



**University of  
Nottingham**

UK | CHINA | MALAYSIA

# **Analysis of food composition by hyperspectral imaging: examples on cocoa beans and coffee beans**

**Nicola (Nick) Caporaso**

[nicola.caporaso1@Nottingham.ac.uk](mailto:nicola.caporaso1@Nottingham.ac.uk)





# Summary

- **Introduction to hyperspectral imaging**
- **Data acquisition and processing**
- **Findings and results for the food commodities investigated:**
  - Cocoa
  - Coffee
- **Conclusions**





University of  
Nottingham

UK | CHINA | MALAYSIA

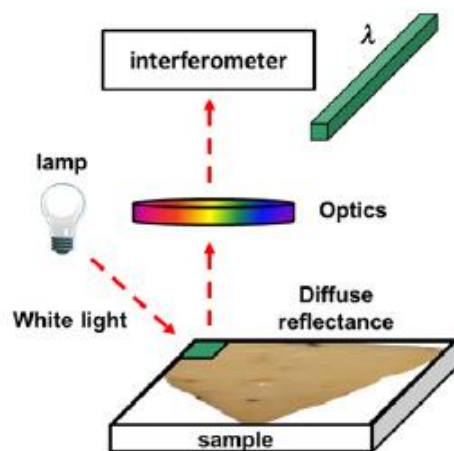
# Hyperspectral imaging

Introduction

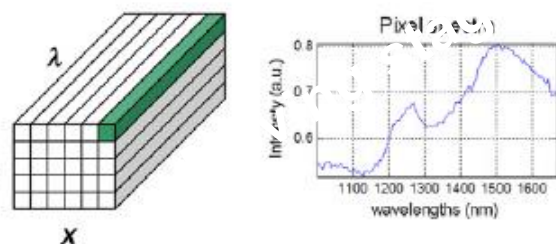


# What is hyperspectral imaging

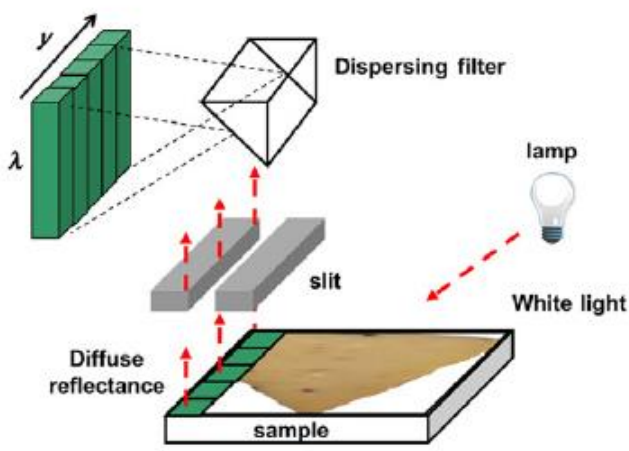
Hyperspectral imaging = spectroscopy (visible or NIR) & image analysis/processing



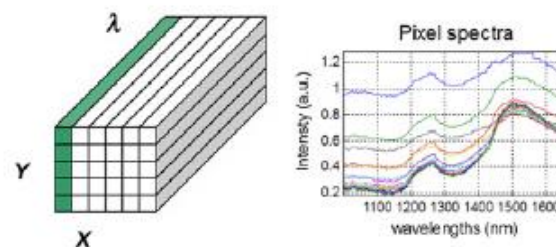
Single spectrum in each displacement



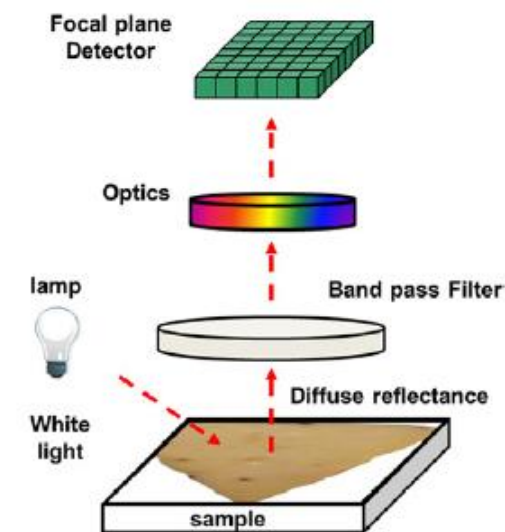
Point scan



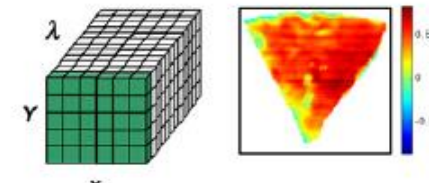
A line of spectra in each displacement



Line scan



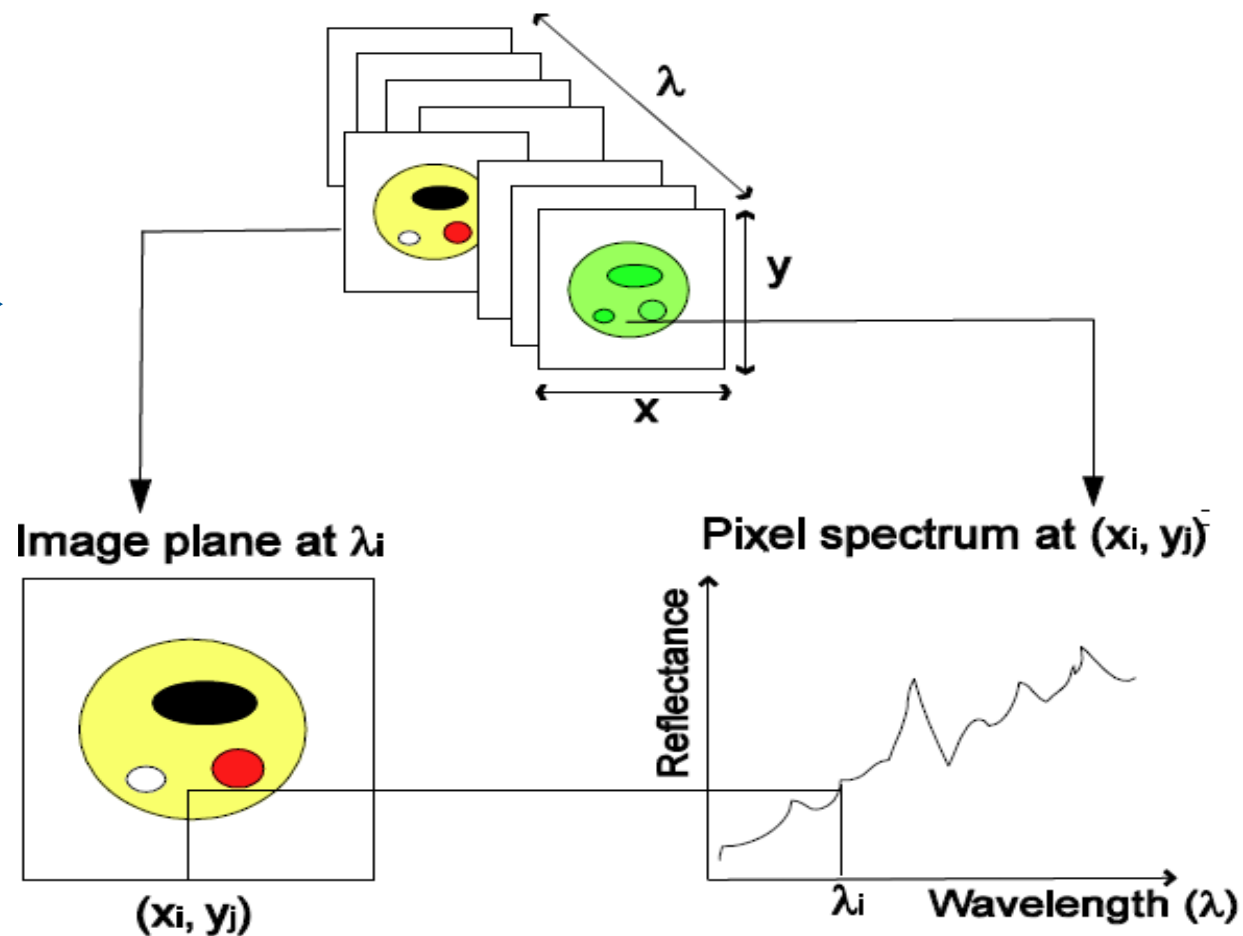
One false color image for each wavenumber



Plane scan

# What is hyperspectral imaging

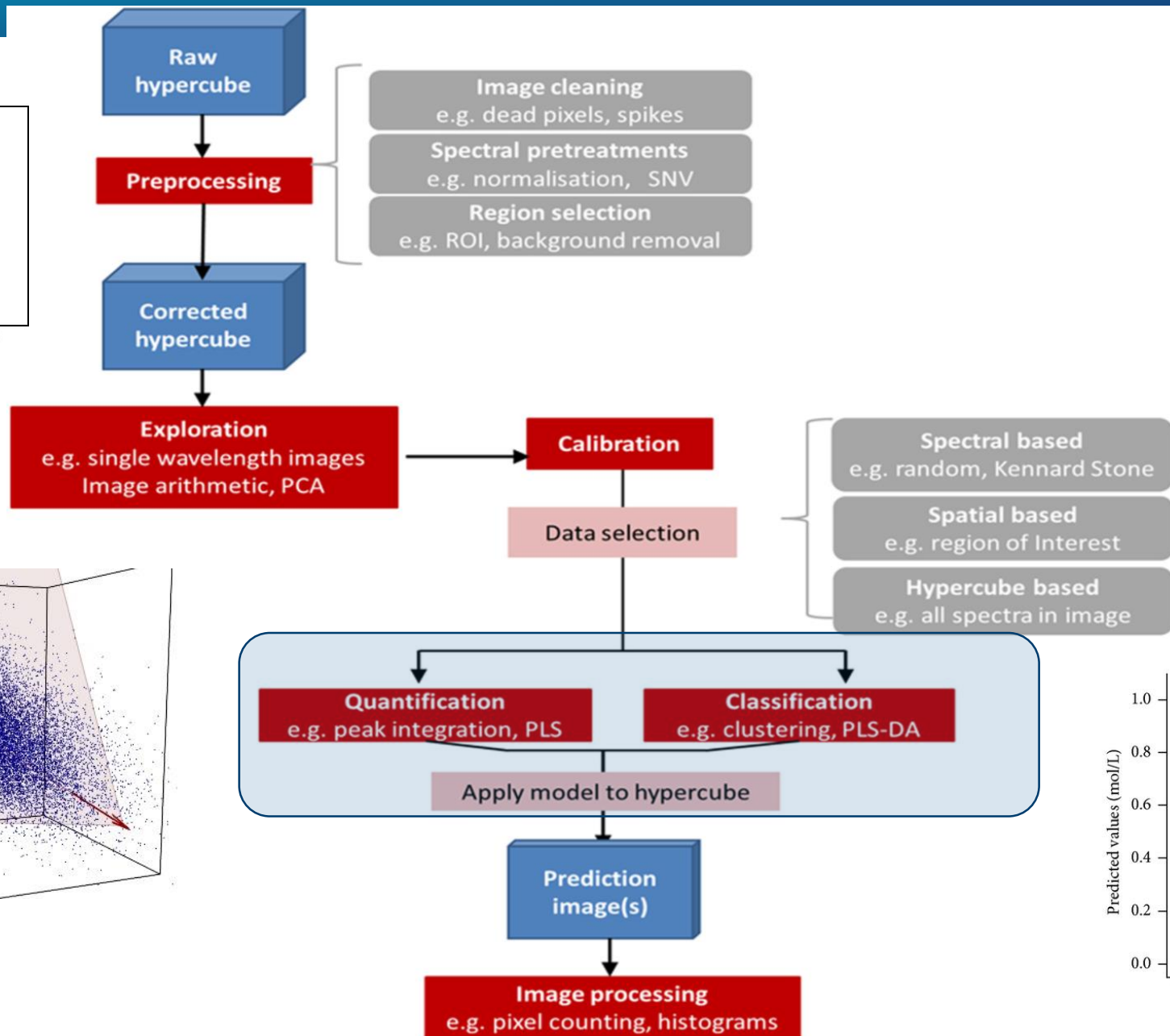
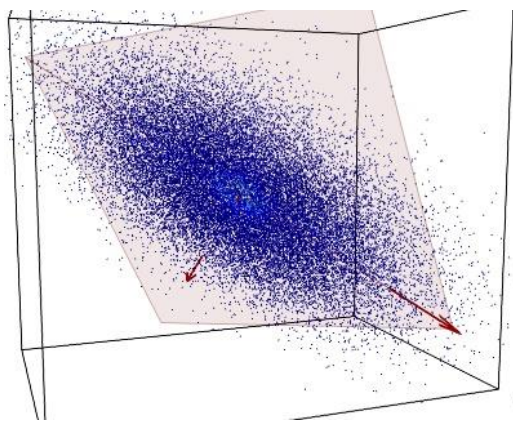
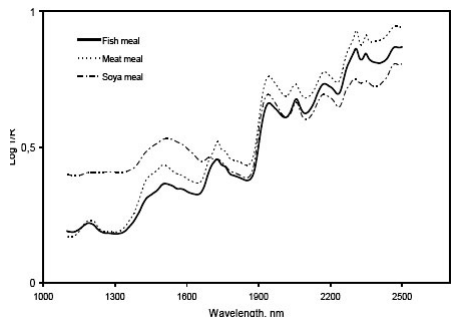
- Spatial information
- Hypercube
- Huge amount of data
- Useful to link the chemical information from NIR to the imaging analysis



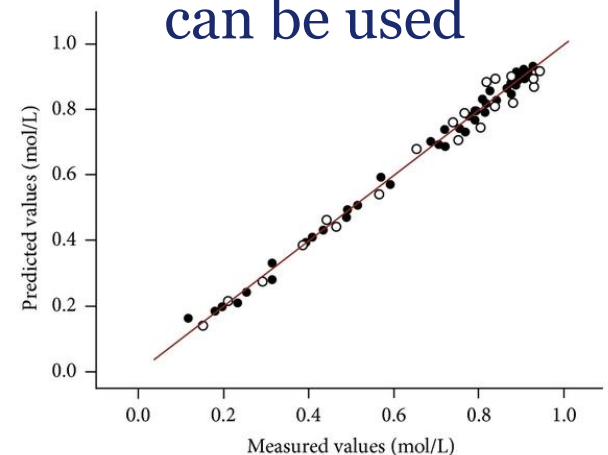
(from Gowen et al., 2007)



# Data processing for Hyperspectral Imaging



- Similarly to NIR spectroscopy, HSI quantification methods require reference measurements to build future predictions
- For classification, either unsupervised or supervised methods can be used



# Specifics of the HSI instrument used

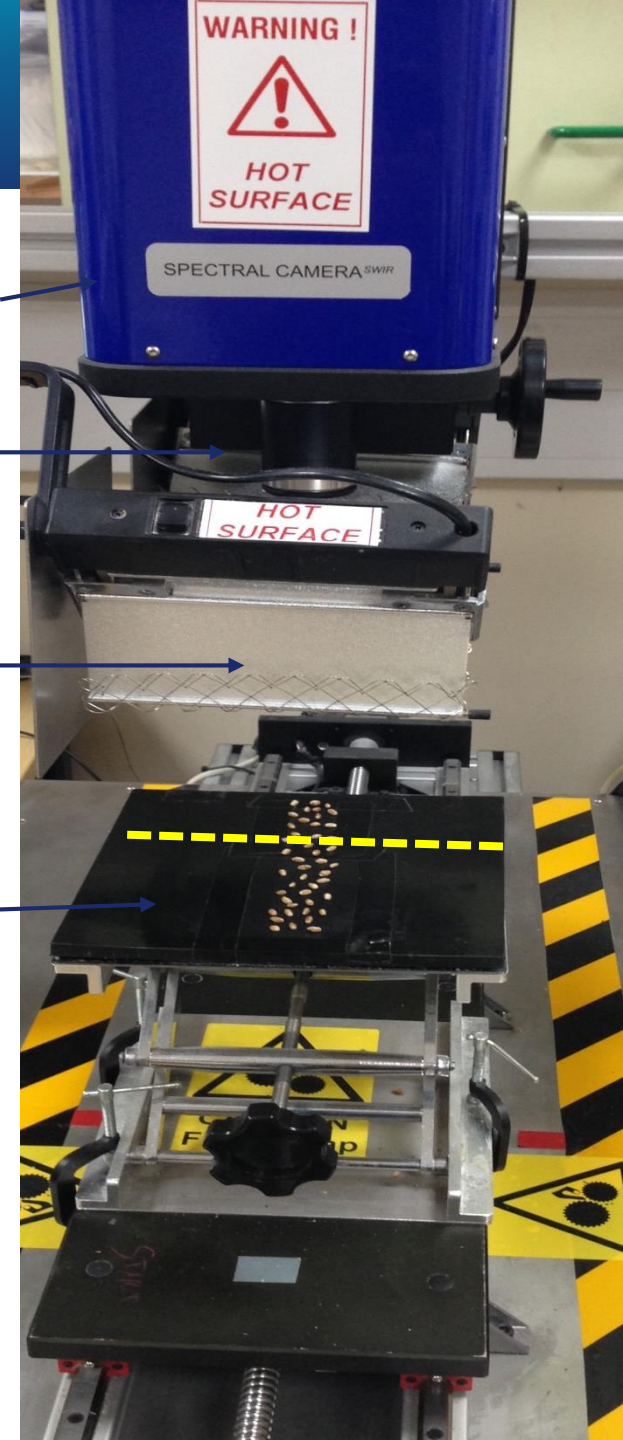
- Push-broom approach
- Specim SWIR spectral camera
  - HgCdTe detector
  - 256 bands (~**970-2500 nm**)
  - Spectral resolution ~6 nm
- Acquisition parameters:
  - Field of view **35 mm** width
  - Scan speed: **10.9 mm/s**
  - resolution 0.1 mm pixel<sup>-1</sup>

Camera

Lens

Light  
source

Moving  
stage



# Image acquisition conditions



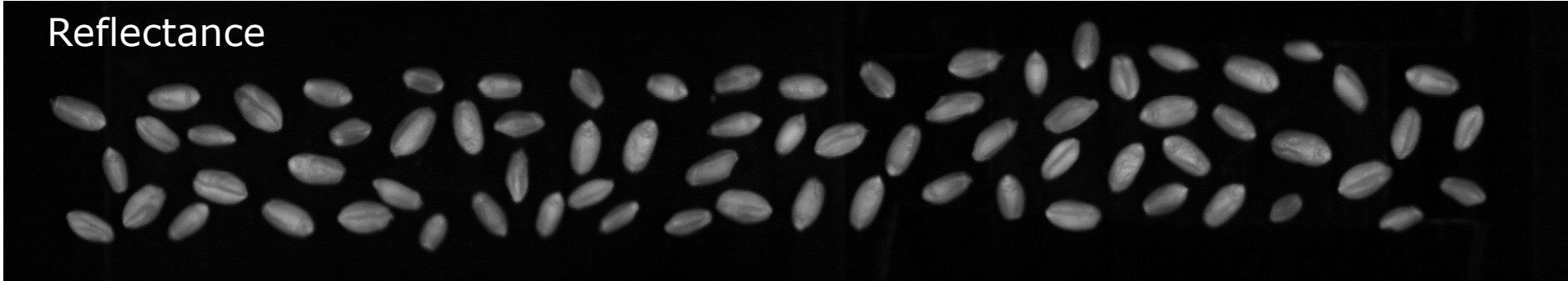
**Specim Camera,  
HgCdTe sensor (1000-2500 nm)**

- Setting up the best spectral acquisition conditions:
  - ▣ Field of view
  - ▣ Camera focusing
  - ▣ Spectral profile (exposure time -> avoid saturation)
  - ▣ Scan speed and camera frequency
  - ▣ 500 W incandescent lamps
- White reference (Teflon)
- Dark current removal
- Image processing (ENVI+IDL)

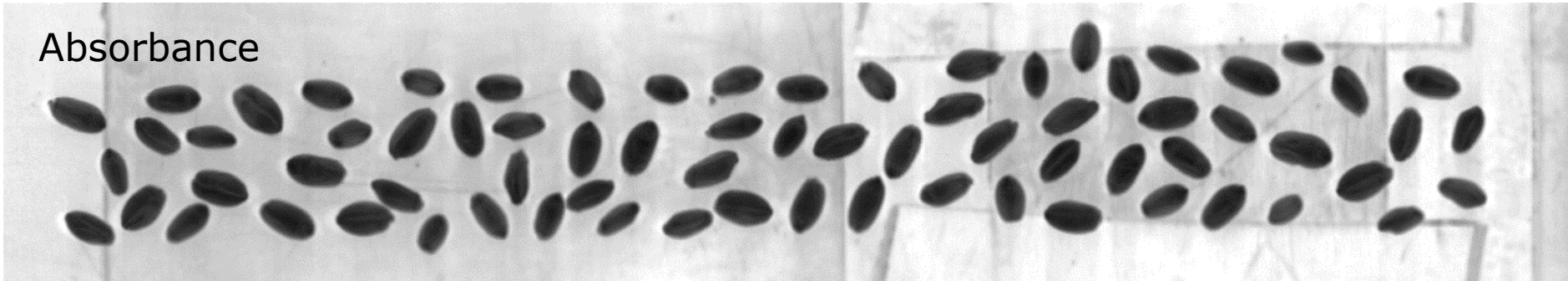


## Normalisation functions

Reflectance



Absorbance



Segmentation

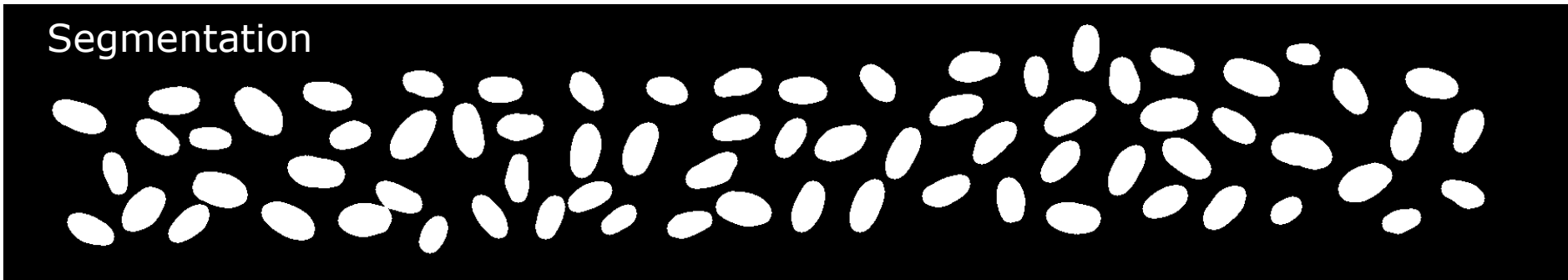
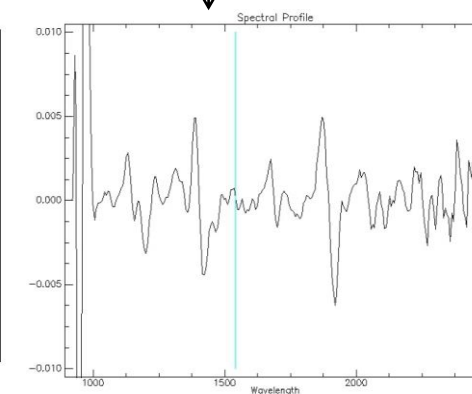
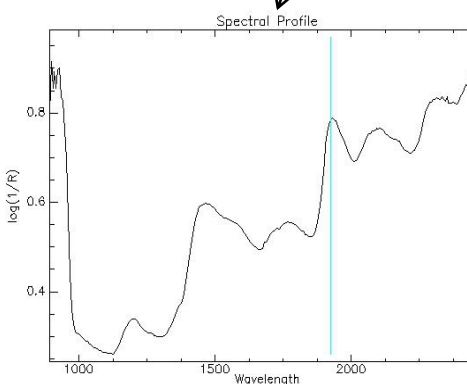
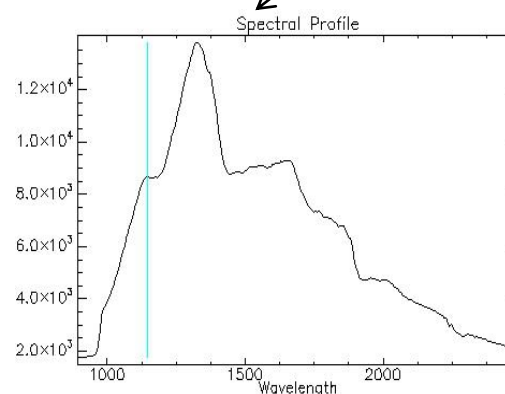
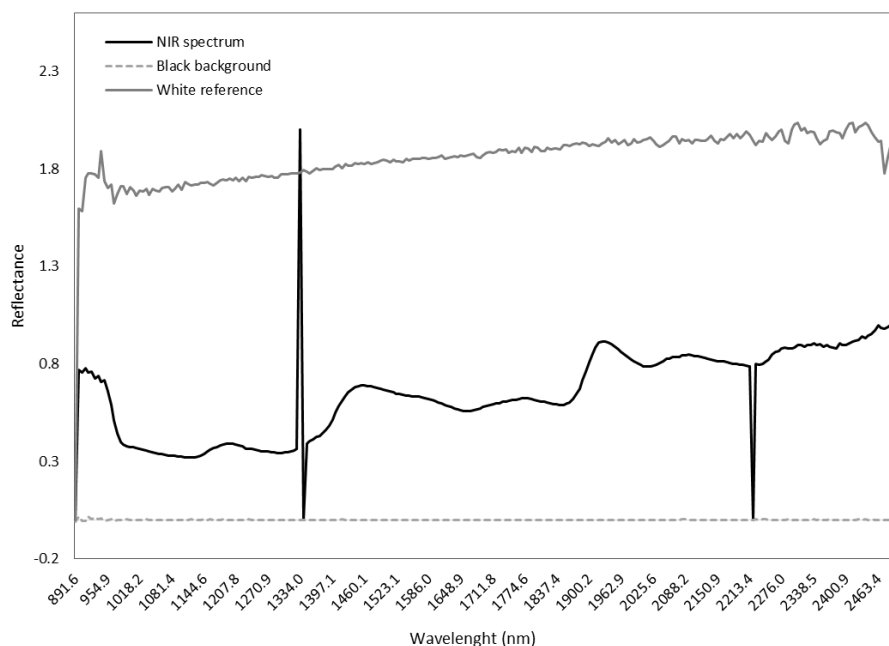
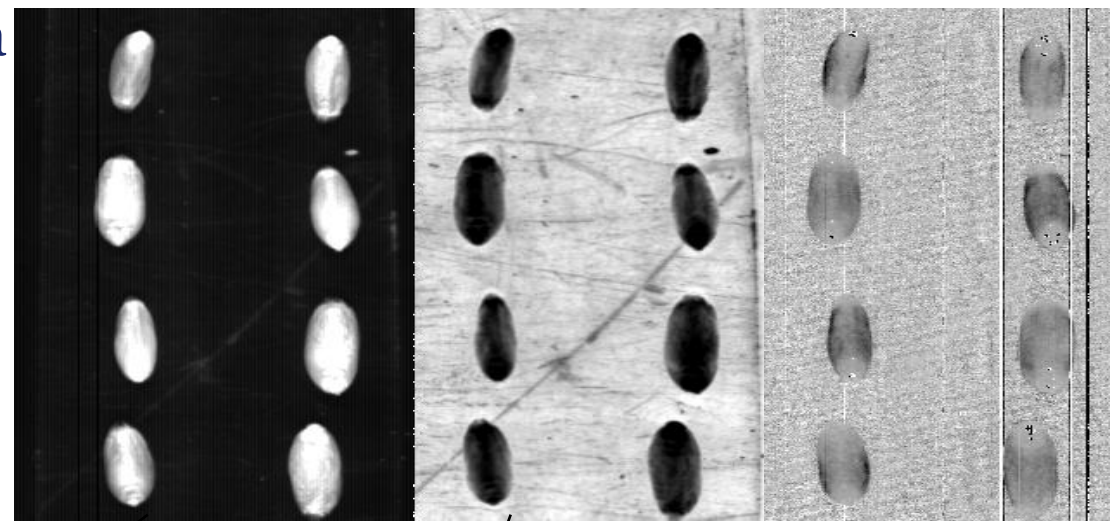


Image cleaning (small objects and fill holes function)

# Treatment of Hyperspectral imaging data

- Reflectance spectra
- Spikes removal
- Background
- White/black stage
- White reference
- Noise removal
- Raw reflectance data
- Absorbance  $\rightarrow \log(1/R)$
- Derivatives







University of  
**Nottingham**

UK | CHINA | MALAYSIA

# HSI for cocoa analysis

Applications



# Cocoa quality and composition

## Fresh cocoa beans:

high moisture content (~30%)  
high polyphenol content  
high antioxidant activity  
lipids ~30%

## Fermented cocoa beans:

moisture 7-8%  
phenolic content 100-150 mg/g  
lower antioxidant activity  
lipids ~50%



Cocoa pod

Fermentation

Cocoa beans  
ready to use





# Quality parameters investigated in this study

## Fermentation index:

- Usually done by visual assessment
- Official method requires methanol
- Overnight extraction and spectroph.
- Ratio 460/530 nm

## Polyphenol content:

- A complex variety of phenolic compounds
- Intense decrease over fermentation
- Important effects on sensory properties
- Impact on consumer's health

## Antioxidant activity:

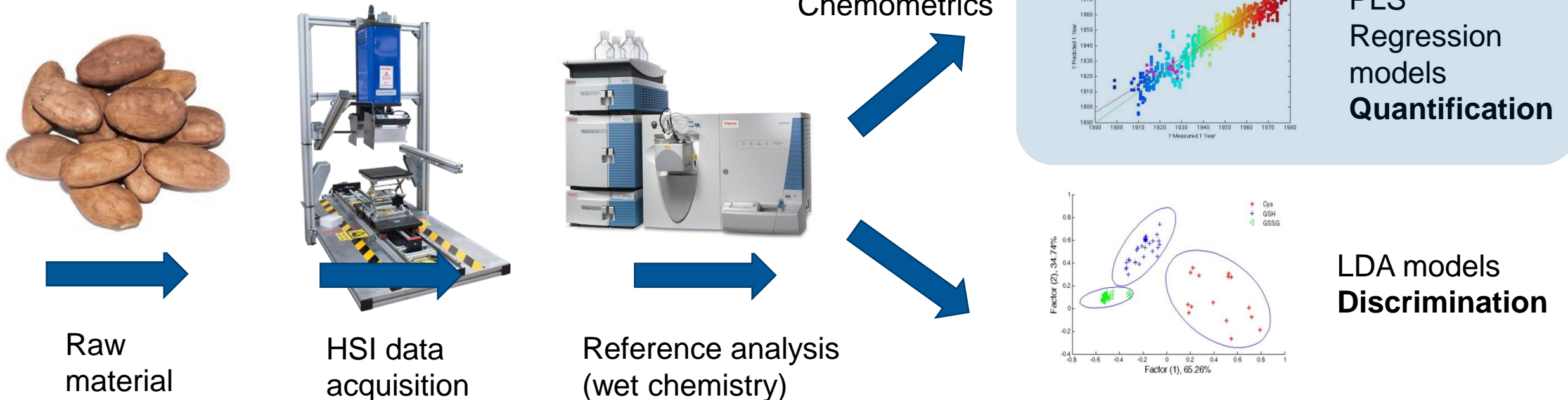
- Impact on consumer's health





# Method used for cocoa analysis

- Randomly sampling 10 cocoa beans from each batch (n=170)
- Shell manually removed
- **Whole cocoa nibs** were scanned by HSI (1000-2500 nm)
- Reference measurement (manual grinding, extraction/analysis)
- PLS regression on single cocoa nibs
- Single-pixel application (“chemical imaging”)







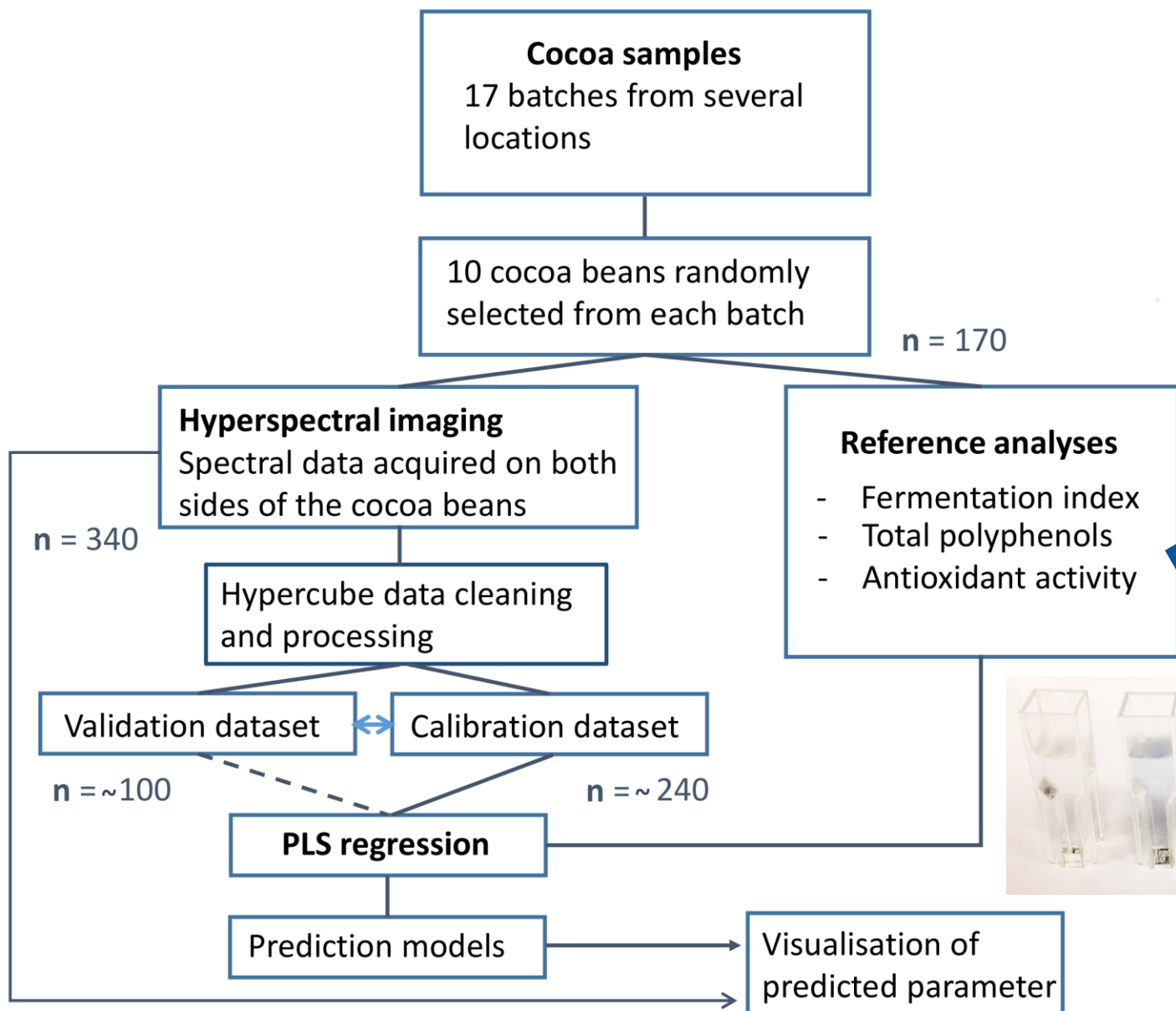
# Prediction of cocoa composition and quality

**Samples: geographical origin**

Number of batches: **17**  
10 cocoa nibs per batch

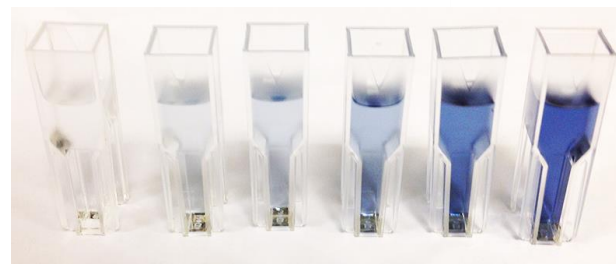


# Reference measurement: fermentation index, polyphenols, a. activity



## Result of reference measurement:

Parameter	Calibration (n=119)			
	Min	Max	Mean	SD
Fermentation index	0.38	2.13	0.98	0.34
Total polyphenols	2.87	204.46	68.41	54.47
Antioxidant activity	13.75	462.71	186.13	116.13

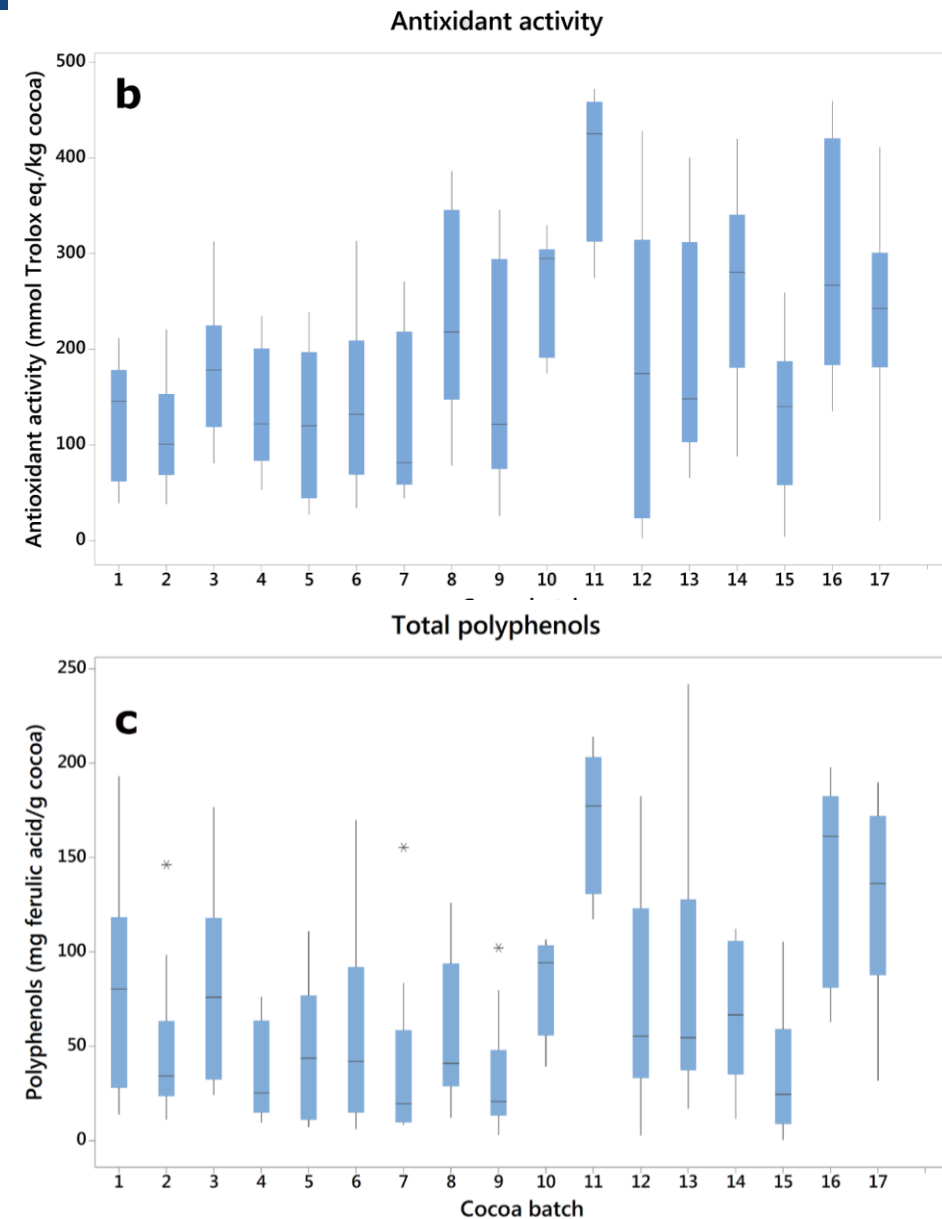
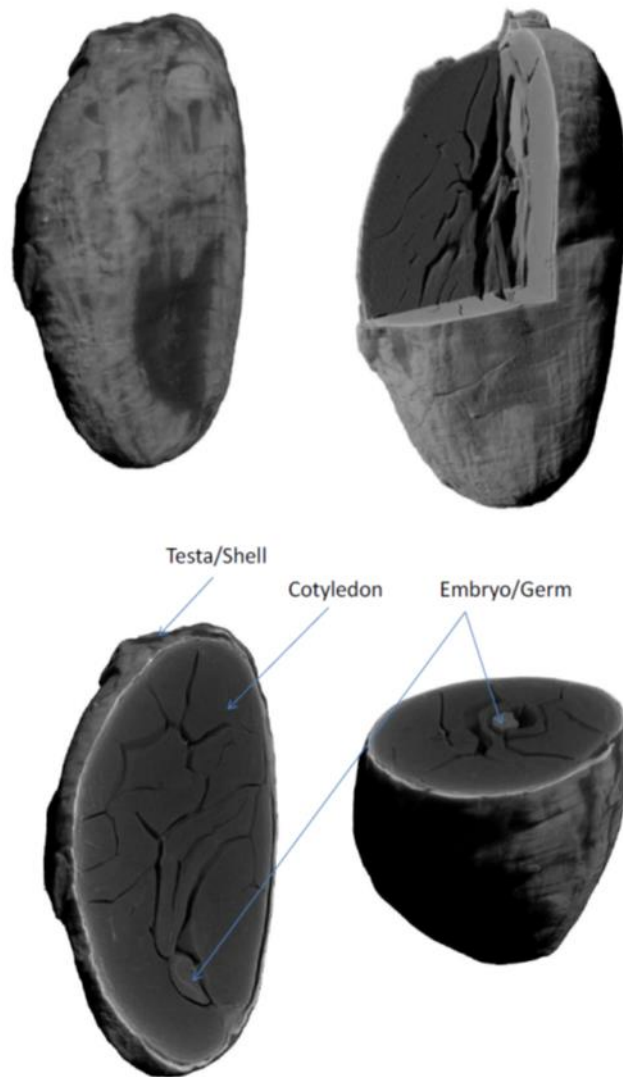
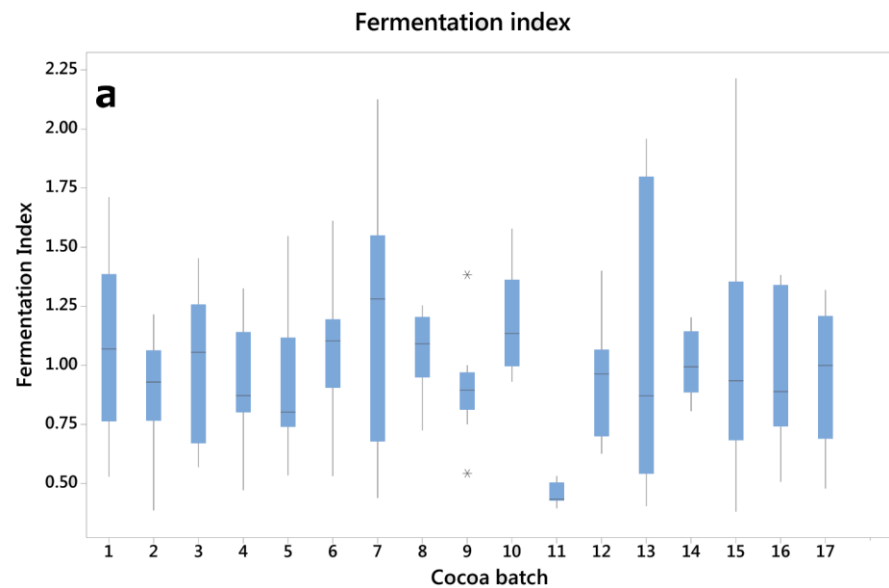


Validation (n=51)			
Min	Max	Mean	SD
0.39	2.21	1.00	0.38
6.43	213.87	85.45	60.53
20.88	472.26	208.77	114.94





## Reference measurements

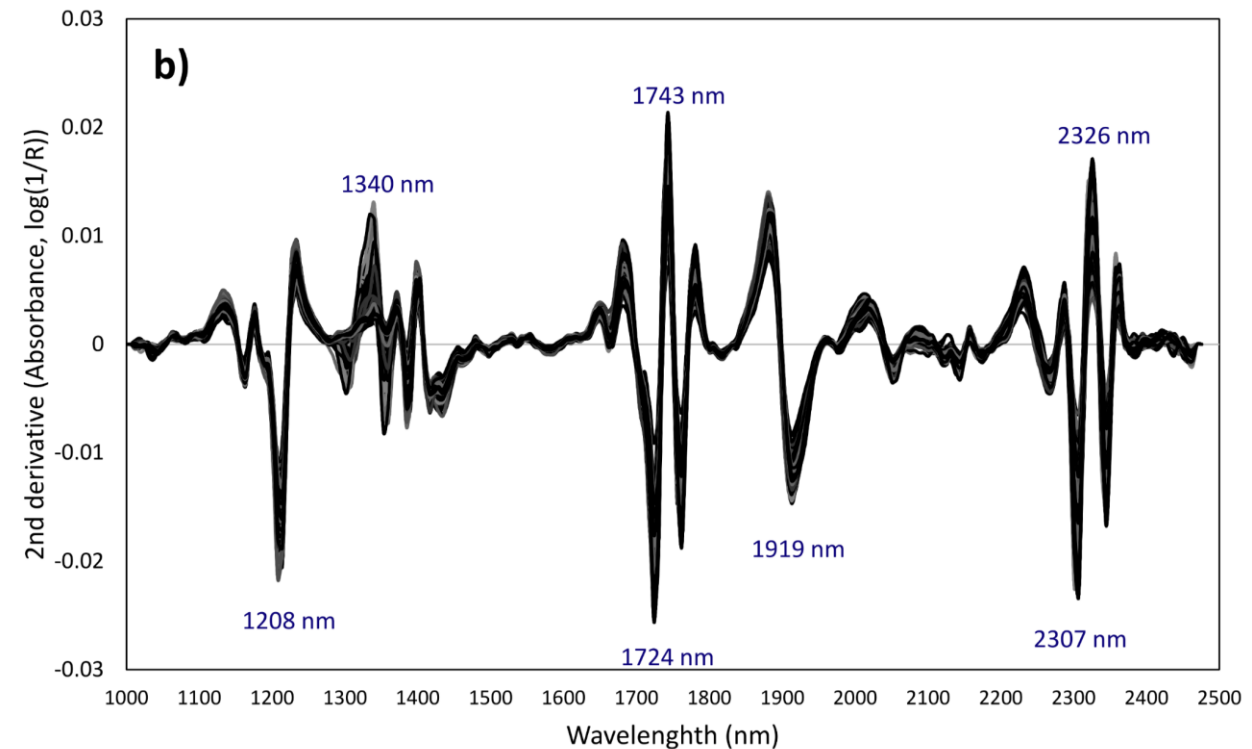
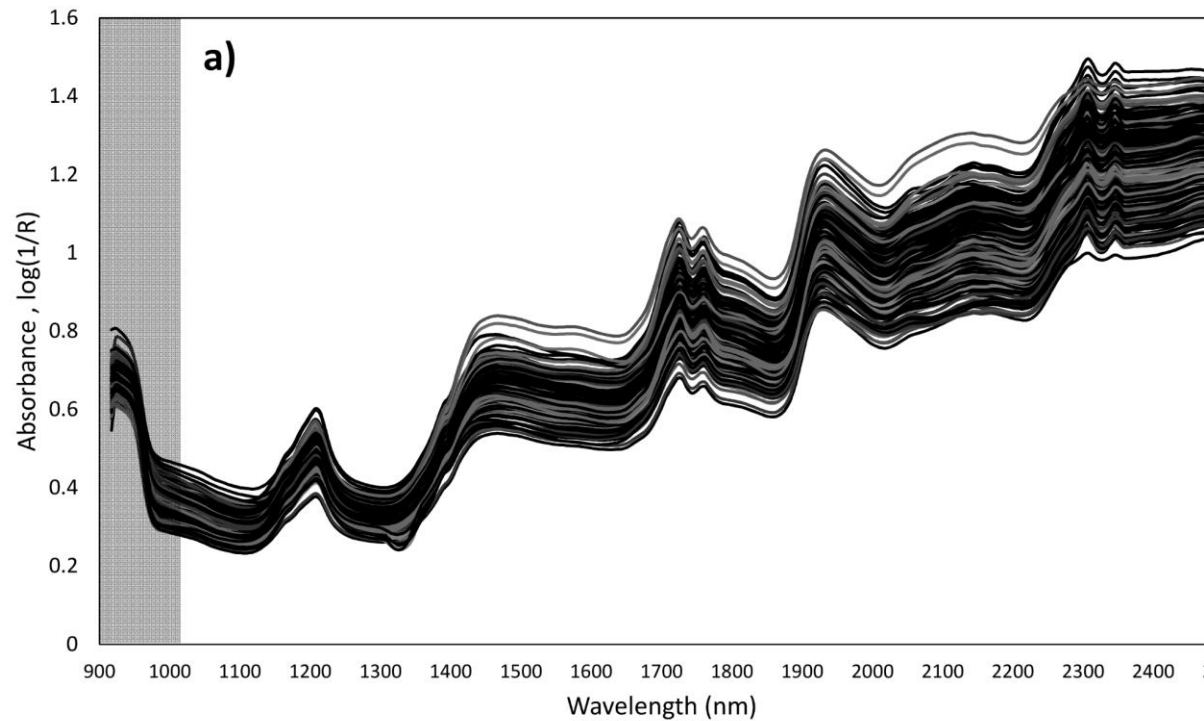




# Prediction of cocoa composition and quality

Reflectance spectra (absorbance and after 2<sup>nd</sup> derivative calculation)

One average spectrum exported for each cocoa nib





# Prediction of cocoa composition and quality

- Results of PLS regression methods
- Spectral normalisation and 2<sup>nd</sup> derivative spectral pre-treatment generally gave the best results

	LV	Calibration		Cross-validation			Prediction		
		R <sup>2</sup>	RMSEC	R <sup>2</sup>	RMSECV	RPD	R <sup>2</sup>	RMSEP	RPD
Fermentation index									
Log(1/R)	15	0.534	0.230	0.502	0.249	1.37	0.489	0.275	1.38
Normalisation	16	0.570	0.223	0.523	0.243	1.40	0.498	0.272	1.40
2 <sup>nd</sup> derivative	8	0.470	0.243	0.360	0.282	1.21	0.300	0.321	1.18
Total polyphenols									
Log(1/R)	15	0.809	23.96	0.727	29.74	1.83	0.660	36.39	1.66
Normalisation	16	0.817	23.35	0.757	28.09	1.94	0.700	34.13	1.77
2 <sup>nd</sup> derivative	8	0.810	23.39	0.697	31.36	1.74	0.573	41.04	1.47
Antioxidant activity									
Log(1/R)	15	0.818	49.37	0.710	62.72	1.85	0.625	71.15	1.62
Normalisation	16	0.825	48.25	0.738	59.55	1.95	0.680	65.88	1.74
2 <sup>nd</sup> derivative	8	0.764	55.25	0.741	59.23	1.96	0.735	60.04	1.91

# Results of Support Vector Machine Regression

- The use of **Support Vector Machine** regression did not lead to better prediction models compared to PLS regression
- Fermentation index had the lowest prediction performance
- The prediction errors for polyphenol content and antioxidant activity on single cocoa beans is acceptable considering the wide range of **natural variability**

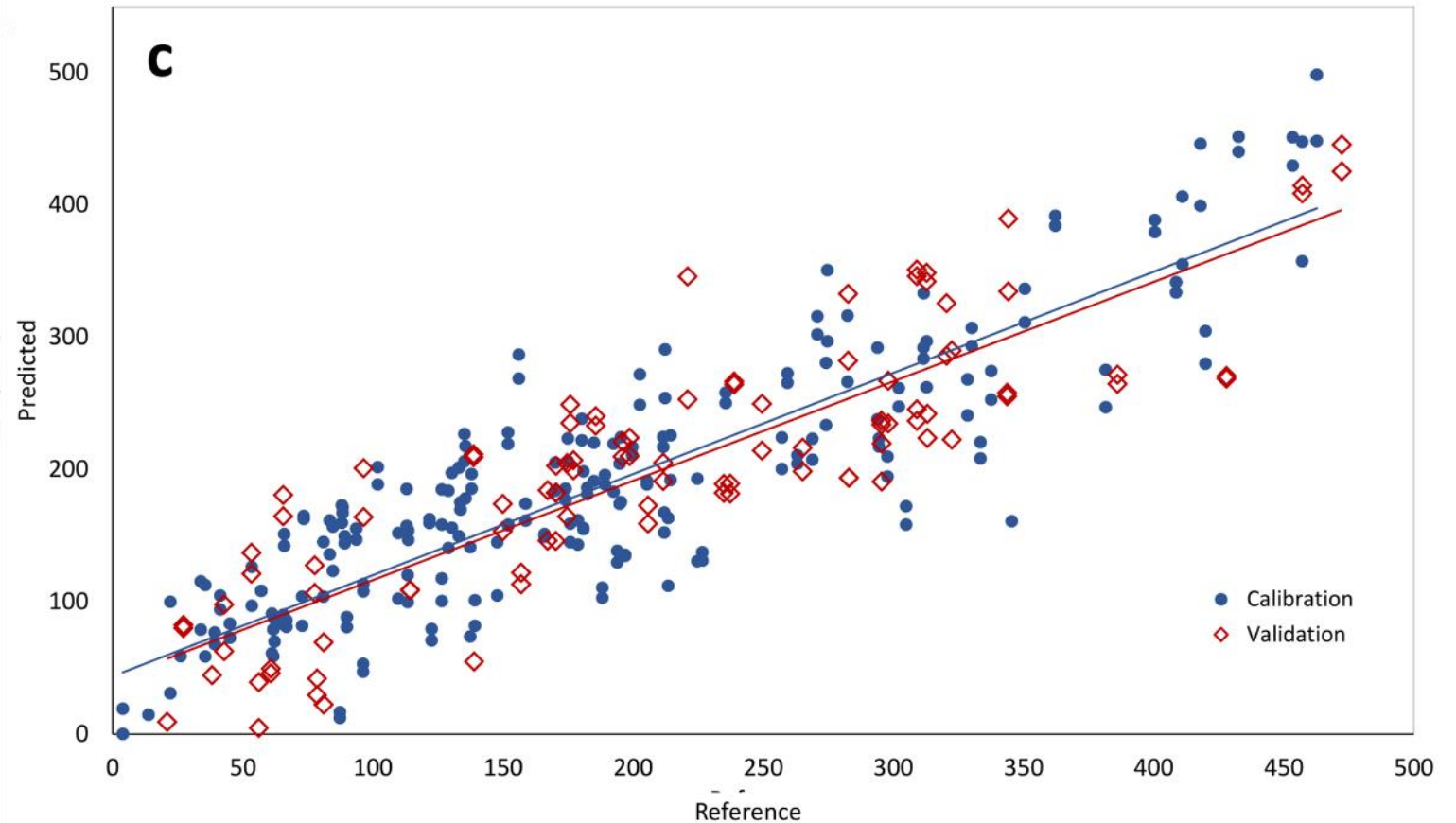
Parameter	Calibration		Prediction	
	R <sup>2</sup>	RMSEC	R <sup>2</sup>	RMSEP
Fermentation index	0.575	0.237	0.511	0.252
Total polyphenols	0.738	31.59	0.678	34.76
Antioxidant activity	0.729	64.65	0.688	69.01





# Prediction of cocoa composition and quality

- The prediction errors for polyphenol content and antioxidant activity on single cocoa beans is acceptable considering the wide range of **natural variability**
- **External validation**
  - a) F. index  $R_v^2 = 0.50$
  - b) Polyphenols  $R_v^2 = 0.70$
  - c) A. activity  $R_v^2 = 0.74$

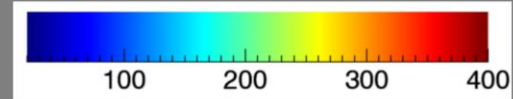
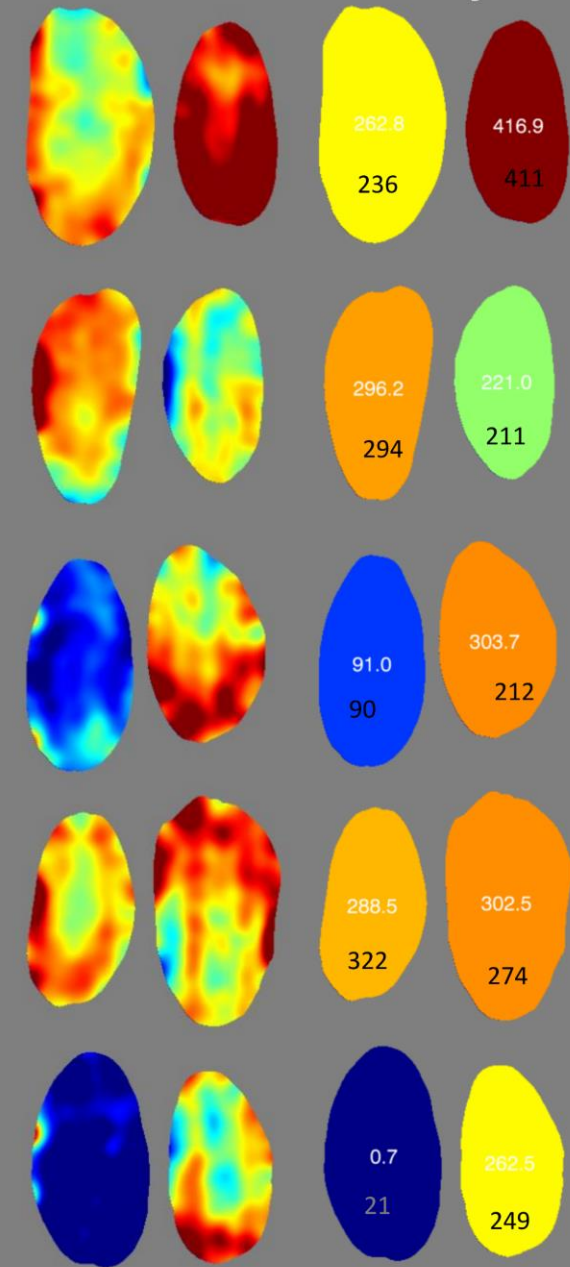
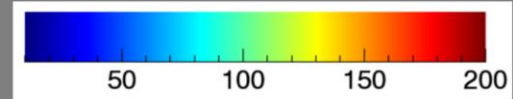
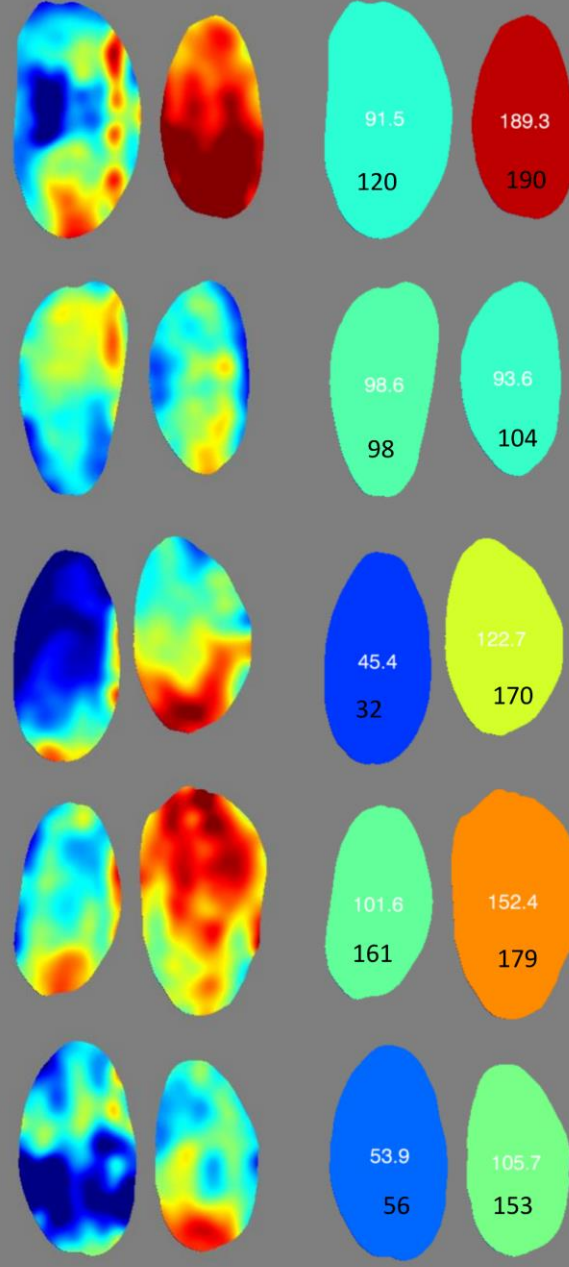
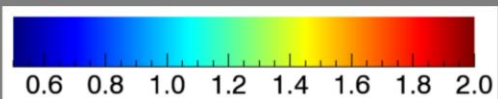
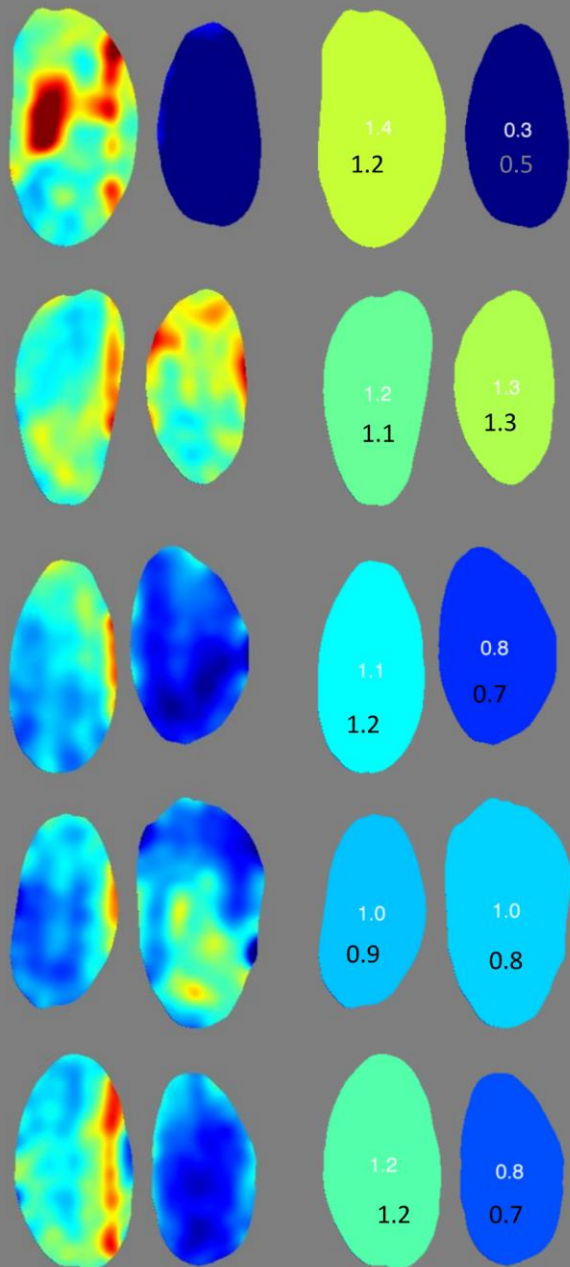
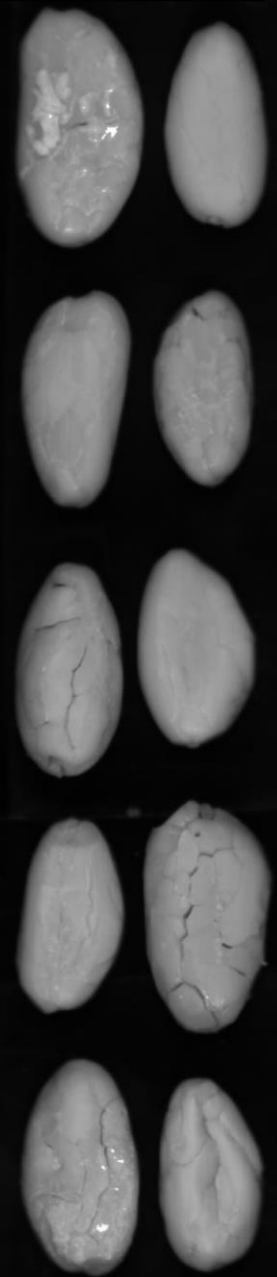


Reflectance

Fermentation index

Total polyphenols

Antioxidant activity







University of  
Nottingham

UK | CHINA | MALAYSIA

# HSI for coffee bean analysis

Applications

# Green coffee composition and quality

Parameter	Green		Roasted	
	<i>Coffea Arabica</i>	<i>Coffea canephora</i>	<i>Coffea Arabica</i>	<i>Coffea canephora</i>
<i>Carbohydrates/fibre</i>				
Sucrose	6.0–9.0	0.9–4.0	4.2-tr	1.6-tr
Reducing sugars	0.1	0.4	0.3	0.3
Polysaccharides (arabinogalactan, mannan, glucan)	34–44	48–55	31–33	37
Lignin	3.0	3.0	3.0	3.0
Pectins	2.0	2.0	2.0	2.0
<i>Nitrogenous compounds</i>				
Protein/Peptides	10.0–11.0	11.0–15.0	7.5–10	7.5–10
Free amino acids	0.5	0.8–1.0	-	-
Caffeine	0.9–1.3	1.5–2.5	1.1–1.3	2.4–2.5
Trigonelline	0.6–2.0	0.6–0.7	1.2–0.2	0.7–0.3
Nicotinic acid	-	-	0.016–0.026	0.014–0.025
<i>Lipids</i>				
Coffee oil (triglycerides with unsaponifiables)	15.0–17.0	7.0–10.0	17	11
Diterpene (free and esterified)	0.5–1.2	0.2–0.8	0.9	0.2
Minerals	3.0–4.2	4.4–4.5	4.5	4.7
<i>Acids and esters</i>				
Chlorogenic acids	4.1–7.9	6.1–11.3	1.9–2.5	3.3–3.8
Aliphatic acids	1.0	1.0	1.6	1.6
Quinic acid	0.4	0.4	0.8	1
Melanoidins	-	-	25	25

Concentrations are expressed as wt %.

Commercial differences between Arabica and Robusta

- **Moisture content:** impact on storage (risk of mould)
- **Lipid content:** impact on body and ‘crema’
- **Other compounds:** related to flavour after roasting



**Post-harvest processing:**  
dry processing  
wet processing



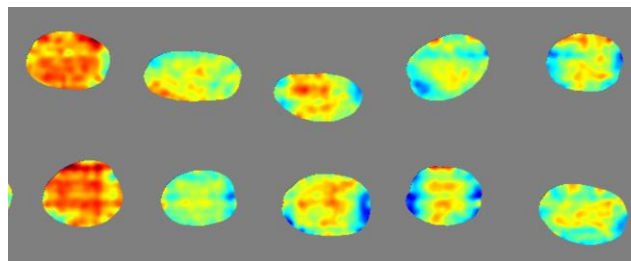




# Materials and methods: experimental design



**26 Batches  
of green  
coffee  
beans**



~6-10 Single  
coffee beans  
selected



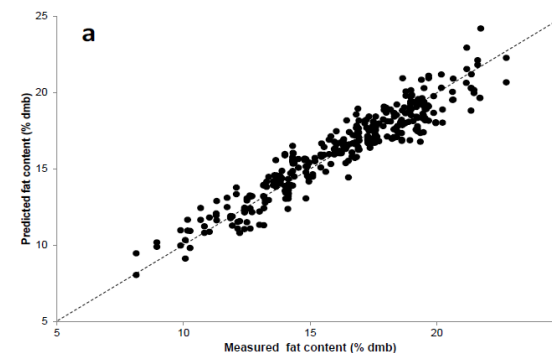
**Parameters evaluated:**  
Moisture, Fat  
Sucrose, caffeine,  
trigonelline

Hyperspectral  
imaging data  
Reference  
measurement  
(oven-drying,  
LC-MS, NMR,  
etc.)

Statistics  
(PLS  
regression)

& **Classification**  
models (SVM,  
etc.)

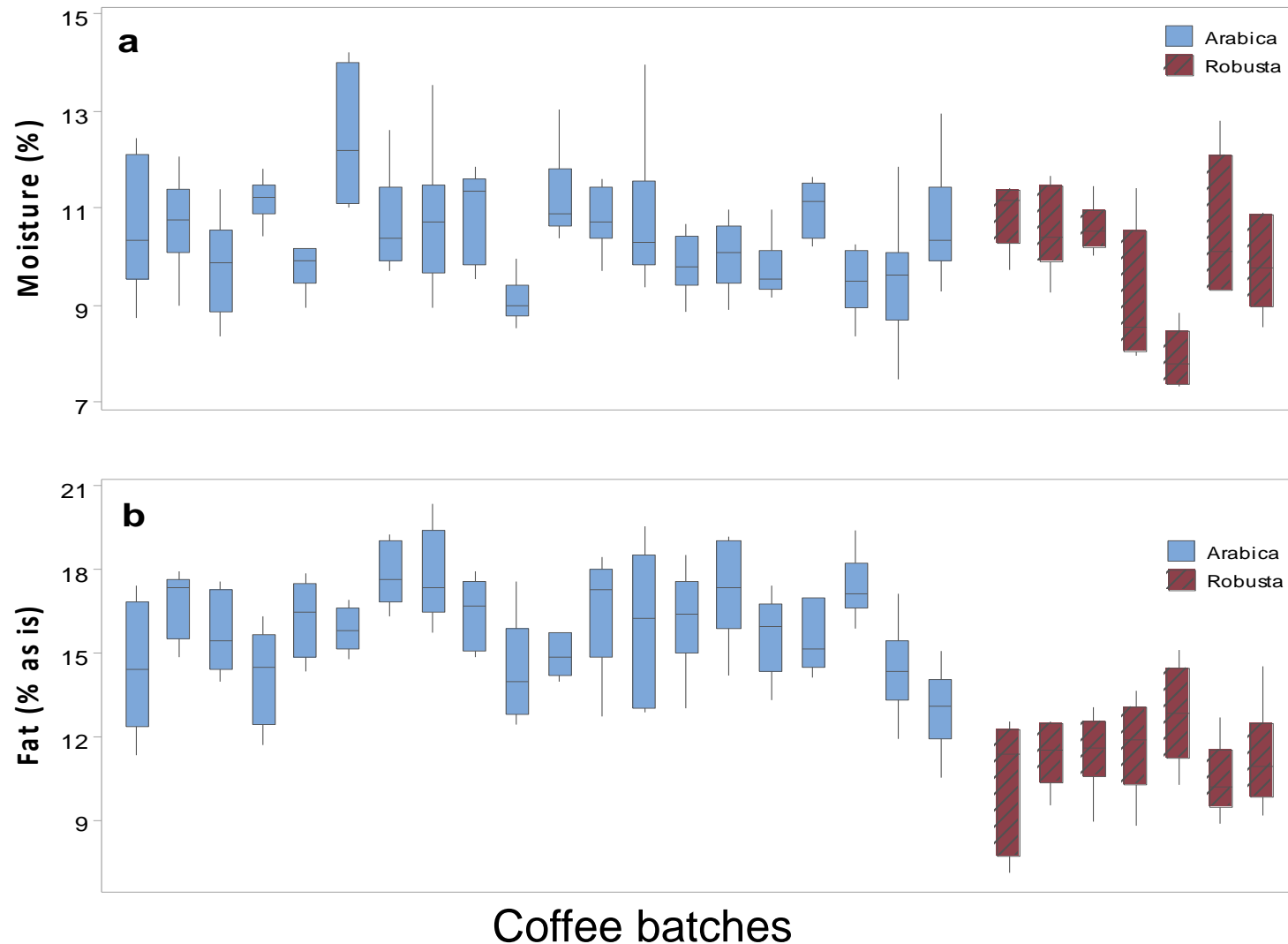
**Non-  
destructive  
prediction**





# Single bean analysis

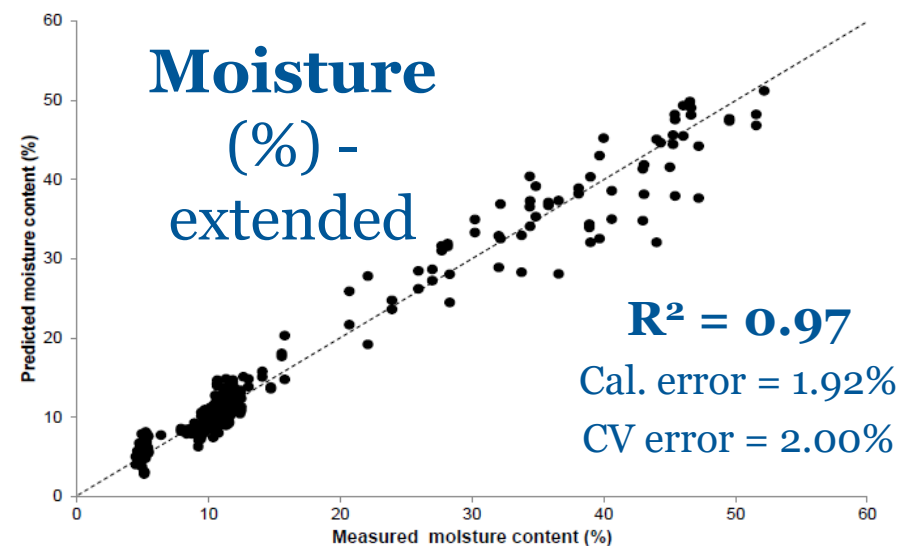
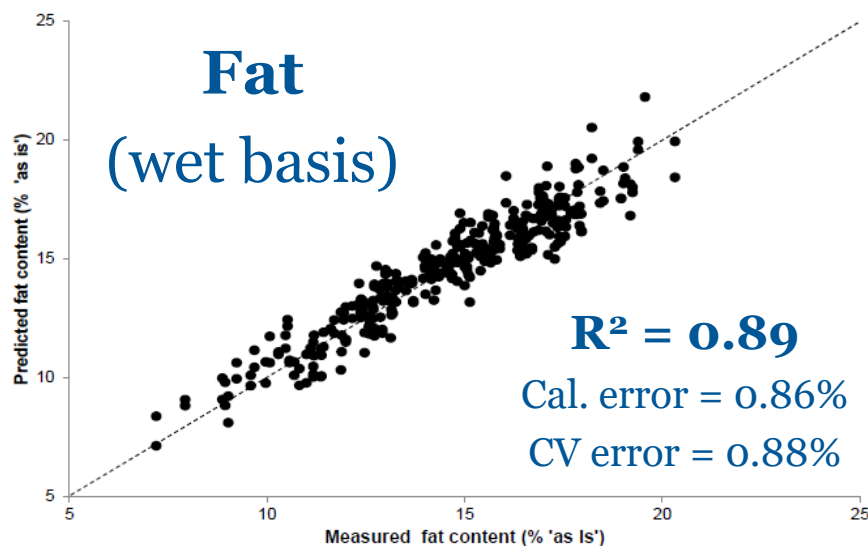
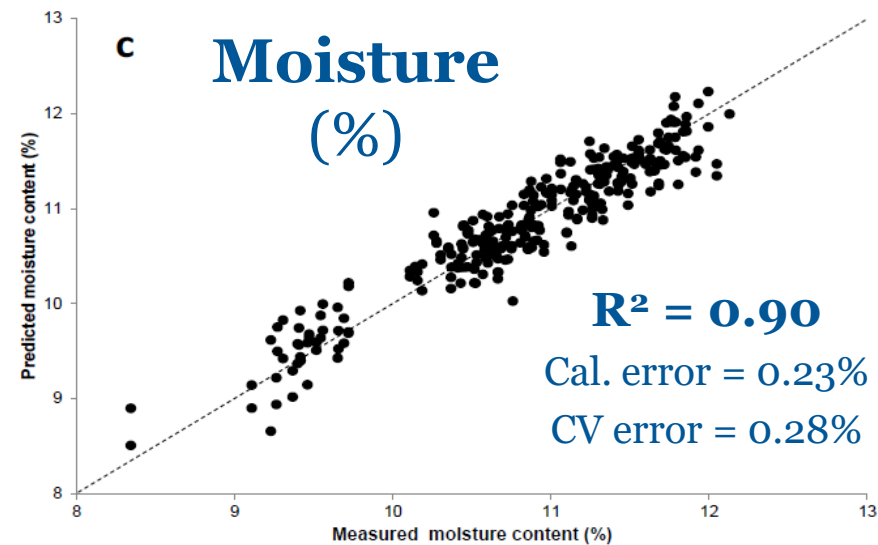
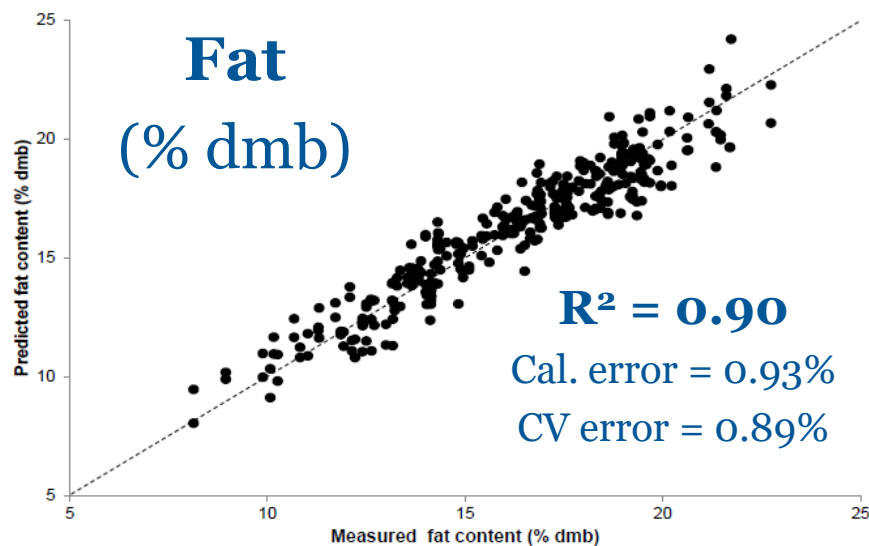
- Reference analysis of single coffee beans
- 10 beans per batch, on 25 batches
- Distribution of **moisture** (~8 to 13%) indicates consistency and high quality of the samples used
- Wide distribution of **lipid** content, ranging from ~7.0% to 20.3%. Expected differences between Arabica and Robusta, but also wide intra-batch variation was observed.
- Impact on the flavour of final coffee brew (e.g. body and “crema”)





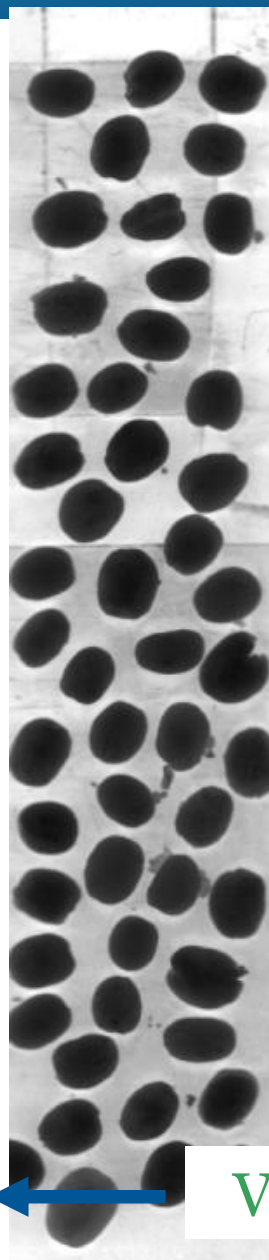
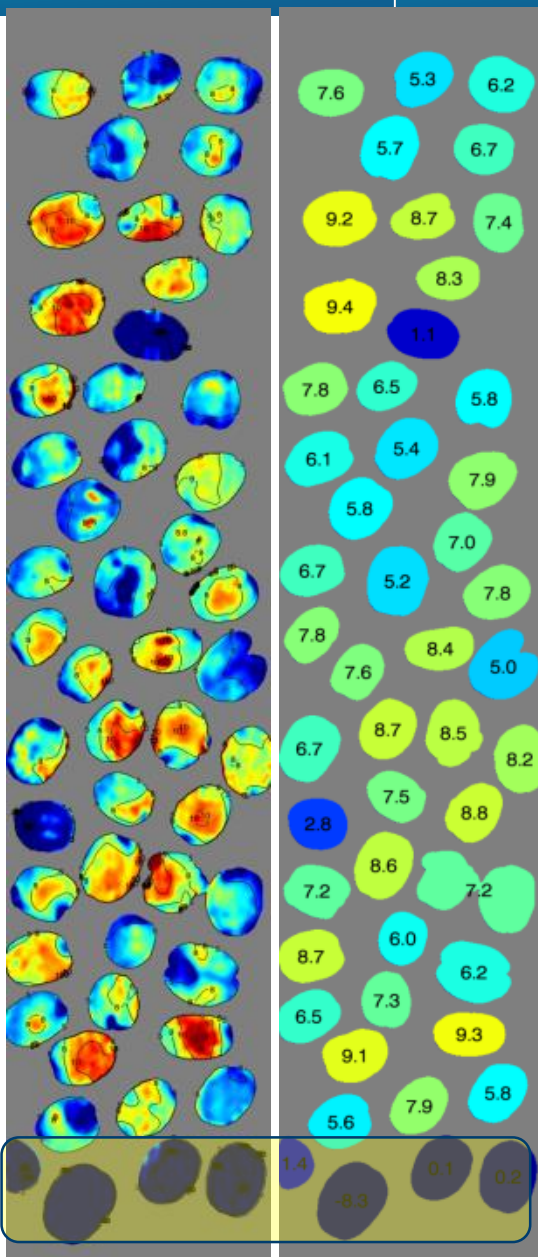


# PLSR models for fat and moisture prediction



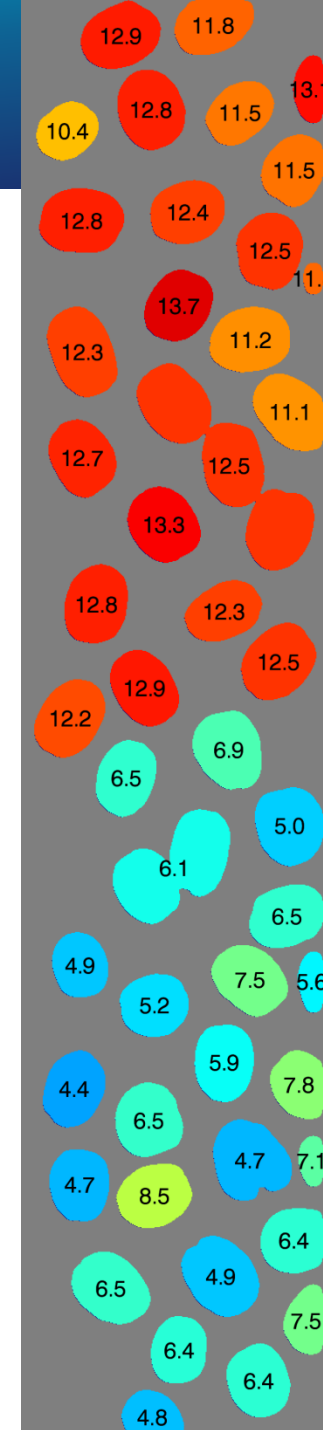
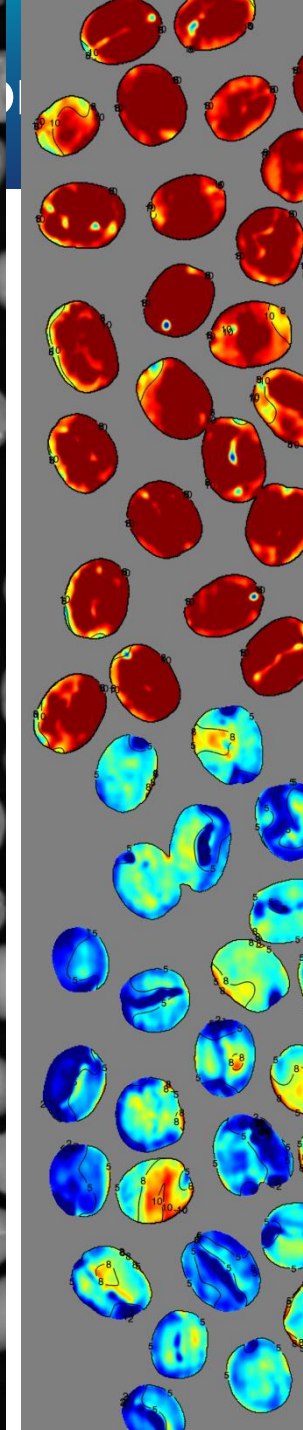
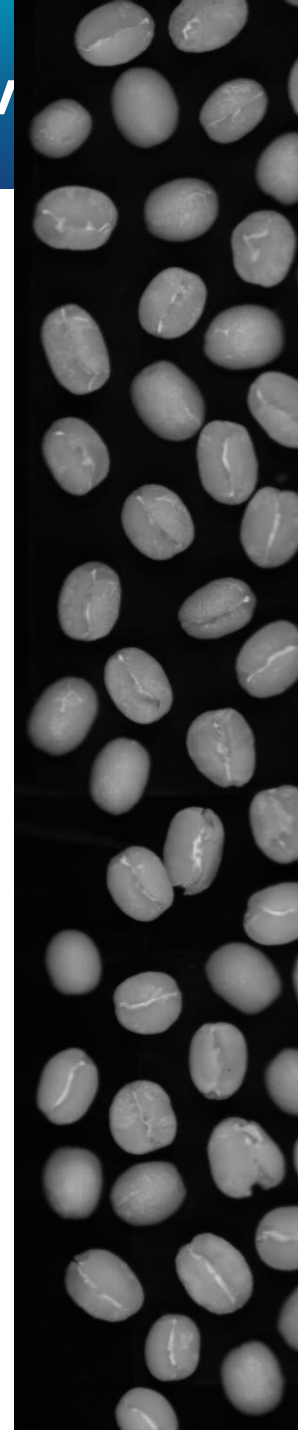


# Green coffee beans – moisture level v



- Example on two batches of Mexican **Arabica** green coffee
- Uneven distribution of moisture within single coffee beans
- The application of PLSR calibration allows to visualise it at single pixel level

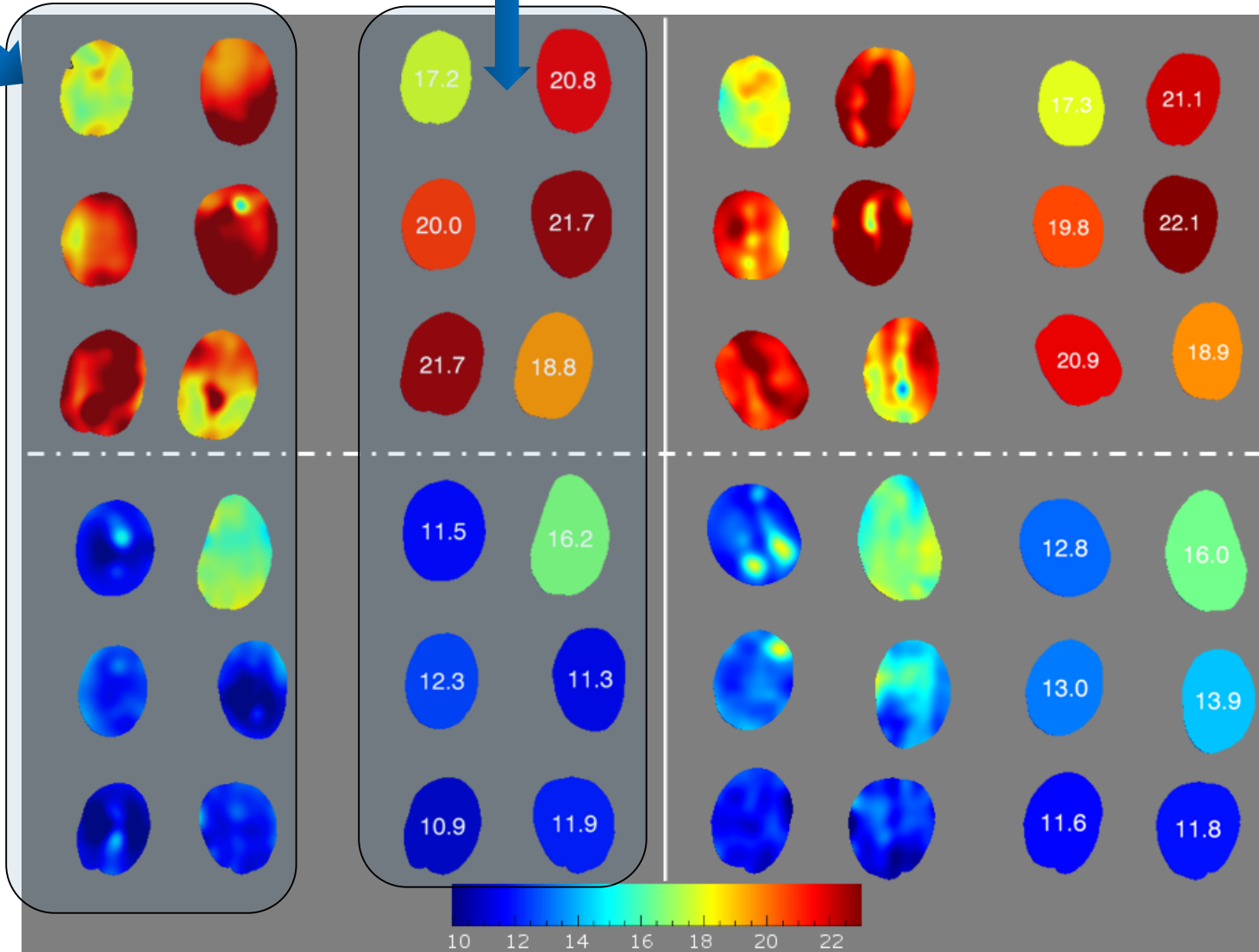
Very dry beans



# Fat (total lipids) prediction in single coffee beans

Single pixels  
(spatial  
distribution)

Single seed prediction



• Arabica

(dorsal vs  
ventral side)

• Robusta





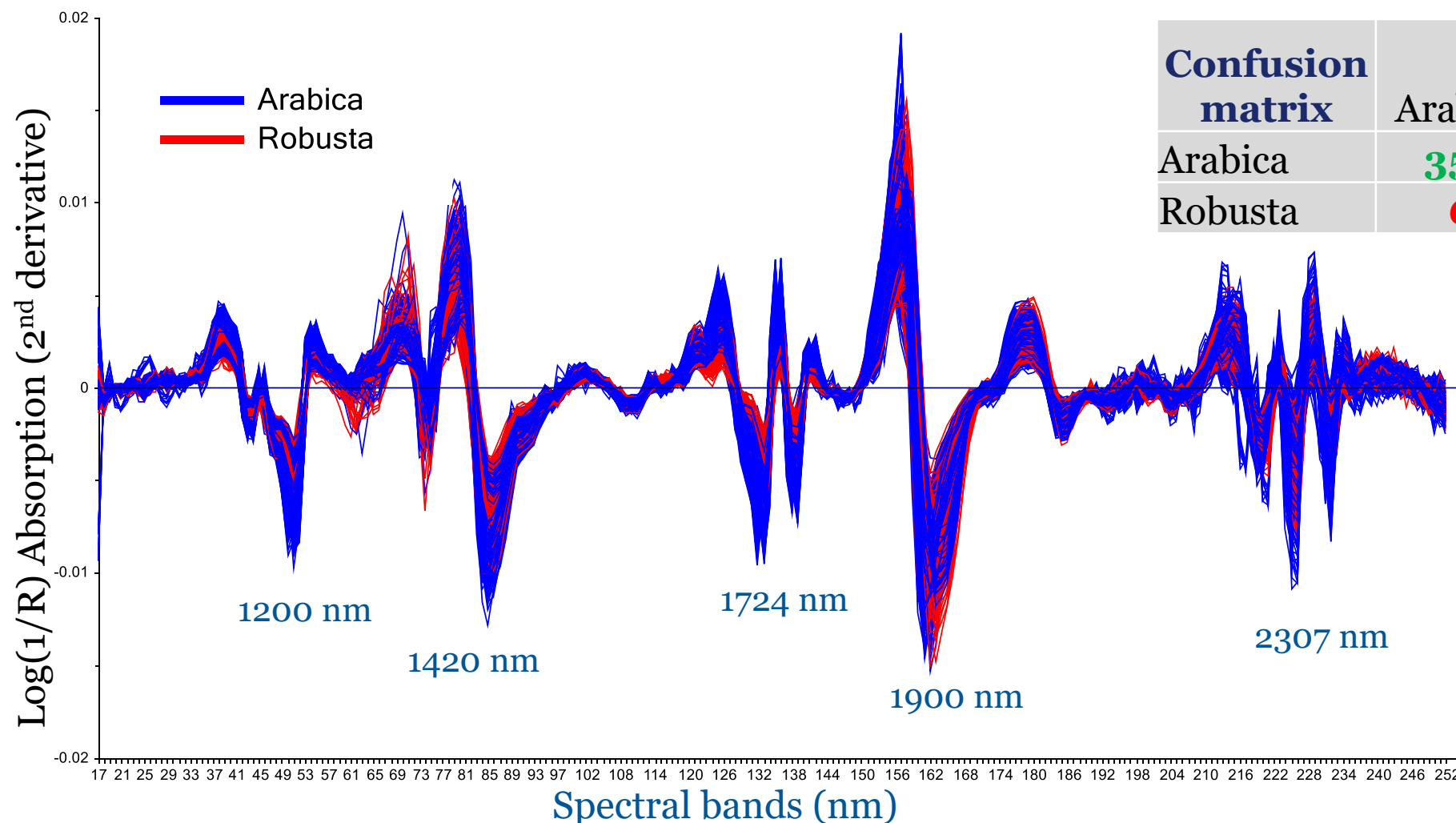
# Arabica / Robusta classification model

## HSI classification for coffee species: Support Vector Machine (SVM)

Species classification

accuracy:

72.5-97.3 %



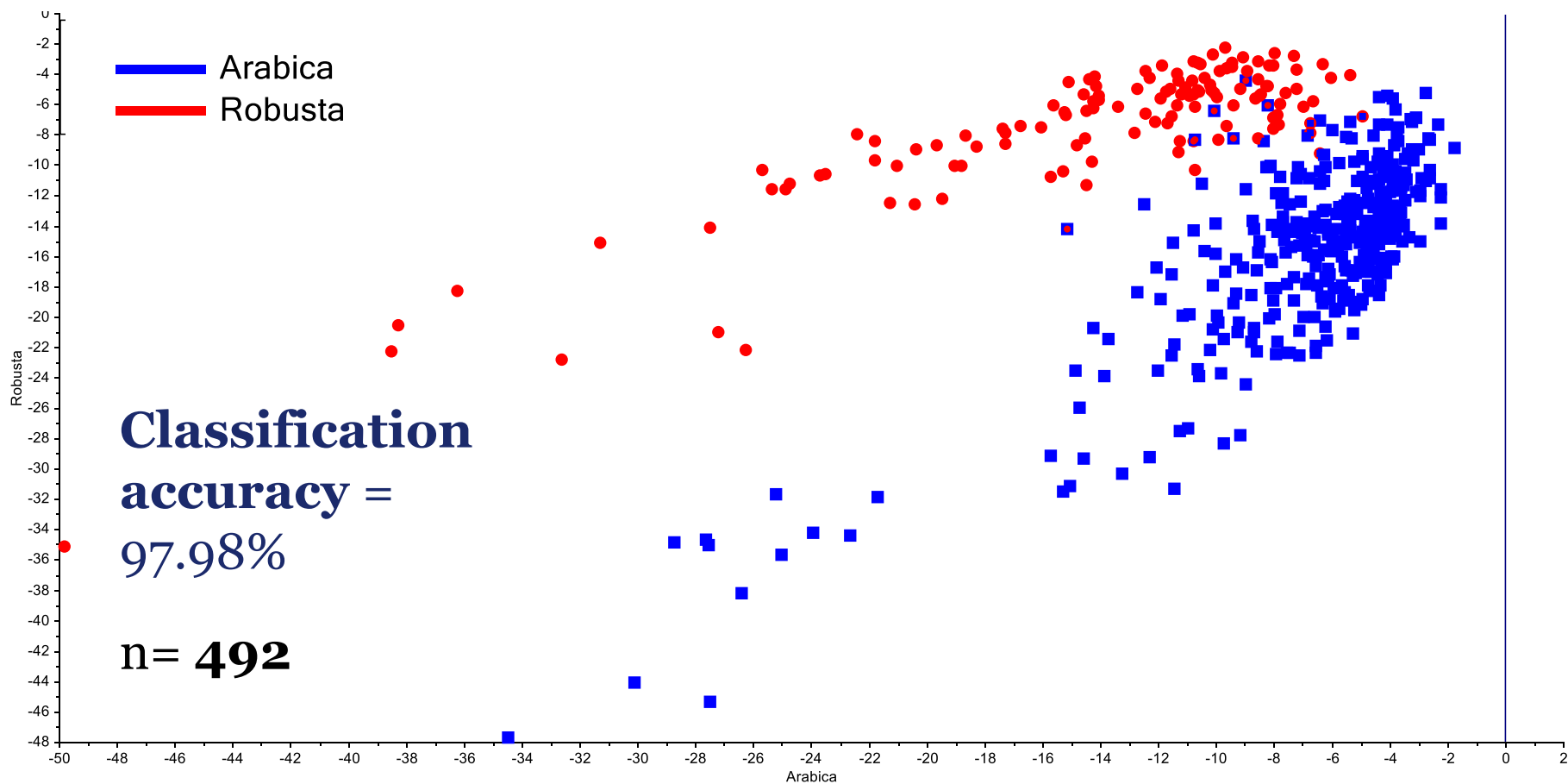
Confusion matrix	Arabica	Robusta
Arabica	358	4
Robusta	6	126



# Classification: Linear Discriminant Analysis

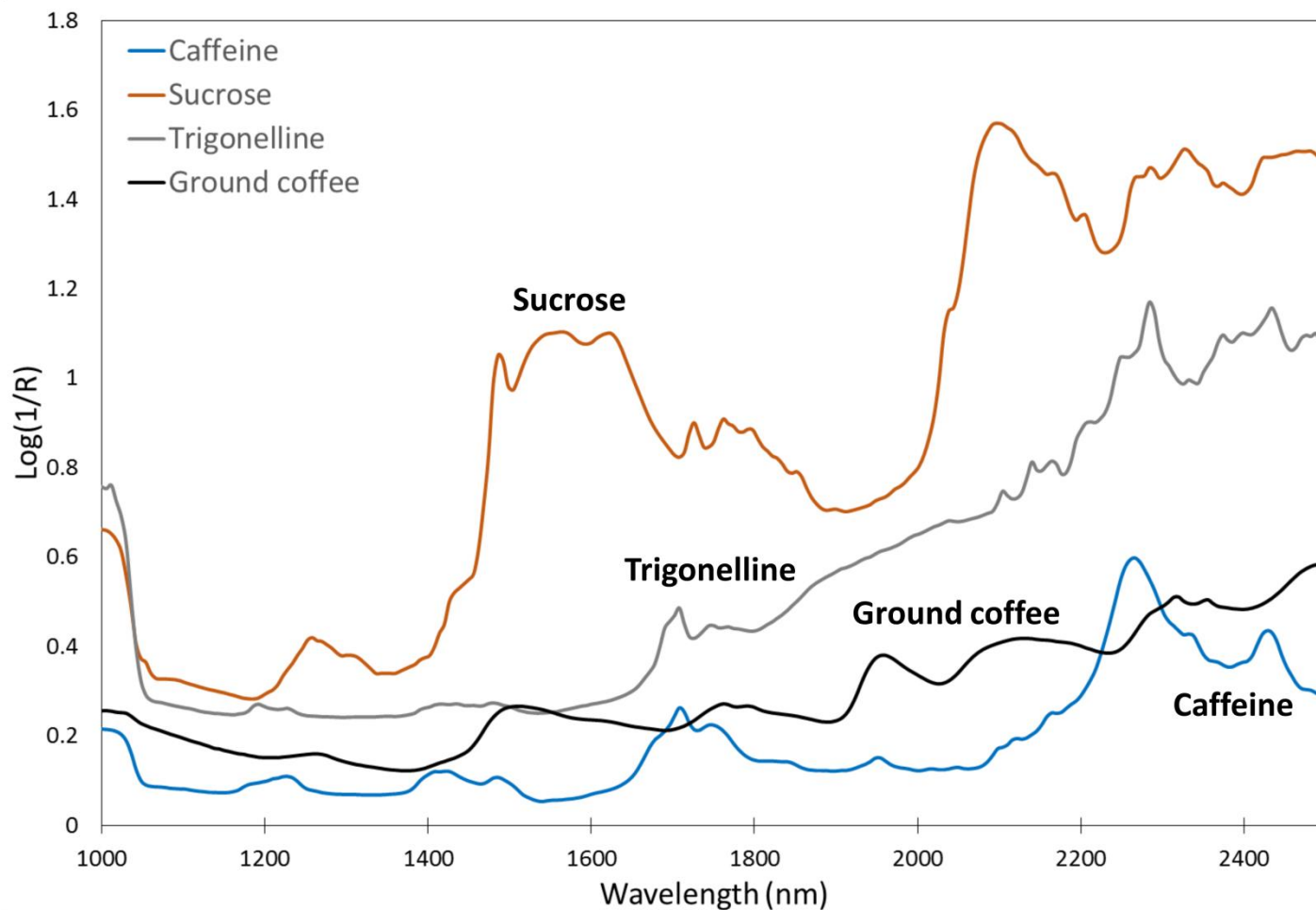
The single bean discrimination between Arabica and Robusta is of high commercial value (despite visual inspection is easy to carry out, it is slow)

The most useful spectral region is 1150-2450 nm

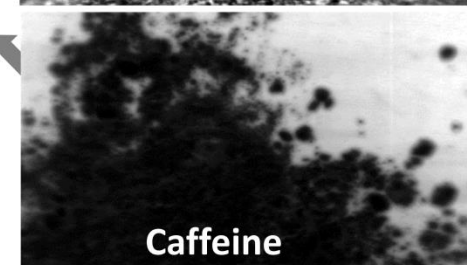
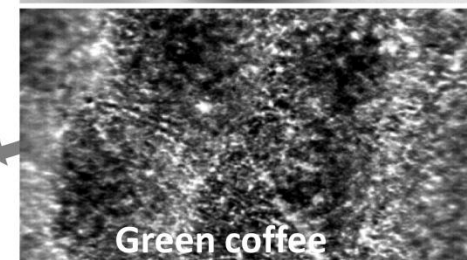
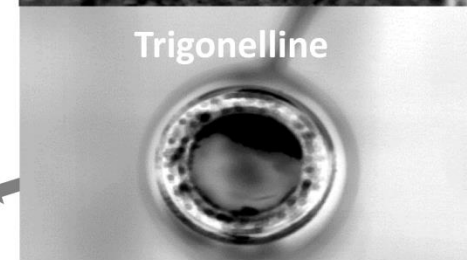
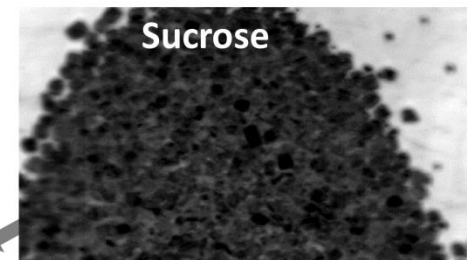




# NIR spectra of other compounds in green coffee



## Hypercubes:



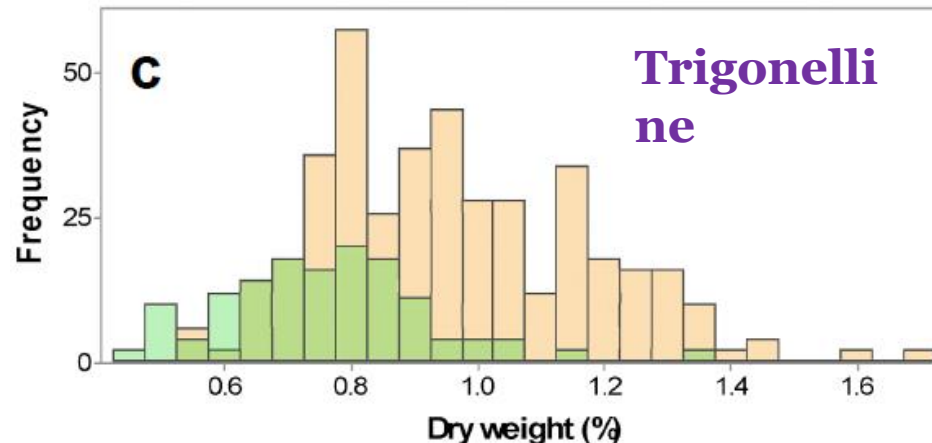
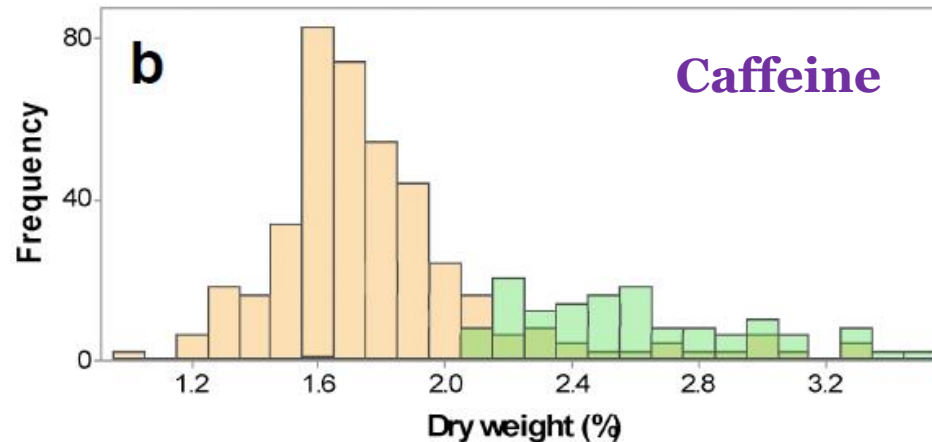
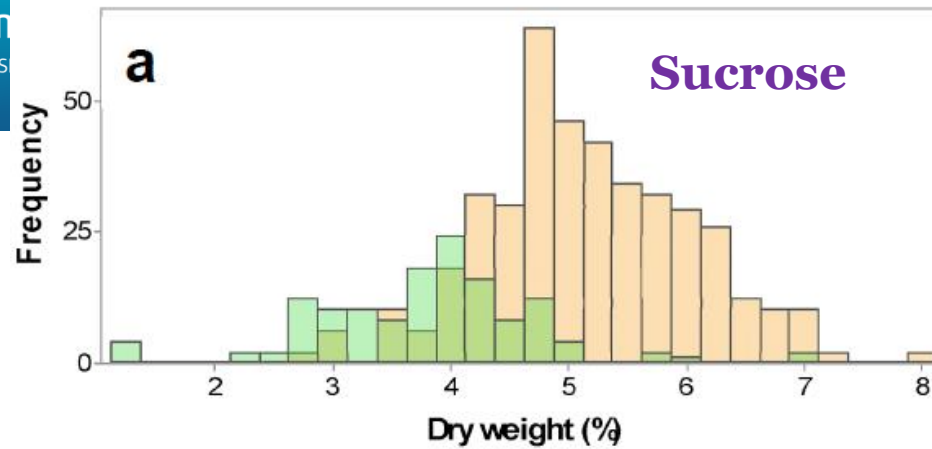


## Statistics of sucrose, caffeine and trigonelline: reference measurement

	Compound	Mean (mg/g)	SD (mg/g)		Min (mg/g)	Max (mg/g)
"As is"	Sucrose	43.28	10.21		5.34	70.76
	Caffeine	18.08	4.66		9.20	31.85
	Trigonelline	8.27	1.96		3.89	15.04
Dry basis	Sucrose	47.77	11.51		5.78	79.96
	Caffeine	19.91	5.07		10.15	34.77
	Trigonelline	9.12	2.18		4.32	17.00



## Analysis of other compounds at single coffee beans level



- LC-MS based method for **simultaneous** analysis
- Wide variation observed
- Obvious difference between Arabica and Robusta (but areas of **overlap**)
- **Single bean** data show within batch variability



# Visualisation of PLSR calibrations

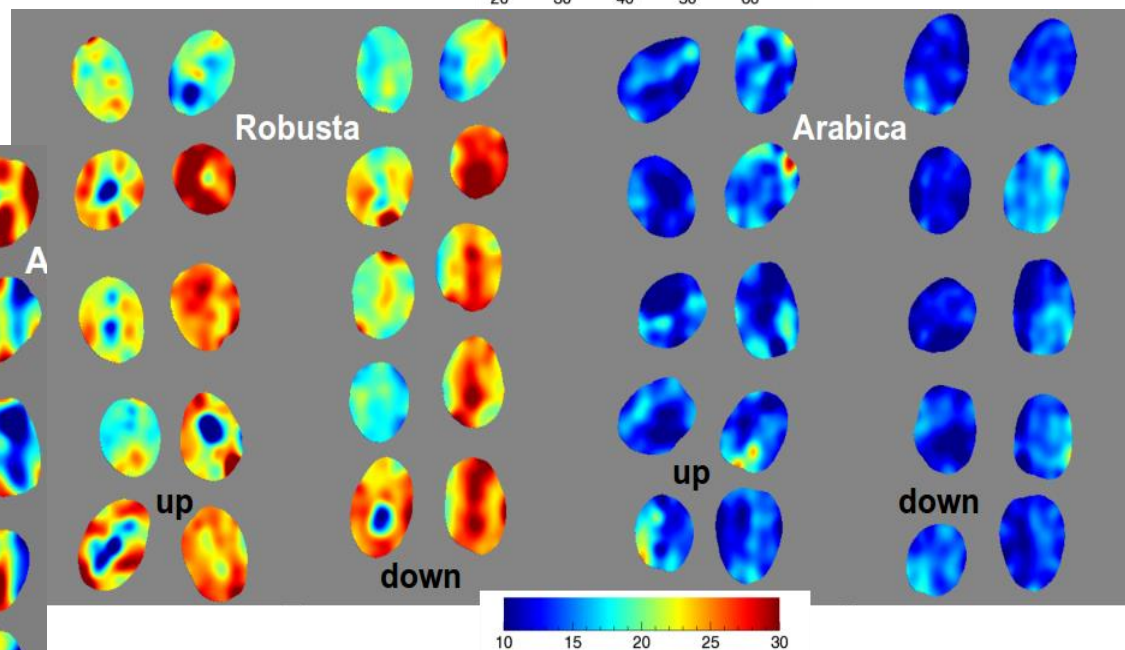
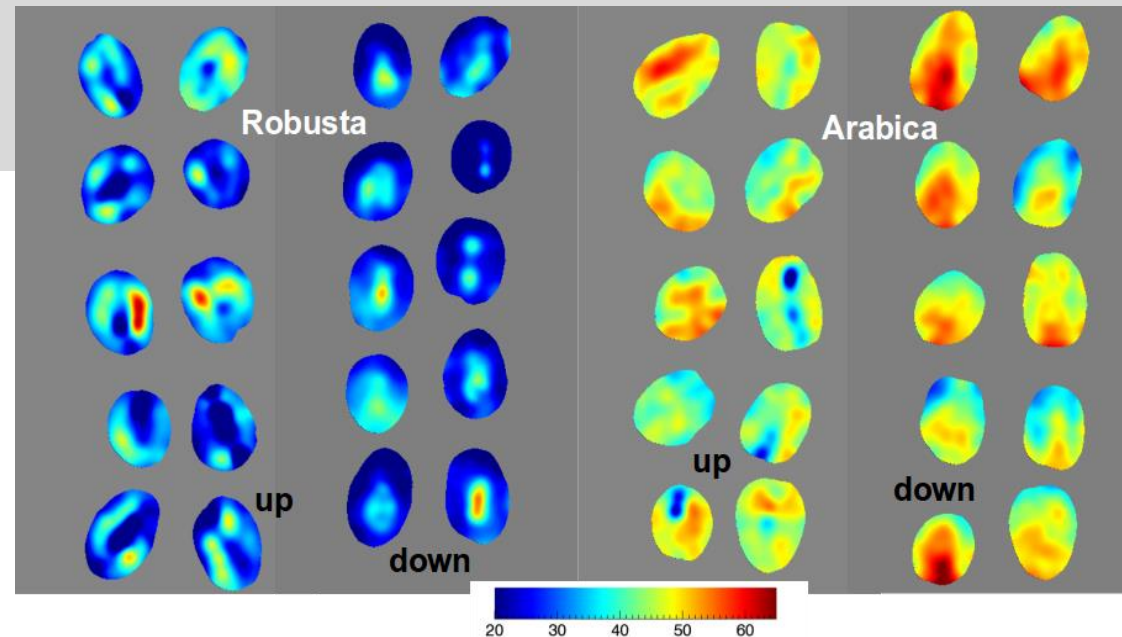
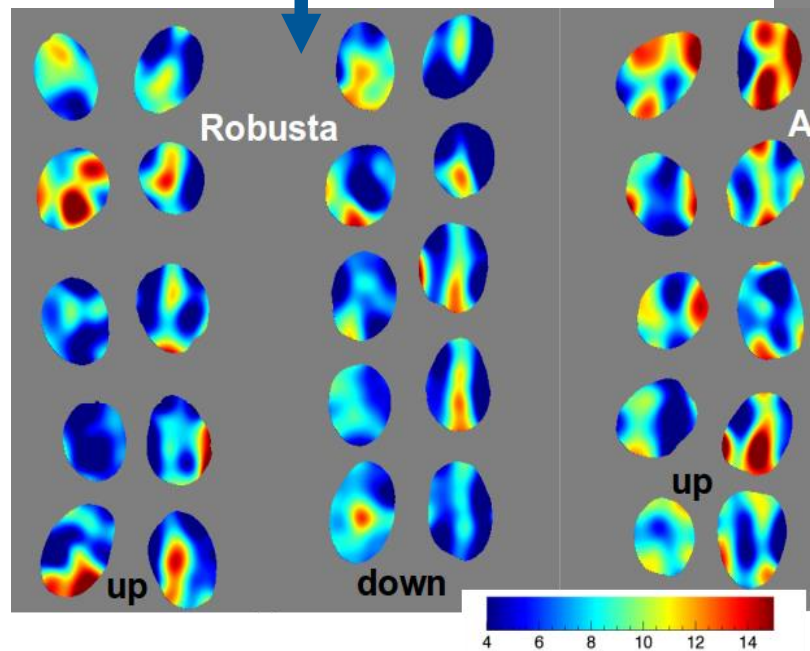
Sucrose



Caffeine



Trigonelline







# Conclusions

Hyperspectral imaging has been shown to be effective in predicting the following parameters in:

## **cocoa beans:**

- Total phenolic content
- Antioxidant activity
- (limited performance for fermentation index)

## **coffee beans:**

- Moisture
- Lipids
- Caffeine
- Trigonelline
- (limited performance for sucrose)

- The analysis of single beans gave an insight into the natural distribution and variability of target constituents
- HSI model performance to be evaluated in relation to:
  - Non-destructive nature
  - Speed of analysis
  - Possibility of single bean analysis
  - Possibility of visualising spatial distribution
  - No preparation needed (e.g. grinding, extraction, etc.)



University of  
Nottingham

UK | CHINA | MALAYSIA

Thank you

Any question?

# Output from my research on HSI

My publications are available on my  
Researchgate page. ORCID 0002-1826-2565

← → ↺ 🏠  
R<sup>6</sup> Home Questions Jobs Search 🔍  
📄 📧 💬 2 👤  
Nicola Caporaso  
29.23 · BSc, MSc, PgCert  
Add new research +  
Overview Contributions Info Stats Scores Research Interests  
Research  
Projects (1)  
Research Items  
All (44)  
Projects (1)  
Hyperspectral chemical imaging for the chemometric assessment of food grain composition  
Project  
Add update

5 research papers  
1 review paper

2 posters  
3 presentations

To contact me:  
**nicola.caporaso1**  
**@nottingham.ac.uk**