INTERNATIONAL WOOL TEXTILE ORGANISATION

TECHNOLOGY & STANDARDS COMMITTEE **CAIRO MEETING**

Raw Wool Group **May 2006**

Chairman: A.C. BOTES (South Africa) **Report No: RWG 02**

A Review of IWTO-19 Appendix K Requirements for the Measurement of Ash Content by Near Infrared Relectance Analysis (NIRA)

By

D. J. Petrie and J. W. Marler

Australian Wool Testing Laboratory Ltd. PO Box 190, Guildford, NSW 2161, Australia

SUMMARY

A review is conducted of the criteria laid out in Appendix K of IWTO-19 for the commercial use of NIRA to predict the ash content of scoured subsamples. The SDDV limit was established from New Zealand experience based on wools of relatively low (0-3%) ash contents compared to what is more common for Australian wools (99% are within the range 0 – 4% ash content while the final 1% can be up to 20% ash content). The research presented demonstrates that to overcome the under prediction of ash content at high levels of ash content, it was necessary to include a number of high ash content samples in the calibration set. Due to the nature and distribution of the contaminant that causes the high ash content, the variation between replicate subsamples drawn at random becomes large for samples with ash contents greater than about 4%. This in turn leads to higher SDDV values than the current limit of 30%. It is recommended that where calibration and validation data sets have ash contents that range 0-6% or higher that the SDDV Limit should be increased to 40%. The implications of this change to the Certified Wool Base are small and in part are dealt with by the Test Method requiring additional subsamples to be tested where differences beyond defined limits are exceeded.

INTRODUCTION

The prediction of Ethyl Alcohol Extractable Matter (LABaem) of laboratory-scoured cores by Near Infrared Reflectance Analysis (NIRA) represented the first routine commercial use of this technology governed by the requirements of IWTO-19 Appendix K. Appendix K was subsequently amended to allow the prediction of ash content (NIRash)[(1)]. Concurrently, a validation requirement based on the standard deviation of differences was introduced in addition to the requirements related to the Geometric Mean (GM) slope and Mean Difference. Currently the Standard Deviation of the Differences expressed as a percentage of the mean of the reference results (SDDV) must be less than 30%. This paper aims to review the background and the appropriateness of the 30% limit.

ASH CALIBRATION

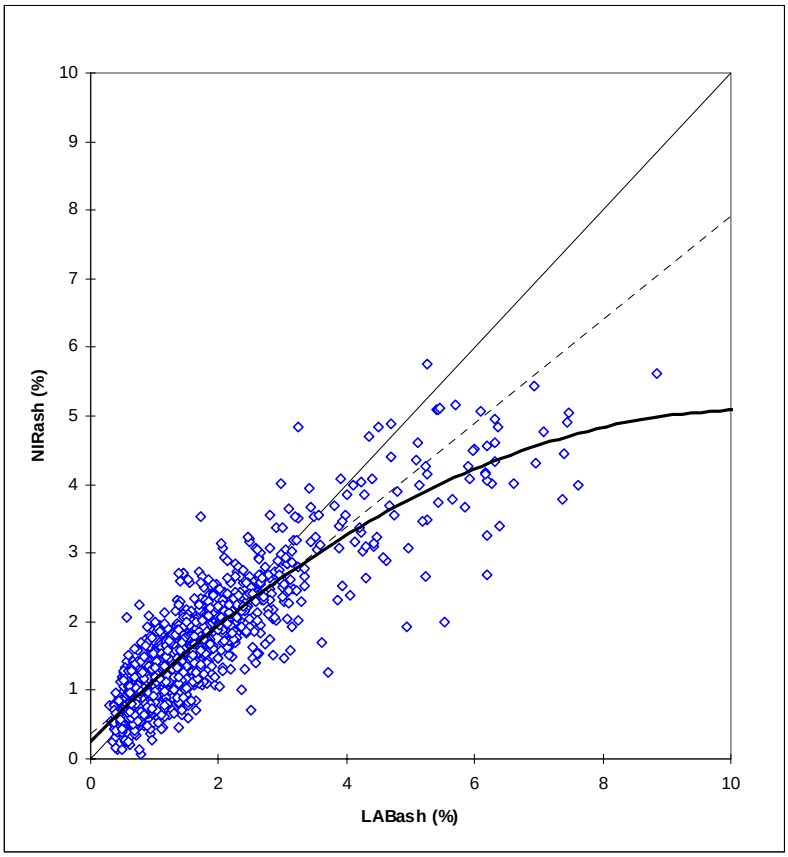
A series of papers investigating the extension of NIRA to ash determination for both wools of New Zealand and Australian origin have been presented at recent IWTO conferences. In New Zealand, Wear reported an inability to accurately predict samples with a high ash content[(2)] and proposed this may be due to a lack of homogeneity of ash in the sample. While further investigations were promising and expanded the measurement range[(3,4)], underestimation of high ash content samples and poor precision required a cut-off point. Subsequent work establishing the equivalence of Wool Base derived from the reference ash content (LABash) and the ash content predicted by NIRA methods (NIRash)[(1)] were based on a commercial set of data that reflected the general low ash content of New Zealand wool. Rather than defining a means of calculating a cut-off point for individual validations it was agreed to implement a proportional 30% limit based on the standard deviation of the difference. Technical data to support the

30% level was not presented at the time. This limit was originally derived from observation of validation sets containing low ash samples, typically LABash values less than 3%, from New Zealand wool[(2)] that had been scoured as part of the traditional Yield test.

Scoured subsamples of Australian wool exhibit a larger range of residual ash content than that observed in the New Zealand data on which the 30% limit was based. While 99% of the results are within the range 0 – 4% LABash, the final 1% can be up to 20% LABash. Petrie et.al.[(5)] reported underestimation of high ash samples and poor precision on wool of Australian origin which was consistent with the New Zealand findings. A typical calibration of that time can be seen in Figure 1. The calibration was unable to predict NIRash above approximately 3% with any certainty. Plotting a quadratic trend line through the data highlighted the curvilinear nature of the relationship.

Figure 1. Ash Calibration Showing Under Prediction of High LABash

(reproduced from Petrie et al 2003)



Investigations into the causes behind the non-linear behaviour concluded that the problem related largely to sampling. Residual matter that affects ash, in particular dag, is discretely distributed throughout the sample. High levels of dag were found to adversely affect the repeatability of both the IWTO-19 reference method and the NIRA method as a direct consequence of the uneven distribution. It was established that the NIRA method can adequately recognise dag as ash up to levels of about 20% ash content providing it was visible to the NIR instrument. However, when scanning the area of the scoured log surface it may be either absent or masked by wool, being hidden just below the surface. The depth that the NIR beam can penetrate a scoured wool sample is not known precisely but is likely to be only a few millimetres. This places important constraints as to how a sample representative of the bulk can be presented to the NIR instrument for analysis. The problem does not impact on measurements of AEM to the same degree as the residual grease tends to be evenly distributed throughout the scoured subsample.

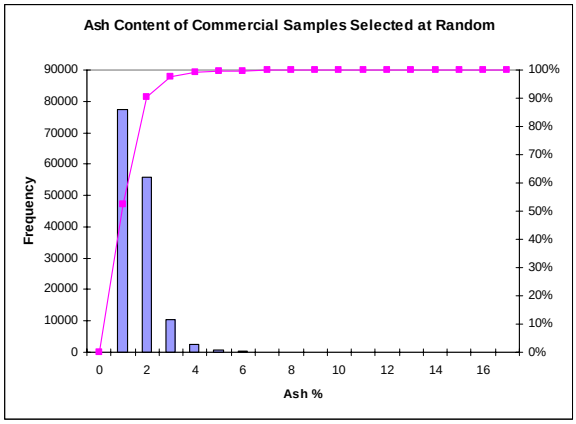
Based on these findings recent calibration development focussed on the following:

- Maximising the exposure of discretely distributed residual contamination by slicing the log perpendicular to its axis to expose its interior and provide a surface for analysis which better represents the bulk,
- Increasing the area scanned by the NIR instrument;
- Closely linking the area scanned by the NIR instrument to the reference method, and
- Increasing the representation of high ash samples up to 17% LABash in the calibration and validation populations.

POPULATION STRUCTURE

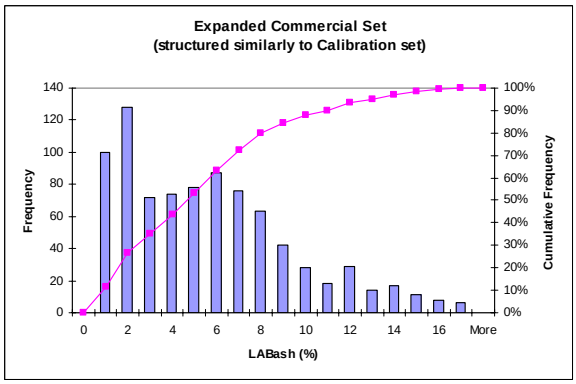
The variation of the Ash Content of samples taken as a random sample from commercial testing is presented in Figure 2.

Figure 2. Ash Content of Commercial Samples Selected at Random.



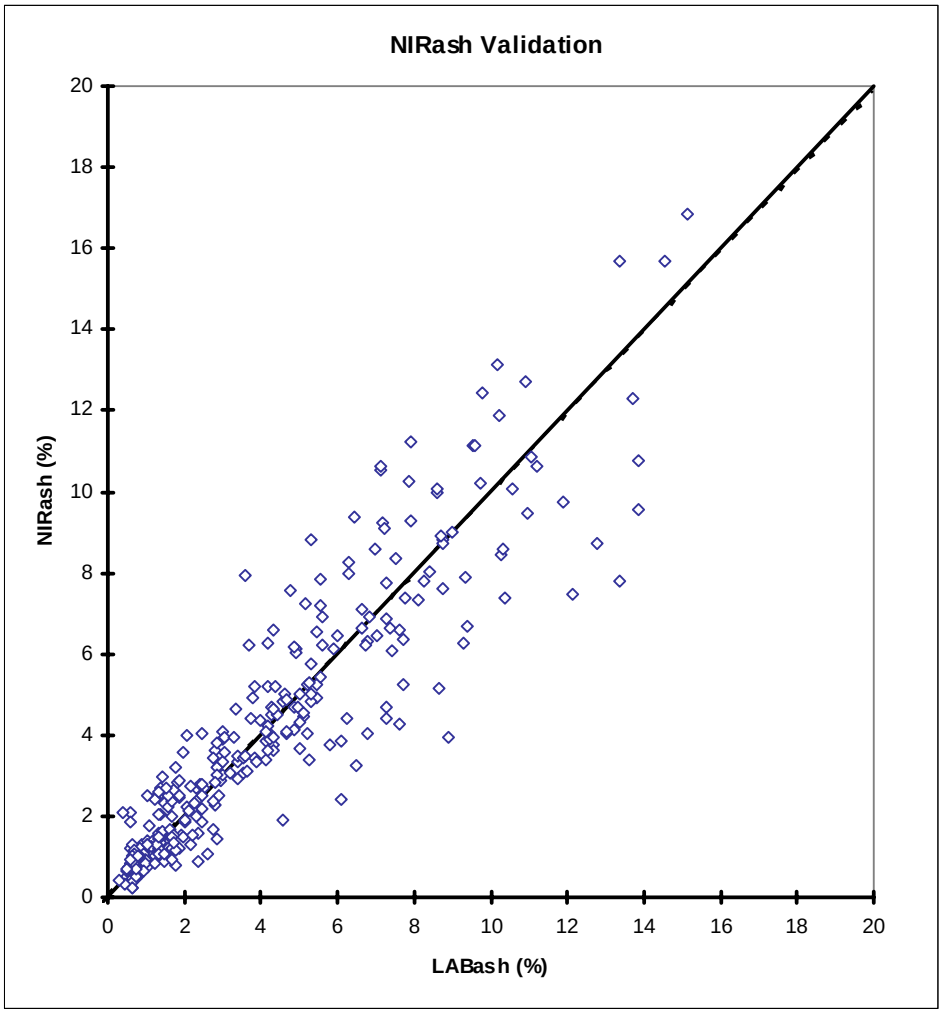
Most of the samples have a low Ash Content. It is a common practice when calibrating NIR instruments to choose samples so that the number chosen across each element over the entire calibration range is the same. For example, this would mean choosing 20 samples from 0% to 5% ash, 20 samples between 5% to 10% ash, 20 samples from 10% to 15% ash and 20 samples between 15% and 20% ash. In NIRA circles, this has typically been referred to as a "Box Car" distribution. In the case of calibrations for scoured wool derived from greasy wool subsamples, there are so many variables (diameter, VM type and content, dag content, grease levels etc.) to take into account that it is impossible to maintain the "Box Car" distribution. The compromise achieved is shown in Figure 3 below.

Figure 3. Distribution of LABash in the calibration samples.



The calibration set comprised 851 samples with LABash ranging from 0.29% - 16.96%. The high ash content samples included some with high dag content and some with high dust content. It was found that by adopting the above approach a significant improvement in the prediction of high ash samples was achieved. Figure 4 displays the validation results.

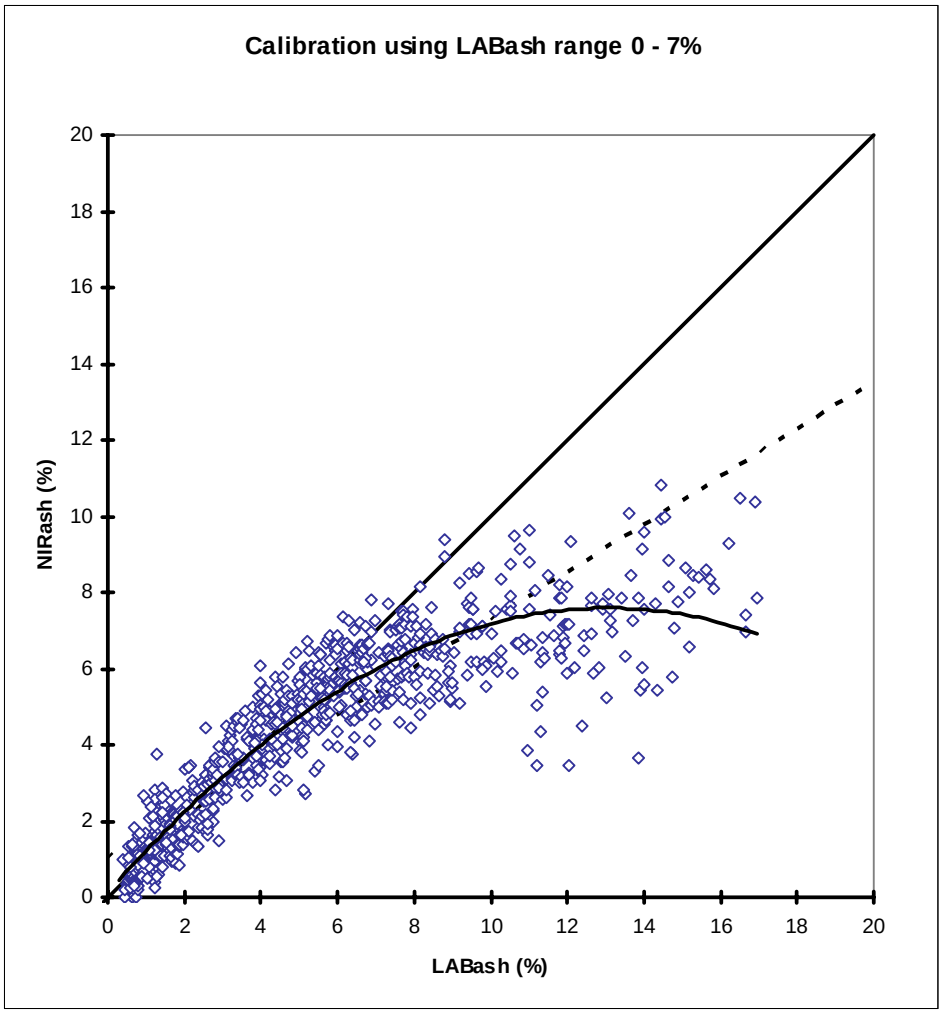
Figure 4. NIRash versus LABash.



All IWTO-19 Appendix K validation criteria were satisfied with the exception of the SDDV, which for this set of data was 34%. Several different calibrations (using different models and different mathematical treatments of the data) and validations were attempted but all failed the current 30% SDDV requirement, ranging in SDDV from 33% to 37%.

Further confirmation that increasing the number of high ash content samples was a significant factor impacting on the under prediction of high ash content samples rather than the additional benefits gained from the slicing and the larger scanned area can be seen in Figure 5. In this case a calibration based on a subset of the 851 samples was prepared to have a reduced range in ash content.

Figure 5. NIRash Calibration Based on a Restricted Range of Ash Content.



To date, commercial NIRash measurement in New Zealand and South Africa has been limited up to approximately 3% ash content in order to eliminate the non-linearity effects and to establish equivalence with the reference method; and in doing so satisfy the SDDV requirement.

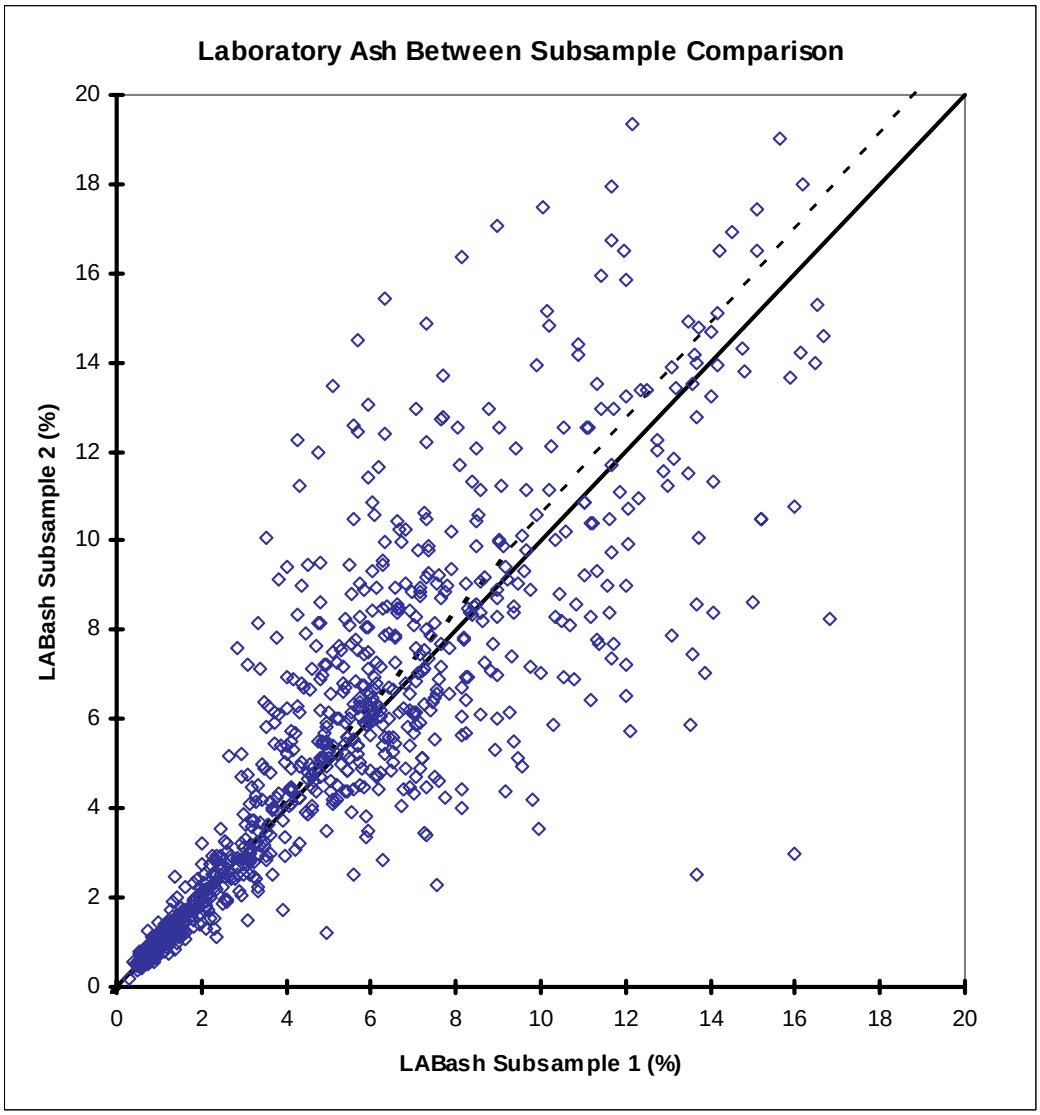
The higher ash content of some Australian wool introduces a requirement to routinely predict ash contents up to at least 5%. If the calibration set is not adequately populated with the full range of high ash samples the calibration cannot predict samples greater than about 3% LABash (see Figures 1 and 5).

As Appendix K requires that the range of the residuals in the validation test specimens be similar to the calibration set, it became necessary to examine the appropriateness of the current SDDV limit of 30% in the context of the wider range of ash contents.

The SDDV is calculated from the Standard Deviation of the Differences (SDD) for a set of validation measurements and the average ash content (determined by the traditional furnace method) for the same set. It is a common phenomenon in wool metrology that the uncertainty in measurement varies with the magnitude of the characteristic being measured. The use of a single SDDV limit is based on an assumption that the uncertainty increases as a fixed proportion of the average. To test this hypothesis a data set was constructed from normal commercial tests that exhibited a similar distribution of ash content

as used for the NIR calibration and validation sets. The ash contents (LABash) of the two subsamples from the same sample plotted against each other provides an indication of the variation that is inherent in the reference method used for calibration (see Figure 6).

Figure 6. A Comparison Between Subsamples for the Conventional Furnace Method (LABash)



It is clear from Figure 6 that the variation in ash content measurements does not increase in direct proportion with the average ash content. For ash contents greater than 4% the variation increased dramatically. The reason for this dramatic increase, as indicated earlier, is thought to arise from the nature of the material contributing to the ash content and its distribution within the scoured subsample. As an example, dag pieces appear as discrete contaminants in the subsample and depending on whether or not such discrete contaminants are randomly included or excluded in the ash test specimen will lead to either a high (e.g.10%) or low (e.g. 4%) ash content. It should be noted that the existing Test Method Range Checks for Wool Base control any errors that arise from this source by ensuring additional subsamples are tested when the Wool Base range between subsamples exceeds defined limits.

The impact of this variation on the calculated SDDV can be determined by applying the same validation analyses to the LABash versus LABash data as is applied to the NIRash versus LABash data. The complete IWTO-0 analyses are presented in an Appendix 2 to the report.

Table 1. The Impact of Ash Content Variability on SDDV.

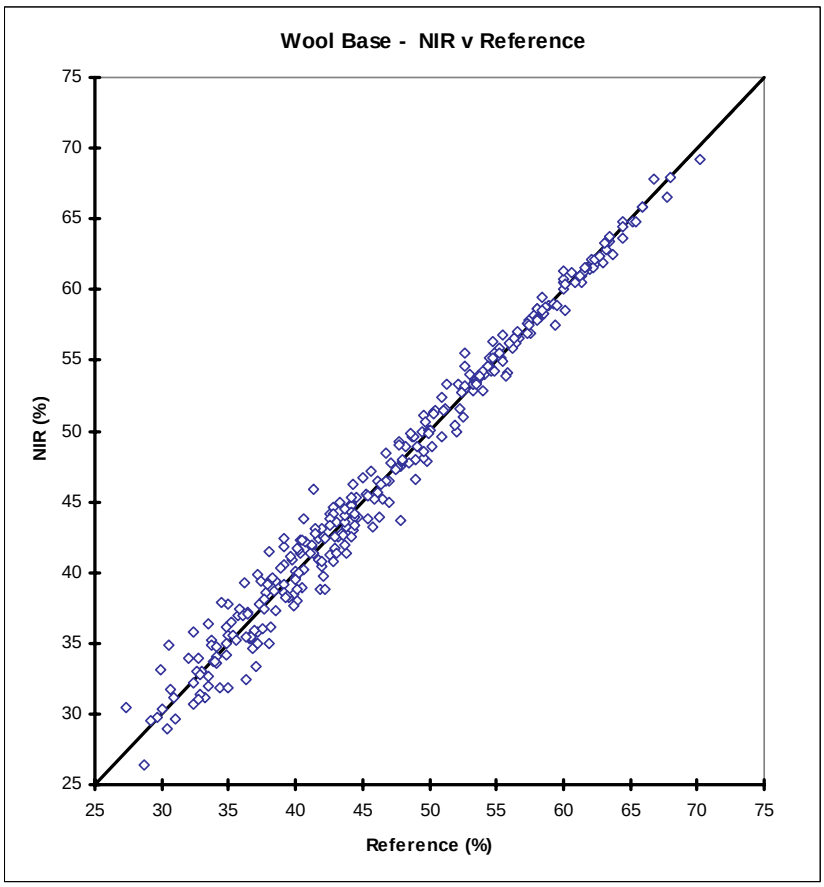
	N	Reference Average Ash (%)	Standard Deviation of Differences (%)	SDDV (%)
Commercial Data: LABash Versus LABash	851	5.17	2.18	42.1
NIR Validation Data: NIRash Versus LABash	316	3.91	1.31	33.5

The above highlights the contradiction that currently exists in the essential requirements. In order to overcome the non-linearity exhibited by calibrations that do not extend the range of ash content in the samples, high ash content samples need to be included in both the calibration and validation sets of samples. However, when this is done one cannot meet the current limit in the Test Method for SDDV.

IMPACT ON WOOL BASE

The impact on Wool Base can be determined from the same validation data set where the Wool Base has been calculated using the LABash values in one case and the NIRash in the other. A comparison of the NIRash Wool Base and the LABash Wool Base is provided in Figure 7.

Figure 7. A Comparison of NIRash Wool Base and Reference Wool Base



The analysis for the existence of a level dependent bias (see IWTO-0 analysis in Appendix 2) reproduced in table 2 below:

Table 2. Differences Between the NIRash Method and the Reference Method.

Reference Wool Base (%)	Calculated NIRash Wool Base (%)	Difference in Wool Base (%)
45.00	45.00	0.00
50.00	49.99	-0.01
55.00	54.98	-0.02
60.00	59.96	-0.04
65.00	64.95	-0.05
70.00	69.94	-0.06

When considered together with the results that:

- the average bias of -0.002% was not significantly different from zero; and
- the slope of the regression line (0.997) was not significantly different from unity

one concludes that the methods are equivalent in estimating Wool Base.

CONCLUSIONS

In order to adequately predict the ash content in scoured subsamples of Australian wool by NIRA, calibration and validation over a large range of ash is required. The statistic SDDV is dynamic and dependent on the LABash range and population structure.

Between subsample variance of the reference method over the range 0.2% to 17% LABash indicates that for a calibration set based on commercial Yield test samples augmented with a number of high LABash samples the current SDDV limit of 30% is not achievable.

An increase in the SDDV limit from 30% to 40% is recommended for ash content predictions where the range of LABash varies from 0-6% or higher.

REFERENCES

1. Wear, J. L. (2003) *The use of NIR technology to predict IWTO-19 ash residuals and its effects on Wool Base.* IWTO T&S Committee, Buenos Aires, Report RWG 03
2. Wear, J. L. (2001) *The use of NIR to predict residual ash in the IWTO-19 Yield test.* IWTO T&S Committee, Nice, Report RWG 02
3. Wear, J. L. (2002) *The use of NIR to predict residual ash in the IWTO-19 Yield test.* IWTO T&S Committee, Barcelona, Report RWG 02
4. Wear, J. L. (2002) *The use of NIR technology for predicting IWTO-19 residual ash in a commercial laboratory.* IWTO T&S Committee, Nice, Report RWG 06
5. Petrie, D. J., Lidgard, J. J., Marler, J. W. and Ireland, A. H. M. (2003) *Measurement of IWTO-19 ash content by Near Infrared Reflectance Analysis.* IWTO T&S Committee, Buenos Aires, Report RWG 02

APPENDIX 1

DO NOT EDIT OR DELETE the above heading. NOTE: This message is formatted as hidden text and will not appear in your published document. It may be deleted at any time once the message is heeded.

RECOMMENDED CHANGES TO APPENDIX K OF IWTO-19.

Recommended changes to Appendix K Clause K3(d)(iii) are underlined and highlighted.

APPENDIX K

Method for Determining the Ethyl Alcohol Extractable Matter and Ash in the Scoured Subsamples using an NIRA Instrument

K1 Scope

This Appendix sets out a suitable method for determining the ethyl alcohol extractable matter and ash content in a scoured subsample of raw wool for determination of Wool Base and Vegetable Matter Base. For the purpose of this appendix the methods for determining the reference measurements will be referred to as the “primary method” and the AEM or Ash results for all methods will be referred to as “the residuals”.

K2 Principle

A Near Infrared Reflectance Analysis (NIRA) instrument calibrated against the primary method may be used to estimate the residuals content directly.

In a retest situation the primary method as described in Appendices E and/or D must be used for determining the residuals.

K3 Essential Requirements

a) The test specimen shall be drawn from the scoured subsample in such a manner as to avoid any change in its residuals.

b) The NIRA instrument shall be calibrated against the primary method.

c) The calibration test specimens must be representative of the commercial population of test specimens and exhibit the full range of variation of all wool characteristics to be expected in routine testing, unless the software utilised by the instrument allows "outlier" spectra to be

identified during measurements. In the latter situation, the calibration data set must initially encompass as wide a range of variation as is feasible, and the calibration data set may then be extended to widen the sample population over a period of time.

d) The calibration of the instrument must be validated prior to commercial use against test specimens drawn from wool samples that have not been used in the calibration. The range of residuals in the validation test specimens shall be similar to the calibration set. The validation data is compared to the primary method using the procedures set out in IWTO-0 Appendix B. Based on the analysis of the individual test specimen validation data, the following requirements must be satisfied before the calibration can be used on commercial tests:

i. The Geometric Mean Slope of the relationship between the NIR predicted Residual and the Primary Method must be within the range 0.9 to 1.1;

ii. The Mean Difference between the NIR predicted Residual and the Primary Method data must be less than 0.1%; and

iii. The Standard Deviation of the Differences must be:

- less than 30% of the mean of the reference results for the validation samples for alcohol extractable matter (AEM); and
- less than 30% of the mean of the reference ash results for the validation samples that cover a range of 0-3% ash content; or
- less than 40% of the mean of the reference ash results for the validation samples that cover a range of 0-6% ash content or higher.

e) The instrument shall not be used to measure residuals outside the range of the validation data.

f) When in commercial use, the instrument performance must be monitored on a regular basis

K4 Method

K4.1 Apparatus

a) An NIRA instrument is required.

K4.2 Procedure

The procedure is as follows:

a) Test specimens presented to the NIRA instrument must be of sufficient mass to give the appropriate packing density and minimum thickness recommended by the instrument manufacturer.

b) The NIRA instrument is calibrated against the primary method, using at least 200 test specimens covering the probable range and population structure of all wool characteristics to be encountered in routine testing, using a multiple linear regression or another suitable statistical technique.

NOTE: The uncertainty in the primary method can impact on the ability to derive a calibration that meets the criteria defined in the Essential Requirements. Where necessary, this can be addressed by either increasing the number of individual test specimens in the calibration and validation or, for the purposes of calibration only, using the mean of quadruplicate residual determinations for the primary method.

c) The calibration is validated against the criteria shown in K3 (d) using at least 200 independent test specimens.

d) The percentage of residuals in the oven-dry scoured subsample, (E_i or A_i), is calculated to the nearest 0.01 % or better from the calibration equation.

Where the resultant value for E_i and/or A_i is greater than the maximum value used in deriving the calibration/validation, the result is discarded and the primary method must be carried out.

e) The calibration/s must be monitored by checking at least 20 test specimens per week. Current records of the quality control cheeks, including the test data and calculations, must be retained by the laboratory for at least 1 year.

K5 Record

Record the following information:

a) The percentage residual matter in the oven-dry scoured subsample (E_i and/or A_i) to the nearest 0.1 %.

APPENDIX 2

DO NOT EDIT OR DELETE the above heading. NOTE: This message is formatted as hidden text and will not appear in your published document. It may be deleted at any time once the message is heeded.

NIRash Validation

(A) Test for Overall Relative Bias & Paired t-test.

	Overall Bias		Paired
	LABash	NIRash	t-test
Number	316	316	316
Average	3.9098	3.9318	0.0220
SD	3.2798	3.2510	1.3095
SE	0.1845	0.1829	0.0737
t value	21.1913	21.4993	0.2988
p value	0.0000	0.0000	0.7653
Significance			**NS**

SDDV (%)
33.5

(B) Test for Correlation.

Number	316
DF	314
R	0.9196
t-value	41.4880
p-value	0.0000
Significance	***

(C) Test for Level Dependent "Bias".

Statistical Significance

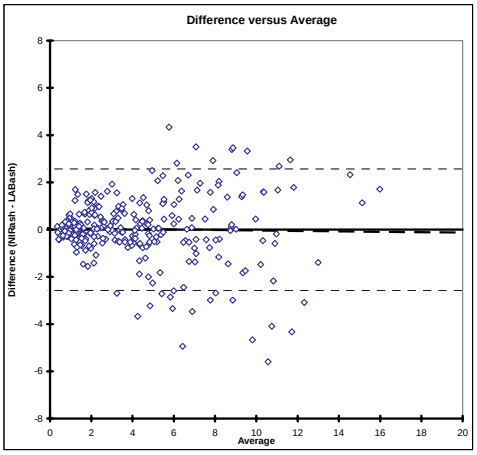
Regression	Slope	Significance	SE of Slope	t-Value	p-value	Rsq
GM	0.9912	**NS**	0.0220	0.3995	0.6898	0.8457
DVA	-0.0092	**NS**	0.0231	0.3977	0.6911	
Intercept(GM)	0.0563					STEYX
Intercept(DVA)	0.0580					1.2903

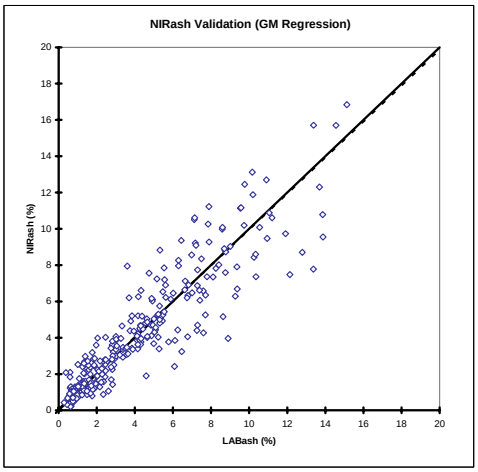
Magnitude of the Level Dependent "Bias".

Nominal	Calculated	Avg. "Bias"
1	1.05	0.05
2	2.04	0.04
4	4.02	0.02
6	6.00	0.00
10	9.97	-0.03
15	14.92	-0.08

Max/Mins.

LABash	max	15.12
	min	0.29
NIRash	max	16.84
	min	0.21
average	max	15.98
	min	0.36
difference	max	4.34
	min	-5.60





Expanded Commercial ash data comparison between subsamples

(A) Test for Overall Relative Bias & Paired t-test.

	Overall Bias		Paired
	LABash Sub 1	LABash Sub 2	t-test
Number	851	851	851
Average	5.1712	5.4005	0.2293
SD	3.7839	4.0431	2.1772
SE	0.1297	0.1386	0.0746
t value	39.8669	38.9663	3.0729
p value	0.0000	0.0000	0.0022
Significance			**

SDDV (%)
42.1

(B) Test for Correlation.

Number	851
DF	849
R	0.8473
t-value	46.4787
p-value	0.0000
Significance	***

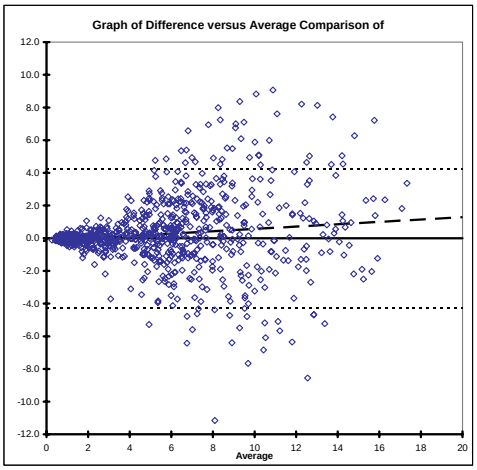
(C) Test for Level Dependent "Bias".

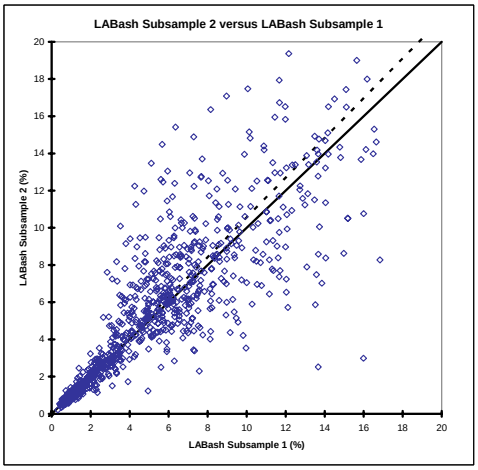
Statistical Significance

Regression	Slope	Significance	SE of Slope	t-Value	p-value	Rsq
GM	1.0685	***	0.0195	3.5162	0.0005	0.7179
DVA	0.0717	***	0.0197	3.6366	0.0003	
Intercept(GM)	-0.1248					STEYX
Intercept(DVA)	-0.1496					2.0111

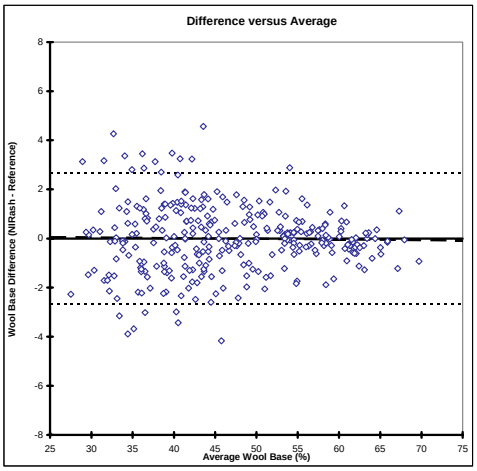
Magnitude of the Level Dependent "Bias".

Nominal	Calculated	Avg. "Bias"
1	0.94	-0.06
4	4.15	0.15
7	7.35	0.35
10	10.56	0.56
13	13.77	0.77
16	16.97	0.97





NIRash Wool Base versus Reference Wool Base



(A) Test for Overall Relative Bias & Paired t-test.

	Overall Bias		Paired
	Reference Wool Base	NIRash Wool Base	t-test
Number	316	316	316
Average	46.6513	46.6495	-0.0017
SD	9.7911	9.7651	1.3515
SE	0.5508	0.5493	0.0760
t value	84.6982	84.9206	-0.0226
p value	0.0000	0.0000	0.9820
Significance			**NS**

SDDV (%)
2.9

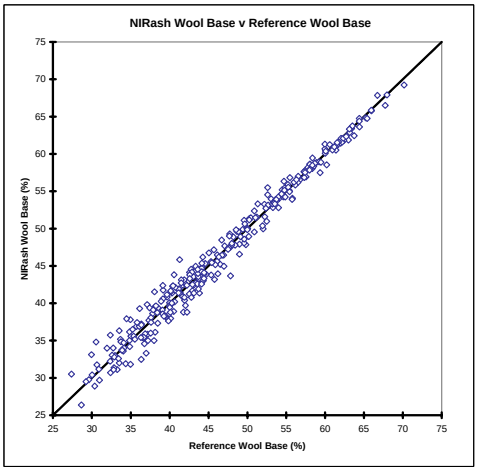
(B) Test for Correlation.

Number	316
DF	314
R	0.9905
t-value	127.3087
p-value	0.0000
Significance	***

(C) Test for Level Dependent "Bias".

Statistical Significance

Regression	Slope	Significance	SE of Slope	t-Value	p-value	Rsq
GM	0.9973	**NS**	0.0078	0.3423	0.7324	0.9810
DVA	-0.0027	**NS**	0.0078	0.3418	0.7327	
Intercept(GM)	0.1222					STEYX
Intercept(DVA)	0.1229					1.3520



Magnitude of the Level Dependent "Bias".

Nominal	Calculated	Avg. "Bias"
45	45.00	0.00
50	49.99	-0.01
55	54.98	-0.02
60	59.96	-0.04
65	64.95	-0.05
70	69.94	-0.06

Max/Mins.

Control	max	70.16
	min	27.37
Treatment	max	69.23
	min	26.37
average	max	69.69
	min	27.51
difference	max	4.56
	min	-4.17