DOI:10.1021/jf100949k

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

Evaluation of the Production Year in Italian and Chinese Tomato Paste for Geographical Determination Using O2PLS Models

R. Consonni,*,† L. R. Cagliani,† M. Stocchero,‡ and S. Porretta§

†Istituto per lo Studio delle Macromolecole, Laboratorio NMR, Consiglio Nazionale delle Ricerche (CNR), via Bassini 15, 20133 Milan, Italy, ‡S-IN Soluzioni Informatiche, via Salvemini 9, 36100 Vicenza, Italy, and §Stazione Sperimentale per l'Industria delle Conserve Alimentari, via F. Tanara 31/A, 43100 Parma, Italy

Nuclear magnetic resonance (NMR) is nowadays largely used as valid tool in metabolomic applications. In this study, the metabolite content of Italian and Chinese tomato paste at different concentration rates of two production years (2007 and 2008) was investigated with the aim of building a robust geographical differentiation statistical model. A total of 119 tomato paste samples were analyzed by ^{1}H NMR and multivariate data analysis tools, in particular using bidirectional orthogonal projection to latent structures−discriminant analysis (O2PLS−DA). This technique is well-suited for noisy and correlated variables and was recently adopted to obtain robust classification models, having a clear interpretation of the systematic variation useful to characterize each class. In the present study, the analysis of latent space underlying the classification model allowed us to understand the role played by the production year on geographical discrimination. The O2PLS−DA model performed considering only tomato paste samples of 2007 was capable of predicting the geographical origin of all analyzed samples. The effect of the production year therefore resulted in not affecting the geographical origin discrimination.

KEYWORDS: Concentrated tomato paste; O2PLS−DA; S-plot; ^{1}H NMR; geographical origin; production year

INTRODUCTION

Most tomato-based foodstuffs, such as sauces, ketchups, puree, and juices, could be sold on the Italian market, such as "Made in Italy" products. Particularly in Italy, considered a worldwide leader in processed tomato quality, several tons of triple-concentrated tomato paste were imported from developing countries, mainly from China. Actually, Italian law (*1*) requires only tomato sauce producers to indicate the grown origin of tomato fruits on the label. In this context, many potential frauds regarding the real origin of tomato products could be made, and consequently, a growing interest from both consumers and producers about food geographical characterization and authenticity is increasing presently (*2*, *3*), with the former asking for more guarantees for what they bought, while the latter asking for better protection of their products.

The industrial production of concentrated tomato paste involves different steps, as already indicated in our previous paper (*4*). After the fruits were harvested, they were washed and treated with a hot- or cold-break process. In the first case, tomatoes are rapidly heated to 90 °C to thermally inactivate enzymes, such as pectin methylesterase and polygalacturonase; this process prevents the pectin breakdown, thus generating a high pectin content and consistency in the final product. In the cold-break process, the temperature achieves only about 65 °C, thus preventing enzymatic inactivation. The final product resulted in this case in a better natural color, fresher tomato flavor, and less density (*5*). After the seeds and peel were removed, a multi-step heat-exchanger evaporation process was performed, reaching different degrees of concentration: semi-concentrated (more than 12% of dry residual), mono-concentrated (more than 18% of dry residual), double-concentrated (more than 28% of dry residual), triple-concentrated (more than 36% of dry residual), and up to sextuple-concentrated (more than 55% of dry residual) (*6*).

It is well-known that several factors, such as environment, climate, and soil, could influence plant metabolism and, consequently, the fruit metabolic content; therefore, differences among seasons could determine differences in fruits and processed tomato products. As a matter of fact, many recently published papers focused their attention on the influence of the tomato season of harvest, sowing, variety, ripening stage, or geographical location on the antioxidant (*7*, *8*), polyphenol, and lipoic acid (*9*) composition, carotenoid content (*10*), or nutritional and organoleptic tomato characteristics (*11*−*14*). Conversely, only two studies faced the influence of the tomato harvest year on the metabolic content. In these latter papers, the tomato antioxidant (*15*) and folate (*16*) contents were monitored by high-performance liquid chromatography (HPLC) measurements. No data concerning the correlation between the year of tomato production and tomato process product geographical origin is present in the literature thus far.

*To whom correspondence should be addressed. Telephone: 02-23699578. Fax: 02-23699620. E-mail: roberto.consonni@ismac.cnr.it.

Published on Web 06/10/2010

Table 1. Confusion Matrix for 28 Concentrated Tomato Paste Samples of 2007 Constituting the Training Set, 27 Samples of 2007 Constituting the Test Set, 37 Samples of 2008, and 27 Validation Set Samples[a]

	training		test		2008		validation	
	predicted C	predicted I	predicted C	predicted I	predicted C	predicted I	predicted C	predicted I
C	14	0	13	0	23	2	2	0
I	0	14	0	14	1	11	2	23

[a] "I" and "C" stand for Italian and Chinese samples, respectively.

In our preliminary study, for the first time, we focused our attention on the geographical discrimination of triple-concentrated tomato paste produced in 2007, obtaining a very clear differentiation between Italian and Chinese samples, analyzing the water-soluble content by ^{1}H nuclear magnetic resonance (NMR) and multivariate statistical protocols (*4*). The NMR technique, in the last few years, has demonstrated its potentiality in food-quality and geographical determination (*17*), as well as tomato characterization (*18−21*). The metabolite content results were feasible with a single experiment, without any sample derivatization, and represent a sort of fingerprint of each analyzed food matrix, reflecting soil and climate of a determined region, cultivar characteristics, as well as particular local process treatments; it could therefore be considered as very useful and precious data for determining the authenticity of food products. Keeping in mind the previous considerations about the metabolic content that could be affected by the production years, in the present study, we analyzed 119 samples of both Italian and Chinese double- and triple-concentrated tomato paste produced in 2007 and 2008 to evaluate if and how the production year could influence the geographical discrimination. This aim was addressed using the bidirectional orthogonal projection to latent structures (O2PLS) technique to obtain robust and clearer classification models and to identify the relationships between the latent space underlying these models.

MATERIALS AND METHODS

NMR Samples. A total of 119 samples of both double- and triple-concentrated tomato paste were analyzed. Among these samples, 92 were of known origin and produced in 2007 or 2008; these samples represented the candidate set, upon which statistical models will be performed. The remaining 27 samples, purchased directly on market during 2007 and 2008 but without any label indication about their production year composed the validation set. Among candidate set samples, 55 were produced in 2007 and 28 were Italian (11 double- and 17 triple-concentrated tomato paste), while 27 were Chinese (all triple-concentrated tomato paste); 37 were produced in 2008 and 12 were Italian, while 25 were Chinese (all triple-concentrated tomato paste). Samples were prepared for NMR analysis following the previously described protocol (*4*); therefore, they were subjected to lyophilization before to be dissolved into buffered deuterated water. ^{1}H NMR spectra were recorded on Bruker DMX 500 spectometer (Bruker Biospin GmbH Rheinstetten, Karlsruhe, Germany) operating at 11.7 T and equipped with a 5 mm reverse probe with z-gradient. All spectra were recorded at 300 K; an exponential function was applied before Fourier transformation, and the phase and baseline were manually corrected with ACD/NMR software (ACD Labs, version 11, Toronto, Ontario, Canada). Spectra were aligned for bucket integration on the α-glucose signal at 5.12 ppm; all spectra were then reduced to integrated regions (buckets) of equal width (0.04 ppm) over the entire spectral region, while the residual water signal between 4.60 and 4.88 ppm and the citrate region from 2.33 to 2.65 ppm were set to zero constant value. The citrate signals were excluded because they could bias the results according to our previous study (*4*). Complete digitalized spectrum areas were internally normalized.

Statistical Methods. Principal component analysis (PCA) and bidirectional orthogonal projection to latent structures−discriminant analysis (O2PLS−DA) were performed with "Pareto" data pretreatment. O2PLS is a multivariate projection method that extracts linear relationships from two data blocks X and Y by removing the so-called structured noise (*22*, *23*). When structured noise is present in a data set X (or Y), traditional projection techniques, such as PLS regression can produce systematic variation of X (or Y), having a component uncorrelated to Y (or X). O2PLS removes this structured noise from both X and Y in a bidirectional way without imposing a particular direction in the prediction model. As a consequence, O2PLS decomposes the systematic variation in the X block (or Y block) into two model parts: the predictive or parallel part, modeling the joint $X-Y$ correlated variation, and the orthogonal part, not related to Y (or X). O2PLS can be used to perform discriminant analysis (DA) by introducing suitable dummy variables. The main benefit using the O2PLS−DA technique is the reduced model complexity (*24*). In the case of N classes, the dimension of the predictive space is $N - 1$ and therefore the model can be explained using only $N - 1$ components. The number of latent components can be determined by cross-validation techniques, and in this study, we used 7-fold cross-validation. In addition, a permutation test on the Y block was performed to safely overcome casualty or overfitting into models. When the dimension of the joint correlated space is one, useful visualization tools, such as, for example, the S-plot, can be used to highlight the role played by the variables in the model or in different models (*25*). The D-optimal onion design (*26*) was applied to select from the candidate set both training and test sets; those were extracted using MODDE 8.0 (Umetrics, Umea, Sweden) from the observation space described by PCA scores of 55 samples produced in 2007 and 92 samples produced in 2007 and 2008. Statistical data analysis was performed with the SIMCA-P+ 12 (Umetrics, Umea, Sweden) program.

RESULTS AND DISCUSSION

In our preliminary study, very good results in geographical differentiation of triple-concentrated tomato paste samples coming from China and Italy were obtained (*4*), with all tested samples produced with tomato fruits harvested in 2007. In the present work, we considered both double- and triple-concentrated tomato paste samples covering 2007 and 2008 productions coming from both China and Italy, and in this respect, a definitive model for Italian and Chinese concentrated tomato paste geographical assessment was investigated and proposed. Because the analyzed samples were previously lyophilized and the same amount resuspended in equal volumes of buffered solution before NMR analysis, the original compound concentration, present into double- or triple-concentrated tomato paste samples, was not influenced; all samples were therefore comparable for our analysis. This simple procedure will enable sample analysis of different formulations of concentrated tomato paste. The most important objective in this work was to test as much as possible samples with certain geographical origin, to build and validate the best statistical model. The acquired ^{1}H NMR spectra were subjected to "bucketing" with 0.04 ppm size; this value allowed little signal shift compensation that occurred, even though the best controlled conditions were employed for NMR samples and applied to all spectra. Citrate buckets from 2.33 to 2.65 ppm were excluded from the data set, according to our previous considerations (*4*). As already pointed out, citrate could be added for both pH correction and bacteria growth inhibition, even though it is not allowed by law, and for this reason, it should not be considered as a variable.

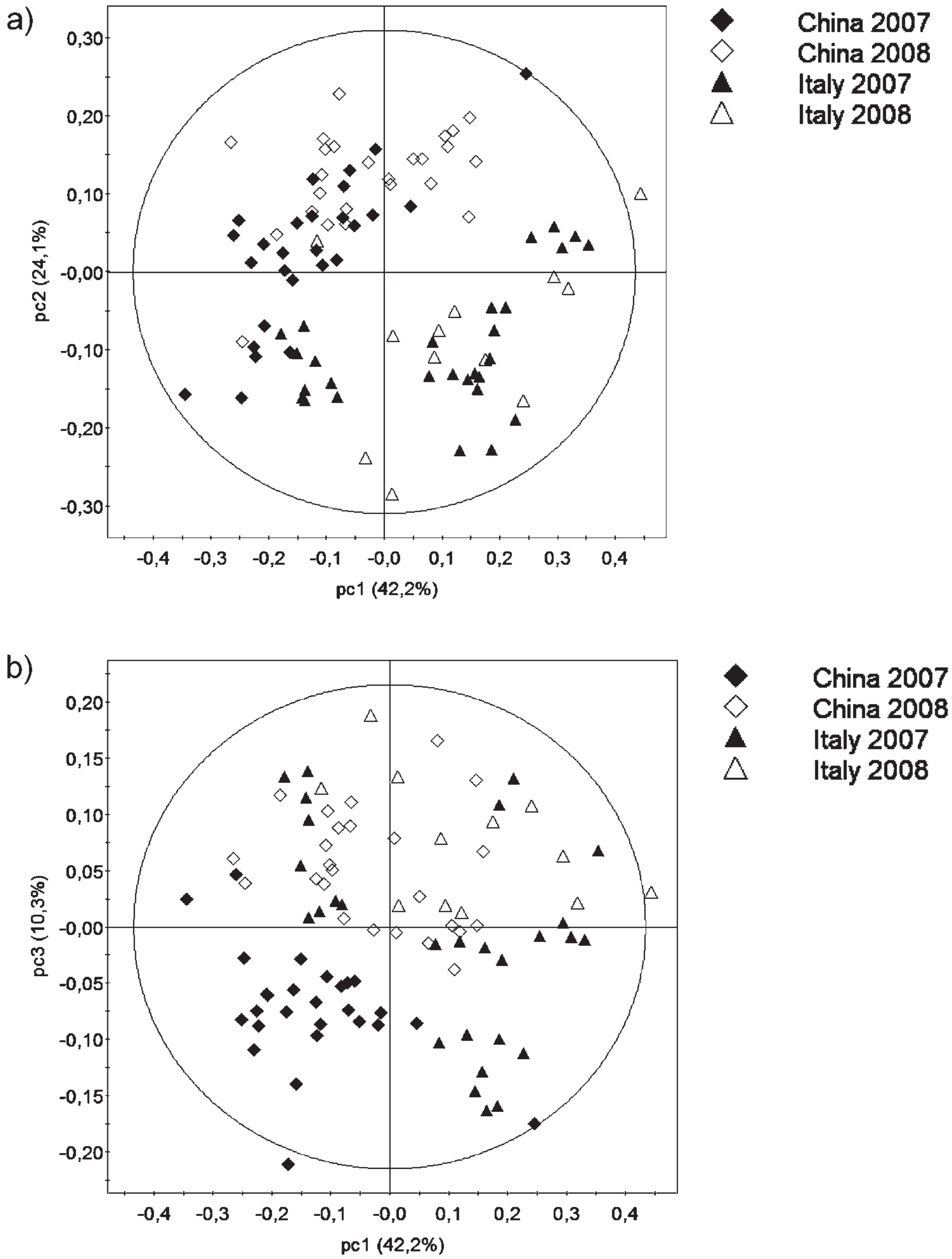

Figure 1. PCA score plot performed by considering all 92 concentrated tomato paste samples of certain origin and production year: (▲ and △) 2007 and 2008 Italian samples, and (◆ and ◇) 2007 and 2008 Chinese samples, respectively. PC1, 42.2%; PC2, 21.4%; PC3, 10.3%. These first three PCs accounted for R^2X = 74% and Q^2 = 67.4%. (A) PC1 versus PC2 and (B) PC1 versus PC3.

The validity of our previous statistical model was checked by performing O2PLS−DA on new double- and triple-concentrated tomato paste samples harvested in 2007. Balanced training (28 samples) and test sets (27 samples) were extracted from the 55 samples of 2007 production by applying D-optimal onion design on PCA scores performed by considering Italian and Chinese samples independently. The O2PLS−DA model let to a very good discrimination between Italian and Chinese samples. A Naïve Bayes (*27*) classifier built using the predictive scores of the model gave a correct geographical origin prediction for both training and test sets (**Table 1**). The related S-plot that allows us to estimate the variable magnitude against its reliability highlighted glucose, γ-aminobutyric acid (GABA), and Ala as the most discriminating variables for Italian concentrated tomato paste samples, while fructose, Gln, and choline were highlighted as the most discriminating variables for Chinese concentrated tomato paste samples, respectively, according to our previous results (*4*) (data not shown). To evaluate whether the production year could influence the sample description, a pattern recognition by PCA on 92 candidate samples (of both 2007 and 2008 productions) was performed, with the first three PCs accounting for R^2X = 74% and Q^2 = 67.4%. The first two PCs led to a clear sample differentiation according to their geographical origin (PCA score plot in **Figure 1A**), while the third latent component (R^2X = 10.3%) routed a possible sample discrimination according to the production year for each country separately considered (PCA score plot in **Figure 1B**).

To better assess year markers for Italian and Chinese concentrated tomato paste samples, two O2PLS−DAs were performed with two classes (2007 and 2008 samples) for each country individually. The O2PLS−DA model carried out on Chinese samples resulted in one predictive and four orthogonal components (R^2Y = 95%, and Q^2 = 88.7%).

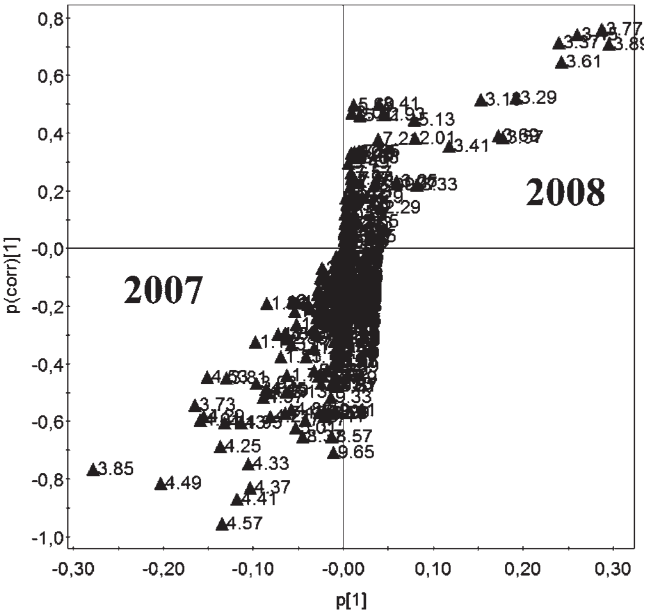

Figure 2. S-plot of O2PLS−DA performed by considering 52 Chinese samples (2007 and 2008) of both certain origin and production year.

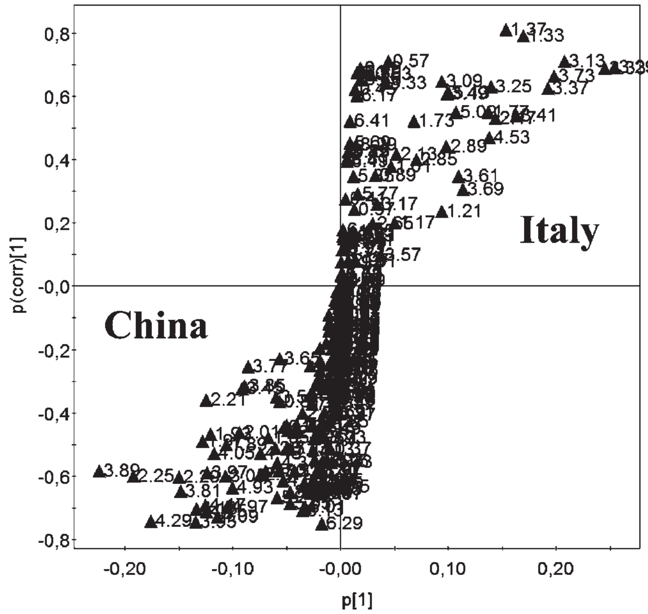

Figure 3. S-plot of O2PLS−DA performed by considering all 92 Chinese and Italian concentrated tomato paste samples of certain production year (2007 and 2008).

This model revealed that single variables could be used to distinguish 2007 and 2008 samples (**Figure 2**), thus suggesting the presence of strong markers, such as β-glucose (buckets at 4.57, 4.53, and 4.49 ppm) and unknown compound (buckets at 4.33, 4.29, and 4.25 ppm) for 2007 Chinese samples and fructose (buckets at 3.89, 3.77, 3.69, 3.57, and 3.45 ppm) for 2008 samples. Interestingly, the bucket at 3.85 ppm, characterizing 2007 samples, included a residual fructose signal and an additional new component, not yet recognized and currently under investigation.

Conversely, the use of O2PLS−DA on Italian samples resulted in one predictive and six orthogonal components (R^2Y = 97.4%, and Q^2 = 88.7%). This latter model revealed that only the combination of several variables allowed sample separation (data not shown); consequently, it was not possible to highlight single variables able to discriminate the production year. Anyhow, for both Italian and Chinese samples, it appeared that the effect of the production year could be evaluated using the information contained into our data set.

On the basis of these results, it was interesting to evaluate to which extent the information about the production year could affect the geographical discrimination. PCA scores performed on the two previous data sets (Chinese samples of 2007 and 2008 and Italian samples of 2007 and 2008) were used to represent the observations space and to sample balanced training and test sets with D-optimal onion design. These sets were characterized by 46 samples for the training set (13 of 2007 and 7 of 2008 for Italian samples and 14 of 2007 and 12 of 2008 for Chinese samples) and 46 samples for the test set (15 of 2007 and 5 of 2008 Italian samples and 13 of 2007 and 13 of 2008 Chinese samples). The training set was used to build a O2PLS−DA model with two classes (Italian and Chinese samples); this model resulted in one predictive latent component and four orthogonal components, with R^2Y = 95.9% and Q^2 = 83.5%. To assess class membership, a Naïve Bayes (*23*) classifier was built using the predictive scores. The class was selected using the highest posterior probability value as the decision rule. Only 1 of 92 samples was "misclassified", thus confirming the model goodness (data not shown). The S-plot of **Figure 3** showed the most relevant variables affecting sample differentiation between the two classes; in particular, Chinese samples were characterized by a higher amount of Gln (buckets at 2.25 and 2.29 ppm), fructose (buckets at 3.89 and 3.97 ppm), and Glu (bucket at 1.97 ppm), while Italian samples were higher in β-glucose (buckets at 4.53, 3.73, 3.41, 3.37, 3.33, 3.29, 3.25, and 3.13 ppm), GABA (buckets at 2.17 and 1.77 ppm), and Ala (buckets at 1.37 and 1.33 ppm).

O2PLS−DA performed on only 2007 or both 2007 and 2008 concentrated tomato paste samples indicated the same discriminant metabolites for Italian and Chinese samples. This result suggested that the two models were comparable from a qualitative point of view. A deeper analysis of this O2PLS−DA model was performed by considering the orthogonal space and the residuals. The predictive part of the model was subtracted from the data set, and the residual data were analyzed by PCA analysis. When the first two PCs were scored (**Figure 4**), a clear-cut sample differentiation according to the production year was evident. On the other hand, the predictive part of the O2PLS−DA model was not able to distinguish the production year (the p value for type I error was equal to 0.74). In other words, information about the production year appeared to be uncorrelated to the predictive space useful to model the geographical origin. This point could be explained in a more rigorous way by studying the relationships between the predictive part of the models useful to distinguish the production year, and the predictive or orthogonal part of the models, to obtain geographical discrimination. The O2PLS technique was thus applied to integrate these parts to quantify the amount of their joint co-variation. Considering the subspace spanned by the Chinese samples, the predictive block of the model for the production year for Chinese samples seemed to be only weakly correlated with the predictive part of the model for the origin discrimination (less than 10% of the total variance), while its variance was completely explained by the orthogonal part of this model (more than 90%). The predictive part of the model for the production year for Italian samples was conversely uncorrelated to the predictive part of the model for the geographical discrimination, while its orthogonal part could be used to completely model the variability produced by different production years.

On the basis of the previous results, the geographical origin of both candidate and validation sample sets was predicted to check the robustness of the classifier built by considering only samples

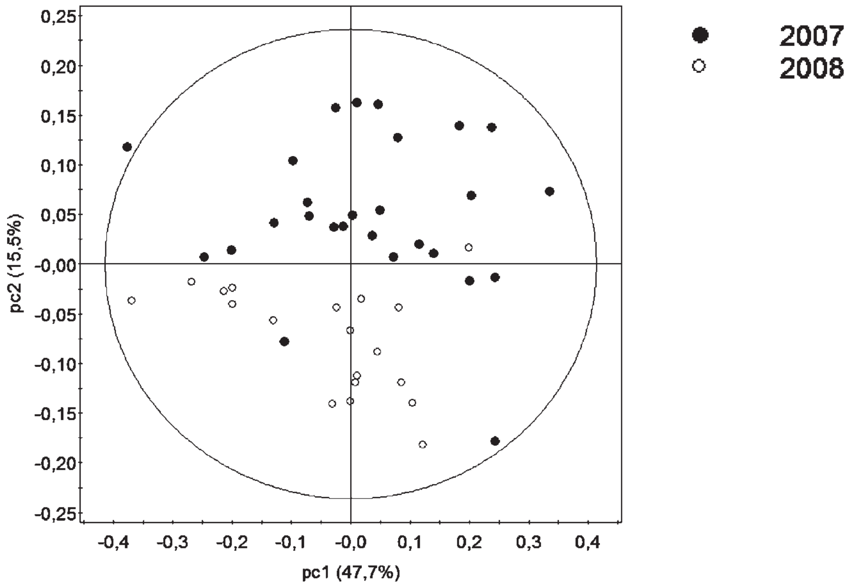

Figure 4. PCA score plot performed on 46 training set samples by considering only residuals and the orthogonal latent space. PC1 = 47.7%, and PC2 = 15.5%. R^2X = 63.2%, and Q^2 = 52.1%.

produced in 2007. A correct classification for more than 95% of all samples was obtained (**Table 1**).

In conclusion, this study showed how the use of NMR spectroscopy and multivariate analysis was well-suited in facing geographical origin determination of concentrated tomato paste. Initially, data analysis by means of an unsupervised PCA method indicated the possibility to obtain information on both the production year and geographical origin of samples. The use of a more powerful statistical procedure, such as O2PLS−DA, allowed us to evaluate distinctively the predictive and orthogonal components of the model, highlighting how the sample variation because of the production year was negligible and dwells mostly in the orthogonal space of the geographical discrimination model. Notwithstanding, some buckets/variables could contribute to both geographical and year discrimination; the final effect of the use of the complete ^{1}H NMR spectra data was to make the geographical and year effect independent of each other. The initial O2PLS−DA performed by considering only concentrated tomato paste samples of 2007 could therefore be applied for geographical origin prediction of all samples.

Our results suggested the possibility of a clear differentiation between Chinese and Italian concentrated tomato paste samples by means of ^{1}H NMR spectroscopy in combination with multivariate statistical data analysis. Interestingly, this sample differentiation was feasible independently from both the concentration rate of samples (double- and triple-concentrated tomato paste) and the tomato production year, at least for 2007 and 2008. Furthermore, NMR was confirmed to be a very useful tool in food characterization and authentication; the importance of detecting several compounds in a single experiment is crucial for sample differentiation.

LITERATURE CITED

(1) Gazzetta Ufficiale number 57, March 9, **2006**. Passata di pomodoro. Origine del pomodoro fresco. Ministero delle Politiche Agricole e Forestali, Feb 17, **2006**.

(2) Luykx, D. M. A. M.; van Ruth, S. M. An overview of analytical methods for determining the geographical origin of food products. *Food Chem.* **2008**, *107*, 897–911.

(3) Arvanitoyanns, J. S.; Vaitsi, O. B. A review on tomato authenticity: Quality control methods in conjunction with multivariate analysis (chemometrics). *Crit. Rev. Food Sci.* **2007**, *47*, 675–699.

(4) Consonni, R.; Cagliani, L. R.; Stocchero, M.; Porretta, S. Triple concentrated tomato paste: Discrimination between Italian and Chinese products. *J. Agric. Food Chem.* **2009**, *57*, 4506–4513.

(5) Chong, H. H.; Simsek, S.; Reuhs, B. L. Analysis of cell-wall pectin from hot and cold break tomato preparations. *Food Res. Int.* **2009**, *42*, 770–772.

(6) Gazzetta Ufficiale number 232, Sept 1, **1975**. Decreto del Presidente della Repubblica, April 11, **1975**; number 428, articolo 1.

(7) Toor, R. K.; Savage, G. P.; Lister, C. E. Seasonal variations in the antioxidant composition of greenhouse grown tomatoes. *J. Food Compos. Anal.* **2006**, *19*, 1–10.

(8) Raffo, A.; La Malfa, G.; Fogliano, V.; Maiani, G.; Quaglia, G. Seasonal variations in antioxidant components of cherry tomatoes (*Lycopersicon esculentum* cv. Naomi F1). *J. Food Compos. Anal.* **2006**, *19*, 11–19.

(9) Incerti, A.; Navari-Izzo, F.; Pardossi, A.; Izzo, R. Seasonal variations in polyphenols and lipoic acid in fruits of tomato irrigated with sea water. *J. Sci. Food Agric.* **2009**, *89*, 1326–1331.

(10) Zanfini, A.; Dreassi, E.; La Rosa, C.; D'Addario, C.; Corti, P. Quantitative variations of the main carotenoids in Italian tomatoes in relation to geographic location, harvest time, varieties and ripening stage. *Ital. J. Food Sci.* **2007**, *19*, 181–190.

(11) Raffo, A.; Leonardi, C.; Fogliano, V.; Ambrosino, P.; Salucci, M.; Gennaro, L.; Bugianesi, R.; Giuffrida, F.; Quaglia, G. Nutritional value of cherry tomatoes (*Lycopersicon esculentum* cv. Naomi F1) harvested at different ripening stages. *J. Agric. Food Chem.* **2002**, *50*, 6550–6556.

(12) Anza, M.; Riga, P.; Garbisu, C. Effects of variety and growth season on the organoleptic and nutritional quality of hydroponically grown tomato. *J. Food Qual.* **2006**, *29*, 16–37.

(13) Slimestad, R.; Verheul, M. J. Seasonal variations in the level of plant constituents in greenhouse production of cherry tomatoes. *J. Agric. Food Chem.* **2005**, *53*, 3114–3119.

(14) Islam, M. S.; Khan, S. Changes in quality characteristics of three tomato cultivars maturing at seven different sowing times. *Trop. Agric.* (*Trinidad*) **2000**, *77*, 236–243.

(15) Chassy, A. W.; Bui, L.; Renaud, E. N. C.; Van Horn, M.; Mitchell, A. E. Three-year comparison of the content of antioxidant microconstituents and several quality characteristics in organic and conventionally managed tomatoes and bell peppers. *J. Agric. Food Chem.* **2006**, *54*, 8244–8252.

(16) Iniesta, M. D.; Pérez-Conesa, J.; Garcia-Alonso, J.; Ros, G.; Periago, M. J. Folate content in tomato (*Lycopersicon esculentum*). Influence of cultivar, ripeness, year of harvest, and pasteurization and storage temperatures. *J. Agric. Food Chem.* **2009**, *57*, 4739–4745.

(17) Consonni, R.; Cagliani, L. R. Nuclear magnetic resonance and chemometrics to assess geographical origin and quality of traditional food products. *Adv. Food Nutr. Res.* **2010**, *59*, DOI: 10.1016/S1043-4526(10)59004-1.

(18) Musse, M.; Quellec, S.; Cambert, M.; Devaux, M. F.; Lahaye, M.; Mariette, F. Monitoring the postharvest ripening of tomato fruit using quantitative MRI and NMR relaxometry. *Postharvest Biol. Technol.* **2009**, *53*, 22–35.

(19) Deborde, C.; Maucourt, M.; Baldet, P.; Bernillon, S.; Biais, B.; Talon, G.; Ferrand, C.; Jacob, D.; Ferry-Dumazet, H.; De Daruvar, A.; Rolin, D.; Moing, A. Proton NMR quantitative profiling for quality assessment of greenhouse-grown tomato fruit. *Metabolomics* **2009**, *5*, 183–198.

(20) Moco, S.; Forshed, J.; De Vos, R. C. H.; Bino, R. J.; Vervoort, J. Intra- and inter-metabolite correlation spectroscopy of tomato metabolomics data obtained by liquid chromatography−mass spectrometry and nuclear magnetic resonance. *Metabolomics* **2008**, *4*, 202–215.

(21) Le Gall, G.; Colquhoun, I. J.; Davis, A. L.; Collins, G. J.; Verhoeyen, M. E. Metabolite profiling of tomato (*Lycopersicon esculentum*) using ^{1}H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *J. Agric. Food Chem.* **2003**, *51*, 2447–2456.

(22) Trygg, J.; Wold, S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* **2002**, *16*, 119–128.

(23) Trygg., J.; Wold, S. O2-PLS, a two-block ($X-Y$) latent variable regression (LVR) method with an integral OSC filter. *J. Chemom.* **2003**, *17*, 53–64.

(24) Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J. OPLS discriminant analysis: Combining the strengths of PLS−DA and SIMCA classification. *J. Chemom.* **2006**, *20*, 341–351.

(25) Wiklund, S.; Johansson, E.; Sjöström, L.; Mellerowicz, E. J.; Edlund, U.; Shockcor, J. P.; Gottfries, J.; Moritz, T.; Trygg, J. Visualization of GC/TOF−MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Anal. Chem.* **2008**, *80*, 115–122.

(26) Olsson, I. M.; Gottfries, J.; Wold, S. D-optimal onion designs in statistical molecular design. *Chemom. Intell. Lab. Syst.* **2004**, *73*, 37–46.

(27) Langseth, H.; Nielsen, T. D. Classification using Hierarchical Naïve Bayes models. *Mach. Learn.* **2006**, *6*, 135–159.

Received for review March 12, 2010. Revised manuscript received May 11, 2010. Accepted May 19, 2010. The Scientific Foundation A. De Marco is acknowledged for financial support.