Data Dive: Sensory Analysis of LAFFORT Product Treatments on Smoke-Exposed Grapes

Daniel A. Dycus

Contents

1 Introduction

This is the notebook for the sensory data from wines made from a study on smoke exposed grapes in 2021. This data was acquired by performing over 120 micro-fermentations in accordance with the micro-fermentation protocol from the Australian Wine Research Institute (AWRI). 32 different wines were made from 4 different varietals from 4 different American Viticulture Areas (AVAs) inside North America. Each region represents an area impacted by the fires on the Western Coast of North America in the Fall of 2020. The fermentations represent controls and treatments from products supplied by Laffort. Quantitative Analysis of compounds related to the sensory character of wines was performed by Excell Laboratories in Bordeaux France using liquid-liquid extraction, acid-mediated hydrolysis, and Gas Chromatography / Mass Spectrometry (GC/MS). Then, in February 2021, we constructed a virtual sensory platform where 288 winemakers, researchers, and wine industry professionals gathered using a web portal. During this virtual wine tasting each panelist rated the intensity of fifteen wine sensory attributes of various treatment sets. This information was gathered and a mainframe was constructed by Daniel A. Dycus, Technical Manager for Laffort in North America. This notebook is constructed using a blend of programming environments including R and Python. This document was written in LaTeX. The code and mainframes are not included in this portion of the study but are hosted on the N. American Laffort server and available upon request.

1.1 Exploratory Data Analysis

The data frame comes in and we can see the wine type and treatments are presented as Factors. The other 15 variables are the Sensory Characters which are listed in Table 1. 1993 rows across 15 attributes is 29,895 data points. However, upon first glance we can see a problem with the data set. Some people didn't vote every time. We don't know why, but maybe the slider on the web app wasn't intuitive enough, or the sensory attribute wasn't a great descriptor, perhaps still another possibility is that the error in recording the intensity rating is completely random. We aren't sure, but before we perform analysis, we'll need to clean the dataset in a way which won't impact the results of our significance testing.

When we examine the main dataframe, we see these NA values which will need to be addressed. Before we can perform any methods for analysis, our dataset must be cleaned and prepared. First, lets look at a summary of the data and then look at the missing values or gaps in the dataset. NA values do not possess the required characteristics for computation. Therefore, in order to proceed, we'll use three methods for filling in these gaps and provide insight into these methods and how they affect the dataframe and the resulting computations and analysis.

Table 1: Descriptive Stats of Data Frame

Aroma	Desc. Stats	Flavor	Desc. Stats	Mouthfeel	Desc. Stats
Fruity	Min. : 0.000	Fruity	Min. : 0.000	Astringent	Min. : 0.000
Fruity	Median : 7.000	Fruity	Median : 7.000	Astringent	Median : 7.000
Fruity	Mean : 6.787	Fruity	Mean : 6.585	Astringent	Mean : 7.236
Fruity	Max. :15.000	Fruity	Max. :15.000	Astringent	Max. :15.000
Fruity	NA's :21	Fruity	NA's :32	Astringent	NA's :42
Spicy	Min. : 0.000	Spicy	Min. : 0.000	Bitter	Min. : 0.00
Spicy	Median : 6.000	Spicy	Median : 5.000	Bitter	Median : 6.00
Spicy	Mean : 5.829	Spicy	Mean : 5.625	Bitter	Mean : 6.48
Spicy	Max. :15.000	Spicy	Max. :15.000	Bitter	Max. :15.00
Spicy	NA's :43	Spicy	NA's :53	Bitter	NA's :44
Smoky	Min. : 0.00	Smoky	Min. : 0.000	Round	Min. : 0.000
Smoky	Median : 5.00	Smoky	Median : 6.000	Round	Median : 6.000
Smoky	Mean : 5.49	Smoky	Mean : 6.104	Round	Mean : 6.009
Smoky	Max. :15.00	Smoky	Max. :15.000	Round	Max. :15.000
Smoky	NA's :73	Smoky	NA's :78	Round	NA's :48
Ashy	Min. : 0.00	Ashy	Min. : 0.000	Lingering.Ash	Min. : 0.000
Ashy	Median : 4.00	Ashy	Median : 6.000	Lingering.Ash	Median : 7.000
Ashy	Mean : 4.92	Ashy	Mean : 6.537	Lingering.Ash	Mean : 6.922
Ashy	Max. :15.00	Ashy	Max. :15.000	Lingering.Ash	Max. :15.000
Ashy	NA's :107	Ashy	NA's :91	Lingering.Ash	NA's :79
Tar	Min. : 0.000	Tar	Min. : 0.000	Metallic	Min. : 0.000
Tar	Median : 4.000	Tar	Median : 5.000	Metallic	Median : 5.000
Tar	Mean : 4.443	Tar	Mean : 5.458	Metallic	Mean : 5.636
Tar	Max. :15.000	Tar	Max. :15.000	Metallic	Max. :15.000
Tar	NA's :146	Tar	NA's :119	Metallic	NA's :132

2 Data Cleaning and Feature Engineering

This section describes three methods for data cleaning. In this section, we perform the following three methods for dealing with gaps in our data set:

- Removing NA Values
- DINEOF functions for interpolating NA values
- Nearest neighbors for imputing NA values

There are 1108 missing data points which are listed as NA. We investigate further by using a counter with a boolean operator known as "is.NA" and then take the sums of each of the columns. Once we see what data is missing, we can determine what to do with the missing values. We could remove the data, that is we could remove the row(s) entirely. We could also impute the data, or replace the missing value with substituted values. That is, we could fill in the missing data with the most common value, the average value, etc. The advantage for using this method is that we don't lose important data which might be important for the row of data. The disadvantage, is the added extra layer of uncertainty to the data set as it is now based on estimates. In this particular case, about 20% of the responses are NA. We can remove them and still possess 80% of the original data. Let's take a look at a couple of other options.

Aromas	Missing Values	Flavors	Missing Values	Mouthfeel	Missing Values
Fruity	21	Fruity	32	Astringent	42
Spicy	43	Spicy	53	Bitter	44
Smoky	73	Smoky	78	Round	48
Ashy	107	Ashy	91	Lingering Ash	79
Tar	146	Tar	119	Metallic	132

2.1 Data Interpolation Empirical Orthogonal Functions (DINEOF)

There are a lot of different ways to perform approximation. In this study, we use the DINEOF (Data Interpolating Empirical Orthogonal Functions) method for optimizing Empirical Orthogonal Function (EOF) analysis on gaps, in this case, NA values in the dataset. This approach gradually solves for EOFs by means of an iterative algorithm to fit EOFS to a given number of non-missing value reference points (small percentage of observations) via RMSE (Root Mean Square Error) minimization. DINEOF is an EOF-based method to fill in missing data from geophysical fields, such as clouds in sea surface temperature. If you want to know more, Beckers and Rixen 2003 provide the basis and origin of this function. Additional papers and the original paper regarding this very interesting topic can be found by clicking the links below:

- The Journal of Atmospheric and Oceanic Technology
- Data interpolating Empirical Orthogonal Functions (DINEOF)
- Applications of DINEOF to Satellite-Derived Chlorophyll-a
- DINEOF Code: March Taylor

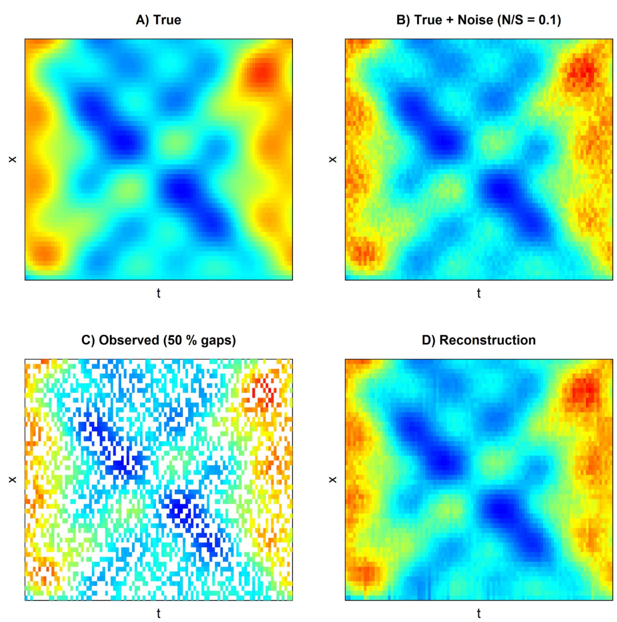


Figure 1: Dineof filling in gaps

2.2 Nearest Neighbors for Imputation

The k-nearest neighbors (kNN) algorithms have grown in popularity especially in clustering and unsupervised machine learning methods. In statistics, kNN is used for classification and regression. In both cases, the input consists of the k closest training examples in the data set. The output depends on whether kNN is used for classification or regression. One problem with kNN is that the algorithm is sensitive to the local structure of the data. In this study, we use kNN to impute the missing values and compare this method to DINEOF as well as ommision of NA values from the mainframe. Figure 2 illustrates the principle of kNN for imputation.

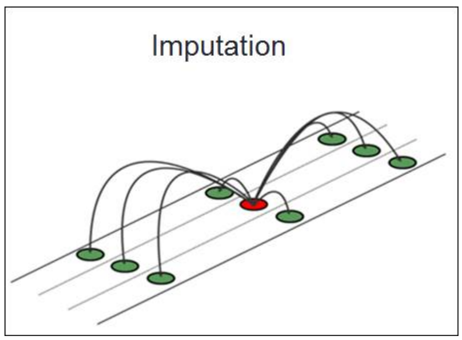
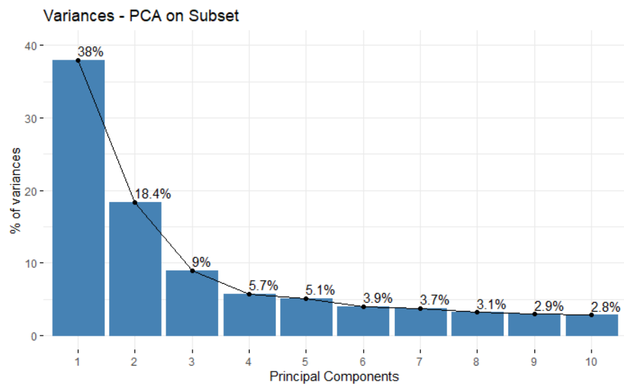


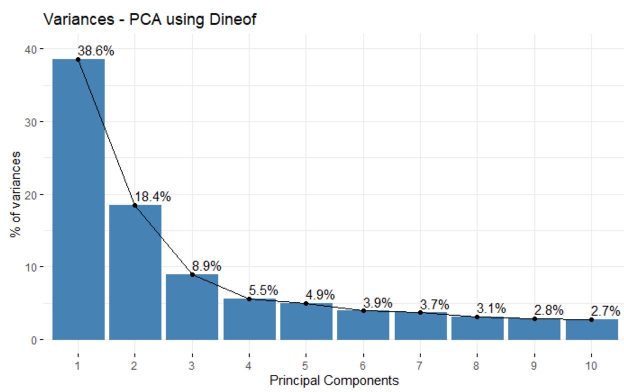
Figure 2: Nearest Neighbor Imputation

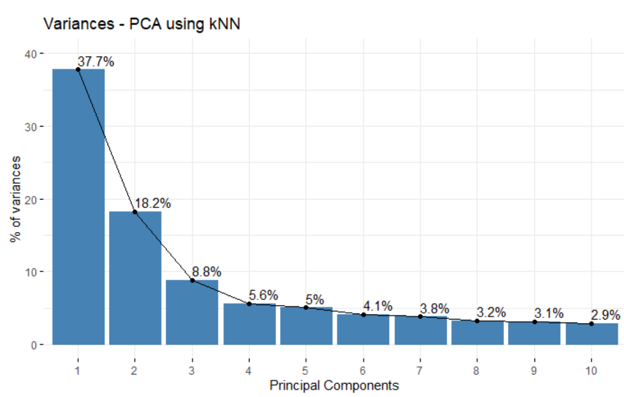
2.3 Conclusions on Data Cleaning

This investigation uses Principal Component Analysis (PCA) in multiple ways. The first way is by using it to validate our data cleaning. So, we compute the principal component variances and examine if our feature engineering and data cleaning caused the dataframe to change. Ideally, filling in the gaps will make little to no difference on the integrity of the dataset. Later, we'll use the centroid data from the PCA to perform significance testing and finally finish with PCA again, this time to provide a biplot to examine correlation.

When we look at principal components one and two we see there are few differences in the amount of variation among all three treatments. We use this information to validate our decision for using DINEOF to fill in all the


Variances - PCA on Subset
% of variances
40
30
20
10
0
38%
18.4%
9%
5.7%
5.1%
3.9%
3.7%
3.1%
2.9%
2.8%
1
2
3
4
5
6
7
8
9
10
Principal Components


Variances - PCA using Dineof
% of variances
40
30
20
10
0
38.6%
18.4%
8.9%
5.5%
4.9%
3.9%
3.7%
3.1%
2.8%
2.7%
1
2
3
4
5
6
7
8
9
10
Principal Components


Variances - PCA using kNN
% of variances
40
30
20
10
0
37.7%
18.2%
8.8%
5.6%
5%
4.1%
3.8%
3.2%
3.1%
2.9%
1
2
3
4
5
6
7
8
9
10
Principal Components

missing values in the dataframe and move forward with Hypothesis Testing, Analysis, and Data Visualization.

3 Hypothesis Testing

Hypothesis testing provides us with a level of significance to our study. It is difficult to speak in these terms as significance is often mistated. However, we perform a couple of Hypothesis tests in this study to test for Normality, and as mentioned before, use the centroid data from the PCA to test for significant differences in our treatments versus the control trials.

3.1 Test I : Shapiro-Wilk Test for Normality

Testing for normality with Shapiro-Wilk consists of our null and alternative hypothesis. The null hypothesis states that the data is non-normal. The alternative hypothesis is that our data is parametric. That is, a p-value > 0.05 impies the data is normal. We find for all attributes, our data is non-parametric.

Shapiro-Wilks	Attribute	F-Stat	Pr(>F)	Sig
1	Aroma.Ashy	0.943	0.0001	****
2	Aroma.Fruity	0.980	0.0001	****
3	Aroma.Smoky	0.969	0.0001	****
4	Aroma.Spicy	0.977	0.0001	****
5	Aroma.Tar	0.935	0.0001	****
6	Flavor.Ashy	0.970	0.0001	****
7	Flavor.Fruity	0.984	0.0001	****
8	Flavor.Smoky	0.976	0.0001	****
9	Flavor.Spicy	0.979	0.0001	****
10	Flavor.Tar	0.955	0.0001	****
11	Mouthfeel.Astringent	0.983	0.0001	****
12	Mouthfeel.Bitter	0.973	0.0001	****
13	Mouthfeel.Lingering.Ash	0.963	0.0001	****
14	Mouthfeel.Metallic	0.955	0.0001	****
15	Mouthfeel.Round	0.980	0.0001	****

However, upon closer examination, we see some attributes are near normal, others possess skewness, while others possess a degree of kurtosis. Without diving into Normality and the Central Limit Theorem, we simply mention large multivariate sets are difficult in this way. It's nearly impossible to adhere to the assumptions of the model when dealing with 30,000 data points. Some small groups might be normal but as a whole, we find the distribution is not Gaussian. We address this in the next section regarding the use of Permutation Multivariate Analysis of Variance (PERMANOVA).

3.2 Test II: PERMANOVA: Centroid Permutations

The logic of ANOVA is as follows: you compare the amount of variation between the groups with the amount of variation within the groups. ANOVA calculates the total sum of squares by taking the distance of each subject's response value from the mean response value, then squares these distances and adds them all up.

The logic of MANOVA is similar, but in two-dimensions: we calculate how far each subjects response vector is from the mean vector. The total sum of squares corresponds to the sum of the squared (euclidean) distances from each point to the mean point (called the centroid). Euclidean distance is just the distance formula from geometry class. It's the hypotenuse of a triangle created by the points from the pythagorean theorem. The centroid is best illustrated in Figure 3.

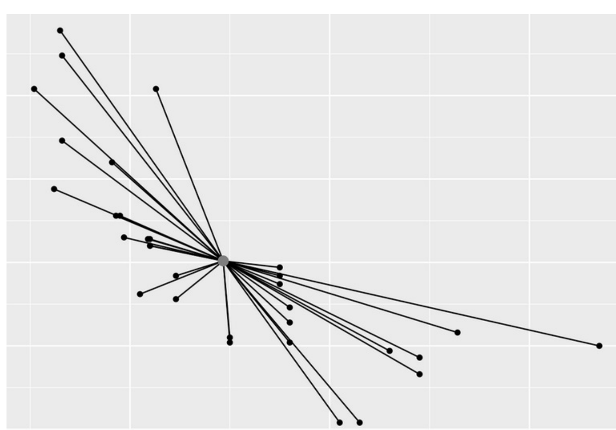


Figure 3: Centroid of Sample Cluster

There is an equality at play here: The sum of the squared distances from every point to the centroid is equal to the sum of the squared distances from each point to each other point, divided by the number of points. So we take this idea a step further, calculating the distance from the group centroid to the grand centroid, square those distances, multiply each distance by the number of subjects in that group, and then add those numbers together (Figure 4).

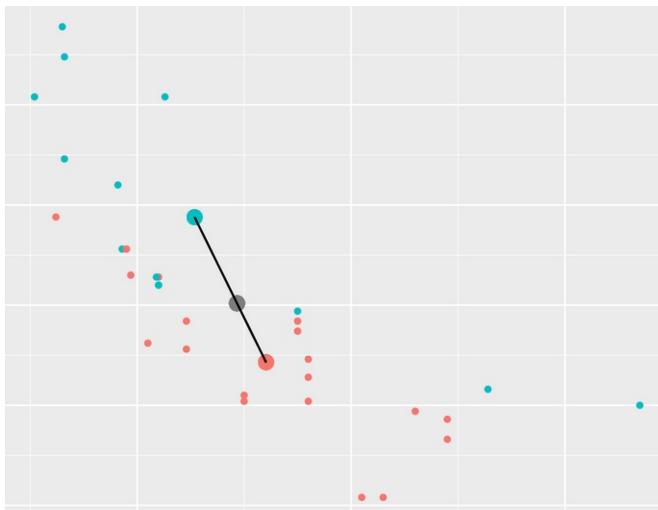


Figure 4: Centroid of the Centroid

Linear algebra concepts become really important here as we use matrix algebra to explain the properties of the identity using eigenvectors. We simply state the crucial part: we use the centroid as a distance measure and permutations to calculate the test statistics. In ANOVA, an F-statistic is a ratio of the variation between the groups to the variation within groups, and thus that ratio is close to 1 when those quantities are similar. This is also true with our pseudo test statistic, and is based on an identity matrix or equality.

PERMANOVA

Permutation is the act of rearranging objects. Understandably, if we were to change the data using permutations, we would disrupt the relationship between our variables. However, in hypothesis testing, we can use this effect of permutations (rearrangements) to create a null space that does not depend on a predefined distribution. So, we use this null space to define our null hypothesis. Obviously, the number of permutations performed determines the minimum probability of rejecting the null hypothesis. For example, using 1,000 permutations, the smallest possible p-value is 0.001.

PERMANOVA Assumptions

Every statistical test relies on assumptions about the distribution of the data. In the case of PERMANOVA, although there is no explicit assumption regarding the homogeneity of spread within each group, this test is sensitive to differences in variability among groups. Thus, if a significant difference between groups is detected, it might be due to differences in location, differences in spread, or a combination of the two. The best approach is to perform a seperate test of homogeneity, including pair-wise comparisons, as well as examining the average within and between-group distances and associated MDS plots. This helps determine if the nature of the difference between any pair of groups, is due to location, spread, or a combination of the two.

In this study, we use the function Beta-Disper to perform our seperate test for homogeneity. Futhermore, we run ANOVA based on the average distance to the median and perform a permutation test for homogeneity of multivariate dispersions. Finally, we used Pillai's trace as our test statistic. So without going too deep into the statistics, our null hypothesis will be that the groups are similar distances from the centroid and the alternative hypothesis will be that the groups are significantly different from each other via the centroid. We'll use $p < 0.05$ to reject the null. Let's see the results!

3.3 Activated Carbon

The Geosorb was added at the following rates during fermentation: 40g/hL, 50g/hL, 60g/hL, and 100g/hL. We saw significant differences between the control and the experimental treatments and suspect differences between dosages may be caused by varietal variation. In the section titled PCA, we look at the biplot and these relationships become more clear. Additional comments are found in the conclusions of this report.

3.4 Yeast Derivatives

This study investigates the addition of two types of Yeast Derivatives: yeast hulls, or cell walls, and inactivated dry yeasts (IDY). Yeast hulls are

Activated Carbon	DF	Sum of Squares	Mean Squares	F. Model	R2	Pr(>F)	Sig
Treatment	4	521.4	130.343	9.0214	0.04138	0.001	***
Residuals	836	12078.6	14.448		0.95862		
Total	840	12600.0			1.00000		

Table 2: Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

Pairwise Comparisons	Control	Geosorb 40g/hL	Geosorb 50g/hL	Geosorb 60g/hL	Geosorb 100g/hL
Control	-	-	-	-	-
Geosorb 40g/hL	0.0025	-	-	-	-
Geosorb 50g/hL	0.0025	0.0117	-	-	-
Geosorb 60g/hL	0.0025	0.5280	0.0040	-	-
Geosorb 100g/hL	0.0025	0.1500	0.5280	0.2425	-

Table 3: Pairwise comparision using PERMANOVA

known to assist stuck fermentations by adsorbing toxins like decanoic acid (Munoz and Ingledew, 1989). Inactivated yeasts possess similar properties to yeast hulls and other qualities, such as the release of polysaccharides. We examined these treatments based on the adsorptive and absorptive properties of yeasts and the well-known positive impacts of yeast derivatives on the sensory character of the wine.

Yeast Derivatives	DF	Sum of Squares	Mean Squares	F. Model	R2	Pr(>F)	Sig
Treatment	2	132.6	66.300	4.5907	0.04753	0.002	***
Residuals	184	2657.4	14.442		0.95247		
Total	186	2790.0			1.00000		

Table 4: Signif codes: '***' 0.001, '**' 0.01, '*' 0.05

We see some significance difference from the control with the yeast derivatives. Here the inactivated yeasts represent a greater distance from the control and both yeast derivatives represent a significant difference from the control cluster at $p < 0.05$. We'll plot these PCA biplots in the next section.

Pairwise Comparisons	Control	Yeast Hulls	Inactivated Dry Yeasts
Control	-	-	-
Yeast Hulls	0.028	-	-
Inactivated Dry Yeasts	0.012	0.900	-

Table 5: Table of test stats using PERMANOVA

3.5 Fining Agents

The third part of our smoke taint study investigates the addition of various fining agents during the pre-fermentation phase of wines made from smoke-exposed grapes in Napa, California. Targeted fining agents are known to reduce polyphenolic content, and studies (Fudge et al., 2012) have shown that only activated carbon effectively reduces smoke taint analytes and improves sensory profiles. In Cabernet Franc (CF), we examined LAFFORT products: GEOSORB, an activated carbon, POLYLACT, a blend of PVPP and Casein, and CASEI PLUS, a 100% casein product. In Cabernet Sauvignon (CS), we compare GEOSORB and Skim Milk.

Fining Agents Cab Sauv	DF	Sum of Squares	Mean Squares	F. Model	R2	Pr(>F)	Sig
Treatment	2	261.7	130.866	9.1584	0.0532	0.001	***
Residuals	326	4658.3	14.289		0.9468		
Total	328	4920.0			1.00000		

Table 6: Significance in Cabernet Sauvignon Treatments

The PERMANOVA pairwise comparisons illustrates both treatments are different from the control wine and that they reside in similar clusters together. This is provided in a graphical format using the PCA biplot in the next section.

Pairwise Comparisons	Control	Geosorb 100g/hL	Skim Milk 10g/hL
Control	-	-	-
Geosorb 100g/hL	0.0015	-	-
Skim Milk 10g/hL	0.0015	0.6170	-

Table 7: Pairwise Comparison on Cabernet Sauvignon

In the Cabernet Franc, we further investigate the effects of fining treatments on the sensory qualities of the wine. There appears to be significance in the treatments again, so we perform pairwise comparisons.

Fining Agents Cab Franc	DF	Sum of Squares	Mean Squares	F. Model	R2	Pr(>F)	Sig
Treatment	3	336.4	112.118	7.8415	0.05364	0.001	***
Residuals	415	5933.6	14.298		0.94636		
Total	418	6270.0			1.00000		

Table 8: Significance testing for Cabernet Franc Treatments

Pairwise comparisons illustrates significant differences between the control and the treatment data while illustrating that each of the fining agents represent a similar distance from the control. Thus, adding a fining agent appears to change the sensory quality of the wine respective to the control cluster. This is counter to current literature published by AWRI and may be due to statistical methods and sensory study design. Additional comments are found in the conclusions of this report.

Pairwise Comparisons	Control	Geosorb 100g/hL	Casei Plus 100g/hL	Polylact 100g/hL
Control	-	-	-	
Geosorb 100g/hL	0.002	-	-	
Casei Plus 100g/hL	0.002	0.882	-	
Polylact 100g/hL	0.002	0.882	0.882	-

Table 9: Pairwise Comparison of Cabernet Franc

3.6 Oak Treatments

The fourth, and final investigation was with NOBILE products. In this portion of the trial, we examine the impact of oak additives, added during fermentation, on wines made from smoke-exposed grapes. This trial was performed on a Cabernet Sauvignon and Cabernet Franc from Napa. We used American Fresh, Fresh Granulars, Original Spice, Base, and Sweet Vanilla products. We also trialed Prototype 7 during this time. We added each oak product at 5 and 10g/L but only include the larger of the two doses in the hypothesis testing and data visualizations.

Oak Cabernet Sauvignon	DF	Sum of Squares	Mean Squares	F. Model	R2	Pr(>F)	Sig
Treatment	4	73.49	18.372	1.242	0.071	0.289	
Residuals	65	961.51	14.793		0.929		
Total	69	1035.00	-		1.00		

Table 10: Significance testing for Cabernet Sauvignon Treatments

Starting with the Cabernet Sauvignon, we don't see much in terms of significance among the treatments but we perform pairwise comparisons anyway. The pairwise comparisons illustrates a near significant relationship with the characters related to the Sweet Vanilla treatment. This information is presented as a PCA biplot in the following section.

Oak Cabernet Sauvignon	Control	Base	Fresh Granular	Original Spice	Sweet Vanilla
Control	-	-	-		
Base	0.79	-	-		
Fresh Granular	0.79	0.92	-		
Original Spice	0.79	0.92	0.79	-	
Sweet Vanilla	0.07	0.88	0.20	0.88	-

Table 11: Pairwise Comparison of Cabernet Sauvignon with NOBILE Oak

We report similar results for the oak treatments on the Cabernet Franc. Adding NOBILE Oak to grapes with heavy smoke-exposure did not have a significant effect on the sensory quality of the wine. We examine the PCA biplot in the next section of this report.

Oak Cabernet Franc	DF	Sum of Squares	Mean Squares	F. Model	R2	Pr(>F)
Treatment	3	51.38	17.128	1.1478	0.03983	0.284
Residuals	0.44	1238.62	14.923		0.96017	
Total	0.78	1290.00			1.00	

Table 12: Significance testing for Cabernet Franc with NOBILE Oak

Oak Cabernet Franc	Control	American Fresh	Fresh Granular	Prototype 7
Control	-	-	-	-
American Fresh	0.44	-	-	-
Fresh Granular	0.78	0.44	-	-
Prototype 7	0.44	0.44	.44	-

Table 13: Pairwise Comparison of Cabernet Franc with NOBILE Oak

4 Principal Component Analysis

This section is dedicated to our Principal Component Analysis (PCA) and the resulting biplots. We've already used PCA for exploratory data analysis in hypothesis testing, and in validation for feature engineering or data cleaning. Our PCA uses Singular Value Decomposition (SVD) to perform the decomposition. SVD is a linear algebra cosntruct which transforms the data by rotating it, flattening it, and then rotating it again. While this sounds complicated, the most important part to understand about the PCA/SVD is how to interpret the results. If two vectors are positively correlated, there is an acute angle between them. If two vectors are negatively correlated, there is a 180 degree angle between them. Looking at the data in this way allows us to explore and project the data into a space which illustrates the trajectory of the sensory character of the wine as products are added.

4.1 Activated Carbon

We begin the investigation with the activated carbon. The ellipses represent a confidence interval of 0.75 meaning 75% of the data in the group resides within the ellipse. The biplot clearly represents positive and negative component grouping. The addition of Geosorb shifts the wine away from those negative attibutes. We speak to the degree of the shift in the previous section regarding significance and the difference between making the wine less bad versus making the wine better using radar charts found in the next section.

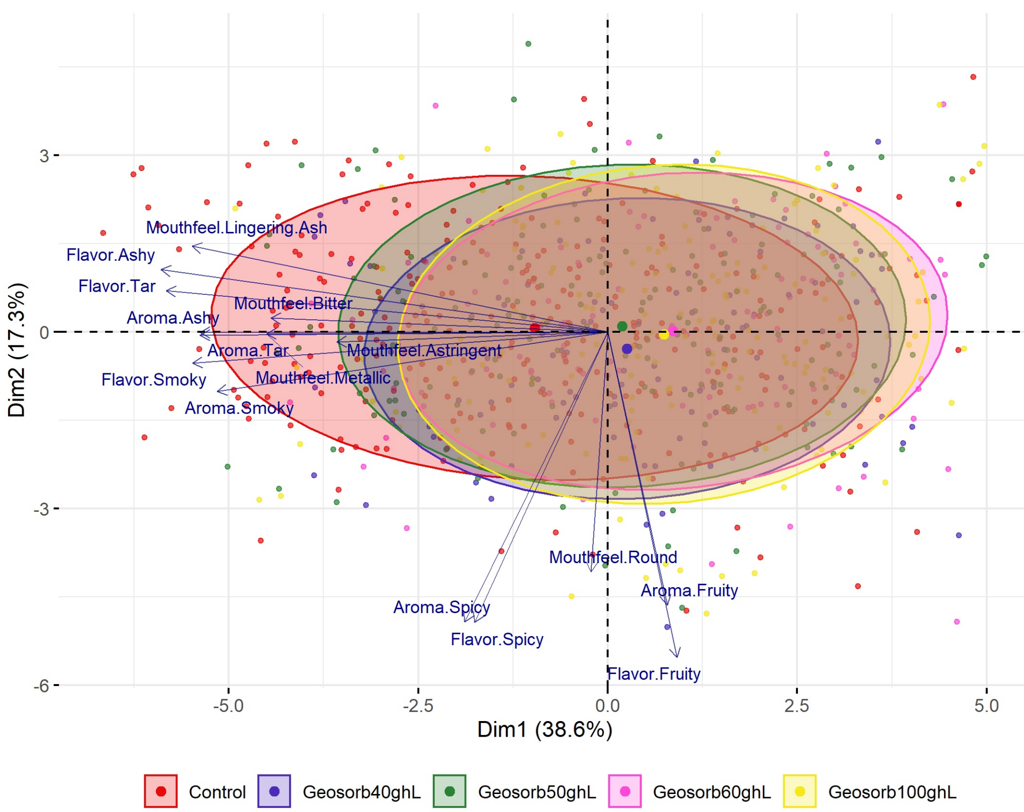


Figure 5: PCA of Geosorb Treatments

4.2 Yeast Derivatives

We continued our investigation with the Yeast Derivatives. Just as in the case with the GEOSORB, we use a 75% ellipse. While we have fewer samples of the yeast derivatives, we still see a shift away from the bad character found in the control and towards the positive characters. In this case, the inactivated yeasts appear to perform slightly better. This claim is substantiated in the hypothesis testing and is also illustrated in the following section titled: Radar Charts.

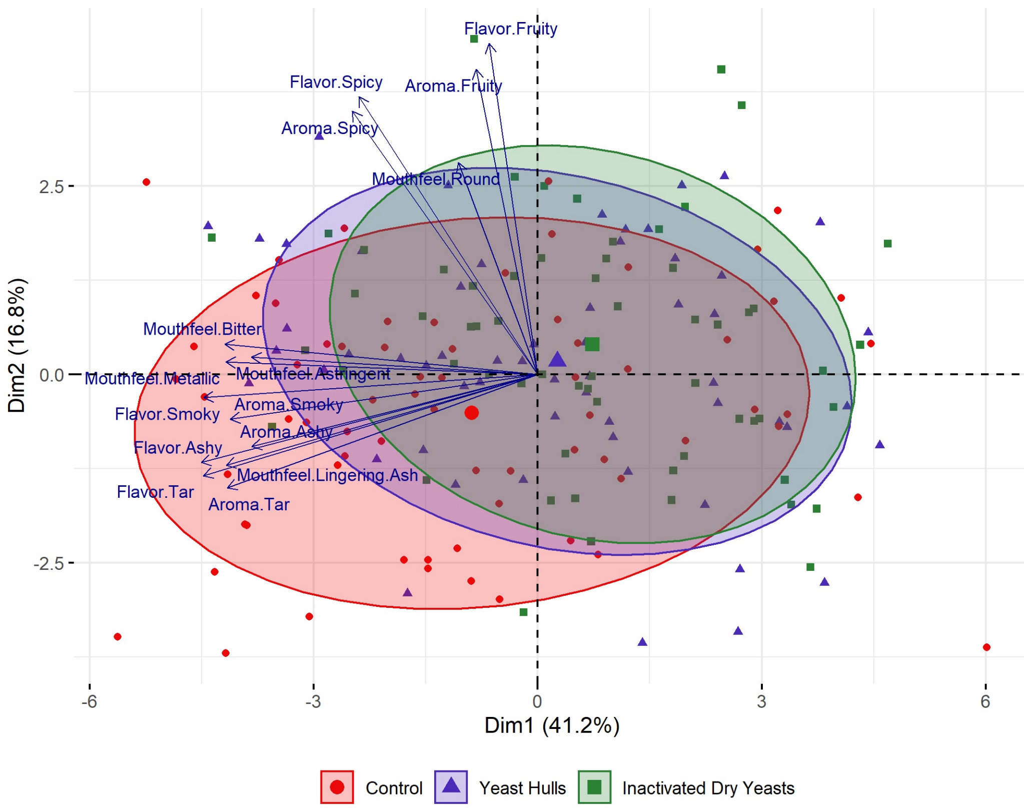


Figure 6: PCA of Yeast Derivative Treatments

4.3 Fining Agents

Next, we examined the fining agents. In this portion of the trial process we had a Cabernet Sauvignon and a Cabernet Franc. We present these findings seperately as they were different varietals. In the Cabernet Franc we examined Geosorb, Casei Plus, and Polylact. This PCA represent our dosages of 100g/hL for each of the products.

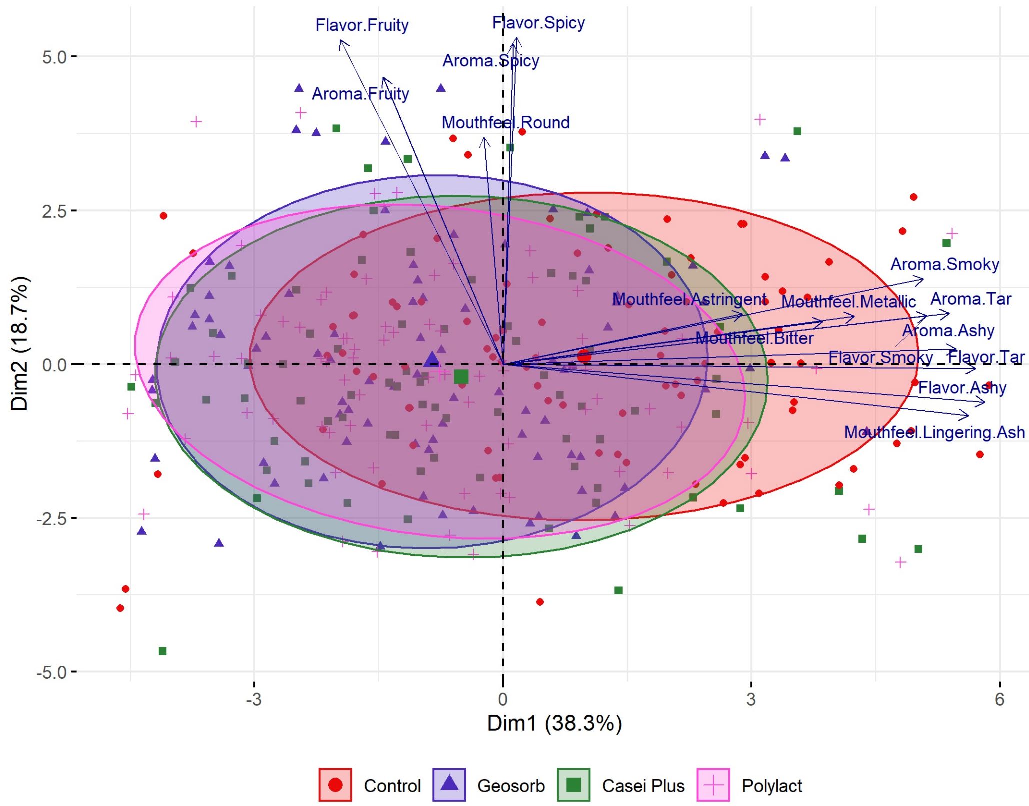


Figure 7: PCA of Fining Agents in Cabernet Franc

The Cabernet Sauvignon is a comparison of a 100g/hL addition of Geosorb versus the addition of Skim Milk at 10m/L. We note the sparsity of the results in the Skim Milk cluster as the Geosorb cluster is tighter and represents a targeted product addition with a greater degree of precision. Additional information is provided in the Conclusions.

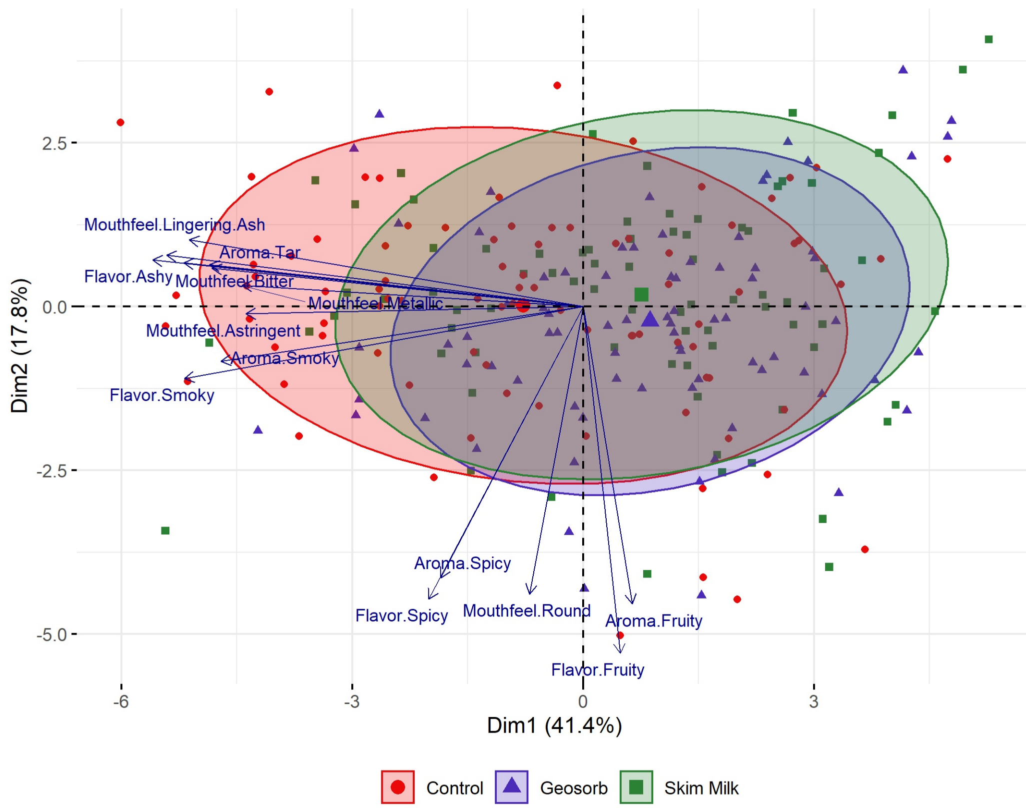


Figure 8: PCA of the Fining Agents in Cabernet Saugvignon

4.4 Oak Treatments

This Oak Treatments represent an array of Nobile treatments at 10g/L. We performed both 5g/L and 10g/L but only include the 10g/L as the data presented was not found to be significant and our desire keep our plots from being overcrowded with information. It is worth noting that Sweet Vanilla was determined to be near significant.

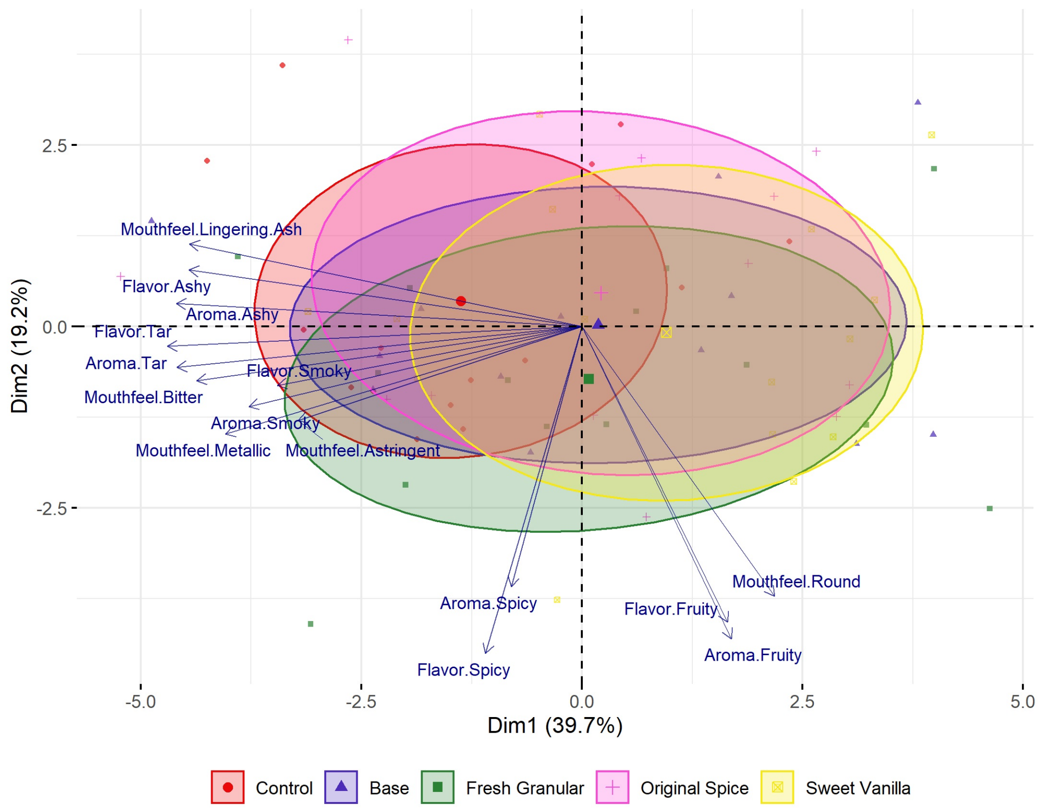


Figure 9: PCA of the NOBILE Oak in Cabernet Sauvignon

Furthermore, these plots illustrate that the impact of adding oak to grapes exposed to oak did not appear to make the smoke character worse. Additional information is included in the conclusions of this report.

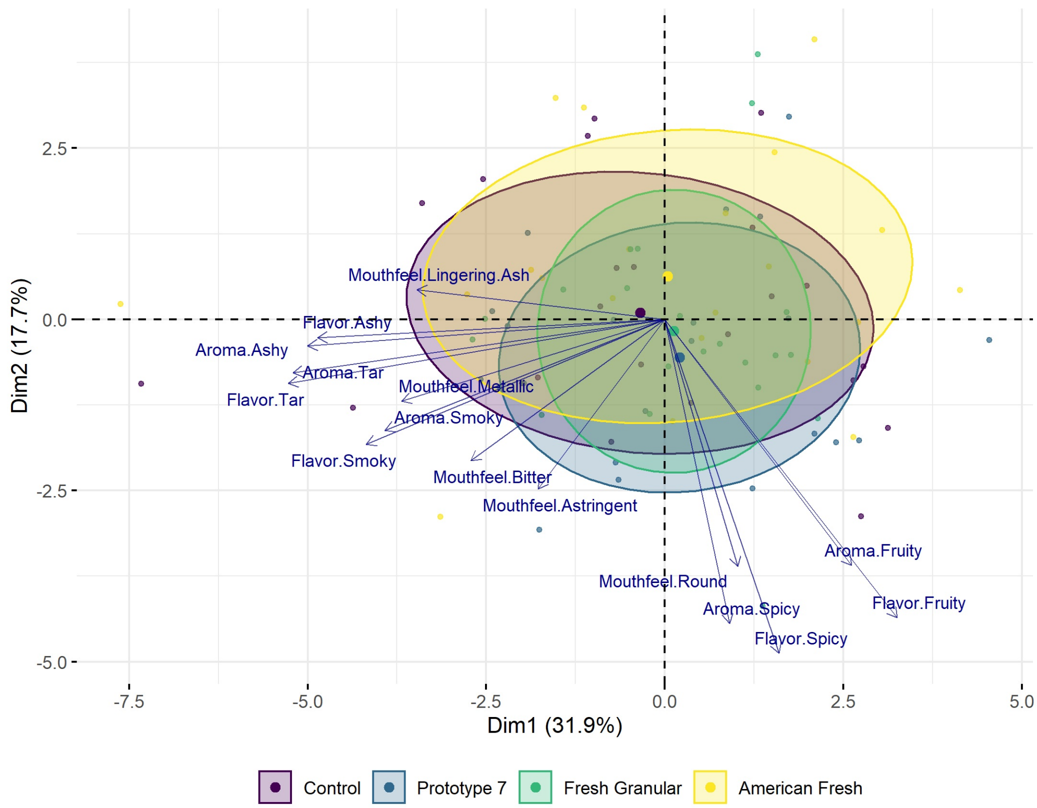


Figure 10: PCA of the NOBILE Oak in Cabernet Franc

5 Radar Charts

Radar charts are commonplace in the wine sensory and often illustrate the impact of products on the final sensory character of the wine. We construct these plots to provide an additional graphical method of displaying our multivariate data. Each of these charts use min-max scaling for each feature. This way, they represent percent reduction or increase as they are scaled between the values of zero and one.

5.1 Activated Carbon

The Geosorb Radar Chart illustrates the range of dosages added to four different varietals including Cabernet Sauvignon, Cabernet Franc, Merlot, and Pinot Noir.

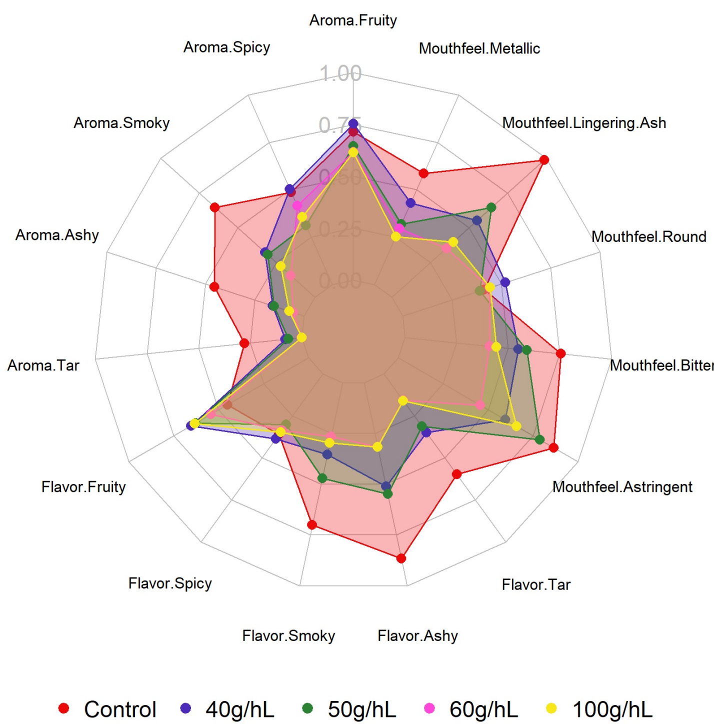


Figure 11: Geosorb Radar Plot

The addition of Geosorb represents greater than 75% reduction of Lingering Ash as a Mouthfeel component. Suprisingly, it also represents increases in fruitiness.

5.2 Yeast Derivatives

The Yeast Derivatives Radar Chart represents the addition of two different yeast derivatives at 50g/hL on a Pinot Noir made from smoke-exposed fruit. In this portion of the trial, we see increases in Fruitiness and Round Mouthfeel, and decreases in the intensity of the smoke character in the Control.

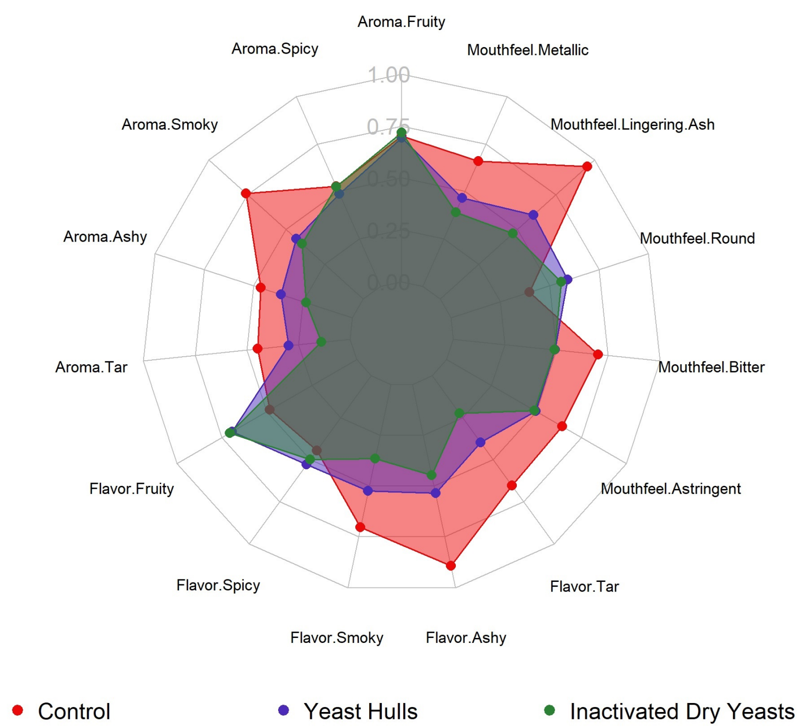


Figure 12: Yeast Radar Plot

5.3 Fining Agents

In the Cabernet Franc fining trial we examined Geosorb, Polylact, and Casei Plus. We find each product makes the wine less bad as there is a clear retraction from smoke related sensory character. This Radar Chart is scaled using Min-Max, meaning for the Mouthfeel sensation of Lingering Ash, we saw just over a 50% reduction of the intensity of this attribute by using Geosorb.

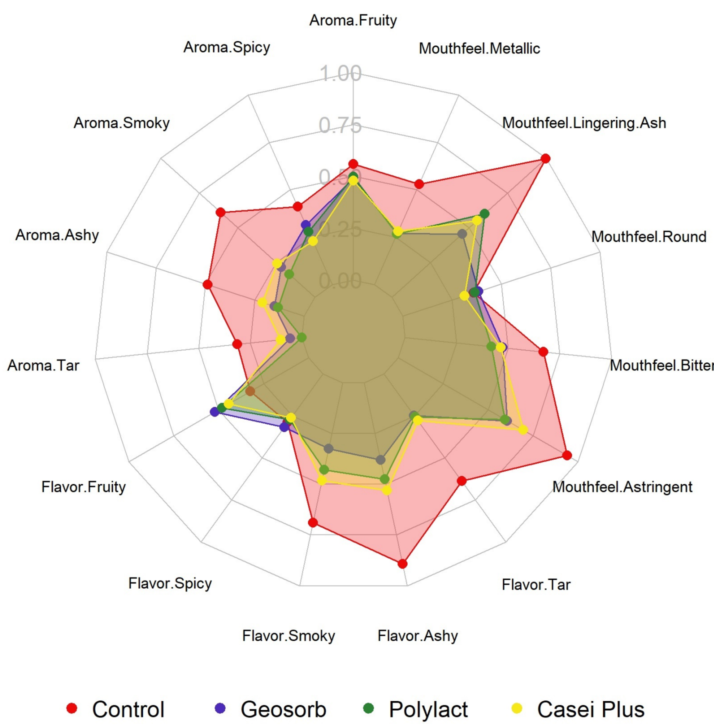


Figure 13: Fining Cab Franc Radar Plot

When we look at the Cabernet Sauvignon, we see Geosorb actually increases the character related to fruity flavors and aromas while Skim Milk is more retractive and therefore a less targeted fining agent. Min-Max scaling has occured in this plot as well, meaning there was a 25% increase in Fruity flavors by adding the Geosorb compared to the Control in Figure 11.

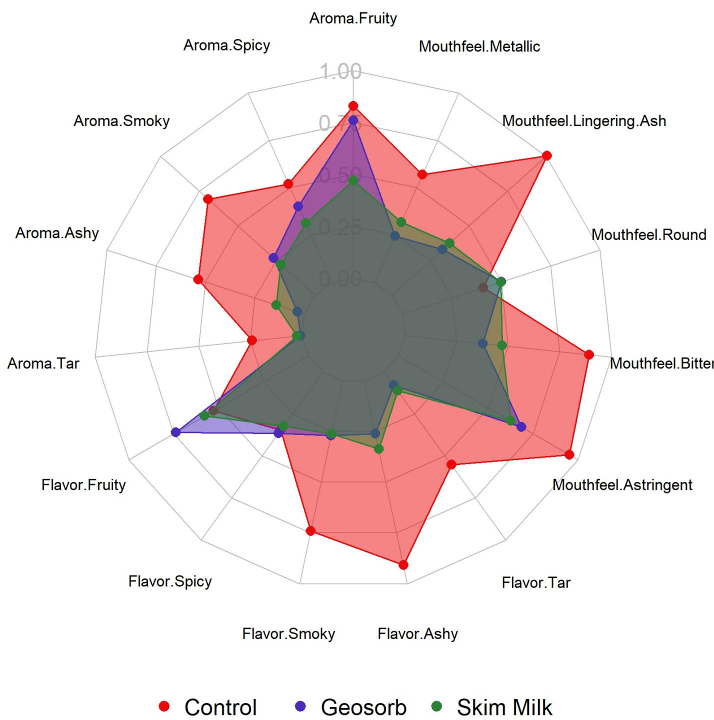


Figure 14: Fining Cab Sauv Radar Plot

5.4 Oak Treatments

When we look at the Oak treatment Radar Plots, we see the addition of oak has marginal effects on some of the sensory characters. These plots are from Oak dosages at 10g/L. These results coincide with our statistics in the previous section and provide insight into the risk of elevated smoke character through the addition of Oak. In our study, we saw no significant increases in smoke related character by adding Oak products.

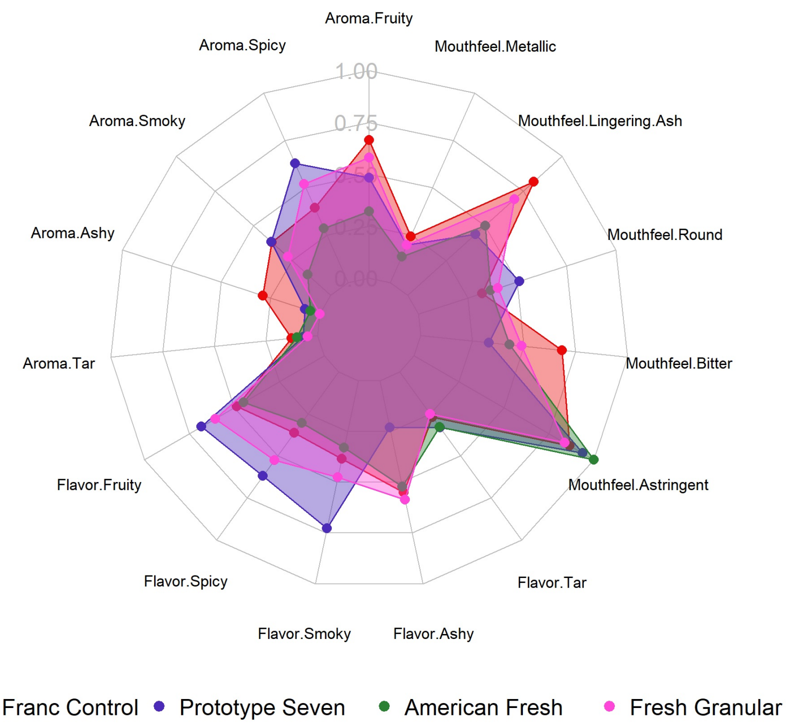


Figure 15: Nobile Cab Franc Radar Plot

In the second portion of the Oak trial we see similar results found in the statistics from the previous section. Here, we see near significance with the Sweet Vanilla addition as the wine possess more roundess. See the aforementioned section on significance to see statistics of these graphs. This portion of the experiment was on adding Oak to the fermentation of smoke-exposed Cabernet Saugvignon.

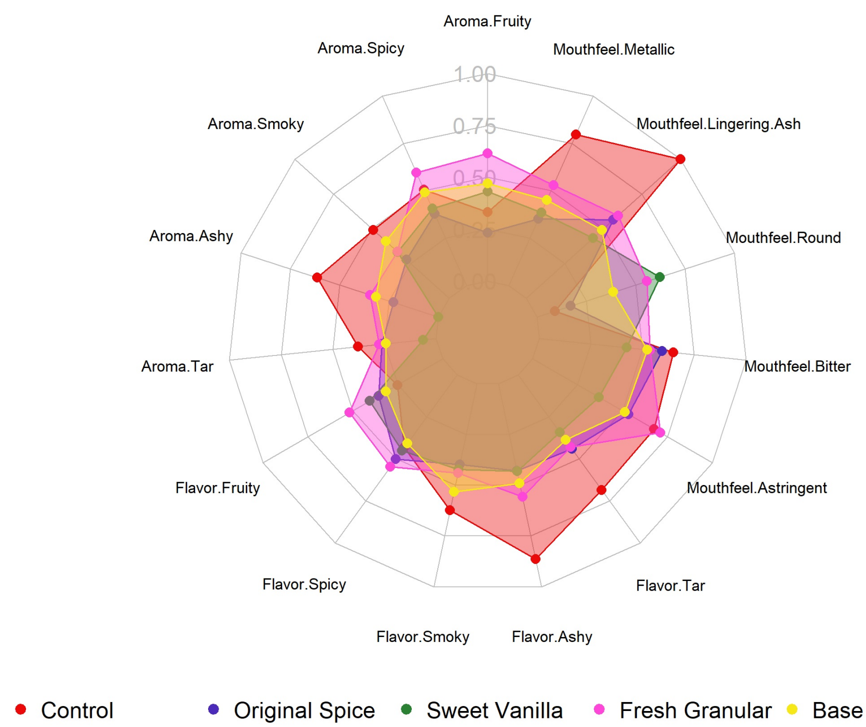


Figure 16: Nobile Cab Sauv Radar Plot

6 Bar Charts

This section is dedicated to the bar charts we created to illustrate the Quantitative Analysis of compounds related to smoke with Excell Laboratories in Bordeaux. We present average reduction of total/bound Guaiacol, 4-Methyl Guaiacol, and Total Phenols in Figure 17. This graph didn't include instances where we saw increases as we only illustrate percent reduction.

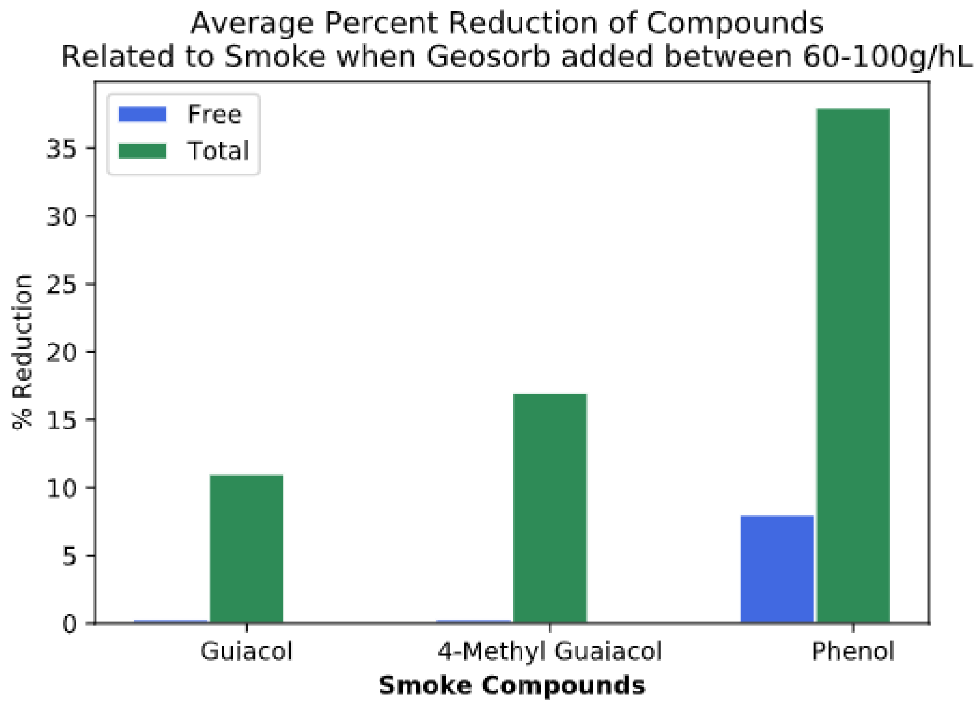


Figure 17: Geosorb reduces bound compounds related to smoke

Furthermore, we examine Geosorb compared to Skim Milk. Not only do we see greater reduction of smoke related compounds, we also find Geosorb is a more targeted addition as it does not fine away positive characters found in the aforementioned Radar Charts.

Our final barchart is for the fining agents we used on the smoke-tainted Cabernet Franc. While these results need additional replication, and may have been impacted by the winemaking methods, we include these results in this report as they indicate fining agents may be playing a role in the reduction of smoke related compounds contrary to reports from AWRI. Additional

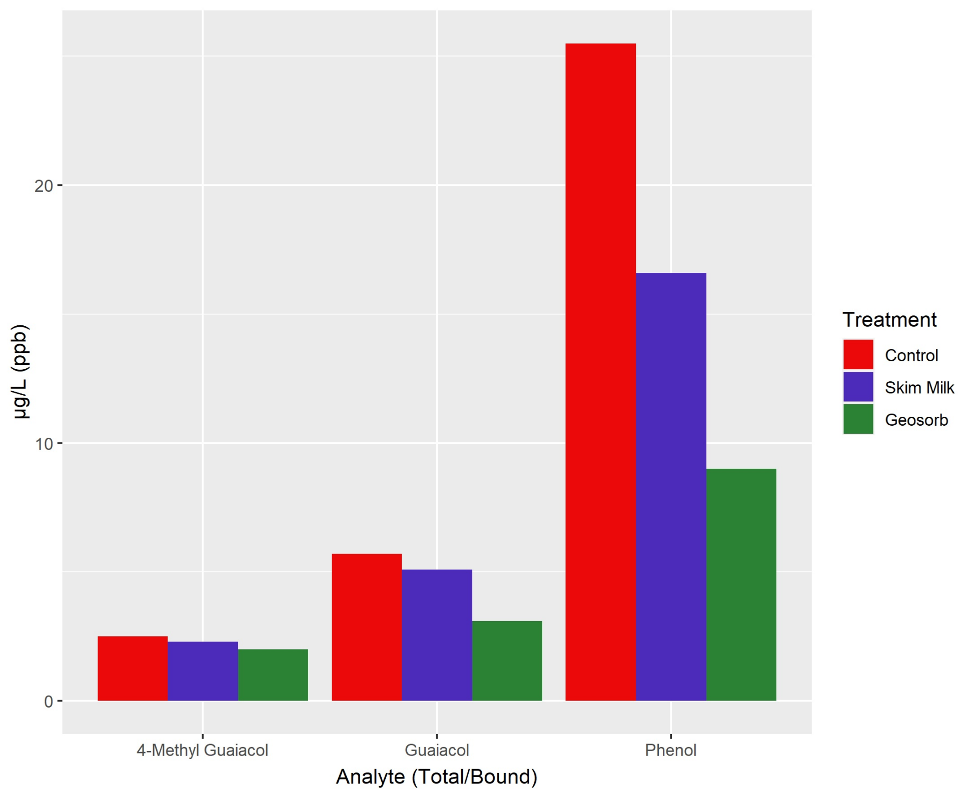


Figure 18: Geosorb compared to Skim Milk in Cabernet Sauvignon

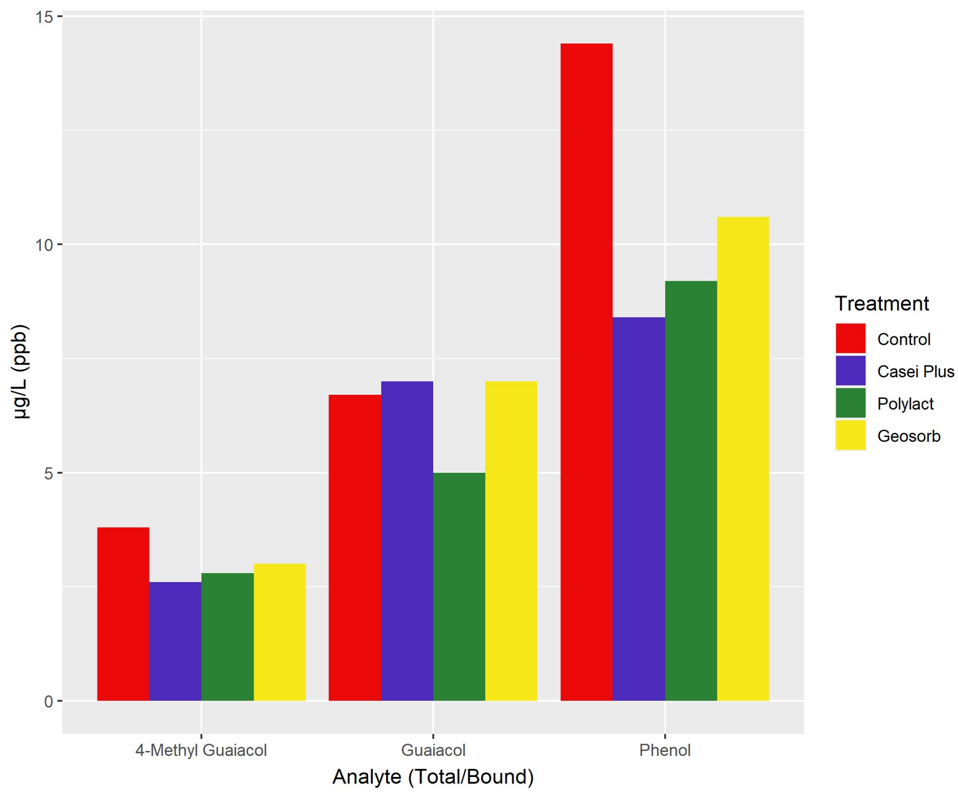


Figure 19: Fining agent efficacy on smoke related compounds

This is contrary to reported findings from Australia and AWRI. Additional information and raw dataset is available upon request. We will want to be as transparent as possible here with our methods and data.

information is included in the conclusions of this report.

7 Conclusions

Our conclusions are broken down into each section where we combine the results from each section providing insight into the product usage. These sections provide our conclusions and recommendations for product usage in future vintages when dealing with wines made from smoke-exposed fruit. This study made over thirty different wines, from four different American Viticulture Areas, using multiple products. Then, in February we conducted the largest known sensory study on smoke-exposed grapes to date where nearly 300 winemakers gathered in a virtual setting to examine the impact of Laffort products added during the fermentation of smoke exposed fruit. This dataset represents nearly 30,000 data points and was acquired using a data pipeline constructed in Python. The analysis was written entirely in R and is available to BioLaffort upon request and backed up on servers at Laffort USA.

7.1 Activated Carbon

When dealing with a smokey vintage we recommend adding 100g/hL to the fermentation and racking off the geosorb after seperating the must from the wine. We find Geosorb removes compounds related to smoke (Figure 17). It performs better than Skim Milk (Figure 18). Geosorb also is more specific and has less of an impact on the sensory quality of the wine (Figure 14). When considering the significance of our findings, we find all of our dosages represent significant changes in the sensory character of the wine (Table 2 and 3). Geosorb works by decreasing the compounds related to smoke and by increasing the fruity character of the wine through the adsorption of the harsh phenolics.

7.2 Yeast Derivatives

When dealing with a possible smokey vintage, we recommend adding yeast derivatives around 50g/hL to the fermentation. Even if the smoke doesn't present itself, by adding the derivatives, we increase the roundness of the wine and thereby limit the harshness of any phenolic bitterness (Figure

6 and Figure 12). While we don't see the yeast derivatives remove any compounds related to smoke, we do find a benefit for their addition (Figure 12).

7.3 Fining Agents

When considering the addition of fining agents during a smokey vintage, we find contrary to AWRI, that adding fining agents provides an impact. We validate this using PERMANOVA and suggest additional studies (Tables 6, 7, 8, 9). Our methods for analysis may be better suited for multivariate analysis and large data sets where simple ANOVA relies on several key assumptions regarding data distribution (Figures 7, 8, 13 and 14).

7.4 Oak Treatments

Finally, when considering whether or not to add Oak to a vintage impacted by smoke. We find adding oak did not significantly change the character of the wine from the controlled smoke wine. While we see near significance with the Sweet Vanilla addition, overall, we don't see significant differences. Adding oak to a smokey wine has been a concern for winemakers in Napa in 2020-2021. We find oak has an impact on the wine as indicated in Figure 15 and 16 but we do not report significance in these differences.