



Assessment of the precision and bias of an online gauge using a single reference instrument

by F. Lombard* and G.J. Lyman†

Synopsis

We consider the Grubbs estimation of the precision of an online gauge. Typically, this type of estimation involves independent results from two or more reference instruments (sampling and laboratory analysis operations). The properties of the estimator are then independent of the product variability. However, the use of more than one reference instrument entails significant additional costs.

The two-instrument Grubbs estimator, which is based on results from a gauge and a single reference instrument, has the disadvantage that its standard error is heavily dependent on the degree of product variability. We propose a new estimator that has a variance that is more or less independent of product variability. In fact, the variance is typically less than that of the Grubbs estimator based on the use of two reference instruments.

In order to function successfully, our methodology requires some prior knowledge of the extent of product variability and of gauge precision. In practice, such prior knowledge is often available and it is a weakness of the traditional two- and three-instrument Grubbs methods that no use is made of such knowledge. The efficacy and robustness of the new method is illustrated by Monte Carlo simulation.

Keywords

precision estimation, Grubbs estimator, Online analyser, ISO 15239, instrument variance.

Introduction

Consider measurement of a quality characteristic of a batch of product, e.g. coal, by each of two instruments. An application of particular interest occurs when one is considering the use of an online gauge as a substitute for a traditional but, in the long run, more costly and time-consuming sampling and laboratory analysis procedure. In order to judge the worth of a gauge, one would want to know its precision and also whether there is a bias between the results it reports and the corresponding results reported by a conventional sampling and analysis operation (the reference instrument). The standard method used to estimate gauge precision is the three-instrument Grubbs¹ method. This method involves a comparison between results produced by the gauge and corresponding results produced by two independent reference

instruments. The landmark paper by Rose² can fairly be said to have led to the establishment of the three-instrument Grubbs estimator as the estimator of choice in the coal industry for dealing with on-line analyser precision. Rose² gives examples of how the reference instruments may be arranged. Invariably, a second, independent sampling operation is required. Clearly, an additional sampling and analysis operation entails significantly increased costs for the duration of the trial. If stopped-belt sampling is involved, there will also be disruption of production.

Rose² discusses briefly the possibility of using a two-instrument Grubbs¹ estimation procedure, that is, where only one reference instrument, instead of two independent ones, is required. A problem with this method, also alluded to by Rose², is that the stability of the resulting estimate is adversely affected by large day-to-day variations in the coal quality. Such variations often leads to an estimate of gauge variance that is negative. The three-instrument Grubbs estimator is not troubled by the effect of day-to-day variation, as this is eliminated by working with the pairwise differences between results rather than with the raw results. An obvious way to increase the stability of the two-instrument Grubbs estimator is to increase the duration of the trial, i.e. the number of samples that are taken. This, in general, is not cost effective. Calculations based on realistic scenarios indicate that hundreds of samples might be needed in order to obtain an acceptably precise estimate of gauge measurement error. The question arises, therefore, whether there is perhaps another way in which the stability of the two-instrument Grubbs estimator might be improved.

* Department of Statistics, University of Johannesburg.

† Materials Sampling & Consulting, Pty. Ltd. Australia.

© The Southern African Institute of Mining and Metallurgy, 2011. SA ISSN 0038-223X/3.00 + 0.00. Paper received Dec. 2010; revised paper received Jun. 2011.

Assessment of the precision and bias of an online gauge using a single reference instrument

The purpose of the present paper is to show that such an improvement can indeed be effected under generally prevailing operating conditions. To understand why such improvement is possible, one should notice that in calculating Grubbs estimators no use whatsoever is made of any prior knowledge that may be available regarding the coal and gauge measurement variances. In other words, there is an implicit assumption that one knows nothing about the properties of the coal analysis technique and of the gauge measurements. In fact, statistical analysis of historical assay results can provide a good prior estimate of coal sampling and analysis variability and overall coal quality variability. Furthermore, the gauge manufacturer invariably supplies some guarantee regarding the precision of the gauge, which implies that there is reasonably accurate information available in that regard too. In any event, one can obtain a preliminary estimate of gauge variance from variographic analysis³ of the run of readings collected during the passage of a number of lots which are used to calibrate the gauge. Under normal circumstances, therefore, it will not be difficult to produce a reasonable prior estimate, w , of the ratio, w_0 , of true coal variance to true gauge measurement variance. We will show that the two-instrument Grubbs estimator of gauge variance can be improved substantially, at no extra cost, by the simple device of sorting the observed pairs of observations into a number of judiciously chosen subsets on the basis of the estimated ratio w . Two-instrument Grubbs estimators are then calculated within each of the subsets.

The effect of the sorting into subsets is that coal quality variation within any subset is smaller than the overall coal quality variation, hence the effect on the variance of the two-instrument Grubbs estimator is similarly smaller. When the estimators from each of the subsets are averaged, an estimator which has greatly reduced variance results. In fact, the new estimator typically performs substantially better than either of the Grubbs estimators. This finding has far-reaching implications regarding the costs involved in an operational comparison between an online gauge and a sampling and analysis procedure. A two-instrument Grubbs estimation using n laboratory results might be judged equivalent in cost to a three-instrument estimation involving $n/2$ results from each of two independent sampling and analysis operations. The actual cost of taking the additional set of $n/2$ samples will, however, be considerably greater because an automatic sampler, which is likely to be available for taking the first sample (instrument 2), is unlikely to be available for taking the sample corresponding to instrument 3. Manual acquisition of the instrument 3 samples will then be more costly than the instrument 2 samples, and may also be biased if cut manually from the stream of coal without using stopped-belt sampling.

The new method proposed here is not a panacea, however. It will be applicable only if reasonably accurate prior knowledge of the extent of coal quality variability and of gauge measurement variance is available. If the prior information regarding either or both of these uncertainties is far off the mark, the method could produce a grossly incorrect estimate of gauge error. However, as remarked above, it is seldom the case that reasonably accurate prior information is not available. Hence, it will generally be safe to implement the new method. The potential savings in respect of time and

additional sampling and analysis costs certainly make the method worth considering as an alternative to the traditional three-instrument Grubbs estimator.

Later we give an overview of the Grubbs estimation methodology and provide examples of the numerical calculations involved. We give a formal definition of the new estimator. We also show how the standard error of the new estimator is calculated. We use Monte Carlo simulation to compare the variance of the new estimator to the variances of the original two-instrument and three-instrument Grubbs estimators. The results suggest that the new estimator can perform substantially better than either of the Grubbs estimators in terms of statistical efficiency. We also give an example of the calculations required to determine ahead of time the number of samples required to estimate gauge precision with a pre-specified standard error. Finally, we discuss how the data might be used to detect biases in the gauge calibration.

The Grubbs estimators

Table I shows a small subset of a larger set of data collected in a gauge evaluation. The rows, which correspond to successive days, are determinations of the specific energy of coal sampled from a moving stream by three instruments: an online gauge, a mechanical sampling and laboratory analysis procedure, and an independent manual sampling and laboratory analysis procedure. (This data is used purely for illustrative purposes and is not intended to reflect any particular reality).

The three numbers (y x z) in any given row can be represented by the following three equations:

$$\begin{aligned} y &= t + g \\ x &= t + l \\ z &= t + m \end{aligned} \quad [1]$$

where t denotes the (unknown) 'true' specific energy value of the coal and g , l , and m denote the measurement errors associated with the three instruments. The measurement errors l and m associated with the two sampling and analysis procedures consist of the sum of the sampling, preparation, and analysis errors. For simplicity of presentation, we

Table I

Illustrative data

| Day | y | x | z |
|-------|--------------------|--------------------|--------------------|
| 1 | 24.23 | 23.57 | 25.17 |
| 2 | 25.14 | 25.06 | 24.54 |
| 3 | 24.74 | 24.40 | 24.54 |
| 4 | 23.12 | 22.54 | 23.46 |
| 5 | 25.95 | 25.47 | 23.50 |
| 6 | 25.57 | 25.66 | 25.24 |
| 7 | 25.46 | 25.21 | 24.78 |
| 8 | 23.21 | 23.38 | 24.81 |
| 9 | 24.56 | 24.44 | 25.82 |
| 10 | 25.99 | 26.36 | 24.32 |
| Total | 247.97 | 246.09 | 246.19 |
| Mean | $\bar{y} = 24.797$ | $\bar{x} = 24.609$ | $\bar{z} = 24.619$ |

Assessment of the precision and bias of an online gauge using a single reference instrument

assume for the moment that no biases are present between the three instruments. Bias is dealt with later. The extents of the measurement errors are quantified by their respective variances, denoted here by σ_g^2 , σ_l^2 , and σ_m^2 , the primary objective being to obtain a reliable estimate of σ_g^2 . Notice that none of the four variables of interest, namely g , l , m , or t are observable. Thus, the challenge is to estimate σ_g^2 using only the three observables y , x , and z . Towards this goal notice that

$$\begin{aligned} y - x &= g - l \\ y - z &= g - m \end{aligned} \quad [2]$$

which eliminates t from consideration, and that

$$(y - x) \times (y - z) = g^2 + \varepsilon \quad [3]$$

where

$$\varepsilon = g \times l + g \times m + l \times m \quad [4]$$

Under the very reasonable assumption that the measurement errors of the instruments are statistically uncorrelated it follows that the expected value $E(\varepsilon)$ is zero, whence

$$E[(y - x) \times (y - z)] = E[g^2] = \sigma_g^2 \quad [5]$$

Here and elsewhere in the paper, the expected value of a quantity denotes its average value over the ensemble consisting of all its possible realizations. Consequently, it is sensible to estimate σ_g^2 by the average of the n ($= 10$ in the particular instance of Table I) observed realizations of the quantity $(y-x) \times (y-z)$:

$$\hat{\sigma}_g^2 = \frac{1}{n} \sum_{i=1}^n (y_i - x_i) \times (y_i - z_i) \quad [6]$$

A small adjustment must be made to the last formula if the presence of constant offsets (biases) between the instruments is to be accommodated. Namely, each of y_i , x_i and z_i must be reduced by the corresponding mean over all days (e.g. y_i must be replaced by $\tilde{y}_i = y_i - \bar{y}$) and n must be replaced by $n - 1$, so that

$$\hat{\sigma}_g^2 = \frac{1}{n-1} \sum_{i=1}^n (\tilde{y}_i - \bar{\tilde{y}}) \times (\tilde{y}_i - \bar{\tilde{z}}) \quad [7]$$

Notice that the variances σ_l^2 and σ_m^2 of the reference instruments can be estimated in exactly the same way by

simply interchanging appropriately the roles of y_i , x_i , and z_i in the last formula. The standard error of $\hat{\sigma}_g^2$ is given by Grubbs¹:

$$se(\hat{\sigma}_g^2) = \sqrt{\frac{1}{n-1} \{ 2\hat{\sigma}_g^4 + \hat{\sigma}_g^2 (\hat{\sigma}_l^2 + \hat{\sigma}_m^2) + \hat{\sigma}_l^2 \hat{\sigma}_m^2 \}} \quad [8]$$

Table II sets out the calculation of $\hat{\sigma}_g^2$ for the data shown in Table I.

We find $\hat{\sigma}_g^2 = -0.217/9 = -0.024$, which is not a useful result and which in practice would be interpreted as saying that the gauge measures specific energy with no error. Using similar calculations, we find $\hat{\sigma}_l^2 = 0.138$ and $\hat{\sigma}_m^2 = 1.645$. One explanation for the negative gauge variance estimate is that the variances of the reference instruments are apparently one to two orders of magnitude larger than the gauge variance. If a reference instrument is of poor quality one can hardly expect to obtain a useful outcome when evaluating against it an instrument (such as a nuclear gauge) that has much better precision, unless the number of samples is increased from 10 to a substantially larger number. However, we will see later that the new estimator gives a sensible result from these 10 observations alone.

Rose² recommends that in general at least 60 samples of data should be gathered in order to obtain a useful estimate of σ_g^2 . If stopped-belt sampling is involved, this recommendation implies a costly interruption of the normal production process for an extended period of time. Accordingly, Rose² considers also Grubbs estimation involving only y and x observations (no data from stopped-belt sampling). That an estimate of σ_g^2 can also be made in this setup follows upon noticing that

$$y \times (y - x) = (t + g) \times (g - l) = g^2 + \varepsilon \quad [9]$$

where now

$$\varepsilon = t \times g - t \times l - g \times l \quad [10]$$

Assuming that the two measurement errors are statistically uncorrelated and also uncorrelated with the true value t , then $E(\varepsilon)$ is again zero and an argument analogous to that used in the three-instrument case shows that σ_g^2 can be estimated by

$$\hat{\sigma}_g^2 = \frac{1}{n-1} \sum_{i=1}^n \tilde{y}_i \times (\tilde{y}_i - \bar{\tilde{z}}) \quad [11]$$

| Day | y/0 | x/0 | z/0 | y/0-x/0 | y/0-z/0 | (y/0-x/0) x (y/0-z/0) |
|-------|--------|--------|--------|---------|---------|-----------------------|
| 1 | -0.567 | -1.039 | 0.551 | 0.47 | -1.118 | -0.528 |
| 2 | 0.343 | 0.451 | -0.079 | -0.108 | 0.422 | -0.046 |
| 3 | -0.057 | -0.209 | -0.079 | 0.152 | 0.022 | 0.003 |
| 4 | -1.677 | -2.069 | -1.158 | 0.392 | -0.519 | -0.203 |
| 5 | 1.153 | 0.861 | -1.116 | 0.292 | 2.269 | 0.662 |
| 6 | 0.773 | 1.051 | 0.618 | -0.278 | 0.155 | -0.043 |
| 7 | 0.663 | 0.601 | 0.159 | 0.062 | 0.504 | 0.031 |
| 8 | -1.587 | -1.229 | 0.194 | -0.358 | -1.781 | 0.638 |
| 9 | -0.237 | -0.169 | 1.205 | -0.068 | -1.442 | 0.098 |
| 10 | 1.193 | 1.751 | -0.296 | -0.558 | 1.488 | -0.830 |
| Total | 0 | 0 | 0 | 0 | 0 | -0.217 |

Assessment of the precision and bias of an online gauge using a single reference instrument

The standard error of $\hat{\sigma}_g^2$ in this context is given by Equation [8] with $\hat{\sigma}_g^2$ there replaced by

$$\hat{\sigma}_t^2 = \frac{1}{n-1} \sum_{i=1}^n \tilde{y}_i \times \tilde{x}_i \quad [12]$$

The calculation of $\hat{\sigma}_g^2$ using only the data in the y and x columns in Table I is set out in Table III.

We find $\hat{\sigma}_g^2 = -0.889/9 = -0.099$, which is again not a useful result. The estimated standard error, calculated using the prescription given Table III is 0.121 (it is a moot point whether it makes sense to quote a standard error in respect of a negative variance estimate). The result is indicative of the statistical price to be paid for the benefit of eliminating a third instrument—many more samples are typically required if a useful estimate of gauge variance is to be made. Here the primary reason for the negative variance estimate is the large coal quality variance σ_t^2 that the Grubbs method estimates as $\hat{\sigma}_t^2 = 1.168$ (see Equation [12]). This happens because in a two-instrument setup σ_t^2 functions essentially as the variance of a fictitious third instrument against which the gauge is being compared. The estimation method described in the next section is a modification of the preceding two-instrument (gauge and one reference instrument) Grubbs method that ameliorates the effect of large coal quality variances.

The new estimator

In the interest of clarity we define our notation anew. A typical value, y , reported by the gauge can be represented as

$$y = t + g \quad [13]$$

where t denotes the true value of the quality characteristic in question as seen by the gauge and g denotes the statistical error intrinsic to gauge measurements. The latter error is assumed to have a distribution with zero mean and variance σ_g^2 . Similarly, a typical value, x , produced by sampling and laboratory analysis can be represented as

$$x = \alpha + t + l \quad [14]$$

where $\alpha + t$ denotes the true value as seen by the sampling and laboratory analysis and where l , the statistical error due to sampling and laboratory analysis, has a distribution with zero mean and variance σ_l^2 . Equation [14] incorporates the possibility of a constant bias, α , between the results produced by the two instruments. Note that neither of these two equations tells us which of the instruments is, in fact, producing biased readings. The standard error of the new estimator will not be affected in any way by the presence of

Table III

Calculation of Grubbs estimate with one reference instrument

| Day | $y/0$ | $y/0 - x/0$ | $y/0 \times (y/0 - z/0)$ |
|-------|--------|-------------|--------------------------|
| 1 | -0.567 | 0.472 | -0.267 |
| 2 | 0.343 | -0.108 | -0.037 |
| 3 | -0.057 | 0.152 | -0.009 |
| 4 | -1.677 | 0.392 | -0.657 |
| 5 | 1.153 | 0.292 | 0.337 |
| 6 | 0.773 | -0.278 | -0.215 |
| 7 | 0.663 | 0.062 | 0.041 |
| 8 | -1.587 | -0.358 | 0.568 |
| 9 | -0.237 | -0.068 | 0.016 |
| 10 | 1.193 | -0.558 | -0.666 |
| Total | 0 | 0 | -0.889 |

such a bias. The data for analysis is a set of n (assumed to be an even number) pairs of observations $(y_1, x_1), \dots, (y_n, x_n)$ obtained from n batches of coal. The true values of the batches are not constant and are assumed to vary in a statistical manner around a mean μ , the average long-term analyte value as seen by the gauge, with a variance σ_t^2 which quantifies the batch-to-batch variation.

Description of the estimation method

A general description of the method is given first, followed by a numerical example. With each pair of observations (y, x) , associate a weighted average

$$d = (1 + w)x - wy \quad [15]$$

where the weight w is an *a priori* estimate of the numerical value of the ratio

$$w_0 = \sigma_t^2 / \sigma_g^2 \quad [16]$$

From (13) and (14) we see that the conditional expected value of d is

$$E(d | t) = (1 + w)\alpha + t \quad [17]$$

so that d serves as an indicator of the quality of the batch in question. We have n such d -values. Arrange these in increasing order of magnitude, $d_1 < \dots < d_n$ say, and form the $m = n/2$ subsets

$$B_1 = \{d_1, d_2\}, \dots, B_m = \{d_{n-1}, d_n\} \quad [18]$$

There are two observation pairs, denoted by (y_1, x_1) and (y_2, x_2) , associated with each subset. For each subset we now calculate the corresponding two-instrument Grubbs estimator using only these two pairs of observations, namely

$$e = \frac{1}{2} \{ (y_1 - y_2)^2 - (y_1 - y_2)(x_1 - x_2) \} \quad [19]$$

Doing this for each subset yields m estimates e_1, \dots, e_m . The new estimate of σ_g^2 is the average of these m estimates:

$$\bar{e} = \frac{e_1 + \dots + e_m}{m} \quad [20]$$

The sampling and analysis variance may be similarly estimated simply by interchanging the roles of y and x and replacing σ_g^2 by σ_l^2 in the preceding algorithm.

The motivation for sorting into subsets is that coal quality variation within any subset is typically substantially smaller than the overall quality variation, hence the effect on the variance of the two-instrument Grubbs estimator is similarly smaller. When the estimators from each of the subsets are averaged, an estimator which has greatly reduced variance results.

Numerical illustration

The calculation of the new estimator will now be illustrated using the small set of specific energy (MJ/kg) determinations shown in Table IV. Prior estimates of σ_t and σ_g are given as 1.15 and 0.23 respectively. (This data is used purely for illustrative purposes and is not intended to reflect any particular reality.) We used $w = (1.15/0.23)^2 = 25$ in the calculation so that $d = 26x - 25y$; see Equation [15].

Table V shows the five subsets of observation pairs and Table VI shows the calculation of the e_i value using Equation [19].

Assessment of the precision and bias of an online gauge using a single reference instrument

The mean of the five e_i values is $\bar{e} = 0.0740$, which is the estimate of the gauge variance σ_g^2 . In this particular instance, we saw earlier that the standard two-instrument Grubbs estimate is $\hat{\sigma}_g^2 = -0.099$, which is uninformative.

Standard error of the estimator

There is a simple formula for the standard error associated with the new estimator, namely

$$\text{std. err.}(\bar{e}) = \hat{\sigma}_e / \sqrt{m} \quad [21]$$

where

$$\hat{\sigma}_e = \sqrt{\frac{\sum_{i=1}^m (e_i - \bar{e})^2}{m-1}} \quad [22]$$

In the numerical example above, $\hat{\sigma}_e = 0.1082$ and the standard error associated with the estimate $\bar{e} = 0.0740$ is $0.1082/\sqrt{5} = 0.0484$. In contrast, the standard error associated with the standard two-instrument Grubbs estimator of gauge variance is 0.121. This is about two and a half times larger than the standard error of the new estimate.

We can also see from the fourth column in Table VII how the improvement by the new estimator comes about. The coal quality variance within a subset is estimated by $(y_1 - y_2)(x_1 - x_2)/2$. The estimates of quality variance within each of the five subsets are thus 0.571, 0.647, 0.347, 1.621, and 0.147, with an average of 0.667. This is a little more than half the overall estimated quality variance of 1.168 found earlier—see below Equation [12].

Equation [21] arises from the fact that the values e_1, K, e_m are, to good approximation, statistically uncorrelated and have a common, albeit unknown, variance σ_e^2 . The mathematical details of the argument leading to Equation [21] are available from the authors as a separate document. Suffice it to point out here that the structure of d in Equation [15] plays a crucial role in the analysis. The simulation results shown later can be used to verify the result empirically in three specific instances.

It will often be more convenient to estimate directly the standard deviation, σ_g , rather than the variance, of the gauge error. Then the estimate of σ_g is $\sqrt{\bar{e}}$ with standard error

$$\text{std. err.}(\sqrt{\bar{e}}) \approx \hat{\sigma}_e / \sqrt{2n\bar{e}} \quad [23]$$

Thus, in the numerical example above, the estimate of gauge error standard deviation is

$$\sqrt{\bar{e}} = \sqrt{0.0740} = 0.27$$

with an associated standard error of:

$$0.1082 / \sqrt{2 \times 10 \times 0.074} = 0.089$$

It is entirely possible, especially if σ_t is an order of magnitude or more larger than σ_g , that some values among the e_i will be negative. When this occurs we simply eliminate from consideration the negative e_i values. Thus, the new estimator is more properly defined as the average of the positive e_i . However, in calculating $\hat{\sigma}_e$ we use all the e_i , both positive and negative. This safeguards a user against gaining an over-optimistic impression of the precision of the estimator. Of course, if σ_t is excessively large, then any two-instrument method will fail because the effect of σ_t cannot be eliminated entirely unless a third instrument is involved. In the remainder of the paper we assume without further mention that this modified version of the estimator is the one under discussion. In particular, then m denotes the number of positive e_i values.

Efficiency of the estimator

The efficiency of the new estimator relative to the standard two- and three-instrument Grubbs estimators will now be illustrated by Monte Carlo experiments. The efficiency of the new estimator is defined as the ratio of the variance of (either of) the Grubbs estimators to that of the new estimator.

Table IV
Ten pairs of specific energy determinations

| Pair | y | x | d | Subset |
|------|-------|-------|------------------|--------|
| 1 | 24.23 | 23.57 | 6.91 = d_1 | B_1 |
| 2 | 25.14 | 25.06 | 23.20 = d_7 | B_4 |
| 3 | 24.74 | 24.40 | 16.08 = d_4 | B_2 |
| 4 | 23.12 | 22.54 | 8.13 = d_2 | B_1 |
| 5 | 25.95 | 25.47 | 13.44 = d_3 | B_2 |
| 6 | 25.57 | 25.66 | 27.85 = d_9 | B_5 |
| 7 | 25.46 | 25.21 | 18.77 = d_5 | B_3 |
| 8 | 23.21 | 23.38 | 27.66 = d_8 | B_4 |
| 9 | 24.56 | 24.44 | 21.49 = d_6 | B_1 |
| 10 | 25.99 | 26.36 | 35.77 = d_{10} | B_5 |

Table V
The five subsets for calculation of the estimate

| Subset | Pairs | y | x |
|--------|-------|------------------|------------------|
| B_1 | 1 | 24.23 (= y_1) | 23.57 (= x_1) |
| | 4 | 23.12 (= y_2) | 22.54 (= x_2) |
| B_2 | 5 | 25.95 | 25.47 |
| | 3 | 24.74 | 24.40 |
| B_3 | 7 | 25.46 | 25.21 |
| | 9 | 24.56 | 24.44 |
| B_4 | 2 | 25.14 | 25.06 |
| | 8 | 23.21 | 23.38 |
| B_5 | 6 | 25.57 | 25.66 |
| | 10 | 25.99 | 26.36 |

Table VI
Calculation of the estimate from Equation [19]

| Subset | $y_1 - y_2$ | $x_1 - x_2$ | $(y_1 - y_2)^2$ | $(y_1 - y_2)(x_1 - x_2)$ | e |
|--------|-------------|-------------|-----------------|--------------------------|-----------------|
| B_1 | 1.11 | 1.03 | 1.2321 | 1.1433 | $e_1 = 0.0444$ |
| B_2 | 1.21 | 1.07 | 1.4641 | 1.2947 | $e_2 = 0.0847$ |
| B_3 | 0.90 | 0.77 | 0.8100 | 0.6930 | $e_3 = 0.0585$ |
| B_4 | 1.93 | 1.68 | 3.7249 | 3.2424 | $e_4 = 0.2413$ |
| B_5 | -0.42 | -0.70 | 0.1764 | 0.2940 | $e_5 = -0.0588$ |

Assessment of the precision and bias of an online gauge using a single reference instrument

Monte Carlo simulations

The following parameter configuration is motivated by data obtained in the evaluation of a PGNAA online coal analyser: $\sigma_g = 0.1179$, $\sigma_l = 0.3162$, $\sigma_t = 1.2808$, and $n = 94$. Thus, the value of w_0 in Equation [16] is 117.85 ($= 1.2802^2/0.1179^2$) in this particular instance. We simulated the estimation procedure using seven different values of w in Equation [15], namely $w = 135.5, 129.6, 123.7, 117.9, 112.0, 106.1,$ and 100.2 .

The first three and last three of the latter values correspond to incorrect specification of the correct value w_0 by 5%, 10%, and 15% respectively. Using normally distributed random numbers, 1 000 samples were generated, each consisting of 94 pairs of observations following the given parameter configuration. The new estimate, \bar{e} , was calculated for each of the 1 000 samples using each of the seven w -values shown above. Table VII gives the average (third column) and standard error (fourth column) of the 1 000 $\sqrt{\bar{e}}$ -values in each instance. If the estimation procedure is any good, then these averages should be close to $\sigma_g = 0.1179$ and the standard errors should also be close to what is predicted by Equation [23]. The latter predicted values are shown in parenthesis in the last column. Also shown in Table VII are the theoretical means and standard deviations of the classical Grubbs estimators based on two and three instruments respectively. In the latter case the second and third instruments are assumed to have the same standard deviation, namely 0.3162, while the sample size is 47. With two independent sampling and analysis procedures a sample of size 47 involves 94 laboratory analyses, which makes such a setup comparable to the two-instrument setup in terms of the amount of available data.

Tables VIII and IX give the results when σ_t is increased to 3.1623 and decreased to 0.1179 respectively.

The preceding three tables show the excellent performance of the new estimator compared to the Grubbs estimators in a range of practically representative situations. The Equation [23] for the standard error of the new estimator seems also to produce standard error estimates that are close to the 'true' standard errors, i.e. those obtained from the simulated results. The following conclusions, which can be established by mathematical calculations, are also supported by the simulation results in the preceding three tables:

- If σ_t is not excessively large compared to σ_g , the new estimator has smaller standard error than the three instrument Grubbs estimator. At any given sample size

and value of σ_g^2 , this superiority diminishes and will be eventually lost when σ_t becomes sufficiently large

- If σ_t is large compared to σ_g (e.g. one order of magnitude, typically the situation in practice), the new estimator has smaller standard error than the two-instrument Grubbs estimator. At any given sample size, this superiority diminishes and will be eventually lost when σ_t becomes sufficiently small (a circumstance that rarely prevails in practice)
- Prior mis-specification of the ratio w_0 (weight) by as much as 15% does not lead to a bias of major significance in the estimation of σ_g .

Sample size determination

An important question in any gauge evaluation is how many batches to interrogate in order to reach a more or less definitive conclusion. In order to give practical content to the term definitive, notice that in checking whether or not a performance guarantee such as that the gauge error has standard deviation less than σ_0 is met, one does not simply compare the estimate $\sqrt{\bar{e}}$ of σ_g to the guarantee value σ_0 and reject the guarantee if $\sqrt{\bar{e}}$ exceeds σ_0 . Since the gauge precision is estimated from a finite amount of data, some margin of error must be allowed regarding a final pronouncement. Typically one would place an upper bound, U , on the observed value of $\sqrt{\bar{e}}$, the estimator of σ_g , and require that the latter should not exceed U . In statistical terms, U is the upper limit of a, say 95%, one-sided confidence interval, $(0, U)$, for σ_g : we wish to be 95%

Table VII

Comparison of means and standard errors of the new estimator with those of two Grubbs estimators ($\sigma_g = 0.1179$, $\sigma_t = 1.2802$, $\sigma_l = 0.3162$)

| Method | Weight w | Mean | Std. err. |
|---------------|------------------------|-------|---------------|
| | 135.5 | 0.109 | 0.018 (0.016) |
| | 129.6 | 0.112 | 0.017 (0.016) |
| | 123.7 | 0.114 | 0.017 (0.016) |
| New | 117.9 | 0.117 | 0.018 (0.016) |
| | 112.0 | 0.120 | 0.018 (0.016) |
| | 106.1 | 0.123 | 0.017 (0.016) |
| | 100.2 | 0.127 | 0.018 (0.016) |
| Grubbs | (2 instruments) | 0.118 | 0.191 |
| Grubbs | (3 instruments) | 0.118 | 0.072 |

Table VIII

Comparison of means and standard errors of the new estimator with those of two Grubbs estimators ($\sigma_g = 0.1179$, $\sigma_t = 3.1623$, $\sigma_l = 0.3162$)

| Method | Weight w | Mean | Std. err. |
|---------------|------------------------|-------|---------------|
| | 82.73 | 0.102 | 0.040 (0.034) |
| | 791.4 | 0.107 | 0.038 (0.035) |
| | 755.4 | 0.111 | 0.037 (0.033) |
| New | 719.4 | 0.113 | 0.039 (0.035) |
| | 683.4 | 0.115 | 0.037 (0.033) |
| | 647.5 | 0.121 | 0.034 (0.031) |
| | 611.5 | 0.123 | 0.033 (0.030) |
| Grubbs | (2 instruments) | 0.118 | 0.026 |
| Grubbs | (3 instruments) | 0.118 | 0.072 |

Table IX

Comparison of means and standard errors of the new estimator with those of two Grubbs estimators ($\sigma_g = \sigma_t = 0.1179$, $\sigma_l = 0.3162$)

| Method | Weight w | Mean | Std. err. |
|---------------|------------------------|-------|---------------|
| | 1.15 | 0.112 | 0.012 (0.011) |
| | 1.10 | 0.114 | 0.012 (0.012) |
| | 1.05 | 0.115 | 0.013 (0.016) |
| New | 1.00 | 0.117 | 0.013 (0.012) |
| | 0.95 | 0.119 | 0.013 (0.012) |
| | 0.90 | 0.120 | 0.013 (0.012) |
| | 0.85 | 0.122 | 0.013 (0.012) |
| Grubbs | (2 instruments) | 0.118 | 0.026 |
| Grubbs | (3 instruments) | 0.118 | 0.072 |

Assessment of the precision and bias of an online gauge using a single reference instrument

confident that σ_g does not exceed the value U . Consider, for instance, the measurement of specific energy which varies on a batch-to-batch basis with a standard deviation $\sigma_t = 1.5 \text{ MJ/kg}$ and a sampling and analysis standard deviation known from past experience to be $\sigma_e \approx 0.6 \text{ MJ/kg}$. The vendor's guarantee is that the standard deviation of the gauge error does not exceed $\sigma_t = 0.2 \text{ MJ/kg}$. How many batches (n) are required to come to an equitable decision if one is willing to accept the guarantee only if $\sqrt{\bar{e}} \leq 0.4$?

From Equation [23] a 95% one-sided confidence interval for σ_g has upper bound

$$U = \sqrt{\bar{e}} + 1.645 \times \sigma_e / \sqrt{2n\bar{e}} \quad [24]$$

Setting $\sqrt{\bar{e}} = \sigma_0 = 0.2$, $U = 0.4$, and $\sigma_e = \sigma_0^2 = 0.04$ and solving for n gives $n = 34$.

It is illuminating to see what results if one contemplates using the three-instrument Grubbs estimator in this case. If each reference instrument interrogates k batches of coal, the total number of assays involved is $n = 2k$. Assume for simplicity that the two reference instruments have the same measurement precision, namely $\sigma_l \approx 0.6 \text{ MJ/kg}$. The standard error of the Grubbs three-instrument estimator of σ_g is then

$$\sqrt{(2\sigma_g^4 + 2\sigma_g^2\sigma_l^2 + \sigma_l^4) / k} / 2\sigma_g = \sqrt{1.01 / k}. \quad [25]$$

Therefore, in this case,

$$U = 0.2 + 1.645 \times \sqrt{1.01 / k} \quad [26]$$

Setting $U = 0.4$ and solving gives $k = 69$, that is, $n = 2k = 138$, about double the number required by the new estimator. This example serves again to illustrate the potential savings to be had from implementing the new method when reasonably accurate prior information on gauge and quality variances are available.

Detection of bias

The method of estimating precision described in the preceding section allows for, and automatically takes account of, a constant bias such as that represented by the term α in Equation [14]. A more pernicious type of bias, namely a bias of scale, must also be considered in the context of a two-instrument setup. In an evaluation of an online gauge one would, presumably, test for bias before proceeding to an estimation of the instrument precision—there is little interest in an instrument which produces an incorrect value with high precision. To incorporate scale bias into the analysis we replace Equation [14] by

$$x = \alpha + \beta t + l \quad [27]$$

Again, Equation [27] does not imply that the sampling and analysis operation is responsible for the scale bias—bias is defined relative to what the analyser is reporting as the 'true' coal quality. Only a separate bias test on the sampling system could establish which of the two instruments is responsible for any bias that may be detected.

A glance at Equation [27] suggests estimating β by least squares from a regression of x (sampling and analysis result) on t (true coal value). However, since t is unknown and since σ_g^2 is typically quite small, it is natural to consider using y in Equation [13] as a surrogate for it. The least squares estimator of β in Equation [27], obtained by regressing x on y , is

$$\hat{\beta} = \frac{\sum(y_j - \bar{y})(x_j - \bar{x})}{\sum(y_j - \bar{y})^2} \quad [28]$$

Now $\hat{\beta}$ is in fact not an estimator of β at all. Instead, it estimates

$$\beta \sigma_l^2 / (\sigma_l^2 + \sigma_g^2) = \beta w_0 / (1 + w_0) \quad [29]$$

rather than β (recall the definition of w_0 from [16]. The situation is ameliorated by using instead the adjusted estimator

$$\hat{\beta}_w = \hat{\beta} \times (1 + w) / w \quad [30]$$

where w is the prior estimate of w_0 . Then $\hat{\beta}_w$ estimates $\phi\beta$, where

$$\phi = \frac{1 + w}{1 + w_0} \frac{w}{w_0} \quad [31]$$

rather than β . In most cases of practical import, that is when w_0 is a relatively large number, the factor ϕ is numerically quite close to 1 even if w over- or underestimates the true value w_0 by as much as 15%. Suppose, for instance, $w_0 = 10$. Using a 15% overestimate of w_0 , namely $w = 11.5$, gives $\phi = 0.988$, while using a 15% underestimate, $w = 8.5$, gives $\phi = 1.016$. Thus in these two circumstances $\hat{\beta}_w$ estimates $0.998 \times \beta$ and $1.016 \times \beta$ respectively (while the unadjusted least squares estimator $\hat{\beta}$ estimates $0.91 \times \beta$). In general, therefore, it is not a serious mis-statement to say that $\hat{\beta}_w$ in Equation [30] estimates β . Setting $X = x \times (1 + w) / w$ one sees that the adjusted estimator $\hat{\beta}_w$ is the least squares estimate of slope in a regression of X on y . In most circumstances of practical relevance, therefore, one can check for scale bias by applying ordinary least squares methods to estimate the regression of X on y . If scale bias is present, σ_g^2 cannot be estimated directly by any of the methods discussed in this paper.

Summary and conclusions

A new statistical method for estimating the precision of an online gauge that requires only one set of comparative laboratory analysis results has been developed. The method is substantially cheaper to implement than the often used three-instrument Grubbs estimation method, which requires two independent sets of comparative sampling and laboratory analysis results. The method is also substantially more efficient than the two-instrument Grubbs method. The efficacy of the new method has been illustrated via Monte Carlo simulation. Examples have been given that demonstrate the calculation of the new estimate and its standard error. Calculations required to determine ahead of time the number of samples required to estimate gauge precision with a pre-specified standard error have also been shown. Finally, it has been shown how the data might be used to check the gauge calibration for biases. The method is applicable only if reasonably accurate prior information is available in respect of the coal and gauge variances.

References

1. GRUBBS, F.E. On estimating precision of measuring instruments and product variability. *Journal of the American Statistical Association*, vol. 43, 1948. pp. 243-264.
2. ROSE, C.D. Methods for assessing the accuracy of on-line coal analyzers. *Journal of Coal Quality*, vol. 10, 1991. pp. 19-28.
3. ISO 15239:2005 - Evaluation of the Measurement Performance of On-line Analysers. Geneva.