Proceedings - 1st International Electronic Conference on Food Science and Functional Foods
10 – 15 November 2020 Online

Application of 13C NMR and Untargeted Multivariate Analysis for Classification of Virgin Coconut Oil

Lolita G. Lagurin, Mark Joseph M. Garrovillas and Fabian M. Dayrit*
Department of Chemistry, Ateneo de Manila University
Quezon City, Philippines 1108
*correspondence: fdayrit@ateneo.edu

ATENEO DE MANILA UNIVERSITY

SCIFORUM-036131

Author Contributions and Affiliations

Fabian M. DAYRIT, PhD

Project Leader and Professor Emeritus

Conceptualization
Funding Acquisition
Project Administration
Supervision
Review and Editing

Mark Joseph M. GARROVILLAS

NMR Instrument Manager and Research Staff

Methodology
Software
Formal Analysis
Validation
Original Draft
Visualization

Lolita G. LAGURIN

Senior Research Staff

Resources
Investigation
Methodology
Data Curation
Validation
Original Draft
Review and Editing

Department of CHEMISTRY

SCHOOL OF SCIENCE AND ENGINEERING

ATENEO DE MANILA UNIVERSITY

Coconut (*Cocos nucifera* L.) Oil Types: Virgin Coconut Oil (VCO) and Refined Bleached Deodorized Coconut Oil (RBDCO)

Fresh Coconuts

Fresh coconut meat

Expeller

Coconut milk

Centrifuge

Fermentation

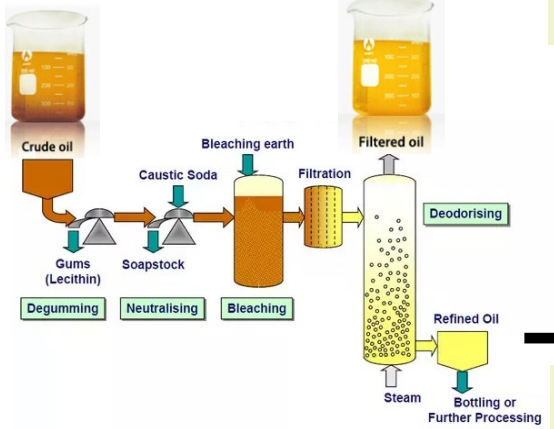
VCO

VCO

VCO

Dried Copra

Expeller



Refined oil

VCO (Control)
- Labor intensive, premium product
- Is colorless
- Has coconut aroma
- Has more phytonutrients
- Shorter Shelf Life

***Old VCO**
- Over two years old at ambient temperature
- Accelerated degradation, 6 months at 40°C

RBDCO
- Mass produced, readily available
- Slight yellow tint
- Odorless
- Phytonutrients have been "refined" out
- More stable, amenable to longer storage

**** Adulterated VCO**
- RBDCO added to Control VCO as adulterant in 25%, 50% and 75% proportions

Instrumental Methods for Differentiating VCO

Method	Analytes	VCO vs RBDCO	VCO Production Processes	Old VCO	VCO + RBDCO Adulteration	Remarks
Headspace / GC-MS	*Volatile organic compounds (VOC)*	Yes	Partial*	No data	No data	*Sample preparation is extensive*
FTIR	*IR Profile*	No data	No data	No data	No data	*Did not account for adulteration of VCO by RBDCO*
DSC	*Thermal Profile and Heat Capacity*	No data	No data	No data	No data	*Did not account for adulteration of VCO by RBDCO*
31P NMR	*Mono- and di-glycerides, free fatty acid (FFA), sterol content*	Yes	Not conclusive	No data	No data	*Needs derivatization step, multi-component NMR solvent mixture with pyridine*

*Only Fermentation VCO was differentiated from the rest

Research **Objectives**

Using **13C NMR Profiling** and **linear methods** are the following situations possible?

Differentiate by Sample Type

- Differentiate Control VCO from Not Control VCO using Binary Classifiers (one vs one)
  - Control VCO vs Not Control VCO, Oil Type - RBDCO
  - Control VCO vs Not Control VCO, Oil Type - Old VCO
  - Control VCO vs Not Control VCO, Oil Type - Adulterated VCO

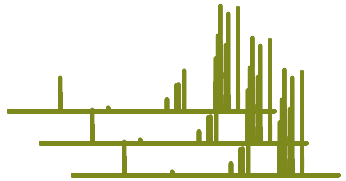
Differentiate by VCO Process

- Differentiate Control VCO by Manufacturing Processes using Binary Classifiers (one vs rest)
  - Fermentation VCO vs Not Fermentation VCO
  - Centrifuge VCO vs Not Centrifuge VCO
  - Expeller VCO vs Not Expeller VCO

Methodology (Metabolomics Workflow)

Instrumental Analysis

- Sampling (*n=98*)
- 13C NMR (*Bruker Avance Neo 400*)



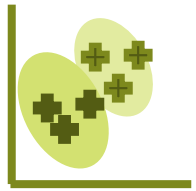
Pre-processing

- NMR Spectra Processing
- Alignment, Peak Shifting (*NMRPROCFLOW*)
- Variable Bucketing (*NMRPROCFLOW*)

Sample	Feature 1	Feature 2	Feature 3
A1	A1X1	A1X2	A1X3
A2	A2Y1	A2Y2	A2Y3
A3	A3Z1	A3Z2	A3Z3

Normalization

- Internal Standard (1,4-dioxane) (*MetaboAnalyst 4.0 – Protocol 1*)
- Auto-scaling (*MetaboAnalyst 4.0 – Protocol 1*)

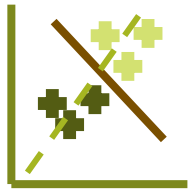


Unsupervised Exploratory Analysis

- PCA (*MetaboAnalyst 4.0 – Protocol 3*)

Optimize # of PLSDA-variables

Optimize # of Features



Supervised Analysis

- PLS–DA Using Binary Classifiers
- Evaluate by R2, Q2, Accuracy (*MetaboAnalyst 4.0 – Protocol 3*)
- Evaluate Accuracy by AUC-ROC Curves (*MetaboAnalyst 4.0 – Protocol 5*)
- Determine Statistical Significance by Permutation Tests (*MetaboAnalyst 4.0 – Protocol 35*)

Control VCO and Not Control VCO **Samples** (*n = 98*)

Classification of Data For Model Development and Validation			Training (Observed) [a]	Validation (Submitted) [b]	Sub-Total per VCO Process	Sub-Total per Sample Type	Total Samples
Sample Type	*Control VCO / VCO Process Type*	*Fermentation VCO*	14	5	19	57	98
		Centrifuge VCO	13	5	18		
		Expeller VCO	15	5	20		
	Not Control VCO / Oil Type	*RBDCO*	11	10 [c]	21	41	
		Old VCO	7	4 [c]	11		
		Adulterated VCO	6	3 [c]	9		

Description of Samples:

[a] **Training (Observed)** : Control VCO Samples

[b] **Validation (Submitted)**: manual holdout submitted Control VCO samples

[c] **Validation (Submitted)** : Manual holdout, RBDCO, Old VCO and Adulterated VCO Samples chosen randomly

Instrumental Analysis - 13C NMR Profiling

Sample Preparation

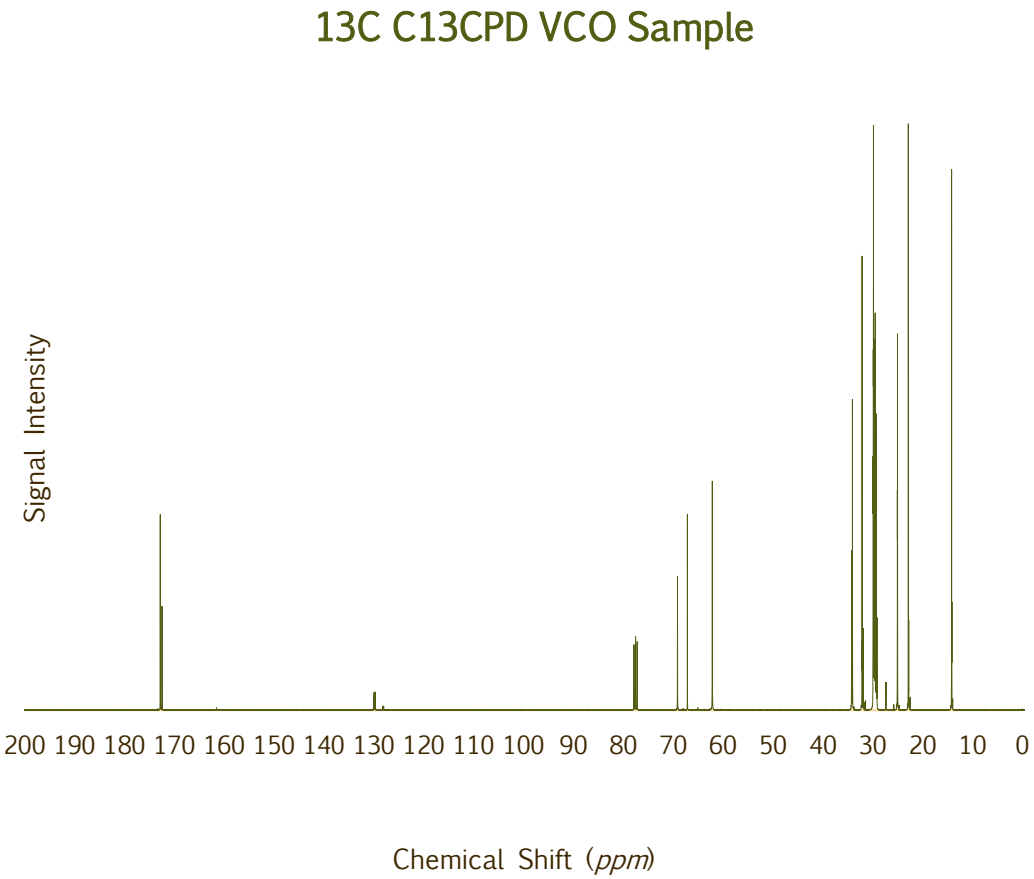
- Oil Sample: 350 µL
- NMR Tube: 5 mm High Throughput
- Solvent: 230 µL (2.9% 1,4-dioxane Internal Standard (IS) in CDCl3

NMR Acquisition

- Bruker Avance Neo 400 NMR
- 13C at 100 MHz standard pulse sequence
- 64K points; 4K scans; SW: -8.6 ppm to 219.3 ppm; autogain; VT = 300K

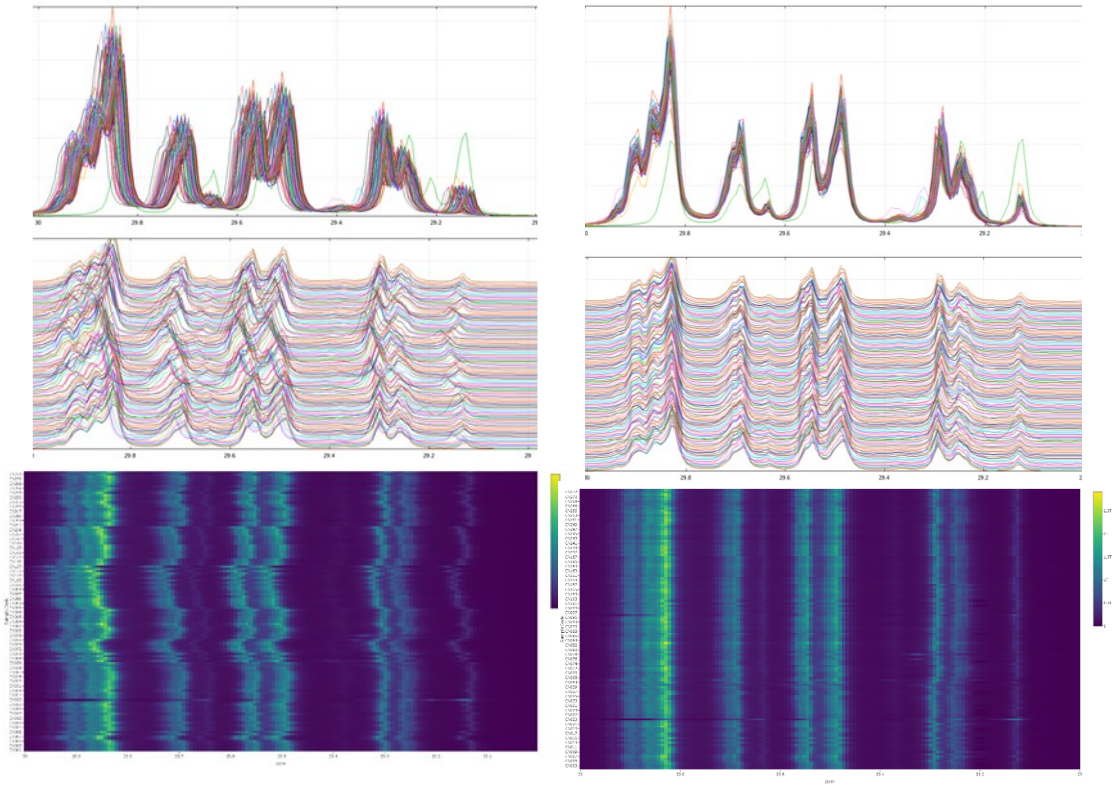
NMR Processing

- Fourier Transformation
- Apodization—exponential multiplication: 1 Hz
- Auto-phasing and baseline optimization—apbk

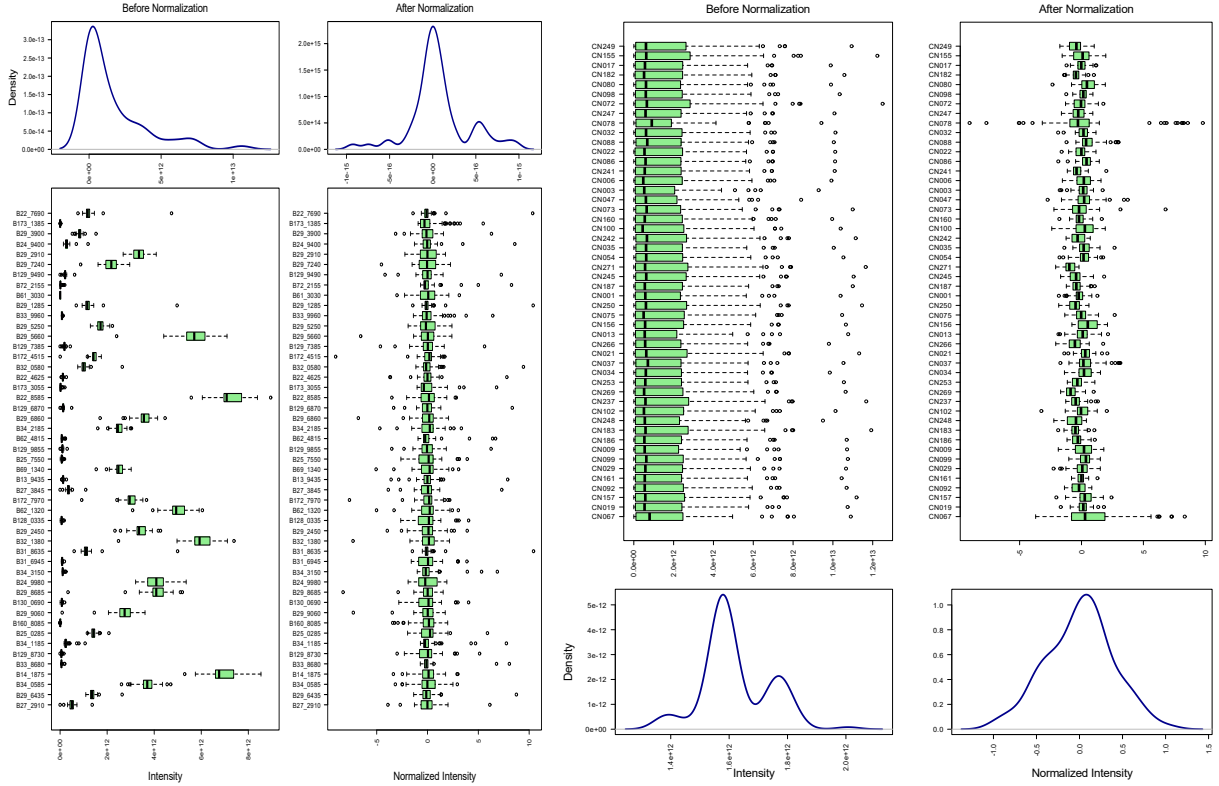


Sample	Feature 1	Feature 2	Feature 3
A1	A1X1	A1X2	A1X3
A2	A2Y1	A2Y2	A2Y3
A3	A3Z1	A3Z2	A3Z3

Spectral **Alignment** and **Normalization**

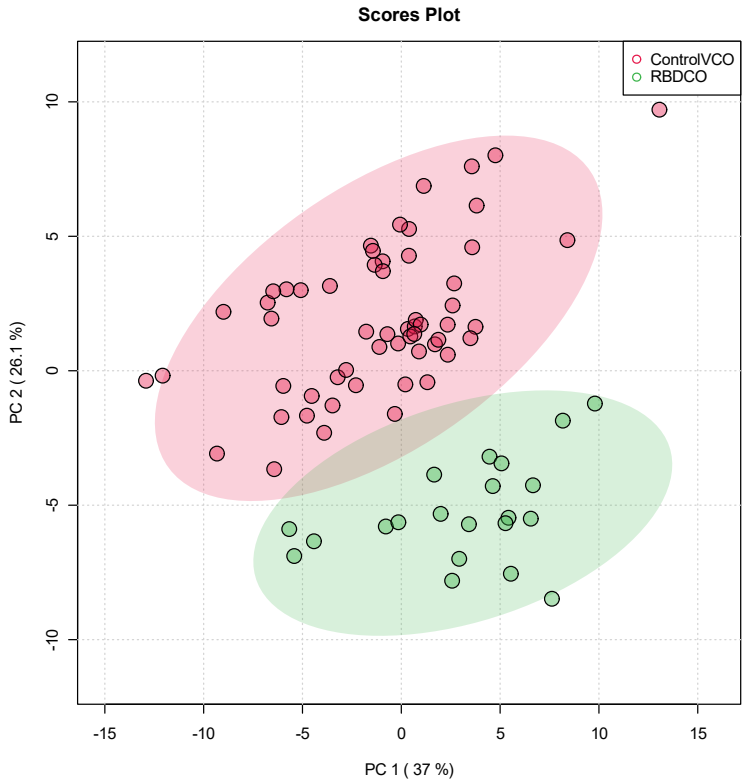


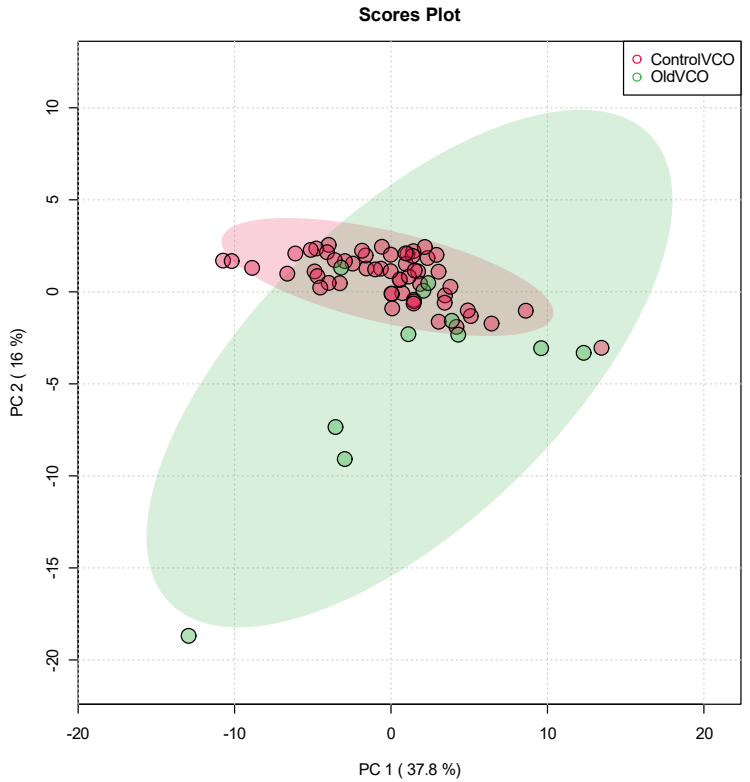
Before and After Alignment
(30.0 to 29.0 ppm)

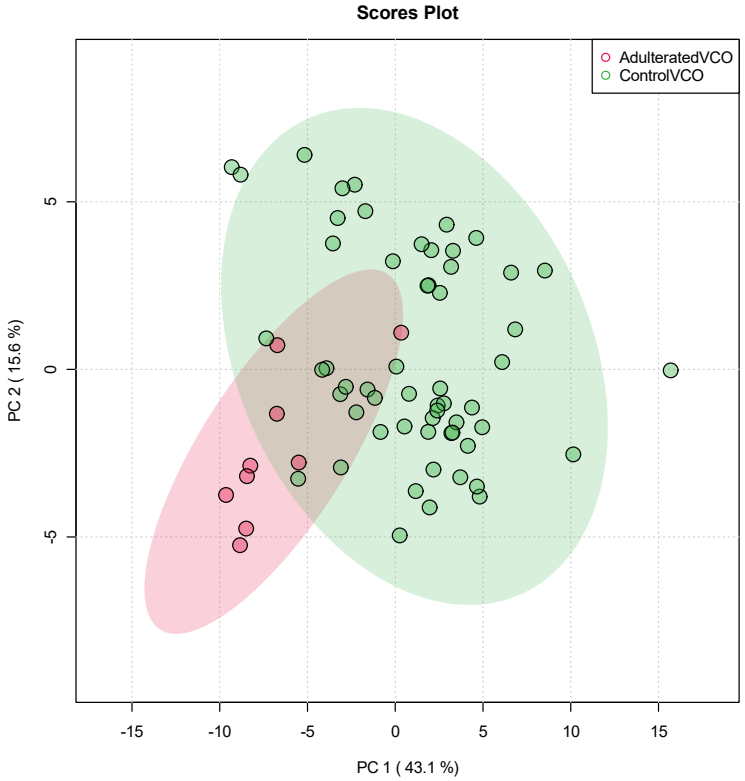


Before and After Normalization
(by Features and by Samples)

PCA of Control VCO vs Not Control VCO

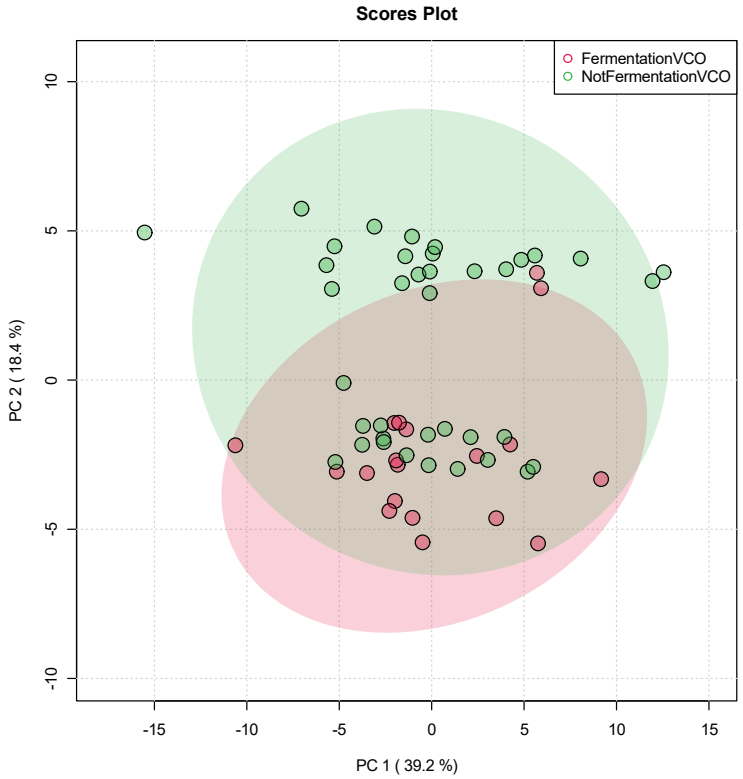


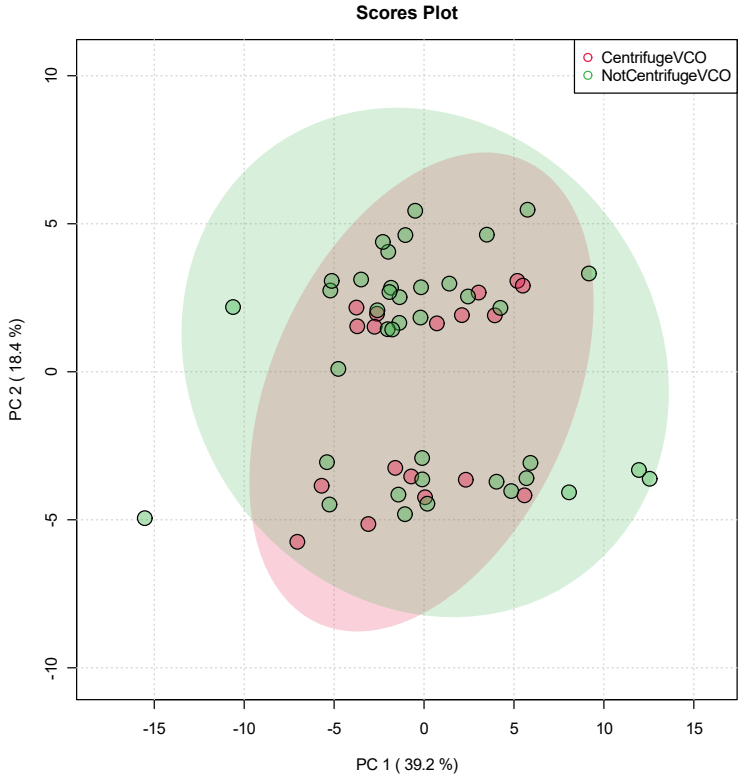


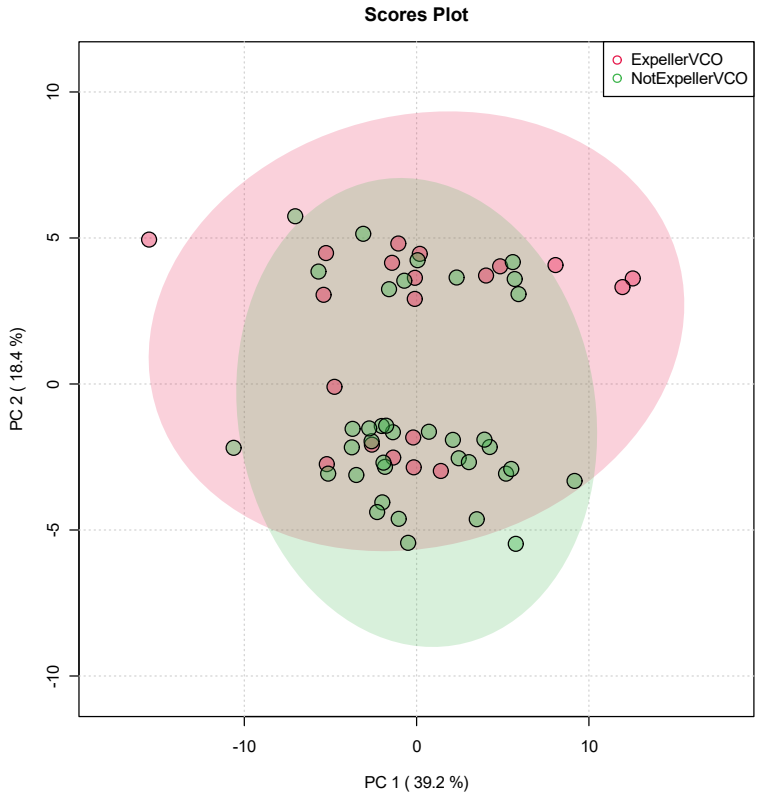


Control VCO vs RBDCO has clear clustering of samples.
Control VCO vs Old VCO and Adulterated VCO may be good models.

PCA: VCO Processes





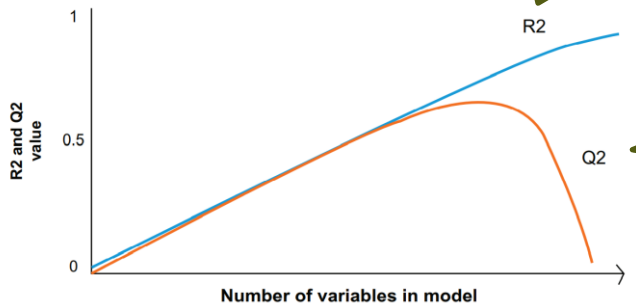


Models based on VCO Processes may NOT be good models; overlapping samples

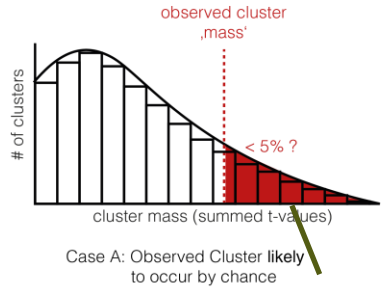
PLS-DA: Evaluation by R2, Q2, Permutation Tests, Receiver Operating Characteristic (ROC) Curves in Binary Classifiers & Cross-Validation

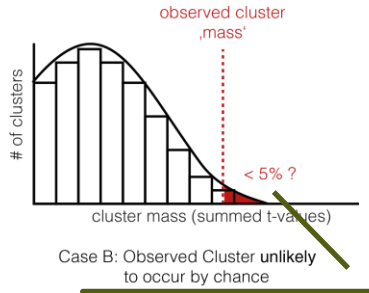
Measure of Model Fit

Maximize R2 and Q2 while minimizing number of PLS-DA variables for optimization



Measure of Model Predictability

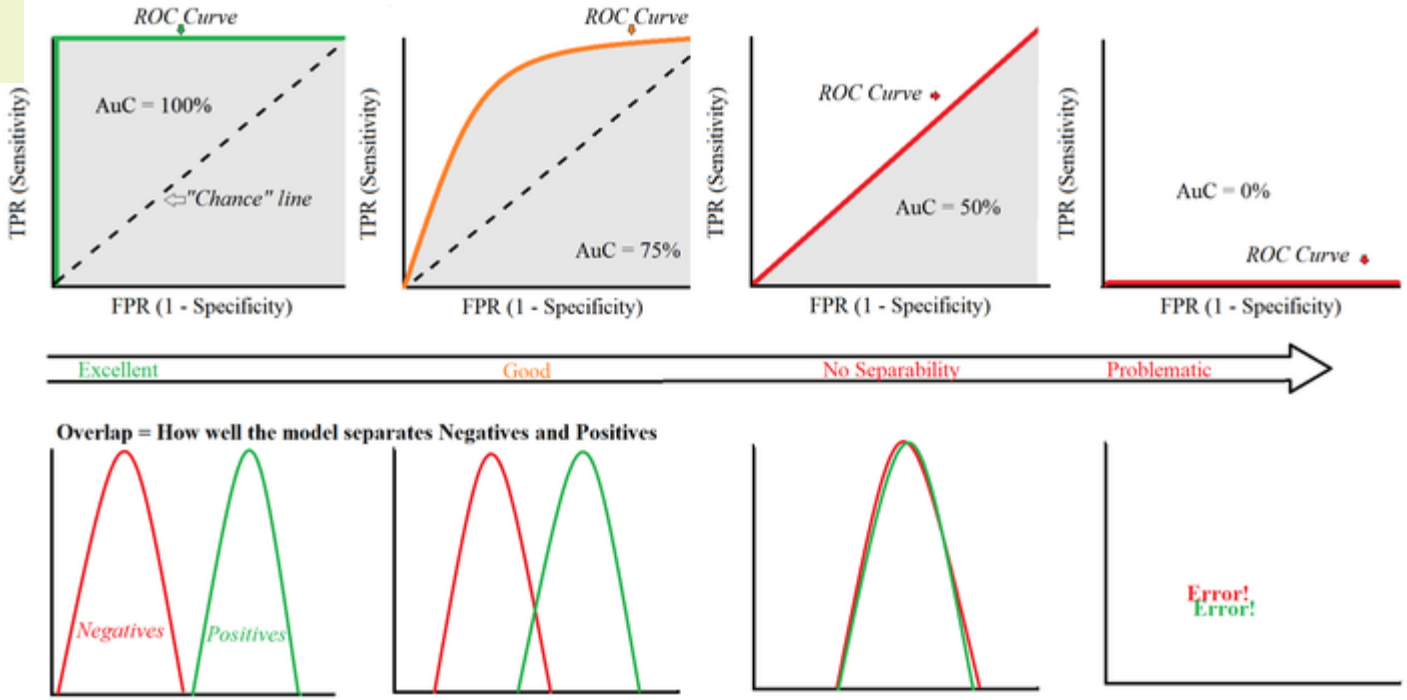




$p > 0.05$
Model may **NOT** be Statistically Significant

$p < 0.05$
Model is Statistically Significant

p-value must be as low as possible to ensure results are statistically significant for Permutation Tests



PLS-DA and AUC-ROC should be high for best model. (PLS-DA gives the ratio of correct predictions vs total predictions = model overall accuracy)

Validated and Cross-Validated PLS-DA Model of **Control VCO** vs **Not Control VCO** indicate they may be **Good** to **Perfect Models**

PLS-DA Binary Models		Overall Accuracy (Preliminary Model)	R2	Q2	Permutation Test (Preliminary Model)	AUC-ROC (MCCV*)	AUC-ROC (100 CV**)	AUC-ROC (Optimized Validation Model)	Overall Accuracy (Optimized Model)	Statistical Significance (Optimized Model)	Remarks
Control VCO / VCO Process Type	Fermentation VCO	> 0.60	> 0.20	< 0.2	p = 0.002	NA	NA	NA	NA	NA	**Poor model** (Statistically significant)
	Centrifuge VCO	> 0.40	< 0.20	< 0.2	p = 0.403	NA	NA	NA	NA	NA	**Poor model** (Not statistically significant)
	Expeller VCO	> 0.60	> 0.20	< 0.2	p = 0.149	NA	NA	NA	NA	NA	**Poor model** (Not statistically significant)
Not Control VCO / Oil Type	RBDCO	> 0.80	> 0.60	> 0.60 (**4** optimal variables)	p < 0.001	**0.996 - 1** (CI: 0.949 - 1)	**1** (CI: 1 - 1)	**1**	**1**	p < 2.502e-05	**Perfect model** (Statistically significant)
	Old VCO	> 0.80	> 0.40	> 0.20 (**3** optimal variables)	p = 0.001	**0.733 - 0.922** (CI: 0.447 - 0.994)	**0.957** (CI: 0.847 - 1)	**0.984**	**0.9**	p = 0.03	**Excellent model** (Marginally statistically significant)
	Adulterated VCO	> 0.80	> 0.40	> 0.20 (**4** optimal variables)	p < 0.001	**0.843 - 0.904** (CI: 0.568 - 1)	**0.819** (CI: 0.595 - 1)	**1**	**0.944**	p = 0.241	**Good model** (May not be statistically significant)

***** Monte-Carlo cross validation (MCCV) with balanced sub-sampling: (2/3) training; (1/3) Validation. The Control VCO / Process Type classifiers did not undergo further cross validation and cross validation due to low Q2 value.

****** The top features with univariate AUC > 0.99 for vs RBDCO, AUC > 0.70 for Old and AUC > 0.90 for vs Adulterated were selected, with 100 cross validations (CV) to generate a smooth ROC curve with 95% confidence interval.

Graphical Presentation of these values (ROC curves, permutation tests are in the extra slides at the end.

Summary of Conclusions

To differentiate Control VCO from Not Control VCO samples, and by VCO Manufacturing Processes, 13C NMR Profiling and binary linear classifier models were evaluated.

Control VCO vs Not Control VCO			*VCO Process*		
Overall Accuracy / AUC-ROC > *0.80* Good Predictive Ability (Q2 > 0.20)			Overall Accuracy < *0.80* Poor Predictive Ability (Q2 < 0.20)		
Control VCO from RBDCO (Perfect Model Performance)	Control VCO from Old VCO (Excellent Model Performance)	Control VCO from Adulterated VCO (Good Model Performance)	Fermentation VCO from Not Fermentation VCO (Unusable Model)	Centrifuge VCO from Not Centrifuge VCO (Unusable Model)	Expeller VCO from Not Expeller VCO (Unusable Model)
Statistically Significant ($p < 0.05$)	Statistically Significant ($p < 0.05$)	NOT Statistically Significant	Statistically Significant ($p < 0.05$)	NOT Statistically Significant	NOT Statistically Significant

Acknowledgements

ATENEO DE MANILA UNIVERSITY

Other Institutions

Funding Agency

REPUBLIC OF THE PHILIPPINES
PHILIPPINE COUNCIL FOR INDUSTRY, ENERGY, AND EMERGING TECHNOLOGY RESEARCH AND DEVELOPMENT (DOST-PCIEERD)
DEPARTMENT OF SCIENCE AND TECHNOLOGY

Industry Partner

VIRGIN COCONUT OIL PHILIPPINES

Research Collaborator

University of the Philippines
College of Home Economics

Research Staff

Grace B. Tantengco

Jerika Mae A. Arceo

Image Credits

https://sc01.alicdn.com/kf/ULB8X023s3nJXKJkSaelq6xUzXXaE.jpg

https://qph.fs.quoracdn.net/main-qimg-9717426905277754ccbd157e0cb7eba88.webp

http://www.bestoilpressmachines.com/uploads/allimg/crude-edible-oil-refined-edible-oil.jpg

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3460701/bin/opth-6-1499f6.jpg

https://storage.ning.com/topology/rest/1.0/file/get/1341805045?profile=RESIZE_710x

References

Marina, A.M.; Che Man, Y.B.; Amin, I. Virgin coconut oil: emerging functional food oil. Trends Food Sci. Technol. 2009, 20, 481–487.

Bawalan, D.D. The Right Technology to Process Quality Virgin Coconut Oil for the Global Market. Coco Info Int. 2016, 23, 13–19.

Pham, L.J. Coconut (Cocos nucifera). In Industrial Oil Crops; McKeon, T.A., Hayes, D.G., Hildebrand, D.F., Weselake, R.J., Eds.; Elsevier, 2016; pp. 231–242 ISBN 978-1-893997-98-1

Rohman, A.; Irnawati; Erwanto, Y.; Lukitaningsih, E.; Rafi, M.; Fadzilah, N.A.; Windarsih, A.; Sulaiman, A.; Zakaria, Z. Virgin Coconut Oil: Extraction, Physicochemical Properties, Biological Activities and Its Authentication Analysis. Food Rev. Int. 2019, 1–21, doi:10.1080/87559129.2019.1687515.

Santos, J.E.R.; Villarino, B.J.; Zosa, A.R.; Dayrit, F.M. Analysis of Volatile Organic Compounds in Virgin Coconut Oil and their Sensory Attibutes. Philipp. J. Sci. 2011, 140, 161–171.

Liland, K.H. Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. TrAC Trends Anal. Chem. 2011, 30, 827–841, doi:10.1016/j.trac.2011.02.007.

Worley, B.; Powers, R. Multivariate Analysis in Metabolomics. Curr. Metabolomics 2012, 1, 92–107, doi:10.2174/2213235X11301010092.

Dayrit, F.M.; Dios, A.C. de 1H and 13C NMR for the Profiling of Natural Product Extracts: Theory and Applications. In Spectroscopic Analyses - Developments and Applications; Sharmin, E., Zafar, F., Eds.; IntechOpen: Rijeka, 2017.

Jacob, D.; Deborde, C.; Lefebvre, M.; Maucourt, M.; Moing, A. NMRProcFlow: a graphical and interactive tool dedicated to 1D spectra processing for NMR-based metabolomics. Metabolomics 2017, 13, 36, doi:10.1007/s11306-017-1178-y.

Chong, J.; Wishart, D.S.; Xia, J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. Curr. Protoc. Bioinforma. 2019, 68, e86, doi:10.1002/cpbi.86.

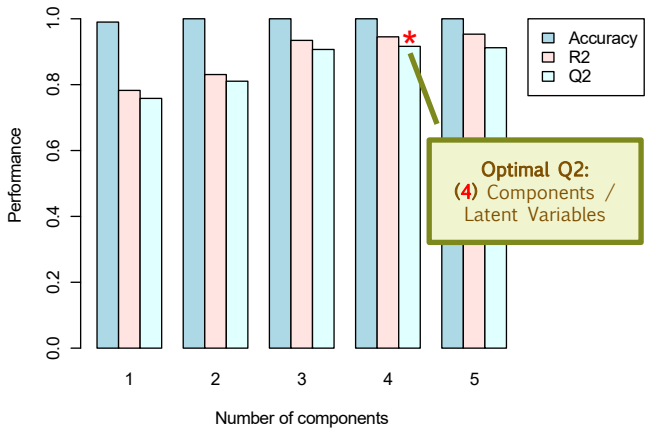
Xia, J.; Broadhurst, D.I.; Wilson, M.; Wishart, D.S. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. Metabolomics 2013, 9, 280–299, doi:10.1007/s11306-012-0482-9

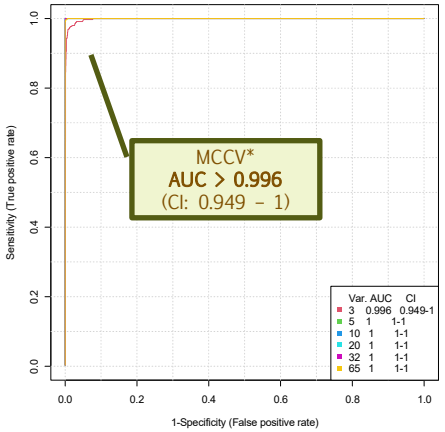
R Core Team R: A Language and Environment for Statistical Computing 2020.

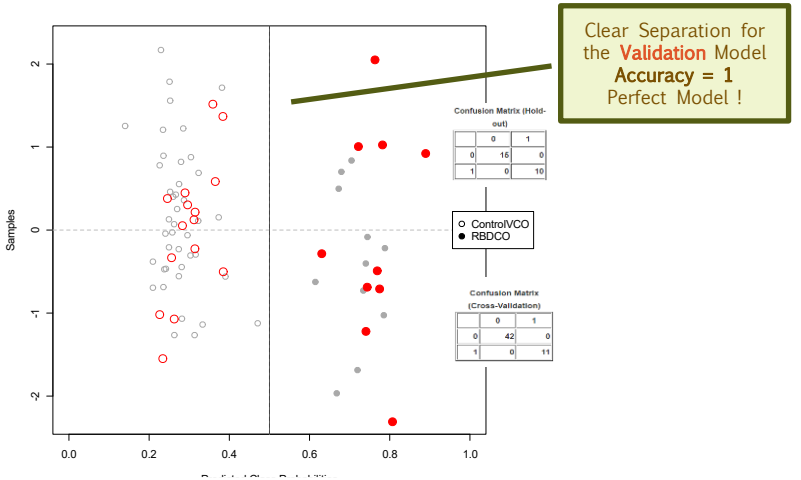
Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the Tidyverse. J. Open Source Softw. 2019, 4, 1686, doi:10.21105/joss.01686.

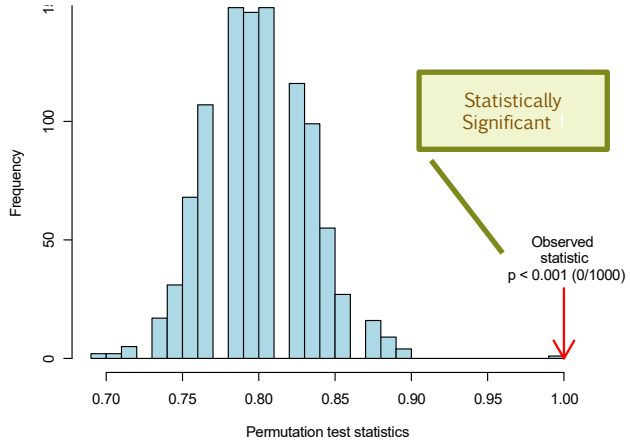
Sievert, C. Interactive Web-Based Data Visualization with R, plotly, and shiny; Chapman and Hall/CRC, 2020; ISBN 9781138331457.

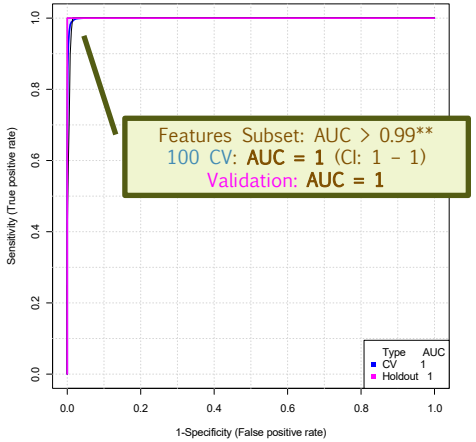
Extra Slide - Validated and Cross-Validated PLS-DA Model Performance of **Control VCO vs RBDCO** indicate a **Perfect Model**

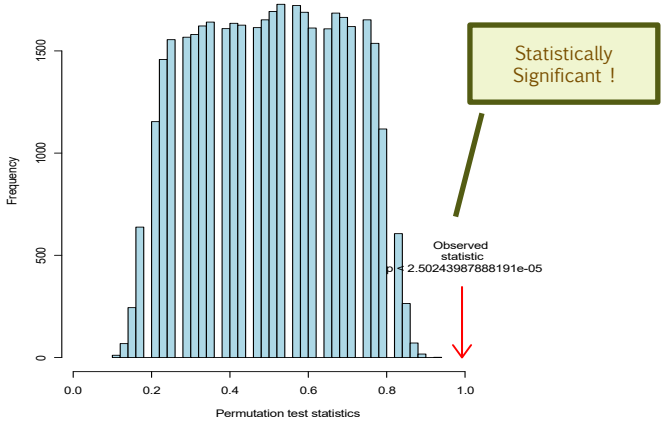










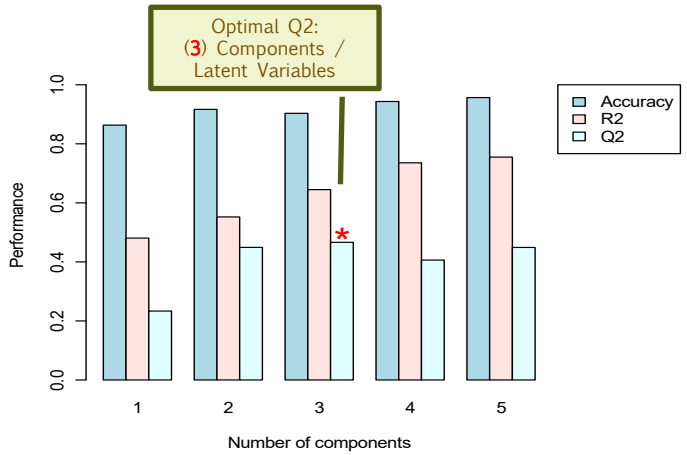


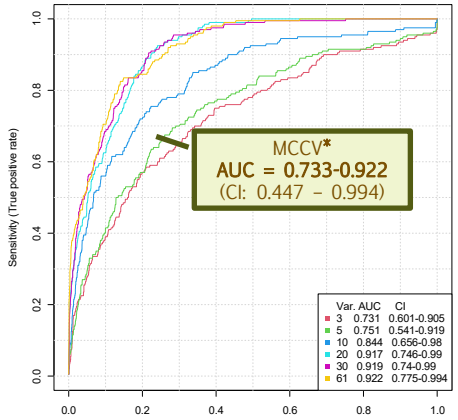
* Monte-Carlo cross validation (MCCV) with balanced sub-sampling: (2/3) training; (1/3) Validation.

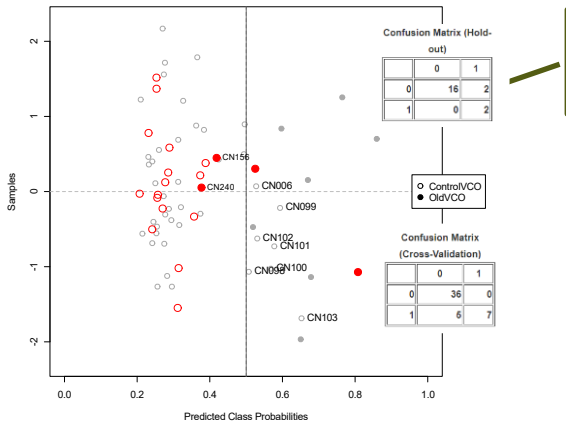
Most models generated would be considered perfect.

** The top features with univariate AUC > 0.99 were selected, with 100 cross validations (CV) to generate a smooth ROC curve with 95% confidence interval.

Extra Slide - Validated and Cross-Validated **PLS-DA Model** of **Control VCO vs Old VCO** indicate an **Excellent** Model Performance





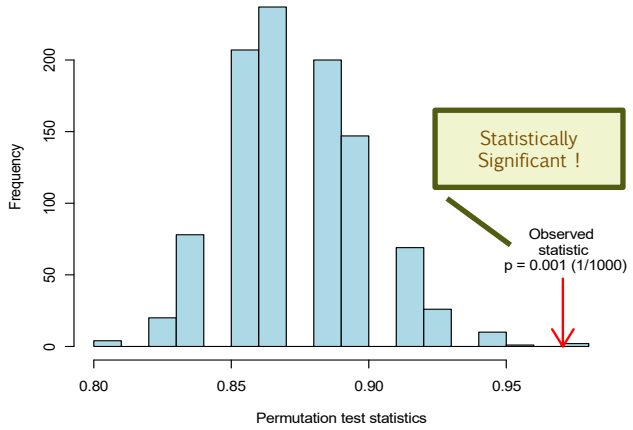


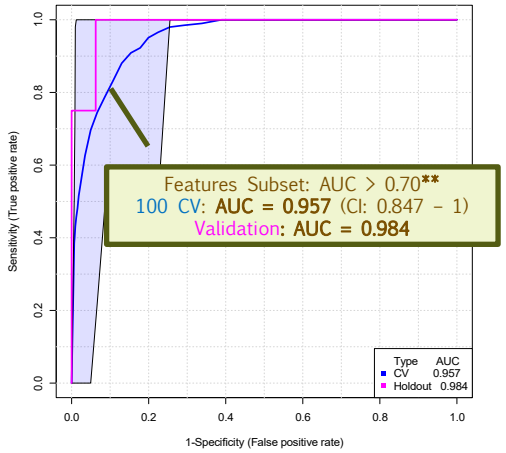
Validation Model:
Misclassifications (n = 2)
Accuracy = 0.9

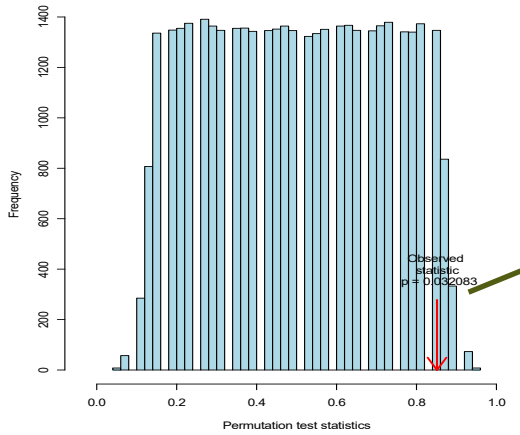
* Monte-Carlo cross validation (MCCV) with balanced sub-sampling: (2/3) training; (1/3) Validation.

AUC ⇑ as # Features ⇑

** The top features with univariate AUC > 0.70 were selected, and 100 cross validations (CV) to generate a smooth ROC curve with 95% confidence interval.

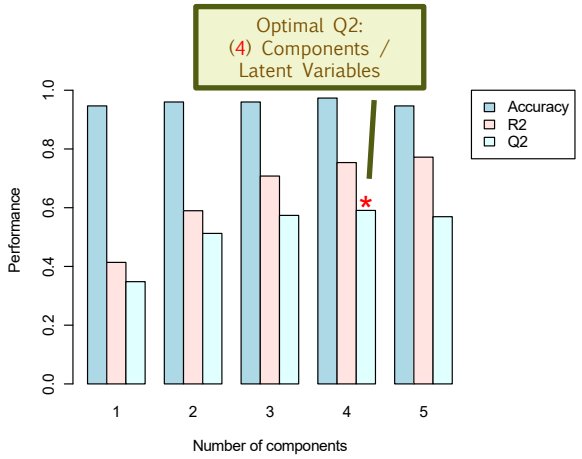


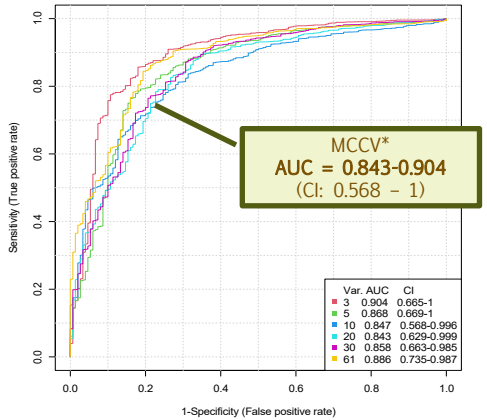


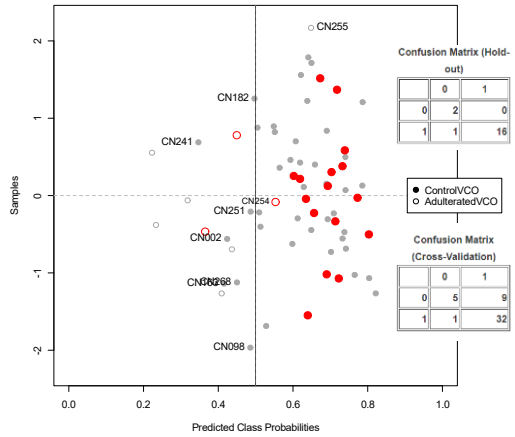


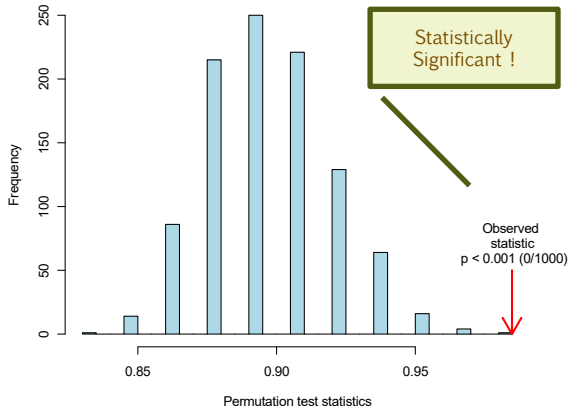
p = 0.03
Optimized Model is Marginally Statistically Significant!

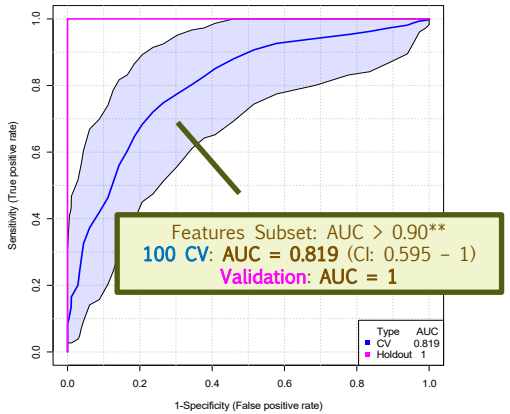
Extra Slide - Validated and Cross-Validated **PLS-DA Model** of **Control VCO vs Adulterated VCO** indicate a **Good** but **NOT Statistically Significant** Model Performance

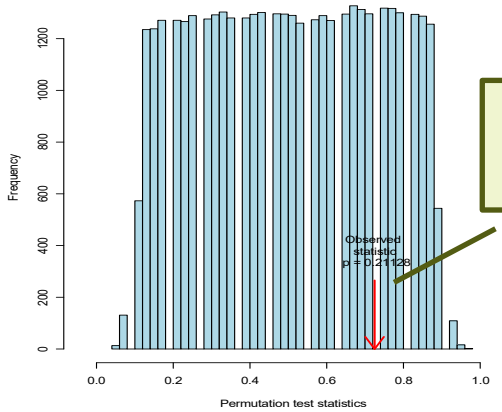










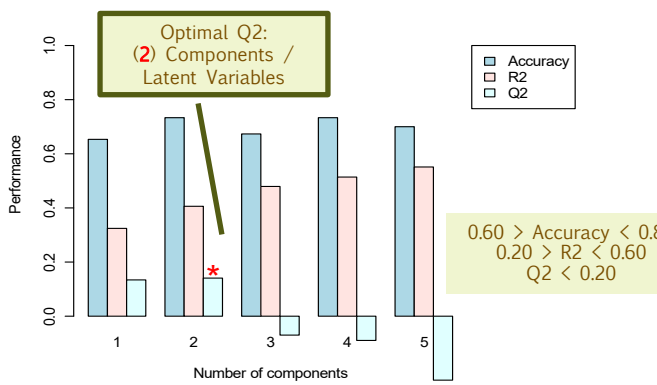


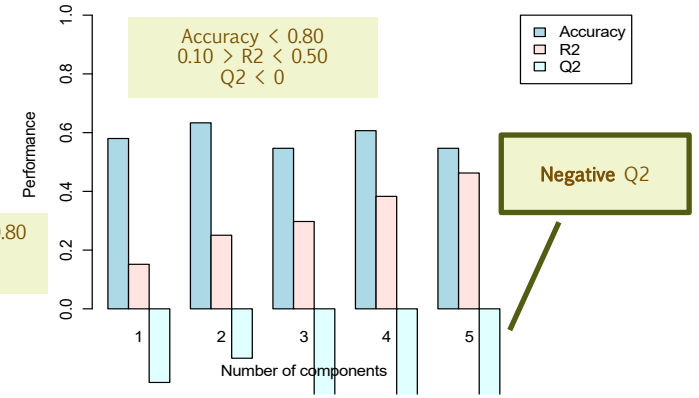
* Monte-Carlo cross validation (MCCV) with balanced sub-sampling: (2/3) training; (1/3) Validation.

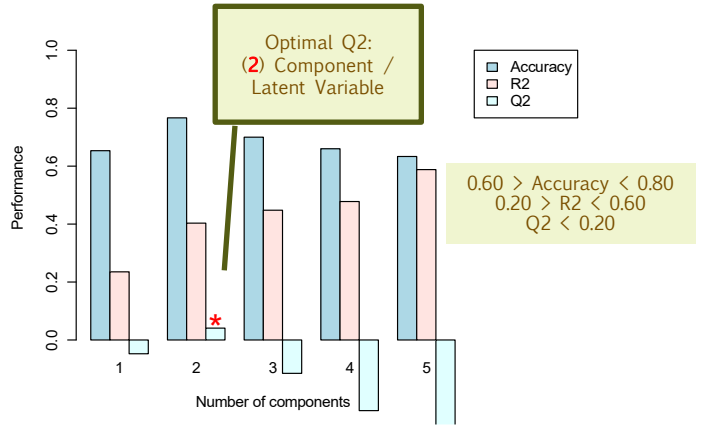
AUC ⇓ then ⇑ as # Features ⇑

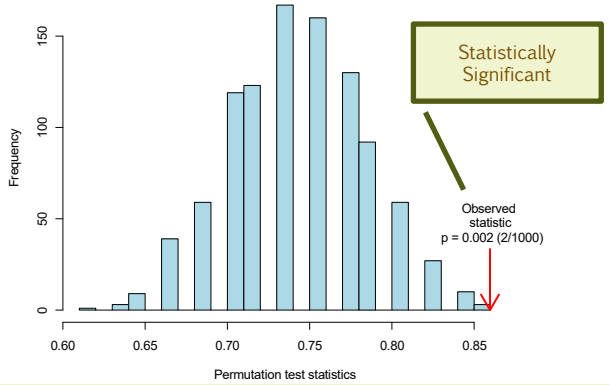
** The top features with univariate AUC > 0.90 were selected, with 100 cross validations (CV) to generate a smooth ROC curve with 95% confidence interval.

Extra Slide - Cross Validated **PLS-DA Model** of **VCO Process** Binary Classifiers Generated **Unusable** Models

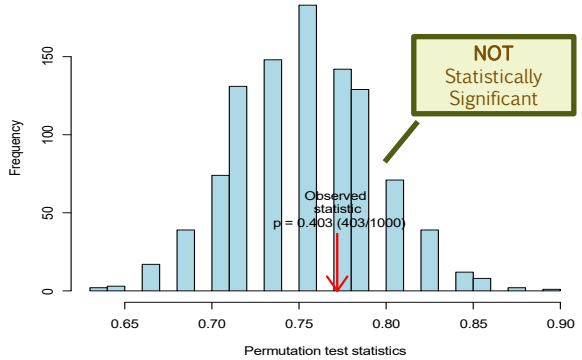




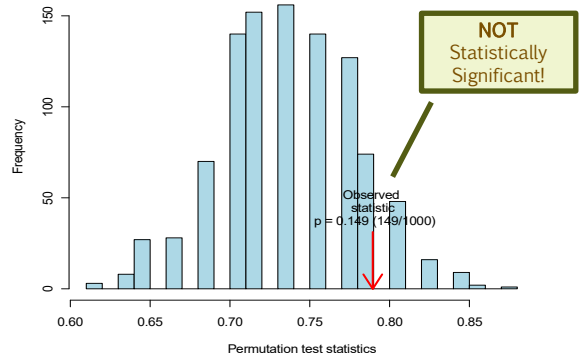




Fermentation VCO vs Not Fermentation VCO
Low Q2 / Predictive Ability



Centrifuge VCO vs Not Centrifuge VCO
Negative Q2 / **No** Predictive Ability
NOT Statistically Significant



Expeller VCO vs Not Expeller VCO
Low Q2 / Predictive Ability
NOT Statistically Significant