doi: 10.1255/tosf.45

Determination of the precision of sampling systems and on-line analysers

Geoffrey J Lyman[a] and James Asbury[b]
[a]Materials Sampling & Consulting, 1902 / 4 Como Crescent, Southport, Queensland, Australia 4215. E-mail: glyman@iprimus.com.au
[b]Realtime Group Limited, Lot J Mackay Marina Village, Mulherin Drive, Mackay QLD 4740, PO Box 9117, Slade Point QLD 4740.
E-mail: james.asbury@rtiaustralia.com

There is a simple and relatively inexpensive way of determining the precision of sampling systems and on-line analysers when a data base of output values from the sampling system or on-line analyser can be accessed and there exists serial correlation in the data sets. For a sampling system, if it is possible to construct a variogram from the routine data collected, it is possible to extract the component of the precision estimate due to material intrinsic heterogeneity, preparation and analysis as this variance is simply given by the intercept (nugget variance) of the variogram. To determine the last component of uncertainty, a punctual variogram determined from a sampling campaign is necessary. The method is much superior to interleaved sampling, which gives incorrect estimates of the precision when serial correlation exists. It is rare to find that there is no serial correlation in plant data. For on-line analysers that interrogate a process stream continuously, the variogram constructed from the gauge output for short time intervals can be used to determine the precision with no additional effort. The gauge ideally should be operated in such a way that the output is not smoothed by some statistical procedure. This paper outlines the methods and illustrates the procedure with data sets from a coal washery.

Introduction

It is very useful to be able to determine the precision with which a sampling system operates. The ISO Standards say that this precision can be found by a process of interleaved sampling, but this statement is incorrect when the assays in the process stream from which the samples are taken show a serial correlation in time (Lyman[1]). Interleaved sampling also demands that sampling be carried out at double the rate of the routine sampling. Building this capability into a sampling system increases the system cost.

What is desired is a simple and cost-effective means of estimating sampling system precision. This can be done by taking advantage of the serial correlation in time that is present in virtually all process streams.

Similarly, it is of great importance to be able to estimate the precision of an output value from an on-line analyser which is interrogating a process stream continuously. A variogram constructed from unfiltered output from the gauge will provide the precision estimate.

This paper provides the mathematical background behind the methods of precision determination and illustrates the method using data from a coal washery.

Mathematical development

When a process stream is observed by intermittently taking increments of material from the process stream and analysing them, the assay of the increment can be modelled as the sum of a random function and a random variable. The random function describes the true value of the assays as a function of time and the random variable describes the uncertainty introduced in the determination of the assay as a result of the intrinsic heterogeneity of the increment and the sample preparation and analysis uncertainties. The relationship can be described as

$$Y(t) = X(t) + \varepsilon \tag{1}$$

where $X(t)$ is the random function describing the true value of the process stream assay at time t and ε is a uncorrelated random variable having a distribution corresponding to that of the intrinsic heterogeneity of the increment plus the distributions due to the sample preparation and analysis. The random variable is statistically independent from the random function.

The random function can be characterised by a covariance function or variogram. Consider increments taken at a set of times $\{t_i\}$ giving rise to a set of measurements $\{y(t_i)\}$. The covariance function of these measurements is then

$$\begin{aligned}\operatorname{cov}\left\{Y(t_i), Y(t_j)\right\} &= \mathrm{E}\left\{\left[X(t_i) - X_0 + \varepsilon_i - \varepsilon_0\right]\left[X(t_j) - X_0 + \varepsilon_j - \varepsilon_0\right]\right\} \\ &= \mathrm{E}\left\{\left[X(t_i) - X_0\right]\left[X(t_j) - X_0\right]\right\} + \mathrm{E}\left\{\left[\varepsilon_i - \varepsilon_0\right]\left[\varepsilon_j - \varepsilon_0\right]\right\}\end{aligned} \tag{2}$$

where X_0 and ε_0 are the expected values of the random function and random variable. Note that the cross-terms between the random function and the random variable vanish due to the independence of the two statistical quantities. Taking the expectations above, those involving the random variable are zero except when $i = j$, that is the covariance is the variance of the random function plus that of the random variable. We have

$$\operatorname{cov}\left\{Y(t_i), Y(t_i)\right\} = \operatorname{var}\left\{Y(t)\right\} = \operatorname{var}\left\{X(t)\right\} + \operatorname{var}\left\{\varepsilon\right\} \tag{3}$$

If the random function is stationary, the value above is the value of the covariance function for Y at the origin, which can be denoted as $C(0)$. The variogram or covariance function estimation will provide a picture of the rest of the function, $C(t)$ which in fact now depends only on the properties of the random function $X(t)$. The covariance function will have the form as shown in Figure 1. The corresponding variogram function is shown at the right of Figure 1.

The relationship between the (semi)variogram and the covariance function is

$$\gamma(t) = C(0) - C(t) \tag{4}$$

so the variogram starts at zero and rises to a sill value equal to the value of the covariance function at the origin.

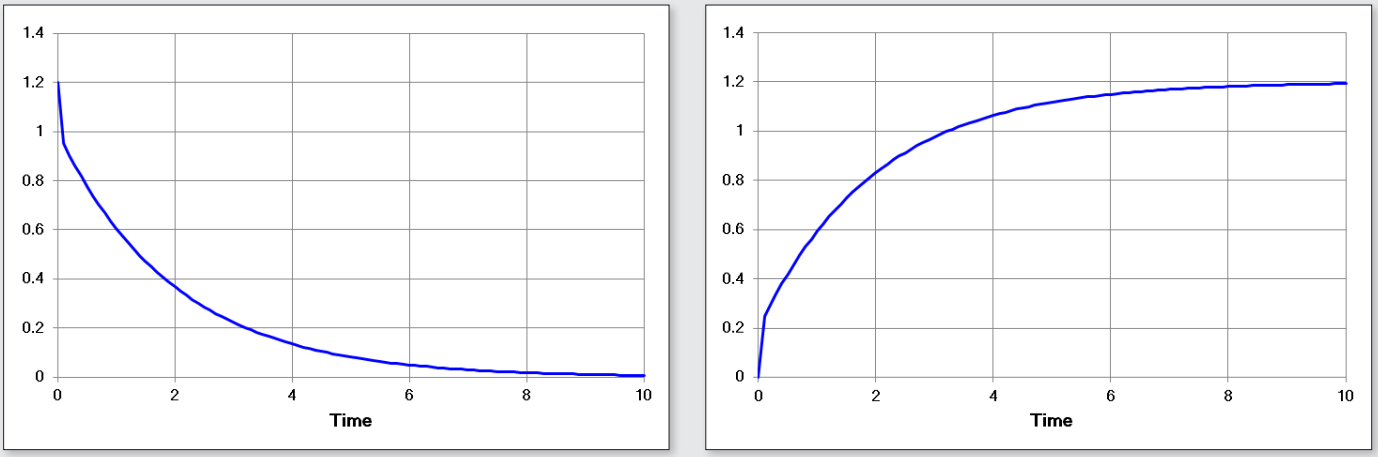

Figure 1. Covariance (left) and variogram (right) functions in the presence of measurement error.

In the operation of an on-line gauge, which may be of the nuclear type (prompt gamma neutron activation), gamma ray (transmission gauges) or x-ray fluorescence, the gauge interrogates the process stream and periodically provides an output which is an estimate of the composition of the stream with respect to one or more analytes. The output of the gauge can be modelled statistically in the same way as above, where $X(t)$ is the true analyte content averaged over some relatively short time period τ and ε is a random measurement error which is uncorrelated from one output value to the next. Some caution must be exercised here as the output from on-line gauges can involve the application of an exponentially weighted moving average process to the raw signals, or some other methodology that smoothes the output values. The use of such smoothing methods will cause serial correlation of the measurement error component of the output and well as modifying the covariance function of the component due to the changing analyte content of the stream.

The last circumstance to be considered is that in which increments are collected from a process stream over a period of time (a shift or day) and then analysed together as a whole. When values from this data stream are analysed variographically, the variogram observed is not the punctual variogram but a punctual variogram which has been regularised (a change of support having been made from single points to a set of points) over the period of sampling. This has important implications for the determination of the total sampling uncertainty as the punctual variogram is obscured.

Application to sampling of process streams

When a sampling system for a process stream is designed and the sampling is mechanically correct, there are three components of uncertainty that must be considered:

- the component due to the fact that there is a difference between the true average analysis of the increments extracted and the true average analysis of the process stream over the entire sampling period
- the component due to the intrinsic heterogeneity of the increments collected and that introduced within the sample preparation protocol
- the final analytical uncertainty

The first component is due to the distributional heterogeneity of the process stream, the second due to intrinsic heterogeneity of the material as sampled and at various stages in the sample preparation protocol and the last due to random error in the analysis procedure be it classical or instrumental.

The first component of uncertainty is determined by the shape and range of the variogram, that is by the time-wise serial correlation of the target analyte content of the stream. The second and third components are uncorrelated with the time variation and together are a measurement uncertainty. With a punctual variogram determined from the analysis of individual increments, the variogram can be extrapolated back to zero to make an estimate of the size of the jump after the origin, as in Figure 2.

The magnitude of the jump is equal to the measurement variance. This determines the sum of the last two components of uncertainty in sampling. The first component can be calculated from the shape of the variogram, providing an estimate of the total sampling uncertainty. There is a potential issue, however, with this procedure, namely that the sample preparation protocol for the individual increments may differ in a significant manner from that for the usual shift or daily sample. While the analytical variance will be the same as for the shift or daily sample (unless multiple assays are routinely carried out and only single assays applied to the individual increment), the second variance component due to the intrinsic heterogeneity of the material sampled may not match that involved in the preparation of the daily sample due to differences in the protocol.

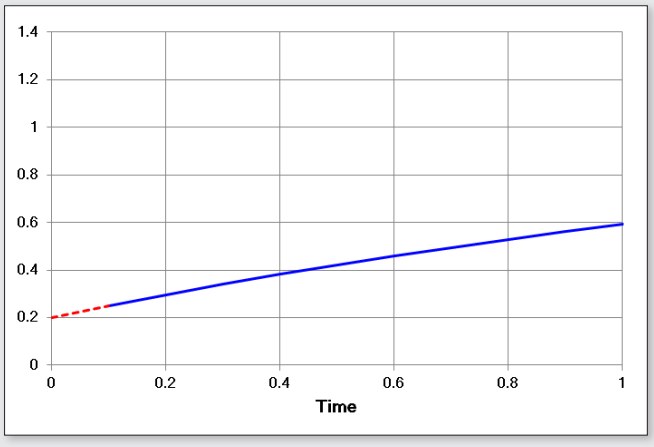

Figure 2. Backward extrapolation of a variogram to estimate the measurement variance.

When dealing with relatively small data sets, as is common when a special sampling program has been carried out, estimation of the value of the variogram and intercept can be made by maximum likelihood methods (Lyman[2]) which are very effective especially when the increments have not been extracted on a strictly constant time base. In such a case, the variance of the estimate of the measurements variance can be calculated as well.

When dealing with a variogram estimated from shift or daily samples, it is still possible to find a variogram and fit or extrapolate to find a measurement variance. However, this variogram cannot be used to find the sampling variance due to distributional heterogeneity as the compositing of the increments taken into a single sample has obscured the original punctual variogram. In particular, the sill of the variogram found will be lower than that for the punctual variogram and the range of the variogram will be longer as a result of the averaging process. It is not possible to work backwards to find a unique variogram that, when regularised using the actual sampling pattern, will match the observed variogram. There are many possible punctual variograms that will match the observed variogram after regularisation. But the intercept of this variogram is equal to the variance due to sample preparation and analysis for the protocol used routinely. This variance can be combined with the variance due to distributional heterogeneity determined from the punctual variogram to arrive at the correct estimate of the sampling system precision.

Therefore the analysis of the data set for shift or daily samples can be combined with the punctual variogram to provide the correct answer for the total sampling variance.

Note that it is an estimation of precision that is made, not an estimation of accuracy; bias cannot be detected in this way.

Application to on-line analysers

If the output from the on-line analyser has not been interfered with by averaging methods, the precision of the analyser on a punctual basis can be estimated. Note again that it is precision that is being estimated, not accuracy.

The current practice in the estimation of the precision of on-line analysers usually rests with the use of the Grubbs estimator (Lyman *et al.*[3]), which requires the use of two reference measurements in addition to the data from the gauge. It is necessary to coordinate the recording of signals from the gauge and the collection of physical samples of the material analysed in two independent ways in order to put this method into place. It is also desirable to ensure that the precision of the two reference measurements are better than that of the gauge; this can be difficult, given sampling problems.

By contrast the variogram approach for estimating analyser precision requires no additional effort. The estimate is derived directly from the gauge output. It is therefore very inexpensive and effective. On-line analysers produce a large volume of data as they generally produce an output value at any desired interval. A largest source of measurement variance may be the counting statistics for nucleonic systems, which ensures that the component of measurement error is independent from one reading to the next.

As for sampling, the estimate of measurement variance involves only the estimation of a variogram with backward extrapolation to the origin to find the intercept. With the large data sets from on-line analysers, the maximum likelihood method of variogram estimation is not practical.

Example

This example is drawn from data collected both from an on-line analyser and a conventional sampling system producing assays about every 6 hours. The operation of the conventional sampling system is somewhat erratic. The washery in question treats a number of types of coal with widely varying ash content. The on-line analyser interrogates all these feed coals on the same belt.

The conventional sampling system data was analysed on a per coal type basis in order to pick up the serial correlation for those coal streams. Figure 3 shows the data for coal type A as a function of tonnes of coal sampled.

The upper trace in Figure 3 is the actual coal ash content as sampled and the solid dark line is the trend line through the data determined by locally weighted regression. The lower trace is the deviation from this trend line. De-trending of the data is mandatory before calculating a variogram as this method can be applied only to stationary data. It is also desirable to apply the method to data that follows a Gaussian distribution as all theory and tools attached to variogram estimation assumes normality of the data. Figure 4 shows the deviation data after having been transformed to z-scores (standard Gaussian deviates of zero mean and unit variance). The

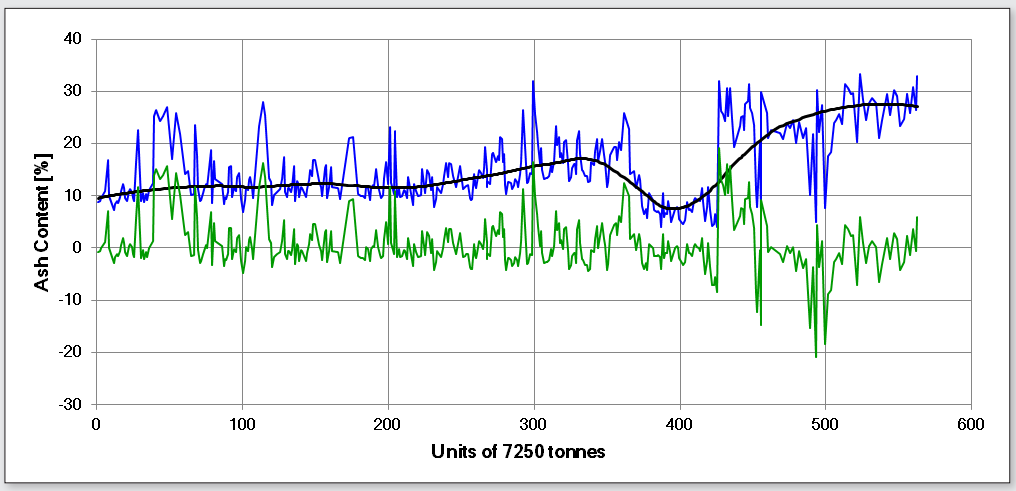

Figure 3. Ash content of coal type A as a function of tonnes of coal sampled. Lower curve shows deviations from trend (black).

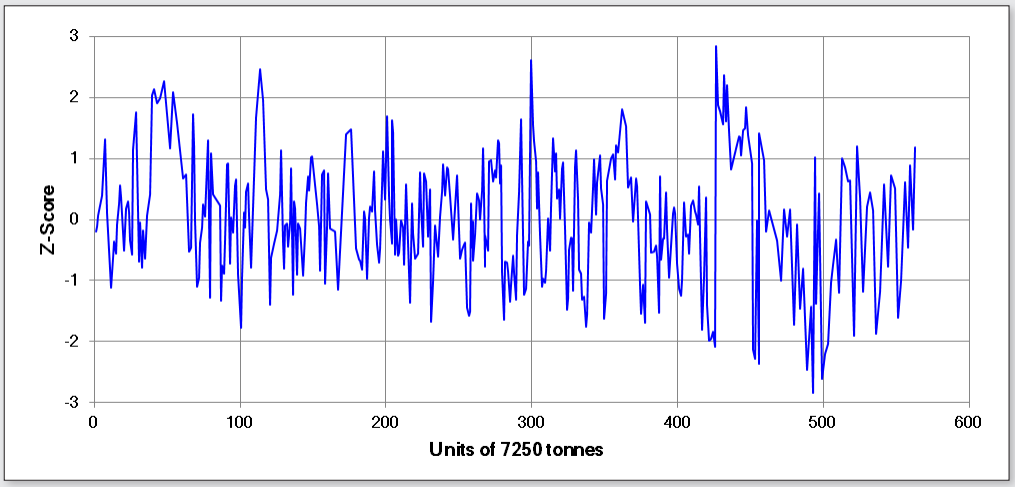

Figure 4. Z-scores for the deviation data of Figure 3.

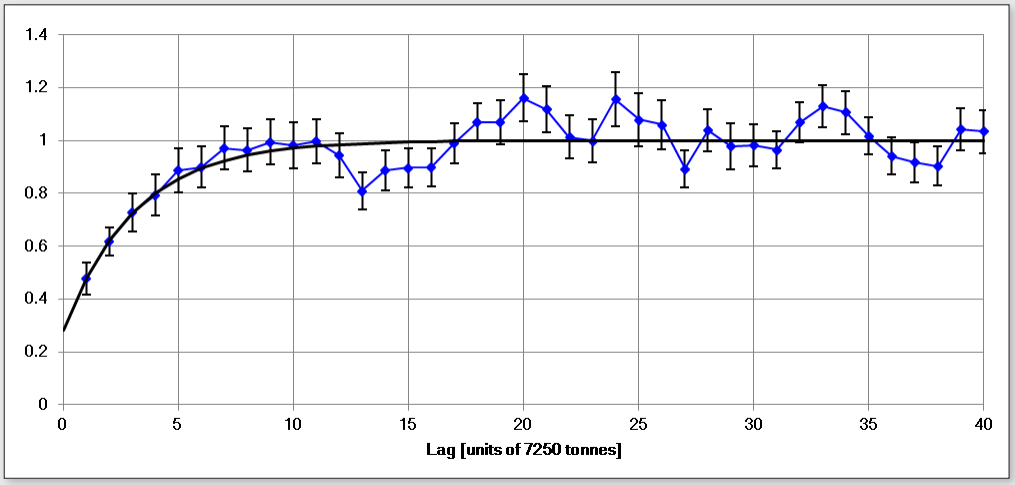

Figure 5. Variogram for coal A as sampled using z-score values, with a fitted exponential variogram. The intercept value is 0.284 from the fit.

variogram is calculated using this transformed data using the conventional Matheron estimator. Note that the variogram is based on actual assay results rather than a measure of heterogeneity as is often done when following Gy's methodology. The interest here is in the assay uncertainty and not a measure of heterogeneity. The error bars on the variogram represent a ±1 SD interval for the variogram estimate at the given lag. The variogram derived from the data of Figure 4 is shown in Figure 5.

The intercept of the z-score variogram can be found either using the first two points on the variogram or by fitting an admissible variogram function to the data. The intercept value is rescaled using the variance of the untransformed deviation data about their mean. Dividing the square root of the scaled intercept value by the mean ash content of the un-detrended data then provides a relative standard deviation for the measurement uncertainty. In this case the RSD is 17.2% ash. This indicates that there are serious problems with the sampling system or the manner in which the sample is prepared and analysed.

The corresponding data for coals B and C are shown in Figures 6 and 7. The RSD for coal B is 14.3% and for coal C 16.0%. The consistency of the estimates of the RSD underlines their validity, given that they are passing through the same sampling system. The ranges of the variograms are similar for coals A and B; that for coal C is longer. However, the last data set is relatively small and the variogram less well-defined.

The analysis of the gauge precision is based on one month of outputs at two minute intervals. The gauge is a prompt gamma neutron activation type (Realtime Group Allscan gauge).

The data for low ash coal is shown in Figure 8; there are just over 7000 data points in the data set. The z-score variogram is shown in Figure 9. The right hand frame shows a closer view of the behaviour of the variogram near the origin. The SD of a two minute reading is 2.12% ash or 23.7% relative.

The corresponding data for high ash coal is shown in Figure 10 with the z-score variogram in Figure 11. The SD of a two-minute reading is 3.04% ash or 12.0% relative.

It is interesting to consider the gauge measurement uncertainty over a period longer than 2 minutes. Because the gauge is measuring continuously, there is no uncertainty due to distributional heterogeneity such as would arise if punctual increments were

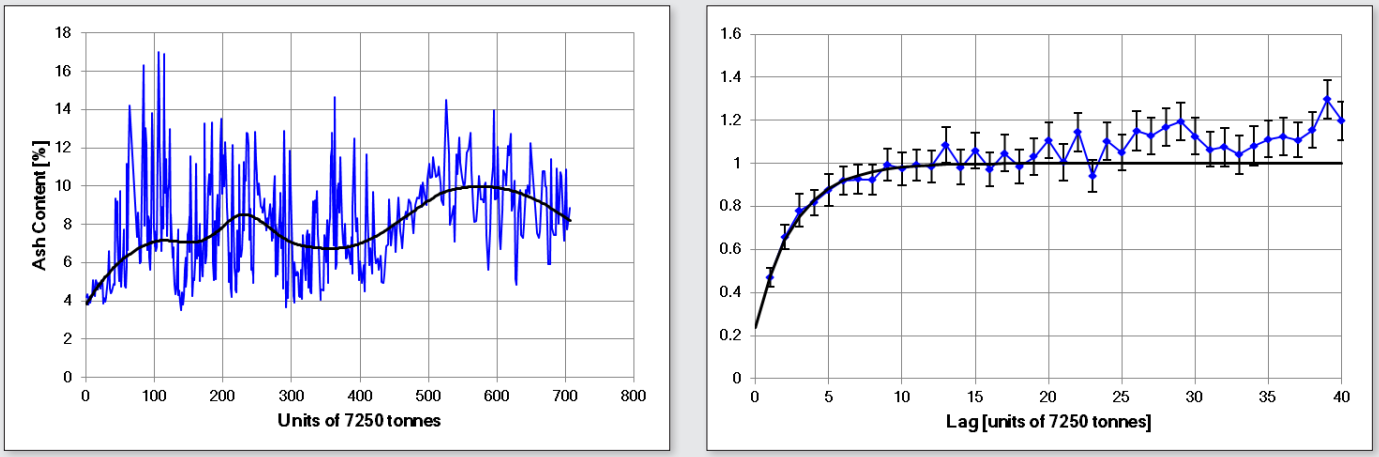

Figure 6. Data, detrending and z-score variogram for coal B.

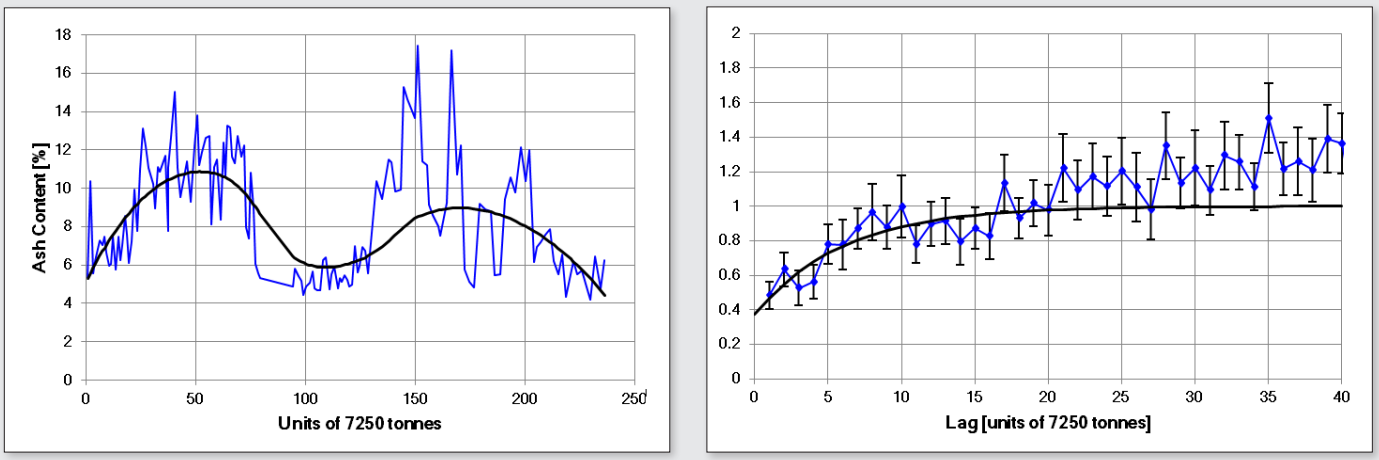

Figure 7. Data, detrending and z-score variogram for coal C.

being taken as in conventional sampling (the gauge misses nothing). Consequently, the measurement variance is simply inversely proportional to the number of two minute readings that are averaged. For a 6 hour period, there are 180 readings so the RSDs are reduced to 1.77% (SD = 0.158% ash) for the low ash coal and 0.894% (SD = 0.227% ash) for the high ash coal. As long as there is no bias in the gauging system, the gauge accuracy over a 6 hour period is extremely good. Over a daily period, the figures above

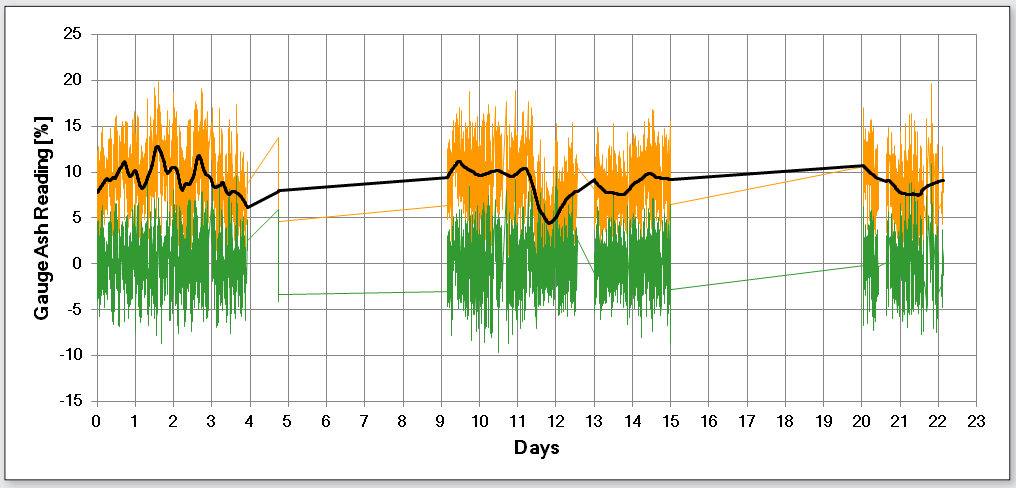

Figure 8. On-line analyser data for low ash coal (yellow) showing de-trending (black) and deviation values from the trend (green).

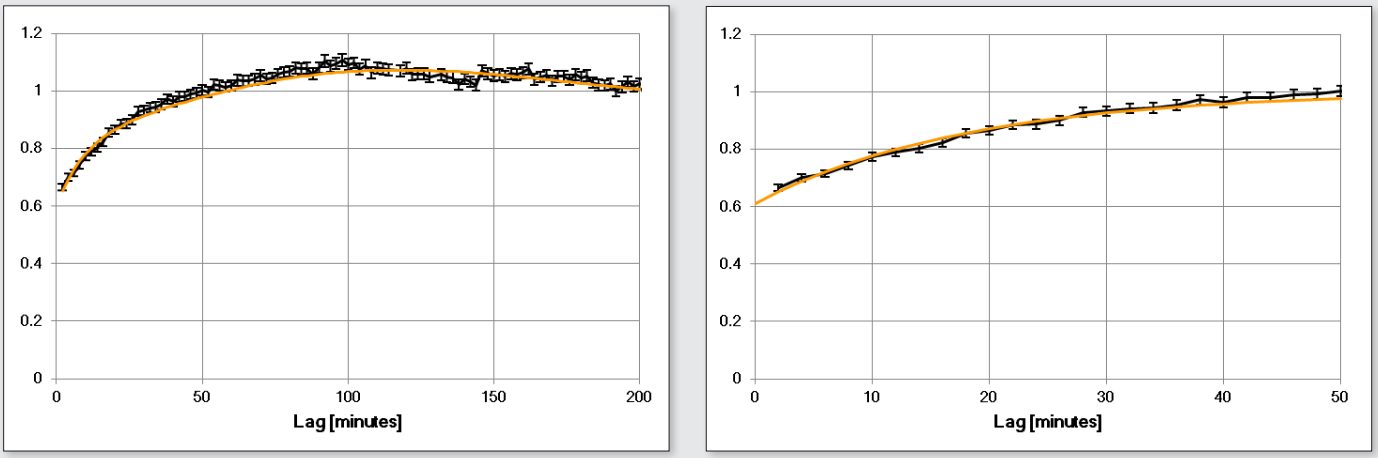

Figure 9. Variogram for low ash coal derived from on-line analyser z-score data. The right hand frame shows a closer view of the behaviour of the variogram near the origin.

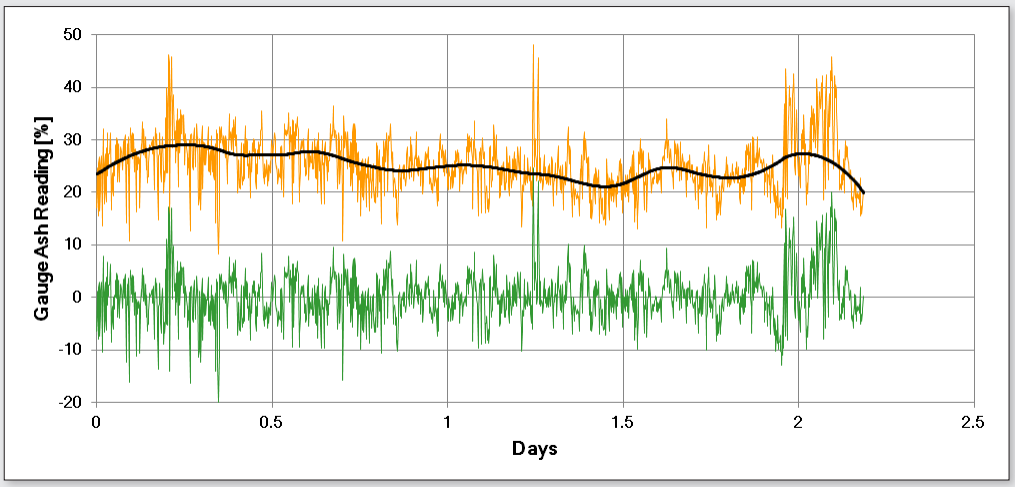

Figure 10. On-line analyser data for high ash coal showing de-trending and deviation values from the trend.

are reduced by half to deliver standard deviations of 0.076% ash and 0.114% ash respectively. These figures can be compared to the standard deviations of a single ash determination by Australian Standard 1038 of 0.05% ash and 0.085% ash.

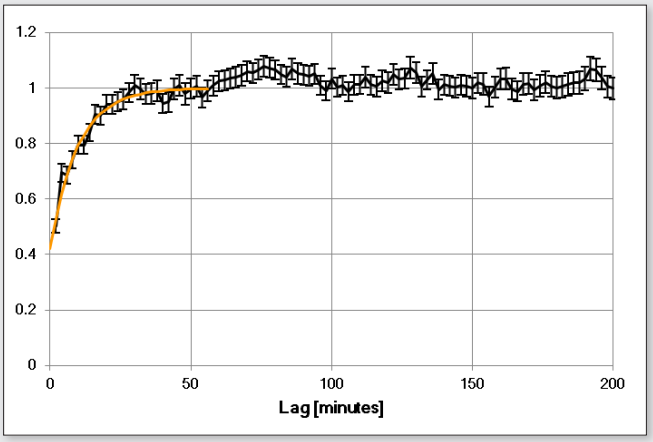

Figure 11. Variogram for high ash coal derived from on-line analyser z-score data.

By comparison, as long as the gauge is bias free (and the sampling system as well), the sampling system performance leaves a great deal to be desired.

This example points up the problems of attempting to calibrate an on-line gauge for coal ash and coal ash constituents against routine samples taken over the measurement period. To be of value in this setting, the sampling system must be unbiased and very precise. Conventional sampling systems rarely deliver this accuracy, so calibration against such sampling systems is impractical. A gauge manufacturer must offer a robust factory calibration procedure and this must be accepted by the buyer.

Conclusions

The determination of the precision of a sampling system requires that the punctual variogram for the process stream be known with some accuracy and that the variance due to the intrinsic heterogeneity of the primary increments as well as the variance added during sample preparation and analysis be known. The latter measurement uncertainty can be determined from analysis of a variogram based on consecutive samples (not increments) taken by the sampling system. With both these sources of information, the total sampling variance can be calculated.

The precision of an on-line analyser can be determined from a variographic analysis of the gauge output, as long as the gauge has not been set up to smooth the output by some statistical procedure such as a moving average. The unadulterated output on a small time interval must be available for construction of the variogram. The precision of the gauge over longer measurement time intervals is not affected by the time variation of the analyte content in the process stream because the gauge 'sees' all of the stream all of the time; there is no error due to distributional heterogeneity. Therefore the precision over longer time intervals can be determined by the classical formula for the standard deviation of the mean of independent quantities. If there are N measurements in the gauging period, the final precision is simply $1/\sqrt{N}$ times the precision determined from the variogram.

References

1. G. J. Lyman, "Estimation of sampling variance and quality variance about the mean by interleaved sampling", Proceedings of the 5th World Conference on Sampling & Blending, Santiago, Chile, Gecamin, 2011, pp 175-184
2. G. J. Lyman, "Variograms: properties and estimation", Proceedings of the 6th World Conference on Sampling & Blending, Lima, Peru, Gecamin, 2013, pp 185-206
3. G. J. Lyman, F. Lombard, D. Edward, and C.J. Clarkson, "Determination of the precision of on-line coal analysers – Theory and practice", Proceedings of the 7th Australian Coal Preparation Conference, J. Smitham (ed), Australian Coal Preparation Society, 1995, pp 324-355.