

Operationally Realistic Validation for Prediction of Cocoa Sensory Qualities by High-Throughput Mass Spectrometry

Jacqueline E. Wood,[†] David Allaway,^{*,§} Emma Boulton,[‡] and Ian M. Scott^{*,†}

Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, SY23 3DA, U.K. and Mars Chocolate UK Ltd, Slough, SL1 4JX, U.K.

The potential of analytical chemistry to predict sensory qualities of food materials is a major current theme. Standard practice is cross-validation (CV), where a set of chemical and associated sensory data is partitioned so chemometric models can be developed on training subsets, and validated on held-out subsets. CV demonstrates prediction, but is an unlikely scenario for industrial operations, where concomitant data acquisition for model development and test materials would be unwieldy. We evaluated cocoa materials of diverse provenance, and analyzed on different dates to those used in model development. Liquor extracts were analyzed by flow-injection electrospray-mass spectrometry (FIE-MS), a novel method for sensory quality prediction. FIE-MS enabled prediction of sensory qualities described by trained human panelists. Optimal models came from the Weka data-mining algorithm *SimpleLinearRegression*, which learns a model for the attribute giving minimal training error, which was (–)-epicatechin. This flavonoid likewise dominated partial least-squares (PLS)-regression models. Refinements of PLS (orthogonal-PLS or orthogonal signal correction) gave poorer generalization to different test sets, as did support vector machines, whose hyperparameters could not be optimized in training to avoid overfitting. In conclusion, if chemometric overfitting is avoided, chemical analysis can predict sensory qualities of food materials under operationally realistic conditions.

Technological prediction of how agricultural products will be perceived by the senses is a long-held goal in analytical chemistry, growing in importance as globalized food production is increasingly consumer-centric.¹ Models for prediction of sensory qualities from chemical analyses of food and beverages appear at a rate of about 10 per year (Table 1). Reported accuracies are promising, and instrumental approaches to complement human sensory panels receive serious consideration.²

Literature review shows sensory quality prediction by analytical chemistry needs more operationally relevant validation. For reasons of sample limitation,³ “cross-validation” (CV) is the de facto standard in relevant food studies (Table 1). In each round of a CV procedure, data are partitioned into complementary subsets, on one of which (the training data), a chemometric model is constructed. Then the model’s predictive accuracy is validated on the held-out subset. In full CV, partitioning is iterated until all data have been held out once. More than half the papers in Table 1 perform leave-one-out (LOO) CV, where each sample is validated individually. More generally, *k*-fold CV refers to partition of the data into *k* subsets for *k* validation rounds.³ Performance over all partitions is often expressed as Q^2 , the fraction of total variation predicted.⁴ Caution has been urged in the use of Q^2 and CV, however, one danger being that with uncritical use of software, validation data may participate in model development^{5,6} (though this is not necessarily the case).

When there are enough samples available to create separate training and test sets, a procedure known as “external” validation can be employed, wherein the test set is independent from all model building processes.³ Among the 49 recent papers invoking prediction of sensory data, only six primarily used validation unequivocally identifiable as external, in this statistical sense of the term (Table 1).

The need to extend validation criteria in instrumental prediction of sensory qualities, however, goes beyond the statistics of CV. For real food production applications, the validation scenario in virtually all papers is too comfortable. No reports state clearly whether validation samples were analyzed on different dates to the training set, for example, even though potential instrument drift limits the adoption of some techniques for process monitoring.⁷ Laboratory workup brings further variability. An extra problem with complex foods is that they may not lend themselves to precise reproduction, or long-term storage, as reference standards. A realistically useful chemometric model of sensory profiles would be serviceable, without critical recalibration, for

* To whom correspondence should be addressed. E-mail: ias@aber.ac.uk.

[†] Aberystwyth University.

[‡] Mars UK Ltd.

[§] Present address: WALTHAM Centre for Pet Nutrition, Melton Mowbray, LE14 4RT, U.K.

(1) Yeretizian, C.; Pollien, P.; Lindinger, C.; Ali, S. *Compr. Rev. Food Sci. Food Saf.* **2004**, *3*, 152–159.

(2) Lindinger, C.; Labbe, D.; Pollien, P.; Rytz, A.; Juillerat, M. A.; Yeretizian, C.; Blank, I. *Anal. Chem.* **2008**, *80*, 1574–1581.

(3) Berrueta, L. A.; Alonso-Salces, R. M.; Héberger, K. J. *Chromatogr., A* **2007**, *1158*, 196–214.

(4) Wold, S.; Sjöström, M.; Eriksson, L. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.

(5) Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; van Velzen, E. J. J.; van Duijnhoven, J. P. M.; van Dorsten, F. A. *Metabolomics* **2008**, *4*, 81–89.

(6) Broadhurst, D. I.; Kell, D. B. *Metabolomics* **2006**, *2*, 171–196.

(7) Coleman, M. D.; Brewer, P. J.; Smith, I. M.; Harris, P. M.; Clift, M. G.; Milton, M. J. T. *Anal. Chim. Acta* **2007**, *601*, 189–195.

Table 1. Validation Criteria in Predictions of Sensory Qualities of Food or Beverages using Chemical Data (2005 to 2009)

product	chemical analysis ^b	chemometrics ^c	number of papers (methods) ^a		
			no validation	cross-validation	external validation ^d
beverages, bread, dairy, meat	13 (near/mid-infrared)	10 (PLSR), 1 (SVMR) 2 (MLR)	1	10	2
wine	2 (GC-olfactometry)	2 (PLSR)		2	
balsamic vinegar, beverages	3 (SPME-GC)	3 (PLSR)		3	
beverages, bread, dairy, olive oil	7 (headspace-MS)	1 (OSC-PLSR), 5 (PLSR), 1 (PCR)		4	3
fish, green tea	3 (GC-MS)	2 (OSC-PLSR), 1 (PLSR)	1	2	
beverages, olive oil	6 (liquid chromatography)	4 (PLSR), 1 (MLR), 1 (LR)	3	3	
beverages, fish, fruit, meat	9 (chemical sensors)	1 (OSC-PLSR), 7 (PLSR), 1 (RBFN)		8	1
cheese, meat, wine	3 (multiple)	2 (PLSR), 1 (GAR)		3	
beverages, fruit	3 (NMR)	1 (OSC-PLSR), 2 (PLSR)		3	

^a Journals (number of papers): *Am. J. Enol. Viticult.* (1), *Anal. Chem.* (2), *Anal. Chim. Acta* (6), *Chemom. Intell. Lab. Syst.* (2), *Eur. Food Res. Technol.* (4), *Food Chem.* (6), *Food Qual. Prefer.* (1), *Int. Dairy J.* (1), *Int. J. Food Microbiol.* (1), *J. Agric. Food Chem.* (10), *J. Chromatogr.* (1), *J. Dairy Sci.* (1), *J. Food Sci.* (2), *LWT* (1), *Meat Sci.* (8), *Sens. Actuat.* (2). ^b GC, gas chromatography; MS, mass spectrometry; NMR, nuclear magnetic resonance; SPME, solid-phase microextraction. ^c GAR, genetic algorithm-regression; LR, linear regression; MLR, multiple linear regression; OSC, orthogonal signal correction; PCR, principal component-regression; PLSR, partial least-squares-regression; RBFN, radial basis function network; SVMR, support vector machine-regression. ^d Validation primarily on samples unequivocally excluded from model development.

chemical analyses acquired over different periods, and encompassing such diversity of materials as might be encountered in a given industry.

This paper evaluates sensory profile prediction from realistically disparate chemical data derived from cocoa. The material analyzed, cocoa, is among the most chemically complex of foods,^{8,9} derived from both fermentation (like wine) and roasting (like coffee). The former occurs as beans, and pulp from the pods, are incubated by traditional methods, usually near the harvest site.¹⁰ Fermented beans are sun-dried before shipment to chocolate manufacturers, who roast and mill them, at temperatures above the melting points of cocoa fats, to “liquors”.¹¹ We evaluated liquors from beans of different geographical origins, crop varieties and fermentation regimes, months apart in chemical analysis. Recent years have seen the chemical identification of many components of cocoa flavor,^{8,9} but their comprehensive separation would be time-consuming for routine monitoring. Chemometric models predicting cocoa sensory qualities have not previously been reported.

The goal of rapidity means chemical models are often based on nonchromatographic techniques (Table 1). Here, we report a robust tool to be nominal-mass flow-injection electrospray-mass spectrometry (FIE-MS),¹² a recent method that has been used in food analysis,¹³ though not sensory attribute prediction. In FIE-MS, analytes enter the ionization chamber in liquid flow without chromatography. Infusion may last only 1 min, and even with between-injection delays to avoid carry-over, FIE-MS is very high-throughput.¹² With no chromatographic alignment needed, data

processing is also expedited.¹⁴ Binning of MS signals to nominal unit m/z reduces dimensionality (i.e., spectral variables), but also resolution, so each signal may integrate several compounds.¹² FIE-MS spectra are thus “fingerprints” with utility as composite patterns for chemometric evaluation, rather than definitive information on chemical constituents.

We compared different chemometric techniques for prediction of cocoa sensory qualities from FIE-MS fingerprints. Each technique was used in a similar supervised modeling procedure.³ Training data, comprising matrices of spectral fingerprints (**X**-data) and associated sensory data (**Y**-data), were provided to the algorithm for production of a model of **Y** from **X**. The resultant models were then evaluated using only the **X** data from the test set, and predictions of **Y** compared with actual values. Given our operational aims, a key criterion was robust *generalization* of models to different analysis dates and materials. As can be seen in Table 1, the current benchmark is standard PLS-regression, used in three-quarters of relevant papers. PLS models comprise multivariate components derived from coordinated algebraic decompositions of the original **X** and **Y** data into scores and weights matrices, in which covariance of **X** and **Y** is maximized.⁴ Successive components describe lower proportions of original variance, as each one is calculated on **X** matrices “deflated” by subtraction of the previous components. Optimization of the number of components to retain in predictive models is a step where CV may be used in model development.⁴

Spectral processing to remove noise such as baseline variation is often desired, and modifications of PLS to accomplish this exist. Ten percent of relevant studies (Table 1) used orthogonal signal correction (OSC), where the first PLS component(s) minimize rather than maximize covariance of **X** and **Y**, so their subtraction removes variation orthogonal to relations between **X** and **Y**.¹⁵ An acknowledged risk in OSC-PLS is overfitting,¹⁶ where fortuitous noise specific to training data is used in model development, and thereby diminishes generalization ability. This problem is report-

(8) Counet, C.; Callemien, D.; Ouwerx, C.; Collin, S. *J. Agric. Food Chem.* **2002**, *50*, 2385–2391.

(9) Frauendorfer, F.; Schieberle, P. *J. Agric. Food Chem.* **2006**, *54*, 5521–5529.

(10) Camu, N.; Gonzalez, A.; De Winter, T.; Van Schoor, A.; De Bruyne, K.; Vandamme, P.; Takrama, J. S.; Addo, S. K.; De Vuyst, L. *Appl. Environ. Microbiol.* **2008**, *74*, 86–98.

(11) Afoakwa, E. O.; Paterson, A.; Fowler, M. *Trends Food Sci. Technol.* **2007**, *18*, 290–298.

(12) Beckmann, M.; Parker, D.; Enot, D. P.; Duval, E.; Draper, J. *Nat. Protoc.* **2008**, *3*, 486–504.

(13) Cevallos-Cevallos, J. M.; Reyes-De-Corcuera, J. I.; Etxeberria, E.; Danyluk, M. D.; Rodrick, G. E. *Trends Food Sci. Tech.* **2009**, *20*, 557–566.

(14) Enot, D. P.; Lin, W.; Beckmann, M.; Parker, D.; Overy, D. P.; Draper, J. *Nat. Protoc.* **2008**, *3*, 446–470.

(15) Wold, S.; Antti, H.; Lindgren, F.; Ohman, J. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 175–185.

(16) Trygg, J.; Wold, S. *J. Chemom.* **2002**, *16*, 119–128.

edly diminished in orthogonal PLS (OPLS), a more recent method with similar objectives, which models \mathbf{Y} -correlated and uncorrelated variation in \mathbf{X} as separate predictive and orthogonal components.¹⁶

A few relevant papers used machine-learning approaches such as support vector machines (SVMs), radial basis function (RBF) networks, or genetic algorithms (Table 1). "Support vectors" are the members of two data classes with least mutual separation, and are used to optimize a separating hyperplane equidistant from each class. Functions called kernels can map the data to higher-dimensional space, where definition of the hyperplane allows linear separation.¹⁷ SVM regression is performed by addition and subtraction of a new parameter, ϵ , to each \mathbf{Y} value, thereby generating two data classes. The hyperplane between the plus- ϵ and minus- ϵ classes is equivalent to a regression function.¹⁸ SVMs interested us as reports and statistical learning theory emphasize their generalization ability,¹⁷ though this is dependent on correct setting of certain model hyperparameters.¹⁹

In summary, we evaluated models, produced by these alternative chemometric techniques, for diverse test materials, subjected to sensory and chemical analysis at different periods. Our central question was whether sensory qualities of materials of different industrial and analytical provenance could be predicted from their chemical data.

EXPERIMENTAL SECTION

The following four sets of test materials of diverse provenance were analyzed.

1. Geographical Origins. Six cocoas (from the Dominican Republic, Ghana, Papua New Guinea, Sulawesi, and two from Ecuador) were sourced as dried fermented beans (the form in which commodity cocoa is traded) to compare commercial market products of diverse geographical origins.

2. Benchmark Varieties. Five well-known varieties or types (Amelonado, Catongo, EET 53, Scavina 6, CCN 51) were grown and the beans microfermented at Fazenda Almirante (a division of Mars, Inc. in Itajupe, Bahia, Brazil). Microfermentation involved a few kg of fresh beans in nylon net bags being buried in larger masses of wet cocoa to effect a bulk fermentation.²⁰ Amelonado, probably derived from an eighteenth century Brazilian introduction,²¹ is the prevalent cocoa type worldwide. Catongo is a similar type, but differs in bean color and certain genetically useful characters.²² EET 53 is a traditional Ecuadorian fine cocoa type called Nacional.²³ These traditional varieties or types were the only ones prior to utilization in the 1950s of germplasm from Upper Amazon expeditions seeking resistance to witches' broom, a devastating fungal disease. Scavina 6 is one of the key clones from

which resistance was obtained.²⁴ CCN 51 is a more recent clone from a breeding program in Ecuador, where it has been planted in the last 30 years for its disease resistance.²⁵

3. Germplasm Accessions. Microfermentation at Fazenda Almirante was performed on 22 accessions of a germplasm collection at the Centro de Pesquisa do Cacau (CEPEC/CEPLAC), Itabuna, Bahia, Brazil:²⁴ AMAZ 5; AMAZ 15; Maria Rosa; NO 34; COCA 3370-5; PH 16; EET 45; SJ 02; GNV 53; SCA 11; Joventina 32; TSH 1188; LCTEEN 37A; LCTEEN 1621010; VB 276; VB 514; VB 636; VB 663; VB 892; VB 900; VB 1117; VB 1151. Those coded AMAZ, LCTEEN, COCA, or SCA, were from Ecuadorian or Peruvian Upper Amazon expeditions; others were hybrids from breeding centers in Trinidad (TSH), Ecuador (EET) and the USA (GNV), or farm selections in Brazil (VB, PH, SJ).^{24,26,27}

4. Fermentations. Twelve cocoas were produced on commercial plantations in Ghana and Sulawesi, in experiments on postharvest practices. In each location, both Sulawesi-style and Ghana-style fermentations were performed on the local beans. In the former, beans (75 kg) in large woven plastic bags were supported in log frames for 2 days. In the latter, palm leaves were used as a base for drainage and air movement below a bean heap, over which banana leaves were placed for a week. From the third day, the heap was opened for sampling every 2 days, then mixed and re-established. After all fermentations, beans were dried, 5 cm deep, on racks for 6–7 days, with turning 3–4 times daily. Unfermented cocoas were also produced by removal of pulp from freshly harvested beans in three cycles of washing and pressing, followed by sun-drying. In addition, two Sulawesi bean samples (unfermented and Sulawesi-style fermented) were treated with a proprietary process.

Cocoa Liquor Preparation. Liquors were prepared from all cocoa beans, by roasting (145 °C, 30 min) in a forced-air oven, breaking and winnowing to remove shells, then milling (75 min) in a mortar and pestle mill preheated to 60 °C.^{20,28}

Sensory Analysis. Liquors were presented to assessors melted on a hot plate at 40–45 °C, under red light to mask color. They were tasted with time for palate recovery between each set of three different cocoas. Sets were taken in random order, but designed so each liquor was tasted as first, second and third in a set.²⁰ The panel had nine assessors, so each panel mean came from 27 evaluations. Quantitative descriptive analysis profiling was used, with 15-point category scales for each attribute, for which assessors were trained and provided with declared anchored references. Thirteen sensory attributes (citrus/fruity, green, bitter, astringent, winery, cocoa, brown fruit, burnt, woody, acetic, nutty, floral, coconut) were analyzed as panel means.²⁸ Test material sets 1 and 2 (see above) were assessed in the same panel sessions, the others in different sessions.

(17) Mahadevan, S.; Shah, S. L.; Marrie, T. J.; Slupsky, C. M. *Anal. Chem.* **2008**, *80*, 7562–7570.

(18) Li, H. D.; Liang, Y. Z.; Xu, Q. S. *Chemom. Intell. Lab. Syst.* **2009**, *95*, 188–198.

(19) Cherkassky, V.; Ma, Y. *Neural Networks* **2004**, *17*, 113–126.

(20) Clapperton, J. F.; Lockwood, G.; Yow, S. T. K.; Lim, D. H. K. *Cocoa Grow. Bull.* **1994**, *48*, 47–63.

(21) Motamayor, J. C.; Risterucci, A. M.; Lopez, P. A.; Ortiz, C. F.; Moreno, A.; Lanaud, C. *Heredity* **2002**, *89*, 380–386.

(22) Crouzilat, D.; Lerceteau, E.; Petiard, V.; Morera, J.; Rodriguez, H.; Walker, D.; Phillips, W.; Ronning, C.; Schnell, R.; Osei, J.; Fritz, P. *Theor. Appl. Genet.* **1996**, *93*, 205–214.

(23) Lerceteau, E.; Flipo, S.; Quiroz, J.; Soria, J.; Petiard, V.; Crouzilat, D. *Euphytica* **1997**, *95*, 77–87.

(24) Marita, J. M.; Nienhuis, J.; Pires, J. L.; Aitken, W. M. *Crop Sci.* **2001**, *41*, 1305–1316.

(25) International Cocoa Germplasm Database, 2010. <http://www.icgd.rdg.ac.uk/>.

(26) Sanches, C.; Pinto, L.; Pomella, A.; Silva, S.; Loguericio, L. *Eur. J. Plant Pathol.* **2008**, *122*, 517–528.

(27) Leal, J. B.; dos Santos, L. M.; dos Santos, C. A. P.; Pires, J. L.; Ahnert, D.; Corrêa, R. X. *Pesqui. Agropecu. Bras.* **2008**, *43*, 851–858.

(28) Sukha, D. A.; Butler, D. R.; Umaharan, P.; Boulton, E. *Eur. Food Res. Technol.* **2008**, *226*, 405–413.

Correlation and hierarchical cluster analyses of the 13 sensory attributes of all 45 test materials were done in PAST v.1.66.²⁹ Principal component analysis (PCA) was performed in SIMCA-P v.11.0 (Umetrics, Umeå, Sweden), on the sensory data after mean-centering. Scores in PCA of all four sets of test materials (45 × 13 matrix), were compared with those obtained when each was predicted using PCA models of the remaining three sets (23–40 × 13 matrices).

The extent to which PCs were informative (“significant” in the manufacturer’s terminology) about sensory attributes was assessed in the default leave-seven-out SIMCA-P CV process that evaluates successive PCs. A PC was included in the model if its associated CV error (squared differences between actual and predicted hold-out data) was less than the residuals following the previous PC.

Flow-Injection Electrospray-Mass Spectrometry. Each set of test materials was analyzed separately by FIE-MS over a 15 month period. Each cocoa liquor was analyzed in nine 20 mg replicates, ultrasonically extracted in 1 mL of 80% methanol: 20% water for 5 min. Supernatants (clarified for 10 min at 13 400g) were loaded in the autosampler of an Alliance 2690 system (Waters, Elstree, UK) delivering the same solvent at 0.5 mL min^{−1}, and 20 μL introduced via a split flow, with no intervening chromatography column, into the Z-spray source of a Waters LCT mass spectrometer. Positive and negative mode spectra were collected in separate analyses with capillary voltages of +3 kV and −2 kV. Other electrospray conditions were: source and desolvation at 80 and 150 °C; sample and extraction cones at 30 and 5 V; nebulizer and desolvation N₂ at 60 and 460 L/h. Spectra (*m/z* 65–1000) were collected in 1 s scan cycles, binned to unit *m/z* values,¹² and normalized to total ion current of the sample infusion. Positive- and negative-ion spectra were concatenated, creating **X** data of 1872 variables for each sample. These data were curated for outliers by PCA,¹² one batch of analytical replicates (16% of all data) being deleted.

Tandem MS involved nanospray analysis on a Waters Q-tof mass spectrometer (capillary 0.8 kV, cone voltages as above, argon collision gas 3.1 × 10^{−5} mBar).

Chemometric Modeling. The *SimpleLinearRegression* algorithm in the Weka v.3.6 data mining suite³⁰ was used to obtain linear regression models based on the single spectral variable (i.e., relative intensity at a unit *m/z*) that the algorithm found to give minimal error on the training set.

SIMCA-P was used for PLS-regression (with OSC if stated), and OPLS-regression, on mean-centered data.^{4,15,16} The number of components for PLS models was optimized on training data by the SIMCA-P CV process with each cocoa as a CV group. OSC components with eigenvalues >2 were removed from models where OSC was used.¹⁵

The *libsvm* interface in the R package e1071³¹ was used for SVM-regression on untransformed data with linear, RBF, and second- and third-degree polynomial kernels.¹⁸ Grid searches with 10-fold CV were used to evaluate training errors with different settings of model hyperparameters.¹⁷ These included ϵ (maximum

deviation from the regression hyperplane to which model optimization is insensitive), and the regularization parameter (c) that determines how deviations > ϵ are tolerated.¹⁹ An RBF and polynomial function hyperparameter (γ) was likewise tuned for these kernels.¹⁷ We also tried estimating appropriate hyperparameters from training data statistics.¹⁹ Models with hyperparameters thereby defined in training were evaluated in prediction of **Y** for the independent validation materials.

Correlation of every individual spectral variable with sensory data was performed by the *corrcoef* function in MATLAB v.6.5 (The MathWorks, Natick, MA). Standard Bonferroni assessments of *P* values were made.⁶

RESULTS AND DISCUSSION

Sensory Panel Profiles of Cocoa Liquor. For 45 liquors from four sets of test cocoas, 13 attributes were evaluated by sensory panel. Certain sensory attributes cocrrelated across all the cocoas, for example, astringent/bitter/green ($r \geq 0.96$) and cocoa/winey ($r = 0.95$). These attributes also grouped in hierarchical cluster analysis by all tested distance measures (Euclidean, Pearson and rank correlation, Manhattan) and clustering rules (unweighted pair-group average, nearest neighbor, Ward’s method). Some attributes varied more than others, for example, variance of winey was 4.6, coconut only 0.02. PCA of all cocoas by their data-vectors of 13 sensory attributes was dominated by the first component, PC[1], representing 82% of overall variance (Figure 1a). Although the relatively minor variation (9.3%) on PC[2] contributed to separation of cocoas, such as the “Germplasm Accessions” set, PC[1] was the only one found informative (“significant”) in SIMCA-P CV tests.

Toward the negative end of the PC[1] axis were cocoas from early stage experimental fermentations, and commercial ones from Sulawesi and Ecuador, where fermentations are brief.³² These were characterized by astringent/bitter/green profiles, reflected in the strong negative weight coefficients (“loadings”) on PC[1] of these attributes (Figure 1b). Toward the positive end of PC[1], defined by cocoa/winey attributes, were most other cocoas, including the later-stage experimental fermentations, with the commercial sample from Ghana, where fermentations are longer,³² at the extreme (Figure 1a and b). Thus, the negative-to-positive direction on PC[1] in general reflected fermentation duration.

To confirm the general applicability of PC[1] to independent materials, we predicted the PC[1] scores for each test set from PCA models of the other three, finding strong correlation with those from combined PCA of all test sets ($r = 0.98$). Coordinates (i.e., “scores”) of cocoas on the sensory spectrum encapsulated by PC[1] were therefore taken as a composite sensory index for modeling from chemical data.

FIE-MS Peaks Correlating with Sensory Qualities. FIE-MS spectral “fingerprints” of methanolic liquor extracts were produced without chromatographic separation. Comprising relative intensities at 1872 nominal unit-mass *m/z* values,¹² these were provided as **X**-data for models of the sensory-PC[1] scores as **Y**-values. These spectra were semiquantitative, given the lack of response calibration for so many peaks (chemical identities of which were often undefined), and the potential for differential suppression of ionization of molecules in complex mixtures.¹²

(29) Hammer, Ø.; Harper, D. A. T.; Ryan, P. D. *Palaeontol. Electron.* **2001**, 4, 1. <http://palaeo-electronica.org>.

(30) Witten, I. H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2 ed.; Morgan Kaufmann: San Francisco, 2005.

(31) Meyer, D. *R News* **2001**, 1/3, 23–26.

(32) Counet, C.; Ouwerx, C.; Rosoux, D.; Collin, S. *J. Agric. Food Chem.* **2004**, 52, 6243–6249.

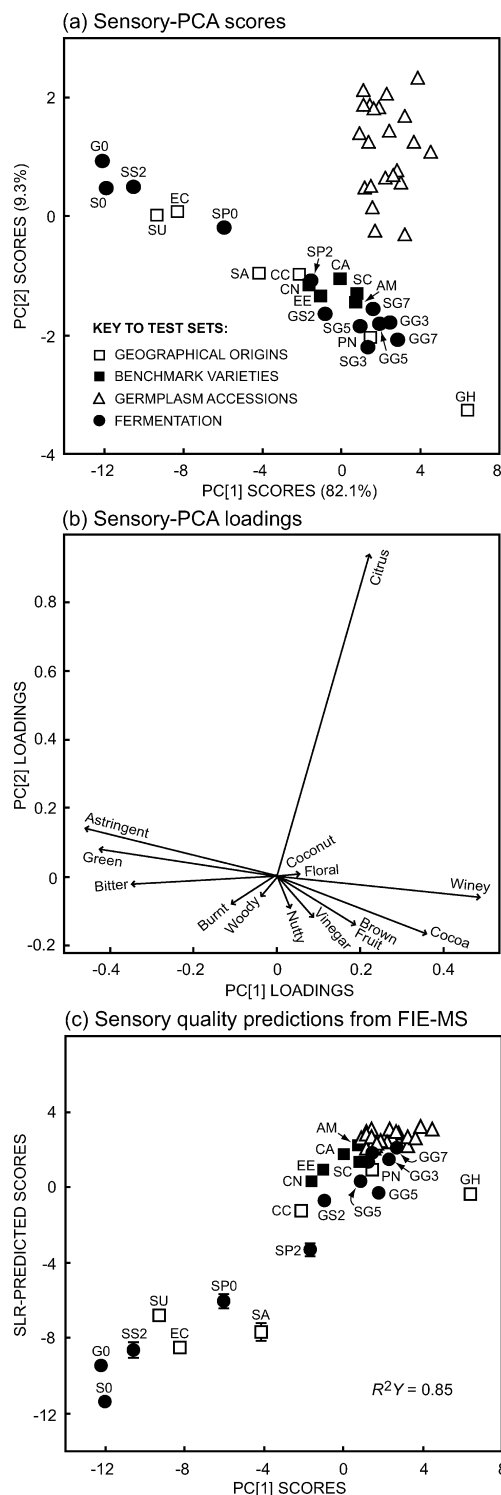


Figure 1. Modeling of cocoa liquor sensory qualities. (a) Scores and (b) loadings of first two components (PC[1], PC[2]), in PCA of 45 cocoas by 13 sensory attributes. Axis label brackets in (a) show variance encapsulated. (c) Prediction of the sensory-PC[1] scores from FIE-MS data (in independent validations) by Weka *SimpleLinearRegression* models. Samples from the four test sets (see Experimental Section) are indicated by data symbols. Open squares, Geographical Origins: EC, Ecuador; SU, Sulawesi; SA, Sanchez (from the Dominican Republic); CC, CCN51 (from Ecuador); PN, Papua New Guinea; GH, Ghana. Black squares, Benchmark Varieties: CN, CCN51; EE, EET53; AM, Amelonado; CA, Catongo; SC, Scavina 6. Open triangles, Germplasm Accessions. Black circles, Fermentations: first letter, location; second (if relevant), fermentation style (G, Ghana; S, Sulawesi; P, proprietary process); numeral, sampling day. Bars: \pm standard errors of analytical replicates ($N = 6 - 9$).

Table 2. Comparison of Models Predicting Sensory Qualities of 45 Cocoas from FIE-MS Data

chemometric methods	independent validations ^a		LOO CV ^b	
univariate	RMSEP	SRD	RMSEP	SRD
Weka SLR (m/z 289) ^c	1.66	314	1.64	310
multivariate				
PLS	2.44	380	1.66	298
linear-SVM ^d	2.89	434	1.95	283
OPLS	2.93	410	1.65	300
OSC-PLS	3.49	494	1.99	344

^a FIE-MS data, from different analysis dates, were used to predict sensory-PC[1] scores (see Figure 1a) of each set of cocoa materials (Geographical Origins, Benchmark Varieties, Germplasm Accessions, Fermentations) by chemometric models developed on the other three sets. ^b LOO CV for same data; sensory-PC[1] scores for each of the 45 cocoa materials were predicted from FIE-MS data on the other 44. ^c SLR, the Weka algorithm *SimpleLinearRegression*, which selected the epicatechin m/z 289 negative-ion (Figure 2) as the single data attribute giving minimal error in linear regression models. ^d SVM hyperparameters were tuned to minimize training error (Figure 3, arrows "1"). Performance parameters summarize all 45 predictions: RMSEP, eq 1; SRD, sum of ranking differences.⁴⁸

We searched for univariate marker signals for sensory qualities in the spectral data, using the Weka algorithm *SimpleLinearRegression*. The Weka software suite comprises an extensive collection of machine learning algorithms for data mining problems.³⁰ Given a multivariate regression problem, *SimpleLinearRegression* fits a simple regression function to each attribute in the training data, selecting only the one yielding the minimum least-squares error.³⁰

Despite the chemical complexity of cocoa flavor, the parsimonious *SimpleLinearRegression* algorithm gave excellent results (Figure 1c, Table 2) in operationally "independent" validations, defined here as prediction of materials of different origins, and dates of chemical analysis, to those used in model construction. For this purpose, models of sensory-PC[1] scores were developed using only three of the four sets of cocoa materials (viz., Geographical Origins, Benchmark Varieties, Germplasm Accessions, Fermentations, see Experimental Section). These models were then used to predict the sensory-PC[1] scores of the excluded set of cocoa materials, on the basis of their FIE-MS fingerprints. Our procedures therefore involved a more meaningful dimension of independence than the minimum statistical definition of "external" validation.³

For all relevant training data sets, *SimpleLinearRegression* selected the negative-ion m/z 289 peak for its regression models. This peak was the most significant in univariate correlations of each m/z variable with sensory-PC[1] scores in the whole data set. Given its spectral prominence in some cocoas (Figure 2), it was an ideal marker for sensory qualities.

Tandem MS identified deprotonated (–)-epicatechin, a cocoa flavan-3-ol,³³ at the negative-ion m/z 289, by known product ions:^{34–37} m/z 271 (H_2O loss), 245 (CH_2-CHOH loss), 205

(33) Ortega, N.; Romero, M. P.; Macià, A.; Reguant, J.; Anglès, N.; Morelló, J. R.; Motilva, M. J. *J. Agric. Food Chem.* **2008**, *56*, 9621–9627.

(34) Pérez-Magariño, S.; Revilla, I.; González-San José, M. L.; Beltrán, S. *J. Chromatogr., A* **1999**, *847*, 75–81.

(35) Bravo, M. N.; Silva, S.; Coelho, A. V.; Boas, L. V.; Bronze, M. R. *Anal. Chim. Acta* **2006**, *563*, 84–92.

(36) Rodríguez-Medina, I. C.; Segura-Carretero, A.; Fernández-Gutiérrez, A. *J. Chromatogr., A* **2009**, *1216*, 4736–4744.

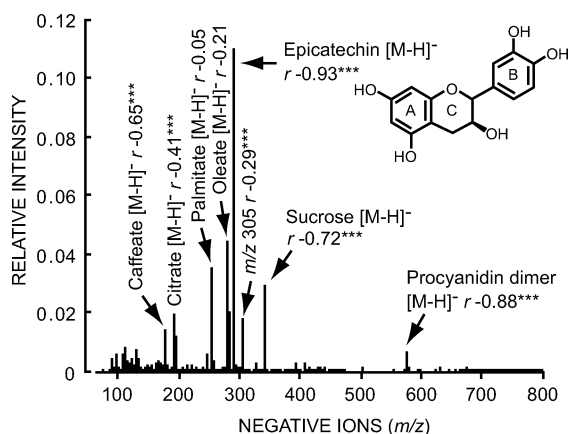


Figure 2. Negative-ion FIE-MS fingerprint of Ecuadorian cocoa liquor (sample EC of Figure 1), with tentative peak attributions (see text for tandem MS evidence). Pearson coefficients (r): correlation of each peak with sensory-PC[1] scores (see Figure 1a) in all FIE-MS data. *** Significant at a Bonferroni-corrected level of $P < 10^{-6}$.

(flavonoid A-ring loss), 203, 179 (flavonoid B-ring loss), 125, 109 (dihydroxybenzene). Identical fragments form with catechin enantiomers,³⁵ which, though less abundant in cocoa,^{33,38} probably also contributed to the m/z 289 peak. Procyanidins, the major polyphenols in cocoa, are epicatechin oligomers.³² The m/z 577 peak, while less prominent than epicatechin, represented the component next most correlated with sensory-PC[1] scores (Figure 2), and was identified as dimeric procyanidin by tandem MS product ions at m/z 451, 425, 407, and 289.³⁹

Cocoa epicatechin is the subject of interest in relation to potential cardiovascular health benefits in humans,⁴⁰ and the procyanidin content of cocoa is under scrutiny.³³ During fermentation, these compounds are liberated from polyphenol-rich cotyledon cells to undergo oxidation and complexation.¹⁰ Their greater abundance in less-fermented and more bitter cocoas,³² was consistent with their correlation with the fermentation and sensory-quality gradient encapsulated in the sensory-PC[1] scores (Figure 1). Since procyanidins have intrinsic bitterness,⁴¹ they will contribute directly to flavor. However, their lower levels in well-fermented cocoas also correlate with greater formation in roasting of key odorants such as pyrazines.³² In this context, m/z 289 would be a “proxy” marker for general processes affecting important flavor volatiles excluded from FIE-MS analyses. Confirmation that epicatechin alone cannot fully encapsulate cocoa sensory profiles is seen in the rather poor prediction of the (commercially important)⁴² Ghanaian sample (Figure 1c).

A definite case of a proxy marker for sensory qualities was m/z 341 (Figure 2), which gave product ions of deprotonated sucrose (m/z 179, 161, 149, 143, 131, 119, 89).⁴³ This correlated

with sensory-PC[1] scores in the same direction as epicatechin, even though sucrose counters the bitterness of phenolics.⁴¹ More relevant was that sucrose also declines in fermentation,⁴⁴ and so is depleted in Ghanaian relative to Ecuadorian cocoa.⁴⁵ Citrate, another key metabolite to decline in fermentation,⁴⁴ was seen at m/z 191 (product ions at m/z 111, 87, 85),⁴⁶ with similar if less marked sensory-correlation (Figure 2). Stronger correlation was shown by m/z 179 (Figure 2), probably the phenolic acid caffeate (product ion at m/z 135).⁴⁷

Comparative Accuracies of Predictive Models of Sensory Qualities. Alternative chemometric approaches were compared in further independent validations with the four sets of cocoa materials. Overall performance statistics for the predictions of all four test sets are presented (Table 2), with regression methods ranked by root mean squared error of prediction (RMSEP), evaluated as

$$\text{RMSEP} = \sqrt{\frac{1}{N} \sum (\hat{y} - y)^2} \quad (1)$$

from the number of samples (N), and predicted (\hat{y}) and actual (y) sensory-PC[1] scores.

Remarkably, none of the multivariate models matched the RMSEP of the *SimpleLinearRegression* models of the epicatechin m/z 289 negative ion (Table 2). This was also found when models were compared using the recently proposed⁴⁸ sum of ranking differences (SRD) measure, for which absolute differences between true and predicted rankings of sensory-PC[1] scores were tallied.

The best multivariate method was standard PLS-regression (Table 2). The m/z 289 peak dominated the PLS-regression coefficients,⁴ its average value being five times the next most prominent. If the m/z 289 variable was masked, performance deteriorated (RMSEP = 3.92). The success of the PLS models was thus due to the high correlation of a prominent spectral peak with the Y data.

OPLS models for independent validations were variable (standard deviations of their coefficients being 5-fold those of PLS), and predictive performance was poorer (Table 2).

Predictive performance was diminished by OSC (Table 2). A basic problem was that deciding which OSC components to remove, on the basis of their eigenvalues¹⁵ in training, did not optimize validation performance. (One, or more often two, OSC components, representing 24–37% of training data variation, were removed in practice.)

SVMs performed better than OSC-PLS (Table 2), despite being similarly subject to difficulties of model adjustment in training. Figure 3 shows the linear-SVM validation results (as RMSEP and regression slope) relative to training RMSEP, which varied with the settings of the ϵ and c hyperparameters.¹⁹ The linear-SVM models in Table 2 were tuned to minimize training RMSEP (arrows labeled “1”, Figure 3). This standard tuning approach did not, however, yield the optimal validation results (arrows labeled

(37) Bastos, D. H. M.; Saldanha, L. A.; Catharino, R. R.; Sawaya, A. C. H. F.; Cunha, I. B. S.; Carvalho, P. O.; Eberlin, M. N. *Molecules* **2007**, *12*, 423–432.

(38) Kofink, M.; Papagiannopoulos, M.; Galens, R. *Molecules* **2007**, *12*, 1274–1288.

(39) Sun, J.; Liang, F.; Bin, Y.; Li, P.; Duan, C. *Molecules* **2007**, *12*, 679–693.

(40) Schroeter, H.; Heiss, C.; Balzer, J.; Kleinbongard, P.; Keen, C. L.; Hollenberg, N. K.; Sies, H.; Kwik-Urbe, C.; Schmitz, H. H.; Kelm, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1024–1029.

(41) Drewnowski, A.; Gomez-Carneros, C. *Am. J. Clin. Nutr.* **2000**, *72*, 1424–1435.

(42) Clapperton, J. F. *Cocoa Grow. Bull.* **1994**, *48*, 7–16.

(43) Taylor, V. F.; March, R. E.; Longerich, H. P.; Stacey, C. J. *Int. J. Mass Spectrom.* **2005**, *243*, 71–84.

(44) Ardhan, M. M.; Fleet, G. H. *Int. J. Food Microbiol.* **2003**, *86*, 87–99.

(45) Redgwell, R. J.; Trovato, V.; Curti, D. *Food Chem.* **2003**, *80*, 511–516.

(46) Bylund, D.; Norström, S. H.; Essén, S. A.; Lundström, U. *S. J. Chromatogr. A* **2007**, *1176*, 89–93.

(47) Sánchez-Rabaneda, F.; Jáuregui, O.; Casals, I.; Andrés-Lacueva, C.; Izquierdo-Pulido, M.; Lamuela-Raventós, R. M. *J. Mass Spectrom.* **2003**, *38*, 35–42.

(48) Héberger, K. *TrAC, Trend. Anal. Chem.* **2010**, *29*, 101–109.

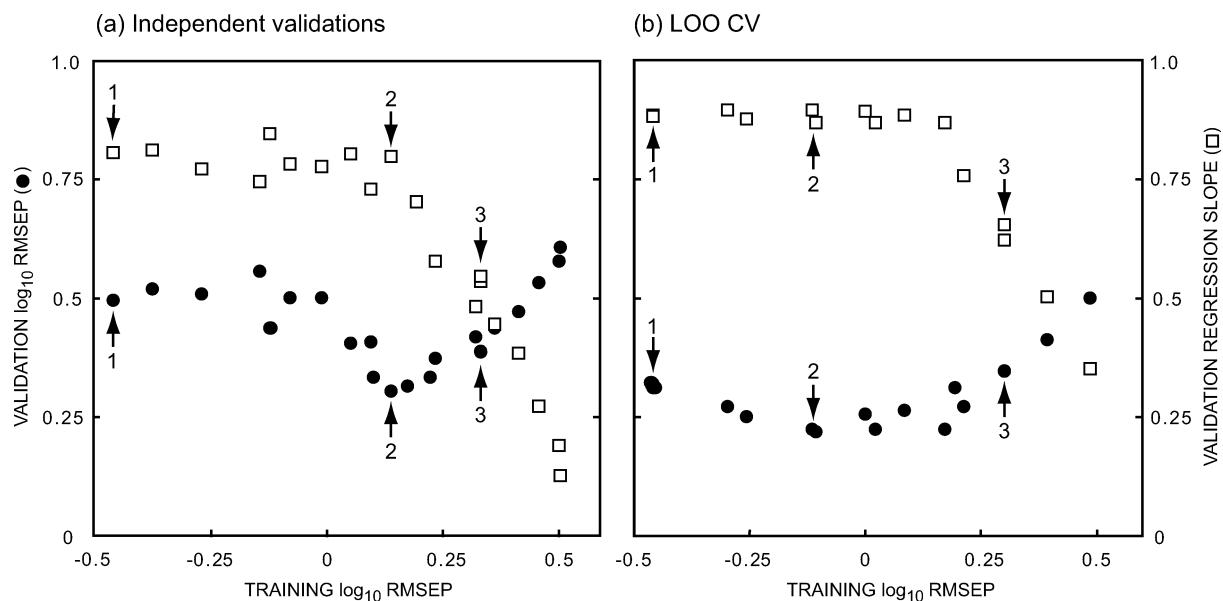


Figure 3. Relations between training and validation of linear-SVM models. Sensory-PC[1] scores were predicted from FIE-MS data. Training RMSEP, as eq 1, varied with SVM hyperparameters ϵ and c . Validation results are shown as RMSEP (black circles), and regression slope (open squares) for predicted versus actual values, collated for (a) independent validations, in which of each set of cocoa materials (Geographical Origins, Benchmark Varieties, Germplasm Accessions, or Fermentations) was predicted by chemometric models developed on the other three sets; and (b) LOO CV. Arrows indicate models: (1) tuned to minimize training RMSEP, as for Table 2; (2) giving optimum validation results; (3) with hyperparameters derived from training data statistics.¹⁹

“2”, Figure 3), which came from models with higher training RMSEP. Since our remit did not permit any use of the validation **Y**-data in model development, we could find no training strategy to identify the optimal models. The same situation occurred with RBF and (second- or third-degree) polynomial kernels (not shown). This was a typical case of overfitting,¹⁷ for which avoidance strategies are needed if SVMs are to achieve their potential in contexts like this. One possibility is to replace empirical tuning by estimation of hyperparameters from training data statistics, and we tested the following proposed derivations:¹⁹

$$\epsilon = 3\sigma\sqrt{\frac{\ln N}{N}} \quad (2)$$

where σ is the standard deviation of **Y** and N the number of samples in the training data, and

$$c = y_i^{\max} - y_i^{\min} \quad (3)$$

where y_i^{\max} and y_i^{\min} are the maximum and minimum values of **Y**. Predictive outcomes with these hyperparameters are shown (arrows labeled “3”, Figure 3): while RMSEP was closer to optimal in the independent validations, the regression slope was poor.

Comparison of Validation Designs for Predictive Models.

For comparison, 45 models, each with one cocoa excluded, were developed. Since the models in a LOO CV encompass nearly all sample variation, predictions were almost always superior to independent validations (Table 2). We also tested 9- and 3-fold CV (not shown), as LOO CV can overfit in small samples,³ but predictions were not better.

For real-world applications, test and training data may unavoidably have separate origins and analysis dates. It was therefore of

paramount importance that even in independent validations some methods produced models that would be serviceable in an industrial context. Indeed, *SimpleLinearRegression* performed as well in independent validation as in LOO CV. In LOO CV other chemometric methods converged toward the same performance as *SimpleLinearRegression*, suggesting this was the best possible performance for these data. Comparing chemometric methods, therefore, the message would be “the simpler the better”. Leaving aside the possibly fortuitous success of *SimpleLinearRegression*, continued predominance of standard PLS (Table 1) seems justified.

The more refined OSC-PLS, OPLS and SVM methods were better in LOO CV than in independent validations, indicating that generalizing to different ranges of variation was a limitation of these approaches. SVM hyperparameter tuning strategies were, nonetheless, no more successful in LOO CV (Figure 3b).

CONCLUSIONS AND OUTLOOK

FIE-MS fingerprinting, despite its rapid, semiquantitative nature, confirmed the practical feasibility of chemical prediction of sensory qualities. This was achieved for materials with diverse provenance and analysis dates, which was essential for operational utility, but has not previously been explicitly demonstrated. The standard practice of CV of the whole data gave overoptimistic accuracies. The significance of this issue might dwindle over the long-term, as industrial applications generate training data resources encompassing the full spectrum of test materials to be encountered.

The application we have described concerns an industry, worth \$5.1 billion per annum,⁴⁹ with scope to benefit from predictive high-throughput chemical monitoring of sensory qualities. In breeding programs, for example, predictive moni-

(49) World Cocoa Foundation, 2010. <http://www.worldcocoaoundation.org/>.

toring could prevent loss of flavor, as brought by past international plantings.⁴² The cocoa crop is set to benefit from genetic advances,⁵⁰ for several purposes. The limited genetic base of traditional planting materials exposes them to epidemics, so Amazonian germplasm, such as analyzed here, has been collected over decades.²⁴ Other breeding targets include yield,⁵⁰ or consumer value via desirable constituents such as flavanols.³³ Our findings showed the fermentation process was also important for sensory qualities. Fermentation is carried out by smallholders, and each continent has its own practices and microbiota.^{10,44} Chemical monitoring projects could improve guidelines on postharvest processes for the fragmented production base of 5–6 million farmers in Latin America, West Africa and South-East Asia.⁴⁹

(50) Cervantes-Martinez, C.; Brown, J. S.; Schnell, R. J.; Phillips-Mora, W.; Takrama, J. F.; Motamayor, J. C. *J. Am. Soc. Hortic. Sci.* **2006**, *131*, 231–241.

ACKNOWLEDGMENT

We thank Caroline Atkinson for laboratory assistance, James Heald for mass spectrometry, Regina Machado for processing the microfermented cocoas, and Eduardo Panadero for preparing all cocoa liquors. Our appreciation also goes to farmers of Soppeng, Sulawesi, and Suhum, Ghana for endeavours in support of the postharvest processing samples, and to Isabella Van Damme, Subianto Mursalin, Arhamuddin Albek, Hussin bin Purung, and Moctar Sangare for help in organizing and running those trials. We also benefited from chemical or statistical advice from Manfred Beckmann, Wanchang Lin, David Broadhurst, David Enot, Roy Goodacre, and Douglas Kell.

Received for review March 10, 2010. Accepted June 4, 2010.

AC1006393