant peaks (in the sense of RF-RFE, refer text for details) for the different classification models: (1) winter versus summer, (2) different milk temperature for winter and (3) summer samples. Estimated mass and corresponding exact mass of the peaks that have been used by RF are reported in the first two columns. The concentrations in the head space are reported in ppb_v. Apexes (a,b,c,d) indicate if there are statistically significant difference (values with different letters are significantly different) according to a Kruskal–Wallis test ($p < 0.001$)

| Estimated m/z	Exact m/z	Chemical formula	1 Season Winter	Summer	2 Winter W_B	W_18	W_12	W_8	3 Summer S_B	S_18	S_12	S_8
42.035	42.0338	C_2H_4N	16.5[a]	29.3[b]	–	–	–	–	–	–	–	–
57.071	57.0699	C_4H_9	136.5[a]	121.8[a]	–	–	–	–	–	–	–	–
59.049	59.0491	C_3H_7O	3238[a]	4710[b]	2419[ab]	2975[b]	2334[a]	1803[c]	–	–	–	–
60.053	60.0525	$C_2{}^{13}CH_7O$	–	–	82.4[ab]	99.2[b]	79.3[a]	172.3[c]	–	–	–	–
73.065	73.0648	C_4H_9O	1203[b]	1050[a]	1044[a]	957[a]	1414[b]	9.4[b]	–	–	–	–
74.069	74.0682	$C_3{}^{13}CH_9O$	–	–	47.0[a]	42.9[a]	60.1[b]	66.6[b]	–	–	–	–
75.045	75.0441	$C_3H_7O_2$	15.4[a]	15.7[a]	–	–	–	–	–	–	–	–
77.059	77.0597	$C_3H_9O_2$	–	–	66.8[a]	85.5[b]	71.0[ab]	42.3[c]	–	–	–	–
78.967	78.9671	CH_3S_2	2.6[a]	3.5[b]	–	–	–	–	2.3[a]	3.0[b]	4.1[c]	4.7[d]
82.066	82.0651	C_5H_8N	–	–	1.1[c]	4.6[ab]	5.5[b]	0.3[a]	0.7[b]	4.4[a]	9.6[c]	5.5[a]
85.065	85.0648	C_5H_9O	15.8[b]	11.8[a]	–	–	–	–	–	–	–	–
94.999	94.9984	$C_2H_7S_2$	–	–	–	–	–	–	4.6[b]	6.1[c]	8.5[a]	9.2[a]
97.102	97.1012	C_7H_{13}	27.7[b]	20.0[a]	–	–	–	–	–	–	–	–
101.060	101.0597	$C_5H_9O_2$	–	–	7.3[a]	8.7[a]	11.3[b]	3.7[b]	–	–	–	–
101.097	101.0961	$C_6H_{13}O$	–	–	37.9[b]	37.5[b]	21.5[a]	9.6[a]	25.0[a]	32.9[c]	25.7[a]	16.6[b]
102.098	102.0995	$C_5{}^{13}CH_{13}O$	–	–	–	–	–	–	2.1[a]	3.7[b]	3.5[b]	2.3[a]
109.075	109.0760	$C_6H_9N_2$	–	–	6.9[a]	9.5[b]	12.9[c]	0.7[d]	–	–	–	–
110.078	110.0788	$C_5{}^{13}CH_9N_2$	–	–	0.5[a]	0.7[b]	1.0[c]	1.3[d]	–	–	–	–
115.112	115.1117	$C_7H_{15}O$	397.7[b]	287.6[a]	–	–	–	–	–	–	–	–
117.091	117.0910	$C_6H_{13}O_2$	–	–	–	–	–	–	390.8[c]	169.3[b]	93.3[a]	79.7[a]
118.092	118.0944	$C_5{}^{13}CH_{13}O_2$	–	–	–	–	–	–	26.2[c]	11.3[b]	6.4[a]	5.5[a]
121.064	121.0648	C_8H_9O	–	–	25.9[b]	22.3[a]	34.167[d]	1.9[c]	–	–	–	–
143.143	143.1430	$C_9H_{19}O$	40.4[b]	31.5[a]	–	–	–	–	–	–	–	–
145.121	145.1223	$C_8H_{17}O_2$	–	–	–	–	–	–	210.1[c]	80.4[b]	44.1[a]	37.4[a]
173.151	173.1536	$C_{10}H_{21}O_2$	–	–	–	–	–	–	18.4[c]	7.2[b]	3.8[a]	3.6[a]
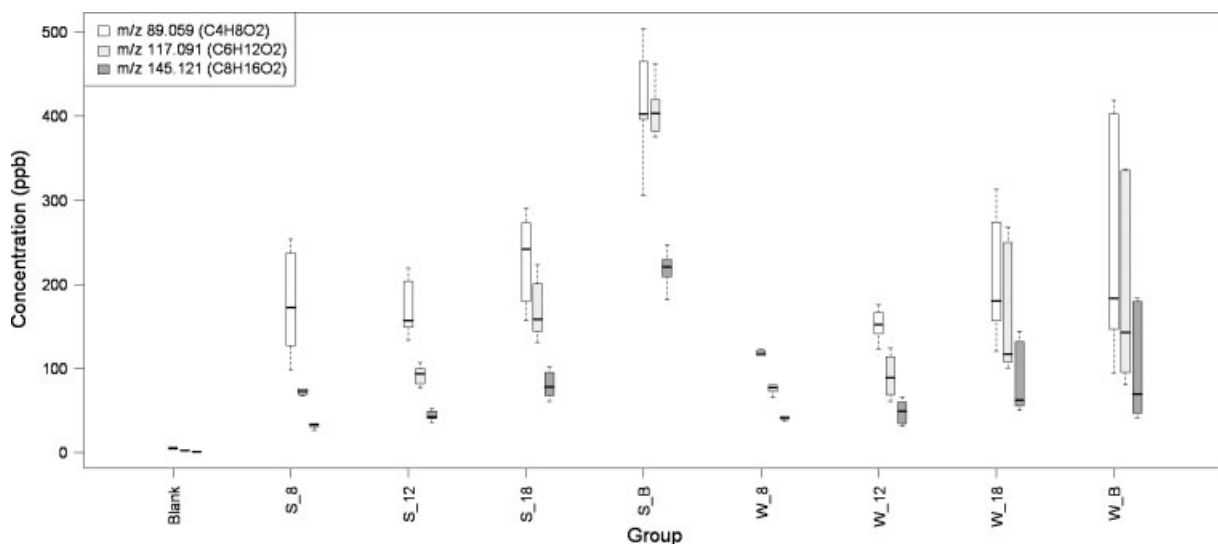
Figure 4. Box and whiskers plot of the measured concentration of peaks related to the discrimination of summer samples and corresponding to the molecular peaks of even carbon chain esters. Box indicate median and inter-quartile range. Whiskers indicate maximum and minimum values.

enzymatically in fruits and vegetables and during the ripening of cheese. Pyrazines appear to be present in unprocessed as well as in heated foods as natural aroma components[23,50] and strongly affect the perceived quality of food. Notably, no other peaks containing nitrogen atoms, beside the case of acetonitrile (discussed below), have been identified by the PTR-TOF-MS.

Measuring the final effect on the matured cheese aims at the description of the VOCs profile that characterises the final product. This is the effect of several complicated processes occurring at
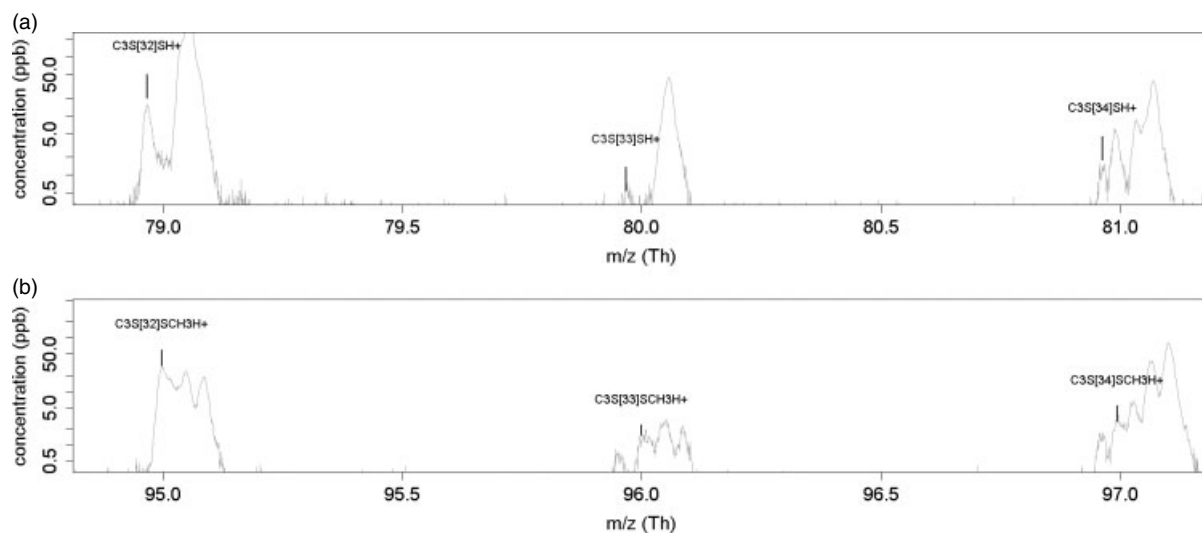
Figure 5. Spectral regions around nominal masses 95 and 79 (b) and (a). The abscissa of ticks indicate the expected position of dimethyl disulphide (nominal mass 95) and its fragment (nominal mass 79) and of their isotope substituted $M + 1$ and $M + 2$, while their ordinate indicate the expected isotopic abundance relative to the base peak.
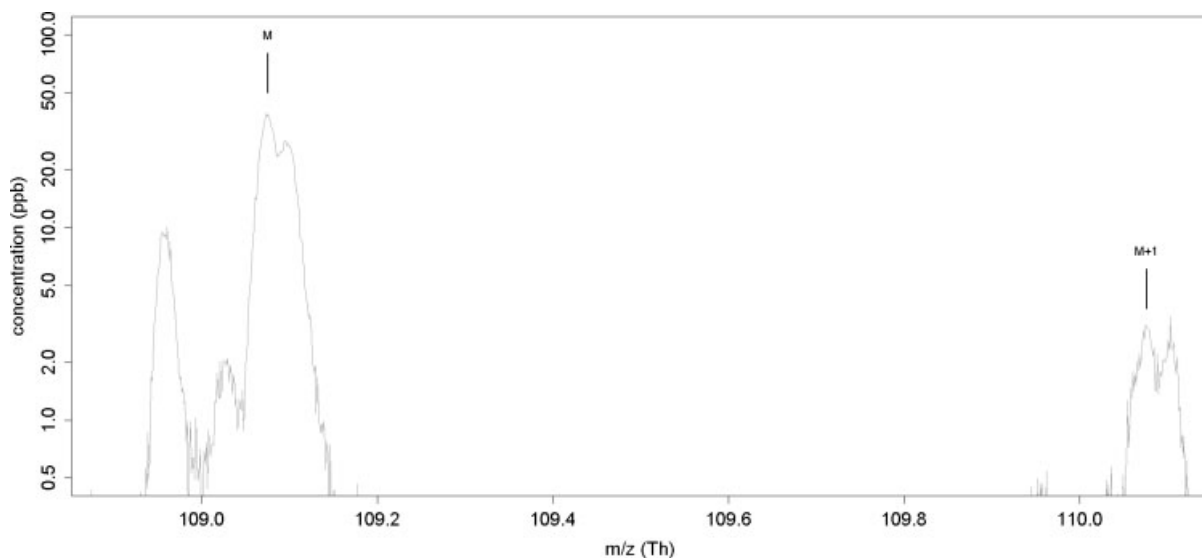

Figure 6. Spectral regions around nominal mass 109. Ticks indicate the expected position of dimethyl pyrazine and its isotope ($M = 109.0760$ and $M + 1 = 110.0789$).

different production stages: milk storage, cheese making and ripening; thus, in general, for the characterisation of the final product, it is not enough to draw detailed conclusions on the history of the measured samples.

Peaks related to possible contaminants

In view of application to process monitoring or quality control it is worth noticing the presence of other peaks associated to contaminants, both in the production and the storage phase. For instance, the peak at $m/z = 42.035$ corresponding to the formula of acetonitrile $CH_3CN\,H^+$ (exact protonated mass $= 42.0338$) is well separated from the generic fragment at $m/z = 42.046$ ($C_3H_6^+$). Being notably higher in summer samples, we assume that the contamination did not occur in the laboratory during sample preparation. Dichloromethane, also easily identified despite its low concentration by the exact mass ($m/z = 82.9455$) and by

its very characteristic isotopic pattern, is present with similar concentrations in all samples and is likely due to contamination at some stage of sample manipulation or storage.

Conclusions

We applied, as far as we know, for the first time, the newly developed PTR-TOF-MS in food science and technology and in particular to the rapid characterisation of cheese samples (Trentin-grana) produced with milk stored under different conditions and in two different seasons (summer and winter). PTR-TOF-MS proves to be a useful tool for the characterisation of cheese. A very rapid screening (20 s) allows the measurement of hundreds of peaks. For many of them, the chemical formula can be unambiguously identified and verified also by isotopic pattern. In the case of low m/z values, attribution to single compounds is often possible.

Unsupervised multivariate analysis suggests that samples can be separated on the basis of PTR-TOF-MS fingerprinting but supervised data-mining methods allow a better classification. This demonstrates that the cheeses investigated are well characterised and that the milk-ripening temperature has a crucial effect on the final product. Other data-mining methods (feature selection) highlighted the peaks that play a major role in the classification: esters for summer samples and, in particular, in the case of milk stored at higher temperature, ketones and aldehydes, on the contrary, are used in the classification of winter samples. The separation of compounds containing sulphur or nitrogen atoms permits their rapid quantification and opens the way to promising applications. Trace contaminants can also be easily detected.

This article indicates a useful strategy, from raw data treatment to data-mining elaboration, for the study of food science related problems and the characterisation of agro-industrial products and highlights the useful information contained in the PTR-TOF-MS spectra. The proposed strategy described in the present work is easily extendible to other technological and scientific themes and not only in the food area.

Acknowledgements

References

[1] W. Lindinger, A. Hansel, A. Jordan. Proton-transfer-reaction mass spectrometry (PTR-MS): on-line monitoring of volatile organic compounds at pptv levels. *Chem. Soc. Rev.* **1998**, *27*, 347.

[2] F. Biasioli, F. Gasperi, E. Aprea, L. Colato, E. Boscaini, T. D. Märk. Fingerprinting mass spectrometry by PTR-MS: heat treatment vs. pressure treatment of red orange juice-a case study. *Int. J. Mass Spectrom.* **2003**, *223–224*, 343.

[3] F. Biasioli, F. Gasperi, E. Aprea, D. Mott, E. Boscaini, D. Mayr, T. D. Märk. Coupling proton transfer reaction-mass spectrometry with linear discriminant analysis: a case study. *J. Agric. Food Chem.* **2003**, *51*, 7227.

[4] A. Hansel, A. Jordan, R. Holzinger, P. Prazeller, W. Vogel, W. Lindenger. Proton transfer reaction mass spectrometry: on-line trace gas analysis at the ppb level. *Int. J. Mass Spectrom. Ion Process.* **1995**, *149–150*, 609.

[5] E. Aprea, F. Biasioli, F. Gasperi, T. D. Märk, S. van Ruth. In vivo monitoring of strawberry flavour release from model custards: effect of texture and oral processing. *Flavour Fragr. J.* **2006**, *21*, 53.

[6] C. Warneke, J. A. de Gouw, E. R. Lovejoy, P. C. Murphy, W. C. Kuster. Development of proton-transfer ion trap mass spectrometry: on-line detection and identification of volatile organic compounds in air. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1316.

[7] C. Ennis, J. Reynold, B. J. Keely, L. J. Carpenter. A hollow cathode proton transfer reaction time of flight mass spectrometer. *Int. J. Mass Spectrom.* **2005**, *247*, 34.

[8] H. Tanimoto, N. Aoki, S. Inamoto, J. Hirokama, Y. Sadamaga. Development of a PTR-TOFMS instrument for real-time measurement of volatile organic compounds in air. *Int. J. Mass Spectrom.* **2007**, *263*, 1.

[9] A. Jordan, S. Haidacher, G. Hanel, E. Hartungen, L. Märk, H. Seehauser, R. Schottkowsky, P. Sulzer, T. D. Märk. A high resolution and high sensitivity proton-transfer-reaction time-of-flight mass spectrometer (PTR-TOF-MS). *Int. J. Mass Spectrom.* **2009**, *286*, 122.

[10] P. M. Granitto, F. Biasioli, E. Aprea, D. Mott, C. Furlanello, T. D. Märk, F. Gasperi. Rapid and non-destructive identification of strawberry cultivars by direct PTR-MS headspace analysis and data mining techniques. *Sens. Actuators B Chem.* **2007**, *121*, 379.

[11] L. Breiman. Random forests. *Mach. Learn.* **2001**, *45*, 5.

[12] T. Hastie, A. Buja, R. Tibshirani. Penalized discriminant analysis. *Ann. Stat.* **1995**, *23*, 73.

[13] P. M. Granitto, C. Furlanello, F. Biasioli, F. Gasperi. Recursive feature elimination with Random forest for PTR-MS analysis of agroindustrial products. *Chemometr. Intell. Lab. Sys.* **2006**, *83*, 83.

[14] EC. Official journal of the European Union, Regulation no. 1107/1996 of 21 June 1996, **1996**.

[15] S. Carini, R. Lodi. Grana Padano. Consideration of starter culture and lysozyme. *Latte* **1997**, *22*, 76.

[16] Y. Ma, D. M. Barbano. Gravity separation of row bovine milk: fat globule size distribution and fat content of milk fractions. *J. Dairy Sci.* **2000**, *83*, 1719.

[17] M. Malacarne, A. Summer, P. Formaggioni, P. Franceschi, S. Sandri, M. Pecorari, P. Vecchia, P. Mariani. Dairy maturation of milk used in the manufacture of Parmiggiano-Reggiano cheese: effects on physico-chemical characteristics, rennet-coagulation aptitude and rheological properties. *J. Dairy Res.* **2008**, *75*, 218.

[18] F. K. Tavaria, P. J. M. Reis, F. Xavier Malcata. Effect of dairy farm and milk refrigeration on microbiological and microstructural characteristics of matured Serra da Estrela cheese. *Int. Dairy J.* **2006**, *16*, 895.

[19] L. Marilley, M. G. Casey. Flavours of cheese products: metabolic pathways, analytical tools and identification of producing strain. *Int. J. Food Microbiol.* **2004**, *90*, 139.

[20] S.-Q. Liu, R. Holland, V. L. Crow. Esters and their biosynthesis in fermented dairy products: a review. *Int. Dairy J.* **2004**, *14*, 923.

[21] L. H. McSweeney, M. J. Sousa. Biochemical pathways for the production of flavour compounds in cheese during ripening: a review. *Lait* **2000**, *80*, 293.

[22] L. H. McSweeney. Biochemistry of cheese ripening. *Int. J. Dairy Technol.* **2004**, *57*, 127.

[23] G. Barbieri, L. Bolzoni, M. Careri, A. Mangia, G. Parolari, S. Spagnoli, R. Virgili. Study of the volatile fraction of Parmesan cheese. *J. Agric. Food Chem.* **1994**, *42*, 1170.

[24] F. Bellesia, A. Pinetti, U. M. Pagnoni, R. Rinaldi, C. Zucchi, L. Caglioti, G. Palyi. Volatile components of Grana-Parmiggiano-Reggiano type hard cheese. *Food Chem.* **2003**, *83*, 55.

[25] J. H. Lee, R. Diono, G.-Y. Kim, D. B. Min. Optimization of solid phase microextraction analysis for the headspace volatile compound of Parmesan cheese. *J. Agric. Food Chem.* **2003**, *51*, 1136.

[26] M. Qian, G. Reineccius. Identification of aroma compounds in Parmiggiano-Reggiano cheese by Gas Chromatography/Olfactometry. *J. Dairy Sci.* **2002**, *85*, 1362.

[27] P. F. Fox, J. Law, P. L. H. McSweeney, J. Wallace. Biochemistry of cheese ripening. In *Cheese: Chemistry, Physics and Microbiology*, vol. 1, 2, 2nd ed, P. F. Fox (Ed). Chapman & Hall: London, UK, **1995**, pp. 389–438.

[28] W. Lindinger, A. Hansel, A. Jordan. On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (PTR-MS) medical applications, food control and environmental research. *Int. J. Mass Spectrom. Ion Process.* **1998**, *173*, 191.

[29] J. Zhao, R. Zhang. Proton transfer reaction rate constants between hydronium ion (H3O+) and volatile organic compounds. *Atmos. Environ.* **2004**, *38*, 2177.

[30] National centre for supercomputing applications. *NCSA HDF specification and developer's guide*. Version 3.2. University of Illinois at Urbana: Champaign, **1993**.

[31] L. Cappelin, F. Biasioli, A. Fabris, E. Scuhfried, C. Soukoulis, T. D. Märk, F. Gasperi. Improved mass accuracy in PTR-TOF-MS: Another step towards better compound identification in PTR-MS. *Int. J. Mass Spectrom.* **2010**, *209*, 60.

[32] V. Vapnik. *The Nature of Statistical Learning Theory*, Springer: New York, **1999**.

[33] S. Wold, M. Sjöström, L. Eriksson. PLS-regression: a basic tool of chemometrics. *Chemometr. Intell. Lab.* **2001**, *58*, 109.

[34] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria, **2009**. ISBN 3-900051-07-0, http://www.R-project.org.

[35] R. Popper, H. Heymann. Analyzing differences among products and panelists by multidimensional scaling. In *Multivariate Analysis of Data in Sensory Science*, T. Naes and E. Risvik (Eds). Elsevier Science: Amsterdam, **1996**, pp. 159, 184.

[36] P. M. Granitto, F. Gasperi, F. Biasioli, E. Trainotti, C. Furlanello. Modern data mining tools in descriptive sensory analysis: a case study with a Random forest approach. *Food Qual. Preference* **2007**, *18*, 681.

[37] E. Aprea, F. Biasioli, G. Sani, C. Cantini, T. D. Märk, F. Gasperi. Proton transfer reaction-mass spectrometry (PTR-MS) headspace analysis for rapid detection of oxidative alteration of olive oil. *J. Agric. Food Chem.* **2006**, *54*, 7635.

[38] H. Burbank, M. C. Qian. Development of volatile sulfur compounds in heat-shocked and pasteurized milk cheese. *Int. Dairy J.* **2008**, *18*, 811.

[39] S. Landaud, S. Helinck, P. Bonnarme. Formation of volatile sulfur compounds and metabolism of methionine and other sulfur compounds in fermented food. *Appl. Microbiol. Biotechnol.* **2008**, *77*, 1191.

[40] E. Aprea, F. Biasioli, T. D. Märk, F. Gasperi. PTR-MS study of esters in water and water/ethanol solutions: fragmentation patterns and partition coefficients. *Int. J. Mass Spectrom.* **2007**, *262*, 114.

[41] E. Fonchy, A. Morin, N. Rodrigue, B. Müller, P. Chalier. Effect of growth temperature on hydrolytic and esterifying activities from Pseudomonas fragi CRDA 037 grown on whey. *Appl. Environ. Microbiol.* **1999**, *65*, 3114.

[42] Y. Raymond, A. Morin, F. Cormier, C. P. Champagne, H. Dubeau. Physical factors influencing the production of strawberry aroma by Pseudomonas fragi grown in skim milk. *Biotechnol. Lett.* **1990**, *12*, 931.

[43] E. Aprea, F. Biasioli, S. Carlin, G. Versini, T. D. Märk, F. Gasperi. Rapid white truffle headspace analysis by proton transfer reaction mass spectrometry and comparison with solid-phase microextraction coupled with gas chromatography/mass spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 2564.

[44] J. Adda, J. C. Gripon, L. Vassal. The chemistry of flavour and texture generation in cheese. *Food Chem.* **1982**, *9*, 115.

[45] P. Molimard, H. E. Spinnler. Review: compounds involved in the flavour of surface mold-ripened cheese: origins and properties. *J. Dairy Sci.* **1996**, *79*, 169.

[46] C. S. Shu. Pyrazine formation from amino acids and reducing sugars, a pathway other than Strecker degradation. *J. Agric. Food Chem.* **1998**, *46*, 1515.

[47] D. Forss. Review of the progress of dairy science: mechanism of formation of aroma compounds in milk and milk products. *J. Dairy Res.* **1979**, *46*, 691.

[48] J. W. Aston, J. R. Dulley. Cheddar cheese flavour. *Austalian J. Dairy Technol.* **1982**, *37*, 59.

[49] R. Liardon, J. O. Bosset, B. Blanc. The aroma composition of Swiss Gruyère cheese. The alkaline volatile components. *Lebensm. Wiss. Technol.* **1982**, *15*, 143.

[50] R. Griffith, E. G. Hammond. Generation of Swiss cheese flavour components by reaction of amino acids with carbonyl components. *J Dairy Sci.* **1989**, *72*, 604.

[51] E. Meinhart, P. Schreier. Study of flavour compounds from Parmiggiano Reggiano cheese. *Milchwissenschaft* **1986**, *41*, 689.