Technical Note 32: Software Improvements for the Cropscan 1000B Whole Grain Analyser

Introduction:

The Cropscan 1000B has been developed to provide a robust and reliable whole grain analyser that uses a flow through sampling system. The specification for precision for this instrument in measuring wheat should be a Standard Deviation of 10 consecutive analyses of 0.15% for protein and moisture in wheat and barley.

Recently barley samples which exhibit high absorbance levels due to weather damage have exhibited higher errors in replicate analyses. The higher the absorbance, the larger the errors and the worse the precision.

In order to remedy this problem, a thorough investigation has been undertaken to examine the parameters that effect precision in the analysis of protein and moisture in grains.

Description:

The Cropscan 1000B uses a 38 silicon photodiode array detector and flat field spectrometer with an integrating amplifier and A to D converter to generate the NIT spectrum of grains. The integrating amplifier collects and stores the energy passing through the sample cell and optics and reaching the detector. The amount of time that the energy is integrated is called the Integration Time, ie, Int_time. This time period varies from 78 micro seconds to 80000 micro seconds depending upon the amount of energy reaching the detector. The more light passing through the optics, the shorter the integration time required to achieve a reliable signal.

The Cropscan 1000B is controlled by a PC board that is programmed to collect the sequential readings from each detector element, called a pixel. The first step is to collect the 0% noise energy curve, then turn on the lamp, wait for 15 seconds for the lamp to stabilise and then collect up to 30 scans of the sample as it is metered through the flow through sample cell. The software has an option called AdaptScan which sets up the electronics to measure the first scan of the sample and decide which is the optimum integration time for each pixel. This is done to make the scanning faster and to optimise the integration time so that samples with low absorbance levels use short integration times where as high absorbing samples use longer integration times. Alternatively the AdaptScan parameter can be set to 0 whereby the integration time is fixed across the entire spectrum. The integration time needs to be set short enough so that the amplifier circuitry does not saturate, ie, overfill with energy, or long enough to collect a stable spectrum.

To fully understand the implications of measuring protein and moisture in whole grains, it is important to go back to basics. The Cropscan 1000B measures three parameters in order to make a measurement, ie, 100% energy curve, noise or 0% energy curve and the sample energy curve. The analyser ultimately measures the amount of light that is absorbed by the grains as the light passes from one side of the cell to the other. The Absorbance is defined as;

Absorbance = Log((100% - 0%)/(Sample% - 0%))

Where:

100% = the amount of light that passes through the sample cell when no grain is present.
0% = the baseline electronic noise of signal that exists in the system when there is no light passing through the sample cell.
Sample% = the amount of light that passes through the sample cell when grain is present.

The 0% or noise is a very small signal and has been traditionally subtracted from the 100% and sample% readings as an issue of being as exact as possible. The 100% readings are very large and vary across the spectrum, ie, 720-1100nm. The sample% readings are small as compared to the 100% reading and vary across the spectrum and as well vary due to the physical characteristics of the grains, ie, size, shape, colour, weathering and type. The packing of the grains in the sample cell also cause variation in the sample% readings. Figures 1, 2 and 3, show typical 0%, 100% and sample% readings. Figure 4. Shows the final NIT spectra of the 15 sub scans collected in the Cropscan 1000B.

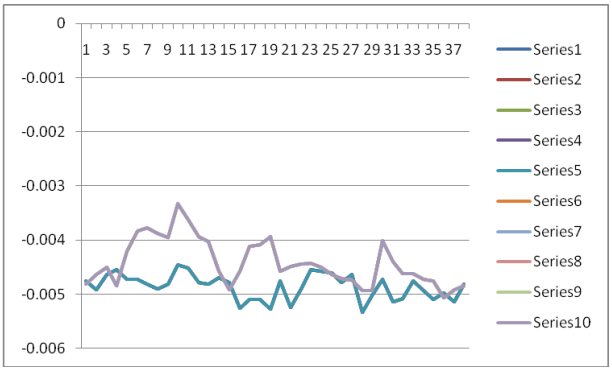

Figure 1. 0% Noise Curve

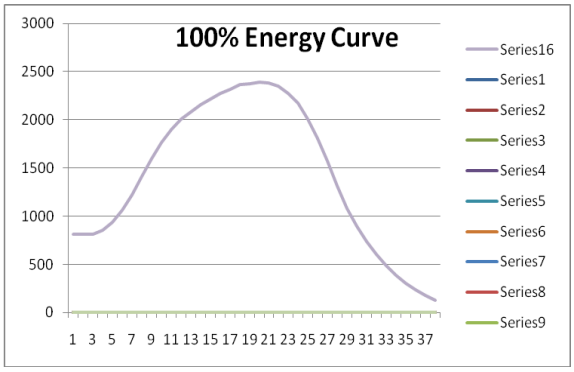

Figure 2. 100% Energy Curve

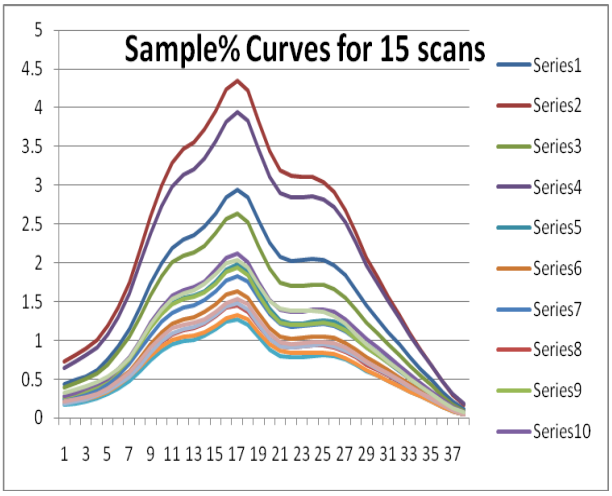

Figure 3. 15 Sample% Scans

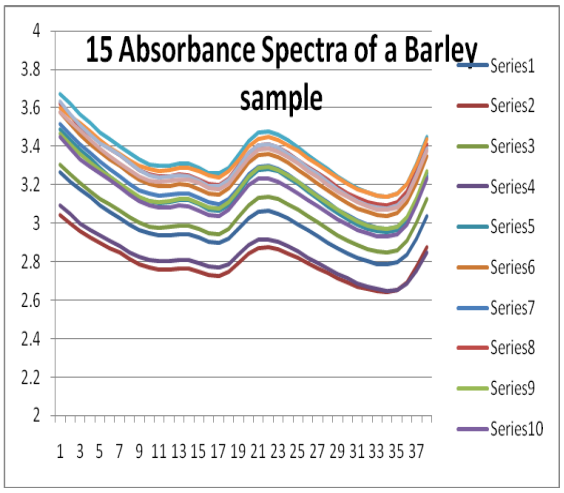

Figure 4. NIT spectra of barley.

To test the affects of the various parameters on the precision of the measurement, 1 sample bag of barley, termed GA18, was analysed 10 times using two fixed integration times, ie, 10000 usecond and 20000 useconds. The 0%, 100% and sample% readings for each of 15 sub scans were saved.

The data was imported into Microsoft Excel whereby the spectra for each sub scan were computed and the PLS model for protein was applied. Post spectral processing is normally averaging the 15 predicted protein. An outlier detection rountine called ZScore is used to identify if subscans are statisically different to the average spectrum and then reject then and recalculate the average.

The following combinations of setups and parameters were examined;

AdaptScan = 0, ie, Int-time: 10000 or 20000usec
IdleZero = 1 ie, 0% Included
IdleZero = 0 ie, 0% excluded
ZScore = 1.5
ZScore = 1.0
HundRef = 1 ie, Constant 100%
HundRef = 0 ie, Individual 100%

The Average, ZScore Average, and Median for each set of 10 analyses were computed and tabulated

Results:

Table 1. presents the predicted protein for Int-time fixed at 10000usec with 0% included.

GA18 10000usec IdleZero = 1

	1	2	3	4	5	6	7	8	9	10	STDev
Ave	9.32	9.44	9.49	9.66	9.43	9.58	9.38	9.13	9.24	9.77	0.19
SD	0.62	0.55	0.70	0.77	0.77	0.62	0.80	0.63	0.81	0.69	
ZScore=1.5	9.32	9.45	9.38	9.68	9.44	9.51	9.37	9.21	9.13	9.92	0.23
Median	9.45	9.52	9.52	9.67	9.63	9.67	9.27	9.23	9.15	9.99	0.25

Table 2. presents the predicted protein for Int-time fixed at 10000usec with 0% excluded.

GA18 10000usec, IdleZero = 0

	1	2	3	4	5	6	7	8	9	10	Stdev
Ave	9.32	9.44	9.49	9.66	9.43	9.58	9.38	9.13	9.24	9.77	0.17
ZScore = 1	9.48	9.29	9.38	9.40	9.43	9.24	9.22	9.39	9.27	9.57	0.09
ZScore = 1.5	9.32	9.38	9.31	9.52	9.43	9.44	9.39	9.37	9.29	9.85	0.07
Median	9.45	9.52	9.52	9.67	9.63	9.67	9.27	9.23	9.15	9.99	0.20

Table 3. presents the predicted protein for Int-time fixed at 20000usec with 0% included.

GA18 20000usec IdleZero 0n Noise Included

Ave	9.94	9.85	9.52	10.04	9.95	9.61	9.60	9.69	9.40	9.66	0.21
SD	0.41	0.50	0.63	0.52	0.57	0.51	0.51	0.66	0.46	0.56	
ZScore	10.00	9.79	9.51	10.10	9.90	9.59	9.63	9.79	9.33	9.58	0.24
Median	9.91	9.76	9.49	10.05	9.93	9.63	9.67	9.71	9.48	9.71	0.19

Table 4. presents the predicted protein for Int-time fixed at 20000usec with 0% excluded.

GA18 20000usec IdleZero 0n Noise Excluded											
Ave	9.72	9.63	9.52	9.85	9.76	9.41	9.45	9.55	9.40	9.66	0.15
ZScore = 1.0	9.67	9.56	9.48	9.89	9.70	9.54	9.48	9.49	9.44	9.58	0.14
ZScore = 1.5	10.00	9.79	9.51	10.10	9.90	9.59	9.63	9.79	9.33	9.58	0.24
Median	9.62	9.52	9.49	9.87	9.67	9.50	9.60	9.55	9.48	9.71	0.12

Note that the use of ZScore of 1.0 provides better exclusion than 1.5. Since 15 sub scans are being collected, the use of ZScore =1.0 results in between 2 and 6 scans rejected as outliers. Whereas ZScore = 1.5 with 10 sub scans allows the rejection of only 1 or 2 scans.

The best precision was obtained using a fixed Int-time of 10000usecs, Noise Excluded and ZScore = 1.0. However the data shows that there can be a step change in the predicted protein after a 5 or more repeat analyses. The step change in the data has been the unresolved problem in the measurement of barley. The question is what causes these step changes. To identify these step changes the GA18 10000 IdleZero = 0 spectra were more closely examined.

Table 5. presents the predicted protein data for several different setups.

1) Indivd Scans	2) Const Scan	3) Indivd Scans	4) Indivd Scans-Noise	5) 5 Scans, 5 100%	6) 5 Scans, 5 100%
Indivd 100%	Variable 100%	Const 100%	Indivd 100% - Noise	1st Noise	2nd Noise
9.5	9.6	9.5	9.8	9.78	9.60
9.7	9.7	9.6	10.0	9.99	9.81
9.6	9.6	9.7	10.0	9.95	9.77
10.0	9.6	10.0	10.2	10.18	10.05
9.5	9.6	9.5	9.8	9.85	9.65
9.8	9.6	9.8	9.9		
9.5	9.7	9.5	9.7		
9.2	9.7	9.2	9.4		
9.5	9.7	9.4	9.6		
9.9	9.7	9.8	10.0		
Stdev 0.22	0.03	0.24	0.23	0.16	0.18
Ave 9.6	9.7	9.6	9.8	9.9	9.8

1) Indivd Scan, Indivd 100%: Normal spectra = log(100%/Sample%)
2) Const Scan, Variable 100%: Used the average of all scans and changed the 100%
3) Indivd Scans, Const 100%: Used each scan and used the average of all 100%
4) Indivd Scan-Noise, Indivd 100%-Noise: Same as 1 but included Noise%
5) 5 Scans,5 100% 1st Noise%: First five scans minus 1st Noise.
6) 5 Scans, 5 100% 2nd Noise%: First five scans minus 2nd Noise%

Table 2. shows that by keeping the keeping the scan constant, the precision is the lowest, ie, 0.03%. All other combinations show that changing 100%, excluding or including noise% or changes in nosie% do not reduce the precision. The major factor that affects precision is the sample scans.

Figures 5, 6 and 7 show the 15 sample% scans for the three analyses that showed the biggest differences from the average of all 10 analyses.

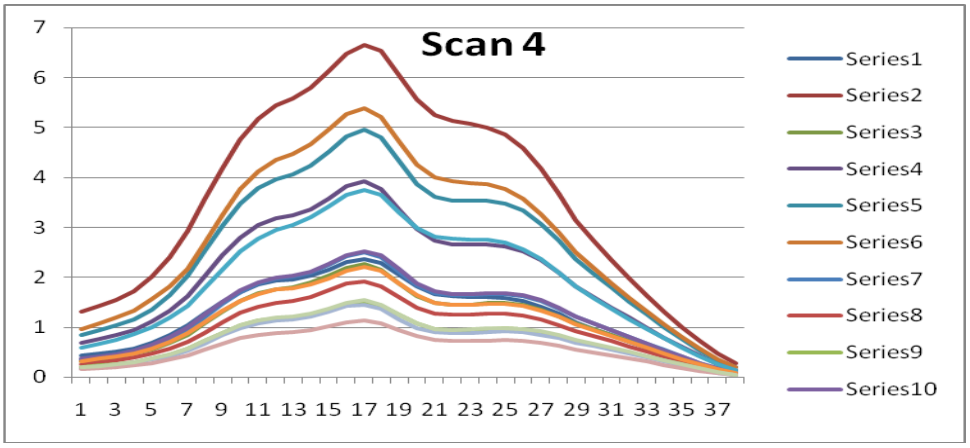

Figure 5. Plot of Sample% Scans for analyses 4.

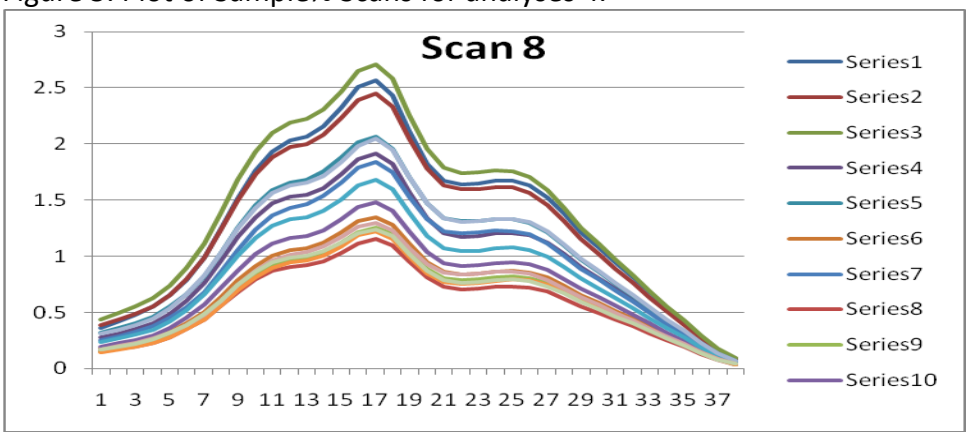

Figure 6. Plot of Sample% Scans for Analyses 8

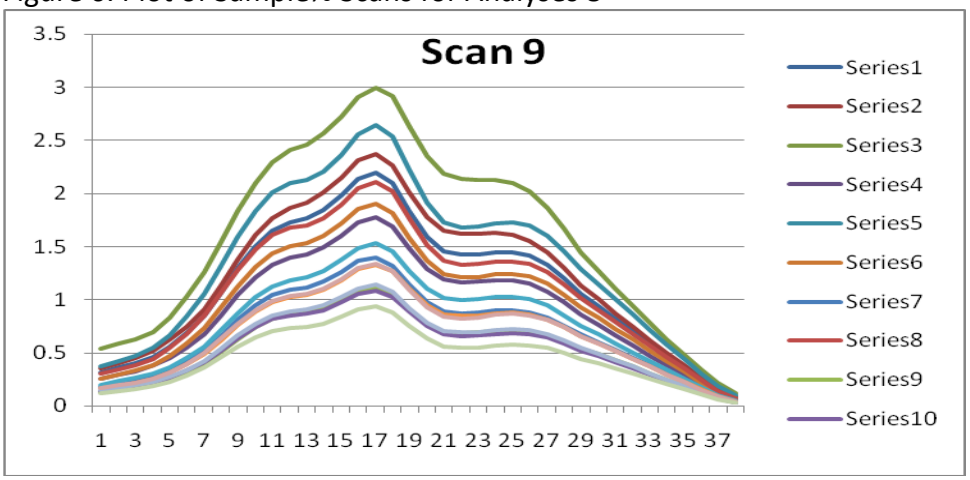

Figure 7. Plot of Sample% Scans for Analyses 9.

Figure 8. shows the plot of the Average Sample% Scans for the 10 analyses.

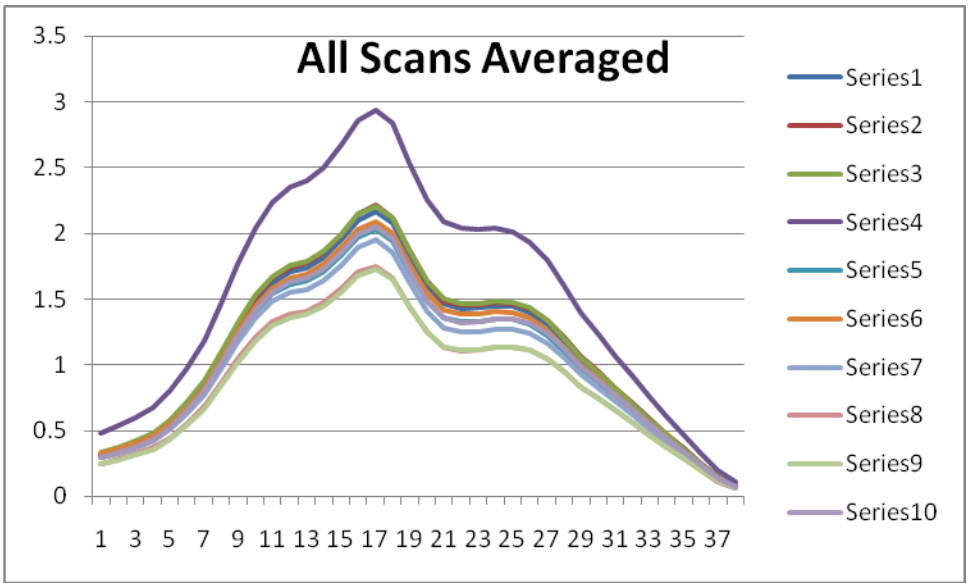

Figure 8. Plot of Average Sample% Scans for all 10 Analyses.

It was observed that the average sample% scans in figure 8. Correlate with the average results as shown in Table 2. The high sample% scan shown in Figure 8. Corresponds to analyses 4, ie, figure 6. The two lowest sample% scans corrspond to analyses 8 and 9, ie, figures 6 and 7.

Conclusions:

The data presented above is quite complex and difficult to explain. However there are several conclusions that can be drawn from the data.

1) Small changes in the 100% reading, ie, 2% or less, do not affect the predicted protein results.
2) Including or excluding Noise% in the calculation of absorbance has a minor affect, ie, 0.1% protein per 0.001% change in noise. Since the Noise% can be influenced by external factors, it would be better to exclude the Noise% from the spectrum.
3) Calculating the Average after using ZScore = 1.0 acts to remove significant outliers and thereby improve the precision. Using Average alone is the least beneficial. However the Median of the 15 scans is generally better than Average. Using ZScore = 1.5 is also beneficial but not quite as effective as 1.0.
4) Integration time is better left fixed rather than setting AdaptScan = 1. To accommodate wheat and barley, the Int-time = 10000usecs was chosen. At Int-time = 20000usecs, some wheat samples passed too much light through the sample and resulted in a saturation of the amplifier. Note that setting Int-time = 10000usecs did show the best precision.
5) Samples that are loosely packed exhibit the largest errors from the average. Samples that are tightly packed also exhibit the largest errors from the average. These observations are obvious, however the importance is that the data shows relationship between precision, average results and sample packing. As such, to improve precision, the only factors that can be optimised are sample packing.

The challenge is to develop a means of compacting the samples in the flow cell in order to present the sample more consistently to the analyser.