

VeriVin experiments

VeriVin and Openvino are teaming up to perform a series of experiments using VeriVin's prototype Raman spectrometer.

These experiments will begin on the 18th of October, 2019, using new installations at Costaflores Organic Vineyard.

Here is a summary of the first three, simultaneous experiments that will be executed:

- Do coloured wine bottles protect wine from oxidation, and what is the measurable effect of lightstrike from different types of lighting?
- Which bottle closures better promote desirable wine evolution?
- Can we create a unique digital fingerprint for a wine, using a spectrometer, and represent this vinoprint on the blockchain as a non-fungible token?

About VeriVin

Experiment Overview

Can we bottle 640 bottles of wine, using four different bottle colours, four different closures, and four different light sources, and perform simultaneous experiments?

1. Do coloured wine bottles protect the wine from oxidation?
 - a. Transparent v Green v Eco v Brown
 - i. Initial Veralia samples (Green, ECO, Brown, transparent) available now
 - ii. Blue bottles not available in Mendoza
 - iii. Working on Veralia QA, PR people for samples and support
 - iv. Meeting, visit to factory, with film person
 - b. Light variables: Natural light (what is this?) v LED v other artificial light v dark, in the box (control)?
 - i. Define what is needed for this - design light chamber
 - ii. Diurnal cycle? (simulating wine shop?)
 - c. This an on-going study...publishing data as we go.
 - d. testing frequency (weekly?)
 - e. sensing/validating other factors - temperature, light, air quality
 - f. sample size

From what I have read in literature, it seems that 'light strike' is mainly due to UV and Blue light. However, considering the bottles absorption spectra and the high complexity of the liquid, I wouldn't be surprised if even green and red played a role. Obviously, intensity of radiation also matters. Also, what part of this light goes through the bottle also matters (so, the absorption spectrum of the glass). I think that as long as the temperature of the various enclosures is consistent, there is no need to have a special dark enclosure, a well sealed box is fine

Given that short wavelength light is incriminated, I would go for bulbs with high blue content, so, COLD WHITE LED and generally cold light. Unfortunately, fluorescence lamps that emit this light are halogen, and they also heat heavily the environment. So they are not advisable. I would rather compromise for something like a tube fluorescent light, as white as possible.

Have a look at these links to check if they may work:

Fluorescence:

<https://www.lightbulbs.com/product/bulbrite-524053>

LED array

https://www.amazon.it/Componente-Circuito-Stampato-Emettitore-Circolare/dp/B07DGVTRH9/ref=sr_1_10?__mk_it_IT=ÅMÅŽŃÑ&keywords=LED%2Barray&qid=1567529917&s=gateway&sr=8-10&th=1

LED panel:

https://www.amazon.it/Pannello-LED-120x60-6500K-incluso/dp/B07DT33SY8/ref=sr_1_7?__mk_it_IT=ÅMÅŽŃÑ&keywords=col+d+white+LED&qid=1567529984&s=gateway&sr=8-7

I would try to keep illumination more or less constant to each bottle, at about 800/1200 lumen, to simulate the illumination of a supermarket, and/or to about 300/500 lumen to simulate a home environment.

This wikiHow is quite well made in my opinion:

<https://www.wikihow.com/Measure-Light-Intensity>

Also, we should illuminate all bottles in the same way, and be sure light distribution is uniform, I imagine the best solution is to stack the bottles in layers and illuminate them from the side, using multiple bulbs at a regular distance, is that something possible?

I am thinking that if we first bottle the wines, in the four different bottles, and then scan them with the spectrometer, should we then keep ALL of them in boxes, in darkness for a few months before our next measurement, to rule out the influence of light for any deviation occurring during that period? or do you think it is enough to have the darkness control collection?

*I think the control group kept in darkness will be enough. We have had meaningful results with 18 bottles, so 32 are a good number already. I am thinking, though, that wine undergoes quite a lot of changes in the first weeks after bottling, and we may risk to affect the experiment if we irradiate the bottles in a way that is heavily different from what the bottle sees in the cellar. So, I guess you have a point there.
It is probably better to put all the bottles to a "zero point" before starting the proper long term experiment, also this could give us some flexibility in case the experiment is delayed for any reason.*

1. Which closures better promote desirable wine evolution?
 - a. natural cork, portocork, synthetic, or screwcap
 - i. http://realdelacruz.com.ar/productos_portocork_natural.php
 - ii. <https://www.tapigroup.com>
 - b. Details about bottling
 - c. Capsule?
 - d. We need to bottle them when the spectrometer is available to validate the initial state of each bottle.
2. Can we create a digital fingerprint for a wine?
 - a. how does this fingerprint evolve over time?
 - i. post-Fermentation (2020)
 - ii. 1-year in stainless steel
 - iii. 1-year in new oak
 - iv. multi-year in bottles, stored at controlled temperature.
 - b. 10000x sample size - cold storage, same light, blockchain-registered temperature
 - c. how can this fingerprint be tracked on the blockchain?
 - d. start experiment with 2018 and 2019, but start in earnest with 2020 vintage
 - e. other wines (10's) available at winery and cold-storage

Shared considerations

1. Location
 - a. new room at winery
 - b. costaflores
2. Staffing for long term testing
 - a. Yamil
 - b. Carla
3. Environmental constraints
 - a. Temperature
 - b. Light
 - c. Air quality
 - d. wine movement
 - e. machine safety
4. Costs
 - a. bottles
 - b. closures
 - c. bottling
 - d. Test staff
 - e. critic testing fees
 - f. shipping
 - g. labeling / packing for critics
 - h. spoofing
5. Funding
6. Logistics
7. Cross evaluation
 - a. organoleptic
 - i. schedule
 - ii. tasters
 - b. chemical
 - i. lab1
 - ii. lab2
 - iii. lab3
 - c. chromatographic
 - i. lab1
 - ii. lab2
 - iii. lab3
8. Research
 - a. similar experiments and results
 - b. positions of critics on numerical rankings
 - c. list of caveats
9. Documentation
 - a. Video presentation (i.e. documentary)
 - b. Public wiki
 - c. Costaflores / VeriVin /
10. Reporting

Experiment / Analysis Walkthrough

In this first experiment all measurements have been taken through the same transparent bottle.

Definition of a measurement: Data gathered with VeriVin's Raman probe

The following data was taken;

- on the 30th of April preliminary measurements of MTB2014-1 through a transparent container
- on the 10th and 13th of May, two sets of measurements of wines MTB2012-1, MTB2014-1, MTB2014-2, MTB2016-1 and MTB2017-1

The first data set is used to create a chemometrics model (PLSDA – Partial least squares discriminant analysis) and the second data set is tested on said model. A data set consists of multiple measurements taken at different times, with different spectral baselines and probing different spots on the test-bottle. We were able to successfully determine which wine was scanned when applying this data to our model – classification successful. -1 and -2 signifies two different bottles, which in the case of MTB2014-1 and -2, had a sensory (taste-able) difference.

Since the final output is a fairly mundane table of classification as shown here, we detail some of the actual analytical components used within the model.

PLSDA Classification Using Rule: Pred Most Probable

MODEL RESULTS

Confusion Matrix:

Class:	TPR	FPR	TNR	FNR	N	Err	P	F1
MTB2012-1	1.00000	0.00000	1.00000	0.00000	25	0.00000	1.00000	1.00000
MTB2014-2,MTB2014-1	1.00000	0.00000	1.00000	0.00000	23	0.00000	1.00000	1.00000
MTB2017-1	1.00000	0.00000	1.00000	0.00000	24	0.00000	1.00000	1.00000
MTB2016-1	1.00000	0.00000	1.00000	0.00000	12	0.00000	1.00000	1.00000

Confusion Table:

	Actual Class			
	MTB2012-	MTB20*2	MTB2017-	MTB2016-
Predicted as MTB2012-1	25	0	0	0
Predicted as MTB2014-2,MTB2014-1	0	23	0	0
Predicted as MTB2017-1	0	0	24	0
Predicted as MTB2016-1	0	0	0	12
Predicted as Unassigned	0	0	0	0

Key:

MTB20*2 = MTB2014-2,MTB2014-1

CV RESULTS

Confusion Matrix (CV):

Class:	TPR	FPR	TNR	FNR	N	Err	P	F1
MTB2012-1	1.00000	0.00000	1.00000	0.00000	25	0.00000	1.00000	1.00000
MTB2014-2,MTB2014-1	1.00000	0.00000	1.00000	0.00000	23	0.00000	1.00000	1.00000
MTB2017-1	1.00000	0.00000	1.00000	0.00000	24	0.00000	1.00000	1.00000
MTB2016-1	1.00000	0.00000	1.00000	0.00000	12	0.00000	1.00000	1.00000

Confusion Table (CV):

	Actual Class			
	MTB2012-	MTB20*2	MTB2017-	MTB2016-
Predicted as MTB2012-1	25	0	0	0
Predicted as MTB2014-2,MTB2014-1	0	23	0	0
Predicted as MTB2017-1	0	0	24	0
Predicted as MTB2016-1	0	0	0	12
Predicted as Unassigned	0	0	0	0

Key:

MTB20*2 = MTB2014-2,MTB2014-1

PREDICTION RESULTS

Confusion Matrix:

Class:	TPR	FPR	TNR	FNR	N	Err	P	F1
MTB2012-1	1.00000	0.00000	1.00000	0.00000	11	0.00000	1.00000	1.00000
MTB2014-2,MTB2014-1	1.00000	0.00000	1.00000	0.00000	16	0.00000	1.00000	1.00000
MTB2017-1	1.00000	0.00000	1.00000	0.00000	11	0.00000	1.00000	1.00000
MTB2016-1	1.00000	0.00000	1.00000	0.00000	11	0.00000	1.00000	1.00000

Confusion Table:

	Actual Class			
	MTB2012-	MTB20*2	MTB2017-	MTB2016-
Predicted as MTB2012-1	11	0	0	0
Predicted as MTB2014-2,MTB2014-1	0	16	0	0
Predicted as MTB2017-1	0	0	11	0
Predicted as MTB2016-1	0	0	0	11
Predicted as Unassigned	0	0	0	0

Key:

MTB20*2 = MTB2014-2,MTB2014-1

A First Simple Model (figures right)

Even when only displaying two of the 3 latent variables this model uses, the measurements cluster into areas of classification. All data points are correctly classified in the results. Even two different bottles of the same vintage are classifiable. Note that even though MTB2014-1 and -2 are shown in different colours the PLSDA model is given the instruction to class them together – visible in the results table. When classifying by both MTB2014-1 and MTB2014-2 there is a slight confusion between the two. This is expected, however if we were to establish a model to probe the difference between individual bottles from the same batch we would fine tune it and depending on actual physical variability the model would succeed. Here, fine-tuning means that spectra of different classes that are too similar would have to be excluded from the model - although they could be used as test data.

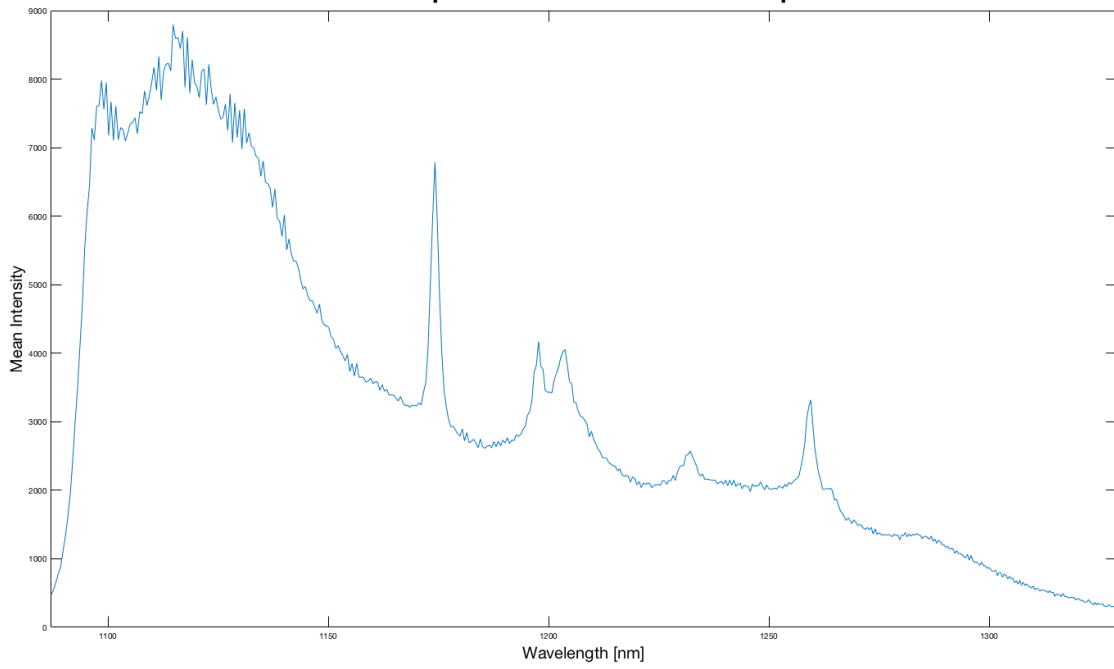
For further clarification; latent variables can be thought of as components, which each of your spectra get broken down into. That means if you sum the right amount of these components, you get back to your original measurement (ideally). How much of these common components each of your measurement (spectrum) contains is displayed in the graphs above. There are more than only three latent variables and depending on the task and size of the model there might be many more. For illustrative purposes a very simple model with only three substances to be classified is shown.

A Spectrum Example

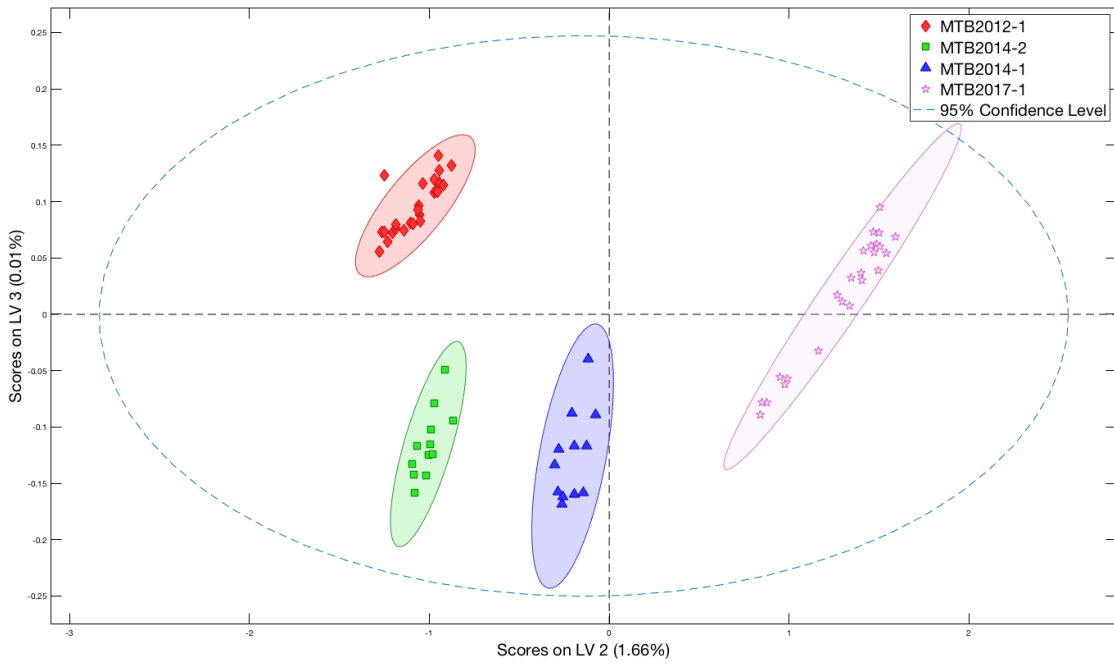
This is what a single data point on these Latent-Variable graphs actually looks like. A spectral Raman response of 512 data points, with a

wavelength assigned to each.

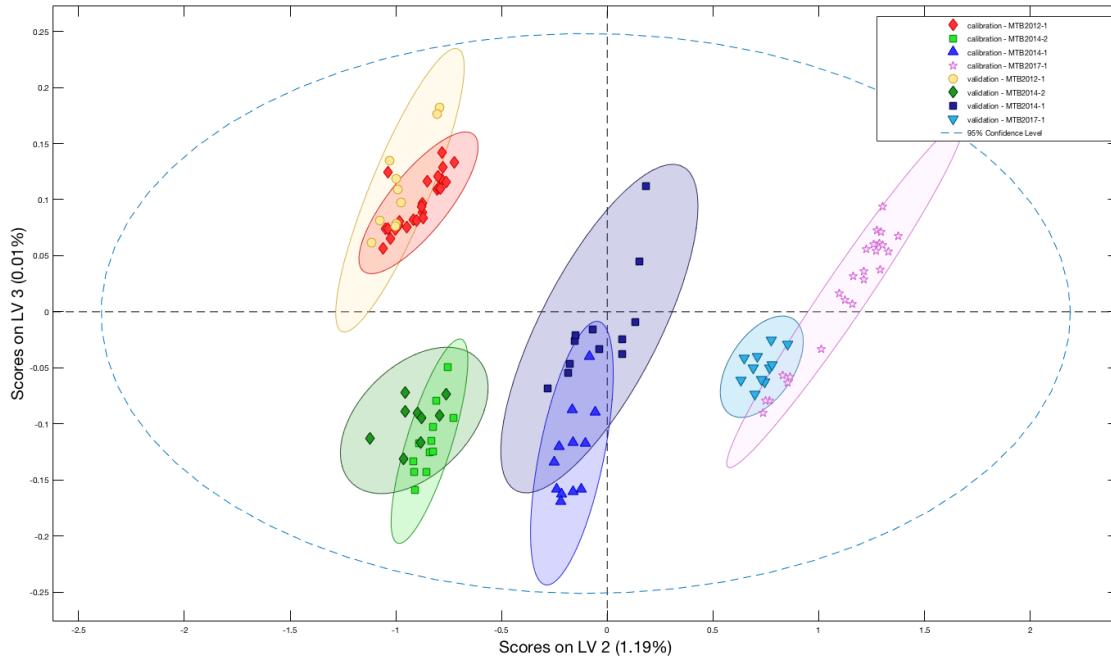
Wine Spectrum in Glass Bottle Example



PLSDA Model - 3 MTB Wines



PLSDA Model and applied Data - 3 MTB Wines

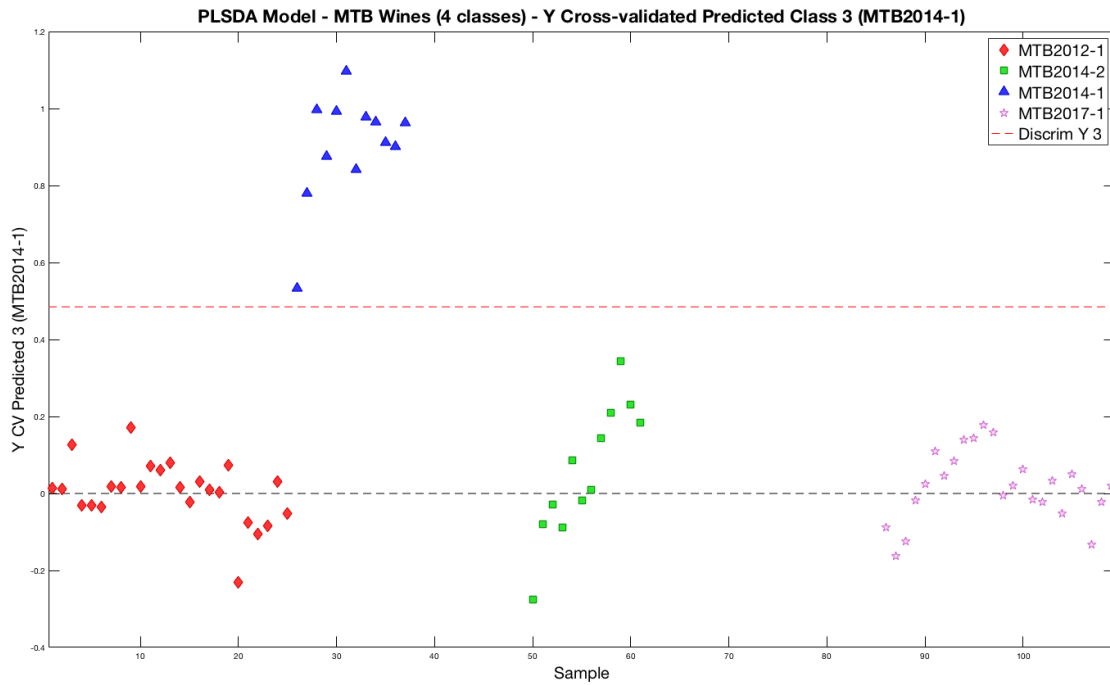


Adding more/ similar wines to the model (figures right)

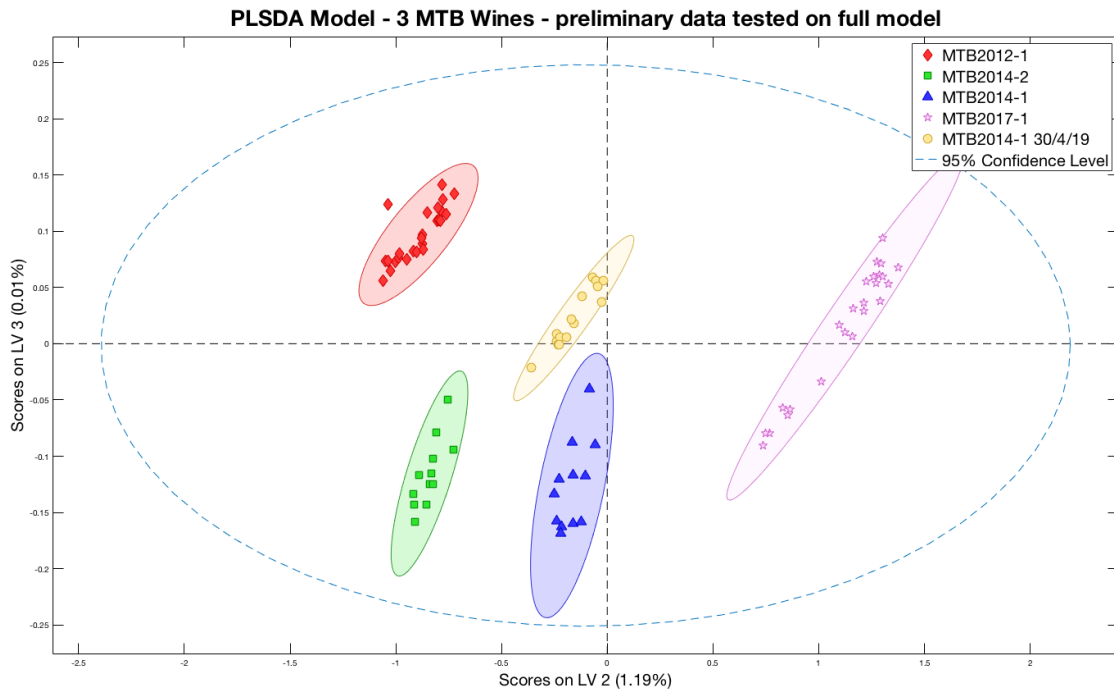
MTB2016 is not as easily distinguishable. Note that the model changes as it is now built with MTB2016 data, meaning that the common components the algorithm searches for also change. A more detailed look into the latent variables is required to visualise the classification. The PLS algorithm in this case looks at five latent variables at the same time but some of the models we use in these small experiments even have up to 8 or more. Of course this cannot be visualised in one graph but it is mathematically necessary to distinguish some test substances. Below an example of how MTB2016 yields similar result to MTB2014 when looking at these LV's. Looking at multiple LVs (3D right), clearly classifies them separately and enables us to build a model.

Lastly, we apply our test data to this model again and in the results table all of them are classified correctly.

Note that measurements can vary and some are more clearly classified than others. To ensure that one measurement is classified correctly, we can set a threshold, which the measurement has to pass to classify as wine X. An ideal result has measurements of one class above the threshold and all others below, as shown here (note, 2014-1 and -2 as separate classes). Keep in mind the lowest (blue) data point here, could be under the threshold for some measurement and then would be marked as unclassified if not going above the threshold for any class.

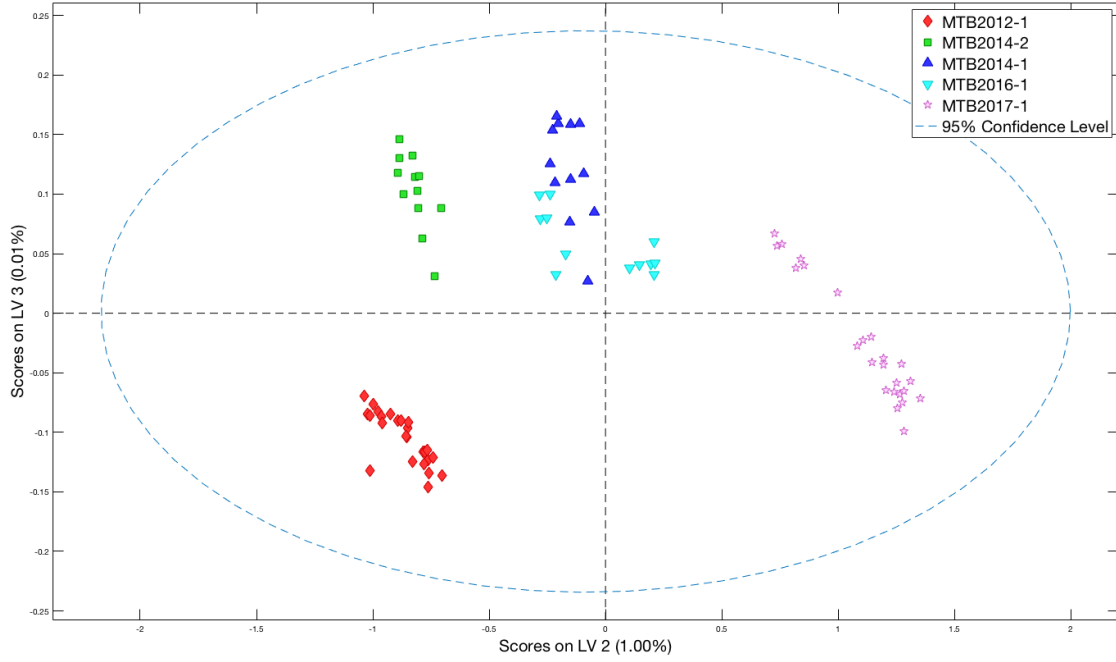


Measurements taken two weeks previously applied to the first simple model (no 2016):

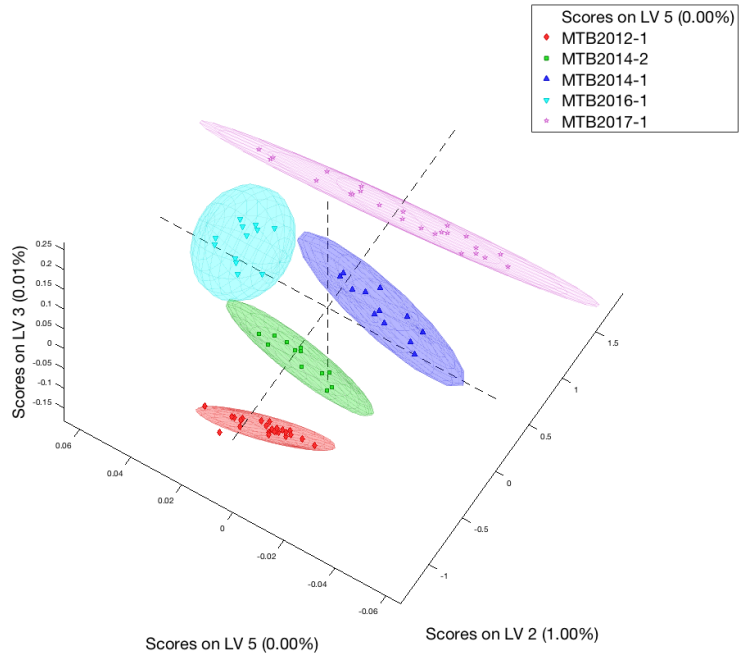


The model results table correctly classifies this wine. However, over the time the bottle was open, the wine may have changed, for it appears outside the calibration MTB2014-1 data shown above. There are other experimental reasons for an overtime drift but these were due to alignment changes, which will clearly not occur in the final device.

PLSDA Model - 4 MTB Wines; (2016 added)



PLSDA Model - 4 MTB Wines - 3 Latent Variables displayed



PLSDA Model and Test data - 4 MTB Wines

