

Fast and reliable quality analysis

On wheat flour protein, moisture, and ash values using FT-NIR

Keywords

TQ Analyst, Ash, Flour, FT-NIR, Moisture, Partial Least Squares, Protein

Introduction

Protein, moisture, and ash content are the principle parameters measured to determine wheat flour quality. These parameters are commonly quantified using either primary methods or Fourier transform near-infrared (FT-NIR) spectroscopy. Replacing the primary methods with FT-NIR provides faster results and accurate quantification, guaranteeing flour meets specifications. Rapid quality testing in the flour milling industry is critical for maximizing production, monitoring extraction efficiency, and producing flour to supply the ever increasing worldwide food demand. In addition, quality testing data is used to segregate flour for its intended use and to meet precise customer specifications. The following section discusses the primary methods commonly used to quantify protein, moisture, and ash content of wheat flour.

Protein content is the basis for judging flour quality and is vitally important to its functionality and finished-product attributes. For example, low protein content is desired for crisp or tender products, such as snacks or cakes, and high protein content is desired for products with a chewy texture, such as breads. The two primary methods for determining protein in flour or wheat are the Kjeldahl and Dumas methods.¹ Both of these methods require lengthy sample preparation, time consuming analysis and, in the case of the Kjeldahl method, involve the use of caustic and toxic chemicals.

Moisture content is important for shelf-life and storage. Very high moisture content (greater than 14.5%) attracts mold, bacteria, and insects, all of which can result in storage issues



Thermo Scientific™ Antaris™ II FT-NIR Analyzer with sample cup spinner accessory

or baking quality deterioration. The primary method for determining moisture content is weight loss by oven drying which requires multiple steps and several hours for results.

Ash is a measure of mineral content and is used to grade flour into different varieties. For example, whole wheat flour has a higher ash level than white flour. By quantifying ash levels during processing, flour millers can maximize extraction efficiencies and optimize blending. For bakers, ash content provides information on both finished product color and flavor. The primary method for determining ash content is gravimetric combustion. This method can be quite lengthy, taking several hours or overnight to complete.

Key Metric	Dumas or Kjeldahl	Combustion	Loss on Oven Drying	FT-NIR
Sampling	Destructive	Destructive	Non-destructive	Non-destructive
Sample Preparation	Yes	Yes	Yes	No
Actual Analysis Time	2–12 hrs	Several hours or overnight	2–4 hrs	<20 seconds
Data Turnaround Time	Hours/Days	Hours/Days	Hours	Minutes
Consumables Cost Per Test	High	Medium	Low	None
Technical Analyst	High	Low	Low	Low
Location of Analysis	Lab	Lab	Lab	Lab or Process Floor
Hazardous Chemicals	Yes	No	No	No
Hazardous Temperatures	Yes, 900 °C for Dumas, 420 °C for Kjeldahl	Yes, 585 °C	Yes, 130 °C	No
Component	Protein	Ash	Moisture	Protein, Ash, Moisture, other

Table 1: FT-NIR advantages vs. primary methods

As summarized in Table 1, the aforementioned primary methods require trained analysts, involve the use of toxic chemicals, entail time consuming sample preparation, and are burdened by time delays waiting for results. In contrast, flour samples can be analyzed by FT-NIR without sample preparation or the use of chemicals, providing faster, more accurate, and reliable results. FT-NIR analysis also provides immediate return on investment by eliminating the cost of consumables, rework, or discarding of product that does not meet specifications. In this application note, we will demonstrate the successful quantitative analysis of wheat flour components including ash, protein, and moisture using the Thermo Scientific Antaris II FT-NIR analyzer.



Figure 1: Spinning cup accessory

Experiment

Development of the calibration models for quantifying the wheat flour components of protein, moisture, and ash began with the collection of the

flour samples from multiple production lots. All samples were held at room temperature and analyzed by diffuse reflection on an Antaris II FT-NIR system equipped with an integrating sphere and a 5 cm spinning cup accessory, as shown in Figure 1. The spinning cup accessory rotates the sample cup filled with the flour over the integrating sphere window while spectra are collected and co-averaged. This provides a single spectrum representing the average diffuse reflectance of the sample for model development. Thirty-two co-averaged scans were collected at 8 cm⁻¹ resolution in less

Component	Method	Parameter
Ash	Loss on combustion at 500 °C	Percent of inorganic material
Protein	Dumas	Percentage of Nitrogen
Moisture	Oven loss on drying at 130 °C	Percent weight of moisture

Table 2: Primary reference methods used for calibration development

than 20 seconds. Primary reference data for each component were obtained from the methods listed in Table 2.

Individual PLS models for the quantitative analysis of ash, protein, and moisture were developed using Thermo Scientific™ TQ Analyst™ Software. PLS models use a statistical approach to quantitative analysis by examining the selected region or regions of the standard spectra to determine which areas vary statistically as a function of component concentration.² PLS was chosen for this analysis because it can account for broad or overlapping peaks which are typically encountered in flour spectra. The concentration ranges for the components quantified are listed in Table 3.

Components	Low	High
Ash	0.27 %	2.12 %
Protein	6.3 %	17.3 %
Moisture	9.0 %	15.2 %

Table 3: Concentration ranges for ash, protein and moisture in ground wheat flour used for calibration development

All three models used Standard Normal Variate (SNV) pathlength treatments to mitigate spectral baseline shifting due to variation in flour particle size and sample cup packing. To enhance spectral features, protein standard spectra were pretreated using a first derivative with Norris smoothing at a segment length of 5 and gap between segments of 5, see Figure 2. Ash and moisture standard spectra were pretreated using a second derivative to enhance spectral features as well using Norris smoothing at a segment length of 3 and gap between segments of 3.

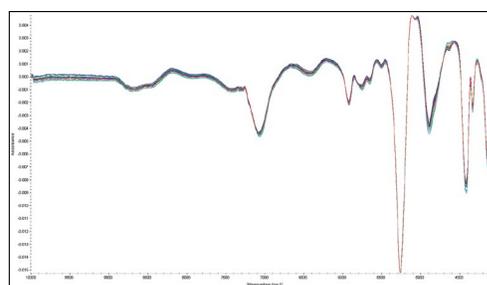


Figure 2: SNV with first derivative treated spectra of flour protein standards used for PLS model development

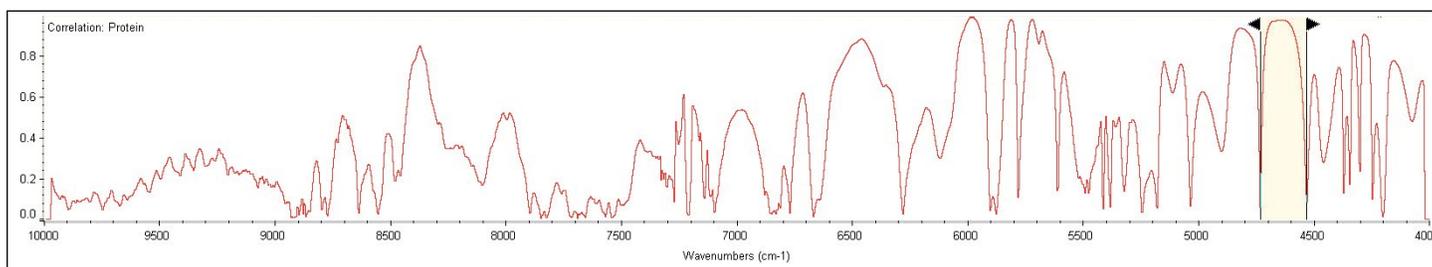


Figure 3: Statistical spectra showing the correlation of spectral variation to changes in component concentration. The region highlighted showing a correlation greater than 0.90.

Spectral analysis regions for ash and moisture models were determined using the Suggest Regions wizard in TQ Analyst. This easy-to-use tool automatically chooses the appropriate spectral regions for analysis. Since region selection is often an iterative process, the Suggested Regions wizard provides a good starting point for model optimization. Manual selection of regions based on pretreated spectra is also easily accomplished using TQ Analyst. For ash and moisture, the Suggested Regions tool was able to provide the best model results. The spectral range suggested for ash was 9,895.12 – 4,053.59 cm^{-1} and for moisture the spectral range was 9,895.12 – 4,053.59 cm^{-1} . The number of calibration standards used were 564 calibration and 68 validation standards for the ash model and 550 calibration and 70 validation standards for the moisture model. Calibration standards are used during calibration model development to relate the variation in spectral features to component concentration. Validation standards are not used in the calibration but are used to provide an unbiased test of calibration model performance.

In contrast, a more advanced visual development tool, Statistical Spectra, was used to optimize the protein model. This tool generates spectra showing the correlation of spectral variation to changes in component concentration, see Figure 3. The correlation of the spectral regions to component concentration ranges from zero to one with one being perfect correlation. This tool also provides both development flexibility and model optimization capability by helping with manual selection of the regions that correlate highest with the changes in component concentration. The statistical spectra tool identified nine regions with high correlation, five peaks above 0.9 and four peaks above 0.8, for model development on 593 calibration and 86 validation standards.

PLS Model	Factors	RMSEC	Correlation Coefficient	RMSECV	RMSEP
Ash	6	0.032	0.989	0.042	0.052
Moisture	5	0.081	0.992	0.096	0.099
Protein	4	0.114	0.998	0.116	0.153

Table 4: Summary of Calibration results for PLS Models

Results and discussion

All three PLS models developed show low Root Mean Square Error of Calibration (RMSEC) and good correlation to the primary method data while using relatively few factors. The correlation coefficient and RMSEC are measures of how well the component concentrations of the calibration standards are predicted by the calibration model. Ideally, the correlation coefficient should have a value close to one, and the RMSEC approaching the standard error of the primary technique. The Factors represent independent sources of variation condensed from the concentrations and spectral information and are ranked by the amount of variation in the data that they explain. Table 4 shows how well the PLS models accurately quantify the components ash, protein, and moisture in wheat flour. In addition, comparing the Root Mean Square Error of Prediction (RMSEP) values to the RMSEC we see a good indication of how the model accurately predicts samples not in the calibration. The RMSEP is computed with the independent set of validation samples that were withheld from the calibration. Another test of model robustness is the Root Mean Square Error of Cross Validation (RMSECV). This diagnostic sequentially removes a specified number of standards from the calibration set, calibrates the method, and then uses the new calibration model to quantify the standards that were removed from the calibration set. This is repeated until all of the standards in the calibration set have been quantified as validation standards. A good measure of model accuracy is for both the RMSECV and the RMSEP to be less than two times the RMSEC. This is the case for all three PLS models discussed here. The following section highlights many of the features provided in TQ Analyst for region selection, diagnostics, and outlier removal that were used for optimizing all three PLS models. We will examine the protein method in detail to demonstrate the tools available in TQ Analyst for Chemometric model development.

The calibration curve for protein in Figure 4 demonstrates very good correlation between the calculated (FT-NIR) to the actual (Dumas) values with a correlation coefficient of 0.998 and an RMSEC of 0.114.

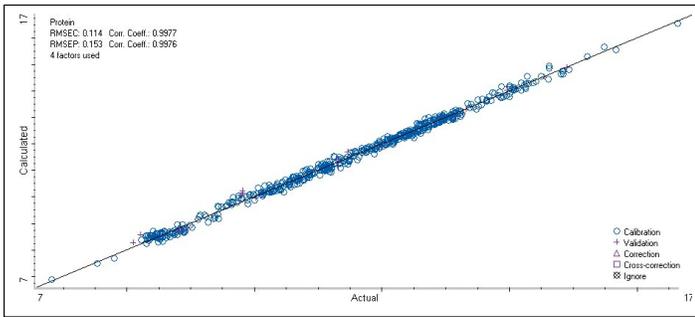


Figure 4: Calibration curve for protein with independent validation (+) samples

If validation standards are selected for the model, the RMSEP is displayed along side the RMSEC on the calibration curve. This provides a quick check on model performance.

Figure 5 is the Residual (percent difference) plot for protein and shows the predicted error distribution of the calibration standards. This plot shows the differences between the calculated and the actual concentration values relative to the actual values.

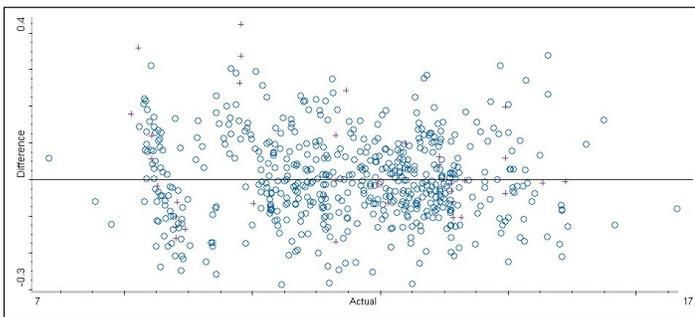


Figure 5: Residual (percent difference) plot for protein

When a model is calibrated in TQ Analyst the residual plot is placed side by side to the calibration plot providing a complimentary diagnostic tool for spotting slope issues and potential outlier samples. Here we expect to see a random distribution of difference values, without any trends or slopes in the values. An examination of Figure 5 demonstrates that the error differences are evenly distributed across the whole component range indicating equal distribution of error for both calibration and validation standards.

The Predicted Residual Error Sum of Squares (PRESS) plot (Figure 6) shows the ranking of the factors and the associated variation. Each time a factor is added that represents useful information to the calibration model, the RMSECV and the PRESS values decrease. The information provided by Figure 6 shows that Factors 1 – 4 explain the

majority of the observed spectral and concentration variation. A well built model will demonstrate a decrease in the amount of error (RMSECV) with successive factors used, as this model does. The PRESS plot diagnostic is used to determine the optimum number of factors for the model. A minimum number of factors were used in the model to avoid over-fitting the data which would result in the model performing poorly with samples not in the model.

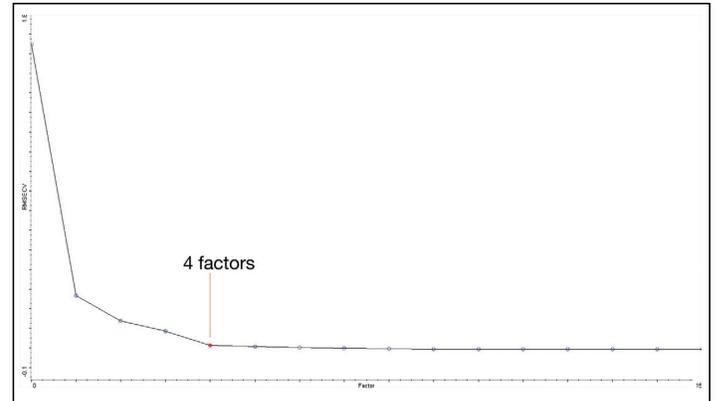


Figure 6: PRESS plot for protein PLS model

TQ Analyst provides both 2 dimensional (2D) and 3 dimensional (3D) principle component scores plots as visual tools for identifying outliers, trends, or patterns. Similar to PLS factors, Principle Components (PCs) are ranked by the amount of spectral variation they describe. However, unlike factors used in PLS calibrations plots, PCs do not factor in the concentration values from the standards. The first principal component describes the most spectral variation, with each subsequent PC describing the remaining variation. Both 2D and 3D PC Scores plots were used for the development of the protein model to identify outliers and trends in the spectra. The 2D scores plot for PC2 vs PC1 in the protein model (Figure 7), shows random distribution of the standards. The plot is ideal for PLS model development since it does not show grouping, trends or outlier standards. This indicates that the spectral variation is random. TQ Analyst also provides the ability to change the usage of individual standards within the principal component scores plot window, reducing development time when optimizing the model. By right clicking on the standard of interest its usage can be toggled between validation, ignore, or calibration, as shown in Figure 7.

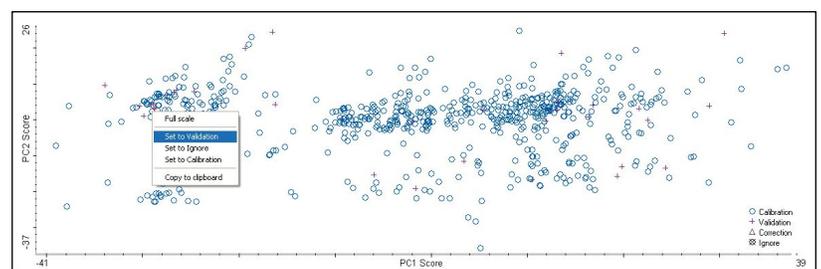


Figure 7: 2D Principal component scores plot for protein standards

The use of 3D scores plot adds a third dimension to the Principal Component Scores Diagnostic by displaying three principal components simultaneously. The 3D modeling in TQ Analyst has features such as the interactive zoom and the ability to rotate the scores plot 360 degrees which greatly improves the ability to see patterns, and ability to display onscreen sample information for efficient data mining. The protein 3D PC scores plot for the first three PCs is displayed in Figure 8, showing TQ Analyst ability to display useful sample information by left clicking individual points.

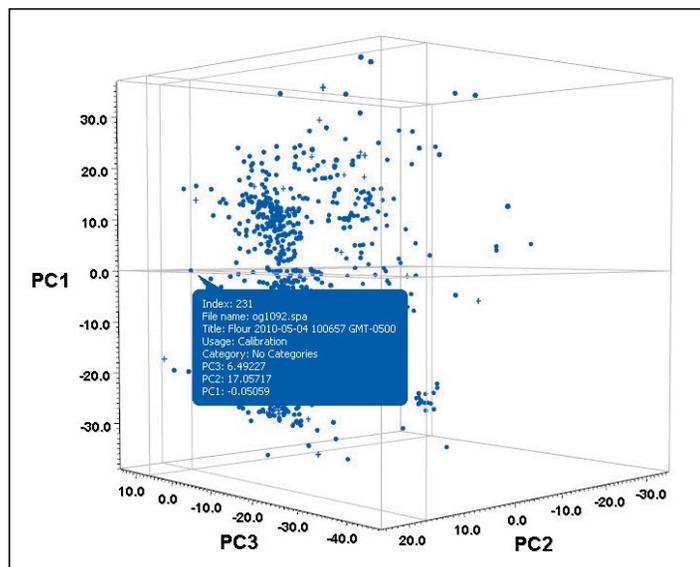


Figure 8: 3D Principal component scores plot for protein standards

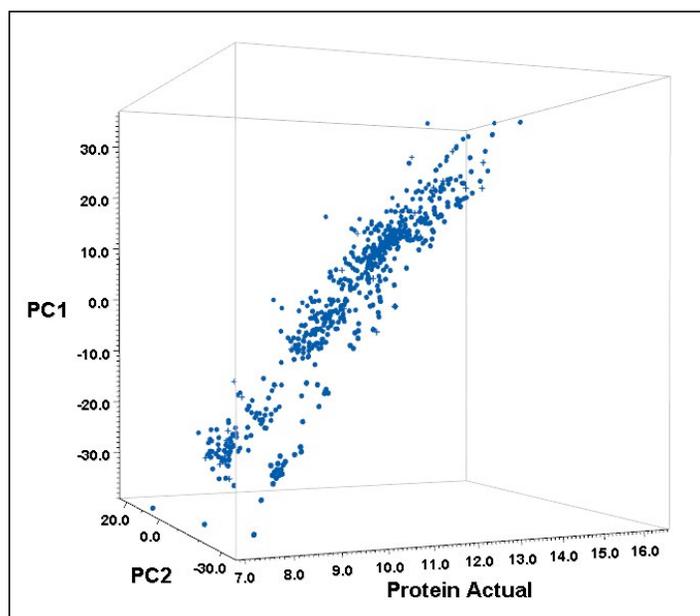


Figure 9: 3D Principal components scores plot showing high correlation between PC1 and protein actual values

TQ Analyst can also plot the actual protein concentrations within the 3D scores plot. This feature allows the analyst to visually explore and quickly determine which PC(s) are most highly correlated to the parameter(s) of interest. A robust model will have PC1 highly correlated to the parameter of interest indicating that the variation explained by the 1st factor in the PLS model is due to the parameter of interest. Figure 9 shows high correlation of PC 1 to the protein concentrations indicating a robust PLS model.

Conclusion

This calibration study has demonstrated that the Thermo Scientific Antaris FT-NIR analyzer provides a rapid solution to accurately quantify the key flour components of moisture, protein, and ash. TQ Analyst offers easy and intuitive PLS calibration optimization and development through its visual and interactive diagnostic tools, such as Statistical Spectra, Residual, PRESS, and 2D and 3D PC scores plots. The traditional quantitative techniques for flour analysis require trained analysts, the use of chemicals, may entail time consuming sample preparation, and are burdened by time delays waiting for results. Alternatively, FT-NIR can analyze flour without the use of chemicals, providing faster, highly accurate, and reliable results. It also provides short term return on investment by eliminating the cost of consumables, rework, or discarding of product that does not meet specifications. In addition, implementing FT-NIR to replace traditional techniques reduces potential delays waiting for test results needed to make critical production decisions. The speed and accuracy of FT-NIR provides quality data for real-time process improvements, and allows flour millers and buyers to verify that their flour specifications are met, thereby maximizing production efficiency and profitability.

References

1. Nielsen, S. Suzanne. Food Analysis (Food Science Text Series), (2010) Springer Science Business Media, ISBN 978-1-4419-1477-4
2. Kramer, R. Chemometric Techniques for Quantitative Analysis, (1998) Marcel-Dekker, ISBN 0-8247-0198-4



Find out more at thermofisher.com/nir

ThermoFisher
SCIENTIFIC