Prediction of single kernel wheat and triticale protein, moisture and kernel hardness – highlighting the applications of NIR hyperspectral imaging

The **purpose** of my MSc project was to predict the protein, moisture and kernel hardness of wheat and triticale single kernels by using the Short Wave Near Infrared Hyperspectral Imaging (SWIR-HSI) system in the vibrational spectroscopy unit.

This **technique involves** imaging samples with a range of each of the properties of interest to obtain a spectral signature for each sample. The samples are then analysed to obtain the true chemical and physical properties. By combining the spectral information and true values or properties and applying data science methods, models can be developed to predict these properties in new samples by collecting their spectra and applying the models.

The **advantage** of this method is that it is non-invasive and non-destructive as it only collects the spectra. The samples can therefore be used for other purposes. The method is a rapid technique. Several models can be developed and applied to the same sample simultaneously to predict different properties. Application of this technique can save time and costs for chemical analysis.

The development of models for the prediction of protein, moisture and kernel hardness of wheat and triticale seeds are of **value** to the wheat and triticale breeding sector as seeds with favourable properties can be identified and re-sown.

METHODOLOGY

Collecting images using the SWIR camera

During the study, a small sample set consisting of 5 g seeds were imaged using the SWIR camera. The spectral imaging setup used in this study is pictured in Fig. 1 and it shows the instrument under operational conditions. The technique utilises light in the near-infrared (NIR) region in the 950 – 2500 nm spectral range and collects the absorption spectra. The protein, moisture and kernel hardness of individual seeds were determined using standard laboratory methods. Calibration models for each of the properties were developed using Evince (Prediktera), MatLab (Mathworks) and the PLS toolbox (Eigenvector) software.

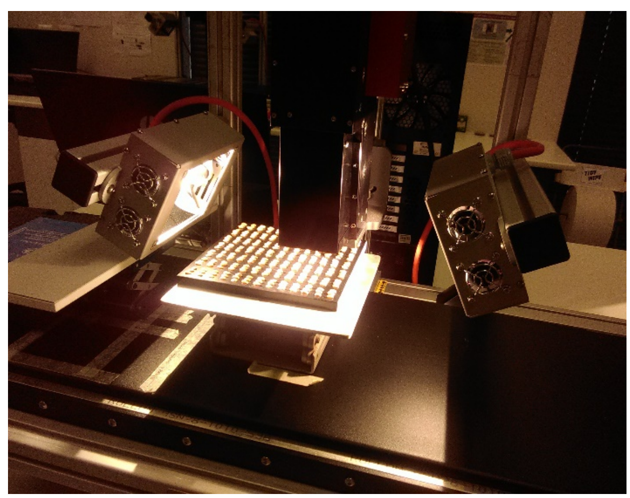

Figure 1 The HySpex SWIR-HSI camera when operational, showing the source of incident radiation (light), the object of analysis (kernels) and the detector. Note that the unit is fitted with a translation stage which moves the sample past the camera. Images are collected according to the pushbroom method which involves scanning and collecting data line-by-line to generate the image.

Using NIR data for model development

The high dimensionality of the data makes spectral imaging a rather powerful technique. The 3-dimensional dataset or 'hypercube' which is obtained is a combination of the spatial (pixel position) and a wavelength dimension (near-infrared light). The dataset can be unfolded so that every pixel in the image has a corresponding spectral absorbance signature as measured by the instrument detector.

The spectral information is extracted from the images using computational software. After extraction, image cleaning takes place in order to only specify the object of interest – in this study, the spectral signature of only the wheat and triticale kernels were considered for predictive modelling. In Fig. 2 the process of removing the background to retain the spectra of the kernels is shown graphically.

Next, two methods of analysis can follow from here, one can either go for a 'pixel wise' approach which would enable prediction of chemical and physical differences within a sample (requires lots of computational power). Alternatively, one can average the spectra of each object and extract the averaged spectra, this approach is useful for predicting differences between samples (requires less computational power).

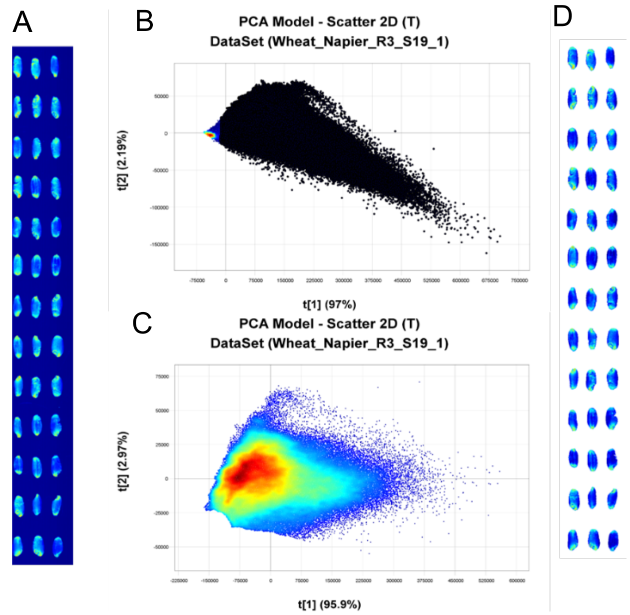

Figure 2 The figure shows the procedure of removing background spectral data from an image (A) by selecting pixels in a PCA score plot (B) which correspond to the background. Exclusion of these pixels creates a new PCA plot (C) and cleaned image (D) only containing the spectra of the objects of interest which can be used for further multivariate modelling.

After image cleaning the spectral and chemical reference data is used to optimize the model by applying a plethora of spectral processing and information extraction techniques to obtain a suitable calibration model for qualitative or quantitative prediction of specific properties. Finally, an external dataset is used to test the accuracy of the model.

Once a suitable model is developed, information such as shown in Fig. 3 can be extracted from the spectral data. The graph indicates which variables (wavebands) are the most important for predicting the properties of interest in this case the protein, moisture and kernel hardness content. This information can further be used to build models and instruments utilising only the wavebands of utmost importance. In so doing spectral filters can be used in order to be selective towards the specific wavebands of interest. This would create a computationally less complex model and more affordable equipment which is also attractive for industrial applications.

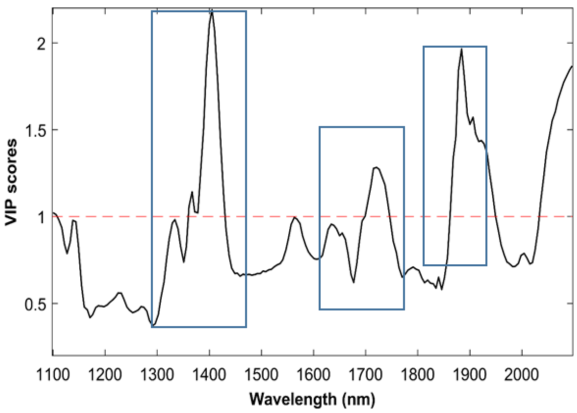

Figure 3 The figure highlights the variables which are the most important for partial least squares regression modelling for protein, moisture and single kernel hardness prediction.

Outcome of this study

For the study I undertook to use two variations of the partial least squares regression algorithms namely SIMPLS and a robust SIMPLS (RSIMPLS) method. I also combined the spectral data of both wheat and triticale single kernels into one data set, showing that prediction of chemical properties in cereal grains of similar type can be achieved. Furthermore, the study showed good robust regression results when the RSIMPLS algorithm was applied to the data set, which was more tolerant to analysis of samples which were atypical (outliers). As with all modelling using biological samples the study showed that seasonal and also geographical variation needs to be added to the calibration set for more accurate and precise predictive modelling. To conclude spectral imaging is a powerful tool that can be used to exploit the wonders of light for non-invasive predictive modelling of chemical and physical properties of samples once the models have been calibrated for with proper spectral and reference data.

Sebastian Orth

(MSc Food Science (*cum laude*))