Original Article

Elemental profile and oxygen isotope ratio (δ^{18}O) for verifying the geographical origin of Chinese wines

Shuangxi Fan [a,b], Qiding Zhong [b,], Hongbo Gao [b], Daobing Wang [b], Guohui Li [b], Zhanbin Huang [a]*

[a] *School of Chemical and Environmental Engineering, China University of Mining and Technology (Beijing), Ding No. 11, Xueyuan road, Haidian District, Beijing, 100083, China*
[b] *China National Institute of Food and Fermentation Industries, Building 6, No. 24 Jiuxianqiao middle road, Chaoyang District, Beijing, 100015, China*

A B S T R A C T

The elemental profile and oxygen isotope ratio (δ^{18}O) of 188 wine samples collected from the Changji, Mile, and Changli regions in China were analyzed by inductively coupled plasma mass spectrometry (ICP-MS), inductively coupled plasma optical emission spectroscopy (ICP-OES) and isotope ratio mass spectrometry (IRMS), respectively. By combining the data of δ^{18}O and the concentration data of 52 elements, the analysis of variance (ANOVA) technique was firstly applied to obtain the important descriptors for the discrimination of the three geographical origins. Ca, Al, Mg, B, Fe, K, Rb, Mn, Na, P, Co, Ga, As, and Sr, and δ^{18}O were identified as the key explanatory factors. In the second step, the key elements were employed as input variables for the subsequent partial least squares discrimination analysis (PLS-DA) and support vector machine (SVM) analyses. Then, cross validation and random data splitting (training set: test set = 70:30, %) were performed to avoid the over-fitting problem. The average correct classification rates of the PLS-DA and SVM models for the training set were both 98%, while for the test set, these values were 95%, 97%, respectively. Thus, it was suggested that the combination of oxygen isotope ratio (δ^{18}O) and elemental profile with multi-step multivariate analysis is a promising approach for the verification of the considered three geographical origins of Chinese wines.

1. Introduction

Many countries and geographical areas are well known for producing high quality wines, underpinning that the geographical origin of a wine is being of great commercial value [1]. After years of development, China has established several important and distinctive geographical origins for wine production such as the Changji, Mile, and Changli regions. However, numerous economical profit-driven fraudulent events have been reported in the Chinese wine

industry concerning the false declaration of geographical origins, thereby significantly damaging the reputation of the Chinese wine market. Unfortunately, there is still no effective official method or national standard to verify the geographical origin of wine produced in China. As such, significant effort has been directed to develop a scientific method for the characterization of Chinese wines according to their geographical origin, which would be particularly beneficial to individual producers, to the Chinese wine industry and to control authentication.

To date, studies concerning the geographical origins of wines produced in China focused mainly on multi-element analysis by inductively coupled plasma mass spectrometry (ICP-MS) coupled with appropriate chemometric techniques [2,3]. In addition, only a relatively low number of Chinese wine samples have been analyzed using site-specific natural isotopic fractionation-nuclear magnetic resonance (SNIF-NMR), isotope ratio mass spectrometry (IRMS) and ^{1}H NMR techniques [4–6]. These previous investigations were laboratory researches with limited sample sizes in China. In addition, although various wine databases in different parts of the world have been established [7], these do not include Chinese wines so far.

In the past few decades, a number of studies for assessing the geographical origin of wines have been reported abroad [8–11]. For example, the analysis of different isotope ratios has been employed, including $^{88}Sr/^{86}Sr$ [12–14] and $^{11}B/^{10}B$ [15] analyses using ICP-MS, and $^{13}C/^{12}C$ and $\delta^{18}O$ analyses in wine ethanol and wine water, respectively, using IRMS [16–19]. Other approaches, such as Fourier transform infrared spectroscopy (FTIR), near infrared spectroscopy (NIR), and nuclear magnetic resonance (NMR), are also suitable [4,20–22].

To date, various studies have demonstrated that multi-elemental trace analysis methods based on ICP-MS are useful to verify wine samples from different origins in Romanian [23,24], Canada [25,26], South Africa [27–29], Spain [30–32], Germany [33,34], Portugal [35], New Zealand [36], the Czech Republic [37], Argentina [38], Portugal, and France [12]. Indeed, the verification of the different geographical origins of wines by multi-element multivariate analysis is demonstrated in Table 1.

A combination of chemical composition determination with subsequent multivariate data analysis can also be regarded as a profiling approach to classify products based on their origin, variety, or other properties [39]. The most common multivariate analysis methods for such verification purposes include principal component analysis (PCA) and hierarchical clustering analysis (HCA) for unsupervised grouping, and discriminant analysis (DA) to select important variables for discrimination [11]. Furthermore, we wish to investigate the non-linear support vector machines (SVM) model and the linear partial least squares discrimination analysis (PLS-DA) technique in our study, as these methods have rarely been used to authenticate the geographical origins of Chinese wines.

In this study, we herein report 52 trace elements and oxygen isotope ratio ($\delta^{18}O$) of water in 188 wine samples collected from the Changji, Mile, and Changli regions in China were determined by ICP-MS, ICP-OES and IRMS, respectively. Multivariate statistical techniques were used to extract significant information relating to the geographical origin of the wine samples based on oxygen isotope ratio ($\delta^{18}O$) and elemental profile. Initially, analysis of variance (ANOVA) was employed to determine the key trace elements required to discriminate between different geographical origins. Subsequent multivariate analysis (PLS-DA and SVM) was then carried out for classification of the wine samples.

2. Materials and methods

2.1. Chemicals and reagents

All reagents used in this study were of analytical grade or better. The HNO3 and ethanol employed were of chromatographic grade (Sigma Aldrich, USA). Ultrapure water (resistivity, 18.2 MΩ cm) was obtained from a Milli-Q system from Millipore (Milford, MA, USA). The microwave digestion system was purchased from CEM MARS, USA. Prior to use, all glassware was washed by immersing in a 20% (v/v) HNO3 solution over 24 h, then rinsing four times with ultrapure water.

The chemicals and regents used for IRMS analysis were described in detail in our previous work for the rapid determination of the stable oxygen isotope ratio ($\delta^{18}O$) of water in alcoholic beverages [40].

2.2. Wine samples

A total of 188 wine samples were collected from three distinctive wine-producing regions in China, namely the Changji (90 samples), Mile (41 samples), and Changli (57 samples) regions. Five grape varieties of red wines (Cabernet Sauvignon, Rose Honey, Merlot, Pinot Noir, Cabernet Gernischt) and four grape varieties of white wines (Long Yan, Chardonnay, Crystal, Riesling) were examined. All wine samples were stored at 4 °C and in the absence of light. To guarantee the authenticity of the samples, all wine samples were taken from different wine-producing manufacturers, collected from the vintage years of 2010, 2011, 2012, 2013 and 2014. Each wine sample was obtained from one variety only and all wine samples met the requirements of the executive national standard of China [41]. The wine samples grown in the three geographical origins are outlined in Table 2. Different wine varieties from the Changji and Mile regions were collected, which covered the major grape varieties in the different geographical origins.

2.3. Sample preparation for ICP-MS and ICP-OES analysis

The organic components in the wine samples were removed using a microwave digestion method. More specifically, the wine sample (5 mL) was transferred into the digestion tube and HNO$_3$ (5 mL) was added. The resulting mixture was then transferred into the microwave and digestion was carried out according to the following temperature program: 120 °C for 5 min, 160 °C for 5 min, and 180 °C for 15 min. After cooling to 25 °C, the mixture was transferred into a 50 mL volumetric flask. The digestion tubes were rinsed four times with ultrapure water, and this water was transferred into the volumetric

Table 1 — ICP-MS method presented in the literature for the verification of geographical origin of wines.

Elements employed	Chemometric methods	Geographical origin	References
Cr, Sr, Rb, Ni, Ag, Cu, Co, Be, V, Pb, Zn, Mn	PCA	Romanian	[23]
As, Cd, Cs, Li, Te, Zr, Mo, Ni, Sb, Ti, U, Y, REEs, Nb, Rb, Be, Co, Ga, TI, W	PCA	Germany	[34]
Al, Cd, V, Ba, Li, Ni, Co, Pb, Sb	PCA and LDA	Spain	[31]
Cs, Ga, Ni, Pb, Rb, Sr, Cd, Co, Mn	LDA	New Zealand	[36]
Bi, Sb, Fe, Mo, Ni, As, Ba, Th, Cs, Cu, Rb, Al, B, Ti, TI, Br, Cd, Co, Se, Sr, P, Pb, U, Ag, Mn, Cl, V, Ca, Ce, Zn, La, Li, Mg	PCA and DA	Canada	[26]
B, Nb, Se, Si, TI, U, Cs, Cl, Mn, Ga, Li, Sr, Ni, Ba, Rb, Sc, W, Mg, La, Al	DA	South Africa	[27]
REEs, Au, Pd, Sb, Zr, Ni, Pb, Co, Cu, Re, Ti, TI, Cd, Ga, Li, Pt, Rb, Sr, Te, V, W, Sn, Cs, As, Ba, Be	SIMCA	Spain (Canary Islands)	[32]
Zn, As, Ba, Co, Li, V, Ni, Sr, Pb, Mo, Rb, Cd, Cu, La, U, Bi, Th, Cs, Ce	MDS	Canada	[25]
Ca, Sr, Mg, Cs, V, Li, Rb, Zn, Co, Mn, B, Fe, Pb	DA	Germany	[33]
Cs, Ag, Z, Ba, Rb, Li, Cu, Cd, Al, Sb, As, V, Ni, Be, Sr, Ti, U, Pb, Co, Cr	PCA, HCA and FA	Czech Republic	[37]
Al, REEs, Y, Hf, B, Sc, Sr, Co, Cr, Cs, Fe, Mn, Mo, Ca, Zn, W, V, U, Cd, Ni, Th, Ti, TI, Ga, Li, Cu, Rb, Nb, Be, As, Ba, Pb, Sb	QDA	Portugal	[35]
Tl, Li, Se, Rb, La, Ga, Cl, Sc, Nb, Cs, Mg, Al, U, Sr, Ba, W, B, Si, Mn, Ni	DA	South Africa	[29]
Li, Be, V, Mn, Co, Ni, Cu, Ge, As, Rb, Sr, Mo, Cd, Ba, Hg, TI, Pb, Bi	PCA and LDA	Argentina	[38]
$^{11}B/^{10}B$	—	South Africa	[27]
$^{87}Sr/^{86}Sr$	—	Portugal and France	[35]

PCA: Principal component analysis, HCA: hierarchical clustering analysis, LDA: linear discriminant analysis, DA: discriminant analysis, SIMCA: soft independent modeling class analogy, QDA: quadratic discriminant analysis, MDS: multidimensional scaling, FA: factor analysis.

Table 2 — Wine samples measured in three geographical origins (Changli, Xinjiang and Yunnan).

Geographical origin	Grape variety								
	Cabernet Sauvignon	Riesling	Pinot Noir	Merlot	Cabernet Gernischet	Chardonnay	Longyan	Crystal	Rose honey
Chanli	57	—	—	—	—	—	—	—	—
Changji	51	7	10	8	10	4	—	—	—
Mile	—	—	—	—	9	—	2	19	11

flask and diluted with ultrapure water to reach the desired final volume of 50 mL. A blank calibration sample which was made of HNO_3 and water (2:98, v/v) and a quality control sample were also prepared using the above procedure.

2.4. Sample preparation for IRMS analysis

The sample preparation procedures for IRMS analysis were described in detail in our previous work [40].

2.5. IRMS analysis

GC-P-IRMS system and analytical conditions for IRMS analysis were the same as our previous work [40].

2.6. ICP-MS and ICP-OES analysis

Experimental measurements were carried out using an iCAP-6300 ICP-OES spectrometer (Thermo Fisher Scientific, USA) and an iCAP Q ICP-MS instrument (Thermo Fisher Scientific, USA). Following the preparation of each wine sample, all measurements were carried out in triplicate and the results were averaged. Detailed instrumental settings regarding the optimized operating conditions for ICP-MS and ICP-OES analyses are presented in Tables 3 and 4.

Table 3 — Optimized ICP-MS operating parameters for the determination of trace elements.

Instrument parameters	Condition
RF power	1500 W
Coolant gas	0 L min^{-1}
Carrier gas	1.17 L min^{-1}
Nebulizer pump	0.1 rsp
Integration time per point	0.1 s (As: 1s, Cd: 1s, Se: 1s)
Scanning mode	Peak jumping
Observation point/peak	3
Sampling depth	8.0 mm
Replicates measured	3
Isotopes measured	^{7}Li, ^{9}Be, ^{47}Ti, ^{51}V, ^{52}Cr, ^{59}Co, ^{60}Ni, ^{63}Cu, ^{69}Ga, ^{72}Ge, ^{75}As, ^{82}Se, ^{88}Sr, ^{89}Y, ^{90}Zr, ^{93}Nb, ^{95}Mo, ^{111}Cd, ^{120}Sn, ^{121}Sb, ^{125}Te, ^{138}Ba, ^{140}Ce, ^{141}Pr, ^{146}Nd, ^{147}Sm, ^{153}Eu, ^{157}Gd, ^{159}Tb, ^{163}Dy, ^{165}Ho, ^{166}Er, ^{169}Tm, ^{172}Yb, ^{175}Lu, ^{205}Tl, ^{206}Pb, ^{209}Bi

External calibrating solutions (0, 1, 2.5, 5, 10, and 20 µg/L) for ICP-MS analysis were prepared by combining the appropriate dilution of a 100 mg/L multi-element stock standard solution (Analytical grade, General Research Institute for Nonferrous Metals, Beijing, China) containing As, Be, Dy, Co,

Table 4 – Optimized ICP-OES operating parameters for the determination of trace elements.

Instrument parameters	Condition
RF power	1150 W
Plasma flow	15 L/min
Coolant gas	1.5 L/min
stabilization time	15 s
Washing time between samples	30 s
Observation time	3 s
Sample uptake rate	1.5 mL/min
Replicates measured	3
Elements measured and wavelengths (nm)	Mg (285.2), Al (396.1), Na (589.5), Si (251.6), P (213.6), K (769.8), Ca (317.9), Mn (257.6), Fe (259.9), Zn (206.2), Rb (780.0), B (249.7), P (178.7), Sr (421.5)

Fig. 1 – The three Chinese wine-growing regions in a map.

Er Cr, Ba, Cu, Nd, Li, V, Ni, Y, Cd, Pb, Ho, Sb, Tl, Sn, Yb, Sr, Ti, Ce, Gd, Pr, Eu, Tb, Tm, Lu, Sm, Ga, Zr, Nb, Te, and Bi, and a 1 g/L multi-element stock standard solution (Analytical grade, General Research Institute for Nonferrous Metals, Beijing, China) containing Ge, Mo, and Se.

External calibrating solutions (0, 1, 2.5, 5, 10, and 25 mg/L) for ICP-OES analysis were prepared by combining the appropriate dilution of a 100 mg/L multi-element stock standard solution (Analytical grade, General Research Institute for Nonferrous Metals, Beijing, China) containing B, Al, Zn, Sr, Cu, Fe, Mg, and Mn, and single standard element solutions (100 mg/L) of P, Rb, Si, Ca, K, and Na (Analytical grade, General Research Institute for Nonferrous Metals, Beijing, China). Blank solutions and working standard solutions were prepared by adding 2% (v/v) HNO_3 and 8% (v/v) ethanol. To correct for instrumental drift, In (2 μg/L) was used as an internal standard. Weighted regression of the calibration curves was employed to quantify the concentrations of trace elements.

A total of 188 wine samples from three different regions of China were investigated by ICP-MS, ICP-OES and IRMS, respectively. Trace analyses were performed by these methods to determine elemental profile and oxygen isotope ratio ($\delta^{18}O$) of water in the wine samples.

2.7. Statistical analysis

To facilitate data processing, all wine samples belonging to the same region were assigned to following codes (CJ for Changji, ML for Mile, and CL for Changli). The geographical locations of the three different Chinese wine-growing regions investigated are herein shown in Fig. 1.

Multivariate data analysis was then performed, where one-way ANOVA was employed to detect any significant differences between the wine samples grouped by their indicated origins [42]. Prior to the ANOVA test in which the level of significance was set at $p < 0.05$, Levene's test and the Shapiro–Wilk test were applied that a typical procedure required for performing ANOVA to assess the homogeneity and normality of variances. Fisher's least significant difference test was also implemented to check for differences between the different groups in the data matrix. The key elements for determining the geographical origins of the samples were identified by
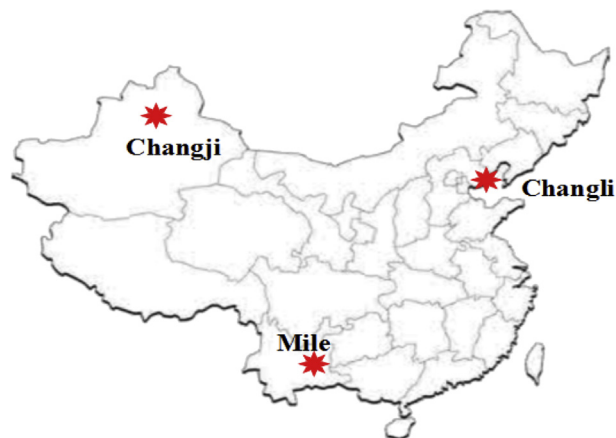
ANOVA in a first step, and were employed as input variables for the subsequent PLS-DA and SVM analyses. The supervised PLS-DA and SVM methods were employed for classification analysis. Before all performed techniques, the data were standardized to zero mean and unit variance using R version 3.2.3 (R Core Team, Vienna, Austria) statistical software.

Creation and evaluation of the classification model were then performed by using splitting data sets: (1) the training set (132 samples, 70%) for building classification model with internal cross validation, and (2) the test set (56 samples, 30%) for external validation. The stratified random function was used for data splitting by R software using the caret package. The SPSS software version 19.0 (SPSS Inc., Chicago, USA) and R software were used for statistical analysis. More specifically, ANOVA analysis was carried out using SPSS, while PLS-DA was carried out by R using the packages: mixOmics [43], mdatools [44], while SVM was performed using the e1071 package [45].

3. Results and discussion

3.1. Quality control

The quality control for IRMS analysis was consistent with our previous work [40]. For analysis by both ICP-MS and ICP-OES, a 3-fold standard deviation of 10 replicate tests of the blank solution was used to estimate the analytical limit of detection (LOD) for each element [38]. The LODs of all measured elements are shown in Table 5. As the concentrations of several elements (Zr, Nb, Te, Bi, and Ge) in all the samples were mostly below their LODs, these elements were not used for subsequent multivariate data analysis. Therefore, the final data matrix contains 188 rows (wine samples) and 48 columns (47 elements and $\delta^{18}O$) (adding up in 9024 data points). The accuracy and reproducibility of the two methods (i.e., ICP-MS and ICP-OES) were also assessed. More specifically, the accuracy was checked by performing recovery experiments, where external calibrating solutions (0.2, 0.4, and 0.8 mg/L) for ICP-OES and external calibrating solutions (2.5, 5, and 10 μg/L) for ICP-MS were added to the quality control sample. The spiked wine samples were then

Table 5 – Analytical characteristics (LOD, recovery) of the determination of elements of wines by ICP-OES and ICP-MS.

Analytical method	Element	LOD (µg/L)	Recovery (%)[a]	Element	LOD (µg/L)	Recovery (%)[a]
ICP-MS	^{7}Li	0.8	120.5 ± 2	^{118}Sn	1.7	104.8 ± 2
	^{9}Be	1.0	120 ± 3	^{125}Te	0.9	93.6 ± 3
	^{47}Ti	6.8	103.5 ± 5	^{137}Ba	1.2	103.7 ± 0.5
	^{52}Cr	0.4	98.5 ± 5	^{140}Ce	0.4	104.5 ± 0.4
	^{51}V	0.2	115.7 ± 1	^{141}Pr	1.8	103.9 ± 2
	^{53}Cr	7.2	106.1 ± 2	^{146}Nd	1.3	94.1 ± 1
	^{59}Co	2.4	105.4 ± 4	^{147}Sm	0.4	114 ± 2
	^{60}Ni	1.6	98.4 ± 4	^{153}Eu	1.8	105.3 ± 3
	^{121}Sb	8.4	101.6 ± 3	^{157}Gd	1.3	107.5 ± 0.9
	^{69}Ga	1.6	106.1 ± 1	^{159}Tb	0.4	105.1 ± 0.8
	^{72}Ge	0.4	93 ± 4	^{163}Dy	1.8	116.4 ± 2
	^{75}As	0.2	110.4 ± 3	^{165}Ho	1.3	104.9 ± 1
	^{82}Se	5.8	99.4 ± 1	^{166}Er	0.4	103.9 ± 3
	^{88}Sr	2.3	112.6 ± 2	^{169}Tm	1.8	103.5 ± 0.7
	^{89}Y	0.7	106 ± 0.6	^{172}Yb	1.3	102.7 ± 2
	^{90}Zr	2.6	117.7 ± 2	^{175}Lu	0.4	104.6 ± 1
	^{93}Nb	1.9	109.6 ± 1	^{205}Tl	1.8	96.6 ± 2
	^{95}Mo	0.7	109.8 ± 2	^{206}Pb	1.3	103.8 ± 4
	^{111}Cd	1.9	101 ± 1	^{209}Bi	1.9	95.6 ± 2
ICP-OES	Na (589.5)	58.6	94.4 ± 1	Mn (257.6)	16.4	97.2 ± 2
	Mg (285.2)	32.8	105.7 ± 3	Fe (259.9)	43.6	111 ± 2
	Al (396.1)	16.2	97.9 ± 1	Zn (206.2)	24.4	102.8 ± 3
	Si (251.6)	56.4	95.5 ± 2	Rb (780.0)	61.2	100.3 ± 1
	P (213.6)	23.7	96.8 ± 0.7	B (249.7)	79.6	101.7 ± 0.8
	K (769.8)	60.2	106.8 ± 2	Cu (327.3)	19.1	112.2 ± 2
	Ca (317.9)	47.1	94.7 ± 2	Sr (421.5)	13.2	96.4 ± 1

[a] Mean ± standard deviation (n = 10).

measured 10 times, and the recoveries of the various elements were calculated. As indicated in Table 5, the recoveries ranged from 78 to 120% for ICP-OES and ranged from 94 to 112% for ICP-MS. Repeatability was then determined by measuring the multi-element concentrations of multiple bottles of the same kind of wine. Two kinds of wine (red and white) were considered, with six bottles of each being obtained. Three bottles of each kind of wine were randomly selected, and these 6 bottles of wine were prepared for the repeatability experiments according to the procedure described in sample preparation section of this study. Each wine sample was analyzed in triplicate. The relative standard deviations (RSDs) of the elemental concentrations in red wine ranged from 6 to 10%, while the RSDs of the elemental concentrations in white wine ranged from 5 to 9%, thereby indicating that the analytical methods employed herein were sufficiently repeatable. As all wine samples were analyzed in a single analysis sequence, the interday repeatability was not examined specifically. However, to ensure the quality of the obtained data, a quality control sample and a blank calibration sample were tested every six samples and monitored carefully.

3.2. Analysis of variance (ANOVA)

Following analysis of the various wine samples from three different regions within China using ICP-OES, ICP-MS and IRMS, ANOVA was employed to identify the key variables from 52 elements and δ^{18}O that may differentiate between the different geographical origins of the wine samples [46]. Based on the ANOVA results, it was figured out that the wine samples from the CJ, ML, and CL regions could be discriminated by their respective Ca, Al, Mg, B, Fe, K, Rb, Mn, Na, P, Co, Ga, As, Sr contents and δ^{18}O values. After application of the ANOVA test, these identified key elements were regarded as reliable descriptors or explanatory factors [47]. The δ^{18}O has strong relationship with the location, which is consistent with the previous investigation of geographical origin for Romanian wines [48]. A list of elements previously described in the literature can be found in Table 1. Statistical analysis of Table 1 shows that elements identified herein were also used as indicative descriptors for geographical origins in a number of previous studies. As indicated in Table 6, Ca was found in three studies. The frequencies of the other descriptors (i.e., Al, Mg, B, Fe, Rb, Mn, P, Co, Ga, As, and Sr) found in previous

Table 6 – The frequencies of important elements used as descriptors in a number of previous studies.

Elements	Ca	Al	Mg	B	Fe	Rb	Mn	P	Co	Ga	As	Sr
Frequencies	3	6	4	6	3	12	8	1	11	6	7	12

ANOVA is not the only one method used in the previous literature to extract important elements for the discrimination of geographical origins, other approaches are also used.
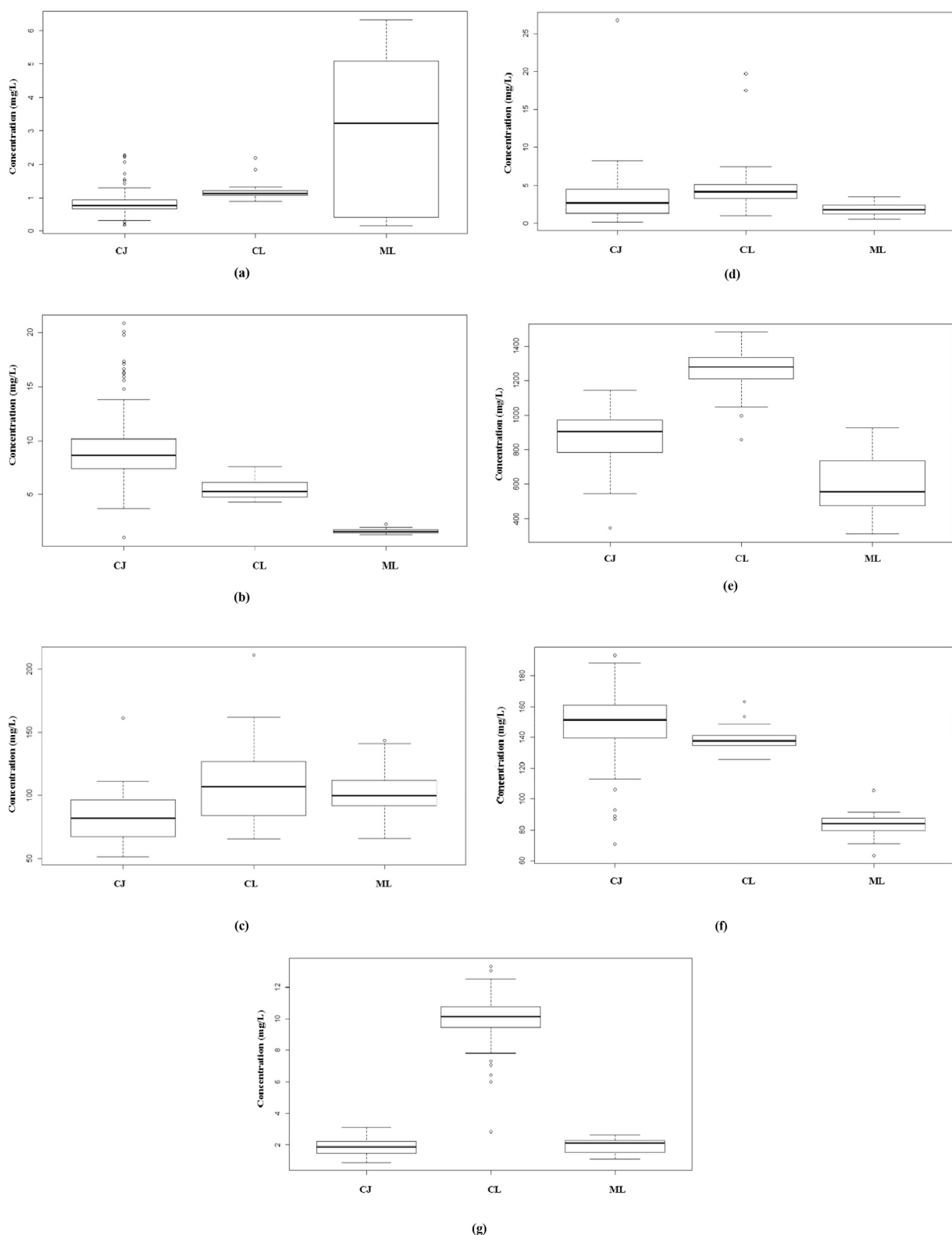
Fig. 2 – Box plot of the concentration of important elements Al (a), B (b), Ca (c), Fe (d), K (e), Mg (f), Mn (g), Na (h), P (i), Rb (j), Co (k), Ga (m), As (n), Sr (o), $\delta^{18}O$ (p) for the discrimination of wines from three geographical origins (Changli (CL), Changji (CJ), Mile (ML)) in China.
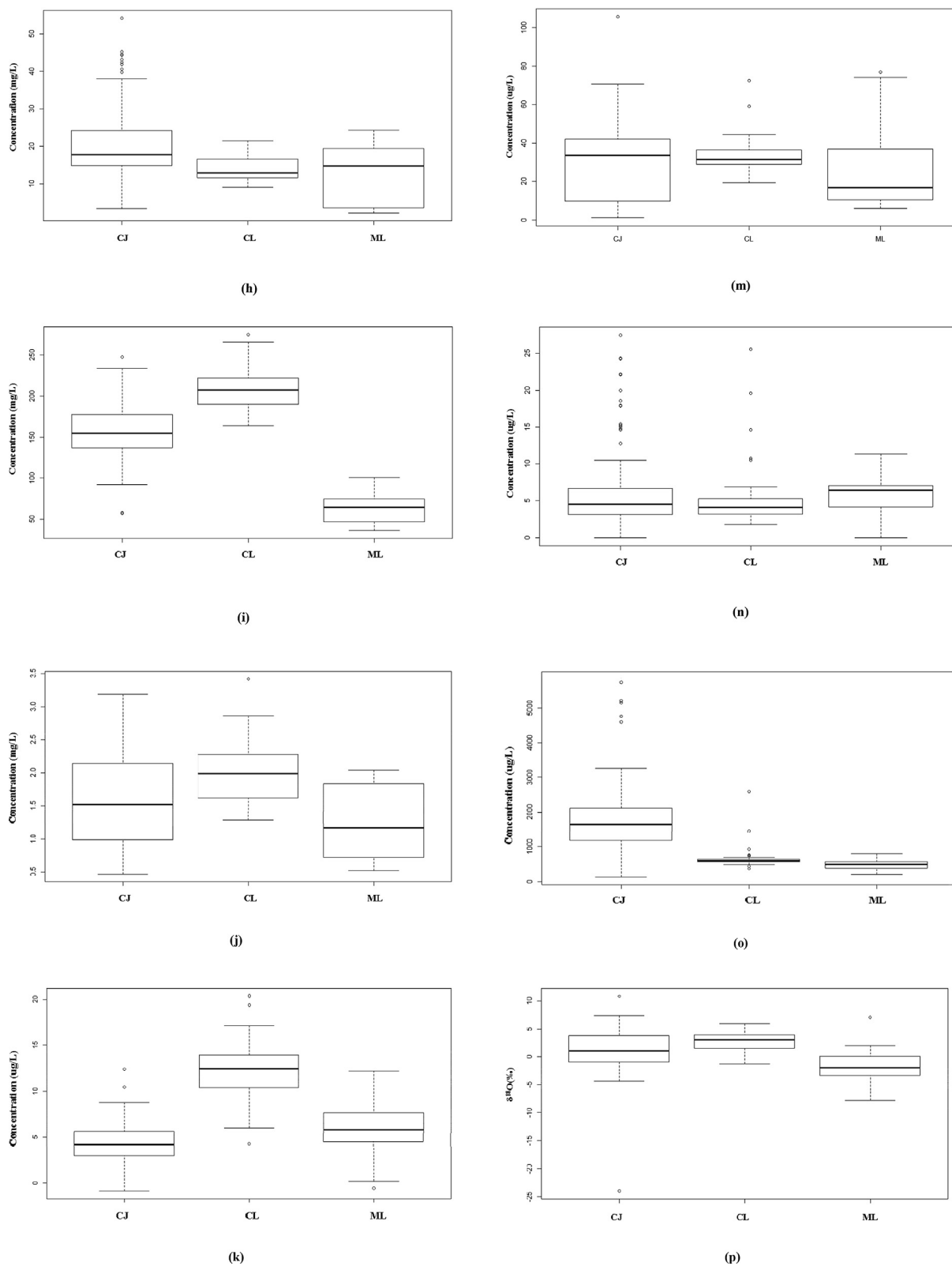
(h)

(m)

(i)

(n)

(j)

(o)

(k)

(p)

Fig. 2 – (continued)

studies were 6, 4, 6, 3, 12, 8, 1, 11, 6, 7, and 12. Furthermore, Table 1 only shows the ICP-MS method presented in the literature; however, as Na and K are usually determined by ICP-OES, these elements have been omitted from the table. Fig. 2 shows a box plot of the concentrations of these key elements found by our present study, where the bottom and top of the box are the first and third quartiles of the data for each geographical origin. From Fig. 2, it is apparent that it was not possible to use each individual element for the discrimination of the different geographical origins. Therefore it is necessary to establish an robust model to accurately discriminate between Chinese wines from different geographical regions.

3.3. Partial least squares discrimination analysis (PLS-DA)

The 14 mineral elements and $\delta^{18}O$ extracted by ANOVA were used as the input variables for subsequent PLS-DA analysis to achieve accurate classification [42,49]. The Bayes method function was selected for the data analysis, as this function has previously been widely used in discrimination analysis. The lowest value of the predicted residual error sum of squares (PRESS) can be calculated by leave-one-out cross validation (LOOCV) and this is subsequently employed to determine the optimum number of latent variables used in the PLS-DA model. The latent variables can then be used for subsequent linear model generation. Based on the cross-validation rule, the first two latent variables (i.e., X-variate 1 and X-variate 2) were used to achieve the optimal prediction accuracy. Fig. 3 demonstrates the score plot of the first two X-variates for the entire data set located in the reduced 2-dimensional space. As indicated, the first two latent variables (X-variate 1: 27% and X-variate 2: 22%) accounted for 49% of the total variance in the raw data, allowing an obvious distinction to be observed between the wine samples from the three different origins. Indeed, PLS-DA is required to maximize the variance between groups rather than within groups,

ultimately providing key criteria for selecting the important variables based on the model prediction ability. The importance of the variables generated from PLS-DA can be classified as follows: Mn > K > Co > P > Rb > Ca > Sr > $\delta^{18}O$ > Fe > Mg > Na > Al > B > Ga > As. Several overlapped wine samples from the CJ and ML regions were also observed. A total of 8 wine samples fell outside the 95% confidence interval (Hotellings T^2 ellipsis) of the respective classes. The reason for these outlier samples may be that the raw data of outlier samples was recorded by artificial errors. The raw data should be checked again. Although a number of samples could not be distinguished in this plane, the overall classification performance figure achieved by PLS-DA was applied for discrimination of the wine samples from different geographical origins in China, as outlined in Table 7, which shows the internal LOOCV classification results for the training set of PLS-DA model. The correct classification rates of the three regions (i.e., CJ, ML, and CL) for the training set were 100, 97, and 97%, respectively, while those for the test set used for external validation were 100, 92, and 94%, respectively. In addition, the average correct classification rates of the PLS-DA model for the training and test sets were 98 and 95%, respectively. From the PLS-DA classification results, it can be concluded that PLS-DA can be considered an efficient method for discriminating between the geographical origins of Chinese wines, although a few samples were misclassified.

3.4. Support vector machine (SVM)

To test a second data evaluation method and verify the misclassified samples from the PLS-DA method, the SVM algorithm was performed using the same training and test sets. Suitable kernel functions and parameters were selected for the SVM model to aid in establishing an optimized model with high stability and prediction accuracy [30,50]. In this study, the radial basis function (RBF) kernel with a Gaussian profile was chosen for non-linear SVM classification analysis [51].
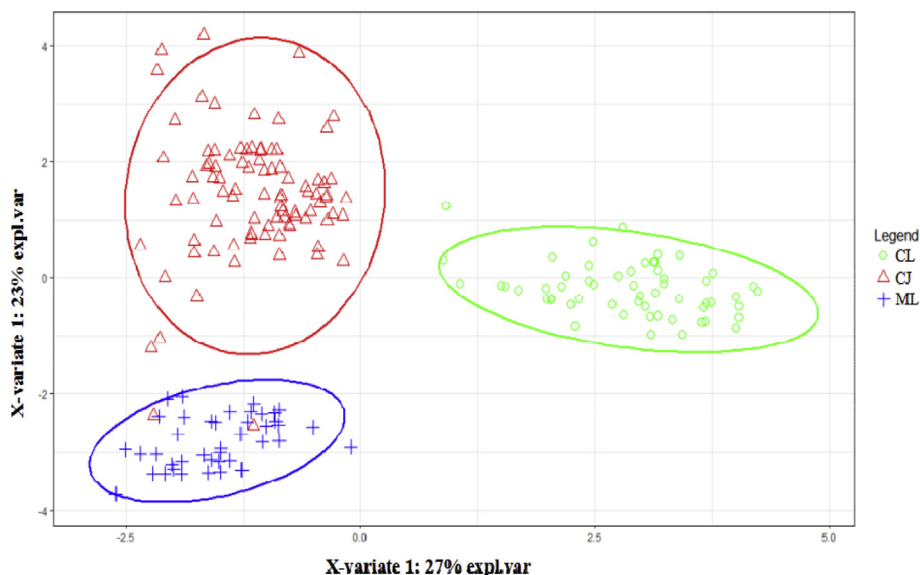

Fig. 3 – Score plot of X-variate 1 vs X-variate 2 for wine samples (188) from three regions (CL, CJ and ML) in China. Solid lines represent the 95% confidence interval (Hotellings T^2 ellipsis).

Data set	Geographical origin	PLS-DA			SVM		
		MS[a]	P (%)[b]	P (%)[c]	MS[a]	P (%)[b]	P (%)[c]
Training set	CJ (n = 61)	0	100	98	2	97	98
	ML (n = 32)	1	97		0	100	
	CL (n = 39)	1	97		1	97	
Test set	CJ (n = 26)	0	100	95	1	96	97
	ML (n = 12)	1	92		0	100	
	CL (n = 18)	1	94		1	94	

Table 7 – PLS-DA and SVM classification results of China regional wines.

[a] Misclassified samples.
[b] Percentage of samples correctly classified.
[c] The average correct classified rates for the training set and the test set.

Two of the most significant parameters for the RBF are C and γ, and so these parameters should be carefully considered. More specifically, C is the cost parameter, i.e., the penalty that can balance the empirical risk minimization and structural risk minimization principles. To a certain extent, C can determine the performance of the SVM model. In addition, γ is the band width of the RBF kernel, and determines the sensitivity of the SVM model. It also affects the function regression error, which plays a vital role in selecting the number of initial eigenvalues/eigenvectors for the SVM model. As such, an inappropriate γ value may lead to overfitting [52]. To reduce overfitting and achieve the optimal error performance, a grid-search algorithm combined with 10-fold cross-validation was employed. In this case, the grid-search algorithm was based on an elaborate search in a restrained range, and reflected a two-dimensional minimization process. In this study, C was optimized in the range of $2^{(-3:3)}$, while γ was optimized in the range of $2^{(-4:1)}$, to give values of C = 0.5 and γ = 0.125 as optimal parameters for the SVM model. Finally, a SVM model has been developed that employs the RBF function to minimize the model error with the optimal parameters (i.e., C and γ) and the optimal hyper plane to differentiate between the different geographical origins of wines. Table 7 shows the internal 10-fold cross-validation classification results for the training set of SVM model. As indicated, the correct classification rates of the CL, CJ, and ML regions for the training set were 97, 97, and 100%, respectively, while those for the test set used for external validation were 94, 96, and 100%, respectively. As such, the average correct classification rate of the CL, CJ, and ML regions for the training set was 98%, while for the test set, the average correct classification rate was 97%. These results suggest that the SVM model is an effective approach for discriminating between different regional wine samples.

3.5. Joint analysis of PLS-DA and SVM

As both PLS-DA and SVM are conducted using the same data set, the misclassified samples for both methods are comparable, and the correct classification performance of the SVM model was found to be similar to that of the PLS-DA model. In addition, the average rate of correct classification for both models was >95%. We therefore propose that the joint data analysis of PLS-DA and SVM techniques could be used to successfully verify the geographical origins of wine samples.

The workflow chart of this joint data analysis is shown in Fig. 4, where the prediction classification results of Chinese wine based on their geographical origin were obtained by PLS-DA in the first step, prior to use of the SVM model in a second data evaluation step. As such, the prediction results from the PLS-DA and SVM models can be used to verify one another, thus if any wine samples are misclassified by both the PLS-DA and SVM models, these samples require further investigation. Therefore, the comprehensive classification results, obtained using the described workflow chart, suggest that joint data analysis using these methods may be preferable to determine the authenticity of the three claimed geographical origins (i.e., CJ, ML, and CL) of Chinese wines.

3.6. Comparing with the results of previous study

The work of Thiel et al. [34], 88 German commercial wines collected from four regions were analyzed for their concentration of 33 elements by ICP-OES and ICP-MS. The important discriminative elements (Li, Mg, Pb, W, Y, Cs, Sr, As, Si, Ti, B, Be, Sn) for distinguishing geographical origins of German white wines were identified by Plackett-Burman design and stepwise discriminant analysis. In our work, 52 elements were determined by ICP-OES and ICP-MS, oxygen isotope ratio ($\delta^{18}O$) was also determined by IRMS. Even though more wine samples collected from three regions were investigated in our study, better classification rates (>95%) were achieved than the work of Thiel et al. [34].

The latest work of Kokkinofta et al. [53], only four grape varieties (including red and white wines) from two regions were used for the geographic discrimination of Cypriot wines. Mg, Na, P, K, Ca, Mn, Fe, Zn, B, and Cu at mg/L were determined by inductively coupled plasma atomic emission spectroscopy (ICP-AES). The trace elements at ug/L were not tested in their study. Some important elements could contribute to the discrimination of wine regions might be missed. In our work, a total of 9 grape varieties (including red and white wines) from three regions were considered. 14 elements at mg/L and 38 elements at ug/L were determined. Therefore more grape varieties in three geographical origins and more mineral elements in our Chinese wine samples were analyzed. SNIF-NMR which was used by Kokkinofta et al. [53], was relative expensive equipment. Therefore it was not possible for every winery in China to use this kind of instrument.
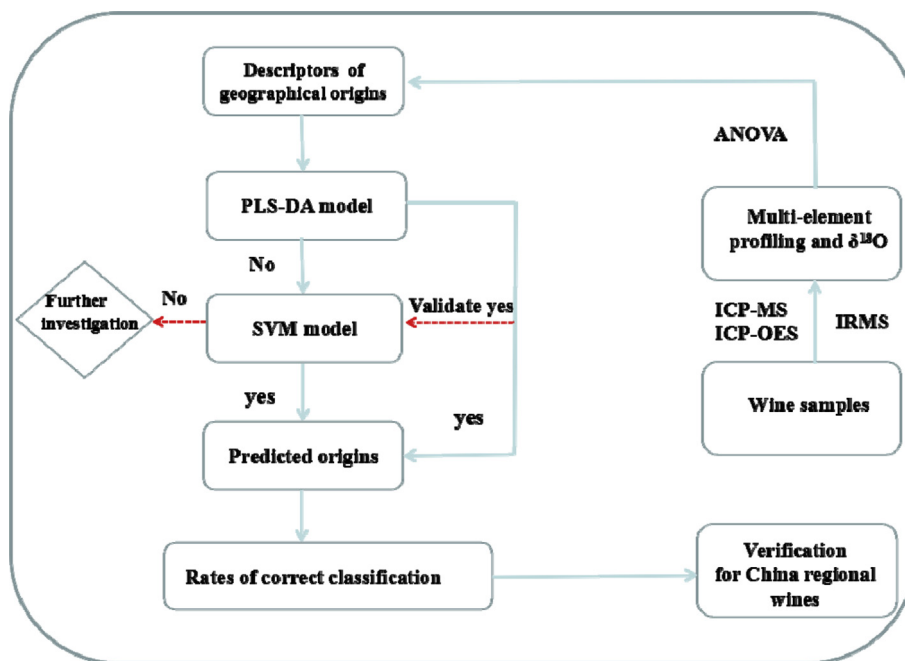
Fig. 4 — The workflow chart of joint analysis of PLS-DA and SVM.

The work of Geană et al. [13], $^{87}Sr/^{86}Sr$ isotope ratio, Ga and Al, were identified as the key indicators for geographic differentiation of the Romanian wines. In our present work, Mn, K and Co which obtained from PLS-DA model, were the most significant descriptors for discriminating Chinese wine regions. Meanwhile, Mg, Al, Na, P, K, Ca, Mn, Fe, Zn, Rb, B, and P which were determined by ICP-OES, are highly abundant elements in Chinese wine. The key important elements in our present study for geographic differentiation also included these elements. The same finding were also confirmed by Geană et al. [13]. Highly abundant elements could contribute to the discrimination of wine regions. Therefore ICP-OES method appears to be advantageous by the fact that it is a simple and low-cost technique. The work of Geană et al. [13], stepwise linear discriminant analysis (LDA) was used to identify key elements for classification of the wine samples, however, in our work PLS-DA was employed. The methods used to extract key elements for Romanian wines in the work of Geană et al. [13] and Chinese wines in our work were totally different. The work of Geană et al. [13], only red wines collected from protected designation of origin (PDO) were investigated. Similarly, only white wines could be classified in the work of Thiel et al. [34]. In our work, both red and white wines were covered for the discrimination of Chinese regional wines.

The recent work of Dinca et al. [48], C and O stable isotope ratios and ten trace elements (Co, V, Zn, Mn, Cu, Cr, Ni, Pb, Sr, and Rb) was used for the discrimination of Romanian regional wines. Only ANOVA and DA were used for multivariate data analysis. Misclassified Romanian wine samples in their study and previous work of Geană et al. [13], no second data evaluation method was employed. On the other hand, only ten trace elements and two stable isotope ratios were used and 100 wine samples collected from five regions were used in the

study by Dinca et al. [48]. In our work 14 mineral elements from 52 elements and $\delta^{18}O$ were extracted by ANOVA to the analysis of the discrimination of 188 wine samples collected from three geographical origins. With large wine sample sizes and more elements used for the multivariate data analysis, better classification rates (>95%) of Chinese regional wines could still be achieved in our study. In addition, in our work SVM was employed as a second data evaluation method for those misclassified samples from the PLS-DA method. The prediction results from the PLS-DA and SVM models can be used to verify against one another. Therefore our workflow for discriminating geographical origins of wines was more accurate, robust and rigorous than the previous work. Our results also indicate that the variables employed for discrimination, the selection of suitable multivariate methods and high-quality data are imperative for achieving accurate and reliable classification results. The same viewpoint is supported by both Cozzolino et al. [20] and Villagra et al. [54].

4. Conclusion

To the best of our knowledge, this is the first work to establish elemental profile and oxygen isotope ratio ($\delta^{18}O$) database for Chinese wines. Herein we also reported first time an innovative approach by the application of joint analysis of PLS-DA and SVM models for the successful classification of Chinese regional wines. A number of important descriptors for indicating geographical origins were extracted by ANOVA, and adequate classification rates (>95%) were obtained by both PLS-DA and SVM models. It was demonstrated that the combination of oxygen isotope ratio ($\delta^{18}O$) and elemental profile with multi-step multivariate analysis is a very promising tool

to verify the authenticity of Chinese wines from the Changji, Mile, and Changli regions.

However, it needs to be underpinned that the research reported herein represents only a starting point for a possible strategy to control the quality and authenticity of Chinese wines. For setting up a profound system for inspecting and controlling the correct labeling of the geographical origin of Chinese wines, additional samples with different vintages from the three investigated regions should be analyzed and samples originating from other geographical origins in China or abroad should be also collected, analyzed, and taken into the statistical model appropriately in the further. However, the data collection methods and the demonstrated approaches for data evaluation proposed in this study are considered to be essential to provide a solid backbone for the establishment of a Chinese wine database.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

REFERENCES

[1] Luykx DMAM, van Ruth SM. An overview of analytical methods for determining the geographical origin of food products. Food Chem 2008;107:897—911.

[2] Rui YK, Yu QQ, Jin YH, Guo J, Luo YB. Application of ICP-MS to the detection of forty elements in wine. Spectrosc Spectr Anal 2007;27:1015—7.

[3] Yang Y, Du H, Li W, Pan Q. Quick determination of 57 mineral elements in wines by inductively coupled plasma mass spectrometry (ICP-MS). Sino Overseas Grapevine Wine 2010;5:4—9.

[4] Du YY, Bai GY, Zhang X, Liu ML. Classification of wines based on combination of ^{1}H NMR spectroscopy and principal component analysis. Chin J Chem 2007;25:930—6.

[5] Jiang W, Xue J, Liu X, Wang DL, Guo Y, Wang L. The application of SNIF-NMR and IRMS combined with C, H and O isotopes for detecting the geographical origin of Chinese wines. Int J Food Sci Technol 2015;50:774—81.

[6] Xue J, Jiang L. The preliminary research on wine geographical origins determination by SNIF-NMR and IRMS methods. Food Ferment Ind 2011;37:154—8 [In Chinese, English abstract].

[7] Charlton AJ, Wrobel MS, Stanimirova I, Daszykowski M, Grundy HH, Walczak B. Multivariate discrimination of wines with respect to their grape varieties and vintages. Eur Food Res Technol 2010;231:733—43.

[8] Selih VS, Sala M, Drgan V. Multi-element analysis of wines by ICP-MS and ICP-OES and their classification according to geographical origin in Slovenia. Food Chem 2014;153:414.

[9] Rossano EC, Szilagyi Z, Malorni A, Pocsfalvi G. Influence of winemaking practices on the concentration of rare earth elements in white wines studied by inductively coupled plasma mass spectrometry. J Agric Food Chem 2007;55:311.

[10] Capron X, Smeyersverbeke J, Massart DL. Multivariate determination of the geographical origin of wines from four different countries. Food Chem 2007;101:1585—97.

[11] Koreňovská M, Suhaj M. Identification of some Slovakian and European wines origin by the use of factor analysis of elemental data. Eur Food Res Technol 2005;221:550—8.

[12] Almeida CM, Vasconcelos MTSD. ICP-MS determination of strontium isotope ratio in wine in order to be used as a fingerprint of its regional origin. J Anal At Spectrom 2001;16:607—11.

[13] Geanä EI, Sandru C, Stanciu V, Ionete RE. Elemental Profile and ^{87}Sr/^{86}Sr isotope ratio as fingerprints for geographical traceability of wines: an approach on Romanian wines. Food Anal Methods 2016:1—11.

[14] Bong YS, Ryu JS, Choi SH, La MR, Lee KS. Investigation of the geographical provenance of the beer available in South Korea using multielements and isotopes. Food Control 2016;60:378—81.

[15] Coetzee PP, Vanhaecke F. Classifying wine according to geographical origin via quadrupole-based ICP—mass spectrometry measurements of boron isotope ratios. Anal Bioanal Chem 2005;383:977—84.

[16] Camin F, Boner M, Bontempo L, Fauhl-Hassek C, Kelly S, Riedl J, et al. Stable isotope techniques for verifying the declared geographical origin of food in legal cases. Trends Food Sci Technol 2017;61:176—87.

[17] Paolanaranjo RDD, Baroni MV, Podio NS, Rubinstein HR, Fabani MP, Badini RG, et al. Fingerprints for main varieties of Argentinean wines: terroir differentiation by inorganic, organic, and stable isotopic analyses coupled to chemometrics. J Agric Food Chem 2011;59:7854.

[18] Dutra SV, Adami L, Marcon AR, Carnieli GJ, Roani CA, Spinelli FR, et al. Determination of the geographical origin of Brazilian wines by isotope and mineral analysis. Anal Bioanal Chem 2011;401:1571—6.

[19] Geana EI, Popescu R, Costinel D, Dinca OR, Stefanescu I, Ionete RE, et al. Verifying the red wines adulteration through isotopic and chromatographic investigations coupled with multivariate statistic interpretation of the data. Food Control 2016;62:1—9.

[20] Cozzolino D, Smyth HE, Cynkar W, Dambergs RG, Gishen M. Usefulness of chemometrics and mass spectrometry-based electronic nose to classify Australian white wines by their varietal origin. Talanta 2005;68:382.

[21] Godelmann R, Fang F, Humpfer E, Schütz B, Bansbach M, Schäfer H, et al. Targeted and nontargeted wine analysis by ^{1}H NMR spectroscopy combined with multivariate statistical analysis. Differentiation of important parameters: grape variety, geographical origin, year of vintage. J Agric Food Chem 2013;61:5610.

[22] Riedl J, Esslinger S, Fauhl-Hassek C. Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. Anal Chim Acta 2015;885:17.

[23] Geana I, Iordache A, Ionete R, Marinescu A, Ranca A, Culea M. Geographical origin identification of Romanian wines by ICP-MS elemental analysis. Food Chem 2013;138:1125–34.

[24] Geana EI, Marinescu A, Iordache AM, Sandru C, Ionete RE, Bala C. Differentiation of Romanian wines on geographical origin and wine variety by elemental composition and phenolic components. Food Anal Methods 2014;7:2064–74.

[25] Greenough JD, Mallory-Greenough LM, Fryer BJ. Geology and wine 9: regional trace element fingerprinting of Canadian wines. Geosci Can 2005;32:129–37.

[26] Taylor VF, Longerich HP, Greenough JD. Multielement analysis of Canadian wines by inductively coupled plasma mass spectrometry (ICP-MS) and multivariate statistics. J Agric Food Chem 2003;51:856.

[27] Coetzee PP, Steffens FE, Eiselen RJ, Augustyn OP, Balcaen L, Vanhaecke F. Multi-element analysis of South African wines by ICP-MS and their classification according to geographical origin. J Agric Food Chem 2005;53:5060.

[28] Bentlin, Spulgati FR, Hdressler F, Lpozebon V, Dirce. Elemental analysis of wines from South America and their classification according to country. J Braz Chem Soc 2011;22:327–36.

[29] Linde GVD, Fischer JL, Coetzee PP. Multi-element analysis of South African wines and their provenance soils by ICP-MS and their classification according to geographical origin using multivariate statistics. S Afr J Enol Vitic 2010;31:143–53.

[30] Jurado JM, Alcázar Á, Palacios-Morillo A, Pablos FD. Classification of Spanish DO white wines according to their elemental profile by means of support vector machines. Food Chem 2012;135:898.

[31] Iglesias M, Besalú E, Anticó E. Internal standardization–atomic spectrometry and geographical pattern recognition techniques for the multielement analysis and classification of Catalonian red wines. J Agric Food Chem 2007;55:219–25.

[32] Barbaste M, Medina B, Sarabia L, Ortiz MC, Pérez-Trujillo JP. Analysis and comparison of SIMCA models for denominations of origin of wines from de Canary Islands (Spain) builds by means of their trace and ultratrace metals content. Anal Chim Acta 2002;472:161–74.

[33] Castiñeira MDM, Feldmann I, Jakubowski N, Andersson JT. Classification of German white wines with certified brand of origin by multielement quantitation and pattern recognition techniques. J Agric Food Chem 2004;52:2962–74.

[34] Thiel G, Geisler G, Blechschmidt I, Danzer K. Determination of trace elements in wines and classification according to their provenance. Anal Bioanal Chem 2004;378:1630.

[35] Almeida CM, Vasconcelos MT. Multielement composition of wines and their precursors including provenance soil and their potentialities as fingerprints of wine origin. J Agric Food Chem 2003;51:4788.

[36] Angus NS, O'Keeffe TJ, Stuart KR, Miskelly GM. Regional classification of New Zealand red wines using inductively-coupled plasma-mass spectrometry (ICP-MS). Aust J Grape Wine Res 2006;12:170–6.

[37] Kment P, Mihaljevic M, Ettler V, Sebek O, Strnad L, Rohlova L. Differentiation of Czech wines using multielement composition–a comparison with vineyard soil. Food Chem 2005;91:157–65.

[38] Azcarate SM, Martinez LD, Savio M, Camiña JM, Gil RA. Classification of monovarietal Argentinean white wines by their elemental profile. Food Control 2015;57:268–74.

[39] Jurado JM, Alcázar A, Pablos F, Martín MJ, González AG. Classification of aniseed drinks by means of cluster, linear discriminant analysis and soft independent modelling of class analogy based on their Zn, B, Fe, Mg, Ca, Na and Si content. Talanta 2005;66:1350–4.

[40] Wang D, Zhong Q, Li G, Huang Z. Rapid method for the determination of the stable oxygen isotope ratio of water in alcoholic beverages. J Agric Food Chem 2015;63.

[41] National standards of the People's Republic of China. Wines. GB/T 15037-2006. Beijing, China: China Standard Press; 2007.

[42] Shen F, Wu J, Ying Y, Li B, Jiang T. Differentiation of Chinese rice wines from different wineries based on mineral elemental fingerprinting. Food Chem 2013;141:4026–30.

[43] Cao KAL, Rohart F, Gonzalez I, Dejean S, Gautier B, Bartolo F, et al. mixOmics: omics data integration Project, v R package version 6.1.3. 2017. https://CRAN.R-project.org/package=mixOmics.

[44] Kucheryavskiy S. Mdatools: multivariate data analysis for Chemometrics, v R package version 0.8.2. 2017. https://CRAN.R-project.org/package=mdatools.

[45] Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. Misc functions of the department of statistics, Probability Theory group (Formerly: E1071), TU Wien, v R package version 1.6-8. 2017. https://CRAN.R-project.org/package=e1071.

[46] Kraic F, Mocák J, Fiket Ž, Kniewald G. ICP MS analysis and classification of potable, spring, and mineral waters. Chem Pap 2008;62:445–50.

[47] Kruzlicova D, Fiket Ž, Kniewald G. Classification of Croatian wine varieties using multivariate analysis of data obtained by high resolution ICP-MS analysis. Food Res Int 2013;54:621–6.

[48] Dinca OR, Ionete RE, Costinel D, Geana IE, Popescu R, Stefanescu I, et al. Regional and vintage discrimination of Romanian wines based on elemental and isotopic fingerprinting. Food Anal Methods 2016;9:2406–17.

[49] Lim DK, Mo C, Lee D-K, Long NP, Lim J, Kwon SW. Non-destructive profiling of volatile organic compounds using HS-SPME/GC–MS and its application for the geographical discrimination of white rice. J Food Drug Anal 2018;26:260–7.

[50] Wu D, He Y, Feng S, Sun DW. Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM. J Food Eng 2008;84:124–31.

[51] Borin A, Ferrão MF, Mello C, Maretto DA, Poppi RJ. Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. Anal Chim Acta 2006;579:25–32.

[52] Tian Y, Wang Z, Han X, Hou H, Zheng R. Comparative investigation of partial least squares discriminant analysis and support vector machines for geological cuttings identification using laser-induced breakdown spectroscopy. Spectrochim Acta Part B At Spectrosc 2014;102:52–7.

[53] Kokkinofta R, Fotakis C, Zervou M, Zoumpoulakis P, Savvidou C, Poulli K, et al. Isotopic and elemental authenticity markers: a case study on Cypriot wines. Food Anal Methods 2017:1–12.

[54] Villagra E, Santos LS, Vaz BG, Eberlin MN, Laurie VF. Varietal discrimination of Chilean wines by direct injection mass spectrometry analysis combined with multivariate statistics. Food Chem 2012;131:692–7.