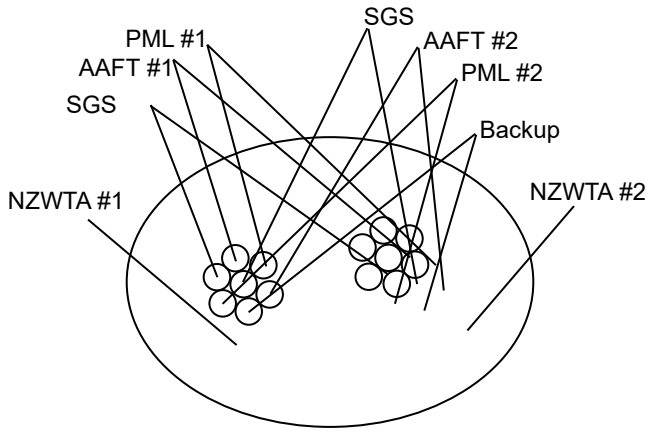
Testing the Testers – Fiber Testing Alpacas
By Stephen Mulholland, Ph.D. for the Camelid Health Trust www.camelidhealth.org

Alpacas are capable of growing fiber of excellent quality, such that many people refer to it as “the fiber of the Gods.” But as we've all experienced, sometimes the Gods have a cruel sense of humor; for every silky-smooth 16 micron fleece, there are plenty of carpet-coarse 38 micron fleeces out there.

In order to breed alpacas with better fleeces, we need to know the quality of the fleeces of our dams and sires. The only quantitative, repeatable, transferrable way to get this information is to submit samples for analysis in one of the many alpaca-fiber-testing laboratories. We are reliant on those results in assessing our animals, yet we rarely ask critical questions about those labs: how consistent and reliable are their test results? What instrument do they use for the test? What information do they provide? Where is the best value for money?

In this study I examined four different labs commonly used by NZ alpaca owners: New Zealand Wool Testing Authority (NZWTA) in Napier, SGS in Timaru, Pastoral Measurement Limited (PML) in Christchurch, and Australia Alpaca Fleece Testing (AAFT) in Australia. For the experiment I gathered fiber samples from 16 alpaca (10 huacaya / 6 suri, 11 females / 5 males), ranging from ~16 to ~36 micron, and from white to blue-black with many colors in between. A larger sample size of 40 to 50 would obviously have been preferred to increase statistical confidence, but AANZ's National Council unfortunately declined funding.

Sample Preparation: In order to measure the consistency of the labs, I had to ensure that the fiber samples were all prepared in a proper manner. This started at shearing where I collected extra-large side samples from the animals to be used in this study. Each such side sample had about 50 grams of fleece. Three of the testing labs (SGS, PML, AAFT) only required a small sample, usually specified as “2 staples”, while NZWTA requires a larger sample on the order of 20 grams, which in the field I interpret as a “generous fistful” of fiber. To reduce error from inconsistencies within a given fleece, I assembled each sample from fiber taken from similar locations in the fleece. For the seven smaller samples (two each for the labs, plus a backup in case a sample was lost or damaged). by taking two clusters of fiber staples about 50 mm apart, and then combining one staple each from each cluster to create the sample for submission. While we know that animals do vary in fleece quality over their bodies, it seemed reasonable to assume that adjacent staples should be relatively consistent. For the NZWTA sample I took the remaining fleece, manually mixed it a bit, then split it in two. A rough graphical representation is provided below:




The samples were submitted to the labs in February and May of 2017. The three month gap between

sample submissions was a way to see how consistent the labs were over time. Scientific instruments can drift out of calibration if not maintained properly, or a different technician might handle the process differently on the day, so checking consistency at different times is a good measure of lab reliability.

I emailed the four laboratories requesting information on how they handled the samples, and how their staff are trained in the use of their instruments, and received replies from all.

In summary:

PML, AAFT and NZWTA wash/degrease all of the samples before testing. This is done with an organic solvent or a quick-drying alcohol. The samples are allowed to dry before scanning. NZWTA specifically allows all samples to "condition" for two hours in a controlled atmosphere before micro-coring, as they have found this produces more consistent results. AAFT commented that they are trialling a data-adjustment "wizard" that may allow alpaca fibers to be scanned without scouring, but as it has not yet been validated they are still scouring the samples they analyze. SGS runs the samples "as is" without degreasing, and apply a software correction factor to the calculated results.

AAFT, SGS and NZWTA have formal internal training methods for their own staff, and only certified operators are allowed to run tests. NZWTA runs four instruments to support the large volumes of wool they study, and they have internal consistency checks between these machines. NZWTA also complies with various ISO-standards for the testing of wool in the international market. PML maintains consistency by conducting all its tests with the same operator (Don Morrison).

AAFT runs their ODFA 2000 with the trim-hi feature turned on, SGS runs with it turned off, unless otherwise instructed by the sample submitter.

The labs had reported expected repeatability of their micron tests as:

| | |
|---|---|
| NZWTA | ~ +/- 1.0 ⌈ |
| AAFT | ~ +/- 0.3 ⌈ |
| ΣΓΣ | ~ +/- 0.3 ⌈ |

The instruments used:

Both AAFT and SGS use the ODFA 2000.

The ODFA 2000 provides the same accuracy and speed as the older ODFA 100, but this instrument can produce more information about each sample analyzed. The instrument uses a microscope video camera to capture images of the fibers which are then computer analyzed to measure the fiber diameter. The accuracy for a single fiber measurement is about 1 micron, but by combining measurements from thousands of fibers the overall accuracy can be as low as 0.01 micron (as per the ODFA specifications, see www.odfa.com) The ODFA 2000 provides data for mean micron, SD, CV, Coarse Edge Micron, comfort factor, spin fineness, staple length, FPFT, SD along, curvature, SDC, as well as micron along the fiber.

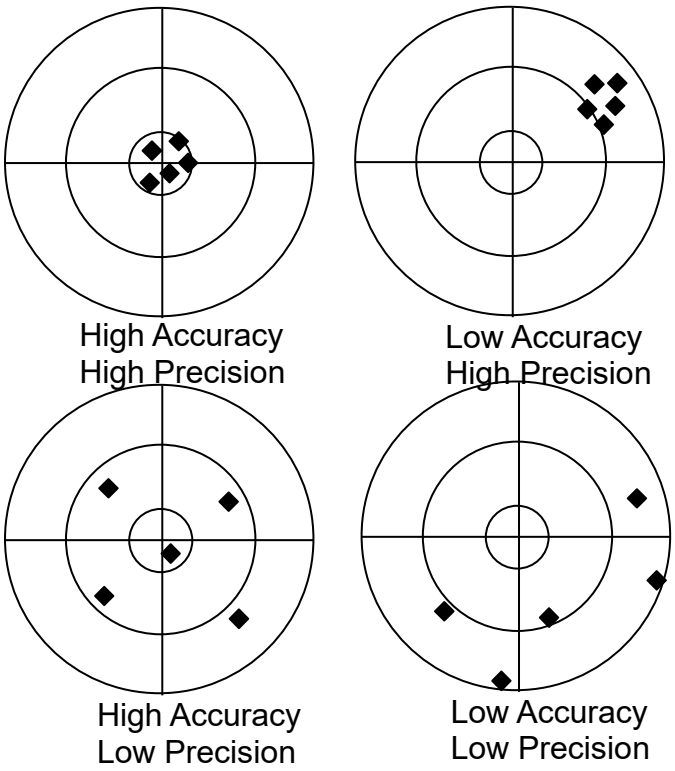
PML uses an instrument of their own design, called Fiberscan Technology, which like the ODFA 2000 captures a high-resolution image of the fibers coupled to a computer analysis of that image. The PML Fiberscan instrument also provides "micron along fiber" data. The Fiberscan instrument provides data on mean micron, mean curvature, staple length, SD, CV, Coarse Edge Micron, Medulation % and comfort factor.

NZWTA typically uses its Laserscan machines for alpaca fiber samples. This instrument was developed by CSIRO for use in the wool industry, and uses "micro-cored" samples. This is when

the fiber is mechanically cut into 2 mm segments before scanning. Because of the sample preparation method the Laserscan can provide neither along-fiber nor staple length measurements. The Laserscan provides information on mean micron, SD, CV and % over 30 micron (which is the inverse of the comfort factor measure). The Laserscan and ODFA 2000 instruments have been long-used in the sheep's wool industry, and IWTO methods have been established for each.

For this paper I'm looking at results for Mean Micron and Standard Deviation of Mean Micron for all four labs, and for the Along-Fiber and Staple Length measurements from SGS, PML and AAFT.

**Measurement Nomenclature – Accuracy versus Precision**




Accuracy is a measure of how closely centered your measurements are in respect to the target. Precision is a measure of how consistent your measurements are. It is possible to be very precise/consistent (a very tight group of measurements), but be way off the target (low accuracy).

I could not measure the absolute accuracy of these instruments, because I have no way of knowing what the "real" microns of the samples I submitted were, so I can't say which answer (if any) was "right." I could measure their precision (tightness of re-tests of the same sample), and their accuracy relative to one another (if a lab gave higher or lower results than the average of all the labs).

It is important to distinguish between accuracy and precision as we look at the results below.

The results in Summary (an average of all 8 tests on each sample):

| Alpaca: | Color | Type | Fleece# | Micron | SD | Staple |
|---|---|---|---|---|---|---|
| H-1 | Fawn | Huacaya | 1 | 15.7 | 3.59 | 92.33 |
| S-1 | Brown | Suri | 1 | 17.7 | 5.2 | 101 |
| S-2 | Lt Fawn | Suri | 1 | 18.6 | 4.16 | 140.5 |
| H-2 | Black | Huacaya | 1 | 19.1 | 4.7 | 76.17 |

| | | | | | | |
|---|---|---|---|---|---|---|
| H-3 | Brown | Huacaya | 5 | 20.3 | 4.79 | 79 |
| H-4 | White | Huacaya | 2 | 20.6 | 5.08 | 103 |
| H-5 | Brown | Huacaya | 2 | 21.7 | 3.94 | 125.5 |
| S-3 | Lt Fawn | Suri | 1 | 22.3 | 5.68 | 141.5 |
| S-4 | Fawn | Suri | 5 | 22.7 | 4.68 | 89.5 |
| H-6 | White | Huacaya | 4 | 23.2 | 4.74 | 81 |
| S-5 | White | Suri | 1 | 23.6 | 5.15 | 137.67 |
| H-7 | Fawn | Huacaya | 3 | 24.3 | 4.53 | 88.33 |
| S-6 | Brown | Suri | 1 | 28.8 | 7.31 | 153.33 |
| H-8 | Black | Huacaya | 8 | 34.4 | 7.96 | 82.33 |
| H-9 | Brown | Huacaya | 9 | 34.7 | 6.23 | 82.33 |
| H-10 | Brown | Huacaya | 6 | 38 | 8.39 | 98.17 |

Where “Black” are Blue/true (double recessive) black alpaca, and where “brown” spans a range from a very dark brown-black to medium brown/dark fawn.

**Overall Measurements – Mean Micron**

Mean micron is probably the most-quoted numerical statistic about an alpaca, and it is a common measure of animal quality. Knowing the variability in measurements is important, as that is how you can differentiate inconsequential difference (when the two numbers are within measurement error of one another) from consequential and meaningful differences.

Comparing micron results between different farms always introduces complicating factors because feed, weather and management practices vary from farm to farm. All of these factors can influence the reported micron of otherwise similar animals. Two of the test providers emphasized that their results are best applied within a herd of animals living on the same farm. Farm to farm comparisons are likely to have larger error bars, such that only proportionally greater differences in measured traits should be considered significant.

**Precision of Measurement of Mean Micron**

We wanted to know first if the labs would show the same result if we sent them the same sample twice. Based on the this difference between the samples mailed in February to those mailed in May, the calculated variation in the results you are likely to see is shown in the table below.

| Lab: | Average Variance | Maximum Variance | Animals with variance > 2 ⌠ |
|---|---|---|---|
| AAFT | +/- 0.41 ⌠ | 1.7 ⌠ | 0 |
| SGS | +/- 0.75 ⌠ | 2.4 ⌠ | 1 |
| PML | +/- 1.50 ⌠ | 5.7 ⌠ | 3 |
| NZWTA | +/- 0.91 ⌠ | 3.6 ⌠ | 2 |

(“⌠” Is the Greek symbol commonly used for “micron.”)

Variance between tests tends to increase as the animals coarsen. Four of the six cases where variance between measurements was above 2 ⌠ happened in animals with fleeces greater than 28 ⌠, though these animals only constituted 25% of the sample set. So where a two micron difference between animals may be significant for fine-fleeced animals, for coarse animals expect much more

noise (variance) in the reported micron numbers.

It is also impossible to consider what role luck played int these results. With such small sample set good or bad luck could have easily made one of the result sets look better or worse than the reality, so do not read these numbers and become overly obsessed in small differences, as they might not reflect reality.

There did not appear to be any great difference in the variance between measurements of white fiber vs blue-black, though the sample set was too small to detect anything less than a major difference. Likewise suri and huacaya fiber appears to have performed similarly in these tests, having about the same level of variance test-to-test.

**Accuracy of Measurement**
This study did not provide for an absolute assessment of the accuracy of the various instruments / laboratories. I could tell roughly which of the samples I sent were coarse and which were fine, but I didn't have an independent, verified answer as to the "true" the mean microns.

What I could measure however was how results of the four labs compared to each other. This was done by averaging the eight results for each animal and generating a mean-of-the-means number, then seeing how the individual results compared against it.

| Lab: | Average Variance from mean of means | Samples above the mean of means | Samples below the mean of means |
|---|---|---|---|
| AAFT | + 0.39 ⌈ | 22 | 10 |
| SGS | + 0.71 ⌈ | 26 | 6 |
| PML | - 1.09 ⌈ | 3 | 29 |
| NZWTA | - 0.01 ⌈ | 15 | 17 |

As you can see, the NZWTA data seemed to sit sight in the middle, on average almost precisely the mean of all the measurements, with half above and half below. The ODFA 2000 instruments used by AAFT and SGS tended to measure higher average microns, while the PML instrument measured lower microns. Which is correct? As I said above, I don't have that answer. But, based on these results, if you are comparing a result from SGS to a a result from PML, you need to account for the fact that, on average, the same fleece will read 2 microns different between the two labs.

**Overall Measurements – Standard Deviation**
Standard deviation is a measure of the variability of fiber diameters within a given fleece. Fleeces with lower SDs will tend to have a better handle, and produce a higher quality garment because a consistent, low SD fleece will have a smaller "tail" of coarse fiber, and thus a better comfort factor.

**Precision of measurement of Standard Deviation**
Average Variance in S.D. between the samples sent in February and May

| Lab: | Variance of S.D. |
|---|---|
| AAFT | +/- 0.32 ⌈ |
| SGS | +/- 0.84 ⌈ |
| PML | +/- 0.66 ⌈ |
| NZWTA | +/- 0.35 ⌈ |

What stood out most in this data was that every sample SGS analyzed in February had a higher measured S.D than the same fleeces analyzed in May. The data from the other three labs had a natural-appearing distribution for S.D. measurements, with an even mix of some up, and some down (7 up &9 down for AAFT, 12&4 for PML, 8&8 for NZWTA). This suggests there might have been a systematic error in the measurements taken by SGS in February (or May), but the sample set was too small to tell if this was just statistical bad luck or a true error.

**Overall Measurements – Staple Length**

The ODFA 2000 and PML instruments also automatically measure the staple length. Compared to the mean micron and SD, this measure seemed have a great deal more variation, and this variation was not limited to the longer suri fleeces.

The six measurements for each alpaca were averaged to generate the average (mean) staple length. Then each individual test was compared to that mean to see how far it fell from the average. The first column in the table below is the average amount each lab varied from that calculated average. The second column shows their most extreme difference from that calculated mean. The final two columns show how often each lab's results were above or below the calculated mean.

| Lab: | Average variance from the calculated mean | Maximum Difference from mean | Measurements Above the mean | Measurements Below the mean |
|---|---|---|---|---|
| AAFT | 6.7 mm | 23.5 mm | 25 | 7 |
| SGS | 10.1 mm | 22.3 mm | 31 | 1 |
| PML | 15.1 mm | 40.5 mm | 4 | 28 |

In this case the PML instrument tended to measure the staple as consistently shorter, while SGS tended to measured it consistently longer. There was no obvious effect of color or fleece type (suri/huacaya) on the accuracy of the measurements.

Independent of this study I also hand-measured the staple length all the samples submitted using a ruler. My results were consistent with the lab-generated results. So this should be something that most people could do on their own consistently and accurately enough.

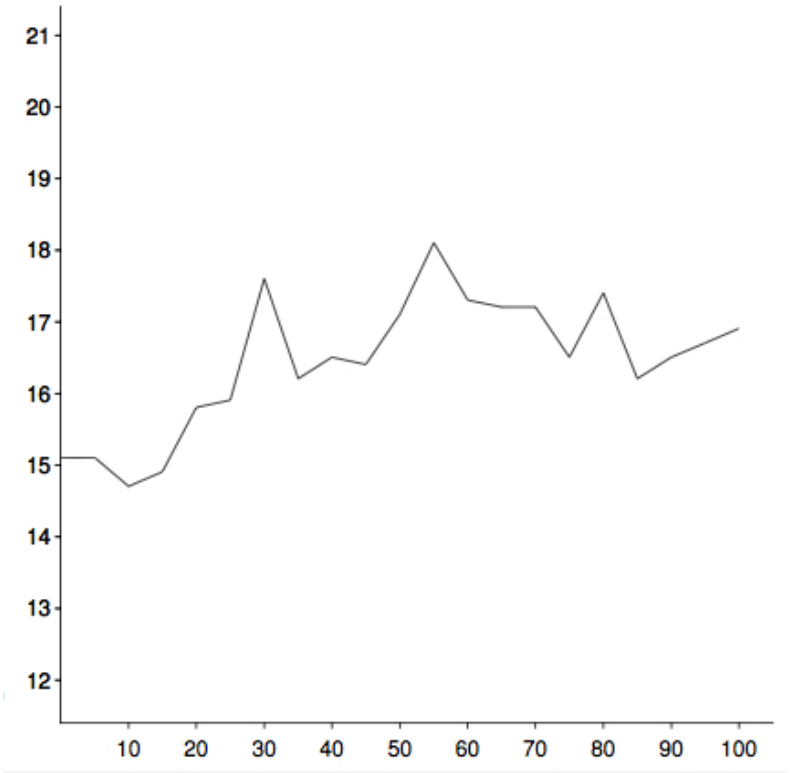
**Measurements of Diameter Along Fiber**

The ODFA 2000 used by SGS and AAFT and PML Fiberscan instrument are also capable of measuring the fiber diameter along the fiber. The along-fiber data can reveal a great deal about an animal, both in terms of genetic and environmental contributions to the mean-micron.

But what about the quality and consistency of the along-fiber data reported by AAFT, SGS and PML?

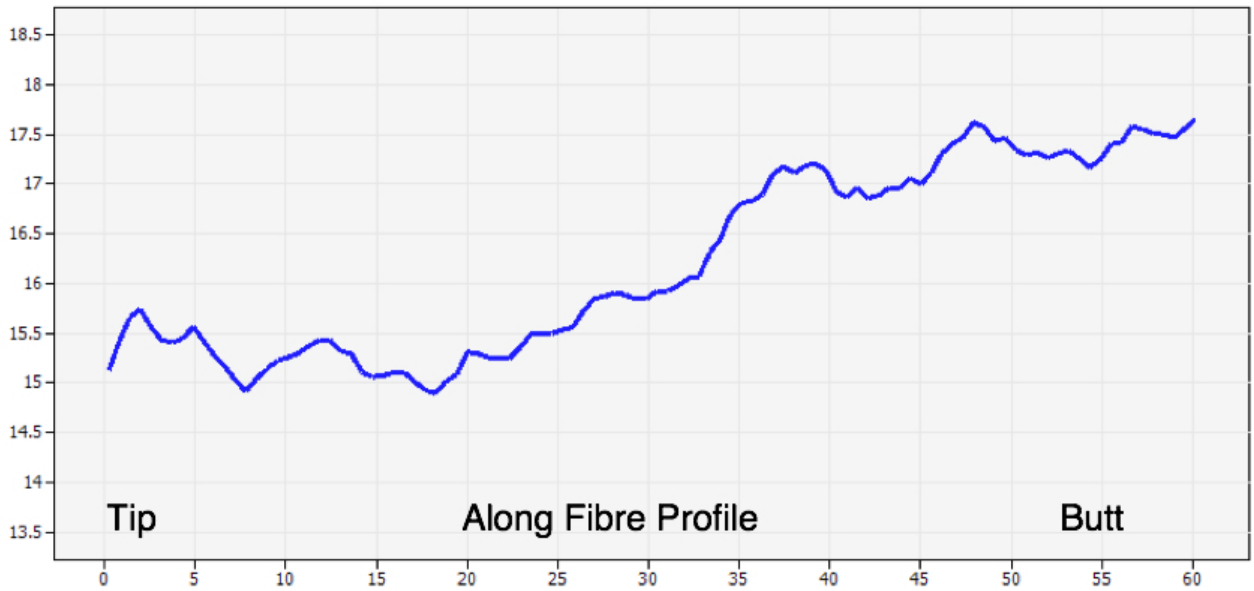
Because along-fiber data can be subject to “noise” (jumps up and down in the measurements along the fiber) it is best suited when looking at samples that have a clear trend. Many of the fiber samples submitted had fiber that didn't change much through the year, as it came from adult animals that had settled into their final adult fleece characteristics. Some of the younger animals had clearer trends in their data. For this example I will use the three along-fiber measurements from May of the animal S-1.

From AAFT we see:



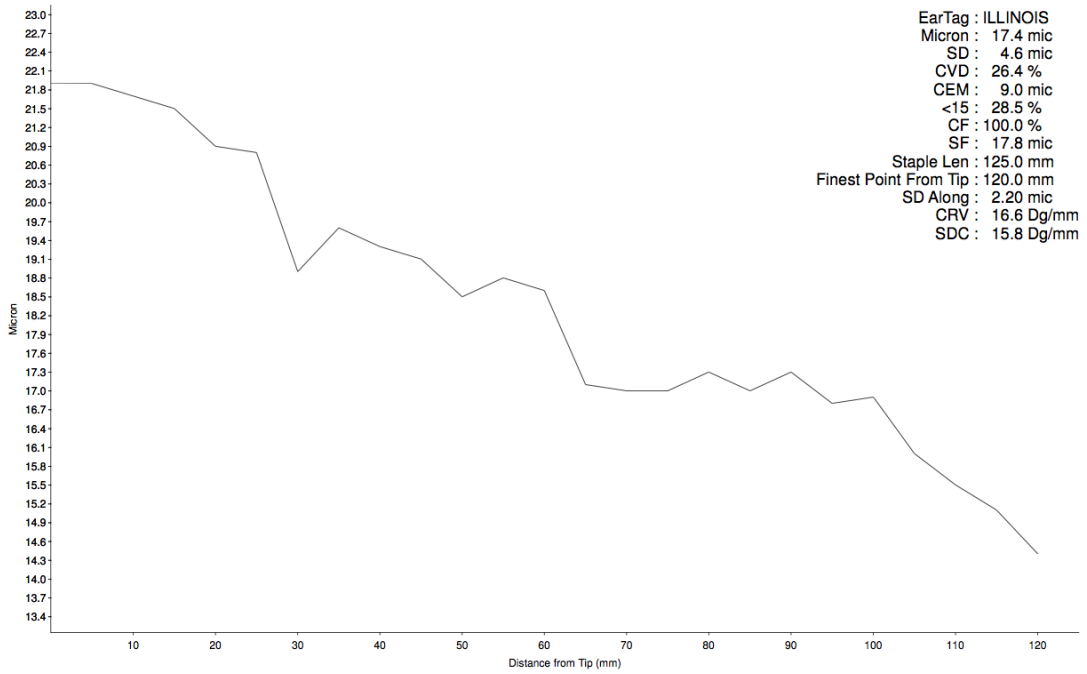

This shows, with some considerable noise, an animal that started with a fleece of about 15 micron and over the year it coarsened to about 17 micron.

From PML we see:




Again the fiber starts fine, about 15 microns in this case, and ends the year it is just over 17 micron. As with the AAFT results, about a 2-3 micron increase over the year. (The PML machine uses a FiberTrac software system which scans the fiber every 19 microns along the staple, the large number of data points this produces is why the PML data looks “smoother” than the ODFA2000 data.)

But when we turn to SGS we see:




Now the fleece started at 22 micron, and ended at 14 micron. Not only is this a large difference in beginning-to-end differential (7 micron compared to the 3 above), but it is moving in the opposite direction! This graph appears to show an animal growing finer during the year, not coarser as we would expect (barring other health or nutrition factors). This is probably an error, or possibly an extreme example of the possible sample-to-sample variability that can happen when running samples.

Serious discrepancies like this is why it is a good idea to double or triple-test animals that may be critical to your breeding program. If the tests don't concur, you know that at least one of them in in error. If all the tests concur it is highly unlikely that they all have suffered from the same error, and thus the result is likely quite trustworthy. In the case above because two of the labs concur, it suggests that the SGS data for alpaca S-1 was probably unreliable.

The “noise” inherent in these graphs meant that broadly speaking it was only safe to draw conclusions about trends in terms of fiber getting finer or coarser if there was a consistent and significant change. For many of the animals sampled there was little change during the year, as they were adult animals on consistent feed whose mean micron simply jumped around within a defined range.

A “backwards” or otherwise seriously wrong answer, if undetected, could cause you reach the wrong conclusions about an animals fleece quality. You can also sometimes spot possibly wrong results by simply comparing to previous years for the same animal. If an animal has been 30 microns for the last five years, but suddenly comes back with a 25-micron result (assuming the animal is not sick or otherwise compromised) you can assume that the outlier result is likely wrong.

**Precision and regularity in along-fiber graphs**

As you can see from the graphs above, there can be a good deal of “noise” in the along-fiber graphs. In an attempt to examine the consistency of these graphs I calculated the “trend” (difference between tip and butt of fiber) and “span” (difference between the minimum maximum values recoded during the year). Where the “span” was equal to or less than the “trend”, there were no large spikes, dips, or slopes in the data.

Comparing pairs of data (e.g. H-1 PML-Feb vs H-1 PML-May) I looked for examples where there was significant deviation (4 or more micron) in the measured "trend" between the two samples. For each of the three along-fiber testing labs, 2 of the 16 sample pairs had a deviation at least this great. So for 1 in 8 of the animals tested at each lab the two results did not concur. In these cases because each sample was in fact tested 6 times (twice for each of the three labs) it was easy to see which result was incorrect. But the high incidence of such errors suggests that a minimum of double testing for critical along-fiber data would be wise, while holding more sample in reserve for re-tests in case of non-consistent data.

**Time for return of results**

All results were received within 7 to 21 days of posting the samples. As I don't know how long it took the samples to arrive, I can't grade any of the labs as being quicker or slower.

**Conclusions**

On our own farm we get all our animals fiber tested every year, and all of our younger animals (at least until the third fleece) and any "core breeding program animals" are double tested at two different lab, at least one of which can provide along-fiber measurements. Double testing the most important breeding animals also provides the advantage that you might catch laboratory errors. If the two test results come back quite different, one is likely incorrect. Holding back sample in reserve can allow of a re-test if it looks as though something has gone wrong; a few dollars well spent when compared to the cost and time investment in our animals!

Because some of the labs appear to produce systematically different results than others (trending finer, coarser, longer, etc), it is good practice to always list what lab you used when displaying any fleece data. If listed stats don't say where they're from, you should probably ask. If they tested at various labs all the different results should be available upon request.

Yes, it is possible to send your fiber samples to multiple labs to "shop" for the "best" results, or to send the many samples of the same fleece to one lab so that normal-statistical-distribution gives you one answer that looks "better." Mis-representing your animals by using unreliable test results is fraud. Faked fiber data can and has resulted in legal actions. Don't do it.

Finally, this sort of "testing the testers" is the sort of thing we as an industry should do regularly. Hopefully in a few years' time the interest can be found to repeat this process, to see which labs are giving the most consistent results, and which are providing the best value for money.

I would like like to thank Phil Cranswick of NZWTA, Eugene O'Sullivan of PML, Jeremy Wear of SGS and Paul Vallely of AAFT for their advice and feedback in the preparation of this article.