Monitoring the lot mean and uncertainty estimates by piecewise local modelling

P. Minkkinen[a, b]
[a]Lappeenranta Lahti University of Technology (LUT), PO Box 20, FI-53851, Lappeenranta, Finland, email: Pentti.Minkkinen@lut.fi
[b]Sirpeka Oy, Lampisuonkatu 9, FI-53850, Lappeenranta, Finland, email: pentti.minkkinen@sirpeka.fi

Variography is an excellent tool for monitoring the long-range trend of continuous processes. Pierre Gy has presented a method that can be used for estimating the measurement variance of a lot mean as function of sampling frequency for different sampling modes: random, stratified, and systematic sample selections. The method involves the estimation of the intercept (also called the nugget effect) of the variogram at the time point zero, and numerical integration of the variogram. The method can also be used for optimising sampling plans. At the time when variography was developed on-line analysers were not available. Samples were extracted from the process streams and analysed in laboratories. It was important to optimise the sampling plans to control the analytical costs and the reliability of the plans in estimating the estimation error. For a reliable variogram more than thirty to forty samples had to be analysed. Consequently, the results could not be used on-line.

Currently process analysers are widely used to monitor continuous processes. Like in variographic estimation of the lot mean this method is based on the theory of stratified sampling. If the lot is divided into N_1 strata of equal sizes (or sublots) of which n_1 are sampled the variance of the lot mean a_L is

$$s^2_{a_L} = \frac{N_1 - n_1}{N_1 - 1} \cdot \frac{s_1^2}{n_1} + \frac{N_2 - n_2}{N_2 - 1} \cdot \frac{s_2^2}{n_1 \cdot n_2} \approx \frac{s_2^2}{n_1 \cdot n_2} \text{, if } N_1 = n_1 \text{ and } N_2 \gg n_2$$

Here s_1^2 is the variance between strata mean values and s_2^2 the within-strata variance, N_2 is the size of strata as the potential number of samples and n_2 the number of samples taken from the stratum. The great advance of stratified sampling is that only the within-strata variance propagates into the lot average if samples are taken from every stratum. With current process analysers measurements can be taken at short time intervals and that is used in the current method to estimate the process average and its variance continuously. Within a short range (or stratum in this case) a continuous process can be locally modelled with a line. With systematic sampling after a minimum of three measurements a line can be fitted to this range and the mean and variance of the range mean calculated. That is the first stratum. When the process progresses, the calculations are repeated for the new strata and values. It is important that the quality of the final lot can be monitored on-line, especially if lots of certain sizes and demanding quality specifications are produced.

The method is tested with different kinds of simulated and real data sets. This method can be easily modified also for 2D and 3D sampling targets.

Introduction

Variography is an excellent tool for monitoring the long-range trend of continuous processes. Pierre Gy[1] presented in his book a method that can be used for estimating the measurement variance of a lot mean as function of sampling frequency for different sampling modes: random, stratified, and systematic sample selections. The method involves the estimation of the intercept (also called the nugget effect) of the variogram at time point zero, and numerical integration of the variogram. The method can also be used for optimizing sampling plans. At the time when variography was developed on-line analysers were not available. Samples were extracted from the process streams and analysed in laboratories. It was important to optimise the sampling plans to control the analytical costs and the reliability of the plans in estimating the estimation error. Variography is an excellent tool for that purpose. More than thirty or forty samples had to be analysed for a reliable variogram. Variograms and the variance estimates derived from variograms present average properties of the investigated lot from the time interval that the variogram covers. Therefore, the results cannot be used on-line.

Minkkinen & Paakkunainen[2] have presented an optional method for variographic analysis. That method has been further developed in this study. Currently process analysers are widely used to monitor continuous processes. Like in the variographic analysis the estimation of the variance of the lot mean is also in this method based on the theory of stratified

doi: https://doi.org/10.1255/tosf.135

sampling. If the lot is divided into N_1 strata of equal sizes (or sublots) of which n_1 are sampled, the variance of the lot mean a_L is

$$s_{\bar{a}_L}^2 = \frac{N_1-n_1}{N_1-1}\cdot\frac{s_1^2}{n_1} + \frac{N_2-n_2}{N_2-1}\cdot\frac{s_2^2}{n_1\cdot n_2} \approx \frac{s_2^2}{n_1\cdot n_2}\ , \text{if } N_1 = n_1 \text{ and } N_2 \gg n_2 \quad (1)$$

Here s_1^2 is the variance between strata mean values and s_2^2 the within-strata variance, N_2 is the size of strata as the potential number of samples and n_2 the number of samples taken from the stratum. If samples are taken from every stratum only the within-strata variance propagates into the lot average, because $N_1 - n_1 = 0$. That is the great advance of stratified sampling. With current process analysers, measurements can be taken at short time intervals and that is used in the current method to estimate the process average and its variance continuously. Within a short range (or stratum in this case) a continuous process can be locally modelled with a line. With systematic sampling after a minimum of three measurements a line can be fitted to this range and the mean and variance of the range mean calculated. That is the first stratum. When the process progresses, the calculations are repeated for the new strata and values. When lots of certain sizes and demanding quality specifications are produced, it is important that the quality of the final lot can be monitored on-line. That makes it possible to classify the product based on the quality.

The method is tested with different kinds of simulated and real data sets. This same method is easy to modify also for 2D and 3D sampling targets.

Principle of the method

With modern process analysers measurements can be taken at short intervals. In chemical processes where large quantities are processed, changes in average process values are slow. As the consequence of this, within short time intervals the changes can be effectively modelled with linear models fitted to the measurement results within the intervals. These short intervals can be treated as sublots forming the total lot monitored. From the predicted values and residuals, i.e., from the differences between the measured and predicted values, the estimates of the residual variances and the variances of the sublot mean values can be estimated as shown below.

Local modelling of continuous data

In the following presentation the equations are written using the MATLAB style. A set of measurements a_i along time or distance axes with a constant lag between measurements can be presented as vector **y** = $[a_1, a_2, \dots, a_{n_L}]$, where n_L is number of measurements within the lot. Fitting a line to sublots of two consecutive measurements (systematic sampling) is done as follows:

$$\mathbf{X} = \begin{bmatrix}1 & 1\\ 1 & 2\end{bmatrix},\quad \mathbf{Y} = \begin{bmatrix}a_1 & a_2 & \cdots & a_{n-1}\\ a_2 & a_3 & \cdots & a_{n_L}\end{bmatrix}, \quad (2)$$

$$\text{Regression coefficients } \mathbf{B} = \mathbf{X}\backslash\mathbf{Y} = \begin{bmatrix}b_{11} & b_{21} & \cdots & b_{n_L-1,1}\\ b_{12} & b_{22} & \cdots & b_{n_L,2}\end{bmatrix} \quad (3)$$

$$\text{Predicted values of } \mathbf{Y_{pred}} = \mathbf{X}\cdot\mathbf{B} \quad (4)$$

The mean values of the substrata of duplicates are: $mean(\mathbf{Y}) = \bar{\mathbf{Y}}_{\boldsymbol{j}} = \left[\bar{a}_1, \bar{a}_2, \dots, \bar{a}_{n_L-1}\right] = mean(\mathbf{X})\cdot\mathbf{B}$ (5)

The variances of substrata j = 1 … n_L -1 are $s_j^2 = var(\mathbf{Y})$ and the relative variances are $s_{r_j}^2 = \frac{s_j^2}{\bar{a}_j}$.

The variance of the mean value of each stratum j for duplicates is: $s_{\bar{a}_j}^2 = s_j^2/2$.

The slopes of the lines fitted to duplicates are equal to the differences: **B**(2,:) = **Y**(1,:) - **Y**(2,:) and $mean(var\ (\mathbf{Y}))$ and $mean(\mathbf{B}(2, j)^2)/2$ are equal to the value of the variogram for lag = 1.

From duplicates it is possible to get the estimates of the total variances of sublots of 1 lag and their pooled value for the total lot.

The variance of the duplicates $var(\mathbf{Y}(:, \boldsymbol{j}) = s_j^2 \approx s_{short}^2 + s_{trend}^2$. If the trend within the lag is approximately constant, i.e., the slope is approximately constant and $s_{trend}^2 \approx 0$, from duplicates it is possible to estimate only s_j^2. The variance estimate $s_{short}^2 \approx s_{FSE}^2 + s_{anal}^2$, is the sum of the fundamental sampling variance and analytical variance. If the properties and concentrations of the particles in the mixture are known, s_{FSE}^2can be estimated theoretically (in variographic analysis this is usually called the variance of the nugget effect, v_0). Many publications recommend that the nugget effect of the variograms is estimated experimentally by fitting a line to five to ten first variogram points and extrapolating it to lag zero. That is an easy but unreliable method as was shown by Heikka & Minkkinen[4] and Minkkinen[3].

The observation vector can be folded into the **Y** matrix containing more than two rows as follows. If n_L is the length of the **y** vector and it is folded to **Y** having i = 1 … j rows**; X** must be modified accordingly. Again, the columns of **X** can be used for all substrata, i.e., the columns of **Y** of the lot.

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & .. \\ .. & .. \\ 1 & j-1 \\ 1 & j \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} a_1, a_2, \dots, a_{n_L-j+1} \\ a_2, a_3, \dots, a_{n_L-j+2} \\ \dots\dots \\ \dots\dots \\ a_{j-1}, a_j, \dots, a_{n_L-1} \\ a_j, a_{j+1}, \dots, a_{n_L} \end{bmatrix} \tag{6}$$

The regression coefficients of lines fitted to substrata (columns of **Y**) together with predicted values are again obtained from equations (3) and (4): $\mathbf{B} = \mathbf{X}\backslash\mathbf{Y} = \begin{bmatrix} b_{11}\ b_{21}, \dots, & b_{n-1,1} \\ b_{12}\ b_{22}, \dots, & b_{n,2} \end{bmatrix}$ and $\mathbf{Y}_{\boldsymbol{pred}} = \mathbf{X} \cdot \mathbf{B}$, and the mean values of the strata (=columns) of **Y** are obtained using Eq. (5). The difference to the previous case is that we can also calculate the residual (measured – predicted) matrix **E**:

$$\mathbf{E} = \mathbf{Y} - \mathbf{Y}_{\boldsymbol{pred}} \tag{7}$$

From the residuals, the residual variance for each substratum *j* can be calculated:

$$s_j^2 = sum(\mathbf{E}(:, j))/(j - 2) \tag{8}$$

The predicted mean values of the substrata *j* are $[\bar{a}_1, \bar{a}_2, \dots, \bar{a}_{n_L-I+1}]$ and are again calculated by using Eq. (5.)

The variances of the strata midpoints are

$$s_{\bar{a}_j}^2 = s_j^2/j \tag{9}$$

After the first *j* measurements of the lot are available, a line can be fitted to the first column of the matrix **Y** and the mean $\bar{a}_{j=1}$, residual variance $s_{j=1}^2$ and variance of the sublot mean, $s_{\bar{a}_{j=1}}^2 = s_{j=1}^2/j$, can be calculated. If the measurement vector is folded according to the Eq. (6) after each new measurement, a new stratum can be added to **Y**. Calculating the mean of the substrata *j* at the midpoint of the range *j* largely removes the effect of autocorrelation. When the process progresses a new mean can be obtained as the mean value of the strata covering the ranges from sublot 1 to the last value included in the lot that is monitored. As every stratum has the mean and standard deviation according to the Eq. (1), the between-strata variance is eliminated from the variance of the lot mean and only the within strata variances propagate into the lot mean. When new sub-lots are completed, cumulative values can easily be calculated, e.g., by using the following short MATLAB code where aLj(1:,*j*) are the predicted mean values of sublot midpoints from Eq. (5) and varaLcum(*j*) the variance of lot mean as function number of strata *j* :

```
for j = 1 : nL- j +1,
    aLcum(j) = mean(aLj(1:,j));
    varaLcum(j) = mean(s2aj)/j;
end
```

In the following section examples with some simulated and real data sets are analysed with this new proposed method. For comparison, results are also calculated ignoring autocorrelation. The notation used in presenting results in the examples are:

Process mean value and variance Ignoring autocorrelation are: $a_L = mean([a_1, a_2, \dots, a_{n_L}])$ and $s_a^2 = var([a_1, a_2, \dots, a_{n_L}])$. If the autocorrelation is not taken into account, the variance of the process mean is calculated as function of the number of samples using Eq. 10.

$$s_{a_L}^2 = s_a^2/n_L \tag{10}$$

Examples

Example 1 is a simulated data set presenting a process with linear drift. The process and the results modelling the process with moving windows of three observations are shown in Fig. 1. It shows the line presenting a time interval from a linearly increasing process and 30 measurements with random noise presenting a nugget effect (sum of the fundamental sampling error + analytical error of the measurements). Fig. 1 shows also the variogram and the variance estimates for systematic sampling estimated from the variogram using Gy's method. When a line was fitted to seven first values of the variogram, extrapolation to lag = 0 gave for V_0 a negative value. As the variance estimate cannot be negative, a value $V_0 = 0$ was used in estimating the variance estimates corrected for autocorrelation for the systematic sampling mode.

If the process is linear, the detrending (or correction for autocorrelation) can be made by fitting the line to all measurement points and calculating the measurement variance from the residuals, $\mathbf{Y} = [a_1; a_2; \dots; a_{n_L}]$ in Eq (6).

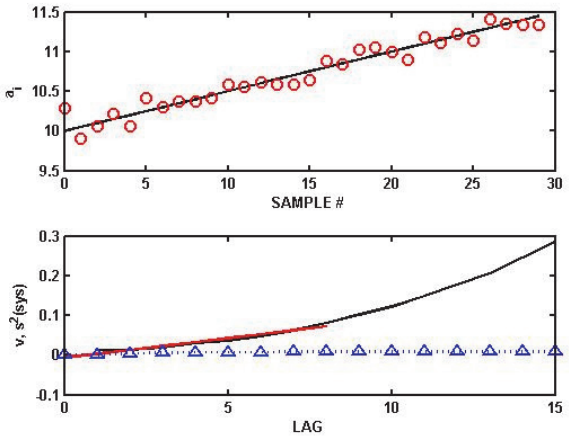

Figure 1. Example of a linear process with measurement noise (upper panel). The lower panel shows the variogram. The variogram extrapolated to lag zero gives a negative value. Therefore V_0 = 0 was used in calculating the variance values for systematic sampling (triangles).

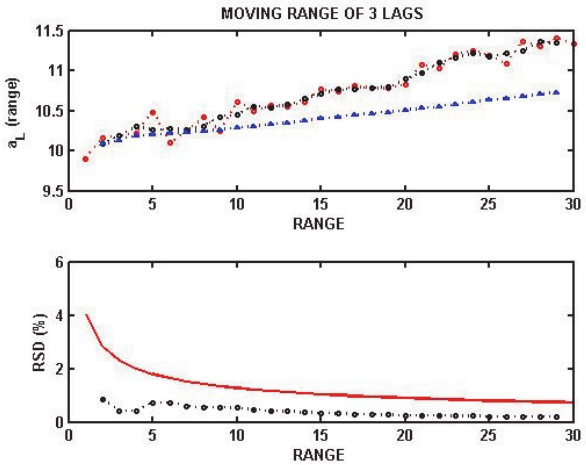

Figure 2. The upper panel shows 30 measurements from the simulated linear process (red dots) and mean of moving windows of 3 lags. The lower line shows the moving average as function of increasing range. The lower panel shows the relative standard deviation as a function of the number of measurements; the upper (red line) ignores the autocorrelation and in the lower panel the relative standard deviation is estimated with the current method from the consecutive moving windows of three measurements.

The upper panel of Fig. 2 shows the measurements, mean values of the moving windows of three measurements, and the moving lot mean as a function of measurement, or increasing range monitored. The lower panel of the figure shows, for comparison, two relative standard deviation estimates as a function of the number of measurements. Many sampling guides even today ignore the effect of autocorrelation and evaluate the mean and the uncertainty of the mean based on random distributions like normal, binomial or Poisson distributions. The figure presents both estimates. Table 1 shows the summary of the results of this experiment. The difference of the uncertainty estimates is large: The relative standard deviation of the mean, a_L, of 30 measurements is clear: 0.73 % ignoring the autocorrelation and 0.0107 % estimated from the residuals from strata of three measurements using the proposed method. The variogram gave a lower value, 0.098 % because the extrapolation of the variogram to lag zero underestimated the value of V_0.

Example 2 presents the results of a simulated periodic process, also contaminated with a random noise. Table 2 gives the summary of this process. This data was analysed by using variography. The variogram up to lag 20 and the estimates of variance of the systematic sampling as function of sampling lag are presented in Fig. 3. In this case extrapolation to lag zero gave a negative value instead of the expected design value of 0.0165. V_0 = 0 was used as the nugget value also in this case for estimating the variance of systematic sampling by using Gy's method. Gy's estimate of the relative standard deviation with lag 1 for the mean of 80 measurements was 0.073%. Estimated from the residuals the result was 0.085 %.

Fig. 4 shows 80 of the first measurements from the simulated process, the mean values of moving windows of three measurements, and the relative standard deviation estimates of the lot of using the current method and ignoring the autocorrelation.

Table 1. Properties of the linear simulated data set analysed in Example 1. Ignoring the autocorrelation, the relative standard deviation of the total Lot of 30 measurements is 0.73 %. Estimated from the residuals of the line fitted to all 30 measurements it is 0.17 %. V_0 is the variance (nugget effect) obtained by extrapolation of the variogram to lag = 0. Estimated from the variogram by Gy's method he relative standard deviation estimate is 0.098 %.

STATISTICS	MEAN	VARIANCE
NOISELESS PROCESS	10.725	0.194
NOISE	0	0.0108
TOTAL, s_a^2	10.729	0.184
FROM RESIDUALS		0.0107
V_O(EXTRAP)	-0.0065	
RSD estimates assuming randomness: $100\frac{\sqrt{0.184/30}}{10.725}\% = 0.73\%$		
RSD from residuals: $100\frac{\sqrt{0.0107/30}}{10.725}\% = 0.17\%$		

Table 2. Properties of the cyclic simulated data set analysed in Example 2. The V_0 estimate was obtained by extrapolation to lag = 0. The relative standard deviation of the total Lot of 80 measurements is 0.085 % estimated by the current method, and 0.073 % estimated by Gy's method from the variogram. If the autocorrelation is ignored the relative standard deviation estimate is 0.81 %.

STATISTICS	MEAN	VARIANCE
NOISELESS PROCESS	10.000	0.5063
NOISE	0.0055	0.0165
TOTAL, s_a^2	10.0055	0.5304
FROM RESIDUALS		7.35×10^{-5}
V_O(EXTRAP)	-0.13	
RSD estimates assuming randomness: $100\frac{\sqrt{\frac{0.5304}{80}}}{10.00555}\% = 0.81\%$		
RSD from residuals: $100\frac{\sqrt{7.35\times10^{-5}/80}}{10.00555}\% = 0.085\%$		

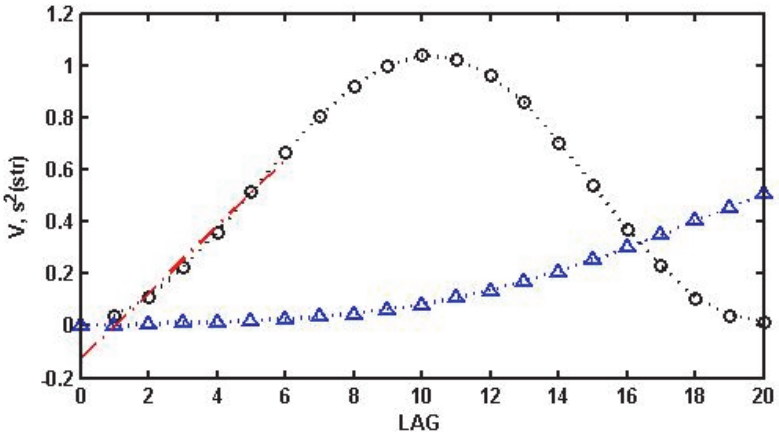

Figure 3. Absolute variogram of the periodic process up to lag 20 and extrapolation to V_0 together with the variance estimates for systematic sampling.

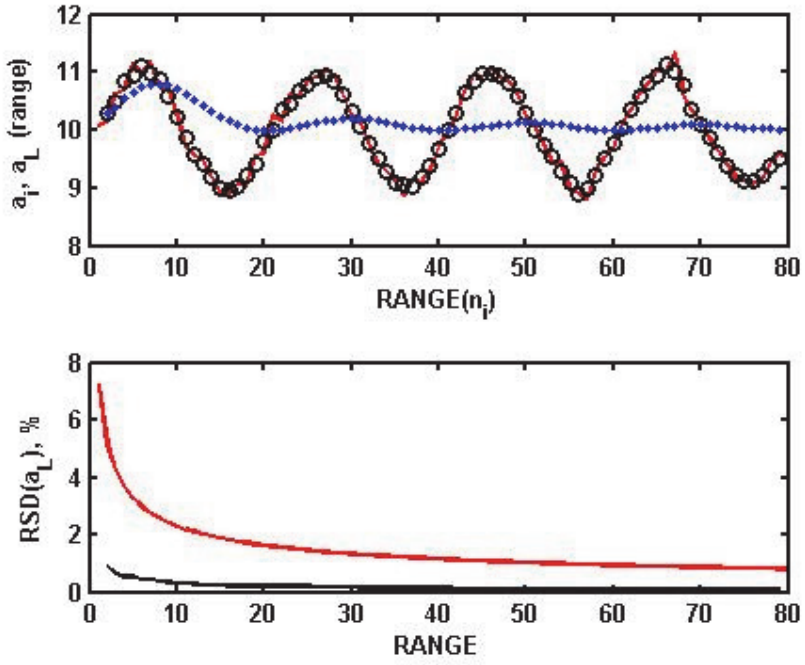

Figure 4. Upper panel: Process values (red line) and mean values of the windows of three measurements (black circles) and moving average as function of increasing range. Lower panel: Relative standard deviation estimates ignoring the autocorrelation (upper line) and the estimates with current method (lower line) as a function of increasing range.

Example 3: Data for this example was taken from the study carefully analysed by using variography in references[5 – 7]. In that study 100 samples were taken during unloading from a shipload of soybeans. Samples were analysed for genetically modified material (GMO). In the European Union the material must be labelled as GMO containing material if the content exceeds 1 %. 0.9 % has been used as the decision limit for acceptance. It is supposed to give a 95% confidence that the true value does not exceed 1 %. Most sampling guides recommend that sample numbers of 4 to 8 is sufficient. That is illusory. With the segregation pattern of this example, in the above-mentioned references[5 – 7], with a sample size of 3000 beans the minimum number of samples (or increments making a composite sample) is 42.

Figure 5 presents the analytical results (a_i) as the relative heterogeneity contributions: $h_i = (a_{i+\text{lag}} - a_i)/a_L$ the relative variogram calculated from the heterogeneities. The concentration of GMO material shows high variability. This is a good data set to demonstrate the value of the method proposed here. If the concentration of the analyte can be measured on-line, the average quality of the process can be evaluated in real time and the product could be classified based on the average quality. This example shows what could be achieved with on-line measurement of the GMO content if such a technology were available. In this example the quality was estimated after each twenty samples taken (equivalent of dividing the cargo during unloading into five sub-loads). The results are presented in Figures 6 -10 and in Table 3.

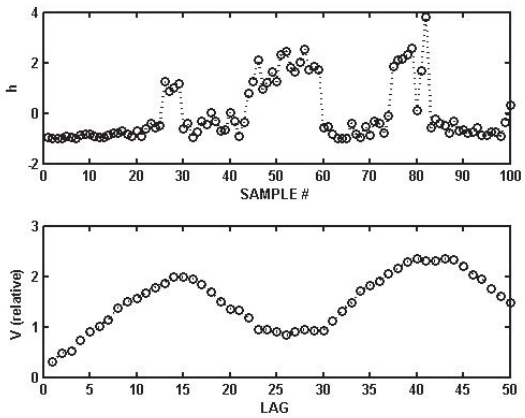

Figure 5. Relative heterogeneity contributions of GMO in 100 samples taken during unloading a shipload of soybeans together with the variogram.

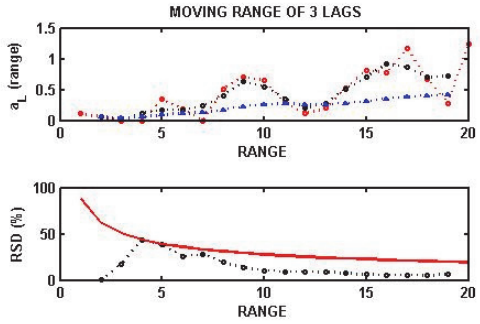

Figure 6. Measurements and moving mean calculated from three sample windows of the first substratum consisting of twenty samples (upper panel). The lower panel shows the relative standard deviation estimate of the moving mean. The continuous line shows the results if autocorrelation is ignored and the dotted line the results calculated from the residuals.

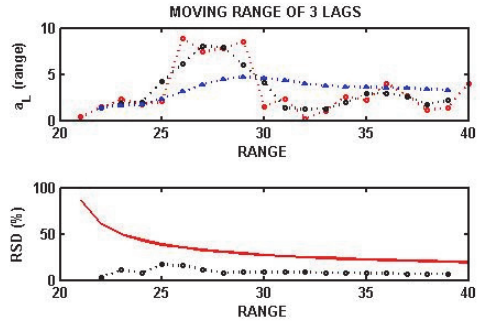

Figure 7. Results of the second stratum of twenty samples.

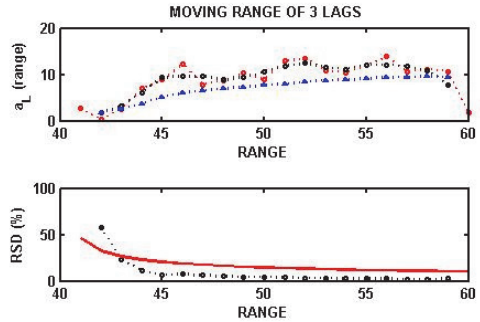

Figure 8. Results of the third stratum of twenty samples.

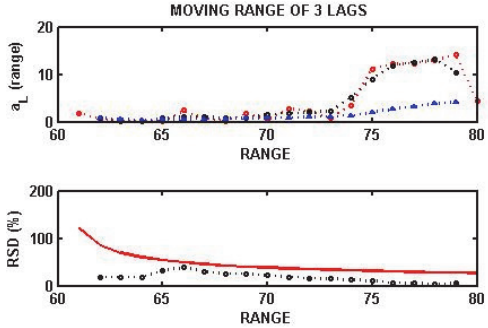

Figure 9. Results of the fourth stratum of twenty samples.

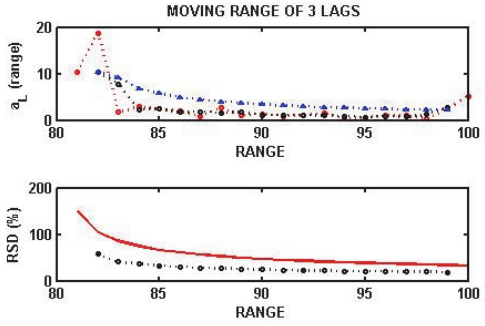

Figure 10. Results of the fifth stratum of twenty samples.

Table 3. Summary of GMO example

RANGE	MEAN (%)	VARIANCES		RSD %	
		var(a_i)	From residuals	From var(a_i)	From residuals
1 - 20	0.423	0.1431	0.0392	19.6	11.0
21 - 40	3.32	7.5	2.62	19.3	11.5
41 - 60	9.47	16.4	2.82	10.2	4.18
61 - 80	4.19	25.9	2.39	26.9	8.70
81 - 100	2.29	19.1	9.44	33.2	31.6
Mean of ranges	3.94	13.8	3.46	9.4	11.1
1 - 100	3.96	21.0	3.80	11.6	13.4
V_O(EXTRAP)	-0.13				

This example shows the advantage of on-line measurements in monitoring continuous processes, raw materials and/or products. When analytical results are obtained with short intervals, the role of random error can be filtered from the estimation of the lot mean. A great advantage is, if a product with high quality specifications is manufactured, it can be continuously classified to product lots according to the quality. Like in this example, the average concentration of the first fifth of the cargo unloaded from the ship is 0.423 %, well below 1 %, and could be labelled as non-GMO material.

Example 4: Data for this example consisted of 100 process analyser measurements of one of the components at 10 min intervals from the feed to a flotation plant. The variance plot clearly shows the noisy parts of the process and the moving mean plot how the noisy parts affect the mean of the lot. That can be valuable diagnostic information for process control.

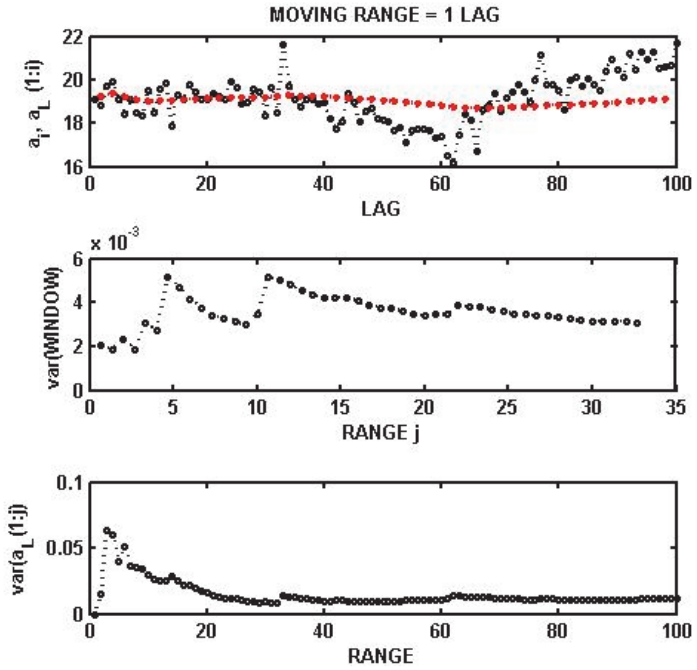

Figure 11. Results of a range of 100 process analyser measurements from a feed to a flotation plant. The sampling interval used was 15 min covering a total of 25 production hours. The top panel shows the analyser results and the middle panel the residual variance, s_j^2 , of the windows of *j* measurements. The lowest panel shows the variance of the mean of the moving average.

Modification of the method to 2-dimensional lots

It is easy to modify this method to handle also 2-D lots. If the samples have been taken at regular grids, a plane can be fitted to squares taken from the grid. **X** and **Y** need to be redefined:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 2 & 2 \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{bmatrix}$$

The second and third row of **X** are the coordinates of the corners of a square on the plane from which the samples are taken and **Y** the vector of the corresponding analytical results. Otherwise, the equations used for the 1-D case apply. Modification for a 3-D case can be done the same way. For a more complicated sampling grid it is possible to use experimental designs.

Summary

Variography is very useful and a well-established method to analyse the long-range variability of 1-D processes. It can provide variance estimates for different sampling modes corrected for autocorrelation and for optimising the sampling interval when samples are cut from the process streams for laboratory analysis. The results are usually available long after a certain lot is produced. The results are also based on the average values of the process.

Process analysers on the other hand can produce results at high frequency. If the traditional variography is used for estimating the sampling variance of the lot, only a part of the data is used. The method proposed here gives the results, variance and mean of the progressing lot in real time. This is especially useful if the product has strict quality specifications (e.g., medicals, foodstuff, fertilizers). If the product is delivered in containers or big bags, they can be assigned individual certificates of their content.

While with increasing number of measurements the effect of the random part of the measurement error becomes insignificant in the lot average, that does not eliminate the systematic errors. A lot of process analysers are based on different spectroscopic techniques which get information only from a thin layer of the process stream. That can be a problem when materials which have a high tendency to segregate are analysed. When such materials, like crushed particles or powders on conveyor belts, are analysed and complete cross-sections cannot be sampled, the only solution is to randomize, if possible, the material at the point where the process analyser is installed. If that is not possible, material balance calculations cannot be based on process analyser results. Still, the results can be useful in monitoring the process behaviour with time.

References

1. P. M. Gy, *Sampling of Heterogeneous and Dynamic Material Systems, Theories of Heterogeneity, Sampling and Homogenizing*. Elsevier, Amsterdam, (1992).
2. P. Minkkinen and M. Paakkunainen**,** "Direct Estimation of Sampling Variance from Time Series Measurements – Comparison to Variographic Analysis", in *Proceedings 2nd World Conference on Sampling and Blending (WCSB2)*. AusIMM, pp.39-44 (2005).
3. R. Heikka and P. Minkkinen, "Comparison of some methods to estimate the limiting value of the variogram $v_h(j)$for the sampling interval j = 0 in sampling error estimation", *ACA,* **346**, 277-283 (1997).
4. P. Minkkinen, "Properties of Process Variograms", in *Proceedings 6th World Conference on Sampling and Blending (WCSB6)*. AusIMM, pp.241-248 (2013).
5. K. H. Esbensen, C. Paoletti and P. Minkkinen, "Representative Sampling of Large Kernel Lots, Part I - Theory of Sampling and variographic analysis", *TrAC* **32**, 154-164 **(**2012).
6. P. Minkkinen, K. H. Esbensen and C. Paoletti, "Representative Sampling of Large Kernel Lots, Part II - Application to Soy Bean Sampling for GMO Control", *TrAC* **32**, 165-177 **(**2012).
7. K. H. Esbensen, C. Paoletti and P. Minkkinen, "Representative Sampling of Large Kernel Lots, Part III - General Considerations on sampling heterogeneous foods", *TrAC* **32**, 178-184 **(**2012).

,