

Research Article

Quality Evaluation Based on Multivariate Statistical Methods

Shen Yin,¹ Xiangping Zhu,¹ and Hamid Reza Karimi²

¹ College of Engineering, Bohai University, Liaoning 121013, China

² Department of Engineering, Faculty of Engineering and Science of the University of Agder, 4898 Grimstad, Norway

Correspondence should be addressed to Xiangping Zhu; xiangping.zhu2010@gmail.com

Received 22 October 2013; Accepted 6 November 2013

Academic Editor: Xiaojie Su

Copyright © 2013 Shen Yin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Quality prediction models are constructed based on multivariate statistical methods, including ordinary least squares regression (OLSR), principal component regression (PCR), partial least squares regression (PLSR), and modified partial least squares regression (MPLSR). The prediction model constructed by MPLSR achieves superior results, compared with the other three methods from both aspects of fitting efficiency and prediction ability. Based on it, further research is dedicated to selecting key variables to directly predict the product quality with satisfactory performance. The prediction models presented are more efficient than tradition ones and can be useful to support human experts in the evaluation and classification of the product quality. The effectiveness of the quality prediction models is finally illustrated and verified based on the practical data set of the red wine.

1. Introduction

An accurate evaluation of the wine quality is of significance for vintners to perform wine classification and target marketing. However, since influenced by numerous factors such as grape varieties, yeast strains, wine making technologies and human experiences [1], evaluating wine quality is the main challenge for both the food industry and wine science community. Traditionally, wine quality is given by human experts or obtained by analyzing chemical compounds in the wine [2, 3]. Nevertheless, besides the issues like time consuming and complexity, these methods sometimes cannot meet the requirements when winemakers want to know the quality before the wine has been vinified, for example, predicting wine quality at the time of selecting grapes.

Amongst the various influence factors of the wine quality, grape is the most basic and important factor for making high quality wine and some grape physicochemical indexes have a strong relation to the wine quality. Since the sugar in grapes is the raw material for yeast to produce alcohol, its content plays a crucial role in the fermentation process and almost determines the alcoholic level of the wine [4]. In addition to the sugar, with studying three kinds of grapes (Cabernet, Sauvignon, and Merlot) which consume the same maturation, Fang et al. have found that the flavonoid varieties

in red wines are greatly affected by the varieties of grapes [5]. Although there exists malate synthesizing by yeast via fumarate or oxaloacetate pathways, the amount of that can be negligible compared with the malate originating from grapes and the high malate levels in grapes follow the high malate levels in wines [6]. All of these research results have revealed the important relationship between the grape physicochemical indexes and the wine quality and the necessity to evaluate the wine quality from such grape indexes.

In this paper, based on a data set of wine qualities and grape physicochemical indexes, wine quality prediction models are constructed with multivariate statistical methods. In order to obtain an efficient wine quality prediction model, a comparison amongst the models established by ordinary least squares regression, principal component regression, partial least squares regression, and a modified partial least square regression is made and the best model is selected out. The calibration set includes 50 grape physicochemical indexes. Dealing with so much data is a time consuming and complex task. With the correlation analysis, it has been found that there exists multicollinearity problems in some grape physicochemical indexes. Under the framework of wine quality prediction model, a suggestion regarding the use of fewer grape physicochemical indexes to predict the wine quality under the promise of prediction accuracy is proposed.

Compared with the majority of the methods reported for wine quality analysis, the method discussed in this paper provides a simpler and more convenient way to predict the wine quality. Besides, what is most remarkable of the proposed method is the increasing possibility for winemakers to predict the wine quality before the complete wine making process and make appropriate decisions in advance, such as the grape selection, wine classification, and target marketing.

Although investigations on evaluating wine quality only based on grape physicochemical indexes are rare in the wine science community, the relationship between grapes and metabolites in wine is available in many literature resources. Since the multiple biochemical process occurs with the grape ripening, the grape harvest time is influential in the wine composition. For example, the yeast-derived metabolites, including volatile esters, dimethyl sulfide, glycerol, and mannoproteins, will increase with harvest date [7]. In addition, the climate also has a profound influence on grape and finally influences the wine metabolites. Among the most important climate change-related effects are advanced harvest times and temperatures, increased grape sugar concentrations that lead to high wine alcohol level, lower levels of acids, and the modification of varietal aroma compounds [8, 9]. Moreover, with the help of ^1H NMR spectroscopic analysis and multivariate statistics methods such as PCA and OPLS-DA, Son et al. further proved that the grape varieties influence the wine quality significantly and they also succeeded in separating the wines vinified from four different grape varieties [10]. Besides, it is known from other literature that the genetic factors are also in a relation with grape and wine quality [11]. Given background of wines and grapes as well as the relevant influential factors achieved by this literature, this paper focuses on studying the relationship between wine quality and grape physicochemical indexes by applying multivariate statistical methods.

The rest of this paper is structured as follows. Section 2 makes a brief description of the multivariate statistical methods, including ordinary least squares regression, principal component regression, partial least squares regression, and a modified partial least square regression. Section 3 discusses the red wine making process with relevant illustrations. The wine prediction models are established based on multivariate statistical methods in Section 4, and a comparison among these models is provided to select the best model. Furthermore, the suggestion about using fewer grape physicochemical indexes to predict the wine quality is proposed as well in this section. Finally, the conclusions are presented in the last section.

2. Modeling Algorithms

Multivariate statistical analysis is a powerful tool for solving the wine analysis problems which often involve large amounts of data and has been widely applied in many relevant studies, such as analyzing the elements in wines by PLS regression [12], investigating the relationship between wine composition and vintage via principal component analysis (PCA) and partial least squares (PLS) [13, 14], and utilizing the visible-near

infrared spectroscopy and chemometrics to classify Riesling wines from different countries with the help of PCA [15]. In this section, besides describing three classical multivariate statistics methods, a modified partial least squares regression (MPLSR) method is also introduced, which has a simpler math description compared with standard PLS regression [16].

2.1. Ordinary Least Squares Regression (OLSR). Ordinary least squares regression (OLSR) is a standard approach to provide the approximate solution for the overdetermined systems through minimizing the sum of the errors created in the results of every single equation. Generally, OLSR often appears in the situation where only one dimension response variable is involved. However, OLSR is also capable of solving the problems of more than one dimension response variables and the algorithm can be briefly introduced as follows.

Step 1. Collect n samples of the measurable variables and response variables and normalize them to zero mean and unit variance, denoted as $X = [x_1 \cdots x_m] \in R^{n \times m}$, $Y = [y_1 \cdots y_p] \in R^{n \times p}$. The following steps will be performed only when $X^T X$ is a full rank matrix.

Step 2. Minimize the sum of errors e_j :

$$\beta_j = \arg \min e_j^T e_j = \arg \min (y_j - X\beta_j)^T (y_j - X\beta_j), \quad (1)$$

$$j = 1 \cdots p.$$

According to the invertibility of $X^T X$, the β_j can also be calculated by

$$\beta_j = (X^T X)^{-1} X^T y_j. \quad (2)$$

Step 3. Use the e_j and β_j to form $E \in R^{n \times p}$ and $M \in R^{m \times p}$, respectively, and the final model can be expressed as

$$Y = XM + E. \quad (3)$$

2.2. Principal Component Regression (PCR). In practice, the independent variables may be highly collinear. This phenomenon is the so-called multicollinearity and it is known that such collinearity problems can sometimes lead to serious stability problems when the OLSR is applied [17]. Alternative methods to OLSR are proposed in order to overcome this problem. The most popular methods are the principal component regression (PCR) and partial least squares regression (PLSR). The collinearity problems are efficiently solved in PCR algorithm by introducing principal components (PCs) which preserve the significant information in the original data set, and the algorithm can be briefly formulated as follows.

Step 1. Perform normalization on the gathered measurable variables and the responses variables, presented as $X = [x_1 \cdots x_m] \in R^{n \times m}$ and $Y = [y_1 \cdots y_p] \in R^{n \times p}$, where x_i and y_i have the zero mean and the unit variance.

Step 2. Implement singular value decomposition (SVD) on the covariance matrix of independent variable set:

$$\frac{1}{n-1}X^T X = P\Lambda P^T, \quad (4)$$

$$\Lambda = \text{diag}(\lambda_1 \cdots \lambda_m), \quad \lambda_1 > \cdots > \lambda_m > 0,$$

in which P is the so-called loading matrix of covariance matrix.

Step 3. Use appropriate criteria [18] to determine the number of principle components, and calculate the score matrix T :

$$T = XP_{\text{pc}} \in R^{n \times l}, \quad (5)$$

where

$$P_{\text{pc}} \subset P = [P_{\text{pc}} P_{\text{res}}] \in R^{m \times m},$$

$$P_{\text{pc}} = [p_1 \cdots p_l] \in R^{m \times l}, \quad (6)$$

$$P_{\text{res}} = [p_{l+1} \cdots p_m] \in R^{m \times (m-l)},$$

l : the number of the principal components.

Step 4. Perform OLS regression between the score matrix T and the dependent matrix Y and obtain the final model:

$$\beta_i = (T^T T)^{-1} T^T y_i, \quad i = 1 \cdots p,$$

$$B = [\beta_1 \cdots \beta_p] \in R^{m \times p}, \quad (7)$$

$$Y = TB + E = XM + E, \quad M = P_{\text{pc}} B.$$

2.3. Partial Least Squares Regression (PLSR). Different from the PCR, which only considers the outer relation of X block, partial least squares regression (PLSR) takes both outer relations (X and Y block individually) and inner relation (linking both blocks) into account and thus the latent variables (LVs) of X and Y have a strong relation which ensures the ability of PLSR model to make prediction from measurable variables [19]. Owing to this property, PLSR is widely used in many fields, such as fault detection [20], neuroimaging [21, 22], and wine analysis [13]. Many improvements have been made in PLS since 1982 when Herman Wold proposed the PLS algorithm [16, 23–25]. In this section, we would like to only consider the standard PLSR method as follows.

Step 1. Normalize the collected data sets of X and Y , expressed as $X = [x_1 \cdots x_m] \in R^{n \times m}$ and $Y = [y_1 \cdots y_p] \in R^{n \times p}$ where x_i and y_i are vectors with zero mean and unit variance.

Step 2. Iteratively calculate the following equations γ times:

$$(p_k, q_k) = \arg \max_{\|p_k\|=1, \|q_k\|=1} p_k^T X^T Y q_k, \quad k = 1 \cdots \gamma,$$

$$u_k = X_k p_k, \quad v_k = Y_k q_k,$$

$$X_k = u_k c_k^T + X_{k+1}, \quad Y_k = u_k r_k^T + Y_{k+1}, \quad (8)$$

$$c_k = \frac{X_k^T u_k}{\|u_k\|^2}, \quad r_k = \frac{Y_k^T u_k}{\|u_k\|^2},$$

$$X_{k+1} = X_k - u_k c_k^T, \quad Y_{k+1} = Y_k - u_k r_k^T,$$

where the p_k, q_k and u_k, v_k are the loading vector and score vector of X and Y , respectively, and the γ is the number of latent variables (LVs) which is usually determined by the cross-validation criteria [26].

Step 3. Store p_k, u_k, c_k , and r_k into P, U, C , and R , and the result of standard PLSR algorithm can be expressed as

$$U = XP, \quad X = UC^T + E, \quad (9)$$

$$Y = UR^T + F = XM + F, \quad M = PR^T.$$

2.4. Modified Partial Least Squares Regression (MPLSR). Both PCR and PLSR project the original data set into principal components space or latent variables space and residue space, and the lower dimension of principal components space or latent variables space overcomes the multicollinearity problem which is a hamper for the OLSR. However, these dimension reduction methods also make PCR and PLSR have the risk that useful information will lose in selected PCs or LVs. Although OLSR builds the model maintaining all of information from the origin data set, it always comes to a halt when the data set exists the problem of multicollinearity or the number of samples smaller than the number of variables, which are two common phenomenons in practice. In order to solve these problems, Yin et al. proposed the MPLSR, which has been validated on the industrial benchmark of Tennessee Eastman process for fault detection and a good result has been obtained [16]. MPLSR also has a simpler math description compared with PCR or PLSR, and the algorithm can be expressed as follows.

Step 1. Gather all the n samples of measurable variables and response variables, and stack them into X and Y , respectively. Normalize them to zero mean and unit variance, denoted as $X = [x_1 \cdots x_m] \in R^{n \times m}$ and $Y = [y_1 \cdots y_p] \in R^{n \times p}$.

Step 2. Calculate the regression coefficient matrix M :

$$\frac{1}{n} Y^T X \approx M^T \frac{X^T X}{n},$$

$$M = (X^T X)^{-1} X^T Y \quad \text{if } \text{rank}(X^T X) = m, \quad (10)$$

$$M = (X^T X)^\dagger X^T Y \quad \text{if } \text{rank}(X^T X) < m,$$

in which the $(X^T X)^\dagger$ is the pseudoinverse of $X^T X$.

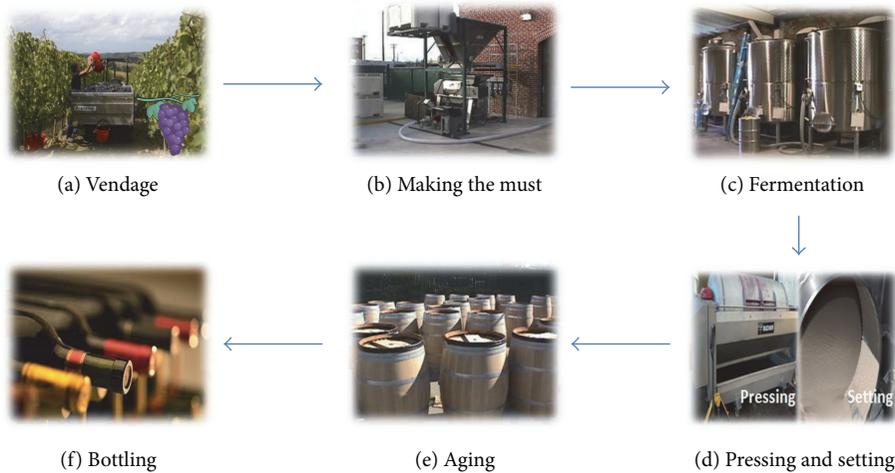


FIGURE 1: Flow sheet of winemaking.

Step 3. Obtain the final model of MPLSR:

$$Y = XM + E_y, \quad (11)$$

where E_y is the residue part of Y which is uncorrelated with X .

3. Winemaking Process

All wines are practically made in a common process from grapes harvesting to bottling. For red wines, the winemaking process can be roughly divided into six steps.

Firstly, grape harvesting is performed to supply the raw material for winemaking. For the grape compositions, which may positively or negatively influence wine chemistry and sensory properties or vary with the maturity of grapes, the harvest dates should be carefully considered. Traditionally, the levels of grape sugar, acids, and PH are used to determine the harvest dates. However, Jackson and Lombard have proposed that such measures alone are not sufficient to accurately predict wine composition because many key grape-derived compounds in wine do not track with sugar accumulation [27]. In many cases, the optimal time for picking lasts only a very few days. Grape harvesting or vendage can be made mechanically or manually. Although mechanical harvester is a cost-effective choice, there are some disadvantages which will affect the wine quality [28]. Thus, in order to guarantee the wine quality, some regions still harvest grapes manually. As soon as grapes have been harvested, separating stems from grapes and crushing grapes are necessarily performed in the winery [29, 30]. Fermentation is the most significant stage of vinification and the main reaction during this stage is converting sugars to ethanol with the assistance of yeast. The chemical reaction can be described as



The natural yeast which is already presented on grapes may give unpredictable results depending on the exact types of yeast; thus the cultured yeast is always added to the must in

order to ensure a successful fermentation. The amount of time consumed by a wine ferment process varies depending on the type of grape and the methods adopted by winemakers. Generally, 10 to 30 days will be consumed during the fermentation which takes place in large vats [31]. After the fermentation, the must (now wine) is drawn off to press away the skins, which are the so-called cap floats contributing to the color and flavour of red wine. In order to acquire a relatively clear wine, statical settlement is necessary to separate the sediments and dead yeast cells from wine. Since the vintage can significantly impact the flavour and quality of wine [32], most winemakers prefer to let wines undergo a period of aging in oak or redwood barrels to improve the wine quality, and the cool underground cellars are perfect for maturing wines. During the period of aging, winemakers will check the wooden barrels at regular intervals. Sometimes moving wine from one barrel to another or from barrel to stainless steel tank is necessary and nitrogen gas is used to protect wine from exposing to large amounts of oxygen [33, 34]. Depending on the type and quality of the wine desired, the aging of red wine may last from several months to several years. However, not all of wines should go through the period of aging. While some wines will improve quality with age, others can and should be drunk immediately because the tannin which gives the wine its flavor will precipitate out. Finally, the wine is prepared for bottling. For getting a good presentation of wines, especially for the commercial wines, filtration, before bottling, is often used to make the wine bright and clear by removing suspended particles. The red wine making process described above can be roughly presented in the flowsheet of Figure 1 (the pictures used in this flowsheet can be downloaded from <http://www.whitman.edu/>).

4. Results and Discussion

In this section, multivariate statistical methods are utilized to establish the models of the relationship between the grape physicochemical indexes and the wine quality. The fitting efficiency and predicting ability of these methods

TABLE 1: The wine quality of one sample given by human experts.

Human expert	1	2	3	4	5	6	7	8	9	10	Mean points
Presentation (15 points)	10	12	10	7	10	12	12	10	11	10	10.4
Fragrance (30 points)	18	24	24	18	24	22	22	18	27	20	21.7
Mouth feel (44 points)	24	33	37	29	28	26	22	26	34	29	28.8
Overall feeling (11 points)	8	9	10	8	8	7	8	8	9	8	8.3
Sum points (100 points)	60	78	81	62	70	67	64	62	81	67	69.2

are compared by corresponding figures or indexes. After analyzing the influence of grape physicochemical indexes on the wine quality, a suggestion is proposed to reduce the usage of grape physicochemical indexes in efficient wine quality prediction.

4.1. Data Preprocessing. The data set (mathematical modeling official site: <http://www.mcm.edu.cn/>) involved in modeling consists of the information of red wine qualities and 50 grape physicochemical indexes. For some grape physicochemical indexes, two or three measurements are preformed and the mean values are taken as the results of such indexes.

Corresponding to grape samples, wine samples have been vinified and collected simultaneously to avoid the differences in wine quality caused by different vintages, which influence the compositions in wine significantly [13]. Ten human experts have been involved in the evaluation of wine qualities (100 points in total) from four aspects: presentation (15 points), fragrance (30 points), mouth feel (44 points), and overall feeling (11 points). The total points are calculated by summing the points of four aspects given by the human experts and the ultimate wine quality of each sample obtained from averaging the 10 total points. Taking one of the samples, for example, the detailed information can be found in Table 1.

It is known that the data set used to construct the models cannot be applied to verify the performance of the models. Based on such guideline, the samples are divided into two parts, which are named calibration set and verification set, respectively. The calibration set formed by 27 samples is used to construct the wine quality prediction models and the verification set formed by 10 samples is used to verify the predicting ability of the models.

4.2. Modeling and Comparison. In order to simplify the problem, some assumptions are made as follows.

- (1) For all of wines involved in this paper, the vinification processes are identical, including the materials added during the wine making process.
- (2) All of wines have been vinified by the winemakers who possessed the same vinification experience.
- (3) The wine qualities have been evaluated by human experts as soon as the wines have been vinified.

All of the above assumptions are made to ensure that the differences of wine qualities are only caused by differences of the grapes. This paper merely aims to study the relationship between the grape physicochemical indexes and the wine quality.

As aforementioned, OLSR is not satisfactory to build the regression equation when the calibration set suffers the problem of multicollinearity or the number of samples is smaller than the number of variables. By applying correlativity analysis on the calibration set, strong correlations have been found in some grape physicochemical indexes, especially the reducing sugar and glucose whose correlation coefficient reaches 0.982. The correlation coefficients which are greater than 0.95 are presented in Table 3. Besides, the number of samples consumed by the grape physicochemical indexes is also smaller than the number of its measure variables. Thus it is obvious that OLSR is incapable of establishing the model based on the grape physicochemical indexes and wine quality by using the calibration set.

Apart from OLSR, all the discussed multivariate statistical methods like PCR, PLSR, and MPLSR have been utilized to construct the relevant models. Two commonly used indexes, that is, the root mean squared error of calibration (RMSEC) and root mean error of prediction (RMSEP), are mainly considered here for evaluating the fitting efficiency and prediction ability of different methods. RMSEC and RMSEP can be described as follows:

$$\begin{aligned} \text{RMSEC} &= \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_{\text{cal},i})^2}{N}}, \\ \text{RMSEP} &= \sqrt{\frac{\sum_{i=1}^M (y_i - \hat{y}_{\text{pred},i})^2}{M}}, \end{aligned} \quad (13)$$

where y_i is the measured value for the i th sample; $\hat{y}_{\text{cal},i}$ and $\hat{y}_{\text{pred},i}$ are the method fitted and model predicted response values for the i th sample, respectively; N and M are the number of calibration samples and verification samples, respectively.

The CPV² (cumulative percent variance) [18] with selecting 95% as CPV is firstly applied to selected the number of PCs and 16 PCs are obtained for PCR. 4 LVs are selected for PLSR by utilizing cross-validation [26, 35]. With maintaining nearly all the information from the origin data set, MPLSR is not necessary to select the PCs or LVs that are the main difficulties for PCR and PLSR.

Figures 2(a), 2(b), 2(c), 2(d), 2(e), and 2(f) show the fitting efficiency of different models obtained by PCR, PLSR, and MPLSR. For only preserving the significant variability information in the grape physicochemical indexes when choosing PCs, the fitting efficiency of PCR is the worst among these three methods. Different from PCR, PLSR selects LVs

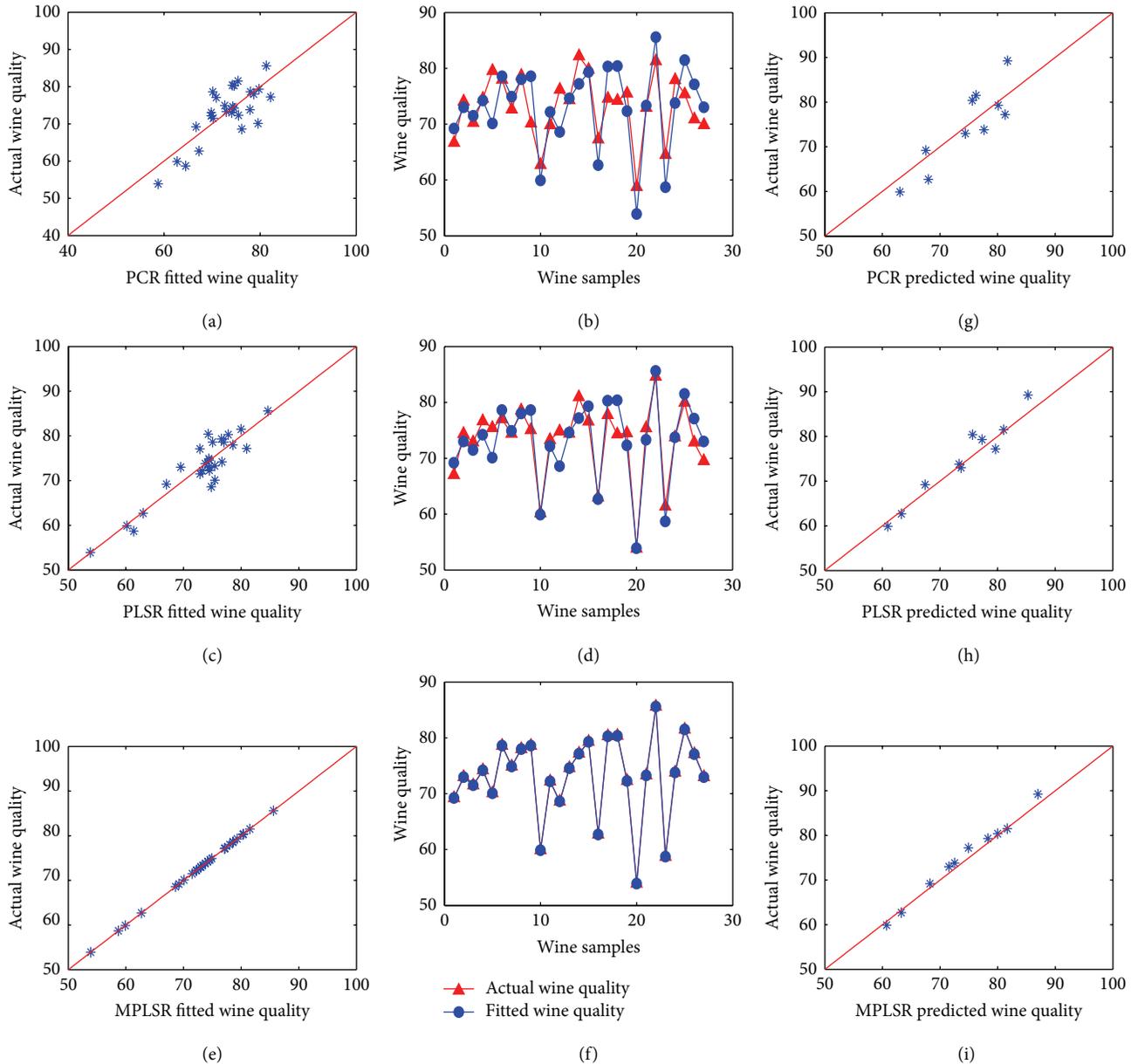


FIGURE 2: (a)–(f) denote fitted quality versus actual quality. (g)–(i) denote predicted quality versus actual quality. (a), (b), (g) are plotted by PCR. (c), (d), (h) are plotted by PLSR. (e), (f), (i) are plotted by MPLSR.

under the consideration of both the relation among physicochemical indexes and the relationship between the grape physicochemical indexes and the wine quality. As it is shown in Figure 2, the model constructed by PLSR is obviously superior to the PCR model. The best fitting efficiency belongs to the MPLSR model in which the fitted wine quality equals the actual quality. For MPLSR model, there exists no problem for selecting the PCs or LVs, which is often the source of errors between the fitted values and actual values due to only keeping the PCs or LVs and ignoring the residues based on some criterion [18, 26]. With the RMSEC values that denote the fitting ability of different models, Table 2 can also show that the fitting efficiency of the MPLSR model is the best among the three models and PCR model is the worst.

TABLE 2: RMSEC and RMSEP values.

	PCR	PLSR	MPLSR
RMSEC	4.53	2.82	0
RMSEP	4.26	2.32	1.31

The goal of regression models is to predict the wine quality from new measured grape physicochemical indexes. In order to compare the prediction ability of these models, the RMSEP values of each model are calculated. Corresponding to the maximum RMSEP value of the PCR model and the minimum RMSEP value of the MPLSR model, the PCR model has the worst prediction ability while the best

TABLE 3: Correlation coefficients of some grape physicochemical indexes.

	Total content of amino acids Proline	Resveratrol CIS resveratrol	Flavonol Quercetin	Reducing sugar Fructose	Reducing sugar Glucose
Correlation coefficient	0.978	0.978	0.968	0.967	0.982

predicting ability belongs to MPLSR model, which is similar to the results of fitting efficiency of these three models. Figures 2(g) and 2(h) give visual results. The blue points in the figures should locate on the red line if the model can precisely predict the wine quality, and based on this rule, it can be seen that the best predicting ability of the MPLSR model can be also derived.

According to Figure 2 and Table 2, it is obvious that different multivariate statistical methods will significantly influence the result of the predicting models. Due to the problems of multicollinearity and the number of samples smaller than the number of variables in the calibration set, the OLSR is absolutely unusable for establishing the prediction model. With the minimal values of RMSEC and RMSEP, MPLSR obtains a satisfying result for both the fitting efficiency and predicting ability.

4.3. *Selecting the Key Grape Physicochemical Indexes.* It is known that 50 grape physicochemical indexes are included in the calibration set and sufficient information has been provided by these indexes. However, it is a time consuming and complex work to measure so many grape physicochemical indexes. From the regression equations, it has been found that some grape physicochemical indexes contribute little to the wine quality. Besides, Table 3 also shows that some grape physicochemical indexes are redundant. From these points, it is necessary to analyze the contribution of every grape physicochemical index and select the optimal grape physicochemical indexes for efficiently predicting the wine quality.

In order to analyze the contribution of the every grape physicochemical index to the wine quality, a contribution ratio (CR) is introduced in and it can be formulated as follows:

$$y - \beta_0 = BX, \quad CR_i = \frac{\beta_i x_i}{y - \beta_0}, \quad i = 1 \cdots m,$$

$$B = [\beta_1, \beta_2 \cdots \beta_m] \in R^{1 \times m}, \quad (14)$$

$$X = [x_1, x_2 \cdots x_m]^T \in R^{1 \times m},$$

where B is the coefficient matrix of regression equation and β_0 is a constant. The benefit of this definition for CR is that the sum of all CR_i ($i = 1 \cdots m$) equals 1.

As the best prediction model, the MPLSR model is applied to analyze the CR of every grape physicochemical index and select the optimal indexes. Utilizing one of the samples, the CR of every grape physicochemical index can be obtained as shown in Table 4. The CR with a negative value means that the index has a negative contribution to the wine quality. From the table, it is obvious that some grape physicochemical indexes influence the wine quality significantly, especially alanine, protein, total sugar, soluble solids, and dry matter content, while some indexes have little

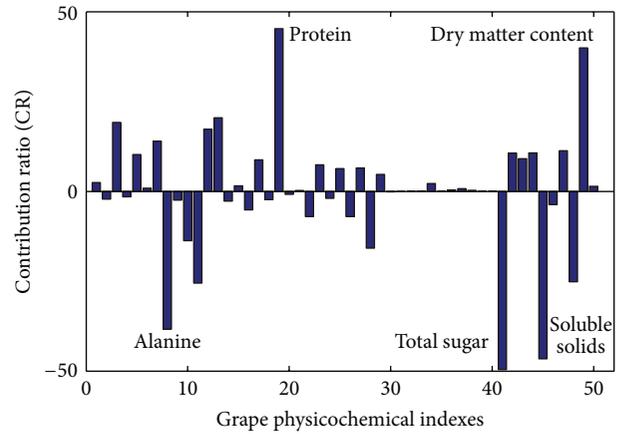


FIGURE 3: The CRs of 50 grape physicochemical indexes.

influence or even no influence on the wine quality, such as trans resveratrol glucoside, CIS resveratrol glucoside, and isorhamnetin whose CRs equal 0. An illustrated expression about the CRs is shown in Figure 3.

To get rid of these useless indexes which contribute little to the wine quality or have a strong correlation with other grape physicochemical indexes, a CR threshold is applied and the grape physicochemical index will be ignored if its CR is lower than the CR threshold. Comparing the regression results with different CR thresholds, 1 is selected as the threshold and 13 indexes are ignored which are marked in Table 5 with asterisk (*). Based on the new calibration set which includes 37 grape physicochemical indexes, a new calibration model is established by using the MPLSR method. The CRs of the selected grape physicochemical indexes in the new model are presented in Figure 4. As can be seen from the figure, alanine, protein, total sugar, and soluble solids still have a great influence on the wine quality. Moreover, fructose, PH, and titratable acid also contribute to the wine quality significantly. However, the dry matter content which influences the wine quality greatly in the old model constructed by 50 grape physicochemical indexes contributes to the wine quality in the new model slightly.

A comparison between the new model and the old model is made in Figure 5. From Figures 5(a), 5(b), 5(d), and 5(e), it is obvious that the fitting efficiency of the new model is nearly equal to the old model, and a same result can also be derived from Table 5 with the equal RMSEC values of both models. The verification set is utilized to test the predicting ability of the new model. As can be seen from Table 5, although 13 grape physicochemical indexes have been ignored, the RMSEP value of the new model does not significantly deviate from that of the model built by all the 50 grape physicochemical

TABLE 4: The CR of 50 grape physicochemical indexes.

Index	CR
Total amino acids	2.49
Aspartic acid	-2.13
Threonine	19.18
Serine	-1.49
Glutamic acid	10.25
Proline*	0.89
Glycine	14.00
Alanine	-38.34
Cystine	-2.40
Valine	-13.73
Methionine	-25.25
Isoleucine	17.28
Leucine	20.47
Tyrosine	-2.68
Phenylalanine	1.56
Lysine	-5.16
Histidine	8.77
Arginine	-2.29
Protein	45.29
VC*	-0.78
Anthocyanin*	0.23
Tartaric acid	-6.97
Malic acid	7.41
Citric acid	-1.90
POA†	6.34
Browning degree	-6.97
DPPH free radical	6.54
Total phenolic	-15.77
Tannin	4.71
Grape flavonoid*	-0.14
Resveratrol*	-0.05
Trans-RG†*	0
CIS-RG†*	0.12
Transresveratrol	2.21
CIS resveratrol*	0
Flavonol*	0.38
MC†*	0.74
Quercetin*	0.29
Kaempferol*	-0.02
Isorhamnetin*	0
Total sugar	-49.56
Reducing sugar	10.67
Fructose	9.08
Glucose	10.72
Soluble solids	-46.56
PH	-3.68
Titrateable acid	11.32
Solid acid ratio	-25.14
Dry matter content	39.92
Ear weight	1.43

*13 grepe physicochemical indexes which are ignored.

†POA: polyphenol oxidase activity; trans-RG: transresveratrol glucoside; CIS-RG: CIS resveratrol glucoside; MC: myricetin cannabiscetin.

TABLE 5: The values of RMSEC and RMSEP.

	Old model	New model
RMSEC	0	0
RMSEP	1.31	1.68

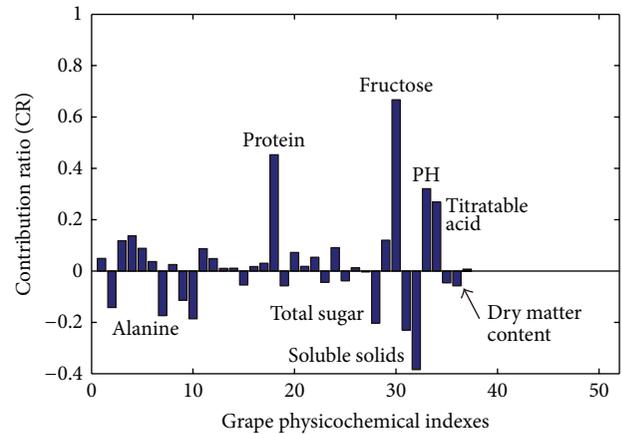


FIGURE 4: The CRs of 37 grape physicochemical indexes in the new model.

indexes. A slight difference also can be found as well by comparing Figures 5(c) and 5(f).

Whether from the results shown in Figure 2 or from the indexes of RMSEC and RMSEP presented in Table 2, the model constructed by MPLSR has a relative satisfying performance for predicting the wine quality based on the grape physicochemical indexes. With the contribution ratio analysis, it is interesting to note that the contributions of some grape physicochemical indexes, that is, trans resveratrol glucoside, CIS resveratrol glucoside, and isorhamnetin, can be negligible, while some grape physicochemical indexes, that is, protein, soluble solids, and total sugar, have a significant influence on the final wine quality. Although 13 grape physicochemical indexes have been ignored, the new model constructed based on the remaining indexes with MPLSR method still presents a good wine quality predicting ability.

5. Conclusion

In this paper, four multivariate statistical methods, that is, OLSR, PCR, PLSR, and MPLSR, are firstly reviewed. Based on these methods and the real data obtained in practice, wine quality prediction models have been constructed. With the superior fitting efficiency and better predicting ability represented by RMSEC and RMSEP, respectively, the model built by MPLSR outperforms the other three models. Several grape physicochemical indexes, such as protein, soluble solids, and total sugar, are found to have significant contributions to the final wine quality while others are insignificant. Through ignoring the insignificant grape physicochemical indexes, the model constructed by key indexes could present a satisfactory wine quality predicting ability. The efficiency

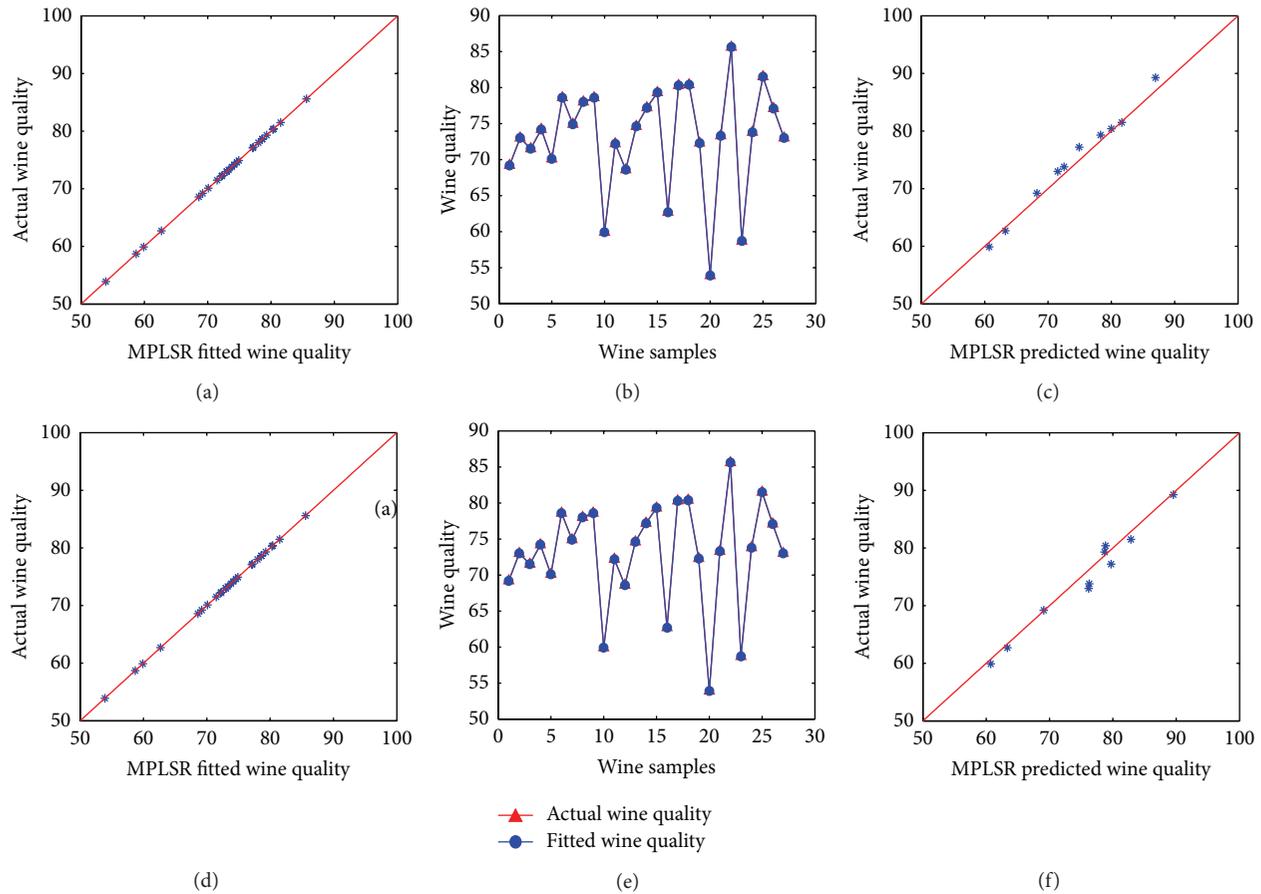


FIGURE 5: A comparison between the model with 50 grape physicochemical indexes and the model with 37 grape physicochemical indexes. (a), (b), (c) are plotted by old model constructed with 50 grape physicochemical indexes. And (d), (e), (f) are plotted by the new model constructed with 37 grape physicochemical indexes.

of the MPLSR model is essential to be validated on a larger data set. Moreover, robust wine quality prediction models are meaningful to be proposed in the future work.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 61304102) and the Natural Science Foundation of Liaoning Province, China (no. 2013020002).

References

[1] S. Charters and S. Pettigrew, “The dimensions of wine quality,” *Food Quality and Preference*, vol. 18, no. 7, pp. 997–1007, 2007.
 [2] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, “Modeling wine preferences by data mining from physicochemical properties,” *Decision Support Systems*, vol. 47, no. 4, pp. 547–553, 2009.

[3] S. Buratti, S. Benedetti, M. Scampicchio, and E. C. Pangerod, “Characterization and classification of Italian Barbera wines by using an electronic nose and an amperometric electronic tongue,” *Analytica Chimica Acta*, vol. 525, no. 1, pp. 133–139, 2004.
 [4] J. Herrera, A. Guesalaga, and E. Agosin, “Shortwave-near infrared spectroscopy for non-destructive determination of maturity of wine grapes,” *Measurement Science and Technology*, vol. 14, no. 5, pp. 689–697, 2003.
 [5] F. Fang, J.-M. Li, P. Zhang et al., “Effects of grape variety, harvest date, fermentation vessel and wine ageing on flavonoid concentration in red wines,” *Food Research International*, vol. 41, no. 1, pp. 53–60, 2008.
 [6] P. R. Gayon, D. Dubourdieu, B. Donèche, and A. Lonvaud, Eds., *Handbook of Enology, the Microbiology of Wine and Vinifications*, John Wiley & Sons, New York, NY, USA, 2006.
 [7] K. Bindon, C. Varela, J. Kennedy, H. Holt, and M. Herderich, “Relationships between harvest time and wine composition in vitis vinifera l. cv. cabernet sauvignon l. grape and wine chemistry,” *Food Chemistry*, vol. 138, no. 1, pp. 1696–1705, 2013.
 [8] R. Mira de Orduña, “Climate change associated effects on grape and wine quality and production,” *Food Research International*, vol. 43, no. 7, pp. 1844–1855, 2010.
 [9] H. Li, J. Yu, C. Hilton, and H. Liu, “Adaptive sliding mode control for nonlinear active suspension vehicle systems using

- t-s fuzzy approach," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3328–3338, 2013.
- [10] H.-S. Son, G.-S. Hwang, H.-J. Ahn, W.-M. Park, C.-H. Lee, and Y.-S. Hong, "Characterization of wines from grape varieties through multivariate statistical analysis of ^1H NMR spectroscopic data," *Food Research International*, vol. 42, no. 10, pp. 1483–1491, 2009.
- [11] G. E. Pereira, J.-P. Gaudillere, C. van Leeuwen et al., " ^1H NMR and chemometrics to characterize mature grape berries in four wine-growing areas in Bordeaux, France," *Journal of Agricultural and Food Chemistry*, vol. 53, no. 16, pp. 6382–6389, 2005.
- [12] D. A. Guillén, M. Palma, R. Natera, R. Romero, and C. G. Barros, "Determination of the age of Sherry wines by regression techniques using routine parameters and phenolic and volatile compounds," *Journal of Agricultural and Food Chemistry*, vol. 53, no. 7, pp. 2412–2417, 2005.
- [13] S. A. Bellomario, R. M. Parker, X. A. Conlan, N. W. Barnett, and M. J. Adams, "Partial least squares and principal components analysis of wine vintage by high performance liquid chromatography with chemiluminescence detection," *Analytica Chimica Acta*, vol. 678, no. 1, pp. 34–38, 2010.
- [14] X. Shi, Y. Lv, Z. Fei, and J. Liang, "A multivariable statistical process monitoring method based on multiscale analysis and principal curves," *International Journal of Innovative Computing, Information and Control*, vol. 9, no. 4, pp. 1781–1800, 2013.
- [15] L. Liu, D. Cozzolino, W. U. Cynkar et al., "Preliminary study on the application of visible-near infrared spectroscopy and chemometrics to classify Riesling wines from different countries," *Food Chemistry*, vol. 106, no. 2, pp. 781–786, 2008.
- [16] S. Yin, S. X. Ding, P. Zhang, A. Hagahni, and A. Naik, "Study on modifications of pls approach for process monitoring," in *Proceedings of the 18th IFAC World Congress*.
- [17] S. Weisberg, *Applied Linear Regression*, vol. 528, John Wiley & Sons, New York, NY, USA, 2005.
- [18] S. Valle, W. Li, and S. J. Qin, "Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods," *Industrial and Engineering Chemistry Research*, vol. 38, no. 11, pp. 4389–4401, 1999.
- [19] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Analytica Chimica Acta*, vol. 185, pp. 1–17, 1986.
- [20] S. Yin, *Data-driven design of fault diagnosis systems [Ph.D. thesis]*, Universität Duisburg-Essen, Fakultät für Ingenieurwissenschaften Ingenieurwissenschaften-Campus Duisburg Abteilung Elektrotechnik und Informationstechnik Automatisierung und komplexe Systeme, 2012.
- [21] A. Krishnan, L. J. Williams, A. R. McIntosh, and H. Abdi, "Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review," *NeuroImage*, vol. 56, no. 2, pp. 455–475, 2011.
- [22] X. Shi, Y. Lv, Z. Fei, and J. Liang, "Fokrul alom mazarbhuiya and muhammad abulaish, clustering periodic frequent patterns using fuzzy statistical parameters," *International Journal of Innovative Computing, Information and Control*, vol. 8, no. 3, pp. 2275–2283, 2012.
- [23] V. E. Vinzi, *Handbook of Partial Least Squares: Concepts, Methods and Applications*, Springer, New York, NY, USA, 2010.
- [24] H. Wold, "Soft modelling: the basic design and some extensions," *Systems Under Indirect Observation*, vol. 2, pp. 589–591, 1982.
- [25] H. Li, X. Jing, and H. R. Karimi, "Output-feedback based h-infinity control for active suspension systems with control delay," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 1, pp. 436–446, 2014.
- [26] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.
- [27] D. I. Jackson and P. B. Lombard, "Environmental and management practices affecting grape composition and wine quality—a review," *American Journal of Enology and Viticulture*, vol. 44, no. 4, pp. 409–430, 1993.
- [28] Q. Zhou, P. Shi, J. Lu, and S. Xu, "Adaptive output-feedback fuzzy tracking control for a class of nonlinear systems," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 5, pp. 972–982, 2011.
- [29] Y. Fukayama, D. Tanaka, and T. Kataoka, "Separation of individual instrument sounds in monaural music signals by applying statistical least-squares criterion," *International Journal of Innovative Computing, Information and Control*, vol. 8, no. 3, pp. 2275–2283, 2012.
- [30] X. Xie, S. Yin, H. Gao, and K. Okyay, "Asymptotic stability and stabilisation of uncertain delta operator systems with time-varying delays," *IET Control Theory & Applications*, vol. 7, no. 8, pp. 1071–1078, 2013.
- [31] D. Li, Z. Wang, and H. Gao, "Distributed h-infinity filtering for a class of markovian jump nonlinear time-delay systems over lossy sensor networks," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 10, pp. 4665–4672, 2013.
- [32] I. S. Hornsey, *The Chemistry and Biology of Winemaking*, Royal Society of Chemistry, 2007.
- [33] Q. Zhou, P. Shi, S. Xu, and H. Li., "Observer-based adaptive neural network control for nonlinear stochastic systems with time-delay," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 1, pp. 71–80, 2013.
- [34] C. J. Tsai, T. Chang, Y. Yang, M. Wu, and Y. Li, "An approach for user authentication on non-keyboard devices using mouse click characteristics and statistical-based classification," *International Journal of Innovative Computing, Information and Control*, vol. 8, no. 11, pp. 7875–7886, 2012.
- [35] X. Su, Z. Li, Y. Feng, and L. Wu, "New global exponential stability criteria for interval-delayed neural networks," *Proceedings of the Institution of Mechanical Engineers I*, vol. 225, no. 1, pp. 125–136, 2011.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

