# Chemical Analysis of Chardonnay Wines: Identification and Analysis of Important Aroma Compounds


## ABSTRACT

Chardonnay wines from leading world producers were selected for analysis and comparison of volatile compounds that contribute to the sensory qualities of the wines. Wines were analyzed in triplicate using a random block design and all analysis were conducted on the same day as sensory descriptive profiling. Samples were extracted using a solid phase sorbent and subject to analysis using a GC-MS configured in scan mode thereby producing a 'fingerprint' of composition. Peak areas representative of compound concentration, and mass spectra representing compound identification, were extracted from the sample profiles. Peak areas of the identified compounds were used to elucidate common groupings (price, country of origin) of wines based upon composition. The method used was deemed to be a suitable approach for analysis based upon the tight grouping of wine sample triplicates in PCA score plots. When samples were plotted and identified based upon country of origin, Australian wines were diverse and no specific groupings were evident. This is in contrast to wines from New Zealand, France, Chile and the United States of America. The lack of apparent clustering of the Australian samples must be considered in light of the selection process of both domestic and imported wines. Australian wines were chosen to represent diverse styles and thus the compositional elements of these wines are not necessarily representative of the styles of wines available from imported producers at the same price points.

## INTRODUCTION

In this investigation we report a on the analysis of a GC-MS data set from chardonnay wines derived from the important wine growing regions of the world. A semi-automated approach to analyzing the GC-MS data profiles of the wines was used to extract peak areas of volatile compounds and ion abundance ratios for compound identification. This effectively establishes a 'fingerprint' of the volatile composition of the wines and can be used for the purposes of Principal Component Analysis to elucidate patterns or clusters of wines with respect to volatile composition.

## Materials and method

Wine samples were collected and extracted for analysis on the day of sensory panel appraisal so that variation between sensory and analytical replicates was minimized. A total of 25 different wines were subjected to descriptive sensory assessment in triplicate over the same period that sensory evaluations were conducted. For the determination of volatile compositions all wines were assessed, extracted and analyzed in random order within replicate blocks. Prior to extraction 50mL of wine was spiked with 50µL of internal standard mixture comprising 3-tert-butyl-4-hydroxyanisole; methylisobutyl ketone and 4-methyl-2-pentanone to give final concentrations of 4000, 1000 and 4000 µg $L^{-1}$ respectively. Samples were extracted with LiChrolute cartridges (Merck) following the method of Lopez et al (*1*) and the eluant spiked with an internal standard mix comprising 2-octanol and 4-hydroxy-4-methyl-2-pentanone to give final concentrations of 6650 and 6650 µg $L^{-1}$ respectively.

## GC/MS analysis

For SPE analysis 1µL of extract was injected into an Varian 3800 gas chromatograph fitted with a 60 m x 0.25 mm i.d. fused-silica capillary column with a 0.25 µm wax (DB-WAXetr) stationary phase (J&W Scientific, Folsom, CA) using a CombiPAL autosampler. The injector temperature was 240°C; and the split ration 10:1. The helium flow rate through

the column was 1.5 mL $min^{-1}$ with an average velocity of 31 cm $sec^{-1}$; the column temperature was held at 40°C for 5 min, then increased at 2°C $min^{-1}$ to 210°C and held for 20 min. Mass spectra were collected using a Saturn 2000 ion trap set to EI AGC ionization mode and a scan time of 0.65 sec, in the range of 35 to 350 *m/z*. A maximum ionization time was set to 25000 microseconds and target TIC at 20000 counts. Mass range segments (10 to 99; 100 to 249; 250 to 350) were set to an ion storage level of 35.0 *m/z*, ionization time at 100% and background mass of 45 *m/z*. Transfer line, manifold and trap temperatures were 230°C; 40°C and 180°C respectively. Spectra were not collected between 7.30 and 10.00 min to exclude solvent peaks in the chromatogram.

## Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) Analysis of GC/MS data

All GC/MS files were exported in three dimensional CSV or ASCII format from MS Workstation SP1 version 6.5 for processing in MATLAB version R2011b (Mathworks, Natick, MA) where all data treatments procedures were performed. The total ion chromatogram (TIC) for each sample was overlaid without spectral alignment to enable visual identification of appropriate time windows for automated processing. Once the time windows were established for each data set, files were processed. Each *m/z* channel was smoothed by subtracting the minimum *m/z* value and then filtered by convolution using a vector of size 7. A smoothed and corrected TIC for each sample was then determined from the sum of each m/z channel in the time window and this profile was used for alignment of elution profiles using the maximum cross-correlation approach. The sample shift from the alignment of the TIC was then applied on a sample by sample basis to each *m/z* channel to align all sample elution profiles within the time window thereby creating an aligned three dimensional matrix (elution time*m/z channel*sample). A two dimensional matrix representing the mean m/z elution profile was then used to determine the number of

interesting features (peaks) within the time window. The number of features was determined by principal component analysis (PCA) using the singular value decomposition of the transformed matrix; an offset of 1 was added to all *m/z* channels which were then down weighted by log10 transform; mean centered and variance scaling using the Pareto equation prior to PCA.

Spectra associated with each peak of interest was estimated with the SIMPLISMA approach using the PURE algorithm. Estimated spectra were then passed along with the column augmented elution profile of all samples (elution time*(samples*m/z)) in the time window to the MCR-ALS algorithm.

Spectral libraries of peaks derived from MCR-ALS were exported from Matlab in a format compatible with the National Institute of Standards and Technology (NIST) Mass Spectral Search Program (version 2.0). Target compound identification of peaks of interest and internal standards was conducted using NIST search results and confirmed using published retention time indices and/or pure compounds run in identical GC/MS conditions.

## Exploratory Data Analysis of GC-MS Peaks to Identify Interesting Sample Groups

Peak areas from the MCR-ALS decomposition of SPE GC-MS data sets were subject to exploratory data analysis. Peak areas were preprocessed by normalizing all peaks in the data sets to the peak area associated with the injection internal standard (SPE: 4-hydroxy-4-methyl-2-pentanone) to eliminate minor injection discrepancies between samples. All peak areas were then normalized to the peak area for the internal standard for extraction (SPE: 3-tert-butyl-4-hydroxyanisole) to eliminate discrepancies associated with extraction efficiency. Preprocessing of peak areas was done by addition of an offset of one, logarithmic transform, mean centre and variance scaling prior to PCA. Peaks associated with the internal standards were excluded from the PCA. The most significant peaks associated with extracted PCs were

identified by the loading values. Peaks in the top 50 percentile of the absolute loading values are considered to be the most significant.

## RESULTS and DISCUSSION

The extracted peak spectra from the GC-MS files were subjected to identification using the National Institute of Standards and Technology (NIST) Mass Spectral Search Program (version 2.0). Excluding the internal standards that had been spiked into each sample to assist in normalizing injection volumes and extraction efficiencies, a total of 19 compounds could be identified with certainty. The identity of these compounds (Table 1) was established using a forward and reverse minimum search result exceeding 800 from the NIST database; matching published retention indices for a similar GC column stationary phase determined using alkane standards; and where available injecting pure standards of the compounds of interest to confirm identity.

The extracted peak areas ranged in several orders of magnitude and thus logarithmic transformation and Pareto scaling was considered appropriate to moderate the influence of compounds with large mass selective device responses relative to concentration and this approach is frequently employed for metabolomic studies (*2*). The results of the PCA are presented in table 2. Five PCs were considered appropriate to model the data based upon the eigenvalues of the correlation matrix decomposition. A cumulative percentage of approximately 55% of data variance is modeled with these PCs (table 2). PCA of the peak area table, excluding the internal standards, shows that wines are positioned along PC1 and PC2 in clusters for each wine triplicate (figures 1 and 2). There is good clustering for each triplicate of wine samples indicating that the extraction, GC-MS analysis and MCR-ALS analysis procedures are robust and free from spurious results. The corresponding loading bar chart for PC1 identifies peaks with the most significant influence upon the PCA. Wine

sample triplicates are plotted blinded (Figure 1) and with wine identity (Figure 2) in scores plots. Apart from the relative positioning and clustering of the wine triplicates used in the study, little insight can be derived from examination of these plots without intimate knowledge of the wine styles and reasoning for inclusion in the study. It can be noted that Australian wines that are unwooded appear in the centre of the scores plot indicating a relative neutral composition in terms of the indentified wine compounds.

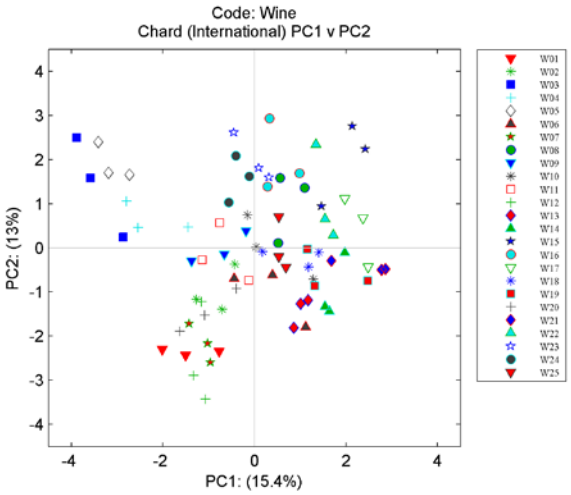


**Figure 1. Scores plot of PCA of peak area table with triplicate samples coded for blind wine code.**

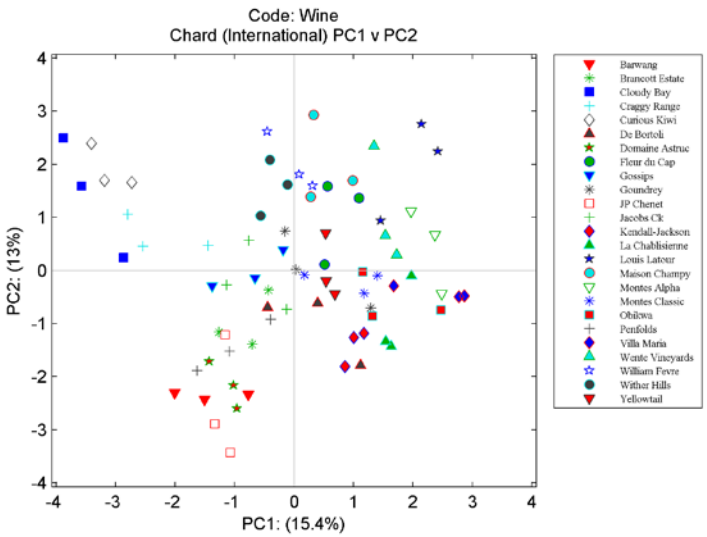

**Figure 2. Scores plot of PCA of peak area table with triplicate samples coded for wine identity.**

Sample scores plots (PC1-PC5; figures 3-6) with samples identified for price and country of origin with the associated loadings plots for variables offer greater insight to the relative compositional aspects of the wines. Clearly evident in figure 3 are specific wine sample groups based on country origin with wines from New Zealand tightly clustered in the negative portion of PC1; Chilean, South African and US wines in the positive dimension of PC1; and French wines appear in two clusters with separation along the PC2 dimension. Of interest is the spread of samples from Australia and this is consistent with the results of sensory evaluation PCA. The relative clusters of these wines according to country of origin must be considered in light of the selection process for wines used the study. Wines from Australia were selected on the basis of expert opinion of wine professional and represent a range of styles. Consideration must also be given to the variance captured within each PC. There is no extracted PCs that capture large amounts of variation in the data; PC1 accounts for only approximately 15% of variance, and this is again suggestive that the diversity of wine styles purposely used in the study has impacted the overall data such that no specific wine style is overly represented. Imported wines were chosen on the basis of style, price point and availability. Loadings for PC1 and PC2 show the relative concentrations of compounds associated with the extracted PCs. Wines positioned in the negative dimension of PC1 have high loadings for phenyl ethanol and ethyl octanoate suggesting these wines have attributes of honey, floral and fat; wines with positive PC1 scores have high attributes of resin, green, apple peel and fruit, characters associated with hexanol and ethyl butyrate. Wines positioned in the positive portion of PC2 may also have high attributes of fruit and fat aroma based upon the association with ethyl octanoate for this PC; and wines with negative

scores in PC2 will have higher grape associated aroma i.e. higher relative concentrations of ethyl decanoate.

Groupings of wines in the PC3 PC4 and PC5 dimensions are not as clear in terms of associated patterns of wine country of origin or price. The results from this investigation must be qualified in terms of the small number of identified compounds in the sample extracts. Further work, including a more comprehensive identification of peaks associated with volatile aromas of these wines is thus warranted to assist in the elucidation of wine composition and how the volatile component of the wine determine style. It can be concluded that Australian styles of Chardonnay are well represented in terms of the volatile components and style diversity when compared to Chardonnay wines of similar price points from other world producers.

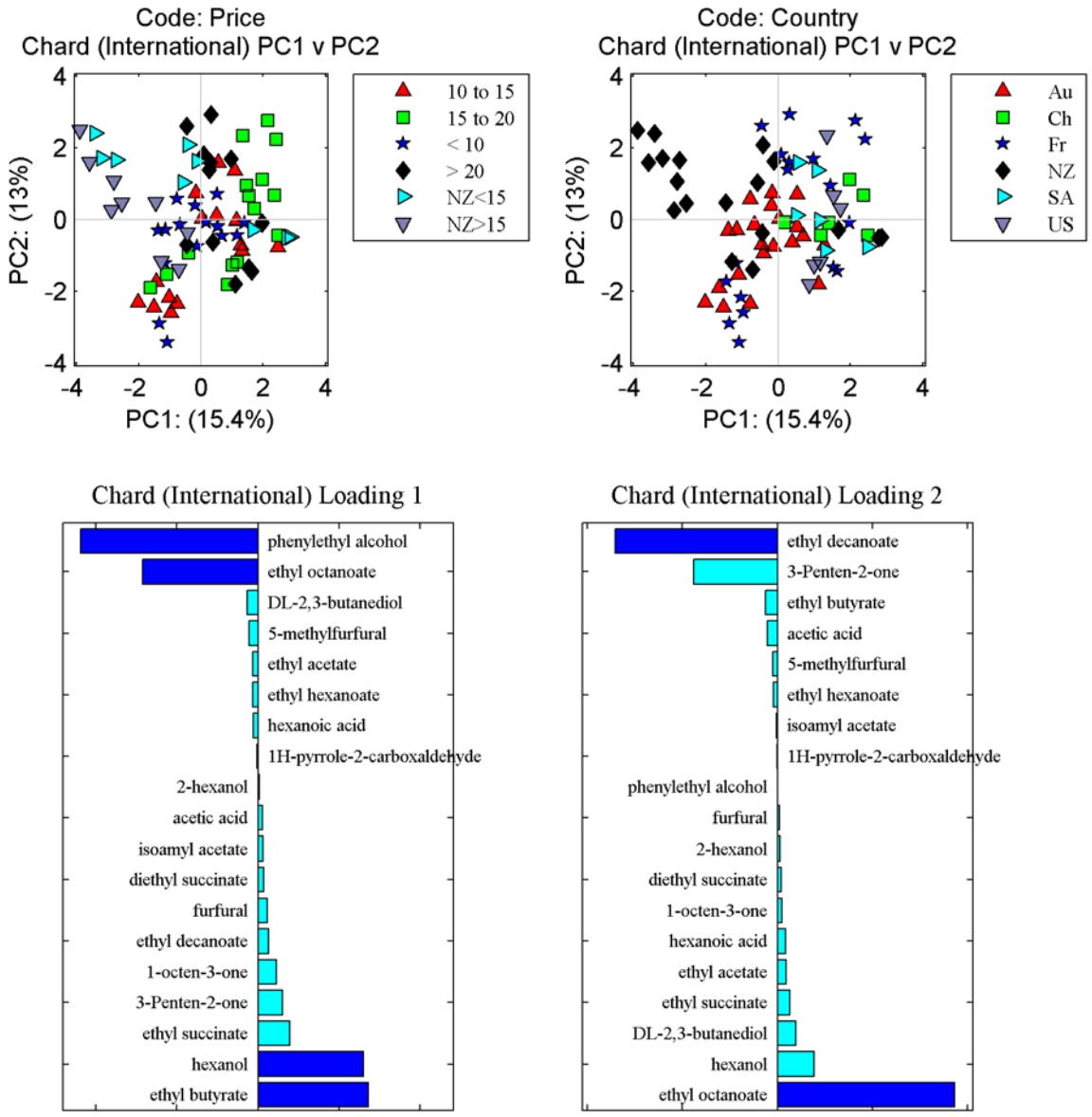


**Figure 3. Sample scores plot (PC1 versus PC2) coded for price and country of origin; and associated loadings plots (PC1 & PC2) for identified compounds. Dark colored loadings are in the top 50 percentile in size.**

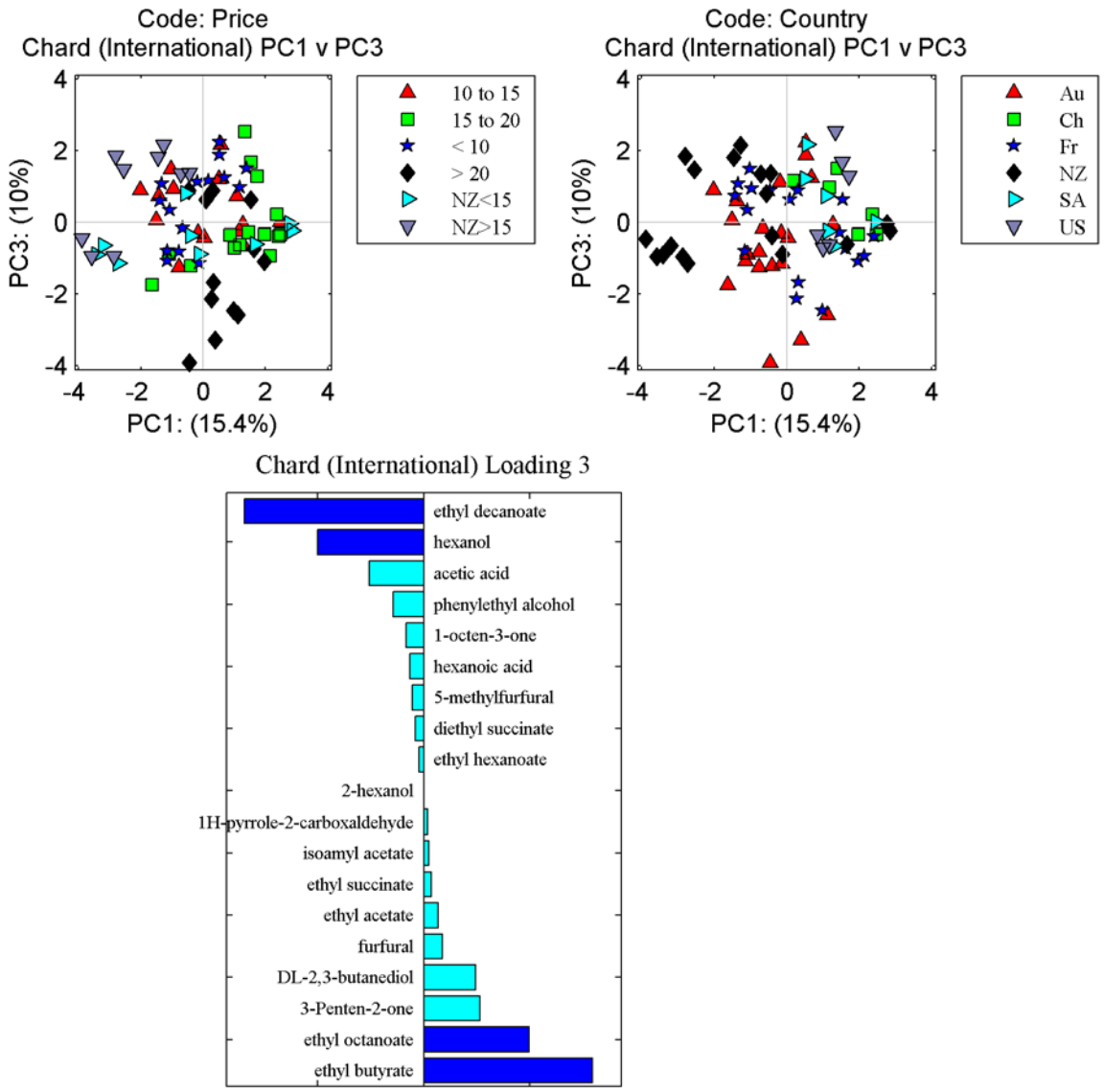


**Figure 4.** **Sample scores plot (PC1 versus PC3) coded for price and country of origin; and loadings plot (PC3) for identified compounds. Dark colored loadings are in the top 50 percentile in size.**

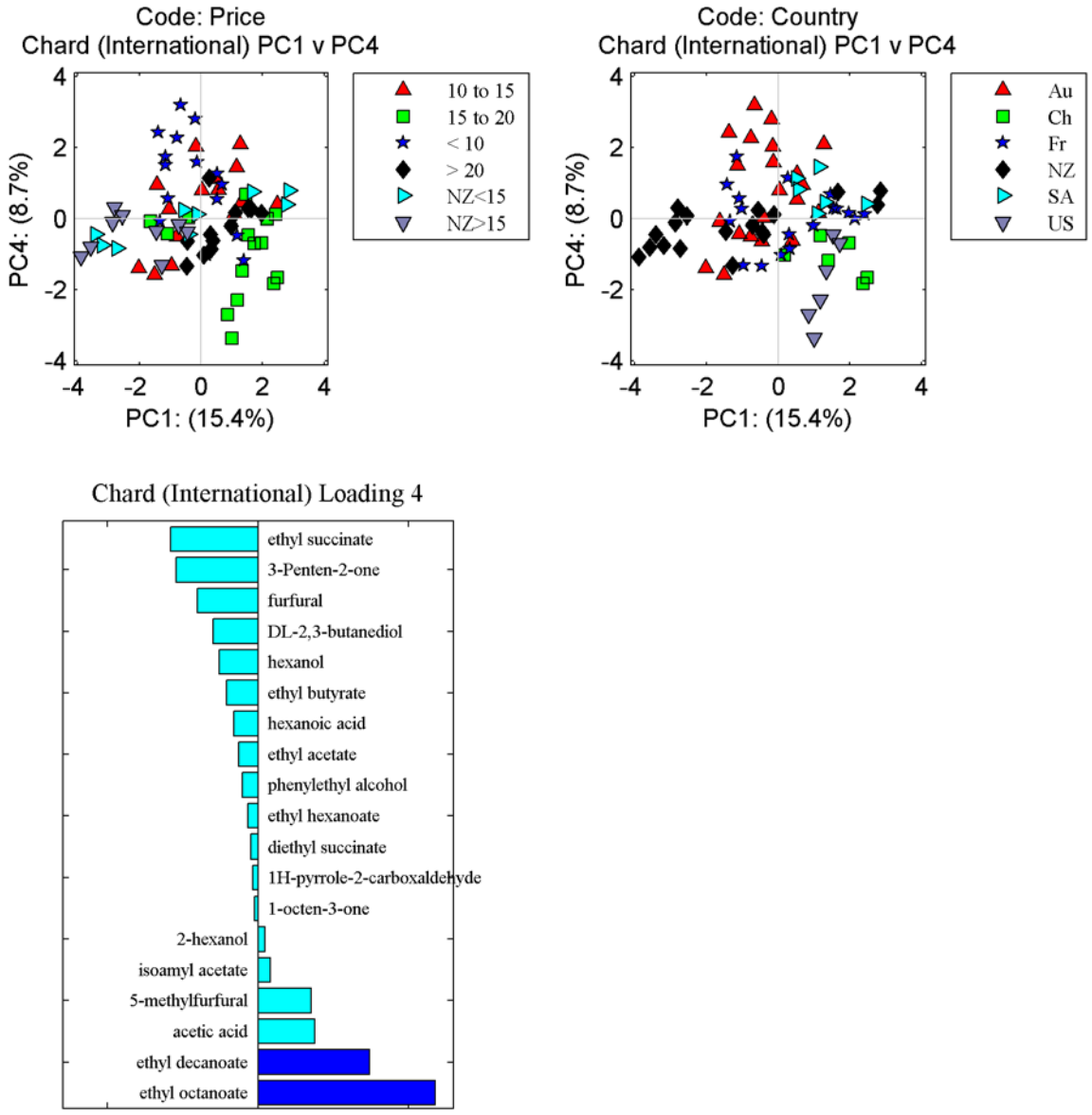


**Figure 5. Sample scores plot (PC1 versus PC4) coded for price and country of origin; and loadings plot (PC4) for identified compounds. Dark colored loadings are in the top 50 percentile in size.**

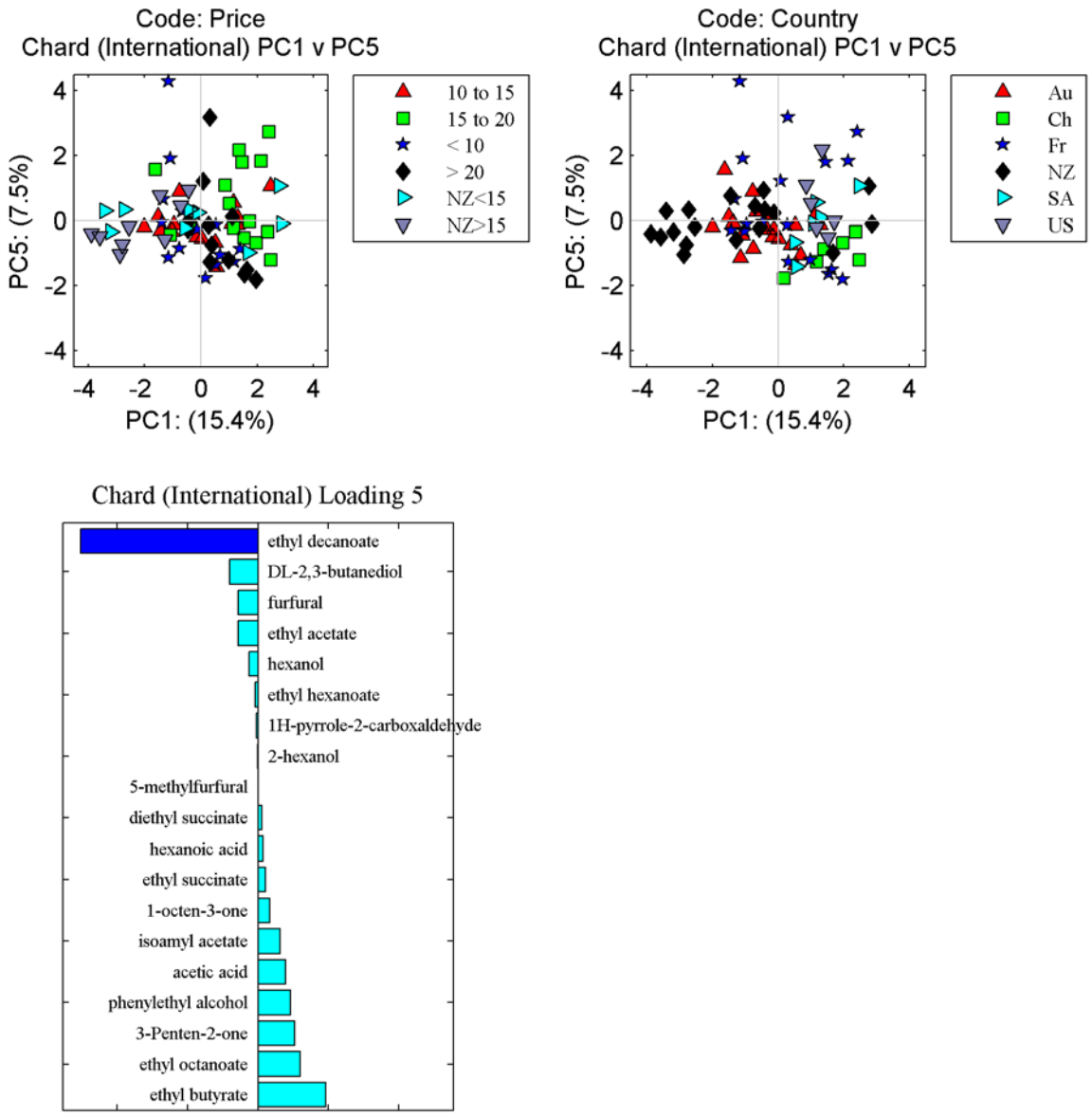


**Figure 6. Sample scores plot (PC1 versus PC4) coded for price and country of origin; and loadings plot (PC4) for identified compounds. Dark colored loadings are in the top 50 percentile in size.**

**Table 1. Peak table identification for compounds extracted from Chardonnay samples (international data set)**

| Compound | CAS | Kovats RI (observed) | Kovats RI (Reported) (*3*) | Identified† | Aroma (www.flavornet.org) |
|---|---|---|---|---|---|
| ethyl acetate | 141-78-6 | 885 | 865 | RI, MS cmp | nail polish, pineapple |
| ethyl butyrate | 105-54-4 | 995 | 1022 | RI, MS cmp | apple |
| isoamyl acetate | 123-92-2 | 1060 | 1118 | RI, MS cmp | banana |
| 2-hexanol | 626-93-7 | 1195 | 1228 | RI, MS cmp | leaf, green fruit |
| 3-penten-2-one | 3102-33-8 | 1145 | 1135 | RI, MS | fruity |
| 1-octen-3-one | 4312-99-6 | 1201 | 1300 | MS, cmp | mushroom, earthy (barata) |
| ethyl hexanoate | 123-66-0 | 1230 | 1224 | RI, MS cmp | apple peel, fruit |
| hexanol | 111-27-3 | 1330 | 1351 | RI, MS cmp | resin, flower, green |
| acetic acid | 94-19-7 | 1420 | 1434 | RI, MS cmp | vinegar |
| furfural | 98-01-1 | 1440 | 1458 | RI, MS cmp | bread, almond |
| ethyl octanoate | 106-32-1 | 1464 | 1422 | RI, MS cmp | fruit, fat |
| DL-2,3-butanediol | 6982-25-8 | 1510 | 1492 | RI, MS cmp | butter |
| 5-methylfurfural | 620-02-0 | 1520 | 1560 | RI, MS cmp | almond, spice, caramel |
| diethyl succinate | 123-25-1 | 1680 | 1705 | RI, MS cmp | |
| ethyl decanoate | 110-38-3 | 1620 | 1641 (*4*) | RI, MS cmp | grape |
| phenylethyl alcohol | 60-12-8 | 1785 | 1795 | RI, MS cmp | honey, floral |
| hexanoic acid | 142-62-1 | 1820 | 1847 | RI, MS cmp | |
| 1H-pyrrole-2-carboxyaldehyde | 1003-29-8 | 2150 | 2032 | RI, MS | |
| ethyl succinate | 1070-34-4 | 2395 | 2440 | RI, MS cmp | |

† RI = matching retention index for similar stationary phase

MS = NIST mass spectral data base match with minimum forward and reverse matching of 800

cmp = matched compound with pure standards

**Table 2. Principal component diagnostics for peak area table.**

| Principal Component Number | Eigenvalue of Correlation Matrix | Percent of Variance Attributed to PC | Percent Cumulative Variance |
|---|---|---|---|
| 1 | 2.59 | 15.40 | 15.40 |
| 2 | 2.18 | 12.97 | 28.37 |
| 3 | 1.69 | 10.03 | 38.40 |
| 4 | 1.47 | 8.75 | 47.15 |
| 5 | 1.26 | 7.46 | 54.61 |

# References


1. López, R.; Aznar, M.; Cacho, J.; Ferreira, V., Determination of minor and trace volatile compounds in wine by solid-phase extraction and gas chromatography with mass spectrometric detection. *J. Chromatogr. A* **2002,** *966*, 167-177.
2. Wiklund, S.; Johansson, E.; Sjostrom, L.; Mellerowicz, E. J.; Edlund, U.; Shockcor, J. P.; Gottfries, J.; Moritz, T.; Trygg, J., Visualization of GC/TOF-MS-Based Metabolomics Data for Identification of Biochemically Interesting Compounds Using OPLS Class Models. *Anal. Chem.* **2008,** *80*, 115-122.
3. El-Sayed, A. M. The Pherobase: Database of Insect Pheromones and Semiochemicals. http://www.pherobase.com/database/kovats/kovats-index.php (07/05/2010),
4. Rebière, L.; Clark, A. C.; Schmidtke, L. M.; Prenzler, P. D.; Scollary, G. R., A robust method for quantification of volatile compounds within and between vintages using headspace-solid-phase micro-extraction coupled with GC-MS - Application on Semillon wines. *Anal. Chim. Acta* **2010,** *660*, 149-157.