

What is the optimum regime for calibration bias adjustments?

Steven Dargaville¹
Joanne Simpson⁴

Mali Abdollahian²
Robert McKibbin⁵

Rob Crawford³
Bob Anderssen⁶

(Received 18 October 2012; revised 16 September 2013)

Abstract

Fonterra, a New Zealand Dairy Company, processes milk into various products, including milk powder and cheese, for export all around the world. The challenge for Fonterra is to guarantee that the composition of a product meets agreed contract specifications. The Study Group was asked to determine if the calibration monitoring scheme currently in place at Fonterra could be improved and what other information could be derived from monitoring. As a result of the Study Group's deliberations, a number of recommendations have been made—one of which has already been implemented (scanning a check-cell daily to monitor the instruments' performance). Fonterra are currently investigating the impact of some of the other recommendations made by the group, including switching to using mean values from median

<http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/6246> gives this article, © Austral. Mathematical Soc. 2013. Published October 19, 2013, as part of the Proceedings of the 2012 Mathematics and Statistics in Industry Study Group. ISSN 1446-8735. (Print two pages per sheet of paper.) Copies of this article must not be made otherwise available on the internet; instead link directly to this URL for this article.

values, and implementing a strategy to monitor laboratory reference values.

Contents

1	Background	M48
2	The issue	M50
3	Calibration/testing	M53
4	Control charts	M55
4.1	Univariate \bar{X} control charts	M56
4.2	Exponentially weighted moving average control chart . . .	M60
4.2.1	Using the EWMA chart for one-step ahead forecasting of the process mean	M66
4.2.2	Comparisons of the mean errors for the three labs using ANOVA	M66
4.3	Multivariate control charts	M67
5	Derivative spectroscopy (an aside)	M72
6	Recommendations and conclusions	M78

1 Background

The dairy industry, like many others, makes multiple measurements to determine product composition (see Figure 1 for an example of Fonterra’s production process). Traditionally, samples have been taken and sent to a laboratory for testing (referred to as the “reference method”). These methods are often slow, sometimes taking hours or days to give a result. As a consequence

of this, rapid instrument methods that are much cheaper and faster to use are becoming increasingly common. However, control charting suggests that these rapid methods are often be unstable relative to the reference method and so the accuracy of the rapid method must be carefully monitored.

For example, the moisture content of milk powder is an important parameter, not least because the moisture content must be less than 5% by weight for the product to be legally called milk powder. In the case of moisture, one method for reference testing involves three weighing stages, followed by the sample being dried for a minimum of three hours in a 102°C oven. This is a time-consuming process, as there are very exacting requirements on the glassware, oven, balance calibration, etc. This gravimetric method only measures free water; water tied up in the powder as water of crystallisation is not included in this moisture result. Other properties that are tested include protein and fat levels.

As technologies have developed, Fonterra now use a range of rapid test methods that are purported to provide test results for the same attributes. There are many technologies that can be used but they have focused their attention on absorbance/reflectance of near infrared (NIR) light by the sample and the development of a quantitative calibration based on NIR. The NIR moisture method essentially measures how much energy is absorbed in O-H bond vibrations at a range of frequencies (an overview is given in Figure 2) and so gives an indirect assessment of how much H₂O there is in the sample. The method must be calibrated: using Moving Least Squares (MLS), Partial Least Squares (PLS) or Artificial Neural Network (ANN) techniques allows the construction of a relationship between the absorbance or reflectance at a range of wavelengths and the measured moisture in each sample within a calibration dataset of representative samples. However, experience shows that neither the reference method nor the NIR method are ever truly stable; seasonality, balance drift, differences in oven temperature and optical variation all contribute to test method variability.

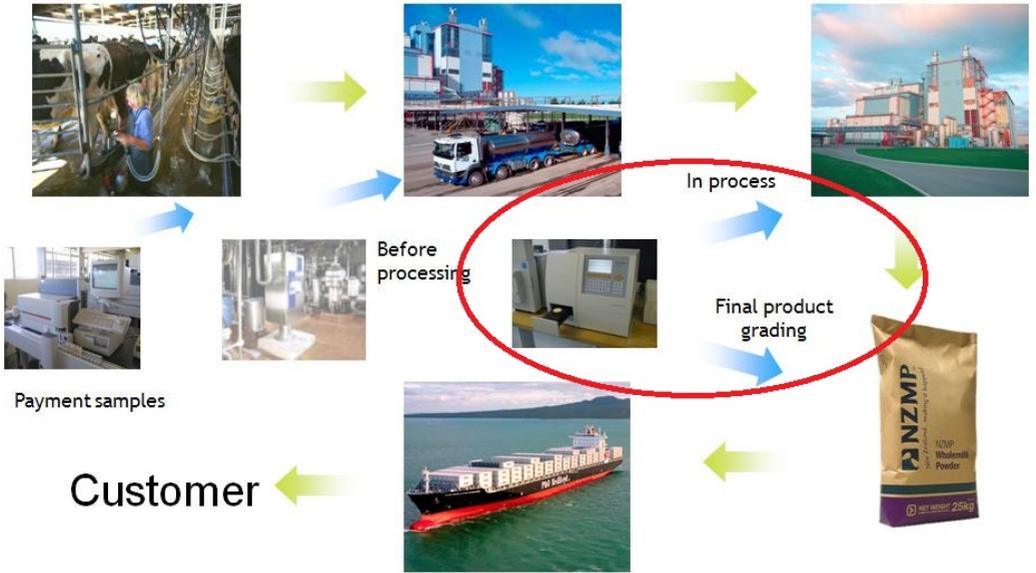


Figure 1: There are many sample testing points in the production process. This project examines part of the final product testing procedure (as circled).

2 The issue

Once a calibration has been developed, validated and released for use, it must be continually monitored and adjustments made periodically. Typically, standard statistical process control (SPC) charts have been used to monitor the difference (or bias) between measurements made with the reference and the NIR instruments. Standard SPC rules are then used to alert the user when it appears that the bias (or variation) is out of control. Unfortunately, dairy products tend to degrade in storage, so it can be impossible to maintain true certified reference measurement (CRM) samples that would be the ideal basis for such SPC tools. Also, a sample might only be available for some short time interval.

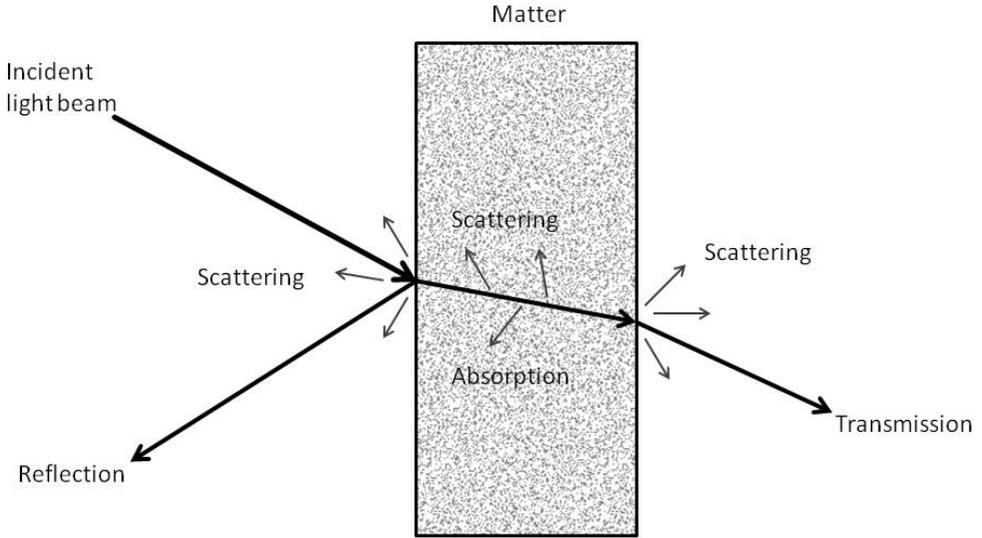


Figure 2: NIR analysis works on the principle of subjecting the material to be analysed to a light source (and varying the wavelengths). The molecular properties of the material determine the amount of light reflected by or transmitted through the material, with minor light scattering taking place throughout the process. Depending on the material being tested or the type of instrument used for analysis, either transmission or reflection analysis is employed. The examination of the transmitted or reflected light reveals how much energy was absorbed at each wavelength, producing spectral information that can then be coupled with laboratory results to form a calibration.

Fonterra have been monitoring the distribution of apparent differences between NIR readings and the reference measurements (the wet-lab results). Typically, the differences might be expected to be normally distributed with a zero mean. However, other factors need to be accounted for, including

- (random) step changes that occur at four-day intervals (due to the staff roster),
- variations due to ambient conditions (with 24 hour periodicity),
- seasonal variations during the dairy season,
- annual step changes (as lab equipment has annual preventative maintenance).

One strategy used by Fonterra to mitigate some of these problems is to introduce a more comprehensive programme of monitoring, with several reference laboratories and multiple NIR instruments each running the same calibration. The multiple readings should provide a better mean value (although the process will still be subject to inherent overall method bias).

Up to now, many aspects of the monitoring scheme have been set intuitively, including

- the frequency of monitoring (e.g., daily or weekly),
- the number of samples used in each period,
- the parameters of the control chart.

The main questions Fonterra were interested in centred around this monitoring process:

- Can a mathematical framework be developed that can be applied to determine the rules of a SPC monitoring scheme for this type of application, that are in some sense optimal?
- What other information can be derived from such a monitoring scheme?

The focus of the Study Group’s activities therefore became

- to investigate the monitoring process currently in use at Fonterra,
- to gain a greater understanding of control charts and statistical process control and how they are used as part of the monitoring process at Fonterra,
- to attempt to quantify some of the uncertainty in the NIR and “reference method” testing,
- to incorporate this uncertainty into control charts, in order to understand when a process that is being monitored is “out-of-control”,
- to investigate the extent to which derivative spectroscopic analysis of the data from the NIR machines, provided by Fonterra, could be utilised in monitoring the “health” of the NIR machines and highlight the background role played by derivative spectroscopy in calibration and prediction.

3 Calibration/testing

As mentioned, Fonterra use NIR spectroscopy at various points in their production chain to detect the levels of properties in samples of their products, for example, measuring fat, moisture and protein levels in milk powders. IR spectroscopy involves exposing a sample to light in the IR spectrum and recording the levels of absorbed light. Fonterra use two main types of NIR spectroscopy, which operate on different wavelengths in the near-IR region (in the range 0.8–1.0 μm). Typically, the instrument provides a spectrum that has been averaged over 100–200 scans, providing highly accurate results. The spectra returned from the NIR machine (Figure 3 shows an example) is then processed to determine measurements of the desired properties. A brief overview of the processing step is given below, taken from [Anderssen, de Hoog & Wesley \(2011\)](#). Beer’s law ([Osborne et al. 1993](#)) states that the

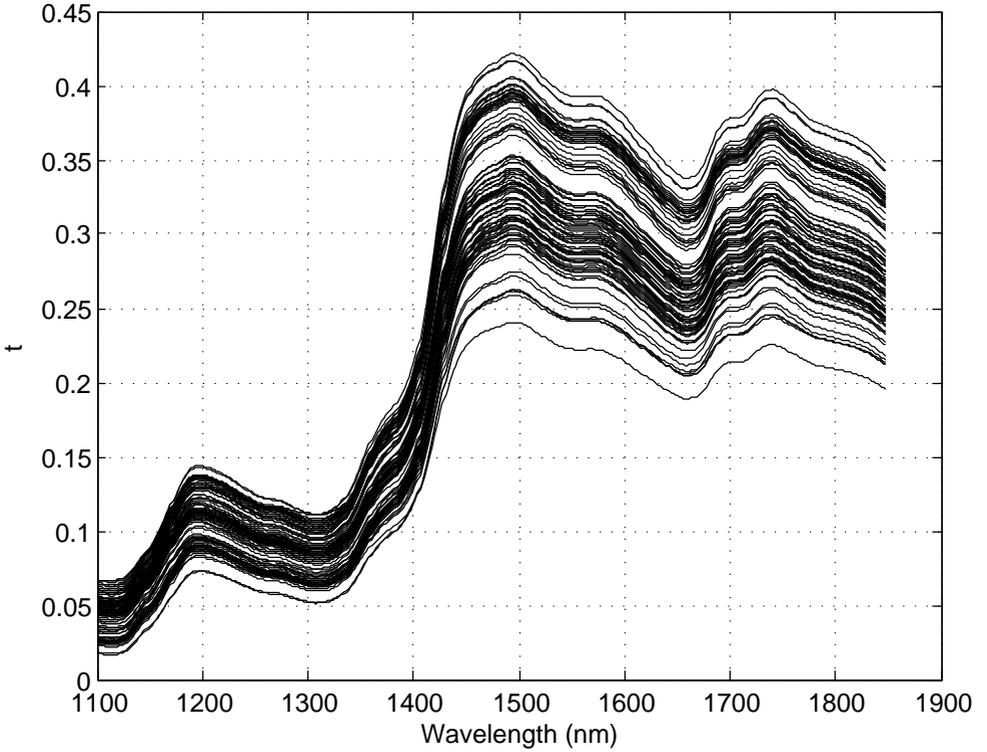


Figure 3: Spectrum for 100 different skim milk powder samples.

absorbance A depends on the path length of radiation through the sample β and the concentration γ in the form $A = \alpha\beta\gamma$, where α is the absorptivity. With this law we form an underdetermined system of equations using a series of spectra from a test set:

$$\mathbf{S}\boldsymbol{\beta} = \mathbf{p},$$

where $\mathbf{S} \in \mathbb{R}^{m \times n}$ is the spectral data given m different samples in a “training” set and n individual wavelength values (with $m \ll n$), $\mathbf{p} \in \mathbb{R}^{m \times 1}$ are the property values (such as fat or moisture) measured using some reference method, and $\boldsymbol{\beta}$ is the unknown (regression) coefficient vector to be determined. Solving this underdetermined linear system (using an ANN or a method like PLS) gives a value for $\boldsymbol{\beta}_k$; this indicates which of the k components in the spectrum are important to determine our property \mathbf{p} . This process is known as calibration, and once this calibration has been performed, this $\boldsymbol{\beta}_k$ is used to determine estimates of a given property from any new sample. Fonterra already have a process in place to solve this calibration problem.

Such estimates are compared with chemistry laboratory reference analysis of the same sample. As mentioned, these measurements are often sent to several different laboratories, with each lab taking several measurements of the same sample. These different measurements are then be combined to obtain an average measurement; a typical arrangement of the averaging process is shown schematically in Figure 4.

4 Control charts

Once calibrated, the NIR machines are then used in the manufacturing process, and control charts are used to both detect out-of-control processes during manufacturing (i.e., NIR instrument compared with lab results) and to monitor machines that are out-of-control with respect to their calibrations.

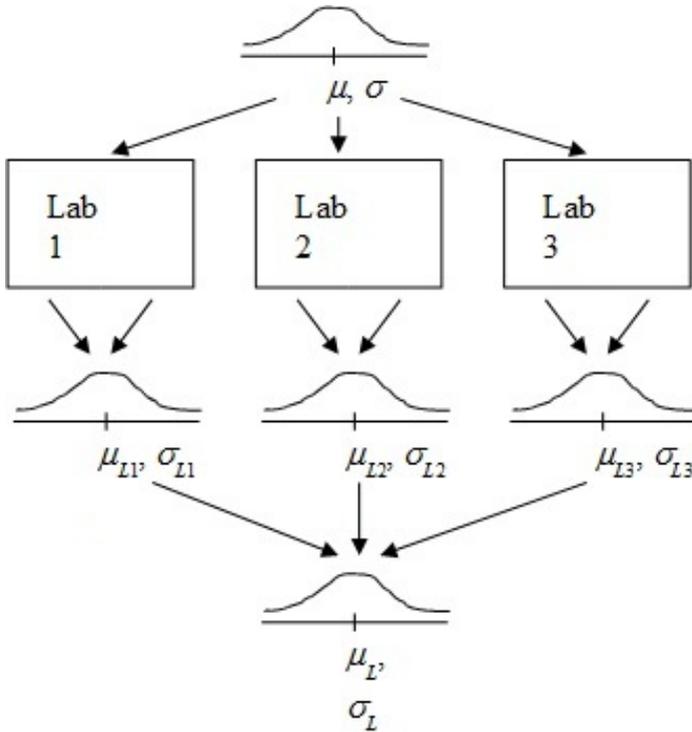


Figure 4: Sample distributions of a particular property vary according to the measurement method.

4.1 Univariate \bar{X} control charts

Univariate control charts are used to monitor statistics including the difference between measurements and a target mean. Control limits are then set on these charts to detect unusual samples. Figure 5 shows an example of a control chart which plots the difference between two measurements on the same specimen taken in one of the laboratories Fonterra use for testing.

There are two types of errors that are used to assess the monitoring process:

- Type I with probability $\alpha \rightarrow$ Assume process is out-of-control when it's

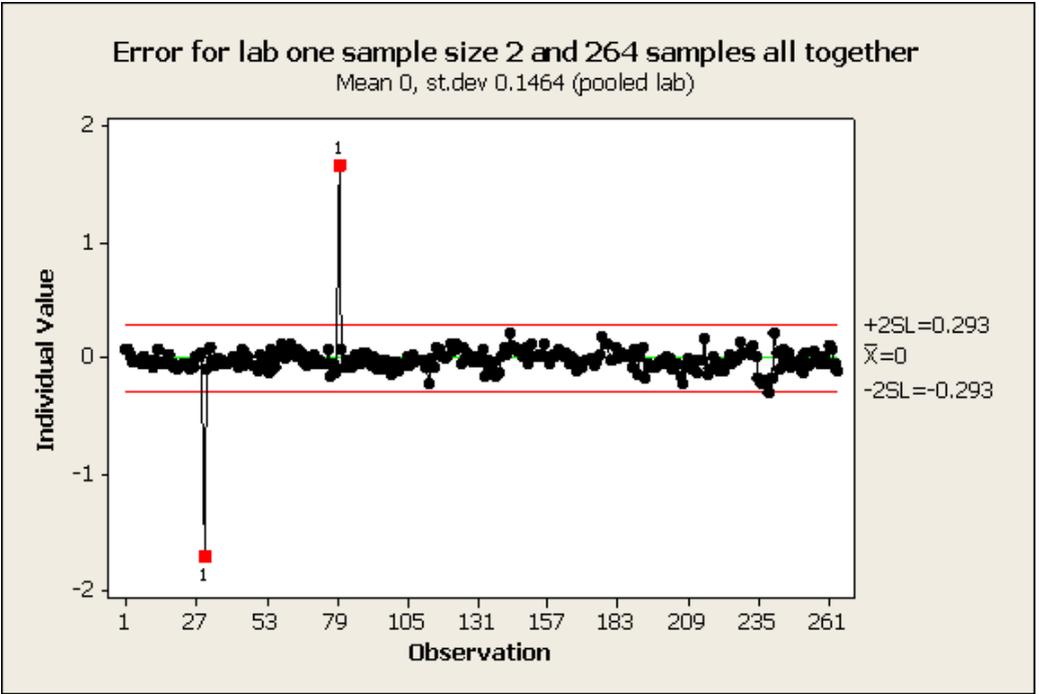


Figure 5: \bar{X} control chart for errors of Lab 1 with $K = 2$.

in-control;

- Type II with probability $\beta \rightarrow$ Assume process is in-control when it's out-of-control.

If the quality characteristics follow a normal distribution, then α and β for \bar{X} control chart are defined by

$$\alpha = 2\Phi(-K), \tag{1}$$

$$\beta = \Phi\left(K - \frac{\Delta\sigma}{\sqrt{n}}\right) - \Phi\left(-K - \frac{\Delta\sigma}{\sqrt{n}}\right), \tag{2}$$

where Φ is the cumulative normal distribution, n is the sample size, $\Delta\sigma$ is

the variation Fonterra would like to detect with probability β , and K is the position of the control limits in units of σ/\sqrt{n} .

One of the most important parameters is the probability β which is a function of the sample size n , variation to be controlled $\Delta\sigma$ and the position of the control limits K . This probability should be small, as Fonterra face large costs, in terms of both money and reputation, if they assume the process is in-control when it is not (i.e., Fonterra would be unaware they are shipping unsatisfactory goods to customers).

Suggestions made by the group include the following.

- Using univariate control charts (\bar{X} and EWMA, see section 4.2) to monitor the reference method, both in-lab variation and between-lab variation (currently not done), that is, the statistics plotted in the chart represent the differences between the two readings on the same specimen for a given lab (in-lab variation) or the differences between the readings of the specimen from two different labs (between-lab variation). Figures 5–7 show the \bar{X} control chart for the in-lab variation of Labs 1–3 with $K = 2$.
- Monitoring NIR results compared to labs (modified) to assess the performance of the individual lab. For example, plot the differences between the mean NIR and mean readings from individual lab. The lab with the smallest differences outperforms the other labs. Figures 8–10 represent the differences between the mean duplicate readings and mean instrument readings for each individual sample in Labs 1–3. The lab with the minimum variation is Lab 2.
- Setting control limits for the NIR results using the sum of variances from the NIR and lab results. This would lead to a more realistic estimate of the true variability in the process.

Figures 5–7 indicate that the reading error for some samples is larger than expected. These samples should be investigated to find the source of the out-of-control signals (large difference between the two readings).

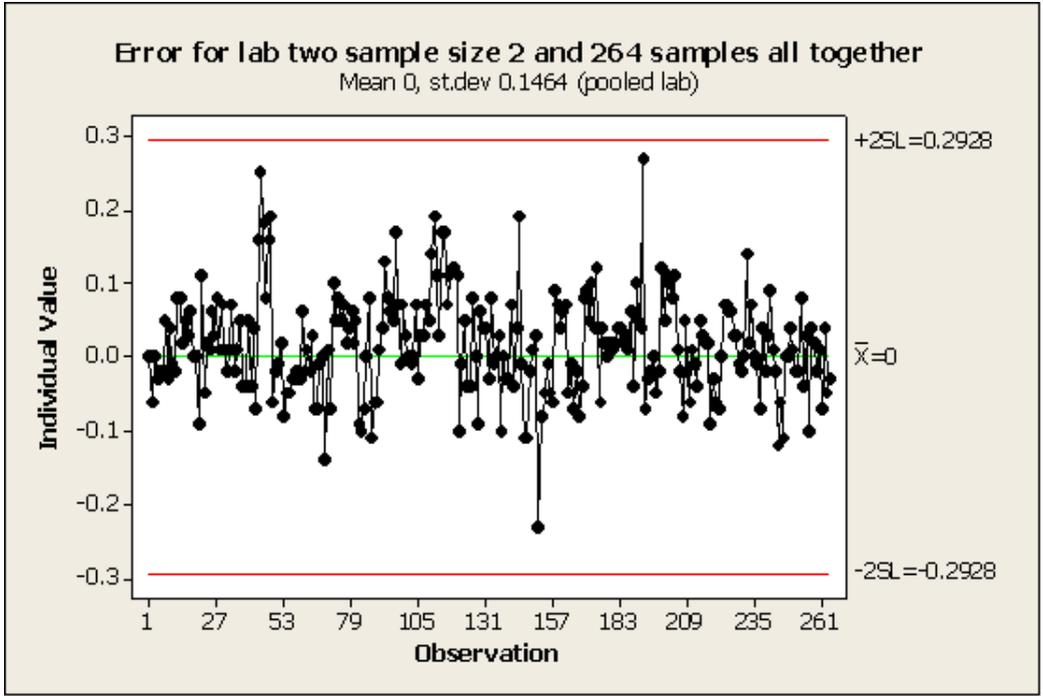


Figure 6: \bar{X} control chart for errors of Lab 2 with $K = 2$.

Figure 7 shows that Lab 3 has recorded the highest number (37 points outside the $2\sigma/\sqrt{n}$ control limits) of the out-of-control signals.

The results presented in Figures 8–10 clearly show that Lab 2 results are the closest to the instrument's results confirming the accuracy of Lab 2 (however, the standard deviation of Lab 2 is higher than that of Lab 3). The above results were also confirmed by comparing the box-and-whisker plots of difference between mean laboratory results and the mean instrument results for the individual lab that is presented in Figure 11.

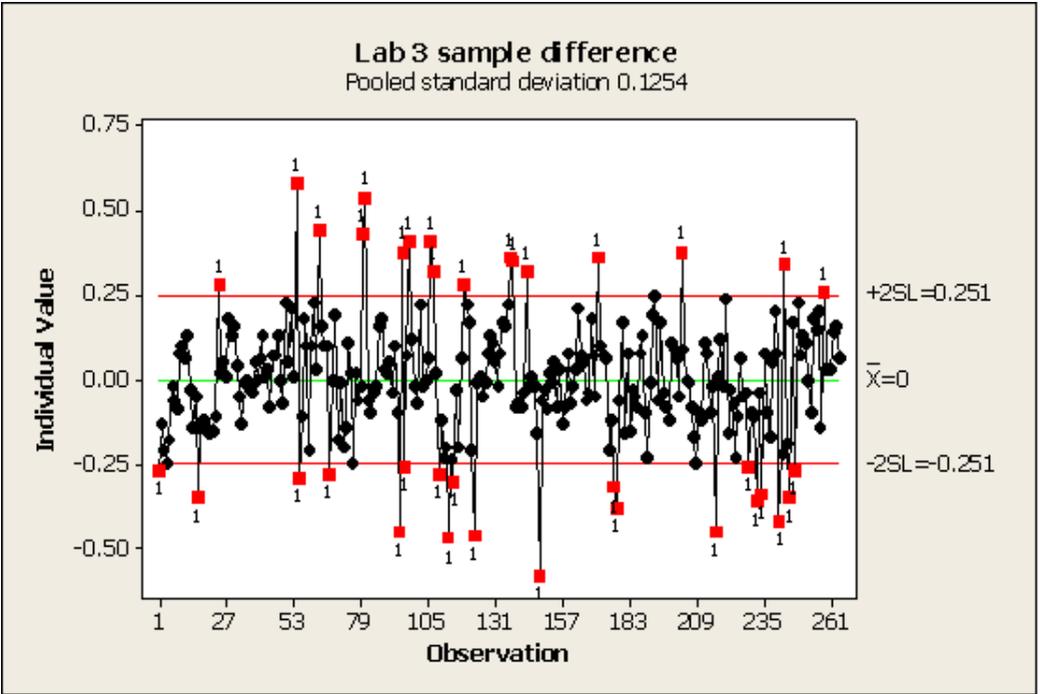


Figure 7: \bar{X} control chart for errors of Lab 3 with $K = 2$.

4.2 Exponentially weighted moving average control chart

One of the major assumptions in deploying the traditional \bar{X} control chart for monitoring a process is that of normality. The probability plot of the individual lab errors (Figure 12) indicates that none of the labs have errors that follow a normal distribution. Since the corresponding p-value for Labs 1–3 is less than 5%, we claim with 95% confidence that Labs 1–3 do not have normally distributed errors. However, the mean error for all three labs is very close to zero.

One of the most effective alternatives to the \bar{X} chart is the Exponentially

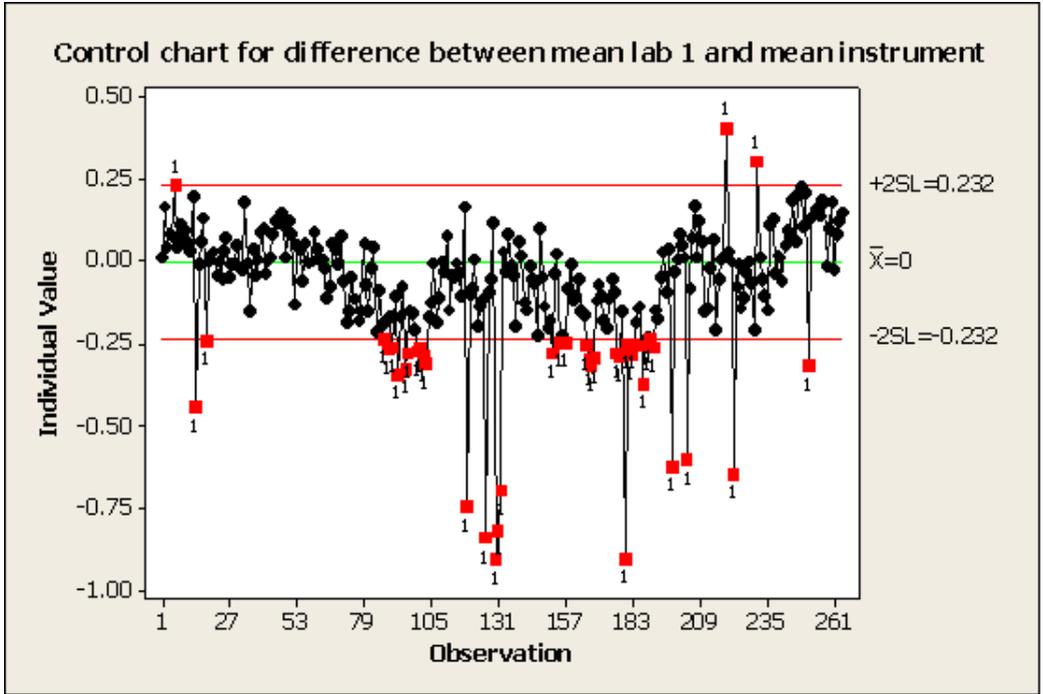


Figure 8: Control chart for the difference between mean Laboratory 1 results and the mean instrument results. The mean is forced to be zero and the standard deviation has been calculated by the software.

Weighted Moving Average (EWMA) chart. This chart was introduced by [Wortham & Ringer \(1971\)](#). This charting method was proposed for applications to the process such as chemical industries, financial and management control systems, particularly when the sample size is one. The chart is effective for detecting small to moderate shifts ($< 1.5\sigma/\sqrt{n}$) and is very insensitive to normality assumptions; therefore it is an ideal chart for individual observations when data do not follow a normal distribution. One of the aims of this study is to monitor the error per sample for an individual lab. The fact that errors do not follow a normal distribution and the sample size for the plotted statistics is one (for an individual sample we monitor the difference between

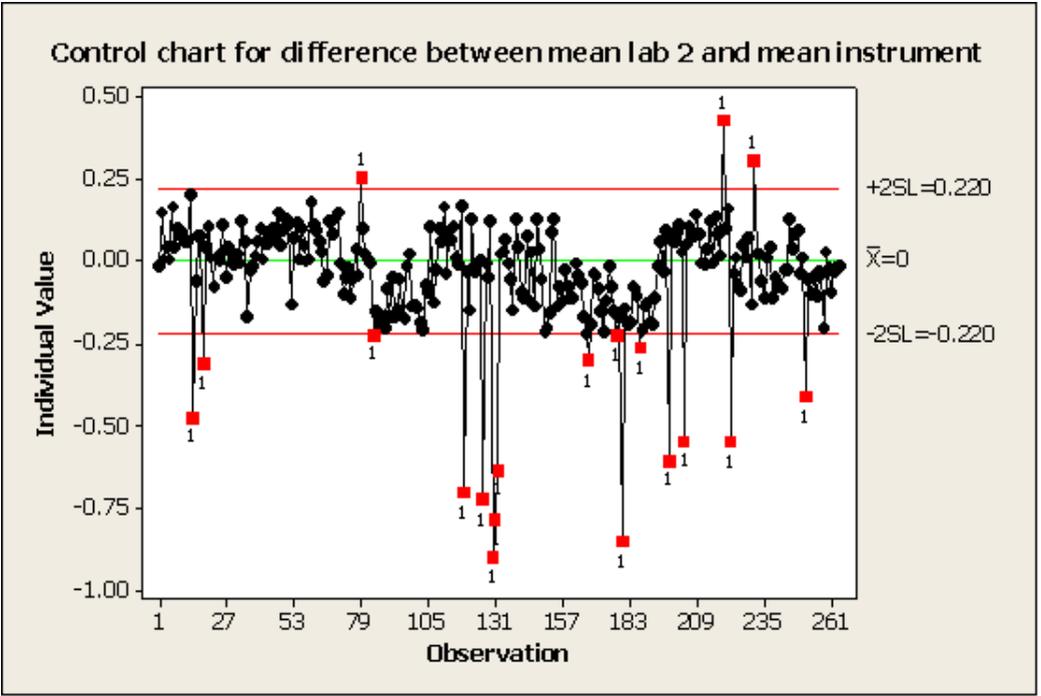


Figure 9: Control chart for the difference between mean Laboratory 2 results and the mean instrument results. The mean is forced to be zero and the standard deviation has been calculated by the software.

the two readings) suggests that the EWMA chart is superior to the traditional \bar{X} chart for monitoring the errors. The chart is used to control: an average (of a sample), the error of an individual sample, a ratio or a proportion (of defects in the sample).

In this chart, the older the observation, the less weight it conveys. The plotted statistics are exponentially weighted moving average defined by

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1}, \quad 0 < \lambda \leq 1, \quad z_0 = \mu_0,$$

where λ is the weight constant and the starting value μ_0 (required with first

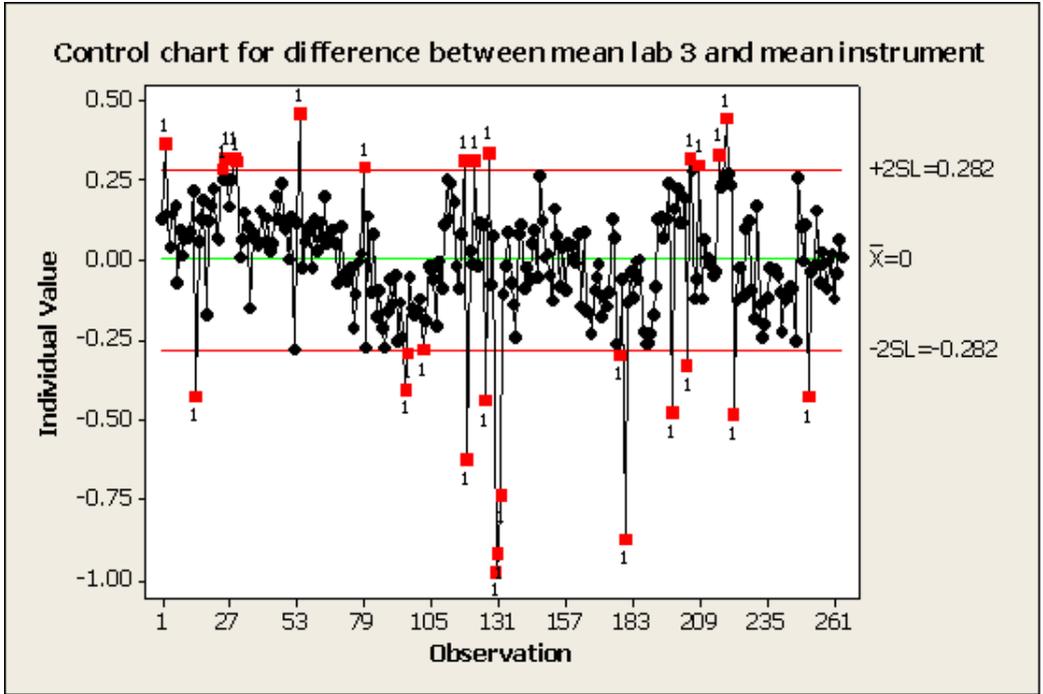


Figure 10: Control chart for the difference between mean Laboratory 3 results and the mean instrument results. The mean is forced to be zero and the standard deviation has been calculated by the software.

sample at $i = 1$) is the process target error (in this case zero). The plotted statistic z_i is a weighted average of all the previous sample statistics, or

$$z_i = \lambda \sum_{j=0}^{i-1} (1 - \lambda)^j x_{i-j} + (1 - \lambda)^i z_0.$$

The variance of z_i is defined by

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}].$$

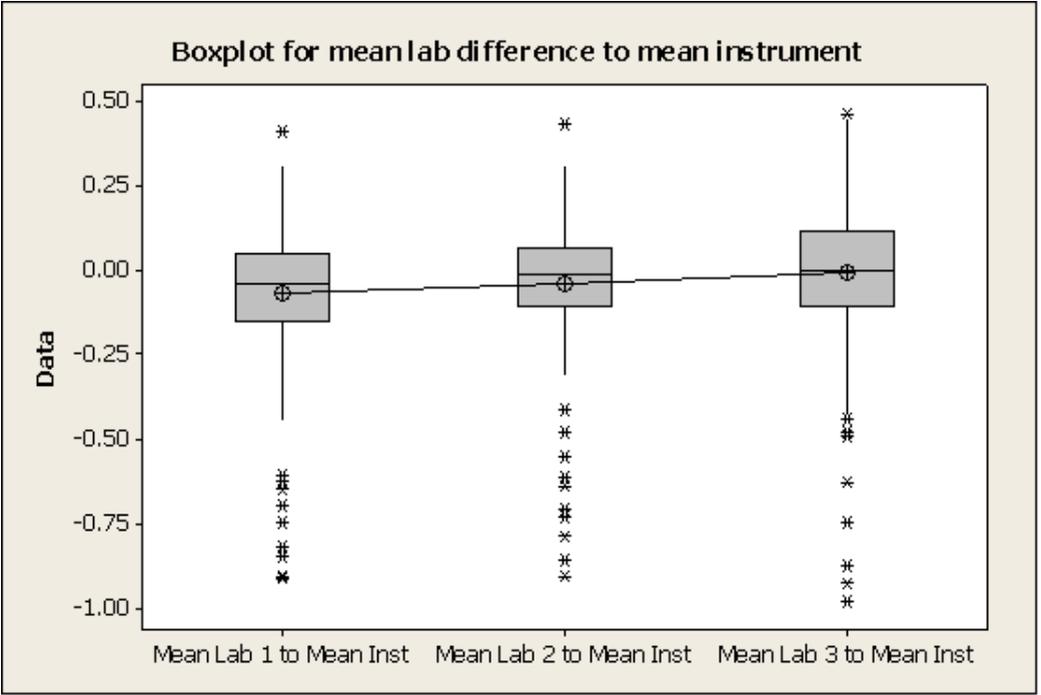


Figure 11: Box-and-whisker plot for the difference between the mean laboratory and the mean instrument readings. The joining line between the boxes indicates the mean for each lab.

where σ is the process standard deviation. When i increases, $(1 - \lambda)^{2i} \rightarrow 0$, therefore

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2 - \lambda} \right),$$

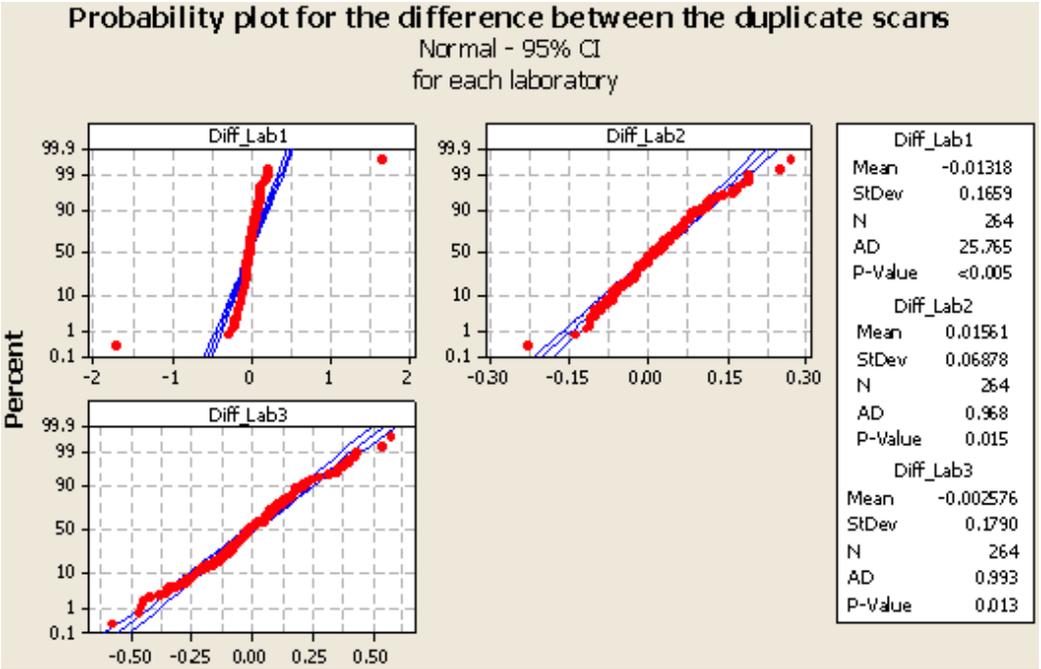


Figure 12: Probability plot of the reading errors for each individual Lab.

The control limits for this chart are as follows:

$$UCL = \mu_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2i}]}, \quad (3)$$

$$CENTER\ LINE = \mu_0, \quad (4)$$

$$LCL = \mu_0 - L\sigma\sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2i}]}. \quad (5)$$

If it is desired to monitor a small error, then the weight parameter should have a value between 0.1–0.2 (Montgomery 2005). All the analyse in this article are carried out using $\lambda = 0.2$ and $L = 2$ (the position of the control limits on the chart).

4.2.1 Using the EWMA chart for one-step ahead forecasting of the process mean

We discussed the EWMA primarily as a statistical process control tool; however, the EWMA provides a one-step ahead forecast of the process mean (Montgomery 2005). That is, z_t is a forecast value of the process mean μ at time $t + 1$. Thus, if the forecast of the mean is different from the required target by a critical amount, then either the operator or some electro-mechanical control system can make the necessary process adjustment. One suggestion for Fonterra is to monitor the labs' errors by deploying the EWMA and use its forecasting ability to decide if the process adjustment is required. The EWMA for Lab 1 errors is given in Figure 13. Figure 13 shows that samples 15, 16, 17, 18, 40, 41 have error much larger than expected in this process. The alternative interpretation (if these were plotted on line) is the mean error for sample 42 (one-step ahead forecast) would be high unless adjustment action is taken.

Also observe that the number of out-of-control errors on the EWMA chart is larger than on the \bar{X} chart (Figure 5). This is due to the \bar{X} chart detecting errors greater than $1.5\sigma/\sqrt{n}$; however, the EWMA with $\lambda = 0.2$ is optimal to detect errors less than $0.5\sigma/\sqrt{n}$.

4.2.2 Comparisons of the mean errors for the three labs using ANOVA

Analyses of Variance (ANOVA) together with Tukey's test are performed to compare the performance of the three labs. The following ANOVA and Tukey's test on the errors of the three labs with the p-value of 0.074, indicate that at 95% confidence level the mean errors for the individual labs are not significantly different. The results also show that the mean errors for Lab 3 are much closer to zero than the other two labs. However, the smallest standard deviation was recorded for Lab 2. This indicates that the error readings in

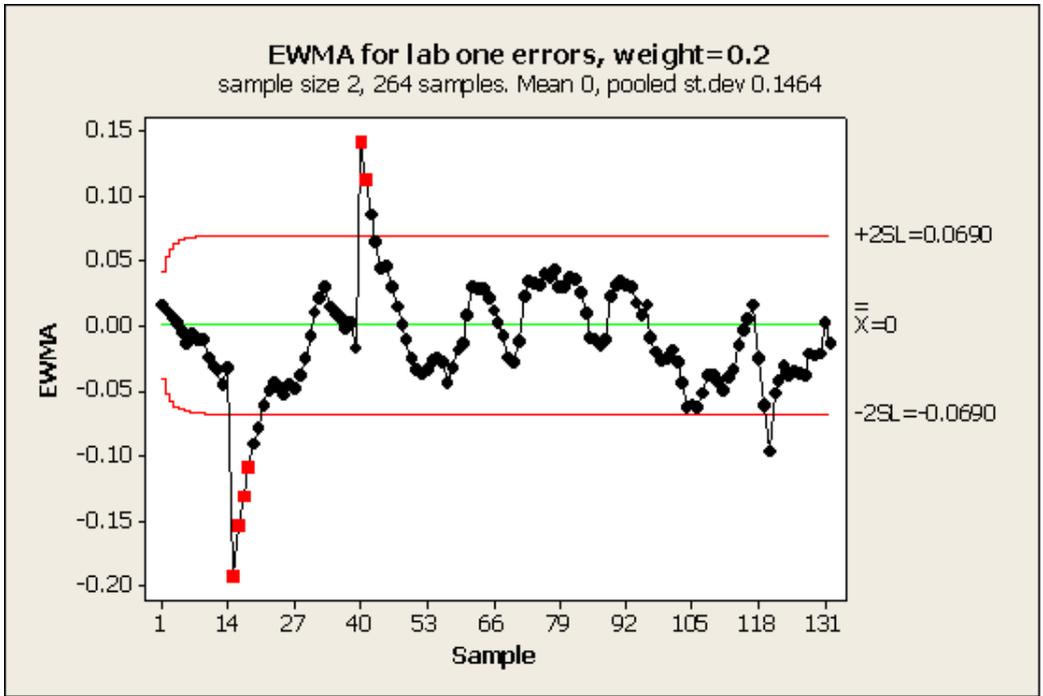


Figure 13: EWMA chart for Lab 1 error.

Lab 2 over all the 264 samples are more uniform than the error readings of the other two labs.

4.3 Multivariate control charts

Multivariate process control techniques were established by [Hotelling \(1947\)](#) in his pioneering article. He introduced the problem of correlation between the quality characteristics of a process and came up with the well-known T^2 statistic to identify whether the whole process is out-of-control. The Hotelling's T^2 statistic is the optimal test statistic for detecting a general shift in the process mean vector for an individual multivariate observation.

Table 1: First part of one-way ANOVA: Diff_Lab1_1, Diff_Lab2_1, Diff_Lab3_1

Source	DF	SS	MS	F	P
Factor	2	0.1119	0.0560	2.61	0.074
Error	789	16.9115	0.0214		
Total	791	17.0234			

S = 0.1464 R-Sq = 0.66% R-Sq(adj) = 0.41%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
Diff_Lab1_1	264	-0.0132	0.1659
Diff_Lab2_1	264	0.0156	0.0688
Diff_Lab3_1	264	-0.0026	0.1790

-0.020 0.000 0.020 0.040

Pooled StDev = 0.1464

However, the technique has several practical drawbacks. One of the most important ones is that when the T^2 statistic indicates that a process is out-of-control, it does not provide information on which variable or set of variables is out-of-control. Moreover, it is difficult to distinguish location shifts from scale shifts since the T^2 statistic is sensitive to both types of process changes.

The difficulty of interpreting an out-of-control signal on a multivariate control chart has been discussed extensively by Alt (1985), Doganaksoy et al. (1991), Murphy (1987), Pignatiello & Runger (1990), Lowry et al. (1992), Hunter (1986), amongst others. When two or more correlated variables are monitored, use of a multivariate chart may cause signals at opposing times to the signals given by a set of univariate charts on the individual variables. This is because the control region for a multivariate chart on correlated variables is represented by a tilted elliptical region as opposed to the non-tilted square region obtained by the use of separate charts. The use of separate charts does not allow for the information concerning the correlation of the variables to be utilized. However, the combination of using a multivariate control chart for signalling purposes and then using separate charts for diagnostic purposes is often effective.

To simultaneously monitor p correlated quality characteristics, the multivariate EWMA (MEWMA) chart is also be applied. The MEWMA was introduced by Lowry et al. (1992) and is a logical extension of the univariate MEWMA. The statistic calculated in the MEWMA is defined as

$$\mathbf{Z}_i = \lambda \mathbf{x}_i + (\mathbf{I} - \lambda) \mathbf{Z}_{i-1}, \quad \mathbf{Z}_0 = \mathbf{0},$$

where $\lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, $0 \leq \lambda_i \leq 1$ and \mathbf{I} is the identity matrix. If there is no a priori reason to weight past observations differently for the p quality characteristics, then $\lambda_1 = \lambda_2 = \dots = \lambda_n$. The plotted statistic on the control chart is

$$T_i^2 = \mathbf{Z}_i' (\boldsymbol{\Sigma}_{\mathbf{Z}_i})^{-1} \mathbf{Z}_i,$$

and the covariance matrix is

$$\boldsymbol{\Sigma}_{\mathbf{Z}_i} = \frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2i}] \boldsymbol{\Sigma},$$

where Σ is the process variance. The upper control limit in this article is produced by the statistical package Minitab 16 based on Average Run Length (ARL) = 200 and number of quality characteristics to be examined. Analogous to the situation in the univariate case, the MEWMA chart is equivalent to T^2 (multivariate \bar{X}) chart if $\lambda = \mathbf{I}$.

Multivariate charts examine all of the correlated relevant quality characteristics such as fat, moisture and protein or the measurements from several labs simultaneously. Figure 14 represents the MEWMA chart for the protein measurements, where for each specimen the sample size is two (one reading from each test in each individual lab) and data from individual labs is treated as different characteristics. Therefore, we have 264 samples of size two from three different labs ($\mathbf{p} = 3$). Here the MEWMA chart is used to detect a number of problems, including transcription errors and out-of-sequence samples produced by individual labs.

The MEWMA in Figure 14 is capable of monitoring all the readings from all three labs simultaneously and identifies which samples are out-of-control and which test in which lab was the cause of the out-of-control signals. The target mean vector and the covariance matrix are estimated based on all the recorded data.

It is a common practice in statistical quality control to use the individual charts to identify the characteristic(s) responsible for the multivariate out-of-control signals. For this reason, the univariate EWMA chart is imposed on each individual quality characteristic by the statistical package Minitab 16. The sources of the out-of-control signal together with their corresponding p -values are presented in Table 3.

The results in Table 3, for example, indicate that the reason for sample 249 being out-of-control is the high first reading of Lab 1 and second reading of Lab 2. Similarly for sample 216 the out-of-control signal was caused by the first reading from Lab 1 and the second reading from Lab 3.

Furthermore, using the one-step ahead forecasting ability of the MEWMA, the

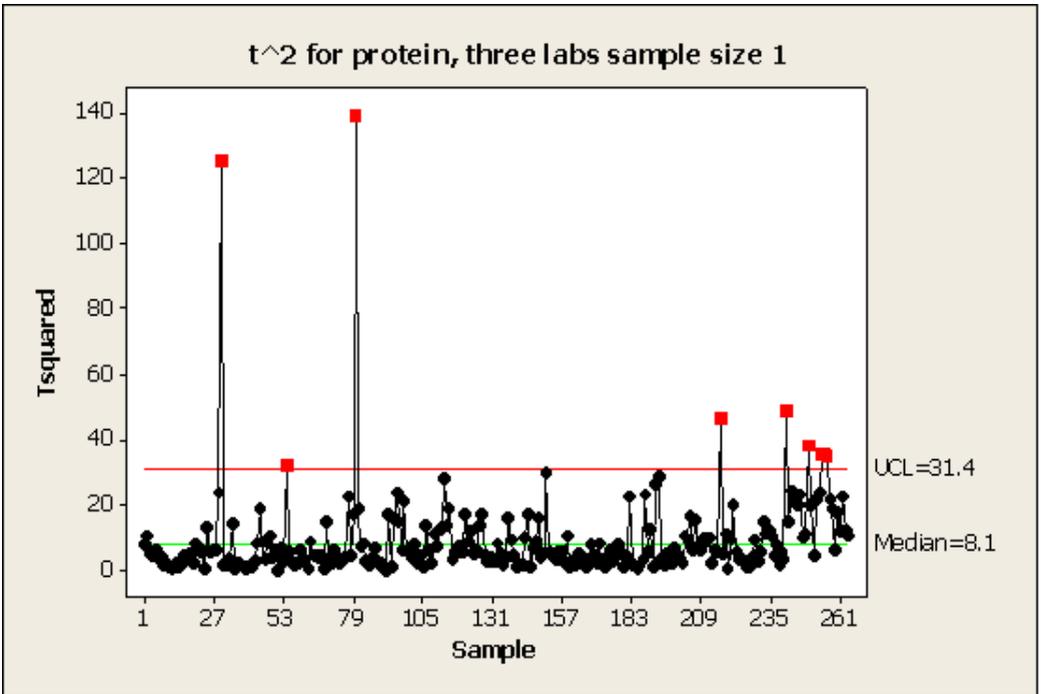


Figure 14: MEWMA chart for the protein measurements of the three labs.

above results warn the producer that the expected multivariate mean vector for, say, sample 250 would be high and the cause would be Lab 1 and Lab 2.

5 Derivative spectroscopy (an aside)

although the focus of this manuscript has been the monitoring of the rapid NIR assessment of product compositions using control charts in terms of assessing instrument health, a key monitoring role is played by derivative spectroscopy because second derivatives of the spectra removes the linear trends due to the scatter effect. Derivative spectroscopy also plays a key background role

Table 3: Test results for T^2 chart, greater than UCL, of Lab1_a, ... Lab3_b

Point	Variable	p-value
30	Lab 1b	0.0000
54	Lab 3a	0.0000
80	Lab 1b	0.0000
	Lab 2a	0.0047
	Lab 3b	0.0195
216	Lab 1a	0.0001
	Lab 3b	0.0002
241	Lab 1a	0.0007
	Lab 3a	0.0003
249	Lab 1a	0.0002
	Lab 2b	0.0355
254	Lab 1a	0.0000
	Lab 2a	0.0012
255	Lab 1a	0.0000
	Lab 2a	0.0134
	Lab 2b	0.0146
256	Lab 1a	0.0000
	Lab 2a	0.0000

related to the calibration and prediction of product composition.

Identifying the peaks in spectral data that correspond to key components in a biological material (e.g., milk powder, wheat) can be quite difficult. However, the spectral data shown in Figure 3 can be enhanced by a variety of methods, in order to highlight the locations of peaks that correspond to certain components more clearly; several techniques can be applied to give this enhancement, including derivative spectroscopy (Osborne et al. 1993, Anderssen, Hegland & Wesley 2011). The enhanced spectra can then be used to form the matrix \mathbf{S} , with the hope of increasing the accuracy of the calibration step, and hence the measured property values. A brief overview is

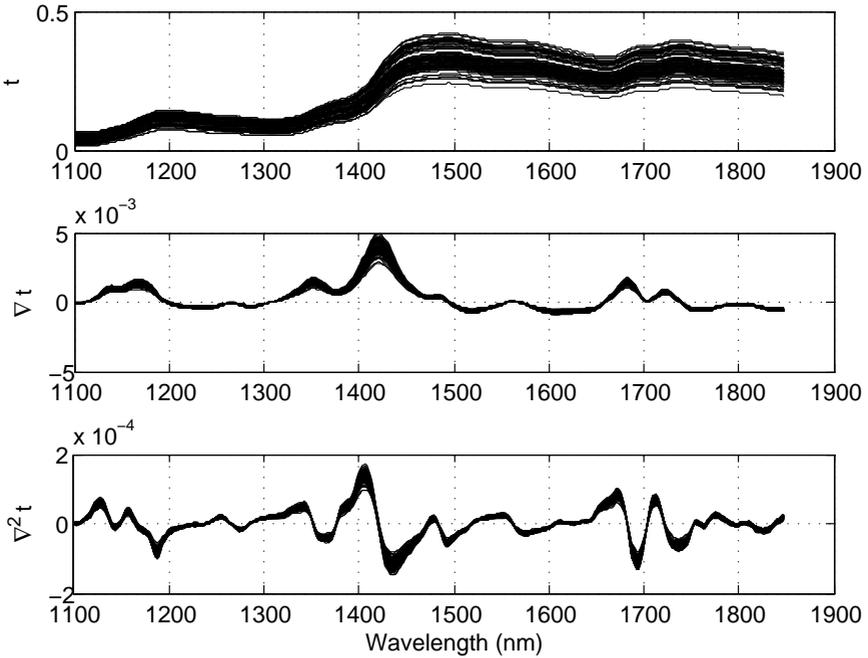


Figure 15: Skim milk powder spectrum for different 100 samples, with the first and second derivatives shown.

given below, taken from [Anderssen, de Hoog & Wesley \(2011\)](#), [Anderssen, Hegland & Wesley \(2011\)](#), [Anderssen et al. \(1998\)](#). A simple example is constructed by considering a spectrum

$$f(x) = \sin(x) + \epsilon \cos(\omega x),$$

where ϵ and ω are constants. By taking successive derivatives of this function, any frequency component that might be hidden (given a large ω and small ϵ),

becomes more prominent, as

$$\frac{d^4f}{dx^4} = \sin(x) + \epsilon\omega^4 \cos(\omega x).$$

Taking the first derivative of the spectrum removes any linear trends in the data, which may often be caused by measurement errors. Other advantages associated with derivative spectroscopy are outlined by [Anderssen, Hegland & Wesley \(2011\)](#).

Derivative spectroscopy clearly relies on the ability to differentiate a spectrum numerically. While in general numerical differentiation is potentially unstable, the information from an NIR machine is generated from an average of about 200 scans over very small wavelength differences and is therefore very smooth. Nevertheless, suitable smoothing techniques are applied to ensure accurate and stable estimates of derivatives of the data. A moving average technique ([Anderssen et al. 1998](#)) was used to differentiate stably. For example, the first derivative $\hat{f}_i^{(1)}$ at a node point x_i is

$$\hat{f}_i^{(1)} = \sum_{j=1}^J w_j f_i^{(1)}[j],$$

where w_j are the weights used for each derivative (which must sum to unity), and the derivative values are given by a standard difference formula, for example,

$$f_i^{(1)}[j] = \frac{f_{i+j} - f_{i-j}}{2jh},$$

where $h = x_{j+1} - x_j$. This algorithm takes linear combinations of J n th derivatives $f_i^{(n)}$ formed at the node point x_i . For each of the J derivatives, the values are computed by only using every J th node point. The method effectively involves computing a series of derivatives using different grid spacing, and taking some weighted average to give a single, stable derivative at the node point x_i .

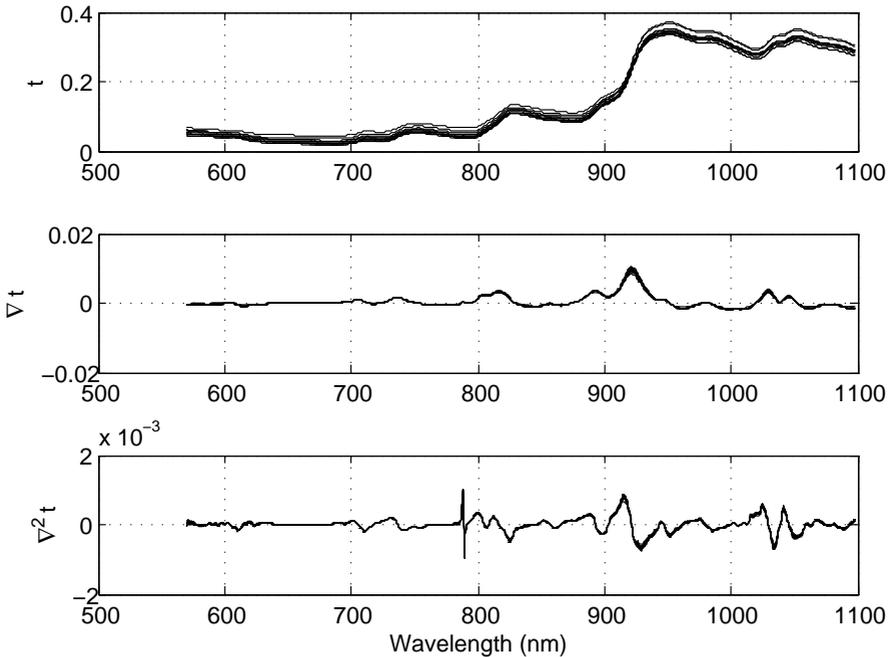


Figure 16: Skim milk powder spectrum for six samples, each scanned twice, before a lamp failure.

Derivative spectroscopy was performed in this manner, using the example spectrum shown in Figure 3, and the results are shown in Figure 15 (the original spectra are also plotted). Figure 15 shows that the peaks in the data are far more visible in the plots of the first and second derivative, when compared with the original spectrum. The peak around 1445 nm for example, represents the levels of moisture in the samples. The enhanced peaks from the derivative spectroscopy are used to improve the accuracy of the calibrations mentioned previously. Fonterra is currently using first derivative data during this process.

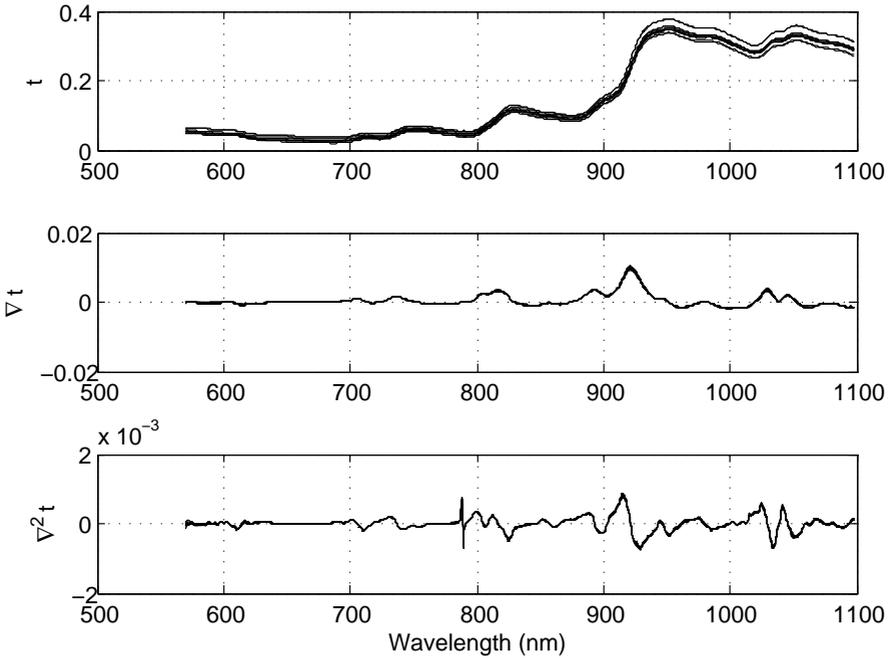


Figure 17: Skim milk powder spectrum for six samples, each scanned twice, after a lamp failure.

The health of the IR instruments is also important, as the lamps (which are the source of the NIR light) periodically fail, and the question was raised at the Study Group about the difference a failing lamp might make in the measured spectrum. Fonterra provided data, plotted in Figures 16 and 17, which shows the spectrum from six samples, each scanned twice, before a lamp failure, and after the lamp failed and was replaced, respectively. Also plotted are the values of the first and second derivatives, calculated as described above. There is a small difference between the spectrum before and after the lamp failure. However, this difference is approximately the same size as

the difference between the duplicate scans taken of the same sample, both before and after the lamp failure. This size implies that variations in the sample between scans, caused by effects like different packing in the sample dish between scans, are more important in terms of the variability seen in the spectrum than a failing lamp reading an incorrect spectrum.

The difference between the mean spectra (the average of duplicate scans) and the individual spectra observed for the six samples is plotted in Figure 18. The differences are roughly centred around zero; the maximum relative magnitude of the difference is approximately 2.8% in the absorbance detected. This gives a rough measure of the maximum difference that can be expected in the spectrum, given the worst-case scenario of a failing lamp, combined with variation in the testing protocols.

6 Recommendations and conclusions

The outcomes of the above investigations resulted in the Study Group making the following recommendations:

- Use control charts to monitor the reference method, with respect to both in-lab variation and between-lab variation (currently not done).
- Monitor NIR results compared with those from the labs (modified).
- Set the control limits in the control charts for the NIR results using the sum of variances from the NIR and lab results.
- Use a Type II error to determine the sample size (n) that should be measured during the NIR machine monitoring process and the position of the control limits in the control charts.
- Use Multivariate MEWMA charts to monitor the performances of all the three labs simultaneously in terms of reading errors and in the lab fault diagnostics task.

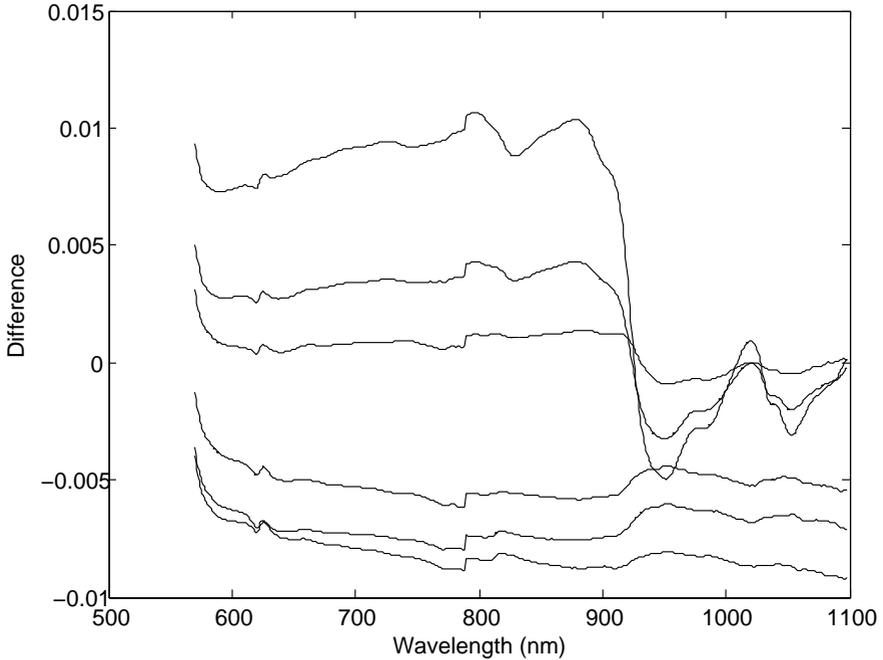


Figure 18: Skim milk powder spectrum for six samples, each scanned twice, after a lamp failure.

- Use Multivariate EWMA charts to monitor the performance of the NIR machines over the variety of the correlated product characteristics simultaneously.
- Use the one-step ahead forecasting ability of the univariate EWMA and Multivariate MEWMA charts to monitor the expected shift in the measurements reading for the next sample and the necessity for the machine adjustment.
- Investigate using quality assurance switching rules, based on the number

of consecutive accepted and rejected lots to switch between intensive and reduced testing.

- Use the mean instead of the median to monitor the differences between NIR and lab measurements, and also while comparing a particular instrument to mean instrument readings.
- Standardise the standard deviation calculation in control charts based on all the available historical data.
- Do not tweak the bias in machines based on monitoring.
- Start daily NIR testing for each machine with a check-cell; this provides a stable test of the machine's "health" before beginning sample testing for the day.

References

- Alt, F. (1985), *Multivariate quality control: Encyclopedia of statistical society*, Wiley: New York. [M70](#)
- Anderssen, R. S., de Hoog, F. R. & Hegland, M. (1998), 'A stable finite difference ansatz for higher order differentiation of non-exact data', *Bull. Aust. Math. Soc.* **58**, 223–232. [M74](#), [M75](#)
- Anderssen, R. S., de Hoog, F. R. & Wesley, I. J. (2011), Information recovery from near infrared data, in W. McLean & A. J. Roberts, eds, 'Proceedings of the 15th Biennial Computational Techniques and Applications Conference, CTAC-2010', Vol. 52 of *ANZIAM J.*, pp. C333–C348.
<http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/3909> [M53](#), [M74](#)
- Anderssen, R. S., Hegland, M. & Wesley, I. J. (2011), Resolution enhancement for infrared spectroscopic data, in 'MODSIM2011, 19th International Congress of Modelling and Simulation', pp. 371–377.

- <http://www.mssanz.org.au/modsim2011/A4/anderssen3.pdf> M73, M74, M75
- Doganaksoy, N., Faltin, F. & Tucker, W. (1991), 'Identification of out of control quality characteristics in a multivariate manufacturing environment', *Communications in Statistics—Theory and Methods* **20**, 2775–2790. M70
- Hotelling, H. (1947), *Multivariate quality control: Techniques of Statistical Analysis*, McGraw-Hill. M67
- Hunter, S. (1986), 'The exponentially weighted moving average', *Journal of Quality Technology* **18**, 203–210. M70
- Lowry, C., Woodall, W., Champ, C. & Rigdon, S. (1992), 'A multivariate exponentially weighted moving average control chart', *Technometrics* **34**, 46–53. M70
- Montgomery, D. (2005), *Introduction to Statistical Quality Control*, John Wiley & Sons. M65, M66
- Murphy, B. (1987), 'Selecting out of control variables with the T^2 multivariate quality control procedure.', *The Statistician* **36**, 571–583. M70
- Osborne, B., Fearn, T. & Hindle, P. (1993), *Practical NIR spectroscopy with applications in food and beverage analysis*, McGraw-Hill Series in Higher Mathematics, Longman Scientific and Technical, Harlow, UK. M53, M73
- Pignatiello, J. & Runger, G. (1990), 'Comparisons of multivariate cusum charts', *Journal of Quality Technology* **22**, 173–186. M70
- Worthham, A. & Ringer, L. (1971), 'Control via exponential smoothing', *The Logistics Review* **7**, 33–40. M61

Author addresses

1. **Steven Dargaville**, Mathematical Sciences, Queensland University of Technology, AUSTRALIA.
<mailto:dargaville.steven@gmail.com>
2. **Mali Abdollahian**, School of Mathematical & Geospatial Sciences, RMIT University, AUSTRALIA.
<mailto:mali.adbollahian@rmit.edu.au>
3. **Rob Crawford**, Fonterra, NEW ZEALAND.
<mailto:rob.crawford@fonterra.com>
4. **Joanne Simpson**, Fonterra, NEW ZEALAND.
<mailto:jo.simpson2@fonterra.com>
5. **Robert McKibbin**, Centre for Mathematical Modelling, Massey University, Auckland, NEW ZEALAND.
<mailto:r.mckibbin@massey.ac.nz>
6. **Bob Anderssen**, CSIRO, AUSTRALIA.
<mailto:bob.anderssen@csiro.au>