8th International IFAC Symposium on
Dynamics and Control of Process Systems
Preprints Vol.1, June 6-8, 2007, Cancún, Mexico

DETECTION AND EFFECT OF QUANTISATION IN DATA-DRIVEN PROCESS ANALYSIS

Margret Bauer [*] **Sethabile Madolo** [**]

[*] *Department of Electronic, Electrical and Computer Engineering, University of Pretoria, South Africa*
[**] *Anglo American, Anglo Technical Division, 45 Main Street, Marshalltown 2107, South Africa*

Abstract: Process data captured from the instrumentation of a plant is frequently sampled and quantised. Quantisation can result in poor quality data if the quantisation level is large. This paper investigates the effect of quantisation on data-driven analysis methods, in particular the effect on first order statistics, the power spectrum and on an oscillation detection method. Analysis shows that a large quantisation level distorts the results obtained from data-driven process analysis. Investigation shows that temperature measurements in particular are prone to quantisation due to low accuracy. An automated algorithm is presented that estimates the quantisation level from historical process data.

Keywords: Chemical process; Uniform quantisation; Fault detection; Process monitoring; Data storage; Frequency analysis; Mean; Variance; Oscillation detection

1. INTRODUCTION

In modern process instrumentation, analogue measurements are converted into digital form. The analogue-to-digital-conversion (ADC) comprises three steps as shown in Figure 1 adapted from (Bellan *et al.*, 1996), namely sampling, quantisation and coding. In the quantisation block, the stored sample $x[k]$ is translated into discrete, equidistant amplitude levels that are separated by quantisation level Δ. The discrete levels are then converted into binary code where the least significant bit represents the quantisation level. In a data historian, the discrete sampled, quantised and coded data $x_Q^c[k]$ is stored.

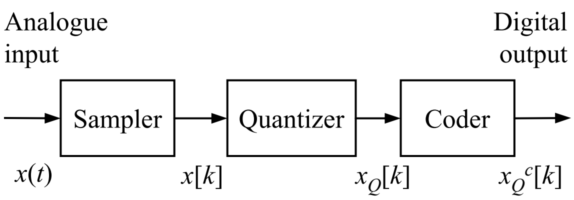

Fig. 1. Analogue-to-digital conversion process model.

Data-driven process analysis methods, such as time series or frequency analysis methods, use quantised data to gain insight into the process operation, to diagnose faults and to identify disturbances. If the quantisation level Δ is set too high then the quantised time trend $x_Q[k]$ will give a distorted representation of the measurement $x[k]$. Analysis of the quantised data might give misleading results. Large quantisation levels can originate from poorly adjusted sensors or sensors with a low accuracy. These causes are common even in modern data acquisition systems when, for example, initially correct sensor calibrations might be unsuitable after changes in process conditions.

Quantisation in the process measurement is also an outside disturbance. Another unwanted effect affecting the stored data and distorting the time trend is data compression (Watson *et al.*, 1998). Compression algorithms are applied to the coded data to save on storage cost but can result in a poor representation of the actual measurement. The impact of compression on data-driven analysis methods has been studied recently (Thornhill *et al.*, 2004; Singhal and Seborg, 2005). Although quantisation and its impact on analysis methods has been studied in detail (Blum, 1995; Mascarenhas *et al.*, 2000; Bellan *et al.*, 1996), there seems to be little attention to the effect of data quantisation on data-driven analysis methods for the process industries. This paper aims at closing the gap by investigating the impact of quantisation in the process analysis context. Furthermore, an automated detection method is proposed to estimate the quantisation level Δ from a time trend and gives a measure for the severity of the quantisation in the signal. If the data quality is too poor due to high quantisation, the data should be discarded and, in the long run, the sensor should be re-calibrated. Information lost in the sampling process cannot be restored.

This paper is organised as follows. First, the background of quantisation methods and developments is discussed. In Section 3, the impact of quantisation on data-driven methods is then reviewed and new insights are gained. An automated quantisation detection method is proposed and applied to industrial data in Section 4.

2. BACKGROUND

2.1 Applications of quantisation

Quantisation is an effect that occurs in almost all electronic and electrical application. The world we live in is analogue while all computational operations and electronic systems are conducted digitally. The information must be converted from analogue to digital in a quantiser. Even before the wide-spread introduction of computers, quantisation was essential for telephony and communication theory and the research on quantisation is dominated by those applications (Gray and Neuhoff, 1998). The main objective in communication theory is to develop a quantisation technique with the lowest loss of accuracy. The quality of the quantizer is measured by the accuracy of the resulting reproduction in comparison to the original.

In other engineering fields, an objective is to deal with and eliminate quantisation effects that have

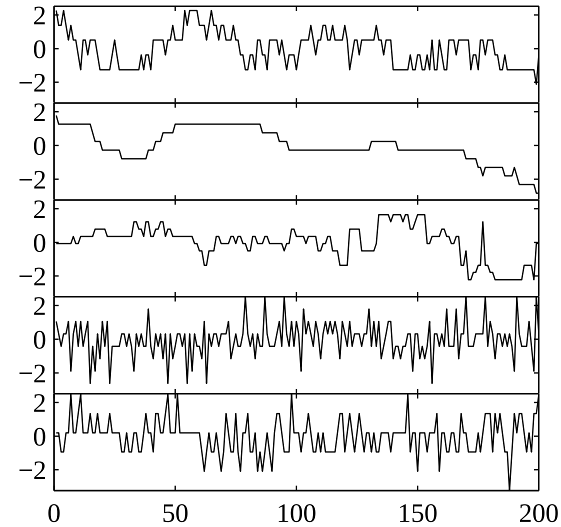

Fig. 2. Five industrial examples of quantised data in measurements (PV1-PV5 from top to bottom).

been introduced by analogue-to-digital conversion. A well researched area is the detection of objects in quantised two- and three-dimensional data for sonar and radar applications (Blum, 1995). In control theory, the presence of quantisation has been recognised since the introduction of computer control technology and many of the consequences, such as instability and round-off errors, have been studied (Böhm *et al.*, 1994). Far less attention has been paid to the effect of quantisation for time series analysis for the purpose of process analysis and monitoring.

2.2 Quantisation in process measurements

Industrial process data is captured from the sensors at the plant and usually quantised by the ADC in the sensor. If the quantisation level is set to high then the logged data will show strong indications of quantisation. Figure 2 shows five process measurements that exhibit clear quantisation signatures. These measurements stem from industrial processes at Anglo American and are normalized to zero mean and unit variance. Only discrete amplitude levels are adopted and step-like jumps between those levels can be observed. For example, in the first time trend PV1 only 6 discrete quantisation levels are adopted for all shown 200 time samples. The linearisation effect gives similar signatures to some compression algorithms (Thornhill *et al.*, 2004). Analysing quantised time trends as shown in sample time trends will result in the loss of information and in some instances will give misleading results as will be discussed in the following sections.

The frequency with which quantisation occurs in industrial data is investigated in a mini-case study. A total of 250 measurements from various

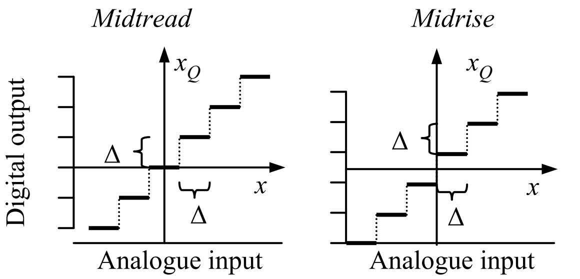

Fig. 3. Transfer functions of uniform midtread midrise quantisers.

plants and companies were inspected for quantisation. Table 1 shows the percentage of measurements that show the effect of quantisation in the time trends according to measurement type (Bauer, 2005). It can be observed that the temperature measurements are more prone to quantisation than any other measurement types. In the case study, 54% of the temperature measurements showed some extent of quantisation. In industrial processes, temperature is usually measured with digital thermometers and thermostats. The accuracy of these sensors lies in the region of $\pm 0.5^0$C (DalSemi, 2006). As the temperature is crucial for the reaction in many processes, it will vary only around a few degrees. The low accuracy of the temperature sensor thus results in quantisation.

3. IMPACT OF QUANTISATION ON DATA-DRIVEN METHODS

The most commonly used quantisers are uniform midtread and midrise quantisers which are both symmetrical around the origin. The transfer functions of those quantisers are shown in Figure 3. The transfer function of a uniform midtread quantisation is as follows.

$$x_Q = \Delta \left[\frac{x}{\Delta} \right] \tag{1}$$

where $[\cdot]$ represent rounding to the next integer and Δ is the quantisation level. The quantisation error is defined as the difference between the original and the quantised sample $e = x_Q - x$. The smaller the quantisation level Δ, the smaller the quantisation error e and the better the quantiser.

Temperature	54%
Pressure	23%
Flow	14%
Level	11%
Total	28%

Table 1. Percentage of quantised measurements in industrial data out of a total of 250 measurements.

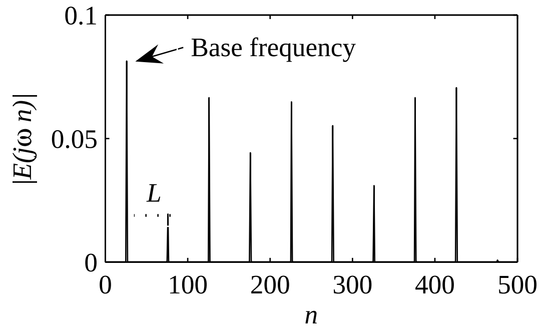

Fig. 4. Power spectrum of a quantised sine wave.

3.1 Mean and variance

The quantisation error e of a quantiser with quantisation level Δ is usually modelled by an additive, input independent noise uniformly distributed in $[-\Delta/2; +\Delta/2]$. The mean μ_e and variance σ_e^2 of the quantisation error can thus be derived from the uniform distribution.

$$\mu_e = \int_{-\Delta/2}^{+\Delta} \frac{x}{\Delta} \mathrm{d}x = 0; \; \sigma_e^2 = \int_{-\Delta/2}^{+\Delta} \frac{x^2}{\Delta} \mathrm{d}x = \frac{\Delta^2}{12} \tag{2}$$

Since the quantisation error is regarded as an additive noise source the mean value μ_x and the variance σ_x^2 of the unquantised data are approximated from the quantised data as follows (Carbone and Petri, 1998).

$$\mu_x \approx \mu_{x_Q}; \quad \sigma_x^2 \approx \sigma_{x_Q}^2 - \frac{\Delta^2}{12} \tag{3}$$

Thus, the larger the quantisation error, the larger the impact of the quantisation error on the variance σ_x^2. The approximation for the mean μ_x is only applicable if the quantisation level Δ is sufficiently small compared to the standard deviation of the input signal, that is, $\Delta \ll \sigma_x$ (Chiorboli, 2003). This is usually assumed for a quantisation level by a factor of ten smaller than the standard deviation. If this is not the case, then the quantisation depends strongly on the offset position of the quantiser which can vary between $-\Delta$ and $+\Delta$. It is therefore important to know the size of the quantisation level to decide whether the mean and variance can be estimated from Eq. 3 or if the time trend should be discarded.

3.2 Power spectrum

The power spectrum gives insight into the frequency contained in a time trend x and is obtained from absolute value of the Fourier transform $|X(j\omega n)|$. Here, n are the discrete frequency bins. The impact of quantisation on the power spectrum has been previously discussed in (Bellan *et al.*, 1996). Quantisation results in step-like features in the time trend. Steps in the time trend will be result in additional terms over all frequencies in the power spectrum. In particular, high frequencies which represent fast changes will be present in the spectrum after quantisation.

For example, a sinusoidal time trend $x_S[k] = A\sin(2\pi k/T)$ with amplitude A and oscillation period T and a midtread quantiser with quantisation level Δ are considered. Bellan *et al.* (1996) show that the quantisation error spectrum consists of periodic repetitions. Figure 4 shows the power spectrum of a sinusoidal time trend $x_S[k]$. The additional frequency peaks resulting from the quantisation are separated by a distance L:

$$L = 2\pi A/\Delta. \tag{4}$$

The smaller the quantisation level Δ, the fewer additional frequency peaks there will be in the power spectrum at larger distances L.

3.3 Oscillation detection

Common disturbances in the time trend that give particular rise for concern are oscillations. Oscillations can be caused by inappropriate tuning or by instrumentation problems. A real-time oscillation detection method presented by Hägglund (1995) and modified in (Thornhill and Hägglund, 1997) investigates the time between zero-crossings and thus determines whether a time trend exhibits oscillation. The main computation of the statistic is the calculation of the integrated absolute error (IAE) defined by the following expression:

$$IAE_i = \sum_{\kappa=\kappa_i}^{\kappa_{i+1}} |Y[\kappa]| \tag{5}$$

where $Y[\kappa]$ is the autocovariance of the controller error signal at time shift κ. Furthermore, κ_i and κ_{i+1} are the times of successive zero crossings of $Y[\kappa]$. The autocovariance is used instead of the time trend to remove high frequency noise effects. It measures the similarity of a signal with a time-shifted version of the same signal and can be derived from the process variable time trend x with N samples. An interval between two zero crossings is defined as $\Delta\kappa = \kappa_{i+1} - \kappa_i$. Regularity is assessed by the use of a statistic q which is defined as follows:

$$q = \frac{\hat{\mu}_R}{\hat{\sigma}_R} \tag{6}$$

where the ratio R between adjacent intervals k is as $R_i = \Delta\kappa_{i+1}/\Delta\kappa_i$ and from which the average value $\hat{\mu}_R$ as well as the standard deviation $\hat{\sigma}_R$ are estimated. For most time trends, the oscillation can be positively identified if the regularity index exceed the 3-σ threshold ($q > 3$). The oscillation

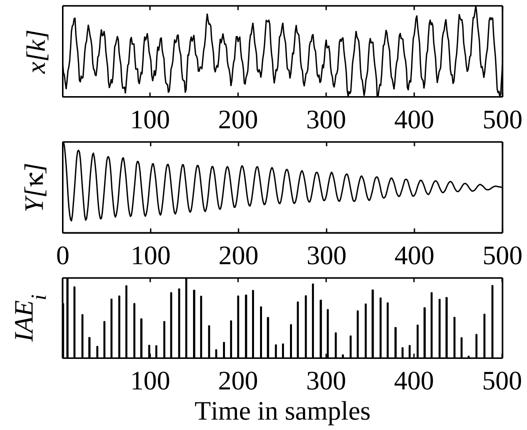

Fig. 5. Oscillatory time trend $x[k]$, autocorrelation function $Y[k]$ and integrated absolute error IAE_i for oscillation detection.

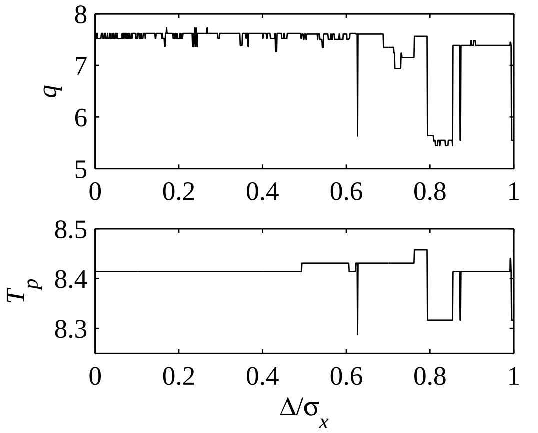

Fig. 6. Regularity q and period of oscillation T_p as function of quantisation level Δ.

period of the signal can be calculated as follows from the detected intervals between the zero-crossings $T_p = 2\hat{\mu}_{\Delta\kappa}$.

To test the impact of quantisation on the oscillation detection mechanism, an oscillating time trend of an industrial data set is quantised into increasingly larger intervals. Figure 5 shows the sample time trend $x[k]$, the autocorrelation function $Y[\kappa]$ and the integrated absolute error (IAE) between the zero crossings of $Y[\kappa]$. Since the peaks of the IAE appear in regular intervals, the regularity index exceeds the threshold $q = 7.5 > 3$. The period of oscillation is $T_p = 8.4$.

The impact of quantisation on oscillation detection is investigated by quantising the input sequence $x[k]$ by an increasing level Δ. The effect of quantisation on the regularity index q and the period of oscillation T_p is shown in Figure 6. Even for a large quantisation level Δ, the period of oscillation is detected quite accurately. The effect is noticed in the detected oscillation period T_p only if Δ exceeds $0.5\sigma_x$. The brief analysis shows that the oscillation index is robust to quantisation.

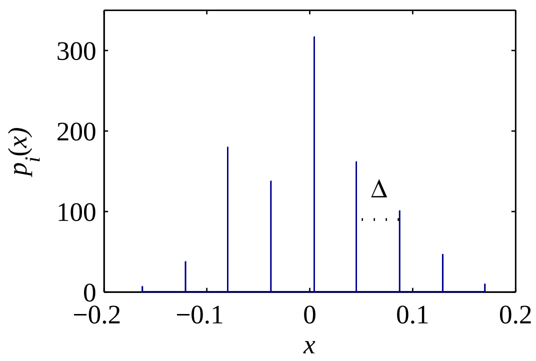

Fig. 7. Histogram of quantised data reveals quantisation level Δ (example for data trend PV1 from Figure 2).

4. AUTOMATED QUANTISATION DETECTION

Since the human sensory system is analogue it can easily differentiate between quantised and unquantised data. For example, visual inspection of the time trend in Figure 2 reveals the effect of quantisation. In this section, an automated algorithm is presented that identifies the quantisation level from historical data.

4.1 Algorithm

After quantisation the data samples adopt only discrete amplitude levels. The amplitude levels can be seen when constructing the histogram of the process variable from historical data. Histograms approximate the probability density function (PDF) of a time trend by counting the number of measurements that lie within a measurement interval. Histograms are also referred to as the relative frequency. Figure 7 shows the histogram of the first time trend from Figure 2. A large number of amplitude bins will give a better resolution but a more computational effort is required. In the example, the number of bins is set to 10,000.

The quantisation detection algorithm is similar to the oscillation detection method described in Section 3.3, only that the regularity analysis is applied to the histogram and not to the autocorrelation function. Assuming a uniform quantiser, the regularity of peaks in the histogram, as shown in Figure 7 is an indication of quantisation. An interval between two peaks is therefore defined as $\Delta x = x[i+1] - x[i]$. The regularity of the peaks can be expressed in a regularity index similar to Equation 6:

$$q_{\Delta x} = \frac{\hat{\mu}_{\Delta x}}{\hat{\sigma}_{\Delta x}} \tag{7}$$

were $\hat{\mu}_{\Delta x}$ and $\hat{\sigma}_{\Delta x}$ are the mean and standard deviation of the interval between two peaks in the histogram. A similar $3-\sigma$ detection threshold as used in the oscillation detection method can be applied, so that the signal is considered as quantised if $q_{\Delta x} > 3$. The quantisation level can be approximated by the average interval Δx:

$$\hat{\Delta} = \hat{\mu}_{\Delta x}. \tag{8}$$

A comparison of the detected quantisation level $\hat{\Delta}$ to the standard deviation σ_x of the data gives an indication of the severity of the quantisation. When estimating mean and variance of quantised data as discussed in Section 3.1, the quantisation level should be only a tenth of the standard deviation $\hat{\Delta} < 10\sigma_x$. In some industrial applications, the coder adds noise to the quantised data so that some samples will not fall on the discrete amplitude levels though the majority does. In this case, a threshold p_{th} can be introduced above which the peaks in the histogram are registered as x_i.

Variable	$q_{\Delta x}$	$\hat{\Delta}/\sigma_x$
PV1	76.1	0.88
PV2	10.1	0.36
PV3	74.8	0.35
PV4	87.5	0.38
PV5	9.2	0.64

Table 2. Regularity index and quantisation level of time trends from Figure 2.

4.2 Quantisation detection examples

The quantisation detection algorithm was applied to the normalised industrial data shown in Figure 2. The resulting regularity index q as well as the estimated quantisation level $\hat{\Delta}$ of those five time trends are listed in Table 2. For all five PVs the regularity index $q_{\Delta x}$ lies above the 3σ threshold and indicate quantisation or all PVs. PV1 exhibits the largest quantisation level of $\hat{\Delta} = 0.88\sigma_x$ but all time trends have a large quantisation level compared to the standard deviation of the time trend. The estimation of mean and variance of this signal should be considered according to the guidelines given in Section 3.1.

There are, however, several instances when the quantisation detection algorithm might not give the desired results. For example, a constant value resulting from a dead sensor will be problematic to identify. A test for zero variance should therefore be conducted first. Also, the data might be nearly constant when a time frame is selected with no or little variation. Pre-processing the data and selecting a representative time frame is therefore necessary. A further shortcoming of the algorithm is that it does not acknowledge that most process variables follow a Gaussian PDF, as for example the time trend of PV1 in Figure 7. The histogram peaks on the outer left and right side of the bell-shaped curve might not all be represented leaving gaps in the otherwise regular histogram. It is therefore advisable to focus on the dominant peaks by introducing a detection threshold. Only peaks above the threshold are considered for the regularity analysis.

In some instances, the quantisation is not caused by the analogue-to-digital conversion but is part of the actual process operation. For example, an electric pump might operate at certain speed levels. The flow rate after the pump will show only these discrete levels though the measurement itself might not be quantised. It will be impossible to tell from the time trend that a measurement is 'deliberately' quantised and expert knowledge has to be used to identify the difference between a step-wise function and a quantised time trend. Expert knowledge from the operator or process engineer rather than an automated algorithm is required to tell the difference between unwanted and real quantization.

5. CONCLUSIONS

In this paper, the effect of quantisation on data-driven process analysis methods has been discussed. Quantisation occurs in industrial process data due to inadequate sensor calibration or low accuracy in sensor. A mini case study of 250 industrial time trends showed that 28% of the measurements exhibit quantisation signatures. Mean and standard deviation of a time trend as well as its power spectrum are affected significantly by quantisation. A frequently used oscillation detection method, on the other hand, is robust when analysing quantised data.

A procedure was introduced for detecting the effect of quantisation in historical data. The procedure is similar to a popular oscillation detection method and establishes the quantisation level as well as a regularity index that gives an indication of the significance of the quantisation in the time trend. With the proposed method, process data can be tested for quantisation and a decision can be made whether to keep or discard the data for process analysis.

ACKNOWLEDGEMENTS

The principal author gratefully acknowledges the financial support of the Claude-Leon-Foundation and of the University of Pretoria Post-doctoral Fellowship Programme.

REFERENCES

Bauer, M. (2005). Data-driven methods for process analysis. PhD thesis. University College London.

Bellan, D., A. Brandolini and A. Gandelli (1996). Quantization effects in sampling processes. In: *IEEE Instrumentation and Measurement Technology Conference*. pp. 721–727.

Blum, R.S. (1995). Quantization in multisensor random signal detection. *IEEE Transactions on Information Theory* **41**, 204–215.

Böhm, B., J.A. Schoonees and J. Tapson (1994). Some non-linear effects of quantization in phase-locked loops. In: *Proceedings of COMSIG*. pp. 116–120.

Carbone, P. and D. Petri (1998). Mean value and variance of noisy quantized data. *Measurement* **23**, 131–144.

Chiorboli, G. (2003). Uncertainty of mean value and variance obtained from quantized data. *IEEE Transactions on Instrumentation and Measurment* **52**, 1273–1278.

DalSemi (2006). DS1620 Digital thermometer and thermostat. Technical Report 072099. Dallas Semiconductors, http://www.dalsemi.com.

Gray, R.M. and D.L. Neuhoff (1998). Quantization. *IEEE Transaction on Information Theory* **44**, 2325–2383.

Mascarenhas, N.D.A., L.V. Dutra and A.C. Frery (2000). The effect of quantization on SAR parameter estimation. In: *International Geoscience and Remote Sensing Symposium (IGARSS) 6*. pp. 2663–2665.

Singhal, A. and D.E. Seborg (2005). Effect of data compression on pattern matching in historical data. *Industrial Engineering Chemistry Research* **44**, 3203–3212.

Thornhill, N.F. and T. Hägglund (1997). Detection and diagnosis of oscillation in control loops. *Control Engineering Practice* **5**, 1343–1354.

Thornhill, N.F., M.A.A.S. Choudhury and S.L. Shah (2004). The impact of compression on data-driven process analysis. *Journal of Process Control* **14**, 389–398.

Watson, M.J., A. Liakopoulos, D. Brzakovic and C. Georgakis (1998). A practical assessment of process data compression techniques. *Industrial and Engineering Chemistry Research* **37**, 267–274.