Spectroscopic Fingerprinting Techniques for Verifying Food

Dr Gerard Downey
Teagasc
Ashtown Food Research Centre
Dublin 15, Ireland

gerard.downey@teagasc.ie

Background

- EU-funded TRACE project: 2005 to 2009
- Workpackage 2: Fingerprint and profiling methods
- Foods studied:
  - Olive oil
  - Honey
  - Trappist beer
  - Aged beef

Initial considerations

- Verification:
  - Confirmation of a claim made on the food
  - May relate to :
    - Geographic origin eg Corsican honey (PDO)
    - Brand name eg Rochefort 8° Trappist beer
    - Processing eg beef aged for 21 days

- Not identification
  - ie what is this?

Question formulation

- Precision in defining the question to be answered at the outset is key to maximising the likelihood of a successful outcome eg
  - This olive oil claims to be extra virgin from the Kolymvari PDO region of Crete – is it?

  - This beer claims to be Trappist Rochefort 8° from Belgium – is it?

Analytical strategy

- Unlike conventional chemistry, fingerprint spectroscopic methods do not rely on the detection of a limited number of analytes which indicate identity or adulteration
- They record an analytical response to a large number of variables and manipulate these mathematically to generate a fingerprint for a specific sample type
- Derived fingerprints are applied to analytical responses measured on unknown samples to indicate whether they match those previously developed

Assumptions

- Spectral data contains useful and relevant information with which to solve the problem
- Samples used to generate models span most of the variability likely to be encountered in the future
- Instrumental measurements are precise and reproducible

Multivariate Method Options

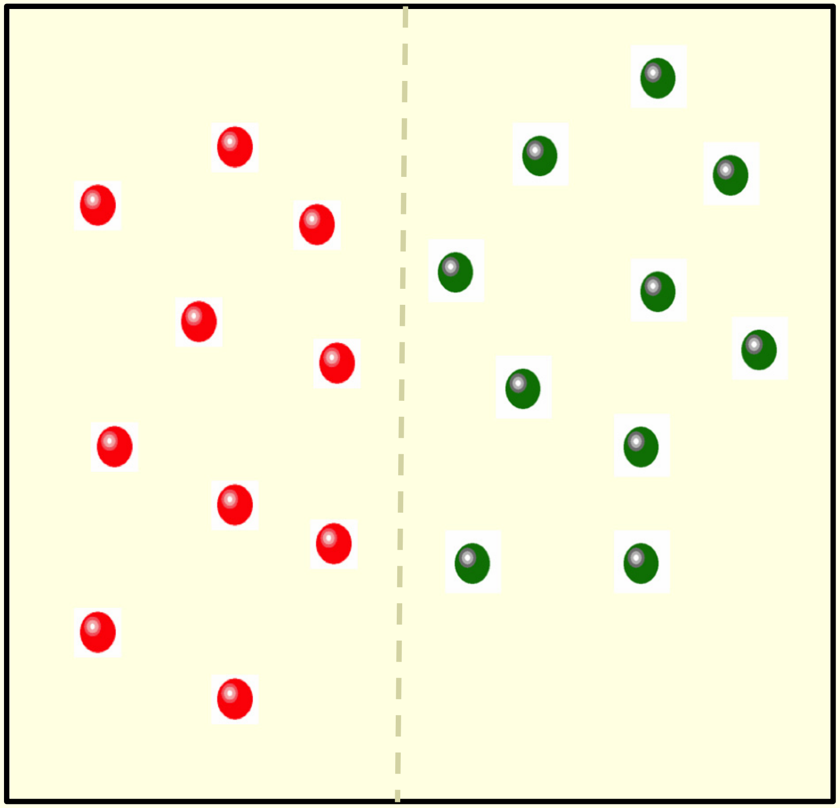
- Classification techniques
  - Suitable for closed verification systems
    - Limited and defined number of possible sample types
    - Discriminant PLS

- Class-modelling techniques
  - Suitable for open verification systems
    - Unlimited number of possible sample types
    - SIMCA, POTFUN, UNEQ

Classification Techniques

Classification techniques build a delimiter between the classes so that they always assign a new object to the class to which it most probably belongs

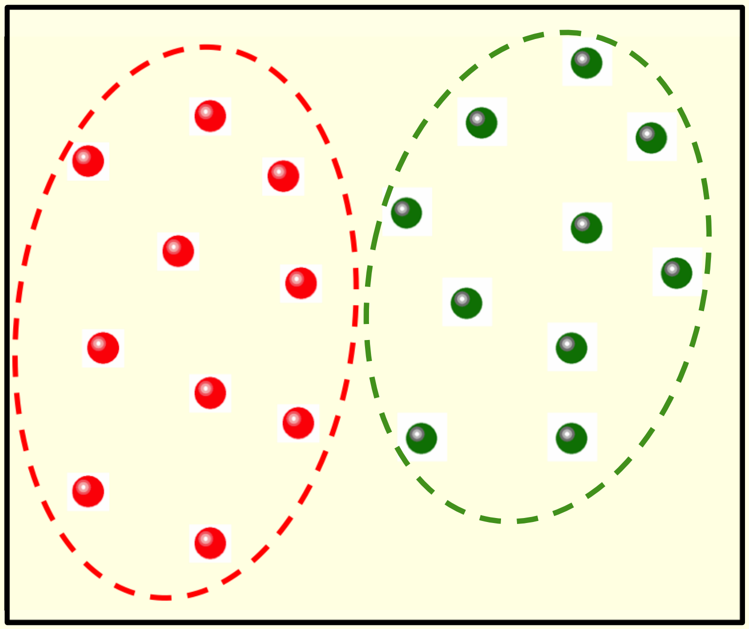
(even in the case of objects extraneous to all the classes studied)



(*P. Oliveri, University of Genoa*)

Class Modelling Techniques

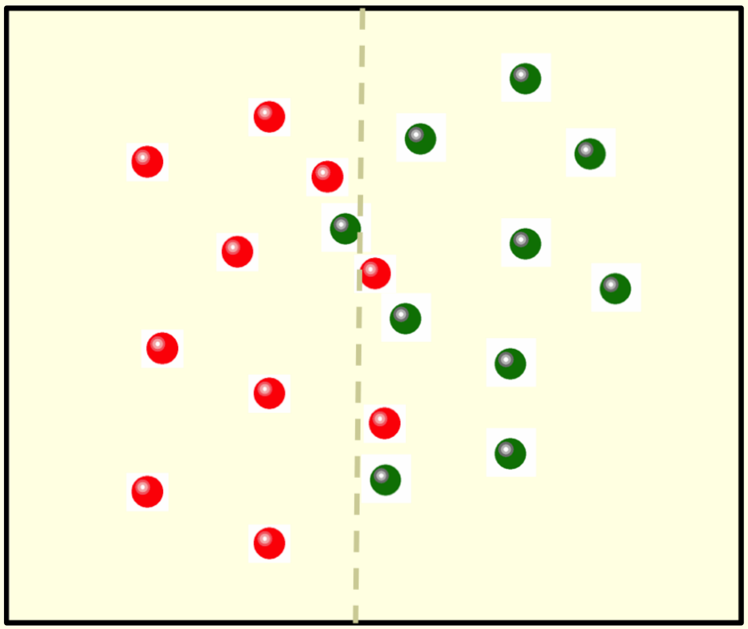
Class modelling techniques build a model for each class studied and then evaluate the fitting of all objects to each model. For this reason, for any given object there is the possibility of assignment to more than one class or to none of the classes studied.



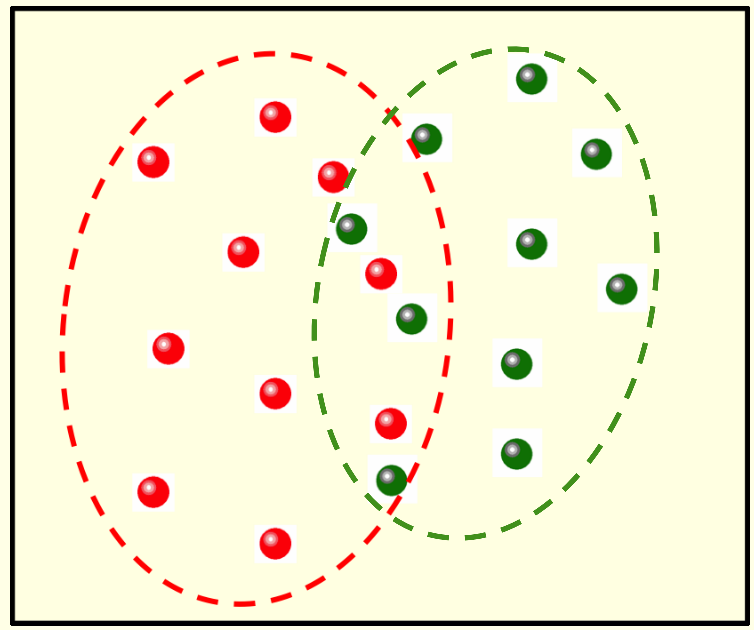
(*P. Oliveri, University of Genoa*)

Real World Situations

Classification



Class Modelling



(P. Oliveri, University of Genoa)

How Do We Assess Model Performance?

Classification Techniques

For each class:

$$\%\text{correct classifications} = \frac{\text{objects of training set correctly classified}}{\text{total objects of training set}} \cdot 100\%$$

$$\%\text{correct predictions} = \frac{\text{objects of validation set correctly classified}}{\text{total objects of validation set}} \cdot 100\%$$

(*P. Oliveri, University of Genoa*)

How Do We Assess Model Performance?

Class Modelling

For each class:

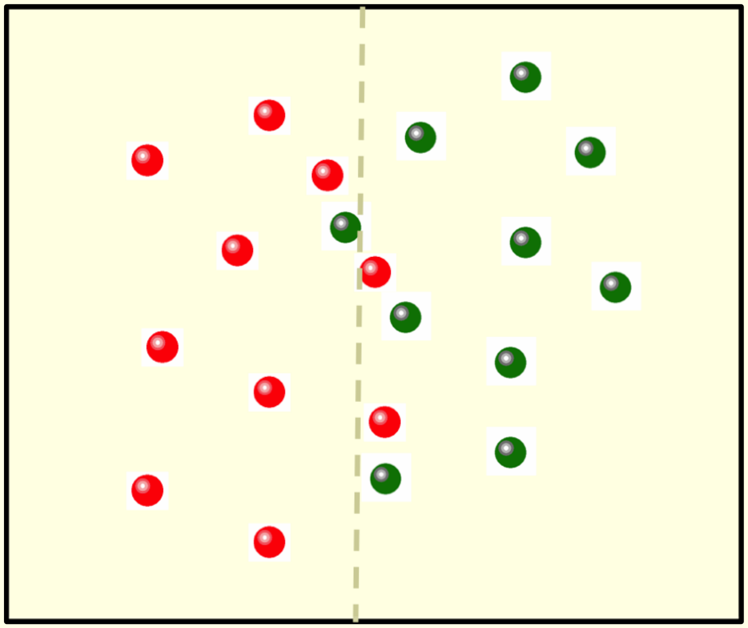
$$\%\text{sensitivity} = \frac{\text{objects of the modelled class correctly accepted by the model}}{\text{total objects of the modelled class}} \cdot 100\%$$

$$\%\text{specificity} = \frac{\text{objects, extraneous to the modelled class, correctly refused by the model}}{\text{total objects extraneous to the modelled class}} \cdot 100\%$$

(*P. Oliveri, University of Genoa*)

Incomplete Separation

Classification



RED

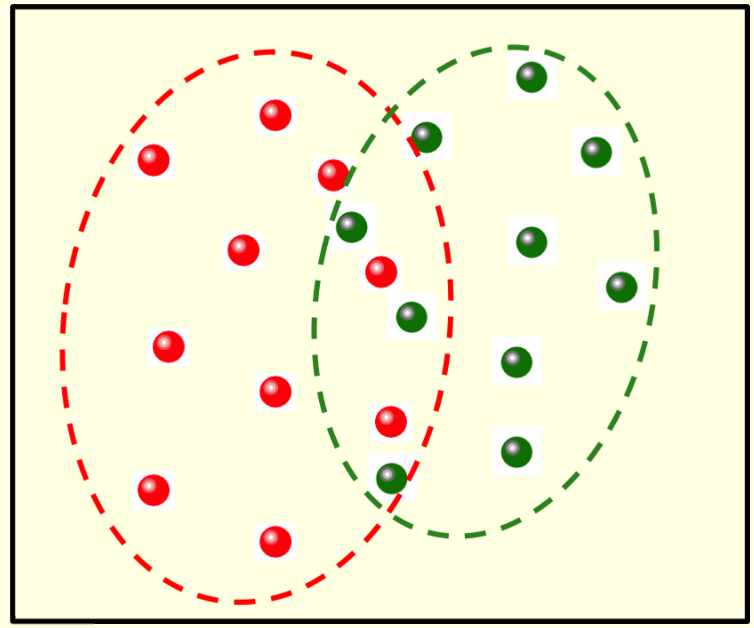
$$\%\text{correct classifications} \Rightarrow \frac{\text{objects correctly classified}}{\text{total objects}} \Rightarrow \frac{8}{10} \Rightarrow 80\%$$

GREEN

$$\%\text{correct classifications} \Rightarrow \frac{\text{objects correctly classified}}{\text{total objects}} \Rightarrow \frac{9}{10} \Rightarrow 90\%$$

(*P. Oliveri, University of Genoa*)

Class Modelling



RED

$$\%\text{sensitivity} \Rightarrow \frac{\text{objects correctly accepted}}{\text{total objects}} \Rightarrow \frac{10}{10} \Rightarrow 100\%$$

$$\left(\%\text{specificity} \Rightarrow \frac{\text{objects correctly refused}}{\text{total objects extraneous}} \Rightarrow \frac{7}{10} \Rightarrow 70\%\right)$$

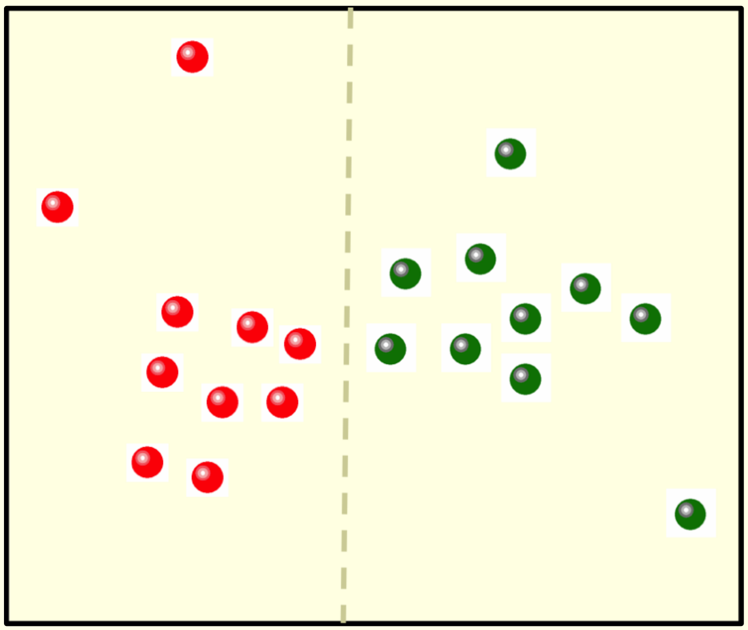
GREEN

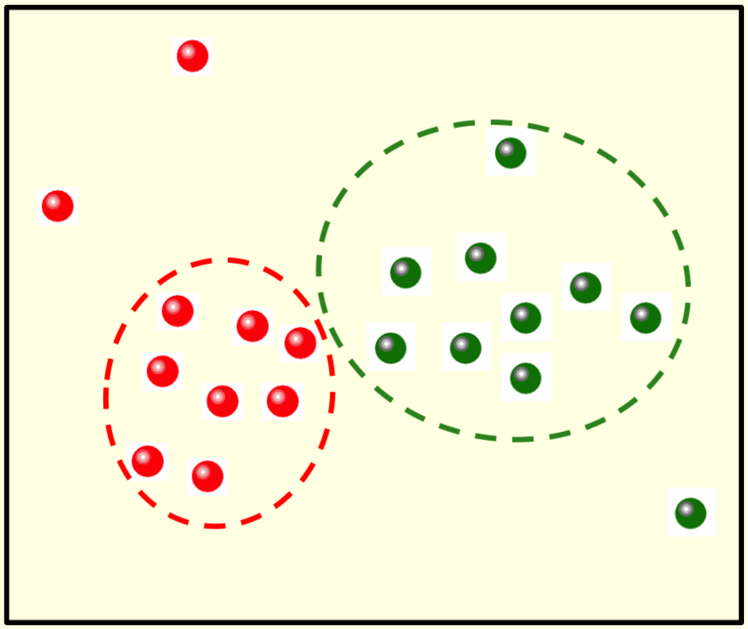
$$\%\text{sensitivity} \Rightarrow \frac{\text{objects correctly accepted}}{\text{total objects}} \Rightarrow \frac{10}{10} \Rightarrow 100\%$$

$$\left(\%\text{specificity} \Rightarrow \frac{\text{objects correctly refused}}{\text{total objects extraneous}} \Rightarrow \frac{8}{10} \Rightarrow 80\%\right)$$

Unusual Samples

Classification	Class modelling





RED

$$\%\text{correct classifications} \Rightarrow \frac{\text{objects correctly classified}}{\text{total objects}} \Rightarrow \frac{10}{10} \Rightarrow 100\%$$

GREEN

$$\%\text{correct classifications} \Rightarrow \frac{\text{objects correctly classified}}{\text{total objects}} \Rightarrow \frac{10}{10} \Rightarrow 100\%$$

RED

$$\%\text{sensitivity} \Rightarrow \frac{\text{objects correctly accepted}}{\text{total objects}} \Rightarrow \frac{8}{10} \Rightarrow 80\%$$

$$\left(\%\text{specificity} \Rightarrow \frac{\text{objects correctly refused}}{\text{total objects extraneous}} \Rightarrow \frac{10}{10} \Rightarrow 100\%\right)$$

GREEN

$$\%\text{sensitivity} \Rightarrow \frac{\text{objects correctly accepted}}{\text{total objects}} \Rightarrow \frac{9}{10} \Rightarrow 90\%$$

$$\left(\%\text{specificity} \Rightarrow \frac{\text{objects correctly refused}}{\text{total objects extraneous}} \Rightarrow \frac{10}{10} \Rightarrow 100\%\right)$$

(*P. Oliveri, University of Genoa*)

Implications

Only if EACH object in class modelling is accepted by ONE and ONLY ONE class model

will

% CORRECT CLASSIFICATION and SENSITIVITY have the same value.

In all other cases, they will differ.

(*P. Oliveri, University of Genoa*)

Practical Considerations and Risk

- Ideally, we achieve 100% correct classification and sensitivity but usually we do not.
- What constitutes the greater risk – incorrect classification of true product or incorrect classification of untrue product?
- How do we adjust class boundary limits to address this issue?

Worked Example

- Trappist Rochefort 8° beer from Belgium: brand identity claim

AUTHENTIC
TRAPPIST
PRODUCT

Issue outline

- Trappist beers originally brewed only by Trappist monks
- Only they are permitted to use a Trappist logo as a mark of authenticity
- Rochefort is a specific beer brand; available as Rochefort 6°, 8° and 10°
- Can we use fingerprint and profiling methods to confirm the identity of a beer which claims to be Rochefort 8°?

Experimental plan

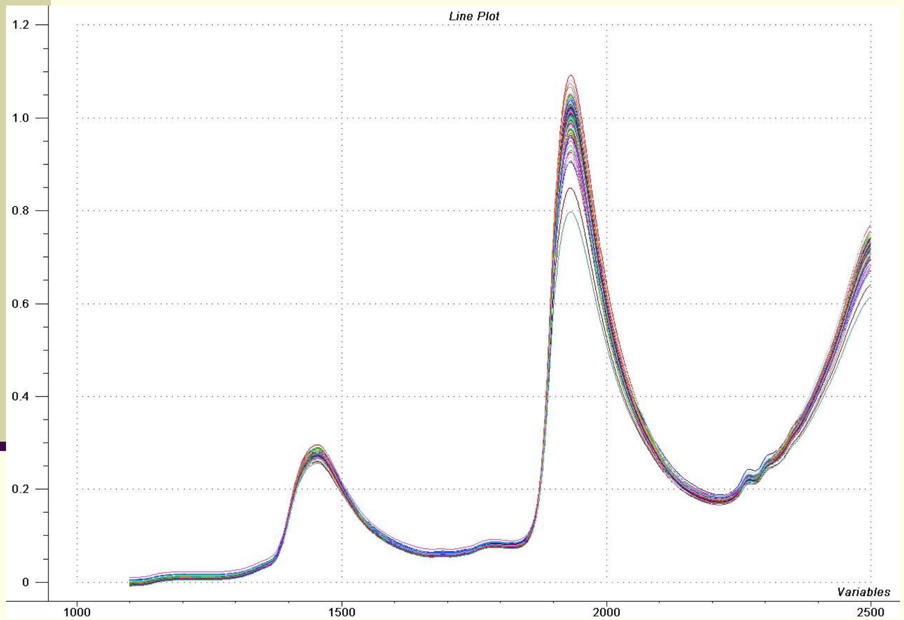
- Collect samples of Trappist and non-Trappist beers from different production batches
- Samples collected, coded, assembled into sets for each laboratory (5), and distributed by courier
- Beers distributed in two lots – 1 in autumn 2008 and 1 just before Christmas.

Beer numbers

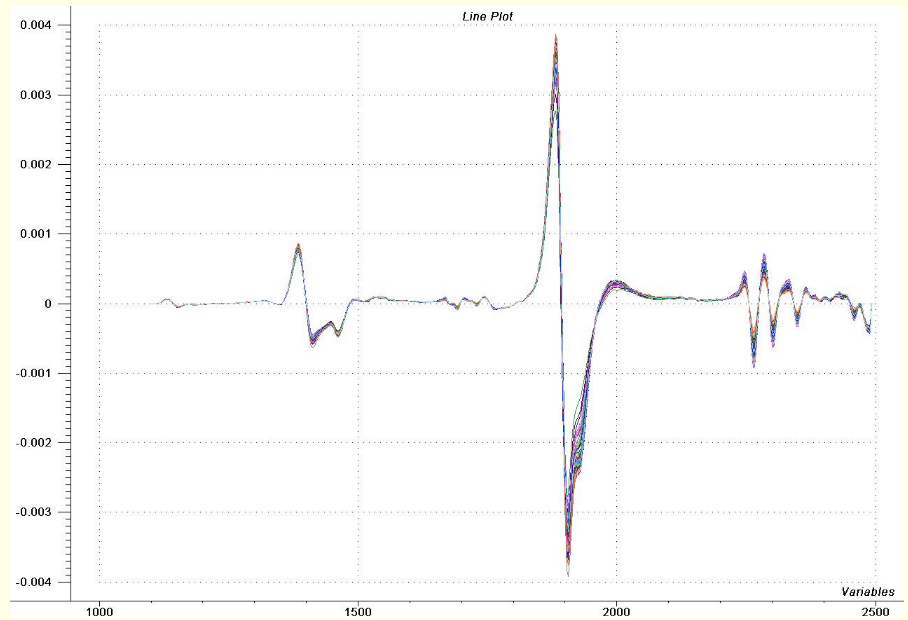
		1st study Month 0	Month 6nd study	3rd study Month 12	
Trappist beer	Rochefort 8°	**16**	**32**	**16**	
	Other trappist around 8° (Chimay triple, Archel brune, Westmalle, Westvleteren, Trappe)) + Other trappist (Rochefort 10°, Orval, Chimay dorée...)	**37**	**0**	**37**	
Other Beers	"special" beers but not trappist (Leffe, Grimbergen, gueuze, Jupiler,…)	**67**	**0**	**67**	
					Total bottles/ team:
	Total	**120**	**32**	**120**	**272**

Total number of samples sourced: 1165!

Infrared methods – NIR

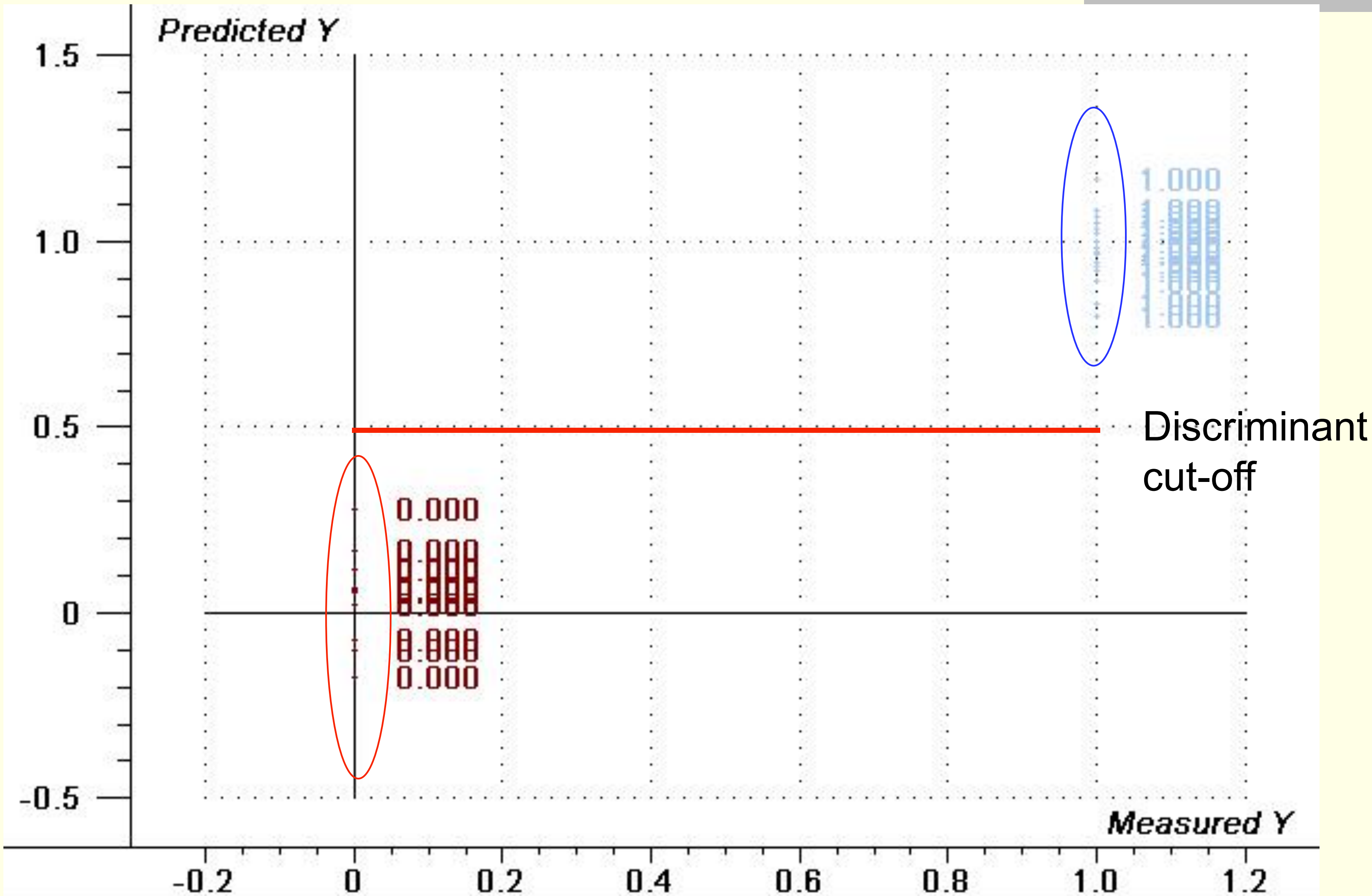


Raw, transflectance spectra

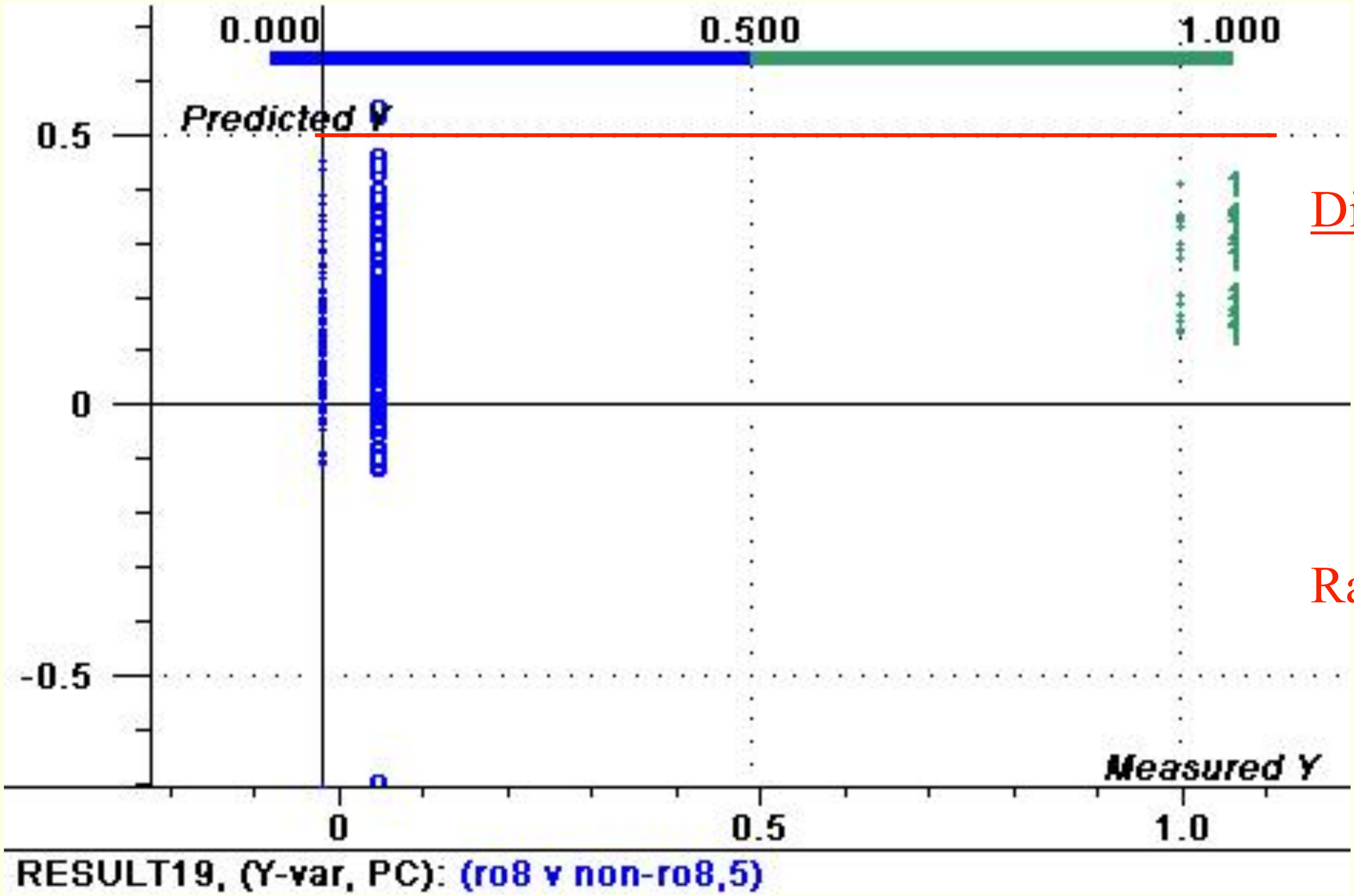


2nd derivative transflectance spectra

Discriminant Partial Least Squares Regression (PLS-DA) - ideal



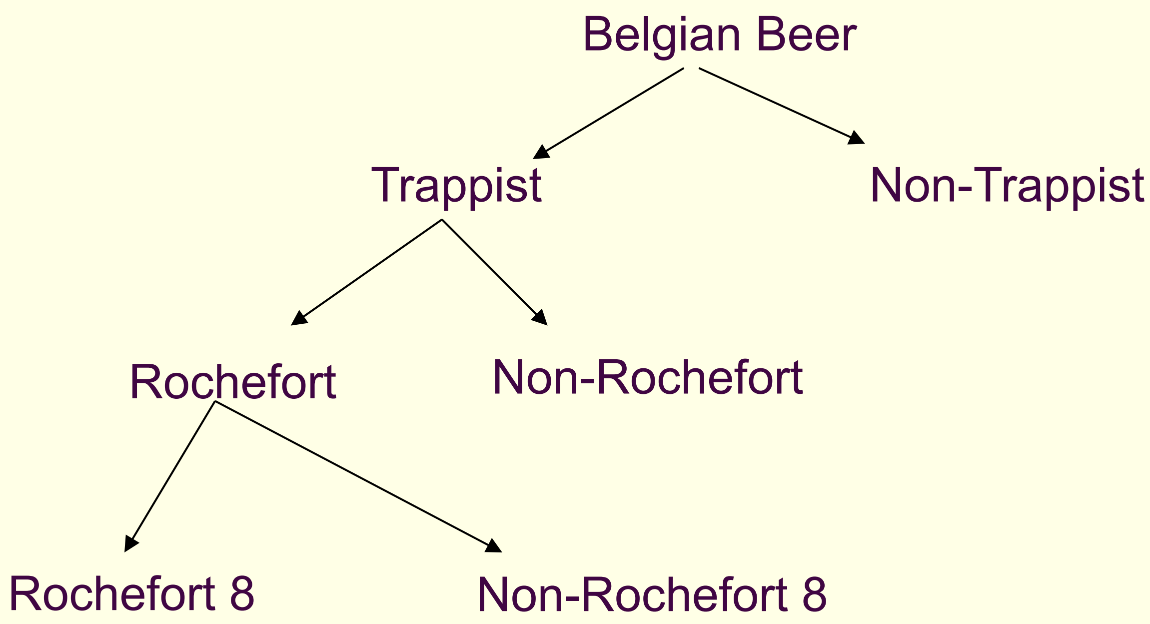
Discriminant Analysis - real
Rochefort 8 *vs* non-Rochefort 8



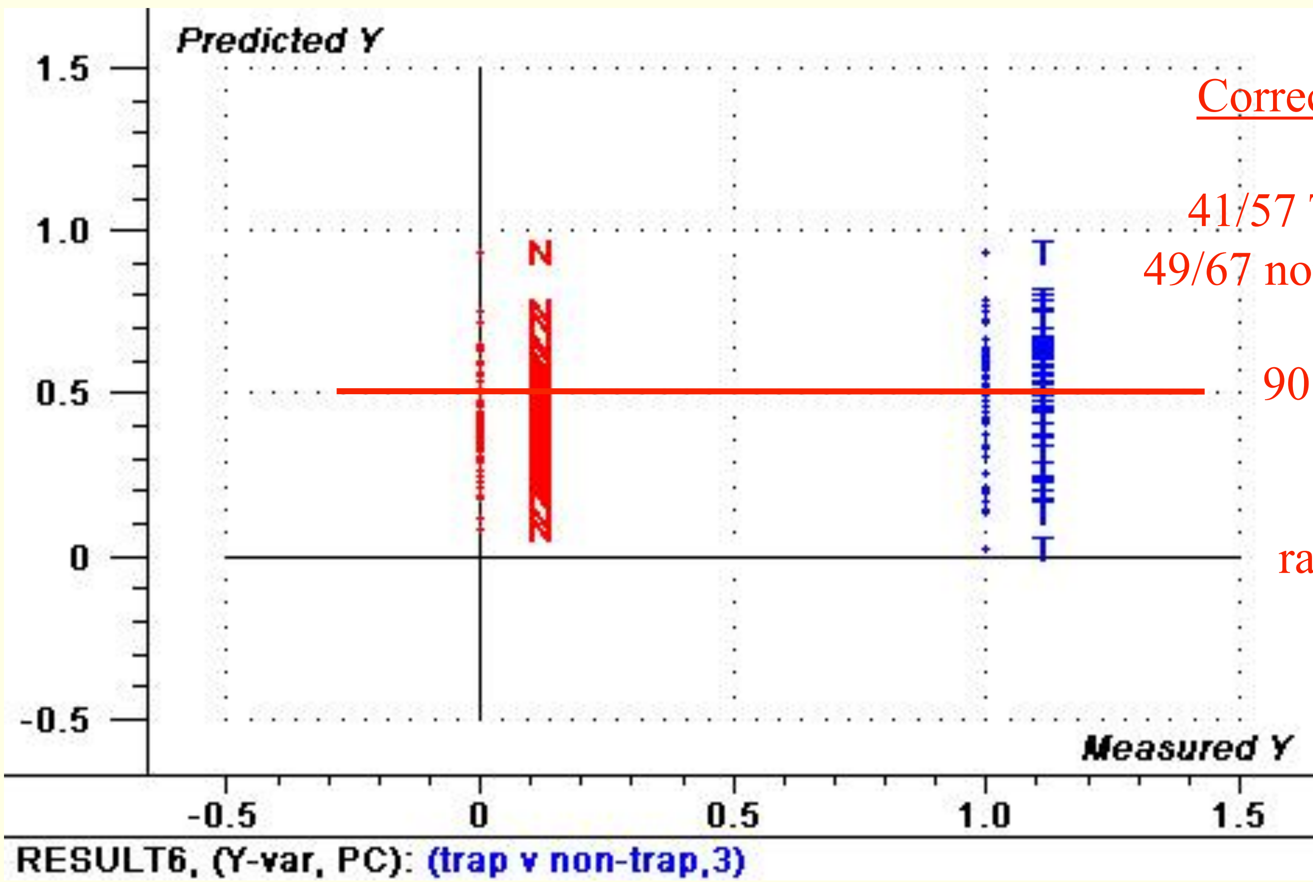
Discriminant PLS

Raw spectral data

NIR Data Analysis - hierarchical



NIR data analysis
Trappist *vs* non-Trappist



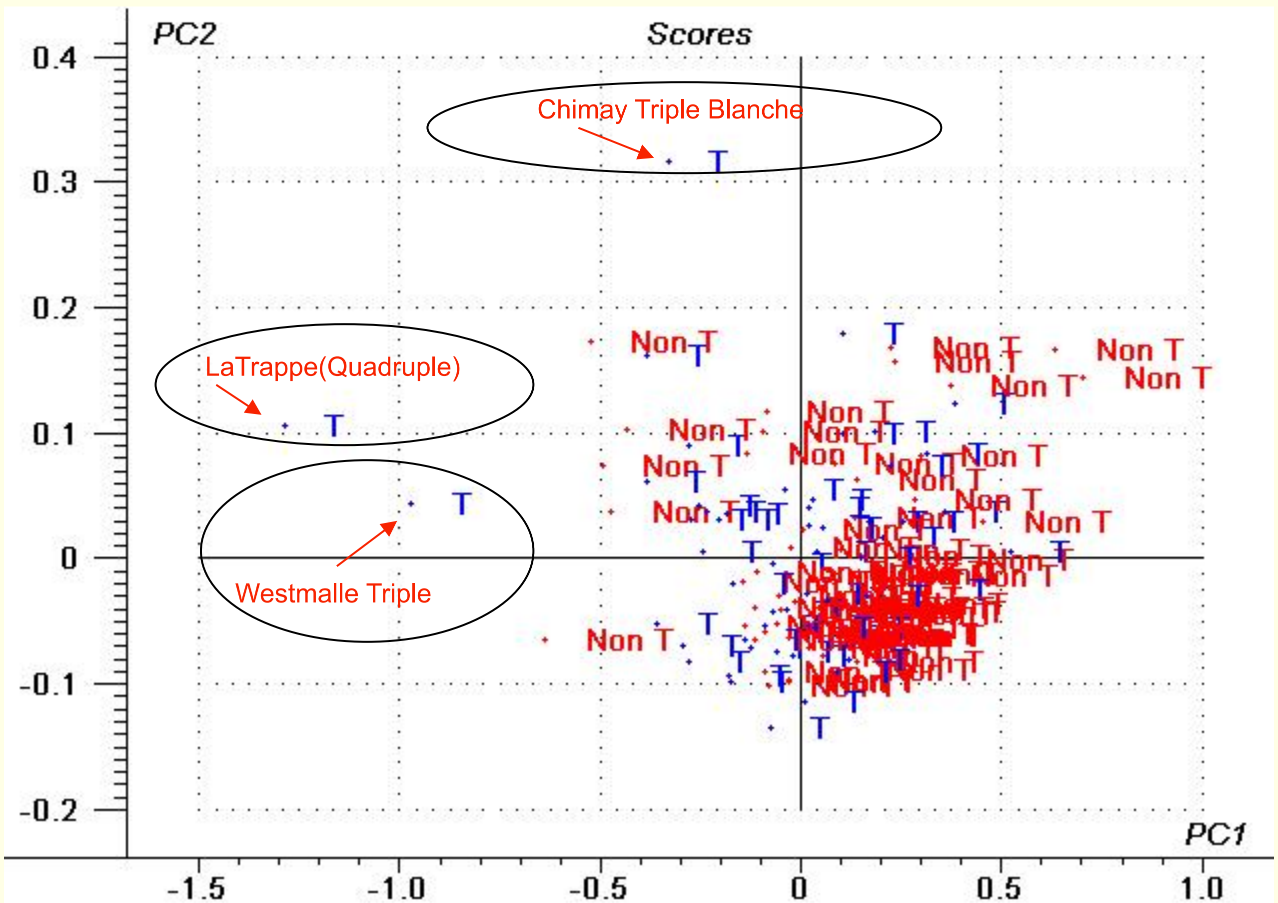
Correct classifications

41/57 Trappist (71.9%)
49/67 non-Trappist (73.1%)

90 of 124 beers
(72.6%)

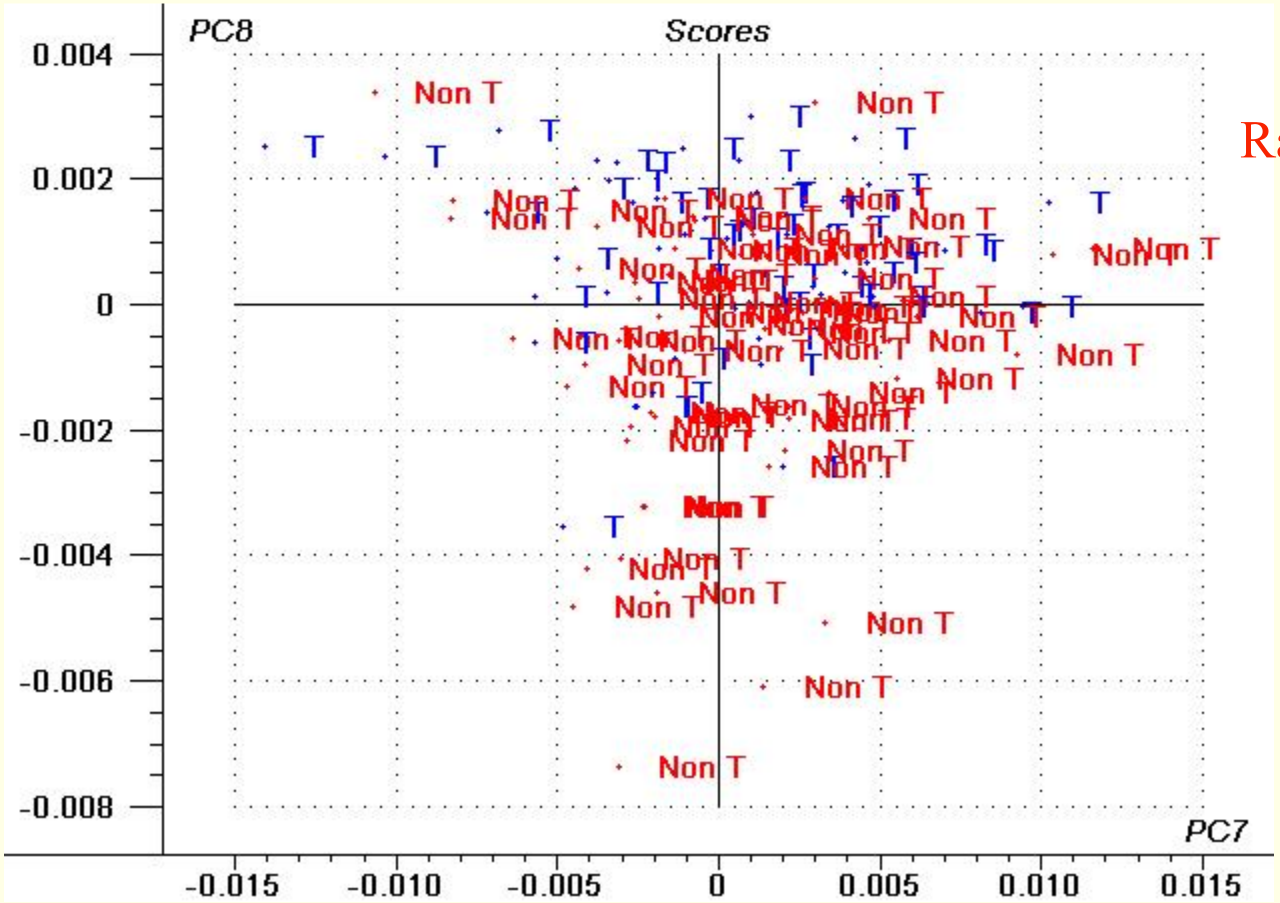
raw spectral data

NIR data analysis
Trappist *vs* non-Trappist



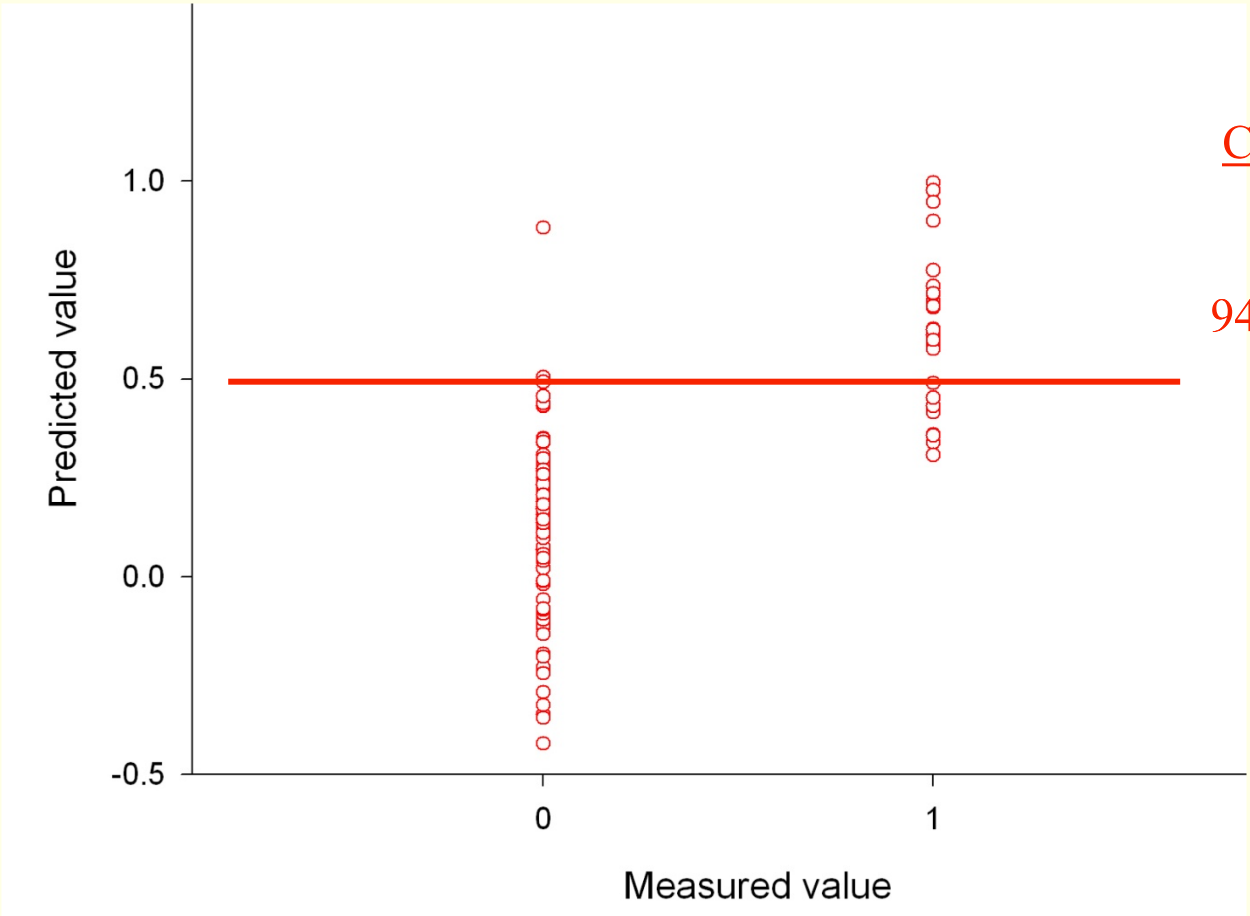
Raw spectral data

NIR data analysis
Trappist *vs* non-Trappist



Raw spectral data

NIR data analysis
Rochefort *vs* non-Rochefort



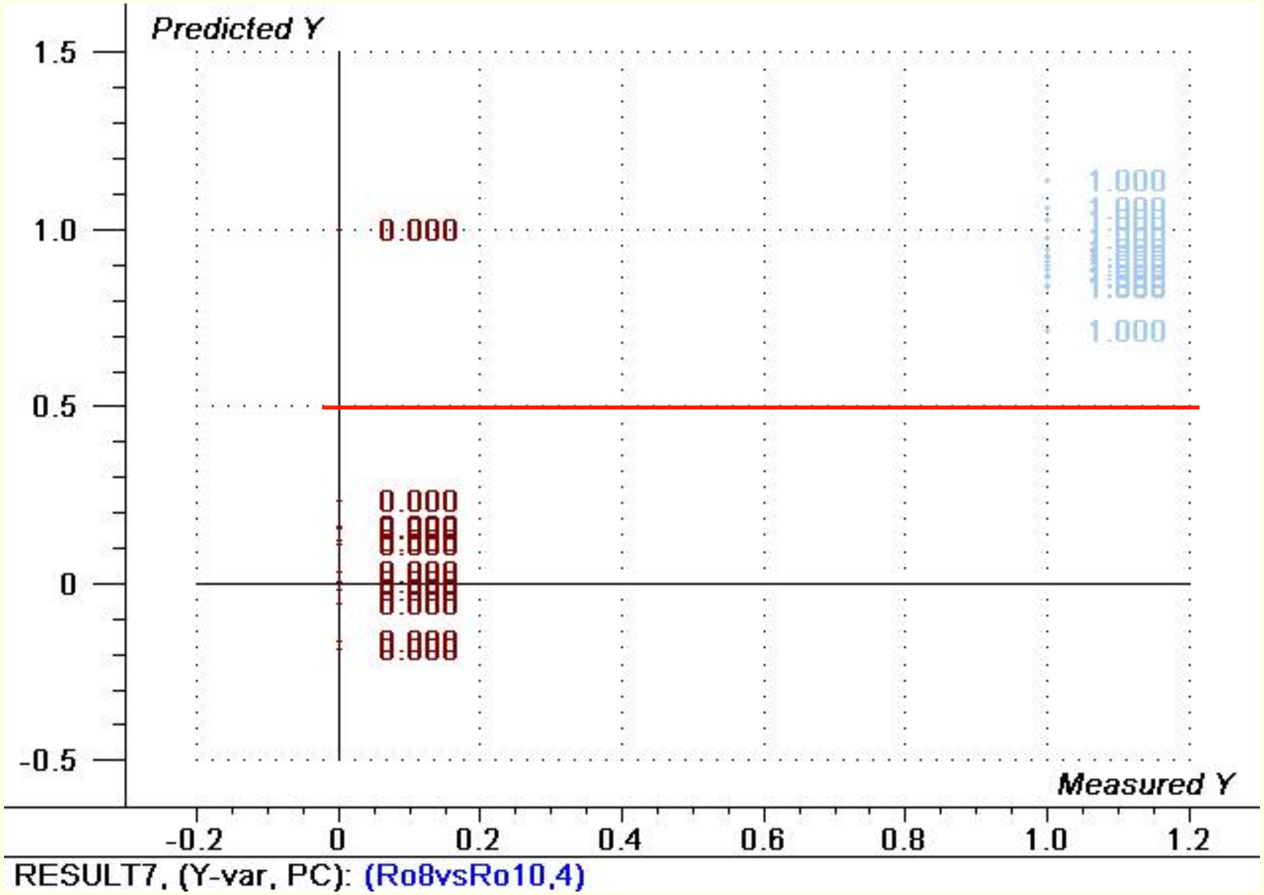
Correct classifications

22 of 28 Rochefort
94 of 96 non-Rochefort

116 of 124 beers
(93.7%)

raw spectral data

NIR data analysis
Rochefort8 *vs* Rochefort 10



Raw spectral data

Summary of Results from TRACE Workpackage 2

		% correct classification		
Technique	**Actual**	**Rochefort 8**	**Non-Rochefort 8**	**Average**
NIR	R8	-	-	-
	Non-R8	-	-	
FTIR	R8	-	-	-
	Non-R8	-	-	
Raman	R8	90.6	9.4	**87.2**
	Non-R8	16.2	83.8	
NMR(CSL)	R8	94.6	5.4	**87.2**
	Non-R8	20.6	79.4	
NMR(CNR)	R8	100	0	**90.5**
	Non-R8	19.0	81.0	
UPLC-QTOF	R8	100	0	**98.1**
	Non-R8	3.9	96.1	

IR spectroscopy results

		% correct classification		
Technique	**Actual**	**Rochefort**	**Non-Rochefort**	**Average**
NIR	Rochefort	78.6	21.4	**88.3**
	Non-Rochefort	2.1	97.9	
FTIR	Rochefort	89.7	10.3	**94.3**
	Non-Rochefort	4.3	95.7	
		Rochefort 8	**Rochefort 10**	
NIR	R8	100	0	**95.9**
	R10	8.3	91.7	
FTIR	R8	100	0	**100**
	R10	0	100	

Class Modelling Results
NIR + SIMCA

Data treatment	#PCs	% correct classification		Sensitivity		Specificity	
		Trappist vs non-Trappist					
		T	*Non-T*	*T*	*Non-T*	*T*	*Non-T*
Raw	8	89.1	87.9	81.8	77.3	83.8	85.5
		Rochefort vs non-Rochefort					
		R	*Non-R*	*R*	*Non-R*	*R*	*Non-R*
SNV+	7	100	100	100	100	100	100
		Rochefort 8° vs Rochefort 10°					
		R8	*R10*	*R8*	*R10*	*R8*	*R10*
SNV+	4	100	100	93.8	100	100	93.8

Class Modelling Results NIR + UNEQ

Data treatment	#PCs	% correct classification		Sensitivity		Specificity	
		Trappist vs non-Trappist					
		T	*Non-T*	*T*	*Non-T*	*T*	*Non-T*
MSC	10	89.1	84.9	91.8	88.6	28.8	18.2
		Rochefort vs non-Rochefort					
		R	*Non-R*	*R*	*Non-R*	*R*	*Non-R*
SNV+	7	100	100	96.4	88.9	92.6	78.6
		Rochefort 8° vs Rochefort 10°					
		R8	*R10*	*R8*	*R10*	*R8*	*R10*
SNV+	4	100	100	100	95.8	100	100

Class Modelling Results NIR + POTFUN

Data treatment	#PCs	% correct classification		Sensitivity		Specificity	
		Trappist vs non-Trappist					
		T	*Non-T*	*T*	*Non-T*	*T*	*Non-T*
Raw+	13	70.9	77.3	98.2	92.4	7.6	9.1
		Rochefort vs non-Rochefort					
		R	*Non-R*	*R*	*Non-R*	*R*	*Non-R*
Raw+	5	92.9	96.3	92.9	88.9	51.9	10.7
		Rochefort 8° vs Rochefort 10°					
		R8	*R10*	*R8*	*R10*	*R8*	*R10*
SNV+	4	93.8	91.7	100	91.7	91.7	31.3

Summary

- Clear definition of the issue to be addressed is essential at the outset
- Comprehensive experimental design needed
- Variety of multivariate options available
- How to incorporate the element of relative risk
- How to translate mathematical result into output consumers can understand?

Acknowledgements

- Tony Woodcock, University College Dublin
- Siobhán Hennessy, University College Dublin
- Paolo Oliveri, University of Genoa
- Valentina Di Egidio, University of Milan
- EU Commission for funding
- FERA/JIFSAN for invitation