Towards smart on-line coffee roasting process control: Feasibility of real-time prediction of coffee roast degree and brew antioxidant capacity by single-photon ionization mass spectrometric (SPI-TOFMS) monitoring of roast gases

Journal:	*Journal of Agricultural and Food Chemistry*
Manuscript ID	jf-2019-06502d.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Heide, Jan; Universität Rostock Mathematisch-Naturwissenschaftliche Fakultät, Analytische und Technische Chemie Czech, Hendryk; Universitat Rostock Mathematisch-Naturwissenschaftliche Fakultat, Analytical and Technical Chemistry; Helmholtz Zentrum Munchen Deutsches Forschungszentrum fur Umwelt und Gesundheit, Comprehensive Molecular Analytics (CMA) Ehlert, Sven; Universitat Rostock, Analytical and Technical Chemistry Koziorowski, Thomas; PROBAT-Werke von Gimborn Maschinenfabrik GmbH Zimmermann, Ralf; Universitat Rostock, Chair of Analytical Chemistry ; Helmholtz Zentrum Munchen Deutsches Forschungszentrum fur Umwelt und Gesundheit

Towards smart on-line coffee roasting process control: Feasibility of real-time prediction of coffee roast degree and brew antioxidant capacity by single-photon ionization mass spectrometric (SPI-TOFMS) monitoring of roast gases

Jan Heide[†], Hendryk Czech[†,‡,], Sven Ehlert[§,⊥], Thomas Koziorowski[‖], Ralf Zimmermann[†,‡,⊥]*

[†]Joint Mass Spectrometry Centre, Chair of Analytical Chemistry, Institute of Chemistry, University of Rostock, Dr.-Lorenz-Weg 2, 18059 Rostock, Germany

[‡]Joint Mass Spectrometry Centre, Cooperation Group "Comprehensive Molecular Analytics", Helmholtz Zentrum München - German Research Center for Environmental Health GmbH, Gmunder Str. 37, 81379 München, Germany

[§]Photonion GmbH, Hagenower Str. 73, 19061 Schwerin, Germany

[⊥]Department Life, Light & Matter, University of Rostock, Albert-Einstein-Straße 25, 18059 Rostock, Germany

[‖]PROBAT-Werke von Gimborn Maschinenfabrik GmbH, Reeser Str. 94, 46446 Emmerich am Rhein, Germany

Keywords: photoionization mass spectrometry, Colorette, FC assay, polyphenols, process control

Abstract

Precise controlling and monitoring the status of the coffee roasting process is essential for consistent product quality and optimization towards targeted coffee properties. In small-scale roasting experiments, the chemical composition of the roasting off-gas was analyzed by on-line single-photon ionization time-of-flight mass spectrometry (SPI-TOFMS) at 118 nm with 5 s time resolution. Subsequently, mass spectra at the drop of the coffee beans were combined with off-line measurements of roast degree, described by color value “Colorette”, and the antioxidant capacity, obtained from Folin-Ciocalteu (FC) assay, in an explanatory PLS regression model. While the roast degree gives an indication of the coffee flavor, antioxidants in brewed coffee are regarded as beneficial for human health.

Colorette and FC values could be derived from the SPI mass spectra with root-mean-square errors from Monte Carlo cross-validation of 6.0 and 139 mg GA-eq. L^{-1}, respectively, and explained covariance (R^2_{CV}) better than 89%. Finally, the regression models were applied to the SPI mass spectra over the entire roast in order to demonstrate the predictive ability for on-line process control in real-time.

▪ Introduction

Coffee roasting is considered rather as art than science, which requires much experience of the roast master. Initial temperature of the roaster and the progression in roasting time from a bean sample and visual inspection are the key parameters to identify the roast status of the beans. Beyond that, a couple of measurements can be conducted after dropping of the roasted beans, including bean color, acidity and taste of the cupped coffee or humidity and loss in weight of the beans. Moreover, direct off-line mass spectrometry from headspace by means of nosespace analysis or with prior gas chromatographic (GC) separation [1–6] and near-infrared spectroscopy (NIR) of the beans [7–9] have been used in particular for general quality control, authentication of type and origin as well as identification of flavor components.

Besides the aspect of quality control, there is a growing market for food and beverages with health benefits. Coffee is known to inversely correlate with the prevalence of many diseases [10,11], which is partially associated with its high content of secondary plant metabolites, such as isoprenoid and phenolic compounds [12,13]. Among other antioxidants from Maillard-type reactions, in recent years polyphenols gained attention as valuable food ingredients [14,15]. Although the exact mode of action of polyphenols remains a subject of research, recent studies have found evidence for more complex mode of action and discarded the hypothesis of a direct antioxidative effect [12]. A small portion of the dietary polyphenols is taken up by the small intestine, whereas the majority enters the large intestine and affect the enzymatic activity of the gut microbiota, releasing bioavailable phenolic metabolites or even modify the gut microbiota composition [16]. Furthermore, it has been found that the time-scale for uptake depends on the overall composition of the phenolic intake, but not the maximum bacterial uptake concentration [17].

The content of phenolic compounds is affected by roasting conditions, but does not follow a steady trend [18]. Hence, tools for online monitoring coffee bean properties during roasting are desirable. However, sampling and subsequent offline measurement of these coffee properties with the aforementioned techniques takes too long for the roasting conditions to be modified during the roasting process for optimizing desired properties. Online techniques with high time

resolution, such as time-of-flight mass spectrometry (TOFMS) with single-photon ionization (SPI) [19–21], resonance-enhanced multi-photon ionization (REMPI) [22,23,21,19] and proton-transfer reaction (PTR) [24–26] as well as NIR [27–29] in principal allow monitoring of coffee roasting process in real-time and were previously used for fundamental investigations on the formation of volatile organic compounds during roasting and for the prediction of coffee properties. Moreover, quasi-online assays [18] and ultrafast gas chromatography (Fast-GC) mass spectrometry were applied to study coffee bean and nut roasting [30] as a compromise between time and chemical resolution.

In the present study, the focus is set on SPI-TOFMS analysis of the roasting off-gas in order to derive the roast degree of the coffee beans by means of color value “Colorette” and antioxidant capacity from Folin-Ciocalteu (FC) assay. An explanatory model based on projection on latent structure (PLS) regression is provided to demonstrate the predictive ability of SPI mass spectra and its application in real-time process monitoring.

▪ Materials and methods

Coffee roasting. Green Arabica coffee beans (*Coffea Arabica*) from Colombia were supplied by PROBAT. The roasting was conducted with an electrically heated single drum roaster (*PRE 1Z*, PROBAT-Werke von Gimborn Maschinenfabrik GmbH, Emmerich am Rhein, Germany) of approximately 100 g batch size. The roaster is equipped with a temperature readout for bean pile temperature in the inner drum. Additionally, the outer drum wall temperature was tracked by an infrared temperature sensor (Figure 1a). In particular, the temperature during the filling of the coffee beans has been shown to be crucial for roasting reproducibility.

Each of the 20 roast experiments was started at an outer drum temperature of (180±2)°C which was achieved by setting the power consumption of the roaster to 480 W (±20 W). The roasting profile (Figure 1c) represents typical roasting conditions, which can be roughly divided into two stages: Firstly, a high energy input at the beginning of the roast removes water and starts caramelization and Maillard reactions in the beans (endothermic phase). After the first crack,

aroma develops by Maillard reactions and pyrolysis during the second exothermic roasting phase. To avoid too fast roasting and aroma development, the intake of thermal energy for roasting was reduced by opening the ventilation damper to produce a higher air flow, which resulted in a small drop in bean pile temperature. Furthermore, the generally high temperature during the last stage of roasting can be explained by the position of the thermocouple. Eventually, 20 roasts were conducted, covering different drop temperatures with roasting times between approximately 7 and 14 min (Table 1).

Colorette value. The roast degree of the ground coffee beans (*Sette 270*, Baratza LLC, Seattle, WA, USA; grinding degree: 12H) was described by Colorette color values (*Colorette 3b*, PROBAT-Werke von Gimborn Maschinenfabrik GmbH, Emmerich am Rhein, Germany), which refer to a reflectance measurement using red and near-infrared light. The dimensionless Colorette scale ranges from 0 to 200 with values decreasing towards darker roasts. Colorette values between approximately 150 and 60, but also down to 35, denote coffee appropriate for the market.

Folin-Ciocalteu (FC) assay. The Folin-Ciocalteu (FC) assay was used as an antioxidant measurement, which is in particular sensitive to (poly)phenolic compounds, and results are expressed as equivalent of gallic acid (GA-eq.)[31]. First, 200 ml (190.5±0.5) g of hot water (82±1) °C were poured over 12 g of the ground coffee beans, and remained for 2 min in a French press. Subsequently, the brewed coffee was filtered through filter paper with a pore size smaller than 2 µm. The filtrate was diluted by a factor of 50, set to pH of approximately 10 by adding sodium carbonate (anhydrous sodium carbonate, purity > 99%, Fluka Chemie GmbH, Buchs, Switzerland) and mixed with the FC reagent containing phosphomolybdate and phosphotungstate (Merck KGaA, Darmstadt, Germany). The resulting blue complex from the reaction of the FC reagent was finally analyzed with a photometer at a wavelength of 765 nm (Hach DR 3900, resolution of 1 nm, Düsseldorf, Germany). Quantitation was performed by external calibration with anhydrous gallic acid (purity > 98%; Merck KGaA, Darmstadt, Germany) and deionized water (electrical conductivity < 1µS cm^{-1}) with a linear response in

the range of 0.34 µg L^{-1} to 8.5 µg L^{-1} (r^2=0.999; residual standard deviation corresponds to 106 GA-eq mg L^{-1} in the final cup, Figure S1).

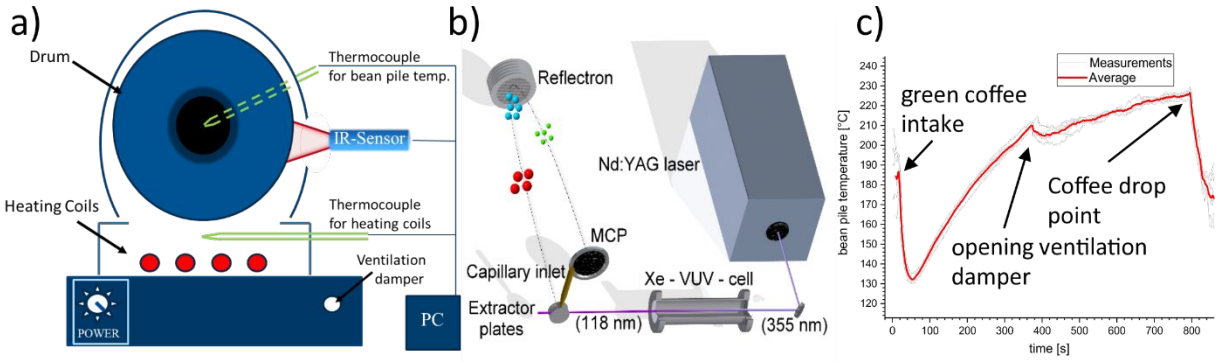


Figure 1. a) Roast experiments for batch sizes of 100 g were conducted with a coffee drum roaster, which was electrically heated and equipped with thermocouple to determine the bean pile temperature. b) The instrumental setup consists of an Nd:YAG laser and non-linear optics to produce 118 nm VUV-radiation for single-photon ionization (SPI) as well as a reflectron time of flight mass spectrometer (TOFMS), which allows monitoring of the roasting off-gas composition down to subsecond time resolution. c) The bean pile temperature shows a typical profile for drum roasters, including a temperature drop after filling and rebound. The second smaller temperature drop is caused by increased air flow from opening of the damper.

Table 1 Number of roasts, mean values and one standard deviation of FC assay and Colorette color values for four different descriptive roast degrees

Descriptive roast degree	Number of roasts	FC value [GA-eq mg L^{-1}]	Colorette value
Light	3	3410±20	148±8
Medium-light	4	3750±70	112±6
Medium-dark	4	3270±80	86±4
Dark	9	2690±130	59±7

Photoionization time-of-flight mass spectrometry. The roasting off-gas was analyzed by single-photon ionization time-of-flight mass spectrometry (SPI-TOFMS, Figure 1b; custom-made by Firma Stefan Kaesdorf - Geräte für Forschung und Industrie, Germany; 1,800 mass resolution at m/z 92; mass range from 1 to 513 with applied settings), which has been described in detail elsewhere [32]. Briefly, the energy of the fundamental radiation of 1064 nm of a pulsed Nd:YAG-laser (*Surelite III*, Continuum, Santa Clara, CA, USA; 10 Hz repetition rate) is tripled by non-linear optical devices to 355 nm and subsequently tripled again in a gas cell filled with xenon to generate 118 nm for SPI. The energy of 118 nm photons (10.49 eV) is slightly above the ionization energy of the majority of organic compounds in the gas emitted during roasting. Due to the low amount of excess energy during ionization, mainly molecular ions and low amounts of fragment ions are formed [33].

Samples of roast gas were taken from the backside of the drum with a flow of 4 L min^{-1}. From this support flow, SPI-TOFMS was sampled through a methyl-deactivated fused silica capillary with an inner diameter of 200 µm. To prevent condensation of volatile and semi-volatile compounds, the entire sampling line was heated to 250 °C.

Data pretreatment and PLS regression model. First, mass spectra from m/z 1 to 350 of the last five seconds before dropping the coffee beans were extracted from the online data and averaged. Subsequently, redundant m/z which are not possible as molecular signals due to the ionization selectivity (m/z from 1 to 16, 18 to 29, 31 to 33 and 35 to 39) as well as m/z for caffeine (m/z 193 to 197) were excluded from the mass spectra. Accordingly, 309 non-zero variables were used for the PLS regression. Moreover, only patterns of the roast gas composition were considered by using the L_1-norm (normalization to total peak intensity) of the mass spectra to eliminate the effect of concentration between the individual roasts.

To create PLS regression models, the mean values from triplicate analyses of Colorette and FC values were used as the dependent variables (predictors, y) and 20 mass spectra from individual roasting experiments were used as independent variables (descriptors, X). PLS regression, which was conducted using libPLS 1.95 toolbox [34] for MATLAB (version 2018; TheMathWorks Inc., MA, USA), refers to the generalization of multiple linear regression and

aims to describe the covariance between X and y [35]. Validation of the PLS model and selection of the optimal number of PLS components were performed by Monte Carlo cross-validation (1000 runs) for 1 to 20 PLS components. 80% of the sample data was randomly selected and used to generate a new model in order to explain the remaining 20% of the sample predictors. Subsequently, the number of PLS components was selected based on the lowest mean RMSE in 1000 Monte Carlo repetitions. In the last step, the PLS models were refined by the Competitive Adaptive Reweighted Sampling method (CARS) [36]. CARS eliminates variables by a survival-of-the-fittest algorithm leading to a possibly lower number of PLS components and root mean-squared errors (RMSE). The importance of the remaining variables for the model performance was examined from the covariance-based target projection (TP) loadings [37].

Due to the low number of roasting experiments, the data set was not partitioned into a calibration data set and a data set for external validation test set as recommended [38]. For that reason, errors and goodness-of-fit were taken from cross-validation, which is, however, not sufficient to assess the predictive ability of a model [39]. Consequently, the results presented in the following section originate from explanatory modelling and should be interpreted as the proof of feasibility for a predictive model rather than a final model [40].

▪ Results and discussion

Components related to roast degree and antioxidant capacity in the roaster gas. SPI-TOFMS enables monitoring of the temporal evolution of volatile and semi-volatile compounds in the roasting off-gas [21,20]. The assignments of chemical structures to detected m/z and their abundance during different stages of the roast was already discussed by Czech et al. (2016) [19]. Briefly, the most intense peaks originate from carbonyls, furans, pyrroles, phenols, fatty acids, as well as aliphatic and aromatic amines. They can be tentatively assigned to volatile reaction products, e.g. formyl-pyrrole (m/z 95), pyridine (m/z 79), methylbutanal (m/z 86), hydroxymethylfurfural (m/z 126), furfuryl alcohol (m/z 98) and vinylguaiacol (m/z 150) from the Maillard reaction, caramelization reactions or decomposition of chlorogenic acids [41,42]. In additional to chemical reactions, some compounds, such as caffeine (m/z 194), evaporate and

appear in the spectra during the entire roast. Even if chemical compounds cannot be identified unambiguously, the chemical information from the mass spectra goes beyond fingerprinting and reduces the number of possible molecular assignments per nominal m/z and the softness of the ionization turns SPI-TOFMS into a chemical sensor for roast gas analysis.

Towards the end of the roast, i.e. higher roast degree, the overall peak intensity increases. However, the pattern of the spectra changes (Figure 2), which is exploited in PLS regression presented in the following section.

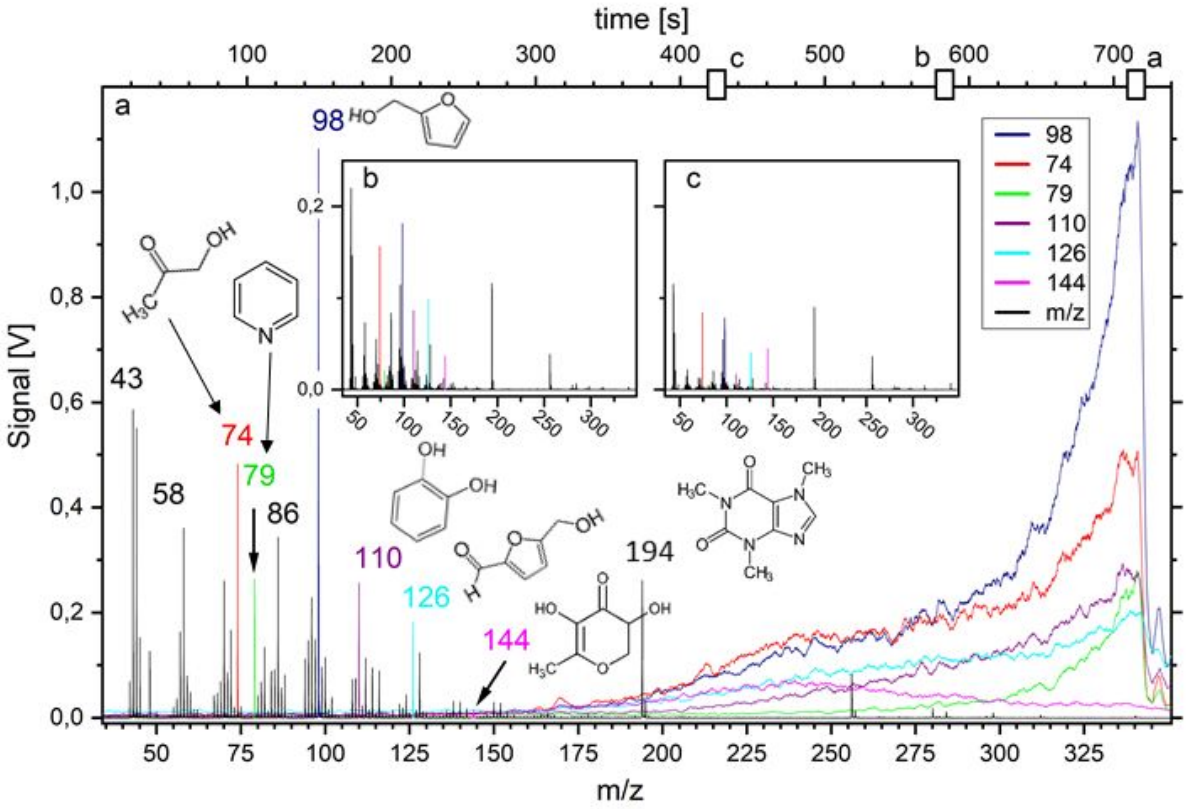


Figure 2. Combined illustration of mass spectra at different points of time (a-c) during roasting and temporal evolution of different m/z (colored). While some m/z, such as 144 (2,3-dihydro-3,5-dihydroxy-6-methyl-4H-pyran-4-one), peak during roasting, others, such as m/z 79 (pyridine), which is known as marker for overroasting, show sharp increases with roasting time. At 300 s, concentrations of VOC generally increase with ongoing roasting time, but with changing VOC pattern, which is exploited for the PLS regression model.

PLS regression parameters. PLS regression aims for the best fit between the descriptor and predictor matrices. If the number of latent variables (PLS components) is consecutively increased, the goodness of fit increases as well. However, the regression coefficients may not work with the same quality for samples not included in the regression, so generalization of the solution is not allowed. On that account, the RMSE was evaluated by Monte Carlo cross-validation, which randomly selects 80% of the sample descriptor (SPI mass spectra) and creates another regression model to calculate predictors (FC and Colorette values) from the remaining descriptors. This procedure was repeated 1000 times for different numbers of PLS components from 1 to 20 to obtain the RMSE from cross-validation ($RMSE_{CV}$), giving a more robust estimate than RMSE from the initial regression ($RMSE_{fit}$). Moreover, $RMSE_{CV}$ and explained variance (R^2_{CV}) reveal minima and maxima at three PLS components, respectively, for both explanation of FC and Colorette values, which is regarded as the optimal number of PLS components (Figure S2). With 0.62 for FC and 0.79 for Colorette values, R^2_{CV} has acceptable values, but the large discrepancy between R^2_{CV} and R^2_{fit} indicates overfitting of the model. Additionally, the $RMSE_{CV}$ of 263 GA-eq. mg L^{-1} for FC and 15 for Colorette values are still substantially higher than the precisions from the direct measurements of FC (±106 GA-eq mg L^{-1}) and Colorette (±1). Hence, the models were refined by applying competitive adaptive reweighted sampling (CARS), which reduces the number of variables, possibly PLS components and $RMSE_{CV}$ as well as increases accordingly R^2_{CV} [36]. The CARS approach was applied with 100 repetitions to investigate the model performance if only subsets of variables are considered. As expected, the number of variables was substantially reduced from 309 to 28 (for FC values) and 16 (for Colorette values), respectively. Accordingly, R^2_{CV} increased and $RMSE_{CV}$ declined close to the values of the PLS regression (R^2_{fit} and $RMSE_{fit}$) (Table 2), so overfitting was minimized. The final regression coefficients give good agreements between measured and calculated values (Figure 3) and led to $RMSE_{CV}$ of 139 GA-eq mg L^{-1} (for FC) and 6.0 (for Colorette values), which is only 31% larger than the precision of FC assay and six times higher than the precision of the Colorette measurement. For 50% of the independent roasts (interquartile range), the relative deviation of the modelled values is less than 3% (FC

values) and 2% (Colorette values) from the measured one without apparent favoring of any roast degree (Figure S3). Based on a pseudo-univariate approach [43], limits of detection were estimated as 687 GA-eq. mg L^{-1} for FC and 17.0 for Colorette values, which likely represent rather upper limits due to higher signal uncertainty from MS than from FC assay and Colorette measurements.

Table 2 Goodness-of-fit from Monte Carlo cross-validation (R^2_{CV}) and PLS fitting (R^2_{fit}), RMSE, number of PLS components (PLScomp) and number of variables (#Var) for initial and final PLS regression models refined by CARS.

	R^2_{CV}	$RMSE_{CV}$ [GA-eq. mg L^{-1}]	PLScomp	#Var	R^2_{fit}	$RMSE_{fit}$
FC_{ini}	0.624	265	3	309	0.936	109
FC_{CARS}	0.895 (0.017)*	139 (11)*	3 (0)*	28 (7)*	0.946	100
$Colorette_{ini}$	0.790	14.9	3	309	0.983	4.3
$Colorette_{CARS}$	0.966 (0.004)*	6.0 (0.3)*	3 (0)*	16 (6)*	0.984	4.2

*numbers in brackets refer to standard deviation of the results from 100 repetitions

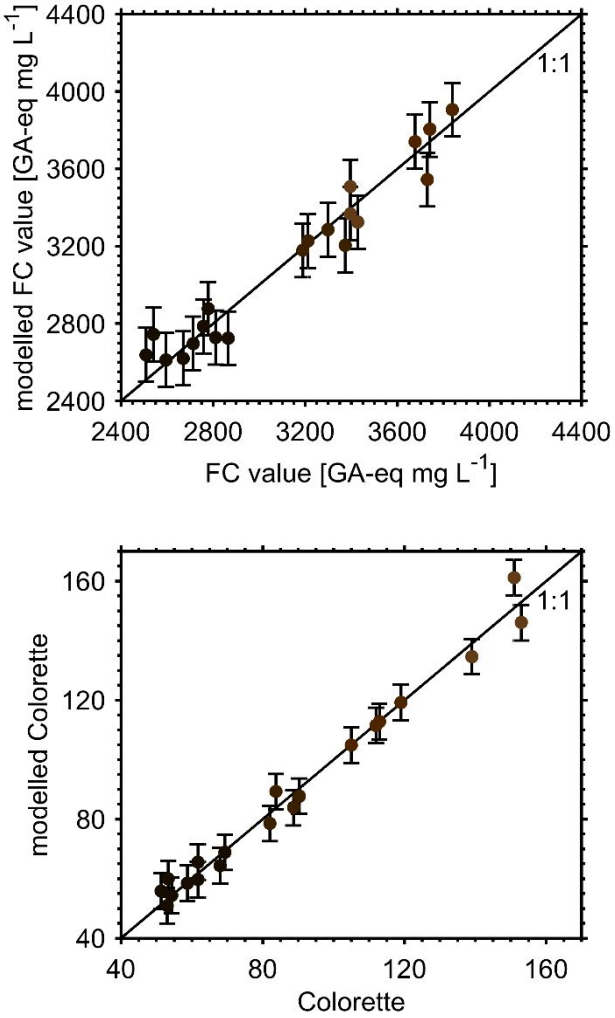


Figure 3. Measured vs calculated FC (top) and Colorette (bottom) values with RMSE from Monte Carlo cross-validation.

Important variables to link SPI mass spectra to FC and Colorette values. In the next step after modelling, the specific variable importance for the model performance was investigated from TP loadings. Both Colorette and FC values show an inverse relation with pyridine (m/z 79), which is a well-known marker for dark or overroast [44]. Apart from the light roasts with Colorette values between 140 and 160, Colorette and FC values exhibit a strong correlation (Figure S4). Therefore, both TP loadings qualitatively contain almost the same variables with differences only in intensity. Most important variables are m/z 126 and m/z 144, which belong

to hydroxymethylfurfural and 2,3-dihydro-3,5-dihydroxy-6-methyl-4H-pyran-4-one, respectively. The latter one has been identified as a strong antioxidant from the Maillard reaction [45] while the first one is also a Maillard reaction product, but without antioxidant activity [46]. For further peaks with positive TP loadings at m/z 97 (2,4-dimethyloxazole), m/z 74 (e.g. hydroxyacetone) and m/z 102 (e.g. 3-methyl-2-buten-1-thiol), no antioxidative effect is known to the best of our knowledge. In the negative TP loadings, the phenolic species (m/z 150 vinylguaiacol, m/z 152 vanillin, m/z 110 benzenediol) originate from the degradation of chlorogenic acids. Chlorogenic acids and other polyphenolic species substantially contribute to the antioxidant capacity of coffee [47] and are recommended micronutritients. However, not the entire class of polyphenols has been associated with positive effects for human health [14] and it should be noted that the brewing alters the chlorogenic acids by water addition and associated molecular rearrangements [48]. The detected phenolic species have a relatively low degree of substituents on the aromatic ring, indicating that they are products from multi-step thermal degradation of chlorogenic acids. Moreover, with longer roasting times, Maillard reaction products replace chlorogenic acids as the prevailing antioxidant in roasted coffee [49]. Hence, negative TP loadings of phenolic species are reasonably linked to lower FC values, and the sharp decreases of FC values with darker roast is explained by the high sensitivity of the FC for (poly)phenolic species [31].

Similar chemical compounds in both FC and Colorette TP loadings indicate that degradation of polyphenols and darkening of coffee beans, i.e. decrease in Colorette value, by reactions of Maillard-type or caramelization appear at the same time. However, the prediction of FC values seems to be more complex than roast degree because more variables are needed for this model than for the Colorette model.

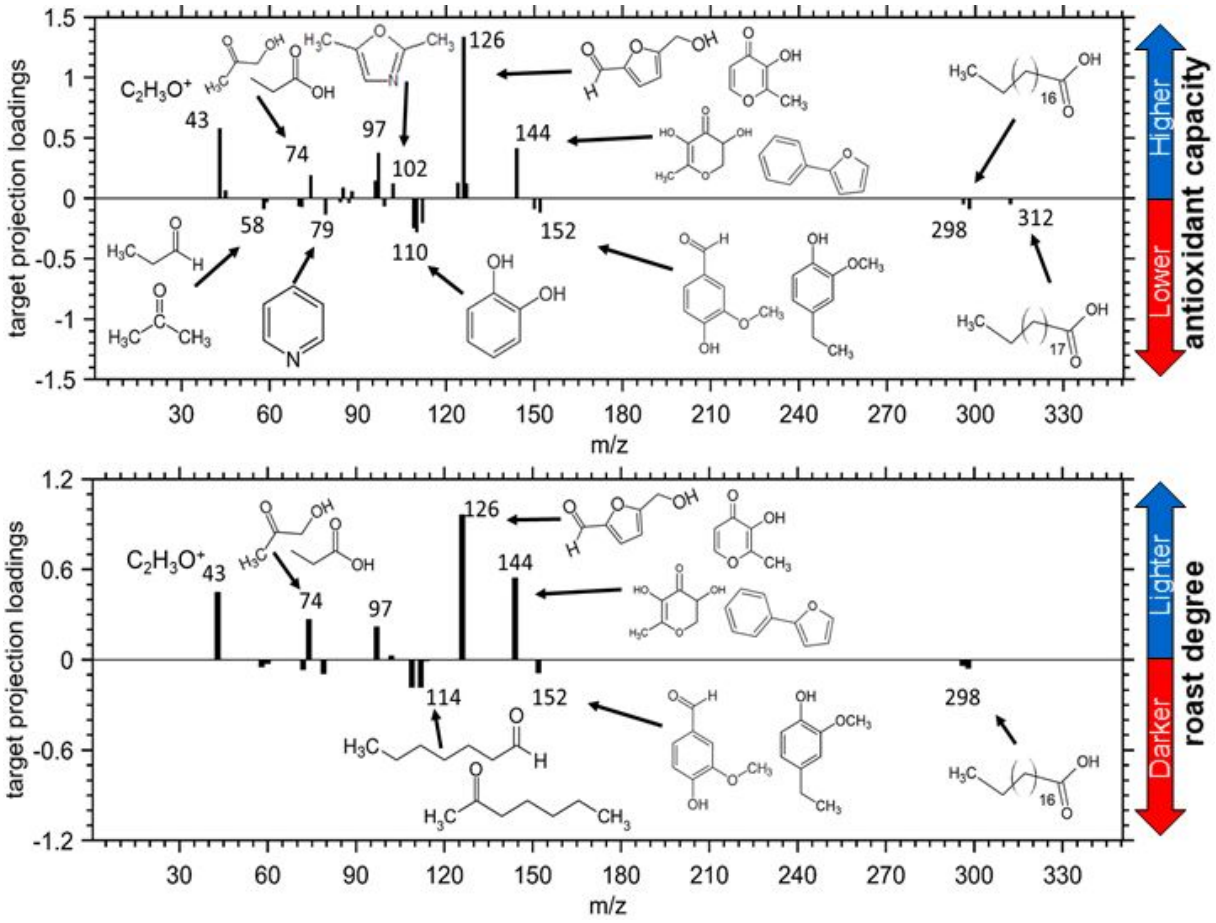


Figure 4. TP loadings, representing variable importance, in FC value (top) and bean color (Colorette) models (bottom) with tentative chemical assignments to m/z

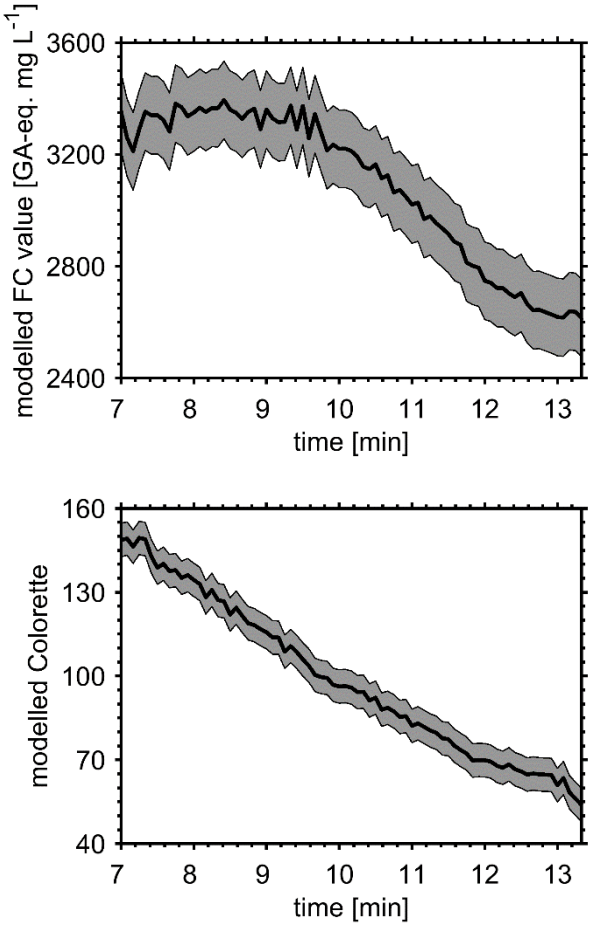
Towards on-line prediction of coffee roast degree and antioxidant capacity in real-time. With bean drops at different temperatures, the models can be applied to real-time analysis of 5 s time resolution within the range of FC and Colorette values used for PLS regression (Figure 5), corresponding to roasting times approximately between 7 and 14 min. For process control in coffee roasting, it is essential to have sufficient time resolution of the measurements because FC and Colorette values in the present experiments change by approximately 300 GA-eq. mg $L^{-1}$and 16 per minute, respectively, which can be gathered by SPI-TOFMS. Regarding Colorette values, it should be noted that trained panelists for sensory evaluation are still able to distinguish coffees having at least a difference in Colorette values of ±3 color values, which is for this roaster and roasting profile equal to a difference in roasting time of

22 s. Thus, from the perspective of time resolution, the same approach could be used for faster roasting profiles, i.e. roasting profiles with higher temperature, and associated faster changes in Colorette value per time.

Also regarding antioxidants, the temporal evolution of the FC values during roasting can be illustrated, including a slight increase between 7 and 8 min followed by apparent decrease after 9 min 30 s (Figure 5, left). However, FC values do not decline as distinctly as Colorette values and have a lower ratio of FC range to $RMSE_{CV}$. Considering the $RMSE_{CV}$ of 139 GA-eq mg L^{-1}, coffee beans differing in 28 s of roasting time can be still distinguished if roasted between 10 min and 12 min 30 s (with a linear dependency between time and FC value) under the presented roasting profile and roaster.

The PLS regression model is only applicable for the presented roast experiments here. In addition to model validation using an external data set and generally higher number of experiments, it would be interesting to explore the applicability of this concept for other coffee bean types, moisture and batch variations as well as further types of roasters and roast conditions, such as temperature, pressure or chemical pretreatment.

For further improvement, the high sensitivity of the FC assay for (poly)phenolic species can be approached by the analysis techniques using laser-based resonance-enhanced multi-photon ionization (REMPI) TOFMS, which is a selective ionization technique for aromatic compounds [50]. Wavelengths of typical industry lasers of 266 nm (fourth harmonic generation of Nd:YAG laser) and 248 nm (KrF excimer laser) enable identification and monitoring of roasting stages [19]. Moreover, optical parametric oscillators, converting light from a pump laser into wavelengths in the UV-range, give control of analyte selectivity even beyond aromatic compounds [19], which might also be of interest for the monitoring and prediction of other coffee properties. Therefore, REMPI might enhance the strength of the correlation between FC values and mass spectra compared to SPI-TOFMS. Finally, assays for total antioxidant capacity could be substituted by more specific analyses of antioxidant molecules, e.g. by high performance thin-layer chromatography mass spectrometry combined with bioassays [51], in order to target specific molecules associated with health benefits.



320

321 **Figure 5.** Real-time prediction of FC (top) and Colorette values (bottom) with absolute errors

322 from $RMSE_{CV}$

323

324 ▪ **Author Information**

325 **Corresponding Author**

326 *Phone: +49 (0)381 498 6533. Fax: +49 (0) 381 498−118 6527

327 E-mail: hendryk.czech@uni-rostock.de

328 **ORCID**

329 Hendryk Czech: 0000-0001-8377-4252

330 Ralf Zimmermann: 0000-0002-6280-3218

331 **Funding**

This study was funded by the German Federal Ministry for Education and Research within the project „Prozessanalyse und -steuerung der industriellen Röstung von Lebens- und Genussmitteln mittels Photoionisationsmassenspektrometrie am Beispiel von Kaffee (PPK)“ (grant 02P16K622)

Notes

The authors declare no competing financial interest.

- **Associated Content**

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website: Calibration function for FC assay, Colorette vs FC value, RMSE and Q^2 in Monte Carlo cross validation, relative deviation between measured and modelled FC and Colorette values.

References

(1) Yeretzian, C.; Jordan, A.; Lindinger, W. Analysing the headspace of coffee by proton-transfer-reaction mass-spectrometry, *Int. J. Mass Spectrom.* **2003**, *223-224,* pp. 115–139.

(2) Liberto, E.; Ruosi, M. R.; Cordero, C.; Rubiolo, P.; Bicchi, C.; Sgorbini, B. Non-separative headspace solid phase microextraction-mass spectrometry profile as a marker to monitor coffee roasting degree, *J. Agric. Food Chem.* **2013**, *61,* pp. 1652–1660.

(3) Ruosi, M. R.; Cordero, C.; Cagliero, C.; Rubiolo, P.; Bicchi, C.; Sgorbini, B.; Liberto, E. A further tool to monitor the coffee roasting process: aroma composition and chemical indices, *J. Agric. Food Chem.* **2012**, *60,* pp. 11283–11291.

(4) Ribeiro, J. S.; Augusto, F.; Salva, T. J. G.; Thomaziello, R. A.; Ferreira, M. M. C. Prediction of sensory properties of Brazilian Arabica roasted coffees by headspace solid phase microextraction-gas chromatography and partial least squares, *Anal. Chim. Acta.* **2009**, *634,* pp. 172–179.

(5) Romano, A.; Cappellin, L.; Ting, V.; Aprea, E.; Navarini, L.; Gasperi, F.; Biasioli, F. Nosespace analysis by PTR-ToF-MS for the characterization of food and tasters. The case study of coffee, *Int. J. Mass Spectrom.* **2014**, *365-366,* pp. 20–27.

(6) Bressanello, D.; Liberto, E.; Cordero, C.; Rubiolo, P.; Pellegrino, G.; Ruosi, M. R.; Bicchi, C. Coffee aroma: Chemometric comparison of the chemical information provided by three different samplings combined with GC-MS to describe the sensory properties in cup, *Food Chem.* **2017**, *214,* pp. 218–226.

(7) Alessandrini, L.; Romani, S.; Pinnavaia, G.; Dalla Rosa, M. Near infrared spectroscopy: an analytical tool to predict coffee roasting degree, *Anal. Chim. Acta.* **2008**, *625,* pp. 95–102.

(8) Tugnolo, A.; Beghi, R.; Giovenzana, V.; Guidetti, R. Characterization of green, roasted beans, and ground coffee using near infrared spectroscopy: A comparison of two devices, *J. Near Infrared Spec.* **2019**, *27,* pp. 93–104.

(9) Barbin, D. F.; Felicio, A. L. d. S. M.; Sun, D.-W.; Nixdorf, S. L.; Hirooka, E. Y. Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview, *Food Res. Int.* **2014**, *61,* pp. 23–32.

(10) Gökcen, B. B.; Şanlier, N. Coffee consumption and disease correlations, *Crit. Rev. Food Sci. Nutr.* **2019**, *59,* pp. 336–348.

(11) Hu, G. L.; Wang, X.; Zhang, L.; Qiu, M. H. The sources and mechanisms of bioactive ingredients in coffee, *Food Funct.* **2019**, *10,* pp. 3113–3126.

(12) Ludwig, I. A.; Clifford, M. N.; Lean, M. E. J.; Ashihara, H.; Crozier, A. Coffee: biochemistry and potential impact on health, *Food Funct.* **2014**, *5,* pp. 1695–1717.

(13) Romualdo, G. R.; Rocha, A. B.; Vinken, M.; Cogliati, B.; Moreno, F. S.; Chaves, M. A. G.; Barbisan, L. F. Drinking for protection? Epidemiological and experimental evidence on the beneficial effects of coffee or major coffee compounds against gastrointestinal and liver carcinogenesis, *Food Res. Int.* **2019**, *123,* pp. 567–589.

(14) Del Bo', C.; Bernardi, S.; Marino, M.; Porrini, M.; Tucci, M.; Guglielmetti, S.; Cherubini, A.; Carrieri, B.; Kirkup, B.; Kroon, P.; Zamora-Ros, R.; Liberona, N. H.; Andres-Lacueva, C.; Riso, P. Systematic Review on Polyphenol Intake and Health Outcomes: Is there Sufficient Evidence to Define a Health-Promoting Polyphenol-Rich Dietary Pattern?, *Nutrients.* **2019**, *11*.

(15) Shahidi, F.; Ambigaipalan, P. Phenolics and polyphenolics in foods, beverages and spices: Antioxidant activity and health effects – A review, *J. Funct. Foods.* **2015**, *18,* pp. 820–897.

(16) Gollücke, A. P. B.; Peres, R. C.; Ribeiro, D. A.; Aguair, O. Polyphenols as Supplements in Foods and Beverages: Recent Discoveries and Health Benefits, an Update. In

Polyphenols: Mechanisms of Action in Human Health and Disease, 2nd ed.; Watson, R. R.; Preedy, V. R.; Zibadi, S., Eds.; Elsevier Academic Press, 2018, pp. 12–18.

(17) Hakem Said, I.; Gencer, S.; Ullrich, M. S.; Kuhnert, N. Tea and coffee time with bacteria – Investigation of uptake of key coffee and tea phenolics by wild type E. coli, *Food Res. Int.* **2018**, *108,* pp. 585–594.

(18) Smrke, S.; Opitz, S. E. W.; Vovk, I.; Yeretzian, C. How does roasting affect the antioxidants of a coffee brew? Exploring the antioxidant capacity of coffee via on-line antioxidant assays coupled with size exclusion chromatography, *Food Funct.* **2013**, *4,* pp. 1082–1092.

(19) Czech, H.; Schepler, C.; Klingbeil, S.; Ehlert, S.; Howell, J.; Zimmermann, R. Resolving Coffee Roasting-Degree Phases Based on the Analysis of Volatile Compounds in the Roasting Off-Gas by Photoionization Time-of-Flight Mass Spectrometry (PI-TOFMS) and Statistical Data Analysis: Toward a PI-TOFMS Roasting Model, *J. Agric. Food Chem.* **2016**, *64,* pp. 5223–5231.

(20) Hertz-Schünemann, R.; Streibel, T.; Ehlert, S.; Zimmermann, R. Looking into individual coffee beans during the roasting process: direct micro-probe sampling on-line photo-ionisation mass spectrometric analysis of coffee roasting gases, *Anal. Bioanal. Chem.* **2013**, *405,* pp. 7083–7096.

(21) Dorfner, R.; Ferge, T.; Yeretzian, C.; Kettrup, A.; Zimmermann, R. Laser mass spectrometry as on-line sensor for industrial process analysis: process control of coffee roasting, *Anal. Chem.* **2004**, *76,* pp. 1386–1402.

(22) Zimmermann, R.; Heger, H.-J.; Yeretzian, C.; Nagel, H.; Boesl, U. Application of Laser Ionization Mass Spectrometry for On-line Monitoring of Volatiles in the Headspace of Food Products: Roasting and Brewing of Coffee, *Rapid Commun. Mass Spectrom.* **1996**, *10,* pp. 1975–1979.

(23) Hertz-Schünemann, R.; Dorfner, R.; Yeretzian, C.; Streibel, T.; Zimmermann, R. On-line process monitoring of coffee roasting by resonant laser ionisation time-of-flight mass spectrometry: bridging the gap from industrial batch roasting to flavour formation inside an individual coffee bean, *J. Mass Spectrom.* **2013**, *48,* pp. 1253–1265.

(24) Gloess, A. N.; Yeretzian, C.; Knochenmuss, R.; Groessl, M. On-line analysis of coffee roasting with ion mobility spectrometry–mass spectrometry (IMS–MS), *Int. J. Mass Spectrom.* **2018**, *424,* pp. 49–57.

(25) Gloess, A. N.; Vietri, A.; Wieland, F.; Smrke, S.; Schönbächler, B.; López, J. A. S.; Petrozzi, S.; Bongers, S.; Koziorowski, T.; Yeretzian, C. Evidence of different flavour formation dynamics by roasting coffee from different origins. On-line analysis with PTR-ToF-MS, *Int. J. Mass Spectrom.* **2014**, *365-366,* pp. 324–337.

(26) Wieland, F.; Gloess, A. N.; Keller, M.; Wetzel, A.; Schenker, S.; Yeretzian, C. Online monitoring of coffee roasting by proton transfer reaction time-of-flight mass spectrometry (PTR-ToF-MS): towards a real-time process control for a consistent roast profile, *Anal. Bioanal. Chem.* **2012**, *402,* pp. 2531–2543.

(27) Catelani, T. A.; Páscoa, R. N. M. J.; Santos, J. R.; Pezza, L.; Pezza, H. R.; Lima, J. L. F. C.; Lopes, J. A. A Non-invasive Real-Time Methodology for the Quantification of Antioxidant Properties in Coffee During the Roasting Process Based on Near-Infrared Spectroscopy, *Food Bioprocess. Technol.* **2017**, *10,* pp. 630–638.

(28) Santos, J. R.; Lopo, M.; Rangel, A. O.S.S.; Lopes, J. A. Exploiting near infrared spectroscopy as an analytical tool for on-line monitoring of acidity during coffee roasting, *Food Control.* **2016**, *60,* pp. 408–415.

(29) Catelani, T. A.; Santos, J. R.; Páscoa, R. N. M. J.; Pezza, L.; Pezza, H. R.; Lopes, J. A. Real-time monitoring of a coffee roasting process with near infrared spectroscopy using multivariate statistical analysis: A feasibility study, *Talanta.* **2018**, *179,* pp. 292–299.

(30) Fischer, M.; Wohlfahrt, S.; Varga, J.; Matuschek, G.; Saraji-Bozorgzad, M. R.; Walte, A.; Denner, T.; Zimmermann, R. Evolution of Volatile Flavor Compounds During Roasting of Nut Seeds by Thermogravimetry Coupled to Fast-Cycling Optical Heating Gas Chromatography-Mass Spectrometry with Electron and Photoionization, *Food Anal. Methods.* **2017**, *10,* pp. 49–62.

(31) Opitz, S. E. W.; Smrke, S.; Goodman, B. A.; Keller, M.; Schenker, S.; Yeretzian, C. Antioxidant Generation during Coffee Roasting: A Comparison and Interpretation from Three Complementary Assays, *Foods.* **2014**, *3,* pp. 586–604.

(32) Mühlberger, F.; Hafner, K.; Kaesdorf, S.; Ferge, T.; Zimmermann, R. Comprehensive on-line characterization of complex gas mixtures by quasi-simultaneous resonance-enhanced multiphoton ionization, vacuum-UV single-photon ionization, and electron impact ionization in a time-of-flight mass spectrometer: setup and instrument characterization, *Anal. Chem.* **2004**, *76,* pp. 6753–6764.

(33) Hanley, L.; Zimmermann, R. Light and molecular ions: the emergence of vacuum UV single-photon ionization in MS, *Anal. Chem.* **2009**, *81,* pp. 4174–4182.

(34) Li, H.-D.; Xu, Q.-S.; Liang, Y.-Z. libPLS: An integrated library for partial least squares regression and linear discriminant analysis, *Chemometr. Intell. Lab.* **2018**, *176,* pp. 34–43.

(35) Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics, *Chemometr. Intell. Lab.* **2001**, *58,* pp. 109–130.

(36) Li, H.; Liang, Y.; Xu, Q.; Cao, D. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration, *Analytica Chimica Acta.* **2009**, *648,* pp. 77–84.

(37) Kvalheim, O. M.; Karstrang, T. V. Interpretation of Latent-Variable Regression Models, *Chemometr. Intell. Lab.* **1989**, *7,* pp. 39–51.

(38) Brereton, R. G. Consequences of sample size, variable selection, and model validation and optimisation, for predicting classification ability from analytical data, *TrAC.* **2006**, *25,* pp. 1103–1111.

(39) Golbraikh, A.; Tropsha, A. Beware of q2!, *J. Mol. Graph. Model.* **2002**, *20,* pp. 269–276.

(40) Shmueli, G. To Explain or to Predict?, *Statist. Sci.* **2010**, *25,* pp. 289–310.

(41) Hellwig, M.; Henle, T. Baking, ageing, diabetes: a short history of the Maillard reaction, *Angew. Chem. Int. Ed.* **2014**, *53,* pp. 10316–10329.

(42) Moon, J.-K.; Shibamoto, T. Formation of volatile chemicals from thermal degradation of less volatile coffee components: quinic acid, caffeic acid, and chlorogenic acid, *J. Agric. Food Chem.* **2010**, *58,* pp. 5465–5470.

(43) Allegrini, F.; Olivieri, A. C. IUPAC-consistent approach to the limit of detection in partial least-squares calibration, *Anal. Chem.* **2014**, *86,* pp. 7858–7866.

(44) Stadler, R. H.; Varga, N.; Hau, J.; Vera, F. A.; Welti, D. H. Alkylpyridiniums. 1. Formation in Model Systems via Thermal Degradation of Trigonelline, *J. Agric. Food Chem.* **2002**, *50,* pp. 1192–1199.

(45) Yu, X.; Zhao, M.; Liu, F.; Zeng, S.; Hu, J. Identification of 2,3-dihydro-3,5-dihydroxy-6-methyl-4H-pyran-4-one as a strong antioxidant in glucose–histidine Maillard reaction products, *Food Res. Int.* **2013**, *51,* pp. 397–403.

(46) Kanzler, C.; Haase, P. T.; Schestkowa, H.; Kroh, L. W. Antioxidant Properties of Heterocyclic Intermediates of the Maillard Reaction and Structurally Related Compounds, *J. Agric. Food Chem.* **2016**, *64,* pp. 7829–7837.

(47) Fujioka, K.; Shibamoto, T. Quantitation of volatiles and nonvolatile acids in an extract from coffee beverages: correlation with antioxidant activity, *J. Agric. Food Chem.* **2006**, *54,* pp. 6054–6058.

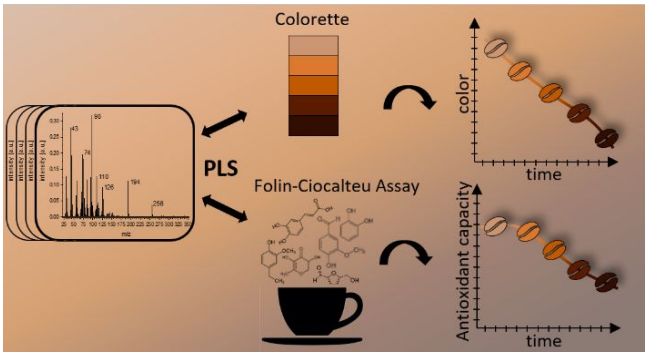
(48) Matei, M. F.; Jaiswal, R.; Kuhnert, N. Investigating the chemical changes of chlorogenic acids during coffee brewing: conjugate addition of water to the olefinic moiety of chlorogenic acids and their quinides, *J. Agric. Food Chem.* **2012**, *60,* pp. 12105–12115.

(49) Liu, Y.; Kitts, D. D. Confirmation that the Maillard reaction is the principle contributor to the antioxidant capacity of coffee brews, *Food Res. Int.* **2011**, *44,* pp. 2418–2424.

(50) Streibel, T.; Zimmermann, R. Resonance-enhanced multiphoton ionization mass spectrometry (REMPI-MS): applications for process analysis, *Annu. Rev. Anal. Chem.* **2014**, *7,* pp. 361–381.

(51) Taha, M. N.; Krawinkel, M. B.; Morlock, G. E. High-performance thin-layer chromatography linked with (bio)assays and mass spectrometry - a suited method for discovery and quantification of bioactive components? Exemplarily shown for turmeric and milk thistle extracts, *J. Chromatogr. A.* **2015**, *1394,* pp. 137–147.

TOC



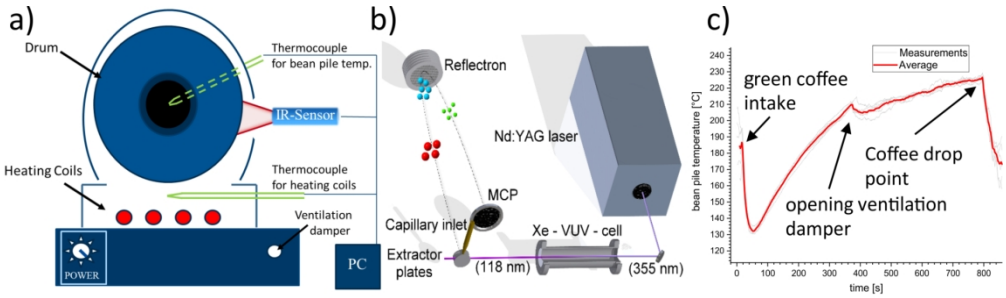


Figure 1. a) Roast experiments for batch sizes of 100 g were conducted with a coffee drum roaster, which was electrically heated and equipped with thermocouple to determine the bean pile temperature. b) The instrumental setup consists of an Nd:YAG laser and non-linear optics to produce 118 nm VUV-radiation for single-photon ionization (SPI) as well as a reflectron time of flight mass spectrometer (TOFMS), which allows monitoring of the roasting off-gas composition down to subsecond time resolution. c) The bean pile temperature shows a typical profile for drum roasters, including a temperature drop after filling and rebound. The second smaller temperature drop is caused by increased air flow from opening of the damper.

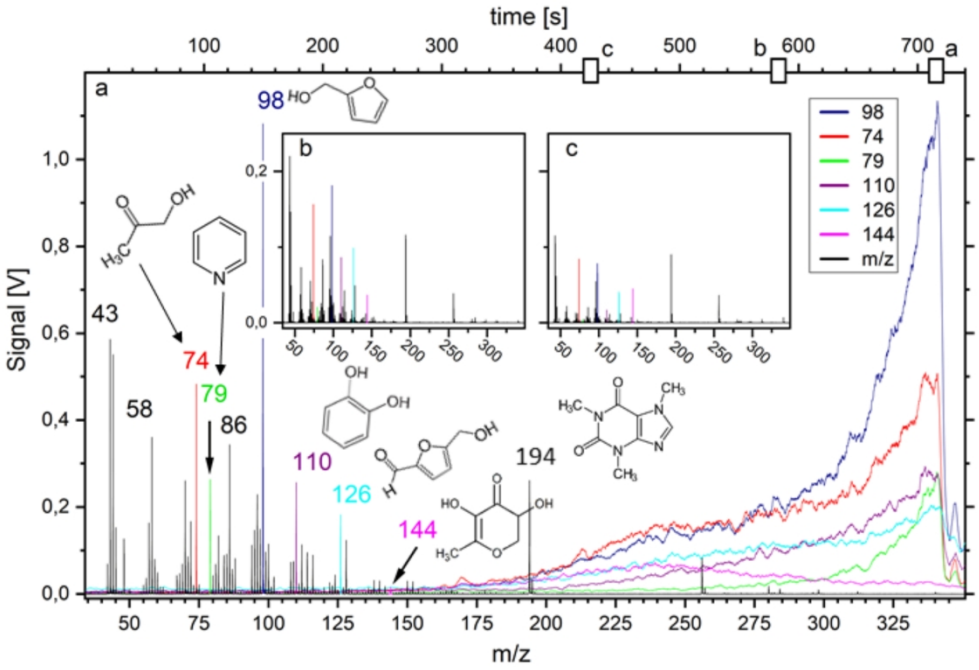


Figure 2. Combined illustration of mass spectra at different points of time (a-c) during roasting and temporal evolution of different m/z (colored). While some m/z, such as 144 (2,3-dihydro-3,5-dihydroxy-6-methyl-4H-pyran-4-one), peak during roasting, others, such as m/z 79 (pyridine), which is known as marker for overroasting, show sharp increases with roasting time. At 300 s, concentrations of VOC generally increase with ongoing roasting time, but with changing VOC pattern, which is exploited for the PLS regression model.

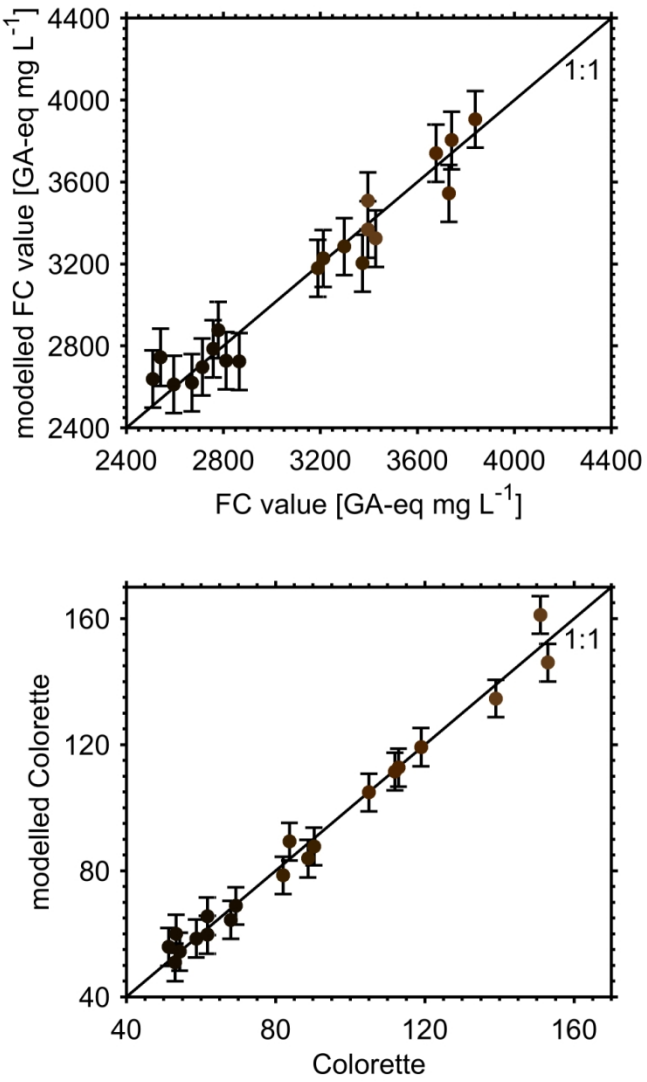


Figure 3. Measured vs calculated FC (top) and Colorette (bottom) values with RMSE from Monte Carlo cross-validation.

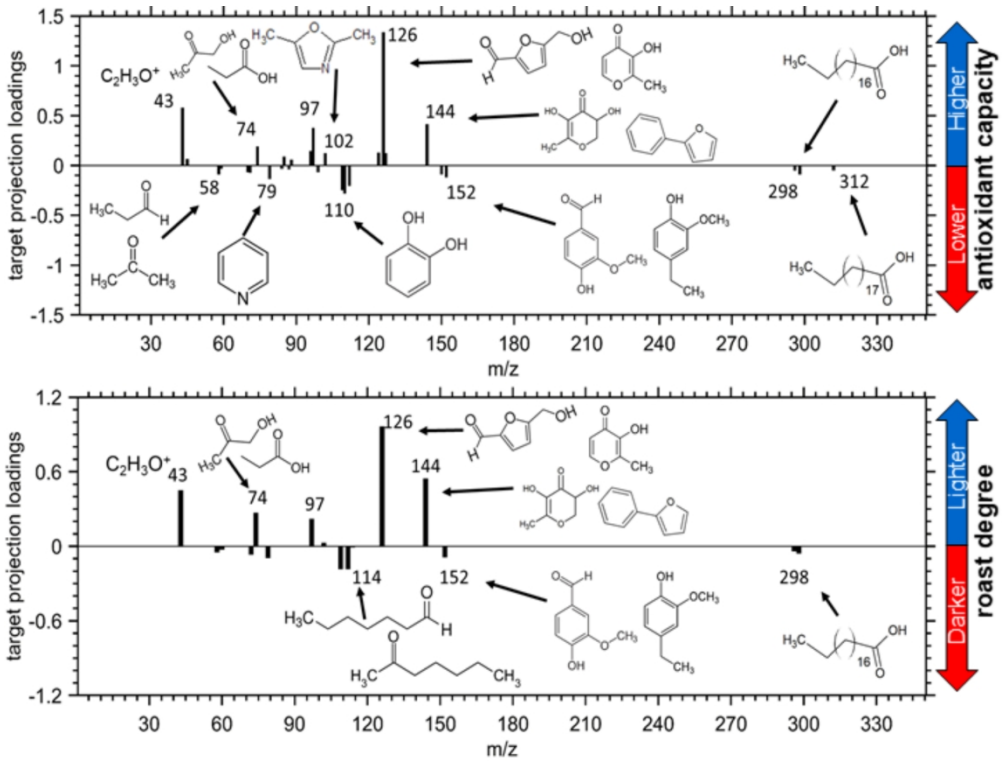


Figure 4. TP loadings, representing variable importance, in FC value (top) and bean color (Colorette) models (bottom) with tentative chemical assignments to m/z

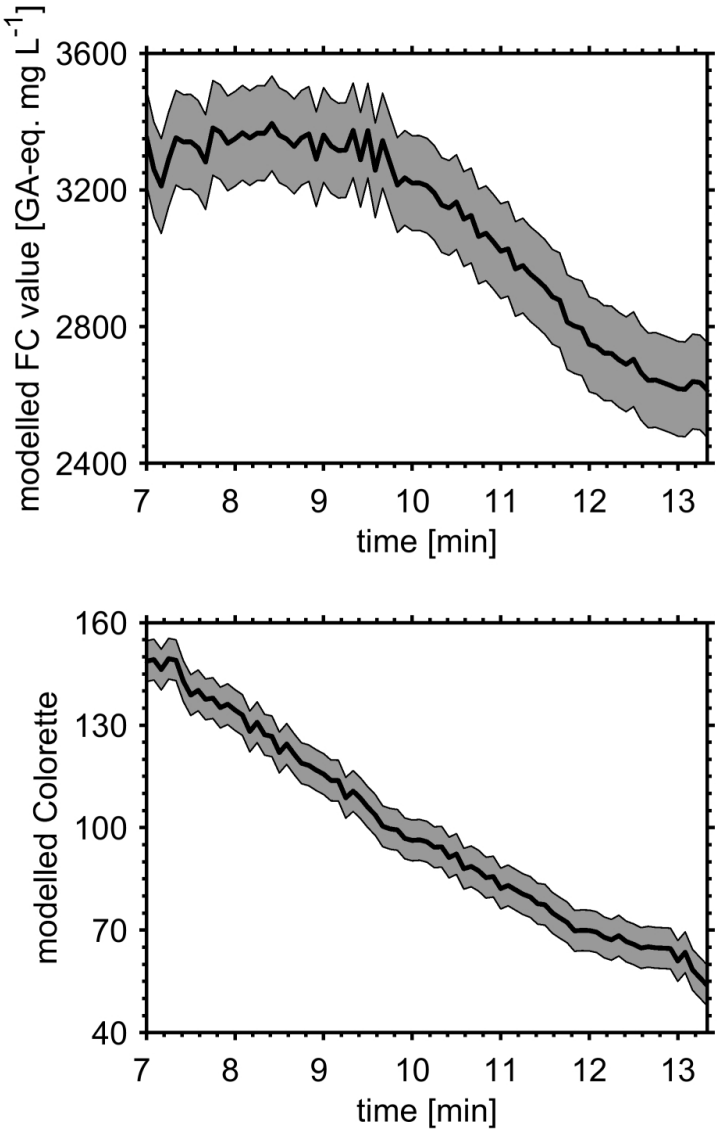


Figure 5. Real-time prediction of FC (top) and Colorette values (bottom) with absolute errors from RMSECV