INTERNATIONAL WOOL TEXTILE ORGANISATION

TECHNOLOGY & STANDARDS COMMITTEE | NICE MEETING

Commercial Technology Forum | November 2002

Chairman: A.G.DE BOOS (Australia) | Report No: CTF 01

OFDA2000 proficiency trials

By

Peter Baxter, Bill Johnston

SGS Wool Testing Services — PO Box 15062, Wellington, New Zealand
Interactive Wool Group — 132 North St., Tamworth, NSW 2340, Australia

SUMMARY

This report summarises the outcomes of three proficiency trials undertaken by the field operators of the OFDA2000 instrument. The numbers of participants in each of the trials were 14, 49 and 18 respectively. The first trial was undertaken to test a draft protocol. The second was a full-scale trial of a slightly-revised protocol. The third trial was undertaken simply to test the accuracy of a representative group of instruments when used in the field. Further routine trials are to be organised on a full-scale basis.

The results from the trials showed that the instruments performed in accordance with published expectations on greasy wool samples, and similar to the OFDA 100 on top slivers. On greasy merino wool samples, the average 95% confidence limits were approximately ± 1.1 to 1.2 µm, ± 2 %, ± 9 °/mm and ± 12 mm respectively for mean fibre diameter, coefficient of variation of diameter, mean curvature and staple length respectively. Overall instrument accuracy was within ± 0.2 µm and 0.5% of the Interwoollabs Standard Values for mean fibre diameter and coefficient of variation of diameter respectively. The environmental compensation algorithm was shown to perform satisfactorily in these trials carried out over a very wide range of field conditions.

INTRODUCTION

Whilst the OFDA2000 instrument may be used in a laboratory, its primary use has been in on-farm measurements of staples taken from fleeces or animals. Interactive Wool Group (IWG) has maintained a policy of ensuring quality assurance in the provision of these services by its licensees using a number of initiatives, one of which has been the requirement for operators to participate in proficiency trials. Preliminary results from these were mentioned at the Barcelona conference, and IWG were requested to provide details to IWTO.

The organisation and logistics of providing suitable samples and instructions to a large number of operators, most of whom have had no experience in participation in such trials, presents a number of problems. In consequence IWG chose to take a cautious approach, and, rather than attempt to impose an inadequately-conceived protocol, took time to develop the program.

At the time of preparing this report, one initial trial had been carried out involving 14 participants, one full-scale trial involving 49 participants had been completed, and a trial of 18 participants undertaken to confirm instrument accuracy. Now that a satisfactory protocol has been confirmed, further trials are underway, and will be undertaken on a regular basis.

TRIAL DESIGNS

Initial trial (#1)

The first proficiency trial was organised in June 2001, with a limited number of participants in both Australia and New Zealand, in order to test the draft protocol.

This preliminary trial was not intended to be onerous on the participants, since many would never have participated in round trials previously, and therefore the initial concept was to "keep it simple". In summary, each participant received two greasy samples, two clean samples and one top sample. Each greasy and clean sample comprised two staples chosen at random from a bulk population of staples. Each bulk sample comprised a large midside sample from a single animal, which was broken down into individual staples. The clean staples were all cleaned by the organiser using the same technique as is specified in the instruction manual for the instrument (washing in a mixture of propanol and hexane). There was no blinding of the replicates in this trial. The greasy staples were to be measured with a default grease correction factor (GCF)[1], whilst the clean staples and top sample were to be measured with the grease correction factor off.

Full-scale trial (#2)

Slight modifications were made to improve the protocol that had been used in the first proficiency trial, and samples distributed to all major users in Australasia during March 2002. Results were obtained from 49 instruments, although data on only some of the samples was available for 6 instruments, due to sample shortages. In summary, large midside patches were taken from 2 animals, and randomly selected greasy staple pairs were sent to participants with blind replication. In all, participants were supplied with 8 greasy staples, 4 of which were to be cleaned before measurement, whereas the other 4 were to be measured using the default GCF correction. Additionally participants were supplied with blind replicates of 2 top samples. All staple measurements were to be carried out in duplicate.

Accuracy trial (#3)

The 3rd trial was organised in August 2002 to examine instrument accuracy. The two earlier trials had utilised commercial top samples for this purpose, but since the nominal values were known only from the topmaker's report, it was considered that accuracy should be re-evaluated using Interwoollabs IH tops. In view of the good performance of the 49 instruments participating in the 2nd trial, and a shortage of time to meet the deadline for preparation of this paper, the number of participants was limited to 18. In total 10 instruments in Australia, 6 in New Zealand and 2 in South Africa participated in this trial.

The 4th trial is intended to be the first routine full-scale trial with the finalised protocol, and involving all or most instruments employed in on-farm work. It is hoped that this will be completed before the end of 2002.

ANALYTICAL METHOD

Data was first screened for obvious errors and outliers. Outliers were identified using the following technique from Statistica 6: "The ranges of outliers and extremes is shown in the "classic" box and whisker plot, where the upper box value (UBV) is the 75th percentile, the lower box value (LBV) is the 25th percentile, and the outlier coefficient is equal to 1.5 (times the LBV minus UBV range)". In the first trial, one complete set of data had to be eliminated when it was discovered that an incorrect calibration had been used (procedures were immediately tightened to prevent this recurring). In the 2nd trial, data from one sample on one instrument was deleted - it was considered that the 2 staples supplied were atypical of the bulk. All other valid data were included in the analyses.

The data for each sample was analysed by standard one-way analysis of variance techniques. Two-way analyses were carried out on the combined data set in order to provide comparative data summary plots

[1] The grease correction factor is a linear equation applied by the instrument to every single fibre measurement to convert the greasy measurements to the equivalent clean measurements. Operators are normally required to determine the constant coefficient used in this equation for each mob of sheep, but the software includes a default equation which can be utilised in appropriate conditions.

and to identify any trends for specific instruments. Between-replicate ranges were analysed using 2-way ANOVA to establish whether data from any specific instrument was more variable than the average.

Mean squares for the error and effect terms were used to calculate within-instrument and between-instrument components of variance for MFD, CvD, curve, staple length, and from these the 95% confidence limits for a single measurement on a single staple were calculated for each sample.

In each case the environmental conditions at the time of measurement were recorded, and an examination of the data made to establish whether these conditions could have contributed to differences.

RESULTS - TRIAL 1

There was no significant correlation between the environmental conditions reported by the instruments and the test results. Over the 14 instruments, the temperatures recorded at measurement varied from 13 to 27 ºC, and the relative humidity values from 43% to 85%.

Precision of measurement

The outcomes of the analyses of variance are shown in the following tables.

MFD - Sample	Grand mean (µm)	Variance (µm²) Within-instrument	Variance (µm²) Between-instruments	95% Confidence limits (±µm)
Greasy 1	19.3	0.064	0.138	0.9
Greasy 2	21.2	0.091	0.324	1.3
Clean 1	19.9	0.031	0.044	0.6
Clean 2	21.7	0.071	0.088	0.8
Top A	21.6	0.038	0.004 (ns)	0.4

CvD - Sample	Grand mean (%)	Variance (%²) Within-instrument	Variance (%²) Between-instruments	95% Confidence limits (± %)
Greasy 1	18.9	0.53	1.44	2.8
Greasy 2	20.5	0.62	0.31 (ns)	1.9
Clean 1	18.0	0.22	0.37	1.5
Clean 2	19.6	0.22	0.51	1.7
Top A	20.1	0.33	0.12 (ns)	1.3

Curve - Sample	Grand mean (º/mm)	Variance (º/mm²) Within-instrument	Variance (º/mm²) Between-instruments	95% Confidence limits (± º/mm)
Greasy 1	64.9	6.0	9.3	7.8
Greasy 2	63.5	5.6	10.5	8.0
Clean 1	65.4	3.0	23.2	10.2
Clean 2	69.7	8.8	14.7	9.7
Top A	59.2	5.5	3.8	6.1

Staple length- Sample	Grand mean (mm)	Variance (mm²) Within-instrument	Variance (mm²) Between-instruments	95% Confidence limits (± mm)
Greasy 1	77.9	18.3	31.3	14.1
Greasy 2	79.5	10.6	24.4	11.8
Clean 1	85.0	30.8	65.9	19.7
Clean 2	83.6	23.1	26.6	14.1

The paired differences between replicates within instruments give an indication of the precision of measurement on a single instrument. The ANOVA's carried out on this data did not indicate any statistically significant difference between the instruments on between-replicate performance.

Discussion

The above precision estimates can be compared with published data. The report given in Aachen on the OFDA2000 precision (Baxter 2001a) suggested that for a single greasy wool staple the 95% confidence limits at 20 μm should be ± 1.2 μm, which is consistent with the values tabulated above for samples Greasy 1 and 2. The equations detailed in the Aachen report can also be used to calculate the precision of an MFD measurement on a single cleaned staple, which at 20 μm would be ± 0.8 μm, which again is totally consistent with the results shown above. There are no comparable historical precision data on top measurements undertaken with this instrument, but the value quoted above is slightly better than the equivalent value (± 0.47μm for a single slide) obtained from the raw data used to derive table D4 in IWTO-47 (after allowing for the number of specimens to be measured under the conditions that the data in this table was obtained).

Published reports have not detailed confidence limit estimates for coefficient of variation of fibre diameter. The validation set used in one published paper (Baxter 2001b) was partially re-analysed to give an estimate of the precision of CvD. This data set omits a component of variance due to between-instrument (as indicated in the notes to Table 2 in that paper), but nevertheless, at approximately 18 μm, a value of ± 1.4% was obtained. This is of the same order of magnitude as the value of ± 1.7% that was published in the original draft of IWTO-47 (however, this estimate was based on the average of measurements on 4 slides). With the exception of the estimate for sample 'Greasy 1', which is heavily influenced by the extreme results from two instruments, the remaining values tabulated above are of the same order.

Estimates of the precision of fibre curvature on greasy staples were published in Baxter 2001b, again with the proviso that between-instrument variance was omitted. The values estimated from this trial were significantly better than the published values of ± 12 and ± 11 °/mm for greasy and clean samples respectively.

Precision of staple length measurements were estimated in both the above papers, although the latter covered a much wider data set. The Aachen values were ± 8.3 and ± 7.2 mm for greasy and clean samples, for which the respective values quoted in the 2nd paper were ± 11 and ± 10 mm. The values obtained in this current trial exceeded both sets of published data, possibly indicating that more care might be needed in order to obtain meaningful results from this measurement. (For comparison, the precision of a certified measurement on approximately 60 staples from a farm lot is ± 4.8 mm. Data from development work carried out with an Agritest staple length meter complying with IWTO-30 suggested that the precision of a single measurement on a single staple on this type of reference instrument is ± 7 mm (Baxter 1996). It should also be noted that since this trial was undertaken, the wire sample slide has been replaced with an improved version that facilitates more precise diameter measurements, but restricts the length increment to 5 mm, whereas the original wire slides allowed varying length increments down to 3.5 mm.)

Conclusions from trial 1

As a preliminary proficiency trial, the results were considered very successful. The results on the top sample show no statistically significant difference between instruments, and the overall precision performance was at least comparable with laboratory instruments, which, considering the range of environments in which the instruments were used, is quite remarkable.

The results on greasy staples also indicate precision values for MFD, CvD and curvature which are consistent with data already published on the performance of the OFDA2000. The precision of the staple length measurements did, however, appear to leave some room for improvement.

The results in this trial were obtained using a range of software versions and possibly two different designs of sample slides. A more consistent approach should be adopted in future trials as a result of the system upgrades that followed this trial.

Participants complied with the instructions to varying degrees. It was recommended that the instructions should be re-drafted so as to be of the maximum clarity, and strict discipline should be attempted in order to obtain the results on time.

It was also recommended that in future trials, the staple replicates should be blinded so that participants are not immediately aware of the staple pairs. Whilst this is not expected to have any significant effect on the results, it will ensure that the trial complies with normal protocols. Future trials should include 2 top samples of different diameters (with two replicates). With this proviso, the tops measurements would be sufficient to confirm calibration compliance, and measurements on IWG-cleaned staples would then not be required. The trial would, however, benefit if each operator were required to clean staples.

RESULTS TRIAL 2

None of the instruments were excluded from the analysis.

The components of variance table is shown below:

Sample	Parameter	Grand mean	GREASY			CLEAN		
			Between	Within	95%CL	Between	Within	95%CL
1, 5	MFD	19.4	0.23	0.06	1.05	0.21	0.05	1.01
1, 5	CVD	20.0	0.38	0.37	1.69	0.46	0.68	2.09
1, 5	Comfort	98.6	0.27	0.25	1.42	0.22	0.24	1.33
1, 5	Curve	69.2	10.44	11.26	9.13	16.50	11.62	10.39
1, 5	Length	64.5	17.13	8.42	9.91	18.15	9.03	10.22
1, 5	AcrossCvD	7.62	1.13	1.49	3.17	1.00	1.25	2.94
1, 5	AlongCvD	21.0	1.23	0.81	2.80	1.00	1.14	2.87
1, 5	MinDAlong	16.5	0.41	0.28	1.64	0.34	0.49	1.79
2, 6	MFD	18.5	0.20	0.08	1.05	0.13	0.11	0.95
2, 6	CVD	18.1	1.12	0.63	2.60	0.93	0.77	2.55
2, 6	Comfort	99.7	0.03	0.05	0.55	0.03	0.05	0.56
2, 6	Curve	67.7	9.32	9.91	8.59	12.09	11.06	9.43
2, 6	Length	63.3	17.66	11.05	10.50	30.43	10.98	12.61
3, 7	MFD	19.0	0.35	0.05	1.24	0.24	0.11	1.16
3, 7	CVD	19.3	1.01	0.43	2.35	0.22 (ns)	0.64	1.81
3, 7	Comfort	99.2	0.31	0.10	1.26	0.20	0.12	1.11
3, 7	Curve	71.2	14.72	9.28	9.60	16.63	8.30	9.79
3, 7	Length	65.9	19.65	11.05	10.86	13.30	7.62	8.97
4, 8	MFD	18.6	0.22	0.16	1.20	0.11	0.11	0.93
4, 8	CVD	18.4	0.41	0.92	2.26	0.65	1.01	2.53
4, 8	Comfort	99.6	0.07	0.06	0.70	0.01 (ns)	0.08	0.58
4, 8	Curve	67.8	13.46	11.28	9.75	9.31	9.71	8.55
4, 8	Length	64.0	28.56	14.53	12.87	11.93	12.91	9.77
	GCF	1.55	0.077	0.482	1.47			
Top A	MFD	21.6				0.01	0.04	0.42
Top A	CVD	20.6				0.04	0.18	0.91
Top A	Comfort	96.0				0.05 (ns)	0.52	1.49
Top A	Curve	47.1				6.97	5.08	6.81
Top B	MFD	19.7				0.01	0.02	0.33
Top B	CVD	19.1				0.04	0.15	0.85
Top B	Comfort	98.8				0.01	0.07	0.54
Top B	Curve	55.9				7.90	5.50	7.17

Due to the manner in which the operators were required to prepare the replicates, the within-instrument component of variance was determined using two portions of a single staple in each case. This gives an estimate of sample presentation and measurement errors, but is not equivalent to taking a separate staple and measuring that. However, the between-instrument component of variance includes a component due to "between-staples within midside", since it is impossible to send the same staples to each operator. It may therefore be concluded that the 95% confidence limits calculated from these two estimates includes contributions from the following sources: measurement, within staple portions, between staples in a midside, between instruments and operators, and in the case of greasy wool, an additional random error component due to the grease factor algorithm (although it excludes any systematic error from this source). The 95%CL therefore includes all major sources of random variation, and is indicative of the overall precision of measurement of a midside sample from an individual animal based on a single staple. It does not, of course, include any component relating to the error associated with assuming that the midside correctly represents the whole or the skirted fleece.

The RH correction factor was also examined. Average corrections of up to 0.7 µm had been automatically applied during measurement, but this had no discernable effect on performance. The range of environmental conditions under which this trial was completed was average temperatures ranging from 14 to 41 ºC, and average relative humidity from 11 to 62% RH.

Discussion

Over the 4 greasy samples, the 95% CL for greasy MFD averaged ±1.16 µm. In the first trial of 14 participants, the average was ±1.05 µm. Both numbers are consistent with published statistics. Over the 4 samples, the 95% CL for clean MFD averaged ±1.01 µm. In the first trial the samples were all cleaned by IWG, so the statistics are not directly comparable.

Over the 2 top samples, top A had an overall mean MFD of 21.6 µm, and top B 19.7 µm. The corresponding global 95% CL values for MFD were ± 0.42 µm and ± 0.33 µm. The value obtained for the 21.6 top used in the first trial was ± 0.4 µm. The data used to determine the precision of the OFDA100 in IWTO-47 can be used to calculate the precision of a single specimen using that method, and the equivalent values would be ± 0.47 µm and ± 0.36 µm respectively for these 2 tops.

The average precision of CvD over both greasy and cleaned samples was ± 2.3 %. This is slightly worse than the results from the first trial (± 2.0 %), although the much greater number of instruments and operators involved is the most probable explanation for this. The precision results for CvD on the top samples were, as expected, very much better, averaging ± 0.9 %.

The precision values for comfort factor were higher on samples 1 and 3 than on samples 2 and 4 (± 1.3 % and 0.6 % respectively for the greasy samples, and for the cleaned samples, 1.2 % and 0.6 % respectively). There was a similar disparity for the 2 top samples, with average CL values of ± 1.5 % and 0.5%. The results on the tops, for which the respective means were 96 % and 99 % suggest that the precision might be strongly level-dependent, a view which is supported by Table D6 in IWTO-12.

The 95% CL values for mean curve were ± 9.3 °/mm and ± 9.5 °/mm for the greasy and clean samples respectively, and ± 7.0 °/mm for the tops. The corresponding figures from the first trial were ± 7.9, 9.9 and 6.1 °/mm respectively. Whilst the later figures are marginally higher, the difference is of doubtful significance. The results tend to confirm that on tops the instrument is virtually as precise as an OFDA100 measuring graticules (Baxter 2002), and the performance on greasy wool is hardly much worse.

The 95% CL values for staple length were ± 11.0 and 10.4 mm for the greasy and clean samples respectively. This is somewhat improved over equivalent figures of 13.0 and 16.9 mm in the first trial. All these values exceed the preliminary figures given in the Aachen paper, but the current trial results are in line with those published in Baxter 2001b. (As noted earlier, the data shown in the Aachen report was obtained with the earlier wire slides, which used a 3.5 mm length increment, as compared with 5 mm used in the current slide design. This limits the precision available for length.) It was encouraging to see a significant improvement from the 1st to the 2nd proficiency trial, especially with the significantly-increased number of participants.

The data for AcrossCvD, Along CvD, and MinDAlong are given for sample 1 simply for interest and comparison with the principal data. The main observation is that these additional data are generally of poorer precision than the main whole-staple averages.

Conclusions from trial 2

The grand means from the measurements on top sliver were within 0.2 µm and 0.1 µm of the nominal values supplied by the mill, indicating relatively accurate overall performance. It was recommended that future trials should use Interwoollabs IH tops for which the nominal values are precisely known.

The performance over the 49 instruments that successfully participated was excellent, especially considering the wide range of environmental conditions under which the instruments were operated (14 to 41 ºC, and 11 to 62 %RH). No instrument/operator had to be excluded on the basis of performance, although the data on one sample from one instrument was deleted - this was most likely a sample rather than an instrument/operator problem. Whilst data from a 50th instrument subsequently became available, this was not included due to lateness, but these results were completely consistent with the others.

The 95% confidence limits determined in this trial were consistent with published information. The performance on tops was exemplary, indicating similar performance to the round trial used for the precision determination in IWTO-47, and confirming that all participating instruments were in satisfactory calibration. No instrument was on average more than 0.4 µm away from the grand mean, although the overall range was 0.7 µm. Nevertheless, it was commented that additional care in re-calibrating instruments would still be worthwhile.

Operator performance on staple cleaning appeared satisfactory overall, and although there were two sets of results at the extremes of the distribution, they were not judged to be significant outliers from the overall distribution and were retained in the data. The precision of a GCF determination on a single staple calculated from this trial was comparable with the precision calculated by summing the individual precision estimates from the greasy and cleaned staple MFD data. It was also comparable with a precision estimate based on the Baxter 2001a model.

RESULTS - TRIAL 3

In this trial, participants were supplied with blind replicates of two IH tops. In order to ensure that reporting took place in time to submit this paper, the number of participants was restricted to 10 in Australia, 6 in New Zealand, and 2 in South Africa. Despite the limited coverage, the trial is a valid indicator of instrument accuracy and precision. The participants were required to prepare 2 slides of what they were supplied with as 4 separate short lengths of top. Samples 1 and 3 were prepared from IH series 14, top 2, and samples 2 and 4 were prepared from IH top 4.

None of the instruments or results were regarded as outliers. The range of environmental conditions under which this trial was performed varied from 15 to 27 ºC and 9 to 50% relative humidity. The summarised precision statistics are shown in the table below:

Sample	Parameter	Grand mean	Component of variance		95%CL
			Between	Within	
Top A	MFD	18.93	0.025	0.013	0.38
Top A	CVD	20.02	0.002 (ns)	0.159	0.79
Top A	Comfort	99.4	0.006 (ns)	0.047	0.45
Top A	Curve	59.8	8.13	5.67	7.3
Top B	MFD	23.64	0.025	0.019	0.41
Top B	CVD	21.66	0.000 (ns)	0.183	0.80
Top B	Comfort	90.2	0.169	0.694	1.82
Top B	Curve	43.1	3.04	1.32	4.1

Discussion

It should be noted that all the instruments would have been originally calibrated some months previously with series 13 IH Standards, whereas the measurements were undertaken on samples prepared from series 14 Standards.

The grand means of the mean fibre diameters were within 0.2 µm and 0.1 µm of the projection microscope Standard Values as supplied by Interwoollabs. The means from individual instruments were all within ± 0.3 µm of the grand means. No instrument was more than 0.5 µm in error when compared with the IH Standard Values.

The grand means of the coefficient of variation of diameter were within 0.5% and 0.1% of the projection microscope Standard Values as supplied by Interwoollabs. The means from individual instruments were all within ± 0.4 % of the grand means. No instrument was more than 0.9% in error when compared with the IH Standard Values.

Using the data used to derive the precision statistics in IWTO-47, the expected precision for a measurement on a single specimen of mean fibre diameter using an OFDA100 is ± 0.31 µm for the 18.9 µm top, and ± 0.55 µm for the 23.6 µm top. These values again compare favourable with the estimates from this trial of ± 0.38 µm and ± 0.41 µm respectively.

The precision estimates for CvD, comfort factor and mean curvature are completely consistent with the estimates obtained in the earlier trials.

Conclusions from trial 3

The results from this trial were totally consistent with the results from previous trials.

CONCLUSIONS

Overall accuracy of measurement of mean fibre diameter on top slivers over a wide range of operators and environments using the OFDA2000 has been shown to be within ± 0.2 µm in two trials. Overall accuracy of CvD has been shown to be within 0.5% of the Standard Values issued by Interwoollabs in this latest trial.

The precision of measurement of greasy mid-side samples has consistently demonstrated agreement with the theoretical model proposed in Aachen.

The precision of measurement of greasy CvD of mid-side samples has been approximately ± 2%.

Precision of measurement of the mean fibre diameter of top samples has been shown to be similar to the precision estimates shown in IWTO-47 for the OFDA100 instrument.

Precision of measurement of mean fibre curvature has also been consistent in these trials, approximating ± 7 °/mm or better on tops, and approximately ± 9 -10 °/mm on greasy wool staples.

The precision estimates for greasy staple length were ± 12 to 14 mm in the first trial and ± 10 to 12 mm in the second trial. Given that the OFDA2000 along-staple measurement interval is 5mm, there will be a probable lower limit to the precision of this measurement of ± 10 mm.

In each of these trials, the participants were commercial operators of the instruments, and the procedures used were typical of field operations. The environmental conditions under which the instruments were operated varied overall from 13 to 41 ºC and 9 to 85% relative humidity, indicating that the environmental compensation used in the instrument software was effective.

REFERENCES

Baxter, B. P. (2001a) *Precision of measurement of diameter, and diameter-length profile, of greasy wool staples on-farm, using the OFDA2000 instrument*. Wool Technology and Sheep Breeding, **49** (1), 42-52

Baxter, B. P. (2001b) *On-farm classing of animals & fleeces with the OFDA2000*. Wool Technology and Sheep Breeding, **49** (2), 133-155

Baxter, B. P. (1996) *Description and performance of the Agritest Staple Breaker model 2*. Wool Technology and Sheep Breeding, **44** (2), 119-137

Baxter, B. P. (2002) *The precision and accuracy of OFDA instruments when measuring the curvature of graticules*. IWTO, Barcelona, Spain, CTF 12

StatSoft, Inc. (2001). STATISTICA (data analysis software system), version 6. www.statsoft.com