Rapid Analysis of Biomass with NIRS

DIBANET

With the support of the Seventh Framework Programme.

DIBANET Workshop

9-17 Dec 2009

Time for Conventional Analysis

With the support of the Seventh Framework Programme.

Sample as Collected

Chop sample
~ 10 mins

Wet Chopped Sample

Air Drying
~ 3+ days

Dry Sample

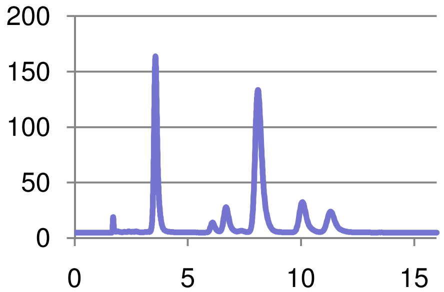
Milling + sieving
~ 1 hour

Dry Sample of
Appropriate Particle Size

Extractives
Removal
~ 3 days

Extractives-free
sample

Hydrolysis and
hydrolysate
analysis
~ 3 days



Completed
Lignocellulosic Analysis
~ **10 days !!!!**

Requirements for Rapid Analysis

With the support of the Seventh Framework Programme.

- Minimal/no sample preparation necessary.
- Ability to analyse wet biomass feedstocks.
- Able to quantify components important to DIBANET, including:
  - C6 Sugars: Glucose, Galactose, Mannose
  - C5 Sugars: Arabinose, Xylose
  - Lignin content (acid soluble and insoluble)
  - Extractives
- Results accurate and reproducible.
- Technology that can potentially be integrated online at a biorefining facility.

Rapid Analysis Techniques

With the support of the Seventh Framework Programme.

- Pyrolysis techniques (PyGC-MS, Py-MBMS).
- Thermo-gravimetric analysis (TGA).
- NMR techniques (e.g. Low Field NMR).
- FTIR (mid-infrared spectroscopy).
- Near Infrared Spectroscopy (NIRS)

Spectroscopy

With the support of the Seventh Framework Programme.

- The study of the interaction between electromagnetic (EM) radiation and matter as a function of wavelength or frequency.
- Spectrometry - a spectroscopic technique used to assess the concentration of a given chemical species. The instruments that carry out this analysis are spectrometers/spectrophotometers.
- EM waves can be described by their frequency (*f*, typically in Hz), their wavelength (λ) and their photon energy (*E*).
- Wavelength is inversely proportional to frequency.
- Photon energy is directly proportional to the frequency.

Effect of EM Radiation on Matter

With the support of the Seventh Framework Programme.

- When EM radiation interacts with single atoms and molecules, its behaviour depends on the amount of energy per quantum (photon) it carries. e.g.:
  - Ultraviolet radiation: Excitation of molecular and atomic valence electrons, including ejection of the electrons.
  - X-rays: Excitation and ejection of core atomic electrons.
  - Gamma rays: Excitation of atomic nuclei.
  - Microwave to far-infrared: Molecular rotations and vibrations.
  - **Near Infrared**: Molecular vibrations (overtones and combination bands).

Theory – Infrared Spectroscopy

With the support of the Seventh Framework Programme.

- Molecules absorb at specific frequencies that are characteristic of their structure. These are resonant frequencies (the frequency of the absorbed radiation matches the frequency of the bond that vibrates).
- These resonant frequencies can be approximated from the strength of the bond and the mass of the two atoms involved.
- For a vibrational mode to absorb IR radiation, it must be associated with a change in the permanent dipole of the bond.

NIRS - History

- Infra = below (i.e. below red in energy).
- NIR region discovered in 1800 by William Herschel. He used a prism to disperse sunlight on to three thermometers and found that just beyond the red light, where there was no visible light, the temperature was greatest.
- Interest in NIRS between then and 1950s was low since it was considered less attractive than MIRS - many people considered that the absorptions were too weak (2 or 3 times less than in the MIR region) and also too confusing (numerous overlapping peaks of overtone and combination bands).

NIRS – History (2)

With the support of the Seventh Framework Programme.

- Research activity increased following the commercial availability in the 1950s of lead sulphide (PbS) as a sensitive heat detector for the NIR region.
- Since then, the advances in microprocessor capabilities and the developments of laboratory and online NIRS systems as well as the expansion of chemometric techniques has greatly increased the presence of NIRS in many sectors.
- Chemometrics is needed to interpret the variations in NIR spectra.

NIRS - Principles

With the support of the Seventh Framework Programme.

- The spectrum of a sample is recorded by passing a beam of infrared light through the sample. Examination of the transmitted/reflected light reveals how much energy was absorbed at each wavelength.
- This can be done with a monochromatic beam, which changes in wavelength over time, or by using a Fourier Transform (FT) instrument to measure all wavelengths at once.
- From this, a transmittance or absorbance spectrum can be produced, showing at which wavelengths the sample absorbs.

NIR - Advantages

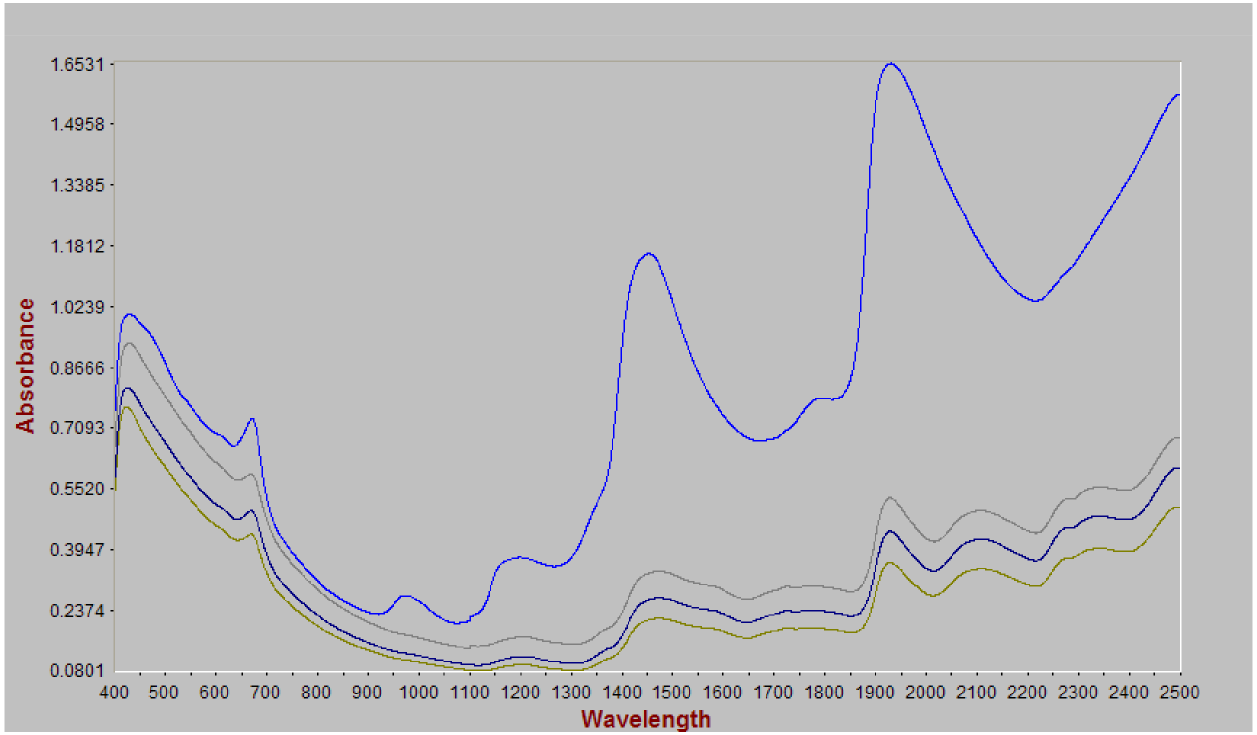
With the support of the Seventh Framework Programme.

- NIR can typically penetrate much farther into a sample than MIR radiation.
- Also, the absorbance of water in the NIR is not as overpowering as in the MIR, hence analysis of wet materials is possible
- Therefore NIRS has utility in analysing materials with **little/no sample preparation**.

NIR – Chemometrics Needed

With the support of the Seventh Framework Programme.

- The molecular overtone and combination bands seen in the near IR are typically very broad, leading to complex spectra; it can be difficult to assign specific features to specific chemical components.
- Multivariate (multiple wavelength) calibration techniques are often employed to extract the desired chemical information.
- Spectral pre-processing (e.g. first derivatives) may also be needed.



Testing Calibrations

With the support of the Seventh Framework Programme.

- Calibration equations need to be tested.
- They can be tested using the same samples that were in the calibration set however that leads to over-fitting and uncertainty as to whether the equation will be accurate for unknown samples.
- Ideally, a separate data set (validation set) will be used to test the calibration and provide informative statistics.
- If there is a limited number of samples cross validation can be used – this takes out one (or more) sample at a time and develops the model. The missing sample(s) are then predicted against this model and the residual determined.
- It is important that the calibration and validation sets are of a similar structure (spectral and reference data variability).

Standard Error of Laboratory - SEL

With the support of the Seventh Framework Programme.

- This can be determined by using a single sample analysed in replicate with the average value being:

$$\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N}$$

- If duplicate analyses are used in a single laboratory the SEL is:

$$\mathrm{SEL} = \sqrt{\left\{\frac{\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2}{N}\right\}} = \sqrt{\left\{\frac{\sum_{i=1}^{N}\left(y_1 - y_2\right)^2}{2N}\right\}}$$

- Where: y_i= individual laboratory result; $\bar{y}$ = mean of two lab results; *N* = total number of samples (not total number of analyses)
- If multiple replicate analyses are preformed, we calculate SEL as:

$$\mathrm{SEL} = \sqrt{\frac{\sum_{j=1}^{M}\sum_{i=1}^{N}\left(y_{ij} - \bar{y}_j\right)^2}{N(r-1)}}$$

- *r* = number of replicate analyses for each sample

Important Statistics - r^2

With the support of the Seventh Framework Programme.

- **Multiple Correlation Coefficent (r^2)**

$$R^2 = \left(\frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \right)$$

- Allows us to determine the amount of variation in the data that is adequately modelled by the calibration equation as a total fraction of 1.0. i.e. an r^2 of 1 would represent a perfect calibration, whereas a value of 0 would indicate there is no link between the model and the actual chemical values.

Statistics – SEC/SECV/SEP

With the support of the Seventh Framework Programme.

$$\text{Standard Error of Calibration, SEC} = \sqrt{\frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{N - K - 1}}$$

- This is the standard deviation for the residuals due to differences between actual and the NIR predicted values for samples within the calibration set. .
- This equation calculates the RMS error, corrected for degrees of freedom. This correction takes into account the fact that when many variables are used for fitting the data, there may be overfitting.
- The SEC is a useful estimate of the theoretical “best” accuracy obtainable for a specified set of variables used to develop a calibration equation.
- As an estimate of the prediction accuracy, **Standard Error of Prediction (SEP)** is needed. Here residuals from the prediction set are used.
- Alternatively, the **Standard Error of Cross Validation (SECV)** uses the residuals from cross validation.

PRESS – Prediction Sum of Squares

With the support of the Seventh Framework Programme.

- **PRESS (Prediction Sum of Squares)**
- Allows for the optimum number of factors in PCR/PLSR.
- Starts by using one factor as a model while leaving a single sample out of the calibration set. After a calibration is developed using the remaining samples the algorithm is programmed to predict the one selected sample and record the difference between the actual versus the predicted values.
- When the procedure is repeated for the entire sample set, the sum of squares for the residuals $\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ is reported. The procedure then adds another factor and

PRESS (2)

- The procedure will stop when the predesignated number of factors is reached and at that time **the model with the least number of factors and smallest sum of squares for residuals** is selected as the best model for the calibration set used.
- A plot of PRESS values (*y* axis) versus the number of factors (*x* axis) is often used to determine the minimal PRESS corresponding with the smallest number of factors.

ASTM Guidelines for NIRS Calibration

With the support of the Seventh Framework Programme.

- The quality of the calibration is determined relative to the natural variation that exists for the components of that feedstock.
- A ratio is determined by dividing the range in concentration by the SEP.
- A ratio of 15+ is considered to be good for quantitative calibration.
- A ratio of 10-15 is considered good for screening control.
- E.g. if the cellulose content of miscanthus varies between 40 and 50%. The maximum SEP permitted to allow a ratio of 15 is 0.67%. This shows the importance of **highly accurate and reproducible lab methods**.

Spectral Pretreatments

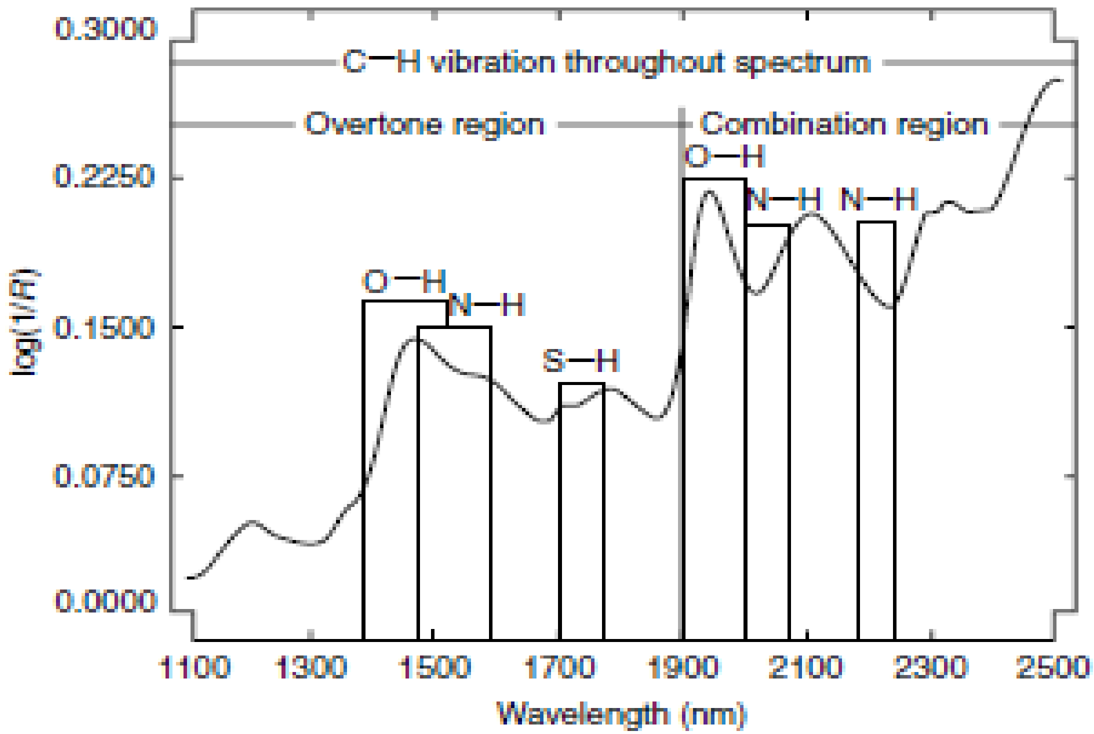
With the support of the Seventh Framework Programme.

- There are various chemometric techniques to treat spectra in order to reduce/remove unwanted effects (e.g. variation in particle size) so that the calibration can be more focused for the analytes of interest.
- Some pretreatments include:
  - Spectral Smoothing
  - Derivatives (1^{st}, 2^{nd}, 3^{rd}, 4^{th}) with gap-segment or SG methods.
  - Multiplicative Scatter Correction
  - Standard Normal Variate
  - Detrend
  - Orthogonal Signal Correction
  - Baseline correction
- These pretreatments can be used alone or in combination.

NIR Spectra for Lignocellulosics

With the support of the Seventh Framework Programme.

- The figure below shows some of the major overtone and combination regions between 1100-2500 nm for a typical agricultural feedstock (analysed dry).



DIBANET

Important Band Assignments

With the support of the Seventh Framework Programme.

Type of material	Test band assignment	Wavelength (nm)
Cellulose	O−H str. first overtone	1490
	C−H str. first overtone	1780
	O−H str. and second over C−O str.	1820
	C−H str. and C−H deform	2335
	CH_2 sym. str. and $=CH_2$ deform	2347
	CH_2 deform. second overtone	2352
	CH str./C−C str. comb.	2488
Lignin	C−H str. second overtone	1170
	O−H str. first overtone	1410
	C−H str. combination	1417
	O−H str. first overtone	1420
	C−H str. combination	1440
	C−H str. first overtone	1685

Effect of Water

With the support of the Seventh Framework Programme.

- Variations in hydrogen bonded molecular subspecies can cause band broadening and peak position shifts.
- Variations in hydrogen bonding (intra- and intermolecular) result in changes in the force constants of X-H bonds.
- The largest change in the force constants of the fundamental vibrations occurs for the X-H stretching vibration. It is known that in the case of O-H vibrations the apparent band location is isolinear and changes position linearly with temperature.
- This information has important implications for agricultural products. The water bands at 1450 and 1930 nm consist of multiple overlapping bands. The apparent location of these composite bands changes as the spectra are measured from one sample to another.
- This apparent shift is due to changes in the relative proportions of the individual bands making up the composite bands.

Application of NIRS in DIBANET

With the support of the Seventh Framework Programme.

- DIBANET targets the development of robust and accurate calibration equations for material that is **wet and of a heterogeneous particle size**.
- The vast majority of existing publications regarding the use of NIRS for the characterisation of lignocellulosic materials have focused on dry samples that have been ground down to a homogeneous particle size (< 1mm).
- They have preferred this method because the presence of water can mask the absorbance signal of some of the analytes of interest.

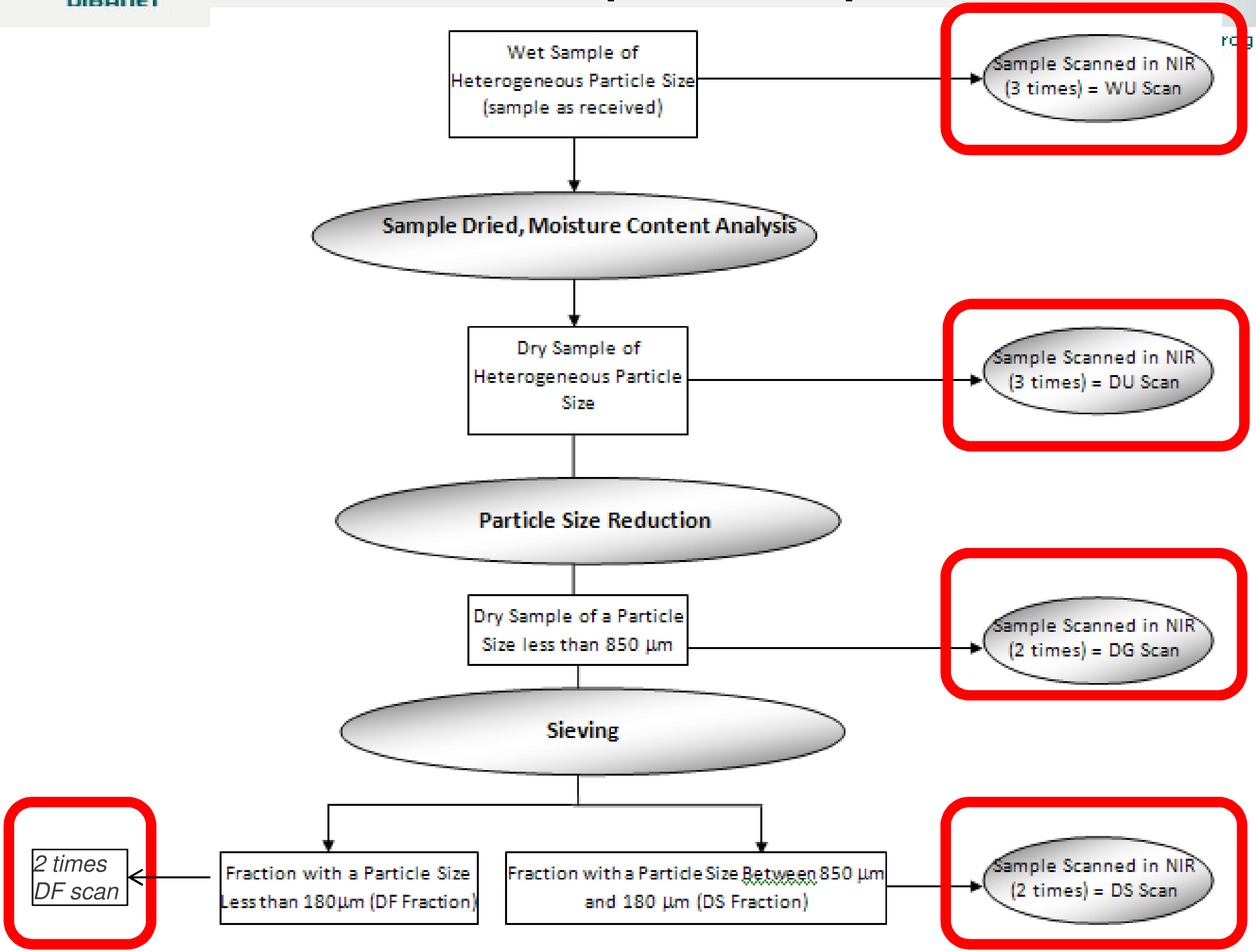
DIBANET

With the support of the Seventh Framework Programme.

NIRS Used for Samples in Various States

- WU (Wet Unground) – Feedstock not dried and minimally processed.
- DU (Dry Unground) – After WU sample has been air dried.
- DG (Dry Ground) – DU sample is comminuted until all is less than <850 υm
- DS (Dry Sieved) – DG sample is seived and DS is the fraction with a particle size 180 υm < x < 850 υm. DS is the fraction used for lignocellulosic analysis in the lab.
- DF (Dry Fines) – The fraction of DG that is < 180 υm

Sample Preparation



Strategy to Develop Good Calibration

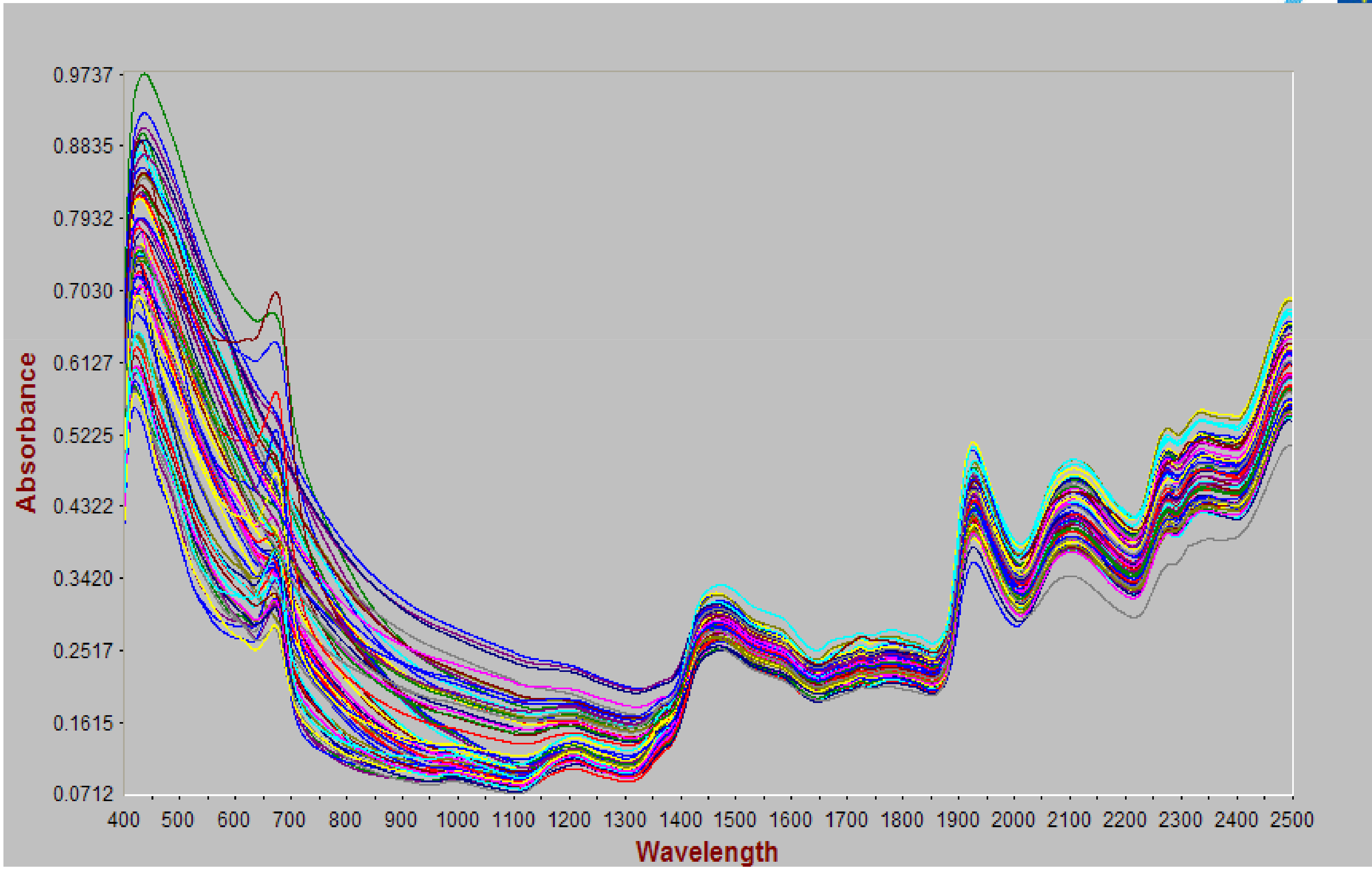
With the support of the Seventh Framework Programme.

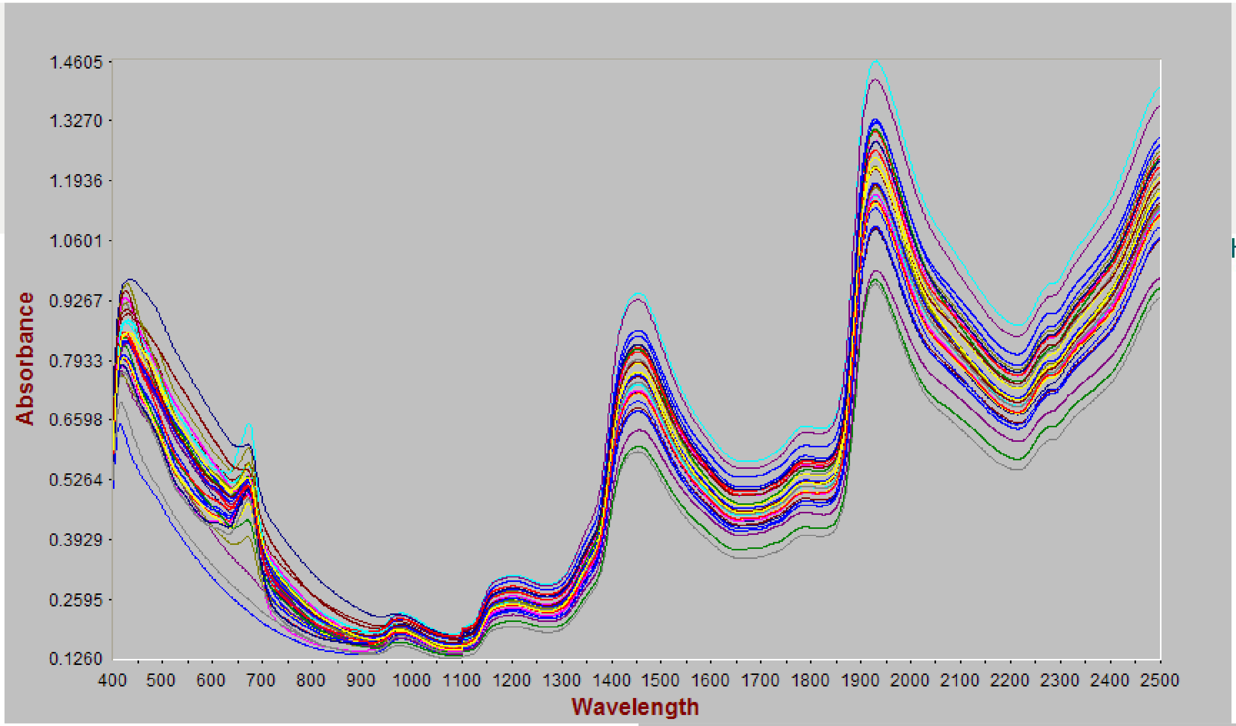
- University of Limerick - focus on miscanthus.
- There are significant physiochemical differences between the various anatomical parts of the plant.
- Incorporating the analytical results for these separate parts into a global analytical equation could improve its predictive ability and also ensure a wide component range exists, allowing good NIRS models with high Range/SEV ratios to be developed.

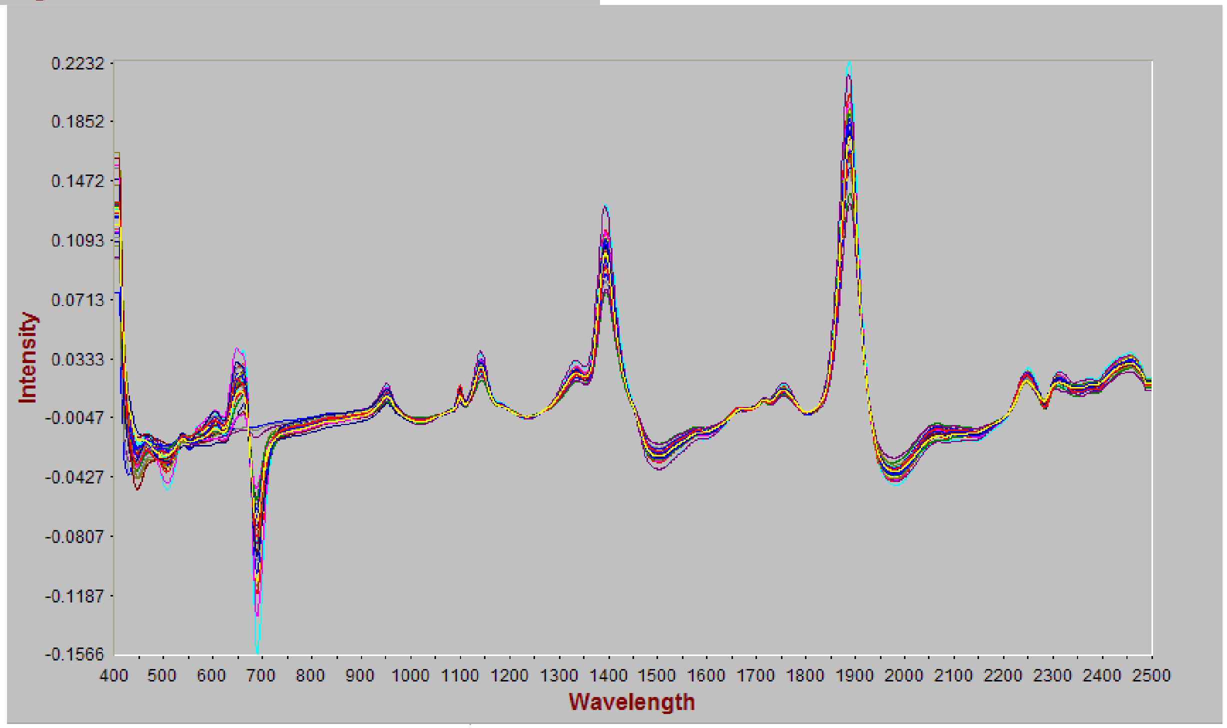
Accurate Analysis is Vital

With the support of the Seventh Framework Programme.

- DIBANET will aim to maximise the Range/SEV ratio for all the lignocellulosic components of interest.
- This necessitates good lab-data – hence strict quality controls are in place at UL – analysis is repeated if the SD between duplicates is greater than 0.25% for the lignin content and greater than 0.3% for the cellulose content.
- Inter-batch analysis variability needs to also be minimised.

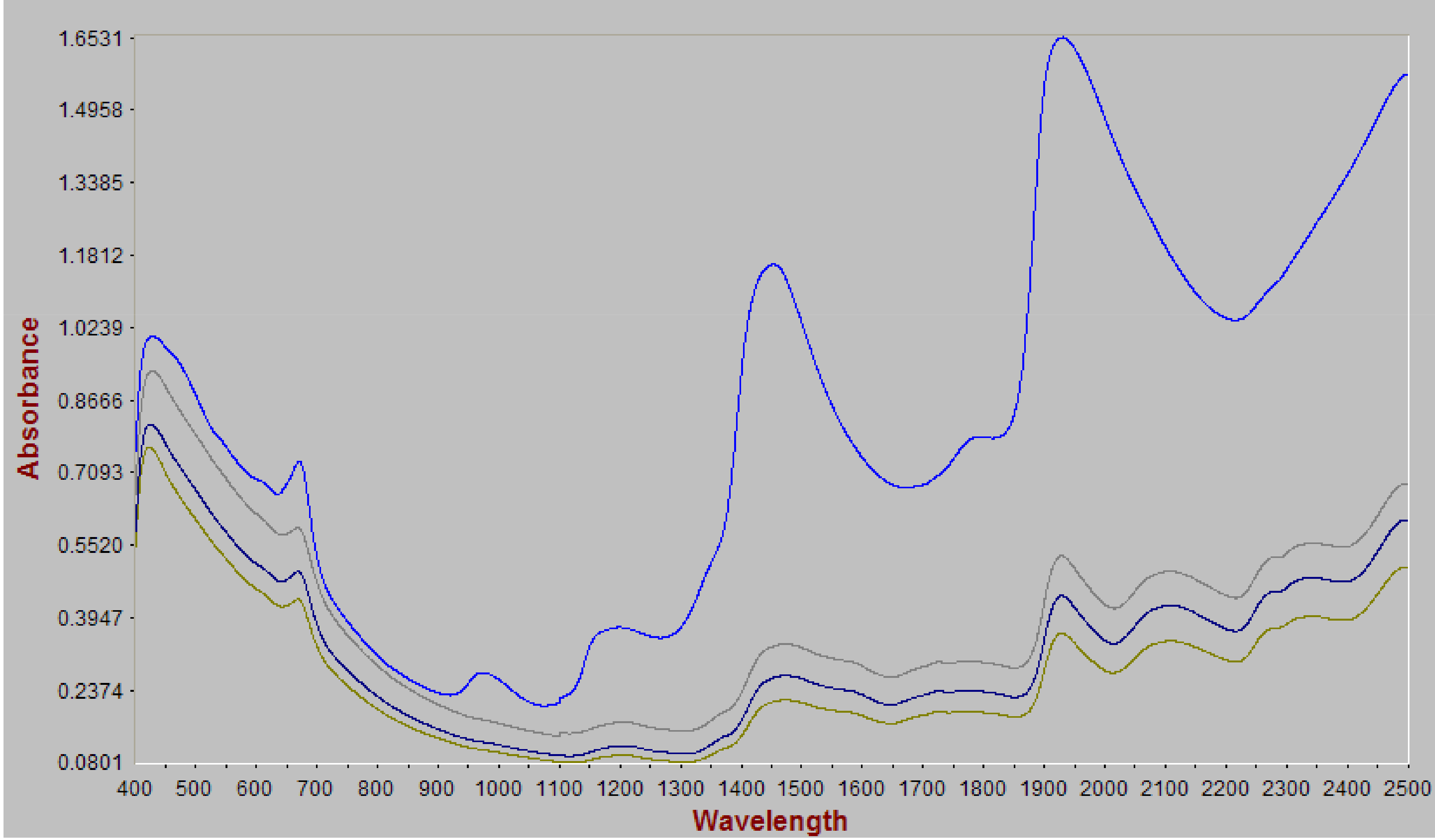

Absorbance
0.9737
0.8835
0.7932
0.7030
0.6127
0.5225
0.4322
0.3420
0.2517
0.1615
0.0712
400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 2400 2500
Wavelength


Absorbance
1.4605
1.3270
1.1936
1.0601
0.9267
0.7933
0.6598
0.5264
0.3929
0.2595
0.1260
400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 2400 2500
Wavelength


Intensity
0.2232
0.1852
0.1472
0.1093
0.0713
0.0333
-0.0047
-0.0427
-0.0807
-0.1187
-0.1566
400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 2400 2500
Wavelength

Scans of The Same Sample in Different States

With the support of the Seventh Framework Programme.



Future Application

With the support of the Seventh Framework Programme.

- Develop calibrations for yields of ethanol, levulinic acid etc. based on feedstock spectra and process efficiencies
- Link to projected feedstock value, other economic indicators.
- DIBANET project involves the use of online NIR equipment at a Brazilian sugar-mill.
- Calibration equations will be transferred and tested at this facility.
- Observations about biorefinery outputs and designs made.