Bias testing of cross-belt samplers

by G. Lyman*, M. Nel†, F. Lombard‡, R. Steinhaus**, and H. Bartlett§

Journal Paper

Synopsis

The cross-belt sampler is a device that is very widely used even though it has not been ratified by the Standards associations. This paper is concerned with the bias testing of two cross-belt sampler designs against stopped belt samples and a cross-stream cutter at the belt end. The bias analysis is carried out using size distribution as the analyte and conventional *t*-testing and the Hotelling T squared test, which are compared for their effectiveness. The paper considers the issue of detectable bias and the statistical planning of bias tests.

Introduction

The cross-belt sampler has been a contentious machine for many years. The first testing of the device was carried out by Siebtechnik in Germany in 1989 (Reiners and Mohrhauer[1]). This test was carried out on coal which was such that the ash content of the coal was almost constant as a function of particle size, and the bias test was carried out using ash content as the test analyte. Consequently, the machine was given a clean bill of health and the cross-belt or 'hammer' sampler became a well-known piece of equipment. The original machine had a cutter head with parallel side plates which were perpendicular to the motion of the belt. The design concept for the sampler was that it should cut a swath of solids off the belt, of a width determined by the aperture of the cutter, and propel these solids into a chute at the side of the belt from where they would be collected as a sample increment. These machines are still common in the mineral industry. They have the advantage of being relatively cheap to manufacture and can be retrofitted to an existing conveyor at low cost. No transfer point is needed for installation as would be needed for a cross-stream sampler.

Subsequent to the development of the first cross-belt sampler, a machine was designed in South Africa which came to be known as a 'go-belt' sampler. This design was patented by JCI (van der Merwe and English[2]) and had parallel plates that were angled with respect to the belt. The idea was that if the speed of the head was equal to the belt speed, the plates would have an angle of 45 degrees. Such a configuration would make the relative velocity of the solids with respect to the head nominally parallel to the side plates.

The authors are unaware of any detailed sampler test data in the public domain other than the German publication, referenced above.

In line with this situation, the Standard associations around the world have never 'ratified' the cross-belt or go-belt sampler as a device suited to sampling for commercial purposes, even though the machines are used extensively for non-commercial sampling (and some commercial sampling) in the coal industry, largely due to the ease and low cost with which they can be retrofitted to an existing conveyor. The machines are also used as a primary sampler for PGM and gold ores in South Africa. While the samplers are mentioned in sampling Standards, they are not recommended, and it is usually stated that the machine can be used after bias testing if the parties involved in the contract which the samples are used to control are in agreement. Generally, only samplers taking samples from a falling stream of material are endorsed by Standards.

The greatest criticism levelled against the cross-belt sampler is that it tends to leave material behind on the belt. In an attempt to rectify this situation, the van der Merwe patent also introduced the idea of equipping the

* *Materials Sampling and Consulting Pty Ltd, Australia.*

† *Kumba Iron Ore, South Africa.*

‡ *University of Johannesburg, South Africa.*

** *Multotec Process Equipment, Johannesburg, South Africa.*

§ *Consultant, South Africa.*

sampler head with a brush on the trailing edge of the head. The brush was intended to ensure that the belt was left clean and that a correct sample, corresponding to a clean diagonal swath of material being removed from the belt, was collected.

A second criticism is that the cross-belt sampler can lead to breakage of the material. It can be shown that the energy imparted to the solids during sampling is significant at high head speeds and that a 10 m/s head speed corresponds to a 10 m drop of the solids and that the energy dissipated in the solids is a quadratic function of the head speed. Of course, if the solids from any sampler are permitted to drop down a chute from an elevated belt, substantial breakage may occur. This is an issue, particularly with coal samples.

This paper discusses the bias testing of cross-belt samplers of the original design and of the 'go-belt' design. The original design machine differs from the original Siebtechnik machine in that a direct drive is used to power the head rather than a clutched system where the sampler head first falls by gravity and is then picked up and driven through the load on the belt to a point at which is starts to decelerate towards its park position. It was shown by internal test work at Sasol, directed by the first author, that the direct drive system for the head produced better samples than the clutch drive system.

The machines were constructed by Multotec Process Equipment with an aperture of 130 mm and installed on a product belt at the Thabazimbi beneficiation plant of Kumba Iron Ore. The product belt carries nominally 300 tph of washed iron ore, but often less, if one of the washing modules is down. Coarse material (approximately –28 mm, +12.5 mm) is deposited on the belt first with finer material (-12.5 mm, +6.3 mm) on top of the coarse. The belt load is, therefore, strongly vertically segregated, which is quite desirable from the point of view of providing a challenging circumstance for bias testing.

The bias testing was carried out on the size distribution of the material.

The analysis of the results was carried out in two ways; individual *t*-tests, and the Hotelling T-squared test, which is a direct generalization of the *t*-test that takes covariance of the random variables into account.

Test system

The product belt at Thabazimbi has a run of 27.7 m where the belt is accessible downstream of the second installed cross-belt sampler (the go-belt type). There is a belt end sampler of the Vezin type which was refurbished by Multotec. This last machine is mechanically correct. The belt speed is 1.47 m/s and the belt width 0.75 m with 35 degree idlers. Figure 1 shows the layout of the samples on the belt. The sampling was initiated by taking three sets of increments with the cross-belt samplers at 2 second intervals; a belt end sample cut was initiated at the same time as the last cross-belt cut. The belt was then stopped as quickly as possible. A stopped belt increment of nominally 0.4 m was taken, as shown in Figure 1. The belt was then started and a second belt end cut taken about 8.7 seconds after the start, giving the sampling pattern as shown. Care was taken to reject any sample sets where there was any overlap of the sample increments; this could be verified visually.

The full set of samples was taken in a 19 m length of belt. Close spacing of the samples was used to avoid, as far as possible, the introduction of variance between the samples due to distributional heterogeneity. Three cross-belt cuts were taken for every stopped belt sample to match the sample volumes and a stopped belt increment was about 3 times the mass of each cross-belt increment. Two belt end samples were taken to bracket the other samples and these samples were split in half by rotary sample division, after drying, to approximately match the sample mass of the other samples.

Samples from each stop of the belt were combined (3 to 6 stops) to make up a sample of nominally 100 kg, which was then sized on 25, 20, 16, 12.5, 9.5, 8.0, 6.3 mm sieves. The results of the sizings are provided in Appendix 1.

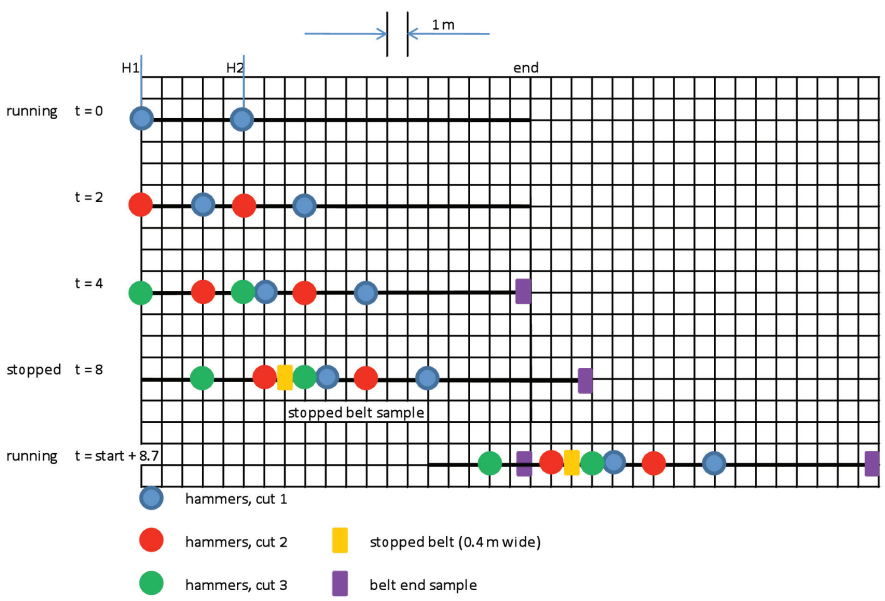

Figure 1—Nominal locations of the samples in the test programme

Bias testing of cross-belt samplers

The detectable level of bias

A bias test has the intention of making comparisons that can be used to decide whether one sampling device or method is biased with respect to another. Implicit within this procedure is the definition of what is an acceptable or unacceptable level of bias. This issue is all too often sidestepped and a test is done on the data at hand and the test system is deemed to be biased if the *t*-tests done on say, ash content or iron content, turns out to be significant. If bias is detected, the detected bias may or may not be economically significant. If one is to go to the expense of carrying out a bias test, it makes a lot of sense to decide before any testing is done just how big a bias must exist before the bias has an economic affect. The test program should be designed to detect the predetermined economically significant bias. The cost of the test program is bound to be inversely proportional to the level of bias to be detected.

Testing a sampling system for bias with respect to particle size, is generally a more powerful method than testing for bias with respect to a chemical constituent, especially when the chosen analyte is present in small concentration. The sample preparation and analysis uncertainties may be sufficiently large under those circumstances that a very large number of paired samples may be needed to arrive at the detectable level of bias that has been targeted.

Samplers commonly show a bias against size if they are in fact biased. A sampler that is biased against particle size is, without doubt, biased against the analytes of interest. Furthermore, if a bias against size shows up in a test when the size distribution is close to the mean distribution expected over time, the sampler will almost certainly be biased when other size distributions are presented to it. Francis Pitard has made the perceptive observation that bias tests can be misleading as biases are by nature not constant in time due to the changing nature of the material sampled and the mass flow rate in the system. However, if a size bias is picked up in one test, it is almost sure to be there, to some extent, in a subsequent test. Actually, as segregation on the belt increases, bias is almost sure to increase. An exception to this is the case of a material having a very wide particle density distribution and a narrow size distribution. Such cases are very rare.

Bias tests that are not planned to detect an economically relevant level of bias are a waste of time and money. Given a knowledge of the size by size distribution of the analytes of interest, the bias induced in the analytes as a consequence of a particular degree of size bias can be estimated.

Consider a case in which a sampler is missing the coarse particles. Such a defect will cause the size distribution to become narrower; a sampler that is missing fines will have the same effect. Therefore, if one is going to design a bias test using sizing as the critical factor, the test should be designed to detect a particular level or extent of narrowing of the size distribution. It is a simple matter to simulate progressively, more narrow size distributions. This narrowing can be readily translated into analyte biases. If the suspected change due to bias is a shift in mean particle size, this effect can also be dealt with by simulation.

Statistical methodologies for data analysis

The comparison to be made in the data analysis is among the stopped belt samples, the belt end samples, and the two samples from the cross-belt samplers. These samples are all paired and so should all have the same expected value of the size distribution. The sample masses are all similar so the variance of the size analysis should be similar for all samples. There should be only two sources of variance attached to a single sample: the difference between the 100 kg samples due to the intrinsic heterogeneity with respect to particle size and the uncertainty associated with the sizing procedure itself. The careful pairing of the samples should eliminate the uncertainty due to distributional heterogeneity along the belt as far as possible.

In making a statistical comparison between two measured size distributions, it must be recognized that the mass fraction reporting to a given size fraction is statistically correlated with the mass fraction reporting to any other size fraction. The correlation comes about as a result of having added up the masses in each size fraction and then dividing each mass by the total to arrive at the mass fraction in the size fraction. Therefore, it is incorrect to treat the amounts in each size fraction (or the amounts passing or retained on a given sieve) as independent random variables.

Treating them as such is, however, a common practice. This practice is justified only when a single measure of bias is involved (e.g a single analyte assay). When more than one analyte and particularly when sizings are involved in making paired comparisons, the covariance structure of the data must be taken into account. The temptation to carry out individual *t*-tests on the mean differences derived from *n* independent measures of the difference should be avoided.

There are three problems associated with such a procedure. First, working only with mean differences ignores that fact that the bias in the sampler may be a bias of scale which is not discovered using mean differences (see Lyman[3]). Secondly, covariance in the data set is ignored, and thirdly, one has to decide how many of the *t*-tests must fail in order for the sampler to be deemed biased.

Does it suffice that one *t* test fail at the 95% level? Or must a certain fraction of all possible *t*-tests fail at the 95% level. With *p* size fractions defined, it is possible to carry out $p - 1$ *t*-tests (the *p*th test is not independent due to the correlations of the mass fractions). Appendix 3 deals with this problem comparing *t*-testing to the Hotelling T-squared test, which fully accounts for the correlation between the size fractions and provides a single test statistic for the entire group of size fractions. It is shown there that determination of the number of *t*-tests that should fail before deciding is not trivial and that deciding, for example, that three tests must fail for the sizing to be different may be too strong a criterion, whereas deciding that two failures are sufficient may be too weak a criterion. Use of the Hotelling statistic entirely circumvents this problem.

Use of the t*-statistic*

Consider first a simple means of testing the difference between two size distributions. Assume that there are *n* pairs of observations on the size distribution which can be denoted as observations x_i, y_i. These vectors contain *p* elements giving, say, the mass fraction within *p* of the $p + 1$ size fractions. The difference vectors, using *p* size fractions are

$$\boldsymbol{\delta}_i = \mathbf{x}_i - \mathbf{y}_i \quad [1]$$

These can be put into a matrix D (n columns x p rows), where the δ form the columns of D.

The mean difference vector is

$$\bar{\boldsymbol{\delta}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\delta}_i \quad [2]$$

The standard deviation of the mean difference of the mass fractions in each size fraction can be calculated as

$$\hat{s}_{\bar{\delta}_j} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\delta_{ij} - \bar{\delta}_j\right)^2} \quad [3]$$

where δ_{ij} is the jth element of the vector δ_i and δ_i is the jth element of the vector δ.

If one were to focus interest on a single size fraction, one would use a t-test and form the statistic

$$t_{\bar{\delta}_j} = \frac{\bar{\delta}_j}{s_{\bar{\delta}_j}} \quad [4]$$

and this has a Student's t distribution with $n - 1$ degrees of freedom. This test could be applied to each size fraction of interest in turn.

The difficulty with carrying out such a test is that while the mass fraction in one size may appear to be biased (that is, it fails to pass the t-test), the material that is apparently missing (or in excess) from that size fraction must consistently appear in some other size fraction or set of size fractions for there to be a consistent bias. Therefore, t-testing based on single size fractions fails to take account of the fact that the missing or excess material must go into or come from some other size fraction(s). When one considers that it is necessary to test more than one size fraction at a time, it should be immediately recognized that one is dealing with two correlated, random variables and that simple t-tests that assume independence of the two random variables are inappropriate. The Hotelling T-squared test intrinsically recognizes the correlation or covariance between mass fractions in the various size fractions as is appropriate to the technical realities of the situation.

Use of the Hotelling T squared statistic

Rather than simply estimating the variance of the amounts in the size fractions, one estimates the covariance matrix of the differences in the amounts in the p size fractions. This matrix is estimated in an unbiased manner as

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-1}\left(\mathbf{D} - \bar{\boldsymbol{\delta}}\right)\left(\mathbf{D} - \bar{\boldsymbol{\delta}}\right)^T \quad [5]$$

The T-squared statistic is then

$$T^2 = \frac{n(n-p)}{p}\bar{\boldsymbol{\delta}}^T\left(\left(\mathbf{D} - \bar{\boldsymbol{\delta}}\right)\left(\mathbf{D} - \bar{\boldsymbol{\delta}}\right)^T\right)^{-1}\bar{\boldsymbol{\delta}} \quad [6]$$

and this is distributed as $F(p, n - p)$.

Note that one must have at least $p + 1$ pairs of observations if the test is to be possible. With $p + 1$ pairs, the F distribution has $(p,1)$ degrees of freedom. This distribution is broad and the power of the test will be poor unless the number of observations exceeds $p + 1$ by some margin.

Note also that the T-squared statistic is proportional to the square of the Mahalanobis distance of the mean difference from the origin. The virtue of the T-squared statistic is that it can be simply tested for significance using the F distribution, if the data distribution is normal.

One must not make the mistake of using all $p + 1$ size fractions in the test as the estimated covariance matrix is singular in such a case as a result of the fact that the mass fraction in any size fraction can be found by subtracting the sum of the amounts in all other size fractions from one. The singular matrix cannot be inverted.

Setting the detectable level of bias in the bias test plan

The planning of a bias test must begin by estimating the size distribution of the material in the process stream sampled by the sampler to be tested. Then, it is necessary to determine the precision of the sizing procedure to be used for the samples taken from the process stream. This must be done by collecting a large sample of material from the process stream, homogenizing and dividing it by rotary, sample division into a substantial number of nominally identical subsamples (10 to 15 is good and there must be more subsamples than there are size fractions defined). The mass of the subsamples to be sized must be large enough that the sizing is reasonably accurate. Smaller samples will carry larger uncertainties due to Poisson error and will lead to a requirement for a larger number of pairs of samples to attain the detectable bias target.

For the test work that is reported in this paper, the covariance matrix derived from the sizings of 15 nominally identical 100 kg samples was compared with the covariance matrix derived from the simulation of sampling of an iron ore with a similar size distribution. The simulation allowed for a Poisson uncertainty in the number of particles in each size fraction and a normally distributed random uncertainty in the mass of material reporting to each size fraction. The former component is what is described by Gy's fundamental error as applied to sampling for size distribution and the latter is intended to account for misplacement of material from a size fraction or loss of material in the sizing process.

The result of sizing of the 15 nominally identical samples is reported in Appendix 1 (see Table A1-II). The size distribution used for simulation is shown in Figure 2. This distribution is slightly different from the observed mean distribution, but the initial distribution used for planning the tests differs from that found in the test work: such a difference is a hazard of the planning, but the differences are not gross. From that Table, with the exception of the coarsest size fraction, it is clear that the fundamental error is smaller than the observed uncertainty. If one allows for an uncertainty of 2.4% relative (1 sigma) due to the 'analytical' uncertainty, the observed uncertainties are adequately modelled by the simulated uncertainties.

Using the size distribution of Figure 2, which is linear on the Rosin Rammler coordinates, it is easy to make the distribution narrower simply by pivoting the line about a point. The 40% passing size was chosen as a pivot.

The determination of detectable bias is described in Appendix 3. Basically, one changes the size distribution from the base distribution and then carries out Monte Carlo simulations of sampling from both the base and biased size

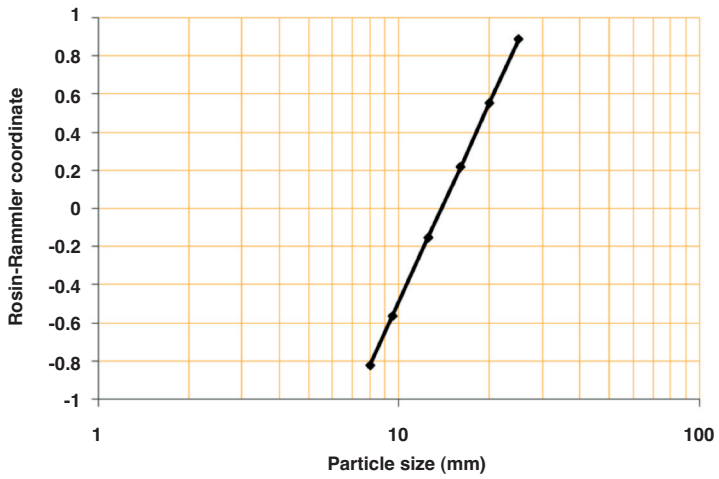

Figure 2—Size distribution used for estimation of detectable bias

distributions, forming the desired numbers of pairs of samples from such samples. For example, 10 pairs of samples were selected for the work in Appendix 3. These ten pairs can be used to calculate a T squared statistic and the value compared to the appropriate *F* statistic. This sampling and comparison procedure can be carried out thousands of times to estimate the probability that the samples will fail the *F* test. When the bias between the size distributions becomes large enough, the probability of the failure of the *F* test will rise to 95%. It can then be said that the bias between the two distributions is detectable at the 95% level.

If the level of bias that can be detected at the 95% level is deemed to be too large, the number of pairs of samples used can be increased (from the 10 used as a trial value) and the Monte Carlo simulation carried out again. The result will be an improvement in the amount of detectable bias. Appendix 3 shows that with 10 paired samples, and with the exception of the coarsest size fraction, a difference of less that 1.0% (absolute) in the amount in a size fraction can be detected. Including the coarsest material raises this figure to 2%. Note also that such differences expressed as relative uncertainties are of order 7 to14%, which exceeds the uncertainty associated with the determination of the size distribution by some margin.

The test was designed to consist of up to 20 paired samples so that the detectable bias would be under 2% (about 1.3%). In the final case, 14 pairs of samples were collected each of, nominally, 100 kg.

Results

It is instructive to note that based on the aperture of the hammer samplers, the actual sample masses were 83 and 85% of the expected masses for samplers H1 and H2. This suggests that both hammer samplers were leaving material on the belt (or that the usual method of estimating the hammer sample increment mass is in error).

The increment masses were checked after drying of the samples to ensure that the samples had not been corrupted. The expected increment masses were compared with the actual increment masses. For a given sampler, these ratios were normally distributed over the 100 samples collected. For the first hammer sampler (H1), it was observed that 10 of the samples were substantially heavier than expected. During the sampling, it happened that when the belt was restarted after collection of a set of samples, the belt was overloaded and there was spillage down the hammer sampler collection chute. This spillage had to be dumped from the sample bucket before collection of the next set of samples. Based on the suspicion that some non sample material might have been present in the bucket, increasing the sample weight, those sets of increments where the hammer 1 sample was too heavy were rejected and not used to make up the nominally 100 kg samples that were sized.

The data collected are provided in Appendix 1. The belt end, hammer 1 (conventional hammer sampler) and the hammer 2 (angled hammer sampler or 'go-belt' sampler) were tested by taking the differences between size distributions that they collected and the belt cut (stopped belt) samples.

The results of making individual *t*-tests are provided in Table I. The belt end sampler is apparently unbiased whereas the other two samplers are significantly biased at the 95% confidence level.

It is relevant to note that the mean differences for the belt end sampler are under 1% absolute, with some about 0.5% absolute, which is of the order of the detectable bias (see Figure A2-2). The differences for the hammer samplers exceed 0.5% in most cases and are, therefore, of a magnitude which will cause the Hotelling test to indicate bias and to cause a number of *t*-tests to fail.

The results of the Hotelling T squared tests are provided in Table II.

The Hotelling T squared tests show all samplers to be biased compared to the belt cuts. To find that the belt end sampler is biased according to the Hotelling test is surprising and one must seek a reason for the apparent assignment of bias. To look for causes of failure of the Hotelling T squared test, it is instructive to make box plots of the distribution of the differences for the 14 samples for each of the comparisons.

Table I

t-test results (significant results shown in red). Differences correspond to 'test' – 'reference' sample. Size fractions run from + 25 mm to 6.3 mm

Belt end				Hammer 1				Hammer 2			
Mean difference (% abs)	SD mean difference	Rel difference (%)	t stat	Mean difference (% abs)	SD mean difference	Rel difference (%)	t stat	Mean difference (% abs)	SD mean difference	Rel difference (%)	t stat
0.649	0.318	5.41	2.04	2.481	0.483	20.71	5.14	2.450	0.490	20.45	5.00
-0.482	0.394	-3.82	-1.22	1.819	0.496	14.42	3.67	-1.241	0.391	-9.83	-3.17
-0.086	0.257	-0.73	-0.33	-0.287	0.169	-2.44	-1.70	-2.877	0.313	-24.45	-9.18
-0.194	0.484	-1.27	-0.40	-1.058	0.343	-6.93	-3.08	0.194	0.496	1.27	0.39
-0.616	0.313	-3.15	-1.97	-1.750	0.336	-8.95	-5.21	1.667	0.536	8.53	3.11
-0.188	0.274	-1.64	-0.69	-0.811	0.227	-7.10	-3.57	-0.012	0.303	-0.11	-0.04
0.029	0.279	0.26	0.10	-0.211	0.350	-1.85	-0.60	-0.150	0.319	-1.31	-0.47
0.160	0.250	2.68	0.64	-0.050	0.267	-0.84	-0.19	0.101	0.310	1.68	0.33

Table II

Results of Hotelling T squared test

Sampler	BE	H1	H2
T squared stat	8.71	7.47	9.60
F probability	0.0053	0.0083	0.0040

Figure 3 shows such plots. The box shows the median (red line) and the 25 and 75% quartiles (limits of box). The 'whiskers' show the value of the last observation that does not differ by more than 1.5 times the inter quartile range (height of the box). Points shown as red crosses can be considered to be outliers if the points are assumed to follow a normal distribution.

It is evident that there are a number of suspected outliers (7) in the comparison between the belt end sampler and the belt cuts. It appears that these outliers have contributed substantially to the result for the Hotelling statistic.

It is possible to devise a ranked version of the Hotelling test in which the observed differences are replaced by their signed rank. In this case, it is not correct to use the F statistic to determine the significance of the outcome of this modified Hotelling test. Instead, the distribution of the 'robustified' statistic must be found using a permutation test and the significance determined in that manner. Note that permutation tests underpin the original derivation of the *t* statistic by Student. Using the ranked or robustified version of the Hotelling test, the hammer samplers are found to have significant bias and the belt end sampler to have no significant bias. The further use of this test and especially a means of connecting the outcome of the test to a detectable level of bias is currently under study by Lyman and Lombard.

The box plots provide a picture of the nature of the bias suffered by the hammer samplers. The +25 mm material is overestimated. For the conventional hammer sampler (hammer 1), the 20 x 25 mm is also overestimated with compensation taking place by an undersampling of the 12 x 8 mm material. For the go-belt sampler, 20 x 12.5 mm material is undersampled and 12.5 x 9.5 material oversampled.

It was observed that the hammer samplers did not cut a clean 'swathe' across the belt. There was a good deal of material left at the side of the belt, despite the fact that the machines were adjusted prior to the tests. It is also to be noted that the material on the belt was heavily segregated with coarse material on the bottom of the load and finer material on top. There was evidence of 'ploughing' of material behind the conventional hammer sampler and loss of material from the increment on the go-belt sampler. The observed bias in these samplers was not unexpected, given the observation of the belt. It appeared that the material on the belt was difficult to remove as the coarse material at the bottom of the load tended to 'lock up'.

Conclusions

Bias testing of a sampler is a critical factor in the testing of a machine. Beyond the collection of the magical 30 pairs of samples suggested by some Standards, it is rare to see any statistically rigorous discussion of the level of detectable bias that will be achieved in the bias test. It is obvious that the detectable bias is a figure of economic significance.

The application of individual *t*-tests to the results of testing a sampler on size distribution is incorrect as such testing ignores correlation between the amounts in size fractions. *t*-testing when more than one analyte is used is also incorrect when the analyte contents in samples are correlated (as they usually are).

The Hotelling T squared test, correctly, takes correlation in the data into account when tests on simple differences are used. However, as apparently illustrated by the results in this paper, the Hotelling test is sensitive to outliers and general non normality of the data. A 'robustified' version of the Hotelling test can be devised based on the signed rank of the differences observed. Having replaced the actual differences by their signed ranks, the calculations for the robustified method do not differ from the method described in here. The critical value of the robustified statistic must be determined by permutation tests.

The important point to note is that replicate analyses are necessary to assess the accuracy of the size analysis or analyte analyses carried out. In spite of working many years

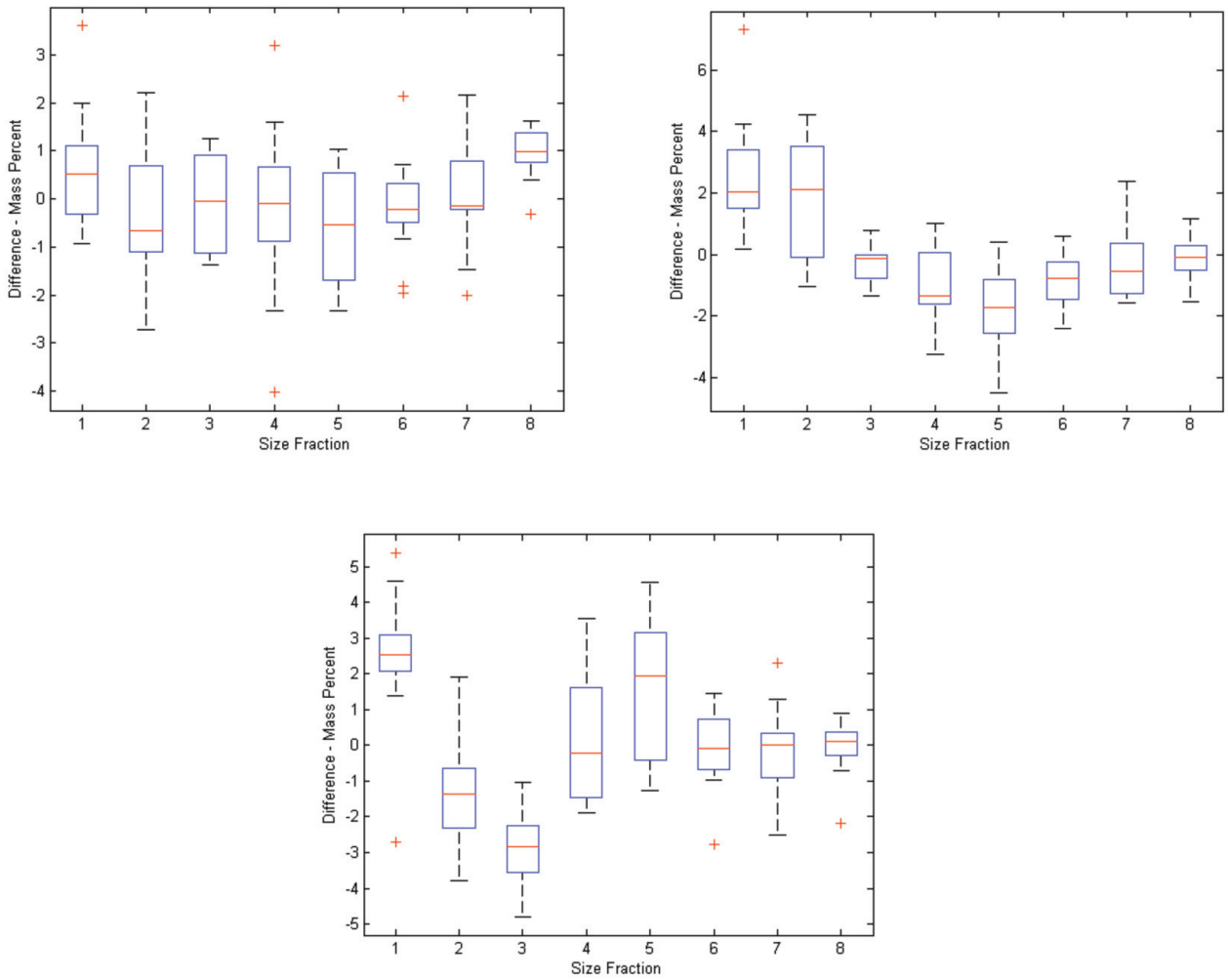

Figure 3—Box plots for samplers: Top left: belt end; top right: hammer 1; bottom, hammer 2. Size fraction 1 is +25 mm and fraction 8 is -6.3 mm. Note the difference in vertical scales on the plots

in association with mineral processors, the first author has rarely seen replicate size analyses carried out with the objective of measuring the accuracy (or at least the precision) of the sizing method. Without this information, the detectable bias in a bias test based on size distribution cannot be reliably estimated. The same is true of tests based on chosen analytes. For the masses used for each pair of samples in the bias test, the total sampling plus preparation plus analytical variance and the covariance between the analytes must be quantified before the detectable bias can be estimated.

The issue of using Hotelling tests in a robust manner is under study by Lyman and Lombard. The issue of estimating both absolute bias and bias of scale for individual size fractions or a set of analytes is being studied by Lyman.

Acknowledgements

- *Thabazimbi mine*—The authors would like to acknowledge Thabazimbi Mine, in particular Cornelia Holtzhausen, for allowing the mine to be used as a test facility.
- *Kumba Iron Ore Management*—We would like to acknowledge the following from Kumba Iron Ore Management for giving their approval to do this very important test work:
  - Chris Griffith (CEO)
  - Christo van Loggerenberg (Head of Technology and Business Development
- *OreProx*—We would like to acknowledge OreProx, Sample Preparation Laboratory based in Kimberley, Northern Cape, South Africa for part sponsoring and conducting the required sample preparation work to generate the source data used in this paper. Their work was done quickly by a very competent crew.

References

1. Reiners, E. and Mohrhauer, P. Hammer Sampler—a proven sampling unit, *Aufbereitungs-Technik*, vol. 30, no. 4. pp. 218–225.
2. Van der Merwe, E. English; Ralph H, Device for sampling material from a moving belt conveyor, US Patent No. 5,115,688, filed 1990.
3. Lyman, G.J. Determination of Bias in Sampling Systems, *Minerals Engineering*, vol. 6. pp. 99–105.

Bias testing of cross-belt samplers

Appendix 1

Data from bias test

Table A1-I

Data from bias test

Sample	Initial mass (g)	Mass % in fraction							
		+25	25 x 20	16 x 20	12.5 x 16	9.5 x 12.5	8 x 9.5	6.3 x 8	-6.3
H1 Sample 1	106890.7	9.182651	13.80045	13.01778	16.95077	20.31739	11.681	10.16693	4.883025
H2 Sample 1	96893.6	9.620759	12.10317	11.486	19.34173	22.55536	9.946787	9.824178	5.12201
BE Sample 1	122211.4	7.983216	11.48379	13.11072	16.53864	20.79912	12.404	11.29723	6.383283
BC Sample 1	131795.2	7.007691	13.87979	13.16315	15.93283	20.37146	12.72854	11.51195	5.404597
H1 Sample 2	116468.7	9.180664	14.11761	13.22973	17.38922	19.04486	10.68244	10.76032	5.595151
H2 Sample 2	107837.8	9.418404	11.02025	10.07365	18.29831	22.62574	12.23078	10.80706	5.525799
BE Sample 2	129272	7.643186	11.15145	13.84646	16.28419	20.89927	11.54225	11.54643	7.086763
BC Sample 2	151180.6	6.729567	11.78438	12.92679	16.70472	21.61607	12.36878	11.77115	6.098534
H1 Sample 3	63533.6	13.29722	18.56608	12.31726	14.04312	16.99762	9.533381	9.210717	6.034602
H2 Sample 3	92819.7	13.56005	12.81862	7.972984	14.46708	21.70035	11.95393	10.79361	6.733377
BE Sample 3	107154	10.85802	13.96607	11.38558	15.56097	19.11277	11.35123	10.63544	7.129925
BC Sample 3	119238.9	11.41934	14.18354	12.75129	13.96549	18.55217	11.23291	10.45892	7.436332
H1 Sample 4	80191.7	10.08758	14.99557	13.12243	13.70616	18.13679	11.1529	12.43346	6.365123
H2 Sample 4	96437.1	8.774424	10.64217	9.684447	15.13277	25.26165	11.78219	12.68132	6.041036
BE Sample 4	111628.5	7.796127	12.11232	13.49593	14.60765	19.97859	11.9723	13.03941	6.997675
BC Sample 4	105801.8	6.677864	10.44103	13.24221	16.93884	20.69747	12.45366	13.59249	5.95642
H1 Sample 5	95208.5	13.05923	12.94412	12.10092	17.38847	19.87669	10.60231	9.890609	4.137656
H2 Sample 5	97659.5	13.64138	11.41589	7.838357	18.96969	23.00903	11.29445	9.522371	4.308849
BE Sample 5	108957.2	11.11813	12.7328	12.61569	15.89101	20.23629	11.46312	10.35875	5.584211
BC Sample 5	118741	10.55086	12.05321	12.64391	19.90635	20.60527	10.75366	9.530491	3.956258
H1 Sample 6	80443.7	23.98808	18.1517	12.17671	11.54447	14.27433	8.095103	7.751135	4.018463
H2 Sample 6	93998.4	23.9652	16.39539	10.05751	11.8202	15.85101	8.94079	8.263864	4.706038
BE Sample 6	110978.6	21.62372	18.09862	11.08313	12.45051	14.45774	8.550117	8.019384	5.716778
BC Sample 6	113439.8	22.05355	19.17325	12.30194	11.78043	13.8707	8.328118	8.226566	4.265434
H1 Sample 7	101970.7	15.92889	16.97664	11.37954	12.15388	16.4409	10.02965	11.05602	6.034479
H2 Sample 7	110455	18.22561	15.86447	9.741343	12.88733	17.05835	9.69182	10.76022	5.770857
BE Sample 7	118381.2	16.01175	14.28419	11.18328	14.57715	15.40388	11.89902	10.64358	5.997152
BC Sample 7	120593.7	15.74875	15.38463	12.29815	14.7686	17.736	9.74769	9.456547	4.859624
H1 Sample 8	96406.8	13.25166	17.58932	11.46071	13.05582	16.54406	11.35646	10.96614	5.775837
H2 Sample 8	106661.3	16.47917	10.15176	8.901917	14.80481	19.91866	12.13055	11.77831	5.834825
BE Sample 8	119804.4	12.34562	13.09676	12.48819	13.7818	17.68975	11.09183	12.06049	7.445553
BC Sample 8	122111.3	11.88948	13.94416	12.65952	14.65827	18.92601	10.75879	11.2621	5.901665
H1 Sample 9	73585.5	11.61003	14.42703	11.08112	13.34652	19.19563	12.78078	10.97716	6.581731
H2 Sample 9	98025.6	12.82451	12.79635	8.645293	13.9078	20.03864	12.23282	12.80614	6.748441
BE Sample 9	107699.7	11.46224	11.65351	12.37153	14.35909	19.43432	11.17905	12.34005	7.200206
BC Sample 9	112416.7	10.00101	10.89509	11.21631	14.66401	21.12782	13.15481	12.5541	6.386862
H1 Sample 10	75343.8	11.45482	12.9716	10.66936	16.68034	18.95591	10.73134	12.26471	6.271916
H2 Sample 10	103269.6	12.07974	9.81189	9.004392	16.50447	22.1863	11.89314	12.17735	6.342718
BE Sample 10	116373.3	9.90442	10.73537	11.29898	15.71314	21.13045	12.10742	12.04108	7.069147
BC Sample 10	117126.8	9.956048	10.76568	10.04569	18.0226	23.45108	12.19593	9.884672	5.678291
H1 Sample 11	77859.4	13.99934	10.84609	11.26184	15.25031	18.18997	10.67591	12.52283	7.253716
H2 Sample 11	104637	13.42948	9.466154	9.431081	13.44763	20.29942	13.21378	13.2672	7.445263
BE Sample 11	118804.5	11.75275	10.44304	11.67464	14.89211	21.25551	9.945415	11.93246	8.10407
BC Sample 11	129356.5	9.756139	11.10729	10.45537	14.90965	20.70735	11.76617	13.9533	7.344741
H1 Sample 12	59259.5	17.46555	11.21761	9.152963	12.36392	17.96134	11.50432	13.40241	6.931884
H2 Sample 12	95509.2	15.06808	7.451324	7.680098	17.36775	23.43575	11.8041	11.48549	5.707408
BE Sample 12	95468.5	12.75363	11.01777	9.285052	14.887	19.02586	11.5349	13.91328	7.582501
BC Sample 12	113510.4	13.67593	8.808444	10.47957	13.80587	19.30396	12.0196	14.01246	7.894167
H1 Sample 13	102442.6	17.9318	13.98412	9.529336	14.25559	18.28985	9.729741	10.63991	5.639646
H2 Sample 13	112105.7	14.84037	9.613963	6.479064	18.11139	23.55563	11.18016	10.74361	5.475814
BE Sample 13	129651.8	17.22876	9.531067	9.593388	19.02434	20.1391	9.506308	9.234812	5.742226
BC Sample 13	132705.7	17.54537	11.92616	9.562815	15.81711	19.10204	9.997008	10.69901	5.350486
H1 Sample 14	116895.2	22.02879	11.51809	10.24045	10.64475	15.02962	9.922991	14.7912	5.824106
H2 Sample 14	107281.2	20.10324	9.71736	7.490781	11.24148	19.60754	11.36583	12.77726	7.696502
BE Sample 14	128267.2	18.33087	9.580158	10.12792	12.30237	15.57935	12.65904	13.13368	8.286608
BC Sample 14	139238.4	14.719	12.29029	11.01729	11.7069	17.69332	12.32663	12.87454	7.372032

Table A1-II

Replicate size distribution data

Sample	Fraction (mm)						
	+25	20 x 25	16 x 20	12.5 x 16	9.5 x 12.5	8 x 9.5	-8
1	2.960	14.057	12.489	12.095	14.350	9.298	34.751
2	2.985	14.073	11.945	12.745	14.309	8.634	35.307
3	3.112	14.719	11.212	12.484	14.333	8.936	35.204
4	3.023	14.905	11.091	11.971	14.568	8.969	35.473
5	2.772	14.376	12.330	13.310	15.449	8.381	33.382
6	2.856	14.308	12.123	12.681	14.213	8.784	35.035
7	2.479	15.018	12.220	12.499	14.000	8.703	35.081
8	2.955	14.966	12.537	12.126	14.342	8.770	34.303
9	3.161	14.689	11.625	11.546	14.754	8.778	35.447
10	3.112	15.155	11.897	11.634	14.669	8.748	34.785
11	3.510	14.583	10.088	12.201	14.545	8.574	36.499
12	2.849	15.233	12.177	12.416	14.104	8.656	34.565
13	3.080	15.020	11.871	11.447	14.788	8.990	34.805
14	3.261	14.717	12.100	11.770	14.025	8.657	35.471
15	3.284	14.957	11.835	11.818	14.073	8.970	35.063
Average.	3.027	14.719	11.836	12.183	14.435	8.790	35.011
SD actual	0.243	0.373	0.635	0.513	0.381	0.218	0.680
RSD actual (%)	8.028	2.534	5.363	4.213	2.637	2.485	1.943
RSD fund. (%)*	13.250	3.811	3.208	2.332	1.611	1.573	0.992

*Based only on the Poisson distribution of particle numbers in the size fractions

Appendix 2—Comparison of the hotelling T-squared test and individual *t*-tests for paired size distribution observations: detectable bias

The objective of this appendix is to compare the behaviour of individual *t*-tests on the difference in mass fractions, within size fractions and the Hotelling T-squared test, on the set of differences in mass fraction between the paired samples and, at the same time, estimate the bias in the size distribution that can be reliably detected.

It is important to determine the amount of bias that can be detected at a given level of confidence when the covariance matrix for the difference in size analyses is known or estimated from the data pairs.

The solution to the problem will be approached by using both analytical results and simulation. The application of the results is intended to be relevant to the bias testing of a sampler which is likely to show an increased or decreased probability of collecting coarse particles. In such a case the proportion of fines in the samples from the test sampler, will correspondingly decrease or increase.

Methodology

To illustrate the concepts behind the comparison of the methods, a particular size distribution will be chosen and the tendency to collect more coarse particles will be modelled by choosing a size distribution that plots linearly on Rosin-Rammler coordinates and pivoting the cumulative passing curve at a particular point. It is convenient to choose the 40% passing point. One 'sampler' will sample from the 'true' size distribution and the second 'sampler' will sample from the distorted distribution.

The sampling and sizing of the samples from the two samplers will be simulated by allowing for a Poisson error in the numbers of particles in each size fraction and a random normally distributed error in the mass of material reporting to each screen. The magnitude of this, normally distributed error, can be changed to make the size analysis of the samples more or less accurate.

In the simulations presented below, the 'true' size distributions from which samples are realised will be taken to be constant as a simplifying assumption.

The amount of bias detectable will be sought by increasing the difference between the true size distributions until a situation arises in which one is 95% certain of detecting bias using a given number of paired samples of a given mass and the Hotelling test. The distribution of the number of failed *t*-tests under the same circumstances, will be determined.

Results

Figure A2-1 shows the statistical results for the base size distribution of A2-I.

The results were collected by simulation of 5 000 datasets and using 10 pairs of samples. There are 7 size fractions that can be *t*-tested and 6 size fractions used in the Hotelling test. The Hotelling test uses one less size fraction than is present as the last size fraction is linearly dependent on the other and provides no additional statistical information.

When there is no difference in size distribution between the two samples (slope factor = 1.00), the critical *F* value for the Hotelling test is chosen as the 95% point for 6 and 4 degrees of freedom corresponding to the test of 6 size fractions and using 10 pairs of samples. The covariance matrix in the test is estimated from the data. At this stage, the probablility that there will be no *t*-test failures out of the 7 is 0.826, but the probability that one *t*-test will fail is 0.168. If the criterion for the detection of bias is a failed *t*-test, then there is a probability of 0.168 (16.8%) of failing the sampler when it is 'perfect'.

Table A2-I

Base size distribution for analysis (slope factor 1.00)

Size fraction (mm)		Mass %
25	31.5	8.86
20	25	8.77
16	20	11.22
12.5	16	13.50
9.5	12.5	14.21
8	9.5	7.80
0.5	8	35.65

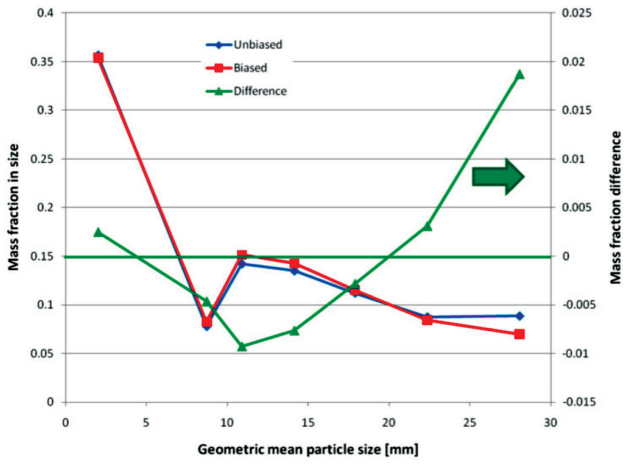

Figure A2-2—Biased and unbiased distributions and difference in distributions at the critical point for detection of bias by the Hotelling T squared test

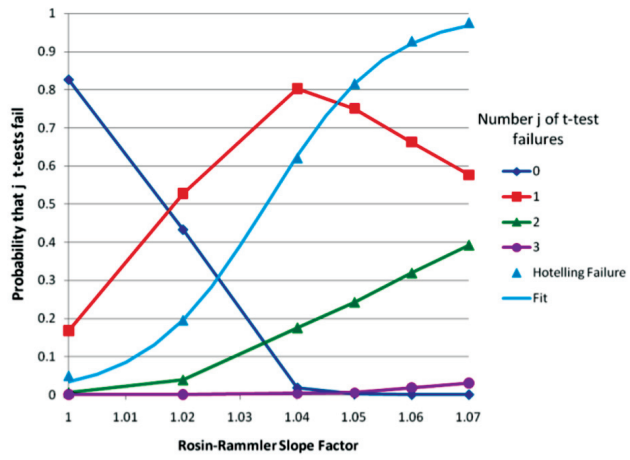

Figure A2-1—Analysis of single *t*-test failures and Hotelling T squared failure probability as a function of difference in test size distributions. The function fitted to the Hotelling probabilities is a logistic function

Table A2-II

Results of Monte Carlo simulation for statistical evaluation of the individual *t*-test method and the Hotelling T squared test method

Rosin-Rammler slope	Probability of Hotelling test failing	Probability that *j* *t*-tests fail			
		0	1	2	3
1.00	0.0494	0.8258	0.1678	0.0064	0
1.02	0.1950	0.4326	0.5278	0.0396	0
1.04	0.6210	0.0178	0.803	0.1762	0.003
1.05	0.8162	0.002	0.7504	0.243	0.0046
1.06	0.9280	0	0.6622	0.3196	0.0182
1.07	0.9766	0	0.5764	0.3926	0.031

Increasing the slope factor to 1.02 to provides some difference in size distribution from the two samplers raises the probability of a failing Hotelling test from 0.05 to 0.195 but raises the probability of a single *t*-test failure to 0.528 and the probability of 2 tests failing to 0.04.

Changing the slope factor to 1.04 raises the Hotelling failure probability to 0.621 but the probability of one *t*-test failure is now 0.803 and the probability of 2 failures is 0.176; the probability of no *t*-test failures is practically zero.

Using the Hotelling test with a critical value place at 95% confidence, the slope factor can increase to 1.065 before there is a probability of 0.95 of the test failing. This point can be regarded as the level of bias that will be detected with a probability of 0.95. The size distributions are shown in Figure A2-2.

It is clear from these results that the discovery of a single *t*-test failure is a poor indicator of size distribution bias. To require the failure of two *t*-tests before concluding the presence of bias is a too-harsh requirement.

When the Hotelling failure rate is 95%, the probability of two *t*-test failures is 0.4, meaning that the probability of making an incorrect decision based on the failure of two tests is 0.6 and this probability of incorrect decision is unacceptably high. The probability of incorrect decision for the Hotelling test is only 0.05.

It would seem that the Hotelling test is a much superior decision making tool when compared to the application of single *t*-tests. ◆