Discrimination of Chinese Yellow Wine from Different Origins Based on Flavor Fingerprint

Tong Chen¹, Xingpu Qi², Mingjie Chen¹, Daoli Lu¹ and Bin Chen¹*

¹Jiangsu University, Zhenjiang 212013, P.R. China
²Jiangsu Agri-animal Husbandry Vocational College, Taizhou, 225300, P.R. China

Received: 22 February 2019; accepted: 23 April 2019

In this study, discrimination of Chinese yellow wines from Shaoxing, Shandong, and Hubei in China has been carried out according to volatile flavor components. A total of 122 yellow wine samples were characterized by gas chromatography–ion mobility spectrometry (GC–IMS). A simple color mixing method was visually used to select characteristic peaks based on the RGB color model. Then, the volatile organic compounds corresponding to the selected characteristic peaks were identified via library searching, and the height values of those peaks were arranged according to their RGB color. Quadratic discriminant analysis was applied to develop a classification model and achieved a correct classified rate of 95.35% for the prediction set. The results prove that the aroma composition combined with chemometric tools can be used as a fingerprinting technique to protect the product of origin and enable the authenticity of Chinese yellow wine.

**Keywords:** GC–IMS, yellow wine, flavor fingerprint, chemometrics

Introduction

Yellow wine, beer, and red wine are known as the three most ancient wines in the world. Traditional yellow wine was initially developed in Shaoxing, China, and brewed from glutinous rice and wheat. It is well known that glutinous rice contains higher protein and lower fat, and wheat can offer rich carbon, nitrogen, and micro-element source for mold and yeast used for fermentation [1]. The unique brewing technology handed down from generation to generation has brought yellow wine with bright brown color, subtle sweet flavor, and low alcoholicity. Therefore, yellow wine has been deeply popular with customers in China for centuries. In addition, yellow wine is widely known for its health care function. Yellow wine is rich in oligos, organic acids, amino acids, vitamins, trace aromatic compounds, and minerals and has been claimed to have beneficial effects on prevention of cancer and cardiovascular diseases [2, 3]. Among the developed yellow wines, Shaoxing yellow wine has the longest history and is the most representative variety. However, since the Shaoxing yellow wine has a huge commercial value, some wineries that do not have the same production conditions fraudulently use this branding, and thus a large number of yellow wine that produced in non-Shaoxing areas are on sales in the market, which is a serious threat to the interest of the customers and spoils the reputation of the Shaoxing yellow wine. Therefore, correct evaluation of wine quality is very important for the credibility of producers and merchants, as well as the right of consumers.

Yellow wine is a complex mixture, which, besides water and alcohol, contains a great variety of inorganic and organic components, such as sugars, organic acids, phenolic compounds, amino acids, metal elements, etc. All those chemical compositions have an important influence on the quality and flavor of yellow wine products and can be used as descriptors to discriminate and verify the quality of wines. However, many factors could affect the flavor of yellow wine in natural ecological environment, which include resource conditions (like different microbes, water quality, climate, air, and conditions) and brewing technology. Therefore, different origins of yellow wine have their own unique flavor composition. Besides conventional physical and chemical analysis of yellow wine [4], the quality of yellow wine mainly depends on sensory evaluation. However, sensory evaluation would be affected by environmental conditions, subjective factors of the evaluator, and physical and mental status and might obtain the different or opposite results. In recent years, more and more attention has been paid to the identification of yellow wine original production. Chromatographic methods, such as gas chromatography (GC) [5], gas chromatography–mass spectrometry (GC–MS) [6], and high-performance liquid chromatography (HPLC) [7], have been reported in the flavor analysis and origin identification of wines. However, the requirement for tedious extraction, long time-consuming, expensive laboratory equipment, and strict working environment significantly limited the wide-spread use of these chromatographic methods. Spectroscopy methods, like near-infrared spectroscopy (NIRS) [8–10] and nuclear magnetic resonance (NMR) [11–13], combined with multivariate statistical methods, were also shown to be useful in the evaluation of food authenticity. Although those methods are affordable, simple, quick, and dependable, the complex analysis of testing data requires specialized software or complex algorithms, which is difficult for ordinary inspectors to master it. In addition, electronic nose is often used to analyze food aroma and has the advantages of simplicity, speed, wide range, good reliability, and reproducibility of detection data, and it still needs an electrode activation process, during which sensor poisoning...
may occur depending on operation and ambient conditions. To the best of our knowledge, there were few studies on wine origin identification in view of aroma of the sample.

Ion mobility spectrometry (IMS) is an emerging approach in electronic nose technology for food control [14–16]. This sensor involves the ionization of aroma compounds and subsequent drift through an electronic field in the drift tube at atmospheric pressure. In order to increase its selectivity for complex mixtures (like food and agriculture products), GC can be coupled to IMS, which provides a 2D map of the flavor compounds as a result of their retention characteristics in a GC column and their drift time when reaching the IMS detector [17, 18]. Therefore, gas chromatography–ion mobility spectrometry (GC–IMS) is a highly sensitive and selective combination of 2 techniques, detecting volatile compounds in the ppb down to the upper ppt range, which can be operated at atmospheric pressure, with no sample pretreatment, fast analysis time, and low detection limits. In recent years, GC–IMS has already been applied in many fields, such as food safety [19], process control [20], bio-marker research [21], drug detection [15], and human breath analysis [22]. These studies have proven that the quality information of samples can be well characterized for quantitative and qualitative analysis by analyzing their volatile organic compounds (VOCs).

However, characteristic variables selection in these studies is usually completed by researcher's subjective perspective without any visual comparison methods or criterion. To the best of our knowledge, no recent work has been conducted by using RGB color model for the guidance of the characteristic peaks selection of 2D data obtained from GC–IMS instrument.

This study aimed at using GC–IMS instrument and chromometric tools to identify the geographical origin of yellow wine based on the fingerprint of volatile organic compounds. In this article, an approach was developed based on a color mixing method to provide visual color differences between different 2D spectra, which could ease the feature selection from GC–IMS plot and further classification study. In addition, volatile organic compounds that are closely related to origin information of yellow wine were screened and identified, and a discriminant model was built for the identification of production places. Finally, a rapid method for recognition of yellow wine origin was developed. The method of processing 2D data could be applied to food quality detection using other combined equipment.

Experiment

Samples. As listed in Table 1, a total of 122 samples were collected from 3 major yellow wine production bases (Shaoxing, Shandong, and Hubei), and those samples were all bought in local market or online in China. Among those, 45 samples were from Shaoxing, which were all labeled with the product of protected designation of origin by Chinese National Standard GB 17946–2008. Forty samples were produced in Shandong, and the remaining 37 samples were from Hubei. Yellow wine samples were all stored in opaque glass bottles in the refrigerator at 5 °C before detection in order to keep their sensory properties.

GC–IMS System and Conditions. Analyses of yellow wine samples were performed on an advanced commercial GC–IMS (FlavourSpec®) system manufactured by Gesellschaft für Analytische Sensorsysteme mbH (G.A.S., Dortmund, Germany). This system was equipped with an automatic sampler unit (CTC-PAL, CTC Analytics AG, Zwingen, Switzerland) for 32 vials and furnished with a 1.0-mL Hamilton syringe (51 mm needle) and a heatable splitless injector with 2 mm ID, 6.5 mm OD × 78.5 mm fused quartz glass, as well as a radioactive ionization source (tritium) of 6.5 KeV.

For VOC analysis, each yellow wine sample (2 mL) was placed into a 20-mL headspace vial and closed with magnetic screw cap. After 10 min of incubation at 60 °C in an incubator, a headspace volume of 100 μL was automatically injected into the injector by a syringe heated to 80 °C for avoiding condensation effects. Before each analysis, the syringe was automatically flushed with a stream of gaseous nitrogen (purity ≥99.999%) for 2 min to avoid cross contamination. The separation was performed using a non-polar column constituted by 95% methyl-5% phenyl with 30 m of length. Nitrogen of 99.999% purity was used as carrier gas with a flow rate set at 2 mL/min for 5 min, and then linearly increased to 150 mL/min in the next 10 min, and maintained this flow until 20 min. After separation of VOCs in the capillary at 40 °C, those analytes were eluted in the isothermal mode and driven into the ionization chamber for ionization, prior to IMS analysis. Molecules were ionized using a tritium source, and the resulting ions were driven to the drift region via a shutter grid under the help of carrier gas. Drift tube length was 20 cm long and operated at a constant voltage of 400 V/cm, a temperature of 45 °C, and a drift gas flow rate of 150 mL/min (nitrogen).

Each spectrum was obtained with an average of 32 scans, the grid pulse width of 100 μs, the sampling frequency of 150 kHz, and the repetition rate of 21 ms. Data were acquired in the positive ion mode via the spectrometer’s built-in computer.

Data Analysis. A color model is an abstract mathematical model describing the way colors can be represented as tuples of numbers, typically as 3 or 4 values or color components. However, a color model with no associated mapping function to an absolute color space is a more or less arbitrary color system with no connection to any globally understood system of color interpretation. Adding a specific mapping function between a color model and a reference color space establishes within the reference color space a definite “footprint”, known as a gamut, and for a given color model, this defines a color space [23]. The RGB color system is an additive color model in which red, green, and blue lights are added together in various ways to reproduce a broad array of colors. To form a color with RGB, three light beams (one red, one green, and one blue) must be superimposed. Each of the three beams is called a component of that color, and each of them can have an arbitrary intensity, from fully off to fully on, in the mixture [24]. As it is known, the data obtained by GC–IMS from a sample was corresponding to a matrix and could be present in the form of a gray image or pseudo-color (or false color)

Table 1. Samples of the test sets

| Place of origin | Processing company | Number of samples |
|-----------------|--------------------|-------------------|
| Shaoxing        | Kuaishanyuan       | 8                 |
| Shandong        | Miaofulao          | 13                |
| Hubei           | Mipo                  | 10               |
|                 | Xiaoxie              | 5                 |
|                 | Fangxian             | 8                 |
|                 | Lulinwang            | 7                 |
|                 | Shenglong            | 3                 |
|                 | Jinggai              | 4                 |
|                 | East Shaoxing        | 8                 |
|                 | Tang-Song            | 7                 |
|                 | Yuepin               | 2                 |
|                 | Weiranlunyu          | 8                 |
|                 | Nu’er Hong           | 2                 |
|                 | Qili                  | 2                 |
|                 | Lanling              | 10                |
|                 | Xinhuaqing           | 7                 |
|                 | Shaoxing Kuaijishan  | 8                 |
|                 | Guyuelongshan        | 12                |
|                 | Lulinwang            | 10                |
|                 | East Shaoxing        | 8                 |
|                 | Yuanfong             | 2                 |

140

Unauthenticated | Downloaded 11/04/20 08:51 AM UTC
image based on computer visualization [25]. Therefore, a matrix (a gray image) could be mapped with a single color; meanwhile, the other matrices could also be painted by any other single color. Then, a new image for observing color differences could be composed by blending those monochrome images, which would ease the selection of characteristic peaks by the effect of the gamut.

Before building the model, the raw data were pretreated with a view to avoid possible variations among samples and a resulting misclassification of them. Firstly, in order to realize peak alignment, a normalization method was applied to each sample with regard to reaction ion peak (RIP, as an internal standard). In the next step, a smoothing procedure based on second order Savitzky–Golay filtering was applied. After that, the zone of each topographic plot, which contained the majority of the information obtained with both devices, was selected (retention time from 147.03 to 413.79 s and drift time from 9.066 to 14.100 ms). Once the pre-treatment was performed on both sample datasets, yellow wine samples were divided into 3 groups according to their origins, and the average spectrum (a matrix) of each group was calculated for color differences, which was used to characterize shared features of each category. Then, an additive color model was used to blend 3 matrices for comparing differences between images. Later, principal component analysis (PCA) was employed for the extraction of the most important relevant information and dimensionality reduction. Finally, a quadratic discriminant analysis (QDA) classifier was constructed to classify 3 different origins, which was used to discriminate origin information of yellow wine samples.

**Characteristics Peaks Selection.** For further analysis, characteristic variables need to be extracted from each topographic plot. However, the changes in all characteristic peaks between different samples were not obvious from the observation, which meant features that could characterize the differences between the 3 groups of yellow wine samples were hard to select. Therefore, a color mixing algorithm was used to blend those images for color difference analysis. Based on the number of the groups, it was easy to use basic primaries (red, green, and blue) to map each matrix. Firstly, the matrix from Shaoxing yellow wine samples were mapped on different red values according to intensity of the characteristic peaks (Figure 2A). On the other hand, the matrices from Shandong and Hubei samples were mapped on green (Figure 2B) and blue colors (Figure 2C), respectively. Then, a blending image was generated by using the additive primary method (Figure 2D). Finally, based on the RGB color model, a broad array of colors can be produced by adding primaries in various ways. This allows the color overlay to show the shared characteristic information and the single-color region to indicate the unique feature for the corresponding group. When red, green, and blue color are mixed or added together with proper intensity, white light was obtained. According to this principle, 19 regions were selected (see Figure 2D), and the height of those regions was used for further analysis. Chemical compounds corresponding to partial characteristic peaks were identified by NIST libraries combined with drift time database (IMS library) from G.A.S. company. The deviation range of retention index was set to ±6, and the drift time deviation range was set to ±0.01 in the library search process, all of which were at default values. In this case, only 12 characteristic peaks were hit, and the corresponding substance information was listed in Table 2. As shown in Figure 2D, yellow wine samples from Shandong area had their typical volatile organic compounds (markers 1, 2, 4, and 6), which were all shown in green color. Some peaks (markers 5, 7, 8, and 9) shared components between different groups. Because of the different values of each color channel, those regions were displayed in light cyan in the mixed image. In addition, the region labeled with red arrow was not captured. One reason for this was that the initial flow of the

---

**Figure 1.** GC–IMS average topographic plot of yellow wine samples from Shaoxing (A), Shandong (B), and Hubei (C)
carrier gas was low, which resulted in incomplete separation of some volatile organic compounds. Another reason was that this region indicated a light color, which means that the total content of those compounds might be the same basically and had no obvious effect on discriminating categories.

**PCA Results.** After characteristic peak selection, a total of 19 regions were marked in each yellow wine sample. Then, the height property set of those peaks of each sample in those characteristic regions was selected as variables for further chemometric analysis, which means that 19 variables were used to characterize the quality of each yellow wine sample. Then, a PCA method was performed on the database (122 samples × 19 variables), and the explained variance by each principal component was shown in Figure 3. As shown, the score plot of the first two principal components accounted for 89.52% of the total variance in raw data (PC1 = 89.52% and PC2 = 17.36%), which explained the variation trend of the dataset and fully complied with the requirement of the PCA. By analyzing the distribution of data mapped on the principal components, it was visible that the data points with different markers belonged to different types of yellow wine samples. The smaller the space occupied by one cluster and the larger the distances among different clusters, the easier the recognition and classification result. A clear separation between samples from Shaoxing and the other 2 groups could be observed, and the cluster region of Shaoxing yellow wine samples was narrow and long, which could be correlated with the differences of origins and the processing technique of the collected samples. At the same time, the loading matrix (the black lines and corresponding peaks in Figure 3) was also visualized. The loading matrix is the projection of features on principal components, which can be used to study the correlation and importance between different features. As shown, compared with other characteristic peaks, the features (markers 5, 6, 12, 13, and 19) were more important aroma compounds, because their positions were far from the coordinate origin and were between different types of origin, indicating that those characteristic aroma substances can be used to distinguish yellow wine from different origins. For example, compounds corresponding to peaks 6 and 19 were the main differences in flavor between Shaoxing and Hubei samples. On the other hand, although there were many characteristic peaks between Shandong and Hubei samples,

![Figure 2. True color images of yellow wine samples from Shaoxing (A), Shandong (B), and Hubei (C), and color mixing of those 3 images (D)](image)

Table 2. Chemical compounds corresponding to partial characteristic peaks

| Marker No. | Compound name          | Drift time (ms) | Retention time (s) |
|------------|------------------------|-----------------|--------------------|
| 1          | Benzaldehyde           | 323.896         | 1.4636             |
| 2          | 1-Hexanol              | 242.333         | 1.6469             |
| 3          | Furfural               | 229.831         | 1.3277             |
| 4          | Acetic acid, butyl ester | 200.272       | 1.6066             |
| 6          | N-propionic acid       | 188.554         | 1.3426             |
| 7          | Methyl Isobutyl Ketone | 181.432         | 1.4742             |
| 8          | Ethyl butanoate        | 212.908         | 1.5538             |
| 9          | (E)-3-Hexen-1-ol       | 221.639         | 1.5337             |
| 12         | Ethyl propionate       | 177.526         | 1.4463             |
| 13         | Butan-2-one            | 148.806         | 1.2418             |
| 15         | (Z)-3-Hexen-1-ol       | 254.494         | 1.2293             |
| 16         | 3-Methylbutanal        | 161.213         | 1.1957             |
those peaks were close to the origin of the coordinate system, showing less importance and high correlation. In order to verify the reliability of above analysis, quality control (QC) samples were incorporated in the PCA analysis. The number of samples from each origin was 8 (Shaoxing), 5 (Shandong), and 6 (Hubei). As shown, QC samples that were added later could also be well distinguished in the new coordinate system and each origin of them had its location in each clustering area. However, the cluster boundary between Shandong and Hubei yellow wine samples was not clear. It could be observed that some samples overlapped each other in two different groups, which might be due to the small differences in aroma composition between them. Therefore, pattern recognition needs to be used to build a non-linear model for the origin identification.

**Classification Results Obtained by QDA.** QDA was applied to sharpen the separation between yellow wine samples from different origins and build a discriminant model. The analysis of the dataset transformed by using PCA method has got the convincing visualization of clusters that correspond to different origins of yellow samples, which gives a good insight for application of a QDA classifier. A total of 122 samples were respectively divided into calibration set (70%) and prediction set (30%) in each group, which meant that 79 samples (31 from Shaoxing, 28 from Shandong, and 20 from Hubei) were selected as the calibration set, and the remaining 43 samples (14 from Shaoxing, 12 from Shandong, and 17 from Hubei) were used as the prediction set. Those samples from calibration and prediction were randomly selected. The calibration set was used to train the QDA classifier, and then the trained classifier was used to test the prediction data set. Finally, the accuracy of classification was used to evaluate the performance of the built classifier. The statistical results were summarized in Table 3. As can be observed, all yellow wine samples have been classified without any error in the calibration set, and only two yellow wine samples from Hubei were misclassified. On the other hand, a sample from Shandong and one from Hubei were misjudged in the prediction set. The accuracy of the trained set was 97.47%, and the accuracy of the tested set was 95.35%. Therefore, our present work verifies that GC–IMS technology combined with the additive color model and chemometrics can be used to recognize different origins of yellow wine samples.

**Conclusion**

This work described a procedure to discriminate Chinese yellow wines from different origins based on the compositions of flavor compounds. An additive color model was used for characteristic-peak selection, and the PCA results indicated the internal differences between samples. Classification model developed by QDA achieved correct classification rate of 95.35% in the prediction set. Overall, it can be concluded that the flavor fingerprint can be used as an effective tool for yellow wine authentication. However, the work reported here is just a feasibility study. In further studies, more samples from other wineries and varieties should be incorporated to build a more robust, reliable and accurate model before practical application.

**References**

1. Zhou, W.; Fang, R.; Chen, Q. *Food Chem.* 2017, 233, 174–181.
2. Wu, P.; Cai, C.; Shen, X.; Wang, L.; Zhang, J.; Tan, Y.; Jiang, W.; Pan, X. *Food Chem.* 2014, 152, 108–112.
3. Que, F.; Mao, L.; Zhu, C.; Xie, G. *LWT - Food Sci. Technol.* 2006, 39, 111–117.
4. Moreno, I. M.; González-Weller, D.; Gutiérrez, V.; Marino, M.; Cameán, A. M.; González, A. G.; Hardisson, A. *Talanta* 2007, 72, 263–268.
5. Luykx, D. M. A. M.; van Ruth, S. M. *Food Chem.* 2008, 107, 897–911.
6. Luo, T.; Fan, W.; Xu, Y. *J. Inst. Brew.* 2008, 114, 172–179.
7. Yu, H.; Zhao, J.; Li, F.; Tian, H.; Ma, X. J. Chromatogr. B 2015, 997, 129–135.
8. Shen, F.; Ying, Y.; Li, B.; Zheng, Y.; Zhuge, Q. Food Chem. 2011, 129, 565–569.
9. Shen, F.; Yang, D.; Ying, Y.; Li, B.; Zheng, Y.; Jiang, T. Food and Bioprocess Technology 2012, 5, 786–795.
10. Shen, F.; Yang, D.; Ying, Y.; Li, B.; Zheng, Y.; Jiang, T. Discrimination Between Shaoxing Wines and Other Chinese Rice Wines by Near-Infrared Spectroscopy and Chemometrics 2012.
11. Mileczarek, R. R.; Liang, P.-S.; Wong, T.; Augustine, M. P.; Smith, J. L.; Woods, R. D.; Sedej, I.; Obien, C. W.; Vilches, A. M.; Haefl, R. P.; Freece, J. E.; Breksa, A. P. Postharvest Biology and Technology 2019, 149, 50–57.
12. Košir, I. J.; Kidrič, J. Anal. Chim. Acta 2002, 438, 77–84.
13. Hayasaka, Y.; Black, C. A.; Hack, J.; Smith, P. Food Chem. 2017, 230, 99–107.
14. Zhang, L. X.; Shuai, Q.; Li, P. W.; Zhang, Q.; Ma, F.; Zhang, W.; Ding, X. X. Food Chem. 2016, 192, 60–66.
15. Wang, W. G.; Liang, X. X.; Cheng, S. S.; Chen, C.; Meng, W.; Peng, L. Y.; Zhou, Q. H.; Haiyang, L. J. Chinese Journal 2014, 59, 1079.
16. Karpas, Z. Food Res. Int. 2013, 54, 1146–1151.
17. Vautz, W.; Franzke, J.; Zampolli, S.; Elmi, I.; Liedtke, S. Anal. Chim. Acta 2018, 1024, 52–64.
18. Cavanna, D.; Zanardi, S.; Dall’Asta, C.; Sunan, M. Food Chem. 2019, 271, 691–696.
19. Garrido-Delgado, R.; Mercader-Trejo, F.; Sielemann, S.; de Bryun, W.; Arce, L.; Valcárcel, M. Anal. Chim. Acta 2011, 696, 108–115.
20. Gerhardt, N.; Birkenmeier, M.; Schwalow, S.; Rohm, S.; Weller, P. Anal. Chem. 2018, 90, 1777–1785.
21. Thompson, R.; Perry, J. D.; Stanforth, S. P.; Dean, J. R. Microchem. J. 2018, 140, 232–240.
22. Vautz, W.; Slodzynski, R.; Hartharan, C.; Seifert, L.; Nolte, J.; Fobbe, R.; Sielemann, S., Lao, B. C.; Huo, R.; Thomas, C. L. Anal. Chem. 2013, 85, 2135–2142.
23. Busin, L.; Vanbendaouwe, N.; Macare, L., "Color Spaces and Image Segmentation", in Advances in Imaging and Electron Physics, Hawkes, P. W., Ed., Elsevier, 2009, pp. 65–168.
24. Hollingsworth, B. V.; Reichenbach, S. E.; Tao, Q.; Visvanathan, A. J. Chromatogr. A 2006, 1105, 51–58.
25. Sheng, Z.; Jiang, X. Y.; Zhen, W. Adv. Mater. Res. 2012, 403, 1618–1621.