2005 Beltwide Cotton Conferences, New Orleans, Louisiana - January 4 - 7, 2005

THE H2SD: INTER-LABORATORY TEST RESULTS

Eric Goze, Richard Frydrych and Jean-Paul Gourlot
Cirad
Montpellier
Serge Lassus and Bruno Bachelier
Montpellier

Abstract

This study involves three High Speed Stickiness Detectors (H2SD) located in three different laboratories. The goal is first, to determine between-measurement variability for each of these devices and second to compare the mean count in each device with the others.

Fifteen cottons covering a wide range of 1 to 60 sticky spots were mixed twice using a laboratory opener, and were then dispatched to the laboratories. Each laboratory made 10 measurements with its H2SD on the 15 cottons following a randomized block design. The entire measurement procedure was replicated twice.

The square of the number of sticky spots was analyzed using a mixed linear model, with fixed laboratory and long-term time effects, and random short-term time effects. The residual variability within the laboratories was approximately the same for all three. Some short term or longer-term drift was observed, depending on the laboratory.

Significant differences were observed between the laboratories, though the correlation between the results was very high (R^2=0.96, 0.97 or 0.99). The usefulness of a calibration is discussed.

Introduction

Six H2SD instruments are now in use in different continents to measure the stickiness of cotton. It is important to verify for commercial as well as for research use, that these different machines yield similar results.

For any given instrument, different measurements repeated on the same material do not yield exactly the same results: some variations are expected from one measurement to another. Further, the more the conditions vary, the greater the expected margin of error. Two notions are commonly used to characterize the quality of an instrument:

Repeatability: the degree of agreement between mutually independent test results produced by the same analyst using the same test method and equipment on random aliquots of the same sample within a short period of time.

Reproducibility: the extent to which a method, test or experiment yields the same or similar results when performed on sub-samples of the same sample by different analysts or laboratories. (EPA, 2005)

Other ground is located between these two extreme notions, e.g. measurements made several months apart by the same analyst in the same laboratory. The aim of this inter-laboratory trial (or round test) is to estimate the margin of error for H2SD measurements made in the same laboratory and the variations from one laboratory to another.

Materials and methods

Experimental design

A set of 15 cottons, covering a range of stickiness from 1 to 60 H2SD sticky spots, was selected from different continents. In accordance with a previous homogeneity study of H2SD counts (Gozé *et al*, 2002), a 900 g sample of each cotton was thoroughly mixed twice using laboratory opener. It was then divided into 3 subsets. At the beginning of each of two measurement sessions, one subset was again divided into 5 to be sent to 5 different laboratories.

In each laboratory, each cotton was tested 10 times on the H2SD. Measurement order was randomized according to a complete block design to control short-term drift effects.

The entire measurement run was replicated two times in each laboratory. The second session was separated from the first by 4 months to estimate the longer term drift effect.

Statistical analysis

As H2SD results are counts, their variance increases with the mean stickiness of the cotton. This precludes the use of such notions as repeatability or reproducibility on raw data.

However, the study of the variability of H2SD counts within cotton bales (Gozé et al, 2002) has shown that, after a cotton has been mixed twice, the variance of the counts in that cotton is proportional to their mean. It can therefore been stabilized by a square root transform. The notions of repeatability and reproducibility are then valid for the square root of the counts. For an experiment such as this, with several factors, a log-linear model would be best, but a more classical linear model is in fact used on the square root of the counts, for the sake of simplicity.

A linear model was constructed by adding together the five possible sources of variations and some interactions:

$$Y_{ij} = a_i + b_j + (ab)_{ij} + c_{jk} + (ac)_{ijk} + D_{jkl}$$

with the following effects:

cotton	a_i
laboratory	b_j
session	c_{jk}
block	D_{jkl}
successive measurements made on the same cotton	E_{ijkl}

(ab) is the interaction between cotton and laboratory, (ac) between cotton and session.

As the session effect was not necessarily the same in different laboratories under different climates, the session effect was nested into the laboratory effect. Likewise, the block effect was nested into the session and laboratory effects.

Repeatability was evaluated by the variance of the measurement error (E). Reproducibility is the variance of the sum of all the environmental effects (laboratory, block and session) plus the variance of the measurement error. It should then be evaluated with all the environmental effects considered as random.

However insufficient laboratories and sessions were involved to determine a laboratory or session variance with precision. Therefore fixed laboratory and session effects were determined and a Fisher-Snedecor test was performed to check whether they were significant or not.

Sufficient blocks were used overall to determine the variance of the block effect D and repeatability E.

This addition of fixed long-term effects and random short-term effects built a linear mixed model. The calculations were made using the restricted maximum likelihood criterion, and the mixed procedure of the Sas® System, version 8.2.

Results

The results were initially analyzed separately for each laboratory. They were then analyzed together.

Within laboratory results

Of the five labs, only three maintained temperature and humidity conditions suitable for reliable results. The results of the remaining two laboratories were discarded.

Fisher-Snedecor tests for the session effect and for cotton x session interaction are presented in Table 1.

Table 1: tests for session effect and its interaction with the cotton effect.

Laboratory		1	2	3
Session effect	F value	F=0.05	**F=8.18**	F=0.09
	P-value	P=0.83	**P=0.01**	P=0.77
	Estimation of the effect	-0.05 ± 0.24	**0.24 ± 0.09**	-0.03 ± 0.11
Cotton x session interaction	F value	F=1.48	F=1.03	F=0.58
	P-value	P=0.12	P=0.42	P=0.88

The session effect was significant in the second laboratory only, at the 1% level. The square root of the H2SD count was shifted by 0.24 from the first session to the second.

On the original scale, using the identity $(A+B)^2 = A^2 + 2AB + B^2$, an original count of X spots was shifted by approximately 2 * 0.24 * the square root of X. For example, a mean count of 10 in the first session was shifted to 11.60 on average in laboratory 2. The session effect was not significant in the other laboratories.

The variances of the block effect and repeatability are shown in table 2.

Table 2: estimated variances for the block and residual effects for each laboratory

Laboratory	1	2	3
Block effect	0.22	0	0.02
Residual =repeatability	0.81	0.55	0.52
Residual, with 6 outliers removed	0.62	0.55	0.52

A block effect was visible in the first laboratory but was negligible in laboratories 2 and 3. Likewise, the residual variance was a little higher in the first laboratory than in the other two. A rapid examination of the residuals showed 6 outliers out of 897 points. The removal of these outliers lowered the residual variance, bringing it closer to those of the other laboratories.

None of the other variances was noticeably affected by the removal of the 6 outliers. As no explanation was found for these (plausible) misrecordings, we kept all the data for the subsequent results.

Inter-laboratory results

F tests for cotton, lab and interaction are shown in table 3.

Table 3: tests for the effects of cotton, laboratory and their interaction

Effect	F	Num df	Den df	P>F
Cotton	504.36	14	747	<.0001
Lab	30.05	2	54	<.0001
Lab x Cotton	2.60	28	747	<.0001

The laboratory x cotton interaction was significant, thereby indicating a laboratory effect, which varies from one cotton to another.

Figure 1 shows for each laboratory the value obtained for each cotton in that laboratory plotted against the average of the three values obtained for the same cotton in the three laboratories. This graphic description shows that laboratory 1 is generally located between the two others, or, depending in the cotton, nearer to laboratory 3. The three laboratories tended to yield more similar results as stickiness increased to very high levels, i.e. above 50 sticky spots.

The tests and the plot show a need for recalibration: either the machines should be adjusted or the results have to be corrected. As an example, calibration equations were computed for laboratories 2 and 3, taking laboratory 1 as a reference.

lab#2 = 0.6413 + 0.9142 lab#1 R^2=0.96
lab#3 = 0.6216 + 0.9684 lab#1 R^2=0.99

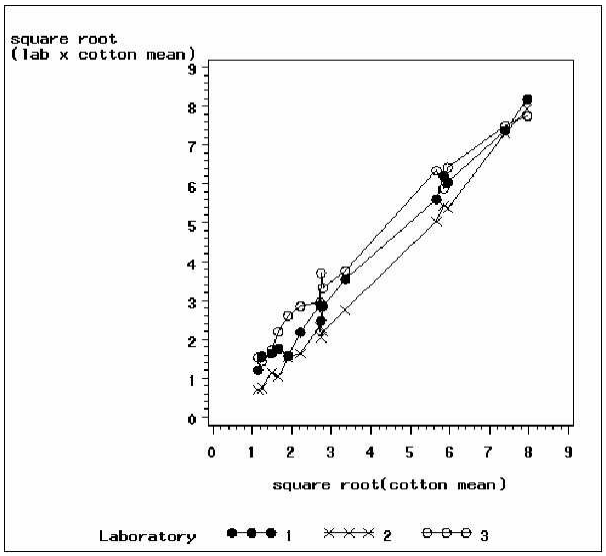

Figure 1: cotton mean for each laboratory plotted against cotton mean over for all laboratories.

Discussion and conclusion

Short-term variability

The short-term variability in 2 laboratories (#2 and #3) was the same as that observed in the within-bale variability study of 2002 (Gozé *et al*, 2002): the residual variance of the square root of the counts was about 0.5. The residual variance in the other laboratory (#1) was higher and superimposed with a block effect: this shows a greater short-term variability.

Longer term variability

One of the three laboratories showed some long-term variability, possibly due to the aging of some components, e.g. lamps or cleaner.

The cotton x session interaction was not significant in any laboratory, meaning that the shift due to the long-time effect, if any, was additive on the square root scale. Hence it can be corrected by calibration.

Discrepancies between laboratories

This round test showed some significant differences between the laboratories. In order to eliminate these discrepancies, either the H2SD can be hardware tuned, or the results can be software calibrated.

The H2SD can be tuned by different means: light intensity, lamp incidence angle, rotary cleaner settings. However a hardware calibration takes some time and is not usually conducted very often, e.g. once a year. Some improvements are about to be made to prevent some of the drifting caused by the aging of instrument components.

Software calibration may still be needed to compensate for the remaining shorter term drifts. Such a calibration has not until now been contemplated because of the good correlation between the results of the different devices, and because it is difficult for a user to admit that after making the necessary corrections, his cotton shows a non-integer number of sticky spots (e.g. 8.56 sticky spots).

Conclusion

We recommend that all H2SD should be tuned using reference cottons so that all yield the same results on average. Check tests should then be conducted on a regular basis to verify that this homogeneity holds. A software calibration should be made possible to compensate any residual drift between the tunings.

Acknowledgements

We wish to thank all the laboratories that participated in this round test. We are particularly grateful to Cotton Inc. (USA) for supporting this project.

References

EPA (2005) EMAP QA Glossary
US Environmental Protection Agency, http://www.epa.gov/emap/html/pubs/docs/resdocs/qa_terms.html

Gozé Eric, Krifa Mourad, Frydrych Richard and Nieweadomski Jean-Charles, 2002, Homogeneity of H2SD Counts Increase with Cotton Mixing - Consequences for Reference Cotton Preparation, ITMF International Committee on Cotton Testing Methods, Bremen 2002 Stickiness, pp. 42.