Fractional Nonconformance Assessment

MASSEY
UNIVERSITY

Xin Zhou

Supervisors: Dr. K. Govindaraju
Prof. Geoff Jones

School of Fundamental Sciences
Massey University

This dissertation is submitted for the degree of
Doctor of Philosophy in Statistics

March 2020

I would like to dedicate this thesis to my loving parents.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Xin Zhou
March 2020

Acknowledgements

First, I would like to express my deepest gratitude to my supervisor Dr. K. Govindaraju for his guidance, patience, advice and support not only in study, but also in life. Thank you, Raj. I am extremely grateful to my co-supervisor Professor Geoff Jones, for his kind guidance and encouragement. Thanks to the members of the Statistics and Bioinformatics Group and friends at postgraduate office. I also would like to acknowledge all support from School of Fundamental Sciences, Massey University.

Secondly, I am sincerely grateful to the Primary Growth Partnership (PGP) scheme, which was sponsored by Fonterra Co-operative Group Limited and the New Zealand Government, for the financial support during my whole study. I would like to extend my appreciation to Roger Kissling and Steve Holroyd from Fonterra, for their valuable suggestions and support during the project.

Last but not the least, I am thankful to my parents for their support, understanding and encouragement during the past few years. I love you.

Abstract

Food quality and safety are important due to the very nature of the product and the potentially severe consequences of a fault in the manufacturing process. Statistical tools are widely used in food quality assurance. However, measurement error, which is inevitable in food manufacturing due to the variation and inaccuracy of the measurement system, affects the performance of statistical quality control activities. The concept *fractional nonconformance* was recently proposed to assess the probability of nonconformance for error-prone individual measurements. This thesis presents five pieces of work for fractional nonconformance assessment mainly suitable for food quality assurance and other applications. The new statistical methods developed pertain to control charting, acceptance sampling, and conformity testing areas. The application of the proposed methods is illustrated with real data from a leading New Zealand dairy product manufacturer. Interactive web-based *Shiny* apps providing step-by-step guidance to implement fractional nonconformance analytic tools are developed for practitioners.

Keywords

Acceptance sampling, Conformity testing, Control chart, Food manufacturing, Fractional nonconformance, Guardbanding, Measurement error, Short run, Statistical quality control.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Food manufacturing and quality assurance

Manufactured food products are required to be safe (e.g. absence of pathogens such as E. Coli) and hygienic (e.g. colony forming units (cfu) count below the set limit). Food products must also meet a large number of quality specifications (such as a minimum protein limit for infant milk powder formula). International and industry bodies provide standards and guidelines, such as the Codex Alimentarius (CAC [13, 14, 15, 16]), the International Commission on Microbiological Specification for Foods (ICMSF [74, 75, 76]) and the Institute of Food Science and Technology (IFST [77]) to help the food manufacturer to establish effective safety/quality assurance systems. Statistical tools are also widely used for quality assurance in food manufacturing. Hubbard [73] provided a comprehensive review of statistical quality control approaches for the food industry. A brief review of common quality control methods and issues encountered in food manufacturing processes follows.

1.1.1 Acceptance sampling

Acceptance sampling is a quality control technique used in industry to make decisions on the quality of lots formed from a production process. In some situations, 100% inspection can be used for lot disposition, where every item in the lot is inspected. Acceptance sampling plans are used in circumstances when 100% inspection is not feasible, for example, the testing is destructive or the costs associated with testing and measurement are too high etc. Acceptance sampling reduces the cost of inspection but it also introduces some risks. The producer's risk is the risk that the sampling plan rejects an acceptable lot and consumer's risk is the risk to accept a lot of rejectable quality. In other words, acceptance sampling achieves a trade-off between risks and costs. Acceptance sampling plans are designed for the specified producer's and consumer's risks. See Schilling and Neubauer [134] for a thorough introduction to acceptance sampling.

The performance of an acceptance sampling plan is evaluated by its Operating Characteristic (OC) curve. The OC curve shows the probability of acceptance for a given proportion nonconforming. The terms AQL and LQL refer to Acceptance Quality Limit and Limiting Quality Level respectively. The producer's risk, α, of a sampling plan corresponds to the risk that the sampling plan rejects the lots of AQL quality and the consumer's risk, β, is the probability of accepting a lot of LQL quality.

Acceptance sampling plans are often classified as *attribute* or *variables* plans. In attribute plans, the quality characteristics are determined on a pass/fail basis given a specification. In a single sampling plan, a random sample of n units is selected from a lot of size N units; the lot is accepted if there are c or fewer nonconforming units, where c is the acceptance number. If the observed number of nonconforming units are more than c, the lot is rejected. The attribute plans are easy to implement as they do not require the knowledge of the underlying distribution of the quality characteristic. See Wortham and Baker [165], Bennett et al. [11] and Hald [56] for examples of different types of attribute plans. In contrast, the quality characteristic is measured on a continuous scale for variables plans; see Lieberman and Resnikoff [98] and Balamurali and Jun [8]. The information obtained from the numerical measurements is greater, and hence a variables plan is more powerful than the attribute plan for the same sample size. An acceptance sampling plan can also be implemented as double/multiple/sequential plans, which can be designed to provide equivalent discrimination between good and poor quality lots.

Examples of acceptance sampling plans applied for food safety can be found in Whitaker [154], Gilbert [47], Starbird [144] and Santos-Fernández et al. [129]. Note that the purpose of acceptance sampling is lot disposition, rather than lot quality estimation. Statistical process control is used to systematically monitor and improve the quality of the process.

1.1.2 Statistical process control

No two leaves are exactly alike, so too the products from a manufacturing process. There are no exactly identical products due to the variations present in the manufacturing processes. Some of the variations are inevitable or natural, which are called *common* or chance causes of variation. These variations, for example, machining operations and temperature conditions, are inherent parts of the process. Certain other sources of variation, such as operator errors, machine failures and defective material, are avoidable and are called *special* or assignable causes of variation.

Statistical process control (SPC) aims to detect the occurrence of special causes and assure that the manufacturing process is free from special causes. A *control chart* is a powerful tool for process monitoring. Grigg [54] discussed the benefits of SPC in food manufacturing. The first sketch of a control chart can be traced back to 1924 in a memorandum from Shewhart [138]. In 1931, Shewhart [139] published the first control chart (which was named as Shewhart

chart later on) in his well-known text "Economic Control of Quality of Manufactured Product". In a control chart, the control statistic is plotted against the subgroup number, together with upper and/or lower control limits and a center line. Some control charts are named based on the control statistic employed. For example, $\bar{X}$ chart, which plots the sample mean, is used to control the process mean and R (sample range) or s (standard deviation) chart is used to monitor the process variation.

Besides the Shewhart charts, exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) charts are also used in practice. The EWMA chart was first introduced by Roberts [126] in 1959 and the CUSUM chart was proposed by Page [116] in 1954. According to Montgomery [111], the performances of the EWMA and CUSUM charts are very similar. The properties of these two charts have been studied by many authors; see Ewan [39], Woodall [160], Hawkins and Wu [61], Lucas and Saccucci [102] and Capizzi and Masarotto [17]. In brief, the EWMA and CUSUM statistics accumulate information selectively from past observations and hence the EWMA and CUSUM charts are more sensitive to detect small shifts compared to the Shewhart chart. Some researchers, such as Hawkins and Olwell [60] and Woodall [161] recommended to combine Shewhart and EWMA or CUSUM charts to achieve faster detection of both small and large shifts.

Control charting is often comprised of two phases: the retrospective analysis phase (Phase I) and the process monitoring phase (Phase II). Woodall and Montgomery [163] highlighted the importance to distinguish between these two phases. Also see Vining [150] for a discussion of Phase I and Phase II control charting. In Phase I, past data from the process are examined to check whether or not the process was in *statistical control*. Montgomery [111] suggested the use of 25-30 subgroups of size five for Phase I to correctly estimate the process parameters and establish control limits. The control limits are usually determined for a desired false alarm rate or Type I error. Establishment of correct control limits is critical. Control charts with misplaced control limits cannot quickly trigger a signal when the process is out-of-control or lead to a higher false alarm rate for an in-control process. After elimination of special causes of variation in Phase I, the process is monitored in Phase II for a state of control.

One of the most popular statistical approaches to evaluate the performance of a control chart is to examine its run length distribution. The average run length (ARL) is the expected number of samples taken until an out-of-control signal is obtained. When successive values of the plotted control statistic are independently and identically distributed, ARL is calculated as $\frac{1}{\epsilon}$, where ϵ is the probability of a single point breaching the control limits. ARL has been widely used in control chart literature; see Crowder [29], Kemp [90], Goldsmith and Whitfield [48], Reynolds [125], Gan [45] and many others. Govindaraju [50] pointed out the limitation of ARL and introduced 'unity' ARL, borrowing the methodology from acceptance sampling. Besides ARL, median run length, other percentage points of the run length distribution and standard deviation of run length (SDRL) are used as supplementary measures to evaluate the performance of control charts. Jones et al. [86] compared the ARL and SDRL for various

EWMA charts with estimated parameters. Jones et al. [87] provided tables of ARL and SDRL for CUSUM schemes.

Short run production has become increasingly widespread as consumer requirements have diverged. Food manufacturing demands frequent stoppage and cleaning of machinery for safety and hygienic reasons; hence food manufacturing processes are often short-run. Phase I is not available in a short run process. As a result, the traditional control charting procedures cannot be applied. Quality improvement becomes harder for short run processes because low fraction nonconforming level changes are difficult to detect with very limited sample sizes. See Elam [36] for an introduction to short-run process control.

Self-starting control charts were initially developed by Hawkins [59] to monitor start-up processes. The running mean and standard deviation are dynamically employed to estimate the process parameters. Quesenberry [122, 123, 124] introduced the Q statistic and applied Q charts to monitor the process mean and variance for short-runs. The Q statistic is a standardized individual measurement, calculated from the first observation of the process and then updated once a new observation is obtained. The statistical properties of Q charts have been studied by many authors. For example, Castillo and Montgomery [19] pointed out that the ARL performance of Q charts is sometimes inefficient and proposed to use EWMA charts and adaptive Kalman filtering when the process target is known and process SD is unknown. He et al. [64] further investigated the ARL performance of Q charts and introduced two alternative Q charts. Zantek [172] studied run length distributions of Q charts and illustrated the importance of reacting immediately to out-of-control signals. Zhang et al. [174] proposed t and EWMA t charts to monitor the process mean when the in-control process standard deviation is unstable or miscalculated, and these charts have been implemented for short run monitoring by Celano et al. [21]. Celano et al. [20] further investigated statistical performances of t charts in the form of Shewhart, EWMA and CUSUM schemes when the shift size is unknown for short run processes. Besides Q and t charts, other kinds of self-starting control charts have also been developed; see Zhang et al. [173], Li et al. [97] and Liu et al. [101].

1.1.3 Measurement error

According to ISO 5725-1 International Standard [80], measurement error is the difference between a measured quantity and its true value. It includes *random* and *systematic* errors. Random errors occur naturally resulting in inconsistent repeated measurements. Systematic errors are the errors, such as bias, due to the inaccuracy of the system. Random errors come from unpredictable fluctuations occurring in the measurement system and these errors cannot be eliminated, while systematic errors can be identified and reduced. Measurement error occurs in food manufacturing due to the variation and inaccuracy of analytical testing.

Measurement errors affect the performance of both acceptance sampling and statistical process control procedures.

When measurement errors are present, an acceptance sampling plan can be adjusted to correct for the error under both attribute and variables plans. Measurement errors are often termed as *inspection error* when misclassification of a conforming item as nonconforming (Type I inspection error) occurs and vice versa (Type II inspection error). It was established long ago by Lavin [96] that Type I inspection error generally increases the producer's risks more than Type II inspection error. Inspection errors occur due to human error, instrument error, or any other measurement process related errors. A thorough introduction to inspection errors in attribute sampling plans can be found in Johnson et al. [85]. Inspection errors can also be corrected by fuzzy set theory. Hryniewicz [70] provided a good review, and illustrated the application of fuzzy set theory in statistical quality control. Jamkhaneh et al. [83] designed a single sampling plan with inspection errors and investigated their effects on a plan with fuzzy parameters. Govindaraju and Jones [51] pointed out the limitations of the fuzzy approach to the design of acceptance sampling plans, and proposed a new fractional acceptance number sampling plan.

For variables sampling plan, the measurement error can be corrected using both parametric and non-parametric approaches. The single sampling plans by variables have been studied by many authors under the assumption that the observations and measurement errors are normally distributed; see Basnet and Case [9], Fang and Zhang [40], Wilrich [159]. Owen and Chou [115] studied the effect of measurement error and instrument bias on the operating characteristics of variables sampling plans. Mei et al. [108] investigated the effect of bias and imprecision on variables plans and suggested to increase the sample size to mitigate the effect. However, the parametric approach is not powerful for non-normal variables sampling plans because of the difficulty in deconvolution of the error and true distributions. Employing the methodology of density deconvolution using weighted kernel estimators of Hazelton and Turlach [62, 63], the estimation of the true distribution of measurements becomes possible. Wu and Govindaraju [166] developed computer-aided variables sampling inspection plans for compositional proportions and measurement error adjustment. This approach is useful for food inspection because many quality characteristics in the food industry, such as percent sugar or fat, are compositional proportions and intrinsically non-normal.

Measurement error effects on the control chart have been studied by many authors. See a recent thorough review by Maleki et al. [103]. In brief, measurement error may result in the loss of power in detecting parameter shifts in the underlying process variable and hence multiple (repeated) measurements are suggested to reduce the effect of measurement error on the chart's performance. Bennett [10] examined the effect of measurement errors on the Shewhart chart using the simple measurement error model $Y = X + Z$, where Y is the observed measurement, X is the true measurement, and Z is the measurement error. Kanazuka [89] and Mittag and Stemann [109] investigated the effect of measurement errors

on the power of joint Shewhart charts using the same error model. Linna and Woodall [99], Maravelakis et al. [105] and Maravelakis [104] investigated the effect of measurement errors on the Shewhart, EWMA and CUSUM chart respectively using a linear covariate model $Y = A + BX + Z$, where A and B are known constants. Hu et al. [71] studied the performance of synthetic $\bar{X}$ chart using the same error model. Noorossana and Zerehsaz [114] investigated the effect of this measurement error model on the most commonly used control charts for monitoring simple linear profiles with random explanatory variable. Costa and Castagliola [28] studied the effect of measurement errors and autocorrelation (separately and combined) on the Shewhart chart. Linna and Woodall [99] addressed the effect of measurement error on the monitoring of multivariate control charts. Hu et al. [72] proposed economic design of one-sided synthetic S^2 chart with measurement errors by considering a quadratic Taguchi loss function. Abbasi [2] investigated the performance of the EWMA mean chart in the presence of two-component measurement errors. Tran et al. [147] studied the performance of the Shewhart-RZ control chart with measurement errors. Cheng and Wang [23, 24] analyzed the performance of the EWMA median, CUSUM median and VSSI median control charts in the presence of measurement error. Zaidi et al. [171] examined the effect of measurement error on control charts for compositional data. No work has focused on the effect of measurement error for short runs in literature, except Farnum [41], who adjusted for measurement errors using the deviation from nominal/target (DNOM) data.

1.1.4 Conformity assessment

Conformity testing, also known as conformity assessment or compliance testing, is employed to assure that an entity meets a specific requirement and/or regulatory standard. The entity could be a process, product, service or personnel measure etc. Examples of conformity testing are provided in the ISO 10576-1 International Standard [79] (Guidelines for the evaluation of conformity with specified requirement). For example, a conformity testing procedure can be used to test of the concentration of some trace elements in the blood of employees for health evaluation, or the analysis of an athlete's urine to detect abuse of xenobiotic anabolic steroids for doping screening. In bulk product manufacturing, conformity testing ensures that the quality characteristic falls within the permissible specification range. According to ISO [78], the conformity testing procedure provides benefits for manufactures, consumers, government regulators as well as for international trade.

Acceptance sampling procedures are related to conformity testing; however, the principles of the two activities are not identical. Acceptance sampling collects a limited sample to determine whether to accept or reject a batch of material (population). In contrast, the scope of conformity testing is limited to an inference only on the sample tested. In other words, the conformity testing procedure is a decision making procedure on the particular sample (rather than the population) and is only concerned with the risk that the measured sample values

conform or not. Moreover, the proportion nonconforming is the main quality measure of interest in acceptance sampling but this is not the case in the conformity testing procedure.

The specification for the quantifiable characteristic is called as limiting value (LV) in the conformity testing protocols. The interval containing all permissible values of the characteristic is called the *region of permissible values*. A conformity testing protocol provides assurance of conformity by checking whether the measurement of interest falls within the region of permissible values or not. Measurement and sampling uncertainties have a crucial impact on conformity testing. According to ISO/IEC Guide 98-3:2008 [81], measurement uncertainty (MU) must be clearly indicated when comparing measurement results with reference values. The practice in conformity testing procedures is to compare the MU interval, usually given in the form of a confidence interval, with the region of permissible values; see Pendrill [119], King [93] and Williams and Hawkins [157] etc. As illustrated in Figure 1.1, the rules for asserting conformity or non-conformity defined in ISO 10576-1 International Standard [79] are:

- Assurance of conformity: The uncertainty interval is inside the region of permissible values. (See for instance, case A)
- Assurance of non-conformity: The uncertainty interval is included in the region of non-permissible values. (See case D)
- Inconclusive result: The uncertainty interval includes LV. (Cases B and C)

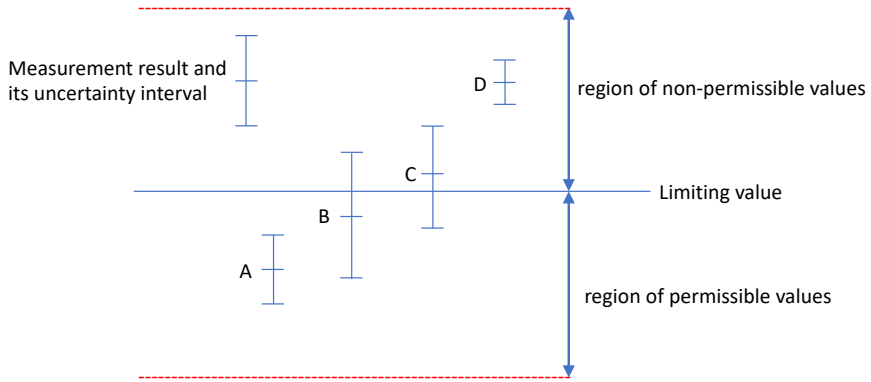


Fig. 1.1 ISO 10576-1 rules for asserting conformity and non-conformity.

When the MU interval includes LV, Hibbert [66] proposed to make a statement on the probability of compliance. Holst et al. [69] suggested a two-stage conformity testing procedure to increase the probability of declaring conformity for those entities with permissible values of the quantity of interest close to the LV and claimed that the probability of obtaining an inconclusive result is reduced for the two-stage procedure. See Desimoni and Brunetti [31] for a thorough review of measurement uncertainty and conformity testing.

1.1.5 Risk management

A probabilistic approach to risk assessment is largely supported by risk analysts compared to using uncertainty intervals or ratios such as relative risk etc. See the discussion in Aven [7]. In acceptance sampling, probabilistic assessment of plan performance plays a dominant role. Risks, including producer's and consumer's risks, arise not only due to sampling but also due to measurement error, human error and operation error etc. In acceptance sampling, the producer's risk is the probability that an acceptable quality product is rejected and the consumer's risk is the probability that poor or limiting quality product is accepted. In statistical process control, the false alarm rate leads to producer's risk, and consumer's risk occurs when the SPC chart fails to trigger a signal when the process is out-of-control. In conformity testing, a conforming product may be incorrectly declared as nonconforming due to measurement errors, which is a risk to the producer. Consumer's risk occurs when a nonconforming entity is declared as conforming. Acceptance sampling, control charting, and the conformity testing procedure require the control of both consumer's and producer's risks as part of their design. Consumer's risk is the priority in the food industry because the economic loss from consumer's risk related failures is several times higher than the cost from producer's risk. Hence the need is to develop quality control tools that ensure better protection for the consumers.

Regulatory recommendations and guidelines for acceptance sampling for food safety and consumer protection have been provided by International bodies such as the Codex Alimentarius (CAC [13, 14, 15, 16]) and the International Commission on Microbiological Specification for Foods (ICMSF [74, 75, 76]). Recent research is focused on developing sampling plans with better consumer protection or similar consumer's risks but with small sample sizes. For example, Liu and Cui [100] designed an attribute double sampling plan for three-class products, which provides the same consumer protection with smaller average sample size and hence reduces the testing cost and the laboratory workload. Santos-Fernández et al. [129] proposed to use compressed specification limits in order to provide better consumer protection even with smaller sample sizes. Further improved sampling plans are given by Zwietering [178], Whiting et al. [155], Powell [121], Gonzales-Barron and Butler [49], Jarvis [84] and Hoelzer and Pouillot [68]. Chun and Rinks [26] discussed three types of producer's

and consumer's risks. Graves et al. [53] reevaluated the producer's and consumer's risks in acceptance sampling with a Bayesian approach.

In Shewhart control charting, an additional warning limit can be used. However, this warning limit is not based on specifications. For example, producers can correct the production process immediately if Good Manufacturing Practices (GMP) limits are breached before significant deviations occur in the process. Guardbanding is an offset technique, which can be used to compensate for measurement and sampling uncertainty. The idea of guardbanding is to reduce the risk of nonconformance to specifications by using a new tightened limit instead of the original specification limit. Williams and Hawkins [158] studied the effect of guardbands on reducing testing errors. More discussion can be found in Easterling et al. [35] and Healy et al. [65]. When a tightened specification is applied, the probability of false acceptance decreases while the false rejection rate increases as a matter of trade-off. In other words, the consumer's risks are reduced at the cost of increased producer's risks. Hence, guardbanding is a strategy to balance the consumer's and producer's risks, which can be selected based on risk priorities and/or economic factors. Chou and Chen [25] proposed a kernel density estimator to correct for the measurement error and obtained the optimal guardband limit balancing both producer's and consumer's risks. Williams and Hawkins [157] discussed the economic placement of the guardband by introducing a cost model and investigated the relationship between profitability, ratio of costs and guardband selection. Kim et al. [92] introduced another cost model incorporating acceptance/rejection costs and measurement precision level to decide the optimal guardband level.

In contrast to the specification limit in acceptance sampling, LV set for conformity testing pertains to a common cause model. Hence, LV is usually a constant, predetermined value, although the methods of obtaining LV may vary between countries and industries. Holst et al. [69] proposed a two-stage conformity testing procedure taking both the producer's and consumer's risks into consideration. Pendrill [117] introduced a cost model and discussed different costs and revenues in the context of conformity testing from a producer's point of view. Forbes [42] suggested a Bayesian decision making approach to quantify and minimize the cost of wrong decisions by introducing a loss function. Examples showing the impact of measurement uncertainty on the producer's and consumer's risks in conformity assessment can be found in Hinrichs [67] and Separovic and Lourenço [136]. For international trade and regulatory purposes, a simple method of quantifying the degree of nonconformance is desirable. This approach is discussed in the next section.

1.2 Fractional nonconformance measurement

Fractional nonconformance measurement was introduced by Govindaraju and Jones [51] to correct for measurement errors in individual measurements. The term *fractional nonconformance* (FNC) refers to the probability of an error-prone observation breaching the

specification limits. Increase in sample size and repeat testing are common practices to mitigate the effect of measurement errors in the literature. In contrast to constructing a confidence interval for the true value using the repeatability standard deviation, the fractional nonconformance measurement quantifies measurement uncertainty in probabilistic terms.

For given specification limits, an observed measurement Y can be classified with certainty as conforming or not only when there are no measurement errors. Analytical testing of food quality characteristics, such as fat content etc. involves considerable measurement uncertainty, which can be up to half of the observed variation. The distribution of the measurement errors, $Z \sim N(0, \sigma_Z^2)$, can be estimated using past calibration studies. The fractional nonconformance statistic quantifies the probability of nonconformance of an individual measurement when the measurement error distribution is known and can be calculated for an observed value y as below:

$$\hat{p}_{iu} = P(x > \text{USL}) = P(z < y - \text{USL}) = \Phi\left(\frac{y - \text{USL}}{\sigma_z}\right) \tag{1.1}$$

Figure 1.2 illustrates the concept of fractional nonconformance. For a given observed measurement and (lower and/or upper) specification limits, (unconditional) fractional nonconformance statistic, $\hat{p}_{iu}$, can be calculated using Equation 1.1. In contrast to the attribute approach of classifying a measurement as conforming or not, FNC approach performs a nonlinear transformation of the apparent observation without losing any information. Also note that an observation $y_i < \text{USL}$ is classifiable as conforming only when measurement errors are absent.

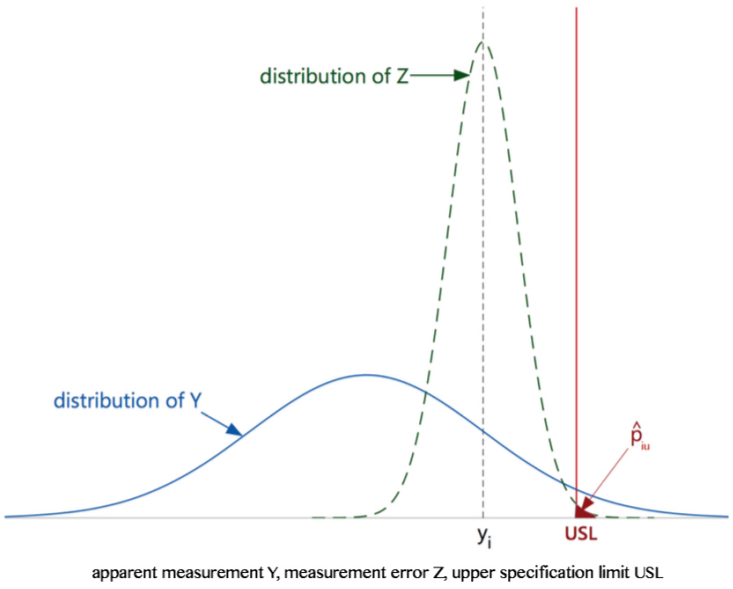


Fig. 1.2 Illustration of fractional nonconformance

A conditional version of the fractional nonconformance measure, $\hat{p}_{ic}$, is also defined in Equation 1.2 by Govindaraju and Jones [51]. The probability distribution of the measurement error Z conditional on the given observation is used to obtain the conditional fractional nonconformance statistic. The additional knowledge of the observed measurement and its distance from the sample mean contains the extra information on its nonconformance.

$$\hat{p}_{ic} = P(x > \text{USL}|Y) = \Phi\left(\frac{(y - \text{USL}) - k(y - \bar{y})}{\sigma_z\sqrt{1-k}}\right) \tag{1.2}$$

For a sample of n measurements, the total or mean FNC statistic can be computed and employed for acceptance sampling or control charting.

The FNC statistic has been successfully employed for milk powder production manufacturing applications. Figure 1.3 (from McHugh et al. [106]) shows a schematic diagram of the short-run production of various types of milk powders. The quality characteristics of interest are mainly compositions such as protein, fat, moisture percentages. There are also many other safety parameters in addition to quality characteristics.

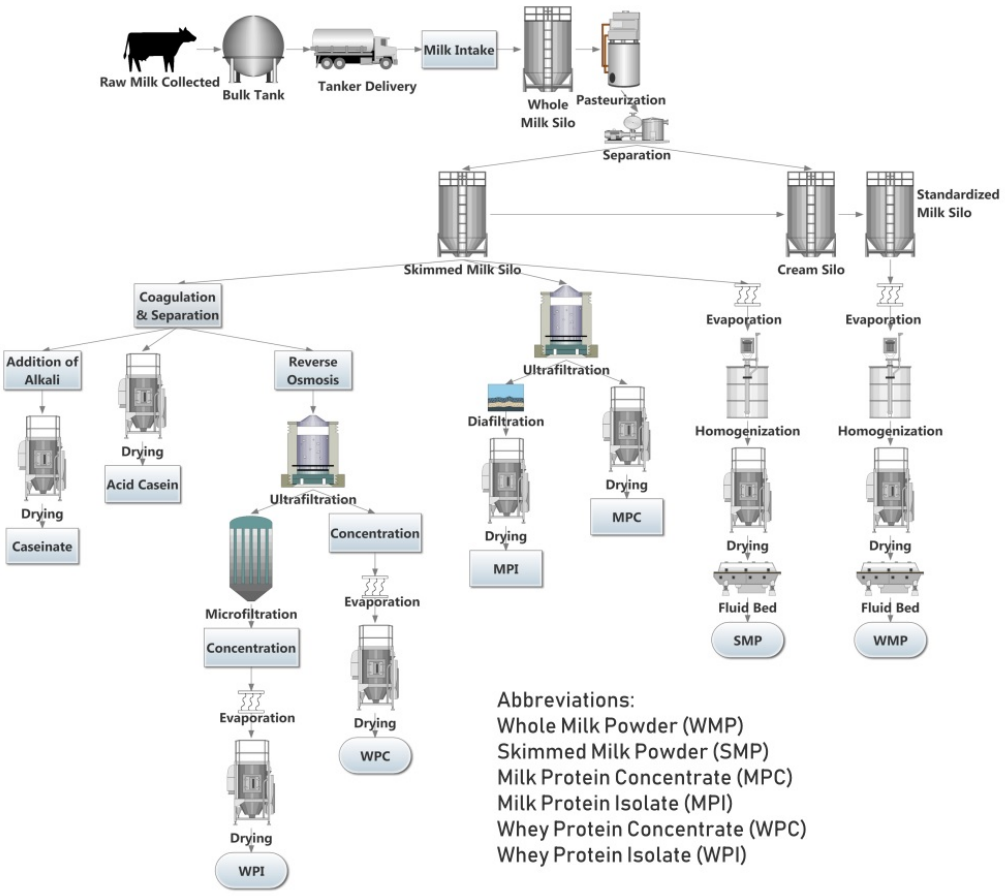


Fig. 1.3 Schematic dairy powder production pipelines

1.3 Research objectives and publications

The term *statistical control* of a process usually refers to the quality characteristic of interest fitting a single probability distribution or probability law in the absence of special causes. It is well-known that a process in a state of statistical control need not comply with the specifications, and hence statistics such as process capability index C_p or a confidence interval of it are computed. For variables control charting, Montgomery [111] even warned against placing specification limits along with control limits. On the other hand, the p or np control chart must use specifications for numerical characteristics in order to compute the number of nonconforming units in the sample. Hence the state of statistical control with reference to a p chart implies that the true fraction is rather a constant, and only sampling variation is present in the process. It should also be noted that the p chart is the first control chart devised by Shewhart; see Schilling [133]. The term *acceptance control* was suggested by Freund [43, 44]. Schilling [132, 133] used this term to imply a state of control where the products comply with the specifications with small producer's and consumer's risks.

The FNC statistic is nothing but a probability of a nonconforming observation when measurement errors are present in obtaining the particular numerical measurement. Hence, monitoring the FNC statistic or the average FNC aims to achieve a state of statistical control in the FNC and thereby leading to a state of acceptance control when supplemented by an inspection plan that controls the consumer's risk.

In the presence of measurement errors, a very low process fraction nonconforming is not demonstrable using small samples. It is known that the observed (or apparent) fraction nonconforming cannot fall below the probability of Type I error of misclassifying a conforming item as nonconforming; see Equation (15) in Lavin [96]. Food products, such as milk powder, are manufactured in high volumes (such as 20 metric tons per hour). Rapid testing for compositional characteristics (such as moisture content) using near infrared techniques involves large measurement errors. A production run would last from 15 to 30 hours for milk powders and the number of samples tested is usually small. As a result, even when the process complies with the specifications, this state of compliance may not be demonstrated adequately with the error-prone observed data. Hence, the process monitoring is driven by setting an AQL value. This is equivalent to setting upper and/or lower specification limits for a numerical characteristic. For example, an AQL of 3% means that the one-sided upper specification limit is set at 1.88 for a N(0,1) process. It should be noted that the AQL is defined as the *maximum* percent nonconforming that is considered satisfactory as a process average. The actual process level would be run well below it.

Mere specification of AQL or specification limits and the design of a control chart may not protect the consumer's interests. In a short run environment, the power of detection improves only after 20 to 30 observations are available but by this time, the high volume production process will end. The output from a single run will be dispatched to many consumers, and

hence each part of the production (from which only very few observations are obtained) needs to be assured to be of not limiting or poor quality. Hence the need for an acceptance sampling plan design with the specification of both AQL and LQL even after monitoring the process using control charts. Consumer protection is one of main objectives of food quality control. The work reported here (in Chapter 4) includes provisions such as guardbanding in addition to employing quality limits such as AQL and/or LQL in the design.

The main research objective of this thesis is to extend the fractional nonconformance measurement approach to control charting of short-run production processes, lot acceptance procedures and conformity assessment problems including testing for doping in sports. The intended area of application of the newly developed procedures is the manufacture of milk products (milk powder, cheese etc.).

- Objective 1: Implement fractional nonconformance control charts to monitor short-run production processes.

  The challenge here is deriving time-varying control limits and computation of the average run length for short production lengths such as 30.

- Objective 2: Global optimal design of control chart and lot grading procedures in the presence of measurement errors.

  For short-run production, the optimum process level settings must be found for both the given fractional nonconformance level and for meeting the designated overall consumer's risk.

- Objective 3: Employ the fractional nonconformance statistic for conformity testing.

  The main challenge is to develop an improved conformity testing procedure with higher sensitivity and/or specificity than the current procedure.

In the course of this PhD study, three articles have been published in peer-reviewed international journals and two manuscripts are yet to be published. The first research objective is fulfilled and the results are presented in Chapters 2 and 3. The accomplishment of the second objective and the relevant results are presented in Chapter 4. The completed work on the last objective appears as Chapters 5 and 6.

The research outputs, which are formed by the papers, are presented in the upcoming Chapters:

- Chapter 2: Zhou, X., Govindaraju, K., and Jones, G. (2018). Monitoring fractional nonconformance for short-run production. *Quality Engineering*, 30, 498–510.

- Chapter 3: Zhou, X., Govindaraju, K., and Jones, G. FNC control charts for bulk material. *Journal of Statistical Computation and Simulation.* (submitted/under review)

- Chapter 4: Zhou, X., Govindaraju, K., and Jones, G. (2019). Acceptance control and guardbanding for error-prone individual measurements. *Quality and Reliability Engineering International*, 35, 517–534.
- Chapter 5: Zhou, X., Govindaraju, K., and Jones, G. (2019). Fractional nonconformance based conformity testing. *Computers & Industrial Engineering*, 135, 402– 411.
- Chapter 6: Zhou, X., Govindaraju, K., and Jones, G. Modified FNC based conformity testing using subject-specific limiting value. *Journal of Quantitative Analysis in Sports*. (submitted/under review)

Chapter 2

Monitoring Fractional Nonconformance for Short-run Production

Xin Zhou, Kondaswamy Govindaraju, Geoff Jones
Quality Engineering, 2018
https://www.tandfonline.com/doi/full/10.1080/08982112.2017.1360499

2.1 Abstract

Quality characteristics observed in industrial processes are not always free from measurement errors. The term *fractional nonconformance* refers to the probability of an error-prone observation breaching the specification limits. Four new control statistics based on the fractional nonconformance concept are defined for process monitoring purposes. This work, motivated by milk products manufacturing, is tailored for short-run productions in which only individual measurements are accumulated over time. The performance of the newly defined control statistics is evaluated using simulation for both independent and autocorrelated processes. The results show that fractional nonconformance charts can be useful to monitor short-run production process, and the choice of the monitoring scheme does not heavily depend on the distribution of the quality characteristics.

Keywords

autocorrelation; control chart; fractional nonconformance; short run production; statistics

2.2 Introduction

Control charting methods, such as Shewhart, cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) involve an initial phase (called Phase I) in which the common cause periods are identified and then the process parameters are estimated. Montgomery [111] suggested a collection of 25-30 subgroups of size five for Phase I to correctly estimate the process parameters and establish control limits for Phase II monitoring. Food manufacturing processes are often short-run or batch production processes limited by time and volume. Food manufacturing demands frequent stoppage and cleaning of machinery for safety, and hence the production becomes a short-run. Phase I estimation is not practical for short-run processes. For an introduction to short-run process control; see Elam [36]. The main challenge for short-run production is to start exercising control of the process with few samples. As soon as new data become available, dynamic update of control statistic of interest is needed for improving the statistical power of the control chart. For example, if the control statistic is the process mean level, the running sample means contain more information than the individual subgroup means for stationary processes. On the other hand, if the process itself is not in-control or non-stationary, dynamic accumulation of sample results can delay detection of sudden big shifts. Quality improvement is also harder to detect for short-run processes because low fraction nonconforming level changes can be detected only with very large sample sizes. In the next paragraph, we briefly present some of the remedies available in the literature for lack of Phase I data.

Due to lack of Phase I, the deviation from nominal/target value (DNOM) is monitored for short-run processes. The use of a nominal value for the central line and setting up control limits partly eliminates the estimation issues but this approach is suitable mainly for individual observations. Self-starting control charts, which begin monitoring a process from the first observation and update the parameter estimates with new observations, have been proposed to monitor start-up processes and short-runs. Quesenberry [122] introduced self-starting Q charts to detect changes in the process mean or variance. Celano et al. [21] proposed the Shewhart t chart and EWMA t chart to monitor short production runs and illustrated that the EWMA t chart is more effective in detection of mean shift.

Linna and Woodall [99], Maravelakis et al. [105] and Maravelakis [104] studied measurement error effects on the Shewhart, EWMA and CUSUM charts respectively. Also see Bennett [10], Mizuno [110], Abraham [3], Yang and Yang [170], Yang et al. [169], Abbasi [2], Costa and Castagliola [28], Haq et al. [57], Hu et al. [72]. For a recent review of measurement error effects on control charts; see Tran et al. [147]. Much of the published work on measurement error adjustment of control charts is not aimed at short production processes, the exception being Farnum [41] who focussed on control chart procedures for short runs adjusted for measurement errors using DNOM data.

Food quality characteristics, such as the percentage sugar or fat, are determined using analytical methods which involve considerable measurement uncertainty. Most food products are bulk in nature and hence the general practice in food process control is to obtain a representative sample at any given time. Only a single lab measurement is made on the whole sampled amount (such as 500g) after lab testing. The Shewhart Individual or I-chart is generally used when only single measurements can be made but to implement this chart, Phase I estimation is required and hence it is not suitable for the short run production environment. The design of individuals charts (I-charts) in the presence of measurement errors is also not addressed in the literature.

The term *fractional nonconformance* refers to quantification of measurement uncertainty in probabilistic terms (as against providing a confidence interval for the true value using the repeatability standard deviation). This term was proposed by Govindaraju and Jones [51] for acceptance sampling inspection problems. In this paper, the fractional nonconformance statistic is investigated further for process monitoring applications in the form of I-charts. Four different monitoring schemes based on fractional nonconformance for short-run productions are evaluated for their control chart performance. The remainder of the paper is organized as follows: Section 2.3 presents a brief review of the fractional nonconformance approach. Performance measurements of four fractional nonconformance monitoring schemes are evaluated in Section 2.4 for both independent and autocorrelated processes. A case study is shown in Section 2.5. A *Shiny* app for fractional nonconformance charting is developed and described in Section 2.6. The last section provides a brief discussion and conclusion.

2.3 Fractional nonconformance

An observed measurement can be classified as conforming or not based on a given lower (L) and/or upper (U) specification limit with certainty when measurement errors are absent. Analytical measurements are often indirect and the associated measurement errors may not be negligible. In order to mitigate the effect of measurement errors, increasing sample size and repeat testing are mostly recommended in the literature. In contrast to the above strategies, Govindaraju and Jones [51] proposed a fractional nonconformance probability measure for each individual observation when the measurement error distribution is known. The common measurement error model $Y = X + Z$ was adopted, where X, Y, Z are the true measurement, the apparent measurement and the measurement error respectively. Let $X \sim N(\mu_X, \sigma_X^2)$, $Z \sim N(0, \sigma_Z^2)$ and $Y \sim N(\mu_X, \sigma_X^2 + \sigma_Z^2)$, assuming that there is no instrument bias or $\mu_X = \mu_Y$. For a random sample with apparent sample measurements $(y_1, y_2, ..., y_n)$ with a given upper specification limit U, (individual unconditional) fractional nonconformance

of each observation after correcting for the measurement error can be defined as below:

$$\hat{p}_{iu} = P(x > U) = P(z < y - U) = \Phi\left(\frac{y-U}{\sigma_z}\right) \tag{2.1}$$

The statistic $\delta = \sum_{i=1}^{n} \hat{p}_{iu}$ is an inverse measure of quality and is used for lot disposition purposes in Govindaraju and Jones [51]. The fractional nonconformance measure $\hat{p}_{iu}$ can also be viewed as a nonlinear transformation of the observed continuous observations which are adjusted for measurement errors. If the observed measurements are directly classified as conforming or not, the information loss is huge for short-run process data. The resulting attribute (binary) data is also misleading because these are not adjusted for measurement errors. In contrast to the traditional attribute method of classifying a measurement as conforming or nonconforming, the fractional nonconformance approach assigns a probability of nonconformance to specifications. The closer an apparent measurement to the specification limit, the higher the fractional nonconformance.

0.5508	0.5503	0.5584	0.5569
0.5545	0.5502	0.5573	0.5518
0.5514	0.5579	0.5580	0.5508
0.5531	0.5575	0.5592	0.5514
0.5563	0.5517	0.5529	0.5583
0.5542	0.5561	0.5576	0.5585
0.5530	0.5598	0.5547	0.5566

(a) Apparent measurement

0.0000	0.0000	0.1431	0.0194
0.0001	0.0000	0.0359	0.0000
0.0000	0.0808	0.0912	0.0000
0.0000	0.0478	0.2969	0.0000
0.0068	0.0000	0.0000	0.1285
0.0001	0.0047	0.0548	0.1587
0.0000	0.4470	0.0002	0.0117

(b) Fractional nonconformance

Table 2.1 Milk cream fat compositions measured for a short-run production process (read data by columns for observation order).

As an example, consider the 28 hourly measurements of fat composition in a milk cream short-run production process shown in Table 2.1(a). The measurement error is known to be normal $N(0, 0.0015)$. For a given upper specification limit of 56% for the fat content, the observed dataset does not contain any nonconforming measurements. By applying Equation 2.1, the fractional nonconformance of each observation is calculated and shown in Table 2.1(b). Note that $\delta = 1.5277$. Even though none of the observations are nonconforming, the total of the fractional nonconformance values exceeds one. This example shows that the fractional nonconformance is a sensitive measure compared to counting nonconforming measurements.

Figure 2.1 shows the Individuals chart for the fractional nonconformance values from Table 2.1(b) for an upper control limit (UCL) of 0.3 corresponding to the false alarm rate of 0.05. An out-of-control signal is triggered by the 14th observation. Note that there is no set nominal or target value for the fat quality characteristic and hence the central line is avoided. Instead of plotting the individual fractional nonconformance $\hat{p}_{iu}$ values, the

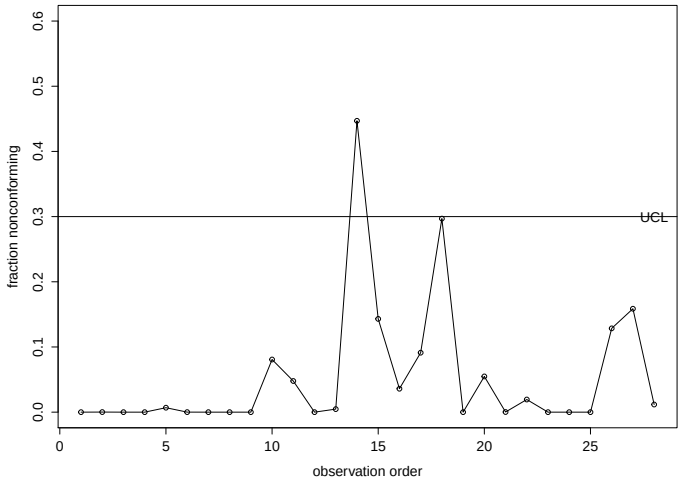


Fig. 2.1 Individuals chart for fractional nonconformance

traditional Individuals chart for apparent measurements can also be set up. This I-chart can only monitor the mean level and not the fraction nonconforming of the process. Since there is no set mathematical or statistical relationship between the control limits and specification limits Montgomery [111, pp.245-246], the I-chart based on apparent measurements cannot monitor the fraction nonconforming level of the process even when specification limits can be displayed on the I-chart. It is also difficult to interpret the chart when observations are subject to measurement errors. The p-chart for fraction nonconforming also cannot be drawn for a sample of size just one. However the I-chart for fractional nonconformance is a good alternative for monitoring the process fraction nonconforming with individual error-prone measurements.

Besides the fractional nonconformance for an individual observation obtained by Equation 2.1, an alternative measure of fractional nonconformance conditional on the given apparent measurement Y can also be calculated. Let $k{=}\sigma_Z^2/\sigma_Y^2$. The ratio k is commonly in the range 0.01 to 0.25 for within lab measurement errors. When between-lab measurement errors are present, k can be as high as 0.5. The probability distribution of the measurement error Z conditional on the given observed measurement value y, say $f(Z|Y=y)$, is shown to follow normal $N\left(k(y-\mu_X),(1-k)\sigma_Z^2\right)$ in Govindaraju and Jones [51]. Hence the fractional nonconformance conditional on the apparent measurement can be estimated as:

$$\hat{p}_{ic} = P(x > U|Y) = \Phi\left(\frac{(y-U)-k(y-\bar{y})}{\sigma_z\sqrt{1-k}}\right) \tag{2.2}$$

Equation 2.2 uses the additional knowledge that an apparent measurement has been made and its distance from the sample mean contains additional information on its nonconformance.

For monitoring short-run production processes subject to measurement errors, the individual *unconditional* fractional nonconformance statistic $\hat{p}_{iu}$ given in Equation 2.1 and the individual *conditional* fractional nonconformance statistic $\hat{p}_{ic}$ given in Equation 2.2 are both useful. In addition to these two individual fractional nonconformance statistics, two more statistics based on the *average* fractional nonconformance at time t can also be defined:

$$\hat{p}_{au} = \frac{1}{t}\sum_{j=1}^{t}\Phi\left(\frac{y_j - U}{\sigma_z}\right) \tag{2.3}$$

$$\hat{p}_{ac} = \frac{1}{t}\sum_{j=1}^{t}\Phi\left(\frac{(y_j - U) - k(y_j - \bar{y})}{\sigma_z\sqrt{1-k}}\right) \tag{2.4}$$

The statistics defined in Equations 2.3 and 2.4 can be updated once a new observation from the process is obtained. We call $\hat{p}_{au}$ the average unconditional fractional nonconformance and $\hat{p}_{ac}$ the average conditional fractional nonconformance.

A Shewhart control chart procedure based on the individual unconditional fractional nonconformance statistic $\hat{p}_{iu}$ has been implemented in a large milk product manufacturing process. In order to investigate further improvement to fractional nonconformance monitoring, the control statistics $\hat{p}_{ic}$, $\hat{p}_{au}$ and $\hat{p}_{ac}$ for monitoring short-run production processes are considered. The control chart performance for the four types of fractional nonconformance statistics, calculated using Equations 2.1-2.4, is discussed in the next section. Both independent and autocorrelated processes are considered.

2.4 Comparison of various FNC statistics for short run production

The most common statistical measure of a control chart's performance is the average run length (ARL), which is the expected number of samples until a signal is triggered. A smaller ARL at an undesirable process level means a higher ability to detect such a shift in process. However, in the case of a short run, the run may end without any signals throughout the process. So the conventional ARL is not fully appropriate for short run control charting, and hence alternative performance measures have been proposed in the literature. Nenes and Tagaras [112] suggested the truncated average run length (TARL), which is the average number of samples until a signal or until the completion of the process without any signal, whichever occurs first. Following Nenes and Tagaras [112], let

$$\text{TARL}_0(N) = \sum_{n=1}^{N} n(1-\alpha)^{n-1}\alpha + (N+1)(1-\alpha)^N = \frac{1-(1-\alpha)^{N+1}}{\alpha} \tag{2.5}$$

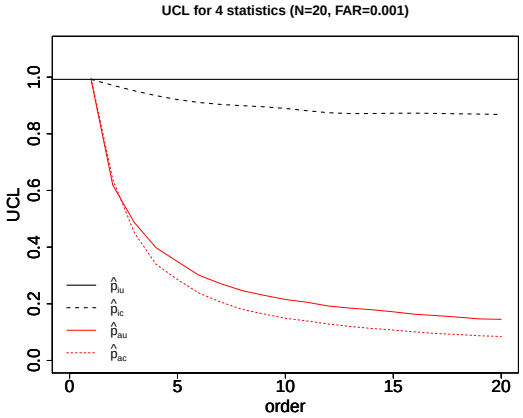
$$\text{TARL}_\delta(N) = \sum_{n=1}^{N} n\beta^{n-1}(1-\beta) + (N+1)\beta^N = \frac{1-\beta^{N+1}}{1-\beta} \tag{2.6}$$

where α is the probability of a Type I error or false alarm rate and β is the probability of a Type II error; δ is the level of process shift and N is the maximum production length. $\text{TARL}_0(N)$ measures the in-control performance whereas $\text{TARL}_\delta(N)$ measures the out-of-control performance. Another useful statistic is the probability q of getting a signal within a specified number of samples N, which was proposed by Quesenberry [124]. The in-control signal probability $q_0(N)$ and the out-of-control signal probability $q_\delta(N)$ are defined as

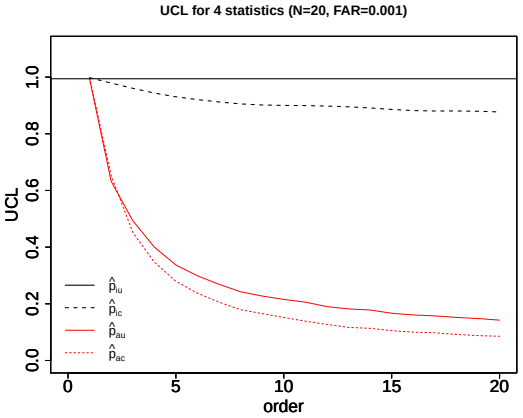
$$q_0(N) = 1-(1-\alpha)^N \tag{2.7}$$

$$q_\delta(N) = 1-\beta^N \tag{2.8}$$

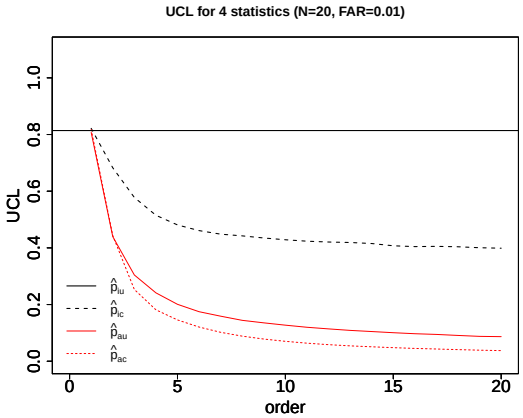
To compare the performance of the four fractional nonconformance monitoring schemes based on $\hat{p}_{iu}$, $\hat{p}_{ic}$, $\hat{p}_{au}$ and $\hat{p}_{ac}$, the maximum production length N was fixed as 20. The specification limits are assumed to correspond to a given acceptance quality limit (AQL) such as 3% fixed for the apparent data. If we adjust for the measurement errors, the true AQL will be rather very low. A simulation study with 10^5 replicates was conducted to obtain the empirical distribution of fractional nonconformance for different false alarm rates ($\alpha = 0.1\%, 1\%$ and 5%). Note that α uniquely defines the UCL for the fractional nonconformance control charts. A signal is detected if any single observation falls above UCL. Note that α here is the probability of getting a signal for each single observation in the run rather than the probability of getting a signal for the whole run of plotted points. The conditional and average fractional nonconformance statistics at time t are not independent of those calculated at time $(t-1)$ and hence time varying UCLs were used (see Figure 2.2) to ensure that every single observation has the same false alarm rate α. Empirical TARL and q values were obtained with 2,000 replicates to compare the performance of the four fractional nonconformance monitoring schemes: individual unconditional ($\hat{p}_{iu}$), individual conditional ($\hat{p}_{ic}$), average unconditional ($\hat{p}_{au}$) and average conditional ($\hat{p}_{ac}$). Independent processes under both normal and beta distributions as well as autocorrelated processes including simple stationary and non-stationary cases were investigated.



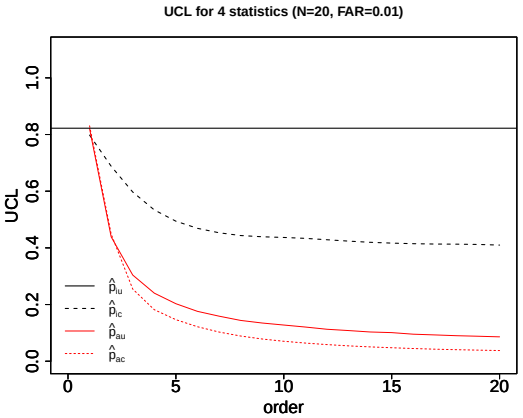
(a) Normal model ($\alpha = 0.1\%$)



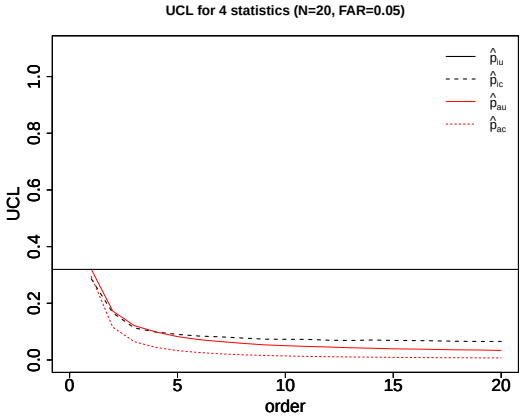
(b) Beta model ($\alpha = 0.1\%$)



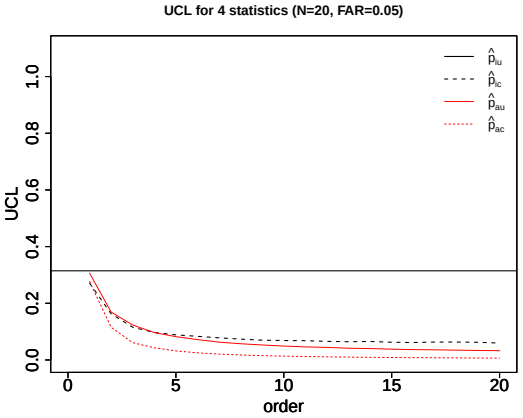
(c) Normal model ($\alpha = 1\%$)



(d) Beta model ($\alpha = 1\%$)



(e) Normal model ($\alpha = 5\%$)



(f) Beta model ($\alpha = 5\%$)

Fig. 2.2 UCL for the four fractional nonconformance based control statistics

2.4.1 Independent processes

Case 1: Normal process

It is assumed that the quality characteristic Y observed in the short-run production process is normally distributed. We further assume that the measurement error Z is also normally distributed. In practice, past calibration study data can be used to justify the normality for Z even when Y is non-normal. Let the in-control $Y \sim N(\mu_0, \sigma)$, whereas an out-of-control condition happens when there is a change in the process parameter, which means $Y \sim N(\mu_1, \tau\sigma)$.

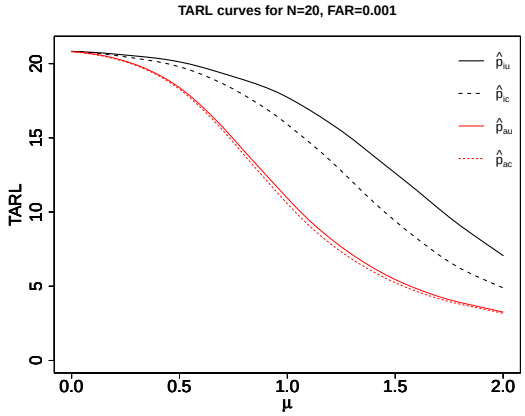
The purpose of this study is to compare the sensitivity of the four fractional nonconformance monitoring schemes in detecting a shift in the mean of the process. We assume that the observed process dispersion is generally constant (τ=1), which can again be validated on empirical grounds. Without loss of generality, let the in-control mean $\mu_0 = 0$ and the process standard deviation $\sigma = 1$ so that $Z \sim N(0, 0.5)$ when $k = 0.25$. For a given false alarm rate α, the UCL values were first obtained (as shown in Figure 2.2). We then allowed the process mean to shift in the range $\mu_0 = 0$ and $\mu_1 = 2$ at time $t = 0$ (initial-state shift) and obtained $\text{TARL}_\delta(N)$ and $\text{q}_\delta(N)$ by simulation. Note that a steady-state shift, say $t = 6$, for a short run with $N = 20$ can be treated as another shorter run ($N = 15$) with initial-state shift.

Figure 2.3 shows no difference in $\text{TARL}_0(N)$ and $q_0(N)$ for the in-control process for the four fractional nonconformance monitoring schemes. If the false alarm rate is set low (say <1%), the $\hat{p}_{ic}$ statistic achieves faster detection of a shift in the mean than its counterpart $\hat{p}_{iu}$ because its TARL and q curves are steeper. The performance of $\hat{p}_{ac}$ and $\hat{p}_{au}$ statistics is similar in terms of their TARL and q performance but they produced sharper curves of $\text{TARL}_\delta(N)$ and $q_\delta(N)$ when compared to the individual fractional nonconformance statistics particularly when the false alarm rate is small (α = 0.1% and 1%). On the other hand, this superior performance is absent for α=5%, which means that the sensitivity in detecting a shift in the mean is almost the same for all the four control statistics. Based on this simulation study, we recommend $\hat{p}_{ic}$ and $\hat{p}_{au}$ control charts to monitor normally distributed short-run processes.

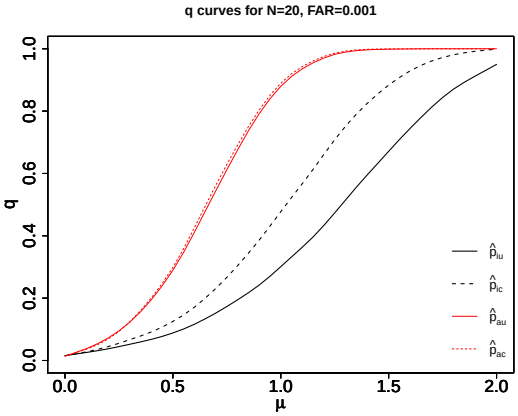
Case 2: Beta process

Compositional fractions such as percentage sugar occurring in the food industry follow the beta distribution rather than normal. Hence we consider the case of the apparent measurement Y following a beta distribution whose density function is given by:

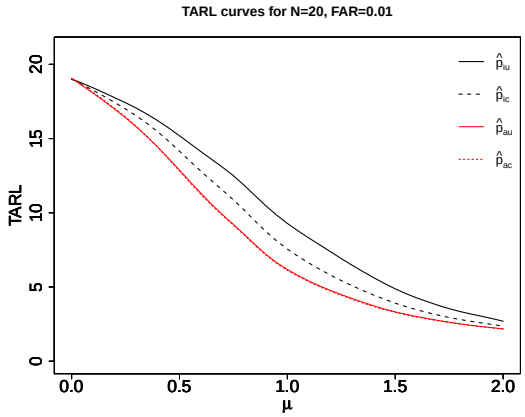
$$\text{Beta}(a,b) = \frac{1}{\text{B}(a,b)} y^{a-1}(1-y)^{b-1}, 0 < y < 1, a, b > 0 \tag{2.9}$$



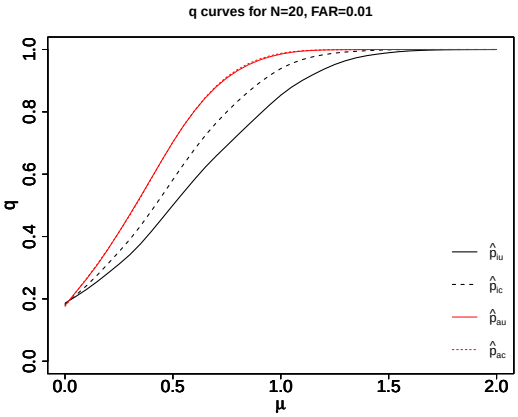
(a) TARL performance ($\alpha = 0.1\%$)



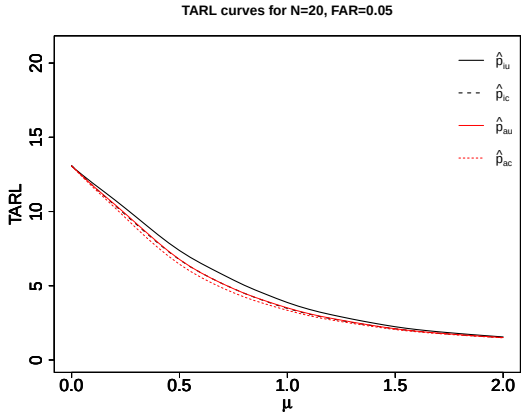
(b) q performance ($\alpha = 0.1\%$)



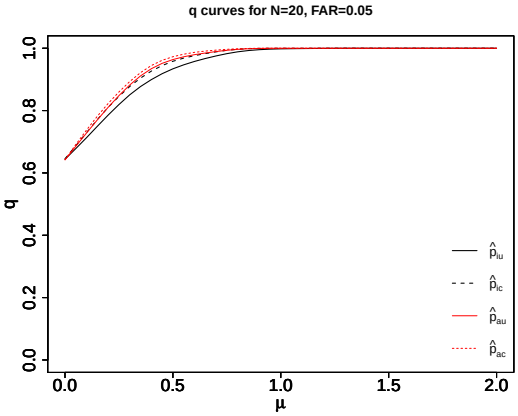
(c) TARL performance ($\alpha = 1\%$)



(d) q performance ($\alpha = 1\%$)



(e) TARL performance ($\alpha = 5\%$)



(f) q performance ($\alpha = 5\%$)

Fig. 2.3 TARL and q performance under the normal model

where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function. The beta density can be also re-parameterized in terms of the expected value $E(Y) = \mu = a/(a+b)$ and precision parameter $\theta = a+b$ as Beta$(\mu\theta, (1-\mu)\theta)$. For large θ, $Var(Y) = ab/(a+b)^2(a+b+1) \approx \mu(1-\mu)/\theta$.

Consider the case $Y \sim$ Beta$(500, 1500)$ representing the in-control process with $E(Y) = \mu_0 = a/(a+b) = 0.25$, $Var(Y) = ab/(a+b)^2(a+b+1) = 0.0001$. For a highly automated production processes, it is reasonable to assume a known precision parameter θ $(= a+b = 2000)$; see Govindaraju and Kissling [52]. We investigated the performance of the four fractional nonconformance control statistics for process mean shifts from $\mu_0 = 0.25$ (no shift) to $\mu_1 = 0.28$ (shift up to 3xSD), which correspond to shifts in terms of a from $a = 500$ to $a = 560$.

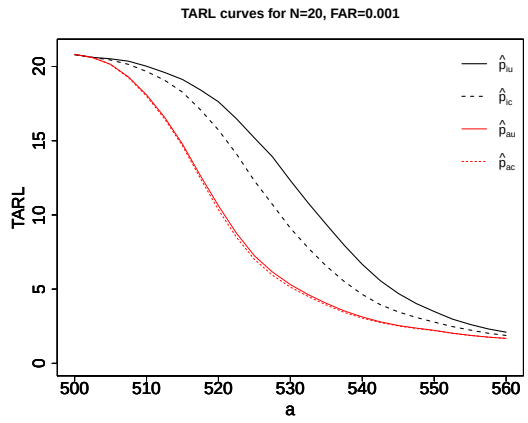
As illustrated in Figure 2.4, all the four monitoring schemes considered in this paper achieve the same in-control performance and their out-of-control performance under the beta distribution is similar to their performance under the normal distribution. For small false alarm rates, 'conditional' type of fractional nonconformance outperformed only when individual statistic is applied. For a higher false alarm rate such as 5%, the average conditional fractional nonconformance approach is the most effective one for detecting small shifts, but all four statistics performed about the same in detecting large shifts in mean. Based on our simulation, we recommend that both $\hat{p}_{ic}$ and $\hat{p}_{au}$ can also be used to monitor short-run beta processes.

Case 3: Effect of production length

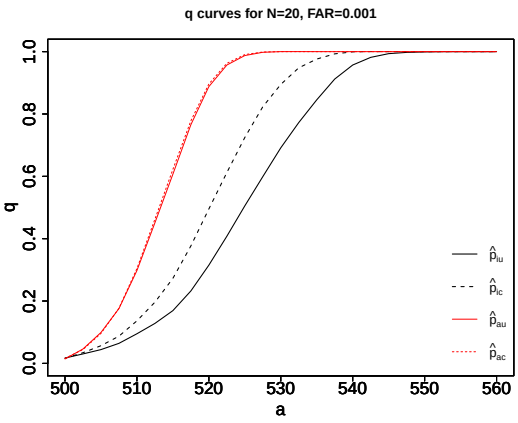
In the previous sections, the properties of the four fractional nonconformance monitoring schemes were studied fixing the maximum production length N as 20. The maximum production length N may vary depending on the customer order size, and hence we investigated whether our choice of the fractional nonconformance statistic is robust to N. Firstly, we investigated the performance of the four fractional nonconformance monitoring schemes for $N = 5$ to $N = 30$. For α=1%, we assumed in-control process means $\mu_0 = 0$ (normal) and $\mu_0 = 0.25$ (beta), and out-of-control process means $\mu_1 = 1$ (normal) and $\mu_1 = 0.26$ (beta). Equations 2.5 and 2.7 clearly show that $TARL_{\mu_0}$ and q_{μ_0} will always increase with N. In other words, the in-control performance depends on N. However, when the process is out-of-control, as illustrated in Figure 2.5, $TARL_{\mu_1}$ and q_{μ_1} curves become flatter for $N \geq 15$. This suggests that the control chart performance becomes stable only when it ages. We repeated our simulation study in Cases 1 and 2 for $N = 15$, 25, and 30. Our investigation confirmed that the $\hat{p}_{ic}$ and $\hat{p}_{au}$ statistics perform better for both normal and beta models (see supplemental material online).

2.4.2 Autocorrelated process

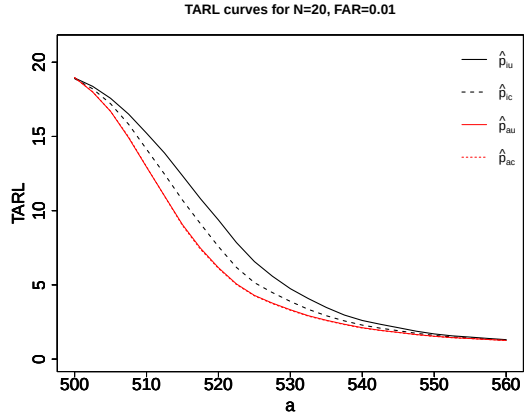
We assumed that the underlying process measurements are independent, and identically distributed in the previous section. However, this assumption may not hold in practice;



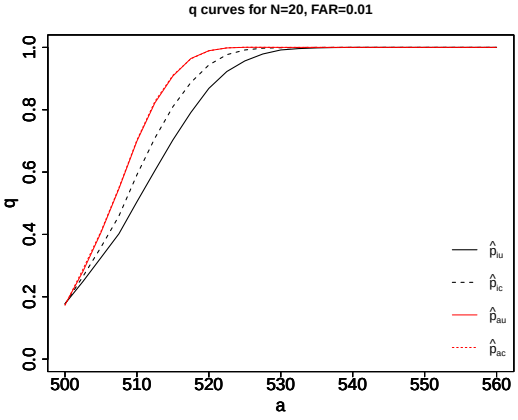
(a) TARL performance ($\alpha = 0.1\%$)



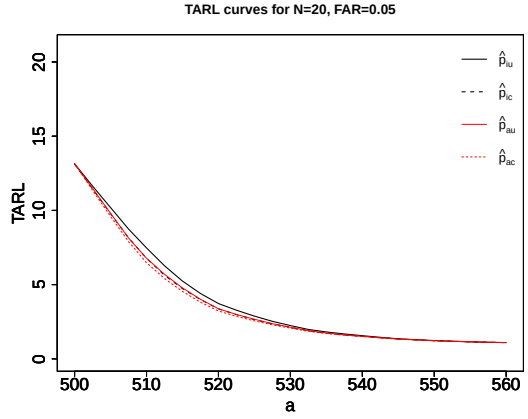
(b) q performance ($\alpha = 0.1\%$)



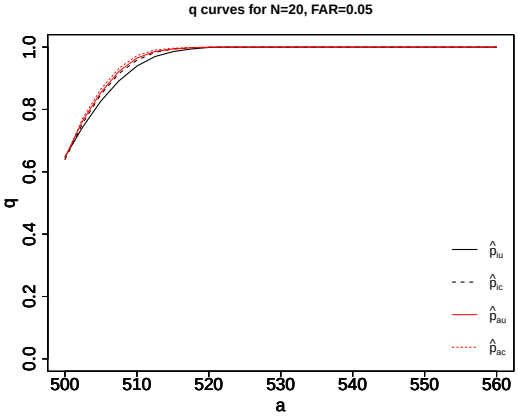
(c) TARL performance ($\alpha = 1\%$)



(d) q performance ($\alpha = 1\%$)



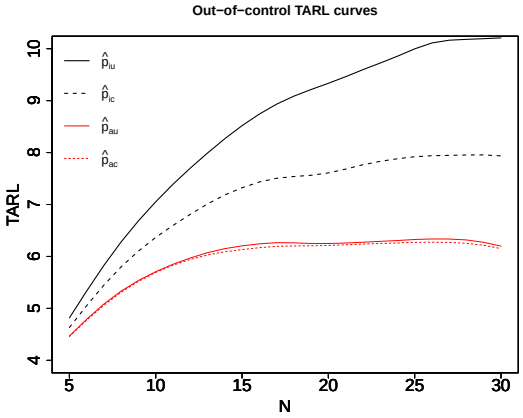
(e) TARL performance ($\alpha = 5\%$)



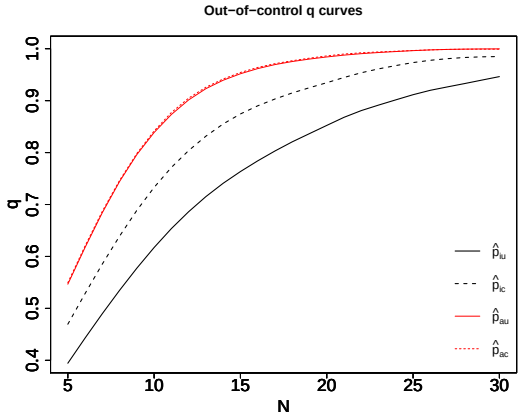
(f) q performance ($\alpha = 5\%$)

Fig. 2.4 TARL and q performance under the beta model

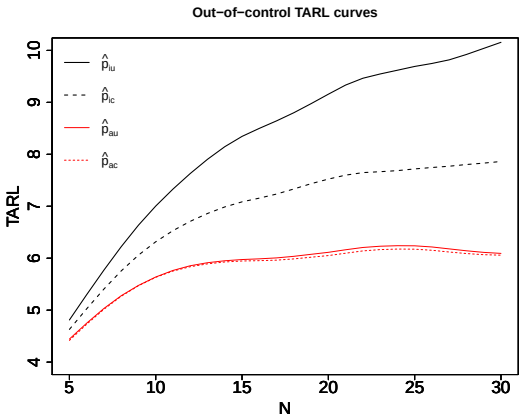
see Alwan [5]. For autocorrelated data, Box and Jenkins autoregressive integrated moving average (ARIMA) models can be fitted; see Box et al. [12]. ARIMA models are compactly



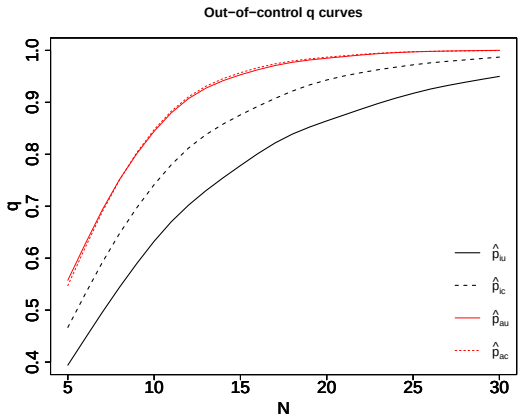
(a) TARL_{μ_1} (normal distribution)



(b) q_{μ_1} (normal distribution)



(c) TARL_{μ_1} (beta distribution)



(d) q_{μ_1} (beta distribution)

Fig. 2.5 TARL_{μ_1} and q_{μ_1} performance with different sample size for process under normal and beta distribution

denoted as ARIMA(p,d,q) and expressed in the form:

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)\epsilon_t \tag{2.10}$$

where $\phi_p(B)$ and $\theta_q(B)$ are the polynomial functions of autoregressive order p and moving-average order q, d is the degree of differencing, B is the backshift operator defined by $BY_t = Y_{t-1}$, and $\epsilon_t \sim N(0,\sigma_\epsilon^2)$ is random error.

A stationary process is one whose properties do not depend on the time at which it is observed, in particular its mean and standard deviation do not change. In contrast, a non-stationary process does not have a fixed mean or standard deviation. Performance of fractional nonconformance on both stationary and non-stationary process are discussed in this section.

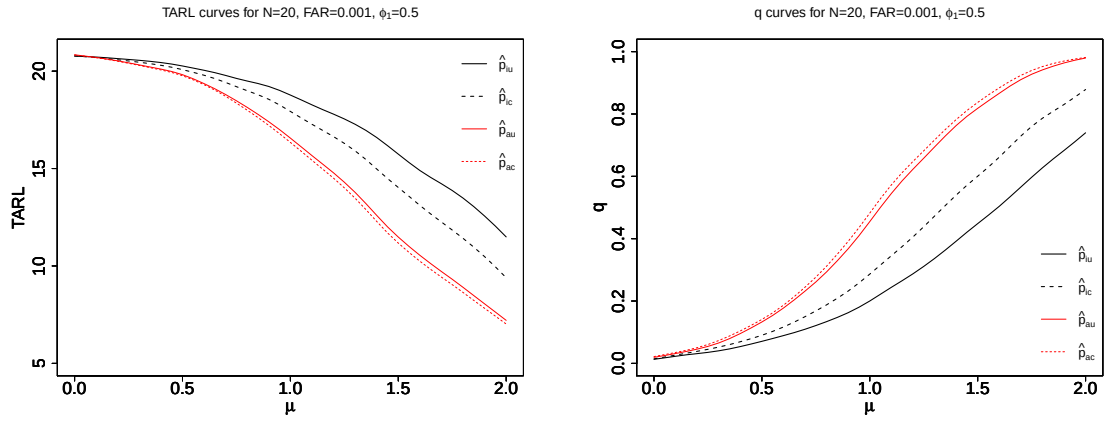


Fig. 2.6 TARL and q performance for stationary process AR(1)

Stationary process

Let the quality characteristic Y_t follow the first order ARIMA(1,0,0) or AR(1) process:

$$y_t = \phi_1 y_{t-1} + \epsilon_t \tag{2.11}$$

where $\epsilon_t \sim N(0, \sigma_\epsilon^2)$. Without loss of generality, let the random error $\epsilon_t \sim N(0,1)$. When $0 < |\phi_1| < 1$, the process mean and variance are given by $\mathrm{E}(Y) = 0$ and $\mathrm{Var}(Y) = \sigma_\epsilon^2/(1-\phi_1^2) = 1/(1-\phi_1^2)$. It should be noted that Y_t is a stationary process. When $\phi_1 = 0$, $y_t = \epsilon_t$ is white noise and Y_t is i.i.d.; when $\phi_1 = 1$, $y_t - y_{t-1} = \epsilon_t$ is a random walk. The variance of a random walk increases linearly with t, which means that the process has no fixed standard deviation. Woodall and Faltin [162] found that positive autocorrelation is more common in manufacturing process. We therefore considered a moderate positive autocorrelation coefficient of $\phi_1 = 0.5$ for the simulation study.

Let the short-run production be a stationary process $y_t = 0.5y_{t-1} + \epsilon_t$ with mean $\mu_0 = 0$ and $\sigma_Y^2 = 1/(1-0.5^2) = 1.33$. For α=0.1%, we varied the process mean from $\mu_0 = 0$ to $\mu_1 = 2$ and then compared the sensitivity of the four monitoring schemes in detecting the mean shifts. Figure 2.6 shows that the individual conditional, and average unconditional fractional nonconformance statistics continue to outperform. This is not surprising because if ϵ_t is normal then so is y_t.

Non-stationary process

The integrated moving average model IMA(d,q), which is also denoted ARIMA$(0,d,q)$, is non-stationary. We considered the simple IMA$(1,1)$ model:

$$y_t - y_{t-1} = \epsilon_t - \theta_1 \epsilon_{t-1} \tag{2.12}$$

where $\epsilon_t \sim N(0,\sigma_\epsilon^2)$ and $0 < |\theta_1| < 1$. When $\theta_1 = 1$, $y_t = \epsilon_t$ is white noise and Y_t is i.i.d.; when $\theta_1 = 0$, $y_t = y_{t-1} + \epsilon_t$ is a random walk.

Let the quality characteristic Y_t be an IMA$(1,1)$ process with $\epsilon_t \sim N(0,1)$. Since Y_t is non-stationary, the process has no fixed mean, so we cannot consider the shifts in the process mean as we did for the i.i.d. and stationary processes. Instead we incorporate a shift function at a random time $t' \in \{0,1,2,...,n\}$ as shown below:

$$y_t' = y_t + \delta_t \tag{2.13}$$

where $\delta_t = 0$ if $t < t'$; otherwise, $\delta_t = \delta_{t'}$ ($\delta_{t'}$ being the magnitude of the shift). That is, the non-stationary process Y_t experiences a step shift of size $\delta_{t'}$ at a random time period t' and this shift persists until the end of the process. Assuming a constant process variance, with $0 \le \delta_{t'} \le 20$, and for α=0.1%, we studied the performance of the four types of fractional nonconformance statistics. We investigated both positive and negative moving-average coefficients $\theta_1 = 0.5$ and $\theta_1 = -0.5$. As shown in Figure 2.7, $\hat{p}_{ic}$ and $\hat{p}_{au}$ statistics are preferred for non-stationary processes. In addition, $\hat{p}_{ic}$ is particularly sensitive in detecting a large shift whereas $\hat{p}_{au}$ is suitable for identification of small shifts.

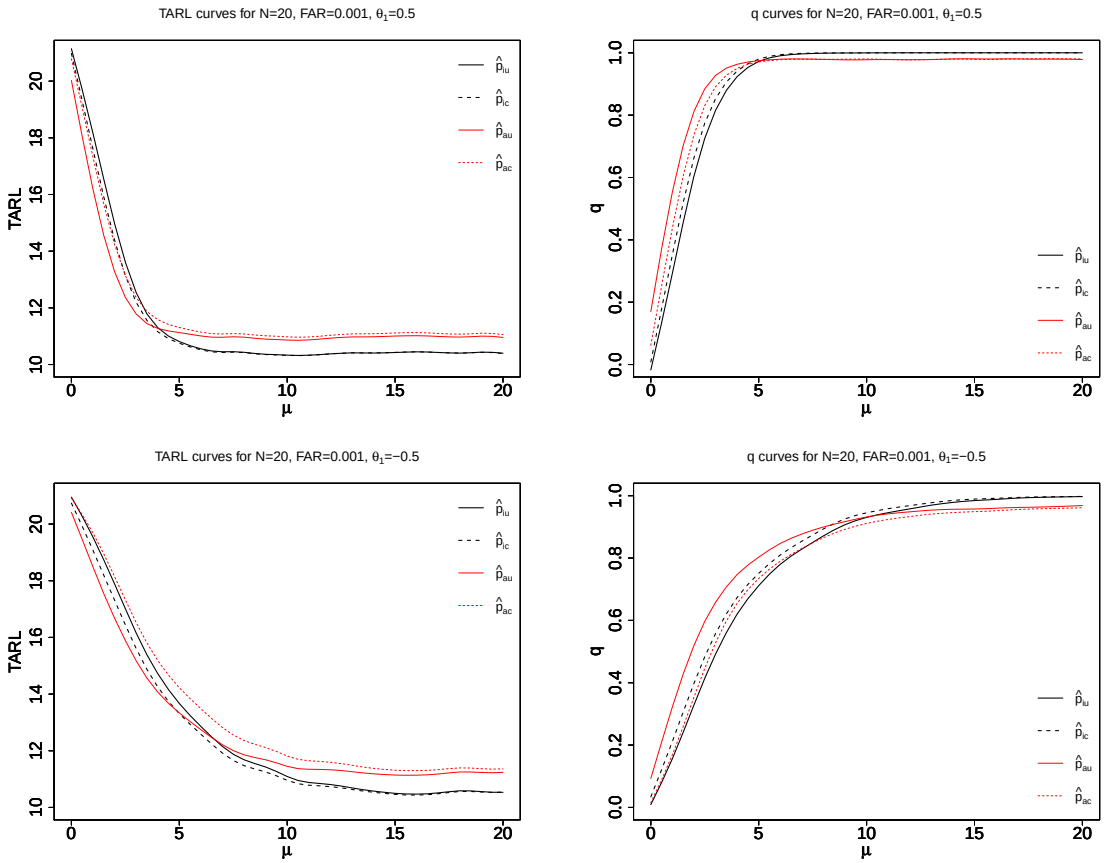


Fig. 2.7 TARL and q performance for non-stationary process IMA(1,1)

2.5 Case study

The four fractional nonconformance monitoring schemes were trialled for monitoring the percentage moisture content in skimmed milk powder in a short-run batch production process. A total of 31 batches of length between 15 and 25 samples were examined. Moisture percentage, being a compositional fraction, is found to follow the beta distribution. The powder production process is a highly automated one and also very precise. This is confirmed by the large precision parameter $\theta = 29384$ fitted for the pooled data from the 31 batches. Individual means and precision parameters of moisture percentage for the 31 batches are presented in Table 2.2. A maximum of 4% of moisture content was prescribed as the upper specification limit. The target in-control process mean is calculated as $\mu = 3.79\%$ to correspond to the fixed apparent AQL of 3%. This means that the moisture percentage is assumed to follow Beta$(\mu\theta, (1-\mu)\theta)$, where $\mu = 3.79\%$ and $\theta = 29384$. Assuming $\alpha = 5\%$ and a historical value of $k = 0.25$, the four fractional nonconformance statistics and the corresponding UCLs were calculated. A sample set of the four fractional nonconformance control charts (batch 12) is presented in Figure 2.8. None of four monitoring schemes triggered a signal in the 31 batches. This is not surprising because the process mean levels for all the 31 batches were kept well below the target. In other words, the process was well under control for all the production runs by maintaining a very low fractional nonconforming level. Then we artificially increased the mean of batch 12 to 3.86%, which is higher than the target process mean. As shown in Figure 2.9, $\hat{p}_{ic}$ is more sensitive than $\hat{p}_{iu}$ while the performance of $\hat{p}_{ac}$ and $\hat{p}_{au}$ statistics are similar. This confirmed our findings in Section 2.4.

Batch	$\hat{\mu}$	$\hat{\theta}$
1	3.67%	28956
2	3.69%	58922
3	3.66%	89799
4	3.58%	39285
5	3.63%	26104
6	3.65%	39290
7	3.65%	82624
8	3.66%	42577
9	3.65%	162220
10	3.71%	66670

(a) Batches 1-10

Batch	$\hat{\mu}$	$\hat{\theta}$
11	3.70%	366158
12	3.79%	91990
13	3.65%	36529
14	3.71%	88945
15	3.70%	99631
16	3.68%	156347
17	3.66%	156115
18	3.74%	79149
19	3.56%	29060
20	3.55%	40176

(b) Batches 11-20

Batch	$\hat{\mu}$	$\hat{\theta}$
21	3.59%	85339
22	3.52%	59474
23	3.61%	68402
24	3.59%	101214
25	3.69%	297662
26	3.62%	60491
27	3.64%	75858
28	3.66%	280560
29	3.48%	97587
30	3.44%	56296
31	3.45%	53795

(c) Batches 21-31

Table 2.2 $\hat{\mu}$ and $\hat{\theta}$ for 31 batches

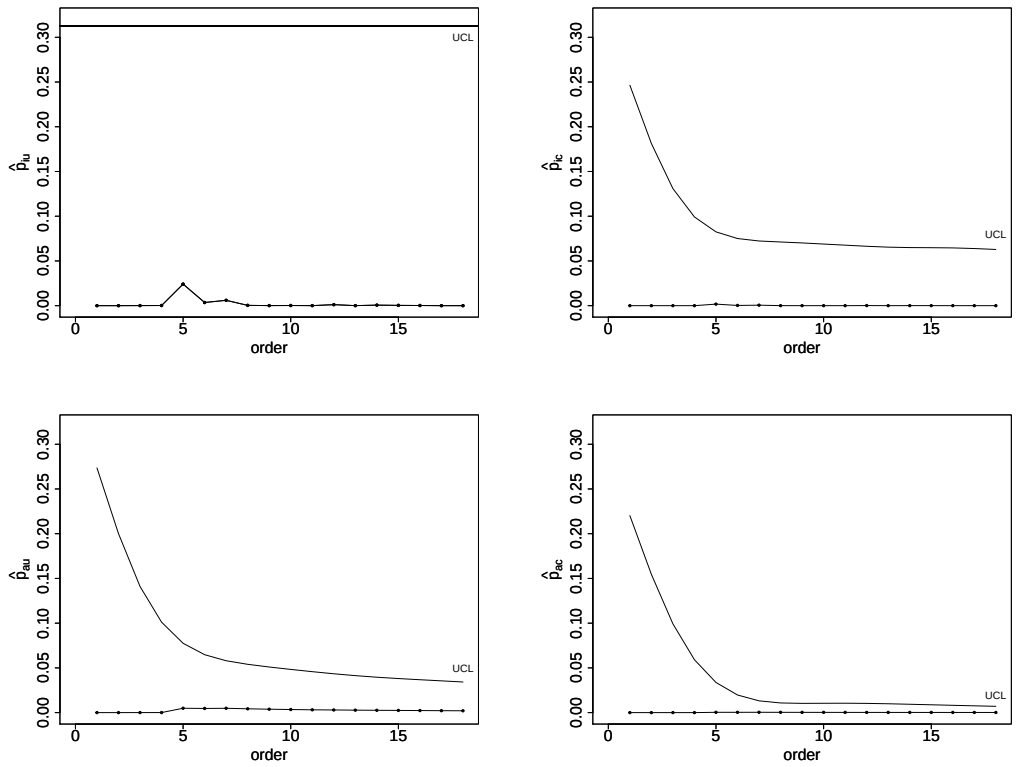


Fig. 2.8 Four fractional nonconformance control charts for batch 12 (without shift)

2.6 *Shiny* web app

An interactive web-based app for fractional nonconformance charting, made with *Shiny* R software package of Chang et al. [22] is hosted at https://zhouxin07.shinyapps.io/fncplot/.

This app will produce control charts based on the four fraction nonconformance statistics as well as the one-sided I-chart for individual measurements. It should be noted that the I-chart can only detect the shifts in the mean process level but not the process fraction nonconforming. Since the quality measure is different for the fractional nonconformance charts and I-chart, the comparison of their average run lengths is not completely meaningful. For example, the true process level may be moving downward from the upper specification limits, which will decrease the fractional nonconformance and the factional nonconformance chart will not trigger a signal. On the other hand, a chart for monitoring the process mean level will trigger a signal. Note that the control charts presented in the app do not incorporate a lower control limit to detect quality improvement. The sample size required to detect downward changes in the fraction nonconforming will be very large (Xie et al. [168]) and hence only one-sided control charts are presented in the app. The main purpose of the app is to compare the performance of the four fractional nonconformance chart with charts useful for process mean level monitoring.

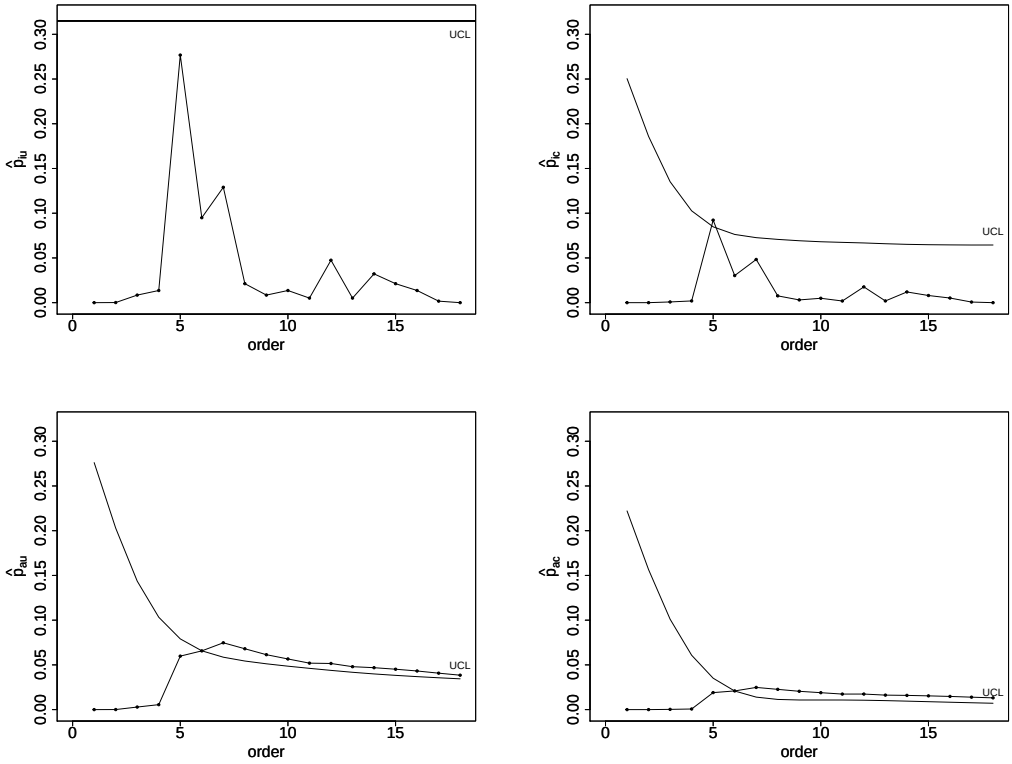


Fig. 2.9 Four fractional nonconformance control charts for batch 12 (with shift)

2.7 Conclusions

In this article, we studied four fractional nonconformance control charts for short-run production under various process models. The simulation results showed that the chart performance becomes stable beyond a production length of 15 or more. For independent normal/beta models as well as stationary autocorrelated processes, individual conditional and average unconditional fractional nonconformance monitoring schemes are preferred. For non-stationary process, individual conditional and average unconditional approaches are recommended for detection of large and small shifts respectively.

The process variance is required to be stable and known for implementing the fractional nonconformance monitoring. This assumption, even though justifiable for bulk product manufacturing processes due to the use of composite sampling techniques, may not be valid for discrete item manufacturing processes. We plan to investigate the case of unknown variance and schemes such as exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) charts in our future research.

Chapter 3

FNC Control Charts for Bulk Material

3.1 Abstract

Bulk material production processes are often monitored using a series of individual lab measurements, which may include considerable measurement errors. *Fractional nonconformance*, which measures the probability of an individual error-prone observation breaching the specification limits, was recently developed to mitigate the effect of measurement errors. Motivated by milk product manufacturing, the fractional nonconformance approach was implemented as a Shewhart chart for short runs in an earlier work. In this study, we developed fractional nonconformance based EWMA and CUSUM charts. We found that the EWMA and CUSUM charts are more sensitive than the Shewhart chart for short run monitoring of fractional nonconformance at all shift sizes, which is contrary to the theory established for traditional control charts in the literature.

Keywords

control chart; fractional nonconformance; individual measurement; measurement error; short-runs

A part of this Chapter was presented at the 2018 International Symposium on Business and Industrial Statistics, held at University of Piraeus, Greece (organized by the International Society for Business and Industrial Statistics, an umbrella organisation of the International Statistical Institute).

3.2 Introduction

Control charts are widely used in industry for monitoring key quality characteristics of products and process performance. Traditional control charting methods, including Shewhart, exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) chart, require two phases of implementation. In Phase I, a minimum of 20-25 samples with subgroups of size 4-5 are drawn, as recommended by Montgomery [111], to estimate the process parameters and set up control limits. The process performance is then progressively monitored in Phase II. Hence, the traditional control charts can only be implemented after a certain period of time, or when the historical data are available. However, this condition may not be met in certain industries. For example, only a low volume might be produced in the manufacture of some semiconductor parts according to customer order size. In milk product manufacturing, thousands of tons of powder are produced with repeated stops and starts as the machines need to be cleaned frequently for safety and hygiene reasons. These production cycles are often short-run, hence the traditional control charting procedures cannot be applied. Self-starting charts were first introduced by Hawkins [59] for start-up processes, where the running mean and standard deviation are employed to estimate the process parameters. Quesenberry [122, 123, 124] proposed Q charts to deal with short runs. The Q statistic is calculated from the first observation of the process and then updated once a new observation is obtained. Q charts are self-starting, hence helpful to monitor short runs or start-up processes without Phase I data. Many authors have studied the statistical properties of Q charts and provided some improvements; see Castillo and Montgomery [19], He et al. [64] and Zantek [172]. Zhang et al. [174] proposed t and EWMA t charts to monitor the process mean when the in-control process standard deviation is unstable or mis-estimated, and these charts have been implemented for short runs by Celano et al. [21].

Measurement errors are common for bulk material testing and may affect the control chart's performance. The effect of measurement error on the Shewhart, EWMA and CUSUM charts has been studied by many researchers; see Linna and Woodall [99], Maravelakis et al. [105], Maravelakis [104] and Hu et al. [72]. However, the effect of measurement errors on short-run control charts is rarely reported in the literature, especially for single measurements. Bulk material testing often yields a single lab measurement with a sizable measurement error. The traditional control charts such as the I-chart, DNOM-chart, and p-chart do not work well when only individual error-prone measurements are made. Govindaraju and Jones [51] proposed a new fractional acceptance number sampling plan, where the *fractional nonconformance* (FNC) approach was first introduced to allow for measurement errors in single measurements when the measurement error distribution is known. This study is motivated by milk powder manufacturing, where monitoring the process fraction nonconforming level is particularly important as every single block of the production must comply with the specifications. For

short-run individual observations involving measurement errors, a series of Shewhart type fractional nonconformance charting has been recently developed by Zhou et al. [175].

Past research has established that the Shewhart chart is more powerful for detecting large shifts, while the EWMA and CUSUM charts are more effective for detecting small shift size. In other words, no unique chart is available to optimally detect small as well as large shifts. In this Chapter, we implement the fractional nonconformance approach in the form of EWMA and CUSUM charts under various process models including autocorrelated processes. Residual chart and other modified control chart schemes are commonly used to accommodate autocorrelated data. Alwan and Roberts [6] proposed to fit a time series model for the common cause effect, and to display the fitted model in a time series plot. The residuals of the fitted model are then used for control charting to detect special causes. However, the residual chart did not perform well for AR(1) and positively autocorrelated processes when compared to the traditional control charts; see Harris and Ross [58], Wardell et al. [152] and Knoth and Schmid [94]. Vasilopoulos and Stamboulis [148] suggested a modified control scheme by adjusting the control chart limits of the traditional SPC chart in the presence of autocorrelation. We employ dynamic control limits to our newly proposed chart to ensure that all the control charts have equivalent in-control performance. Our study shows that the EWMA and CUSUM fractional nonconformance charts are superior to the Shewhart chart for detecting both small and large shifts. Section 3.3 provides a brief review of the concept of fractional nonconformance. Various types of control charts for short runs are compared in Section 3.4 to monitor fractional nonconformance and in Section 3.5 to monitor the fractional nonconformance as well as observed measurements when historical data is available. A case study is provided in Section 3.6, followed by a brief discussion in the last section.

3.3 Review of fractional nonconformance

A brief review of fractional nonconformance is provided in this section (to preserve the publication format). Table 3.1 lists all the symbols used.

Measurement errors are often not ignorable for bulk materials due to indirect analytical measurements made on them. Hence, an error-prone measurement cannot be classified as conforming or nonconforming with 100% confidence. In contrast to direct categorization of an observed measurement as conforming or not, Govindaraju and Jones [51] proposed the fractional nonconformance approach. This approach allocates a probability of nonconformance to each observed measurement.

Let $Y = X + Z$ be the measurement error model, where Z is the error in measurement with known distribution $Z \sim N(0, \sigma_Z)$, and X and Y are the true (unobserved) and observed measurements respectively. For a random sample of observed measurements $(y_1, y_2, ..., y_n)$ and a given upper specification limit U, the fractional nonconformance for y_i is defined as follows:

Table 3.1 Notations

Symbol	Description
Y	Observed measurement, $Y \sim N(\mu_Y, \sigma_Y)$ or Beta(a,b) or ARIMA(p,d,q)
Z	Measurement error, $Z \sim N(0, \sigma_Z)$
X	True measurement
N	Production length of short-run process
AQL	Acceptance quality limit
U	Upper specification limit (set corresponding to an AQL=3%)
k	Variance ratio, $k = \sigma_Z^2/\sigma_Y^2$
μ_0	In-control process mean
μ_1	Out-of-control process mean
t	Process shift time
$\hat{p}_i$	Fractional nonconformance
UCL_S	Upper control limit of the Shewhart chart
UCL_E	Upper control limit of the EWMA chart
TARL	Truncated average run length
q	Probability of getting a signal
α_S	Probability of a Type I error of the Shewhart chart
β_S	Probability of a Type II error of the Shewhart chart
λ	Smoothing parameter of the EWMA chart
L	Width of control limits of the EWMA chart
H	Decision constant of the CUSUM chart
K	Reference value of the CUSUM chart

$$\hat{p}_i = P(x_i > U) = P(z_i < y_i - U) = \Phi\left(\frac{y_i - U}{\sigma_z}\right) \tag{3.1}$$

As illustrated, $\hat{p}_i$ measures the probability of each true measurement x_i exceeding U. Thus the fractional nonconformance quantifies an observation's nonconformance level in probabilistic terms, rather than the observations being classified as conforming or nonconforming. When the observed measurement Y is close to U, the probability of nonconformance increases or the fractional nonconformance statistic increases.

The fractional nonconformance approach was initially proposed for acceptance sampling inspection by Govindaraju and Jones [51]. For example, Table 3.2(a) shows 16 hourly measurements of moisture content in a whole milk powder short run production process. The upper specification limit of moisture is 4% and none of the observed measurements is nonconforming. Assuming measurement error $Z \sim N(0, 0.0003)$ based upon calibration studies, the fractional nonconformance statistic of each measurement is calculated using Equation 3.1 and listed in Table 3.2(b).

Zhou et al. [175] applied the fractional nonconformance approach in the form of a Shewhart chart to monitor the process fraction nonconforming. None of the observations in Table

0.0388	0.0391	0.0390	0.0390
0.0389	0.0394	0.0396	0.0388
0.0394	0.0396	0.0398	0.0386
0.0391	0.0395	0.0393	0.0387

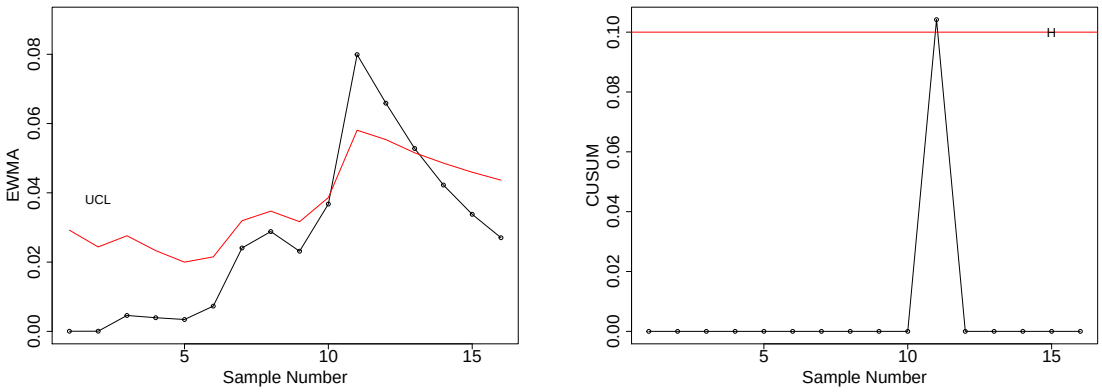
(a) Observed measurement

0.0000	0.0013	0.0004	0.0004
0.0001	0.0228	0.0912	0.0000
0.0228	0.0912	0.2525	0.0000
0.0013	0.0478	0.0098	0.0000

(b) Fractional nonconformance

Table 3.2 Moisture measurement in a short-run whole milk powder production process with upper specification limit of 4% (read data by columns for observation order).

3.2(b) falls beyond the upper control limit (UCL=0.32, $\alpha_S = 5\%$) of the Shewhart fractional nonconformance control chart, which implies that the process fraction nonconforming is in-control, so every part of the production lot can be shipped to the customer. The properties of Shewhart-type fractional nonconformance control charts were investigated by Zhou et al. [175].



(a) EWMA Chart

(b) CUSUM Chart

Fig. 3.1 The EWMA and CUSUM charts for fractional nonconformance of moisture

Besides the Shewhart chart, the EWMA and CUSUM charts are also commonly used because these charts are shown in the literature to be more powerful for detecting small shifts. Considering the example in Table 3.2(a) again, the EWMA and CUSUM charts for fractional conformance statistics in Table 3.2(b) are plotted in Figure 3.1. An out-of-control signal is observed in the 11th measurement by both EWMA and CUSUM charts. To improve fractional nonconformance monitoring, we further investigate the fractional nonconformance statistics in the form of EWMA and CUSUM charts. The comparison of Shewhart vs EWMA/CUSUM charts for short run monitoring of fractional nonconformance, as well as observed measurements, are presented in the next two sections. A comparison is not made between EWMA and CUSUM charts because their properties are quite similar according to Montgomery [111].

3.4 Comparison of control charts for short run monitoring of FNC

One of the key performance measures of a control chart procedure is average run length (ARL), which is the average number of observations until any signal is triggered. However the traditional ARL is not very meaningful for short runs because there might be no signal throughout the entire short run production. For such situations, the probability of getting a signal (q) within a specified number of N samples was suggested by Quesenberry [124] for short-run monitoring, and is defined for the Shewhart chart as follows:

$$q_0(N) = 1 - (1 - \alpha_S)^N \tag{3.2}$$

$$q_\delta(N) = 1 - \beta_S^N \tag{3.3}$$

where α_S is the probability of a Type I error or false alarm rate and β_S is the probability of a Type II error; N is the short run production length and δ denotes the level of process shift. Here $q_0(N)$ and $q_\delta(N)$ are used to evaluate the in-control and out-of-control performance of a control chart respectively. When the in-control $q_0(N)$ is matched, the control chart with highest out-of-control $q_\delta(N)$ is considered most sensitive to monitor the process. Note that Equations 3.2 and 3.3 are not valid for the EWMA/CUSUM charts, due to lack of independence in successive values, but the q performance can be obtained by a large simulation after allowing for the lack of independence in the plotted statistic.

To compare the performance of different fractional nonconformance control charts, $N = 20$ is fixed in this section. The production length N for short run may vary between industries, and hence, we also consider $N = 10$ and 40. These results are given in section B of the online supplemental material, available at https://github.com/zhouxin07/EWMA.

3.4.1 EWMA chart vs Shewhart chart

The ordinary EWMA statistic is defined as below:

$$Z_n = \lambda X_n + (1 - \lambda) Z_{n-1}, \qquad 0 < \lambda \leqslant 1 \tag{3.4}$$

where λ is the smoothing parameter and $Z_0 = \mu_0$ is the target in-control process mean. When X_n are i.i.d. with variance σ_X^2, the standard deviation of the EWMA statistic is given by:

$$\sigma_{Z_n} = \sigma_X \sqrt{\frac{\lambda}{2 - \lambda}[1 - (1 - \lambda)^{2n}]}, \qquad n \geqslant 1 \tag{3.5}$$

However, the available knowledge of the true mean level is limited for short-run production and it can be only estimated after obtaining the first observation, where we assume $Z_0 = X_1$. Hence the standard deviation of the EWMA statistic becomes:

$$\sigma_{Z_1} = \sigma_X \tag{3.6a}$$

$$\sigma_{Z_n} = \sigma_X \sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2(n-1)}] + (1-\lambda)^{2(n-1)}}, n \geqslant 2 \tag{3.6b}$$

The upper control limit of the short run EWMA control chart is therefore defined as:

$$\mathrm{UCL}_E = \mu_0 + L\sigma_{Z_n} \tag{3.7}$$

where L is the width of the control limit.

For the traditional EWMA chart, σ_{Z_n} calculated from Equation 3.5 increases and approaches $\sigma_{Z_n} = \sigma_X \sqrt{\frac{\lambda}{2-\lambda}}$ when n gets larger, which means the control limits for the traditional EWMA chart will converge to steady-state values after a few time periods. According to Equation 3.6, σ_{Z_n} obtained for short-run EWMA statistics drops when n increases. Therefore, in contrast to the traditional EWMA chart, control limits for the short-run EWMA chart become more stringent when n accumulates. Moreover, UCL_E also depends on initial observation of the process and hence is updated after another new observation is obtained. In other words, the limits in Equation 3.7 are not only variance-adjusted, but also dependent on past data. The EWMA chart for short-run monitoring of fractional nonconformance is implemented by replacing X_n with $\hat{p}_i$.

Of the two parameters, L controls the in-control performance q_0 and λ assigns the weight to the current observation. Generally speaking, the smaller the λ, the better the ability to detect a small level shift. A large λ is suitable for detecting a large shift in the process level. The performance of the EWMA chart may be affected by many factors: the magnitude of the shift, the smoothing parameter λ, and the set false alarm rate. We only investigate a process level shift at time $t = 0$ (initial-state shift) in this study as we are more interested in how quickly the control chart can detect the shift after its occurrence; the time at which the process goes out of control is not considered.

A false alarm rate for a single plotted point is not meaningful when consecutive values of the control statistic are not independent. The control limits should be adjusted to maintain an overall constant false alarm rate for the control chart. For the Shewhart chart, we adopt a fixed UCL for the independent process and dynamic UCLs for autocorrelated processes. However, this approach is not applicable for the EWMA design in our study because μ_0 is not available in the short-run process and is estimated dynamically. Hence, we cannot obtain L_i, $i = 1, 2, ..., n$, in Equation 3.7 keeping the false alarm rate constant at each observation. Alternatively, we design the EWMA chart based on the q performance, which means L is selected so that the EWMA chart has an equivalent in-control q_0 to the Shewhart chart. Therefore the chart with higher out-of-control q_δ is superior. The same performance evaluation approach is adopted for the CUSUM chart.

Magnitudes of the shift ranging from 0 to 3σ, with a step size of 0.1, are considered along with λ=0.2 (small), 0.5 (medium) and 0.8 (large). Note that $\lambda = 1$ yields a Shewhart chart. The EWMA chart under various process models, including normal, beta, stationary and non-stationary are studied and compared with the Shewhart chart.

Monte Carlo simulation Monte Carlo simulation was employed since the distribution of fractional nonconformance is intractable algebraically. The simulation algorithms adopted to design fractional nonconformance based Shewhart/EWMA charts, and obtain their q performance are described below:

Step 1. Design for the Shewhart chart

- Step 1.1 Generate a random sample of data with the assumed in-control process mean μ_0 for a given $N = 20$ samples each of size one. Given the upper specification limit, compute the fractional nonconformance statistics $\hat{p}_i$ for each single observation using Equation 3.1.
- Step 1.2 Obtain the empirical distribution of $\hat{p}_i$ by repeating Step 1.1 (100,000 runs). UCL_S of the Shewhart chart is uniquely defined by α_S. For a given α_S, calculate in-control q_0 of the Shewhart chart with Equation 3.2.
- Step 1.3 Generate random samples of data from a shifted process with shift size from 0 to 3σ. Calculate $\hat{p}_i$ and compare it with UCL_S to see if there is any signal for the Shewhart chart.

Step 2. Design for the EWMA chart

- Step 2.1 Given the empirical distribution of $\hat{p}_i$ in Step 1.2, calculate the standard deviation of the EWMA using Equation 3.6.
- Step 2.2 In-control q_0 of the EWMA chart can be simulated with the same 100,000 runs as in Step 1.2. Given q_0 of the Shewhart chart, grid search L to ensure the EWMA chart has an equivalent q_0. Calculate UCL_E of the EWMA chart with Equation 3.7.
- Step 2.3 Calculate the EWMA $\hat{p}_i$ using the data from Step 1.3 and compare it with UCL_E to see if there is any signal for the EWMA chart.

Step 3. Repeat Steps 1.3 and 2.3 100,000 times for each shift size to obtain the probability of getting a signal at each process shift level. Plot q against process shift size for both Shewhart and EWMA charts.

Normal process

First, we consider a normal process. For simplicity, it is assumed that the quality characteristic observed in the short-run process $Y \sim N(0,1)$ when the process is in-control, and the measurement error $Z \sim N(0,0.5)$ also follows a normal distribution, so that variance ratio $k = \sigma_Z^2/\sigma_Y^2 = 0.25$. The in-control performance q_0 of the Shewhart chart is calculated as 2% with Equation 3.2 when $\alpha_S = 0.1\%$. (Other q_0 values corresponding to different α_S are also considered and presented in the online supplemental material.) An equivalent q_0 for the EWMA chart is matched by simulation when L is set for each different λ; see Table 3.3. Hence, we only need to compare the out-of-control performance of the two charting methods. In other words, we aim to find the more sensitive approach with higher q_δ when the process shifts at the beginning ($t = 0$) of the process.

Table 3.3 λ and L for the EWMA chart

λ	0.2	0.5	0.8
L	3.2	5.9	6.9

By following the steps in the simulation algorithms, the q curves of the EWMA chart with different control chart parameters, likewise the Shewhart chart, were obtained as shown in Figure 3.2. We found that the EWMA charts are more sensitive than the Shewhart chart for detection of both small and big shifts (up to 3σ), no matter which λ is adopted. In other words, the EWMA charts are superior to the Shewhart chart for short run monitoring of fractional nonconformance. However, this advantage weakens when the in-control q_0 is large, as the q curves of the EWMA and Shewhart charts become less distinguishable; see section A in the online supplemental material.

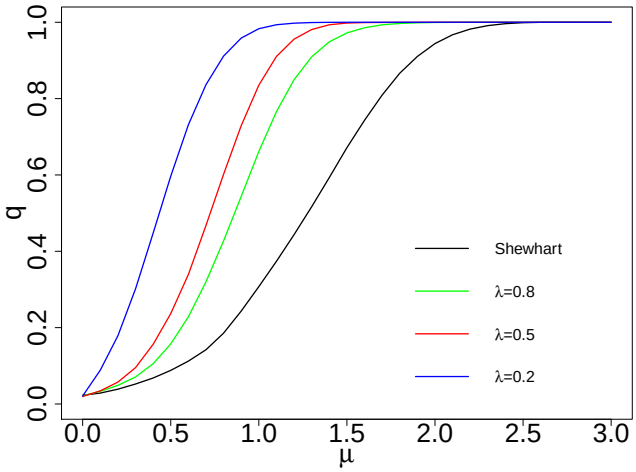


Fig. 3.2 q performance of EWMA vs Shewhart chart for normal process $N(0,1)$

Beta process

In the food industry, compositional fractions such as percentage protein or fat in milk powder are often found to follow a beta distribution. Hence, we also examine the performance of the EWMA chart in monitoring beta processes. Let the observed measurement $Y \sim \text{Beta}(a,b)$ with the following density function:

$$\text{Beta}(a,b) = \frac{1}{\Gamma(a)\Gamma(b)/\Gamma(a+b)} y^{a-1}(1-y)^{b-1} \tag{3.8}$$

where $0 < y < 1$, $a,b > 0$. $\text{Beta}(a,b)$ can also be re-parameterized as $\text{Beta}(\mu\theta,(1-\mu)\theta)$, where $\mu = \text{E}(Y) = \frac{a}{a+b}$ is the mean and $\theta = a+b$ is the precision parameter.

Suppose that the quality characteristic $Y \sim \text{Beta}(500,1500)$ represents an in-control process with process mean $\text{E}(Y) = \mu_0 = \frac{a}{(a+b)} = 0.25$ and process variance $\text{Var}(Y) = \frac{ab}{(a+b)^2(a+b+1)} \approx 0.0001$. Because the precision θ is large, the beta process $Y \sim \text{Beta}(500,1500)$ has similar process properties to the normal process $Y \sim N(0.25, 0.01)$. This assumption is justifiable because the dairy production process is highly automated with relatively high precision parameter θ; see Govindaraju and Kissling [52]. We also assume the variance ratio $k = 0.25$ and measurement error $Z \sim N(0, 0.005)$. The mean of a beta process is related to a, for example $a = 500$ corresponds to the in-control process with $\mu_0 = 0.25$, while $a = 560$ represents a process with a 3σ shift, with out-of-control process mean $\mu_1 = 0.28$.

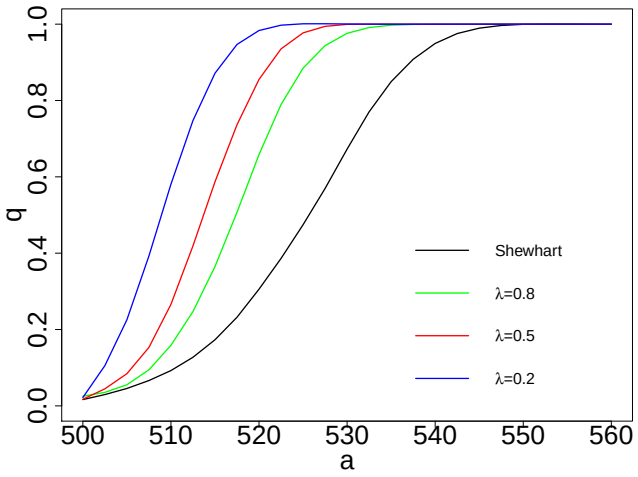


Fig. 3.3 q performance of EWMA vs Shewhart chart for beta process $\text{Beta}(500,1500)$

The performance of the EWMA and Shewhart charts under a beta process can be obtained by following the same simulation steps. As illustrated in Figure 3.3, the EWMA charts are always superior to the Shewhart chart in detecting both small and large shifts, no matter what the choice of λ is. Moreover, q curves of the EWMA and Shewhart charts under the

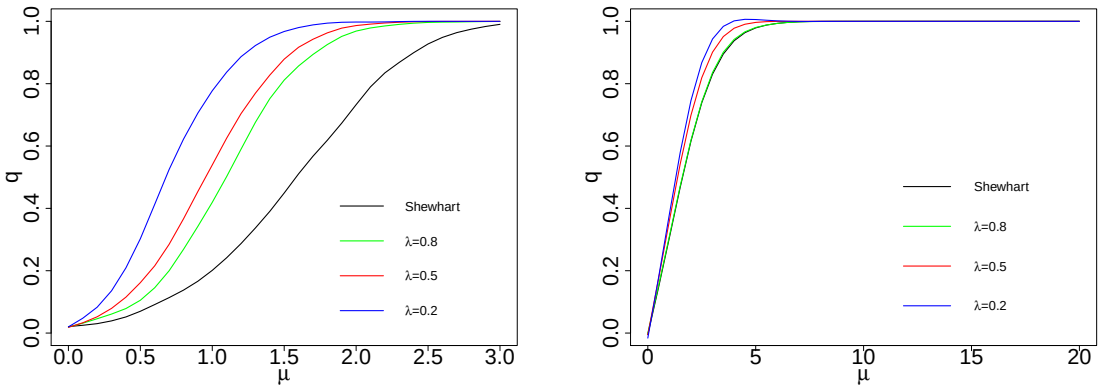
beta process are consistent with the curves under the normal process. In other words, the chart performance does not depend much on the normality assumption.

The beta process with a lower precision parameter θ, say $Y \sim \text{Beta}(1,49)$ with process mean $\text{E}(Y) = \mu_0 = 0.02$ and process variance $\text{Var}(Y) \approx 0.0004$ is also investigated. Similar results are obtained and provided in section C in the online supplemental material.

Autocorrelated processes

In the previous sections, we compared the performance of the EWMA and Shewhart charts for normal and beta processes, where we assumed the measurements are independently and identically distributed. However, process measurements may not be independent, in particular when samples are taken very frequently. Autocorrelated processes including stationary and non-stationary processes are investigated in this section.

Assume the quality characteristic Y_t to be a first order autocorrelated process (AR(1)) defined as $y_t = \phi_1 y_{t-1} + \epsilon_t$ (stationary) or integrated moving average process (IMA(1,1)) with $y_t - y_{t-1} = \epsilon_t - \theta_1 \epsilon_{t-1}$ (non-stationary). For simplicity, let $\epsilon_t \sim N(0,1)$, $\phi_1 = 0.5$ and $\theta_1 = -0.5$, so that the in-control process mean for the AR(1) model is $\mu_0 = 0$ and process variance is $\sigma_Y^2 = 1/(1-0.5^2) = 1.33$. When Y_t is non-stationary (IMA(1,1)), the process mean and variance are not fixed. We compare the q performance of the EWMA and Shewhart charts by shifting the process mean from $\mu_0 = 0$ to $\mu_1 = 3$ for the stationary process or inserting a shift function with size between [0, 20] for the non-stationary process.



(a) AR(1) with $\phi_1 = 0.5$

(b) IMA(1,1) with $\theta_1 = -0.5$

Fig. 3.4 q performance of EWMA vs Shewhart chart for autocorrelated processes

Figure 3.4(a) delivers similar information to Figure 3.2, which shows that the EWMA charts are also more powerful than the Shewhart chart for detecting any shift size when the process is stationary. This demonstrates the similar properties for normal and stationary autocorrelated processes because y_t also follows a normal distribution when $\epsilon_t \sim N(0,1)$. In

contrast to independent and correlated stationary processes, Figure 3.4(b) demonstrates that the q curves of the EWMA and Shewhart charts are almost overlapped for the assumed non-stationary process, which indicates similar sensitivity for both charting methods.

Besides q, the truncated ARL (TARL) is another useful statistical measure for short run monitoring; see Nenes and Tagaras [113] and Celano et al. [21]. When the production length is limited to N, TARL is defined as the actual length at which the first signal occurred or $N+1$ if no signal is observed, and TARL can be approximated using a very large scale simulation. We evaluated the TARL performance of the EWMA and Shewhart charts and presented the results in section D of the online supplemental material. Both the TARL and q performances were found to be very similar.

3.4.2 CUSUM chart vs Shewhart chart

The cumulative sum (CUSUM) statistic is defined and operated as below:

$$C_n^+ = \max\{0, C_{n-1}^+ + (X_n - \mu_0) - K\} \tag{3.9}$$

with the process declared out-of-control when $C_n^+ > H$, where H is the decision constant, X_n denotes the measurement (here the FNC statistic $\hat{p}_i$), μ_0 is the in-control process mean, K is half of the expected shift size; $C_0^+ = 0$ when there is no head start while $C_0^+ = c > 0$ denotes a head start CUSUM. For short-run conditions, because the historical in-control process mean is not available, μ_0 is replaced with a moving average of the process.

Like the EWMA chart, the CUSUM chart also has two parameters: K, the reference value, controls the anticipated shift size, and H sets the in-control performance q_0. We considered reference values of $K = 0.1, 0.2, 0.4$ and the corresponding values of H, as listed in Table 3.4, to match the in-control q_0 of the Shewhart chart. For different combinations of K and H, the q performance of the CUSUM charts under various process models are obtained by a simulation algorithm similar to the one employed for the EWMA charts.

Table 3.4 K and H for the CUSUM chart

K	0.1	0.2	0.4
H	0.91	0.69	0.48

Normal process

As before, assume $Y \sim N(0,1)$ as the in-control process and measurement error $Z \sim N(0, 0.5)$. We add a shift ranging from 0 to 3 at the beginning of the process. The q curves of the CUSUM chart and the Shewhart chart are plotted against shift size in Figure 3.5. It shows that the CUSUM charts outperform the Shewhart chart for both small and large shifts, and the smaller the reference value K, the better the performance of the CUSUM

chart. As for the EWMA chart, the superiority of the CUSUM chart becomes weakened when the in-control q_0 gets larger; see section A in the online supplemental material at https://github.com/zhouxin07/EWMA.

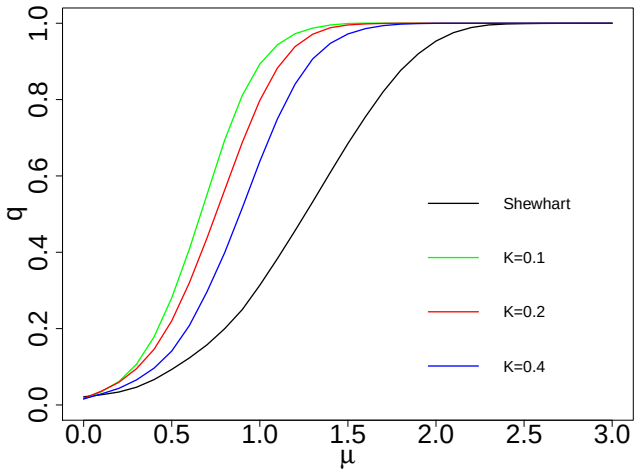
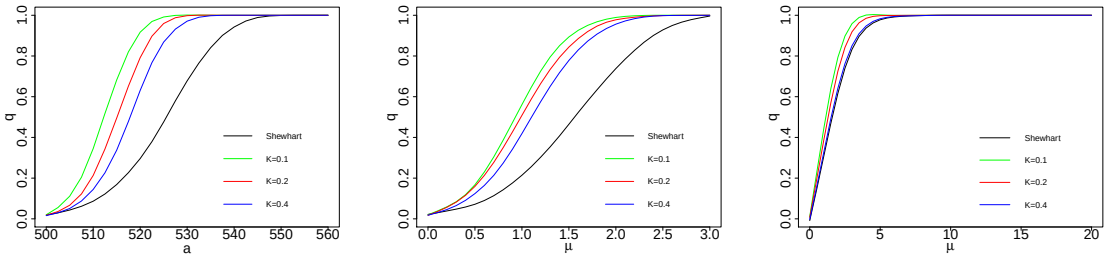


Fig. 3.5 q performance of CUSUM vs Shewhart chart for normal process $N(0,1)$

Other processes

The performance of the CUSUM chart is also investigated under beta, AR(1), and IMA(1,1) processes. The comparisons are displayed in Figure 3.6. These results further demonstrate that the Shewhart chart is also not as sensitive as the CUSUM chart when an initial shift occurs in the process and this inferiority does not depend on the size of shift or the distribution of the quality characteristics.



(a) Beta(500, 1500) (b) AR(1) with $\phi_1 = 0.5$ (c) IMA(1,1) with $\theta_1 = -0.5$

Fig. 3.6 q performance of CUSUM vs Shewhart chart for other processes

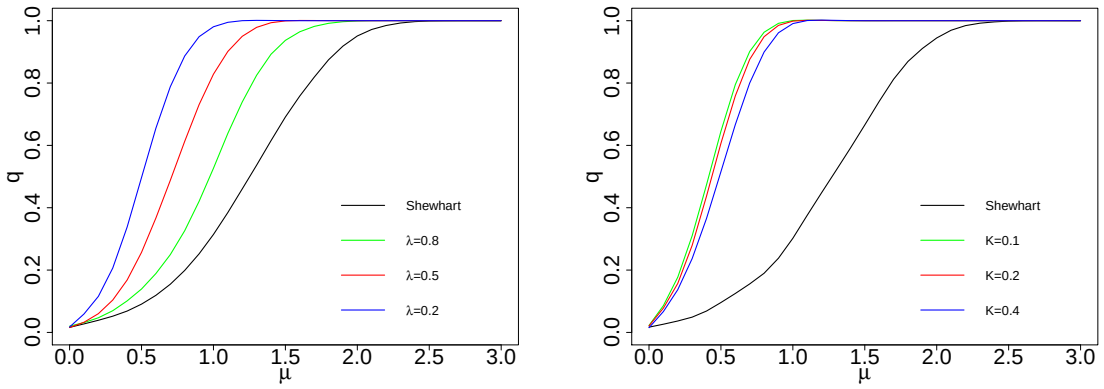
3.5 Monitoring individual values and FNC with historical data

The general agreement in traditional control charting is that the Shewhart chart is less competitive for detecting small shifts since it does not accumulate the available past information (as the EWMA/CUSUM chart does) but is more sensitive for detecting large shifts; see Montgomery [111] and Abbas et al. [1]. In the previous sections, fractional nonconformance $\hat{p}_i$ was implemented in the form of EWMA and CUSUM charts for short run process monitoring. Surprisingly, the comparison of the EWMA and CUSUM charts vs the Shewhart chart showed that the expected superiority of the Shewhart chart in detection of large shifts is absent. In other words, the EWMA and CUSUM charts are always superior to the Shewhart chart, no matter the shift size for short run monitoring of fractional nonconformance.

To investigate whether this is a universal advantage of the EWMA/CUSUM chart in short run monitoring or the unique property of fractional nonconformance, we applied the EWMA, CUSUM and Shewhart control charting procedures for short run monitoring of individual values, as well as the traditional control charting procedure for fractional nonconformance.

3.5.1 Short-run monitoring of individual values (I-chart)

The fractional nonconformance statistics $\hat{p}_i$ were monitored in the previous section in the form of EWMA, CUSUM and Shewhart charts by following the simulation steps in Section 3.4. Alternatively, the observed individual measurements Y can be monitored directly using the EWMA, CUSUM and Shewhart charts. However, in contrast to fractional nonconformance, the in-control distribution of the observed measurement Y is assumed to be known to establish control limits for the control charts. This assumption is valid when extensive in-control historical data is available.



(a) EWMA (b) CUSUM

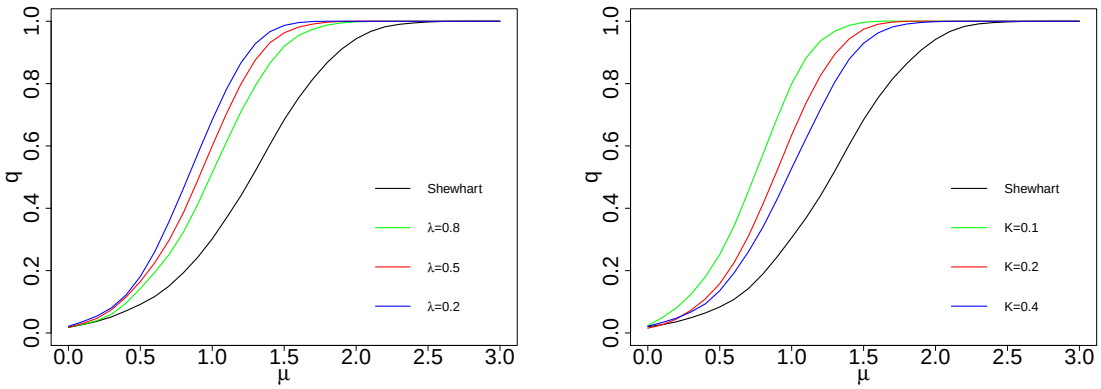
Fig. 3.7 q performance of EWMA/CUSUM vs Shewhart chart (Individual values)

Let $Y \sim N(0,1)$ be the quality characteristics observed in the short-run process, which is also the true measurement when measurement error is absent. We shift the process from $Y \sim N(0,1)$ to $Y \sim N(3,1)$. The q performance of various EWMA, CUSUM and Shewhart charts are shown in Figure 3.7. It is observed that the EWMA/CUSUM chart can identify any shift size quicker than the Shewhart chart. This is congruent with the results obtained in Section 3.4.

3.5.2 Traditional control charting for fractional nonconformance

In contrast to the short-run condition discussed in Section 3.4, traditional control charting procedures can be applied to monitor fractional nonconformance when historical data is available. In Phase I, process parameters and control limits are estimated for the FNC statistic and process performance is monitored in Phase II.

By following the simulation algorithms in Section 3.4, the OC curves of the Shewhart and EWMA charts are plotted in Figure 3.8(a). It shows that EWMA chart is always more sensitive than the Shewhart chart no matter the shift size. Similar results are observed in Figure 3.8(b) for the CUSUM chart. The results show that the superiority of EWMA and CUSUM charts for monitoring of fractional nonconformance is consistent and does not depend on the availability of Phase I data.



(a) EWMA (b) CUSUM

Fig. 3.8 q performance of EWMA/CUSUM vs Shewhart chart (FNC with Phase I data)

Hence, we can conclude that the EWMA and CUSUM charts are more powerful than the Shewhart chart in detecting both small and large shifts when a production process is monitored on a short-run basis (a universal advantage).

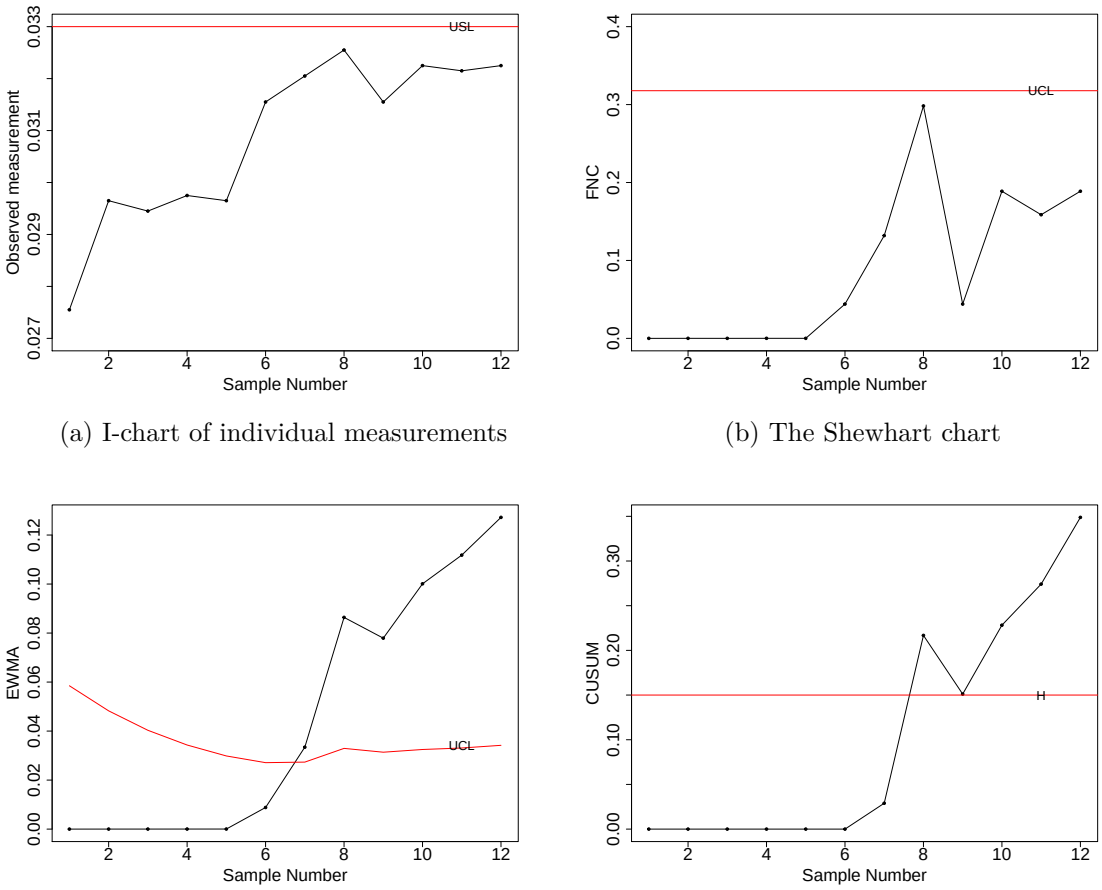
3.6 Case study

We present data from 24 batches of whole milk powder from a large dairy manufacturing plant in New Zealand, with run lengths between 4 and 20 samples, to evaluate the performance of the Shewhart, EWMA and CUSUM charts for short run monitoring of fractional nonconformance. Quality characteristics of whole milk powder, including moisture, protein, fat and P:SNF (Protein to milk Solids-Non-Fat ratio) were trialled and the summary statistics of the 24 batches are shown in section E in the online supplemental material. For moisture, the upper specification limit was set as 3.3%, and the desired in-control process level was 2.98%, which corresponds to AQL=3%. Fractional nonconformance was monitored by the Shewhart, EWMA and CUSUM charts respectively.

Table 3.5 Number of signals detected by various control charts in the 24 batches of whole milk powder after artificially introducing a shift

Shift size	Method	Quality Characteristics			
		Moisture	Protein	Fat	P:SNF
0.5σ	Shewhart	1	6	5	5
	EWMA	7	16	11	17
	CUSUM	2	8	8	10
2σ	Shewhart	19	24	23	23
	EWMA	22	24	24	24
	CUSUM	21	24	23	24

There were no signals in any of the control charts in the 24 batches, because the process level of these 24 batches is maintained well and far below the desired in-control process mean corresponding to an AQL of 3%. We artificially introduced shifts of size 0.5σ (small shift) and 2σ (large shift) to all of the data. For the small shift, the Shewhart chart triggered only 1 signal in the 24 batches, while the EWMA and CUSUM charts detected 7 and 2 signals respectively. For the large shift, the EWMA and CUSUM charts captured all those signals issued by the Shewhart chart. Two sample batches of observed measurements as well as various fractional nonconformance control charts are presented in Figures 3.9 and 3.10, which illustrate the sensitivities of the EWMA/CUSUM chart in detecting small and large shifts respectively. A similar comparison was performed for protein, fat and P:SNF compositional quality characteristics. The summary analysis results of 192 sets of data in Table 3.5 suggest that the EWMA and CUSUM charts are more powerful for short run monitoring of fractional nonconformance. This agrees with our findings in Sections 3.4 and 3.5. The implementation of the fractional nonconformance based Shewhart, EWMA and CUSUM short-run control charting is enabled using a web based *Shiny* app, which is hosted at https://zhouxin07.shinyapps.io/ewma/ (also at http://shiny.massey.ac.nz/kgovinda/ewma/).



(a) I-chart of individual measurements

(b) The Shewhart chart

(c) The EWMA chart

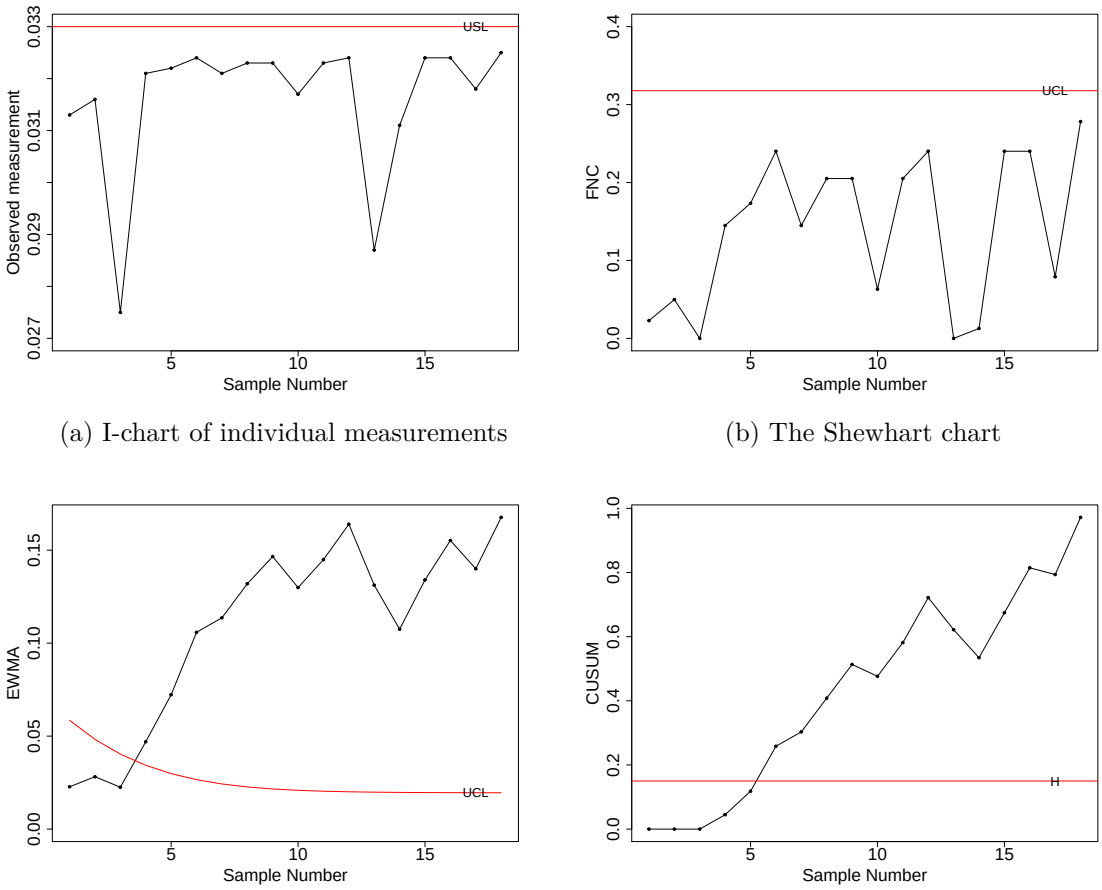
(d) The CUSUM chart

Fig. 3.9 I-chart and FNC control charts for batch 5 with small shift (0.5σ)

3.7 Discussion

The advantage of the fractional nonconformance approach for short run production is that prior knowledge of the process mean or variance is less critical so that various control charts can be implemented to monitor fraction nonconforming level from the first observation of the process. The main assumption to be met is the distribution of the measurement error, which can be reliably found from past calibration studies. Prescription of an AQL value such as 3% is also required.

In this Chapter, fractional nonconformance monitoring was implemented in the form of EWMA and CUSUM charts, and their performances were compared with the Shewhart chart. The EWMA and CUSUM charts are recommended for short run monitoring of fractional nonconformance statistic because of their sensitivity in detection of both small and large shifts when the false alarm rate is small, say $\alpha_S \leq 5\%$. When α_S is large or the process is



(a) I-chart of individual measurements

(b) The Shewhart chart

(c) The EWMA chart

(d) The CUSUM chart

Fig. 3.10 I-chart and FNC control charts for batch 21 with large shift (2σ)

non-stationary, the performances of the EWMA, CUSUM and Shewhart charts become less distinguishable, hence the simpler Shewhart chart is adequate.

This Chapter has only considered simple autocorrelated processes and more complicated ARIMA models for short run monitoring could be investigated in future research.

Chapter 4

Acceptance Control and Guardbanding for Error-prone Individual Measurements

Xin Zhou, Kondaswamy Govindaraju, Geoff Jones
Quality and Reliability Engineering International, 2019
https://onlinelibrary.wiley.com/doi/full/10.1002/qre.2416

4.1 Abstract

The concept of fractional nonconformance was recently proposed to assess the probability of conformance when measurements are error-prone. Applications of fractional nonconformance assessment include acceptance sampling inspection and short run process control. In this study, we introduce a fractional nonconformance based one-sided acceptance control chart to monitor a short run process as well as decide the acceptability of products manufactured from the process. The guardband technique is also incorporated in the proposed approach to reduce the impact of measurement errors. Guardband selection is investigated for both independent and autocorrelated processes. Our analysis shows that guardbanding is beneficial for short run production environments. The optimum guardbands obtained under risk and cost models are also found to be consistent.

Keywords

acceptance control chart; fractional nonconformance; guardbanding; measurement error; short-runs

4.2 Introduction

Statistical process control (SPC) is widely used to ensure that a production process is free from special causes of variation and achieve quality improvement. Shewhart, cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) charts are commonly used SPC tools. These traditional control charting methods require an initial phase (Phase I) to estimate the in-control state process parameters. After Phase I study, control limits are established for the prospective monitoring of the process (Phase II). This two-stage approach is difficult to implement for short production runs. Instead the process parameters are estimated dynamically and/or nominal values for parameters are employed to establish the control limits. A good introduction to short run process control can be found in Elam [36]. Bulk material production processes, such as production of dairy products, are often *short run* processes due to the need to maintain hygiene and safety. There are also numerous product specifications which limit the production batch size. For such processes, the underlying parameters are estimated dynamically, and the in-control state is then monitored in such a way that the proportion nonconforming is kept below the undesirable levels.

For bulk materials, *measurement error* is inevitable due to the variation and inaccuracy of analytical testing as well as difficulty in collecting representative samples. Only individual lab measurements are obtained in a bulk materials production process. Measurement error may cause misclassification and affect a control chart's performance. See Tran et al. [147] for a recent review of measurement error effects on control charts.

Guardbanding was initially proposed by Eagle [34] as a precautionary strategy so that the chance of breaching the specifications is kept low. Grubbs and Coon [55] used guardbanding to reduce the impact of measurement errors using a new tightened limit instead of the original specification limit. This guardband limit decreases the probability of false acceptance at the cost of raising the false rejection rate. Guardband selection has been studied by many authors; see Eagle [34], Easterling et al. [35], Chou and Chen [25], Williams and Hawkins [157] and Kim et al. [92].

Acceptance control chart, originally developed by Freund [43], is a hybrid approach incorporating control limits for process control as well as decision limits to control the process proportion nonconforming so that the resulting batches from the process can be accepted without further inspection. An introduction and review of acceptance control charts are available in Duncan [33] and Wu [167]. The traditional acceptance control methodology requires an indifference zone to be set and it does not allow for measurement errors. As a result, these traditional acceptance control chart procedures cannot be used for bulk products. Much of the current literature on acceptance control chart and guardbanding assumes that the process characteristic is normally distributed and does not deal with short run process monitoring.

Dairy products such as milk powder, butter and cheese, are naturally bulk in nature. Production of dairy and other bulk products such as pharmaceutical drugs are short-run processes. A production run of bulk products is usually capacity constrained. In food product manufacturing, the machinery must be cleaned due to safety and other reasons which also limits the production batch size or volume. Even though milk powder production process runs for about 30 hours, several thousand tons of powder are rapidly produced. Frequent sampling of the end product is also difficult because packaging is done using robotic machinery. A milk powder production process is usually sampled less than 50 times at regular intervals. We call the number of times the short-run process is sampled as the production length N. A single production run will form many consumer lots for further inspection. In other words, a production batch (run) need not be homogeneous in quality but the inspection lots formed are required to be homogeneous. For each sample of material, a single individual lab measurement, such as percentage protein, is made. For a production batch of length N, a single test result outside the specification will render the production batch inhomogeneous but parts of the production conforming to specifications can be made into inspection lots for customers. Hence the assessment of individual proportion nonconforming at the time of sampling as well as the overall proportion nonconforming are both important for the consumer's protection in the dairy industry.

Govindaraju and Jones [51] proposed a probabilistic measure for quantifying nonconformance after adjusting for measurement uncertainty when the underlying measurement error distribution is known. This fractional nonconformance statistic (FNC) was initially applied for acceptance sampling inspection, and was further implemented for short run process monitoring by Zhou et al. [175]. The objective of this work is to employ fractional nonconformance principles to acceptance control charting applications in dairy production. We develop a modified acceptance control chart scheme (MACCS) to monitor the short run process and to simultaneously dispose the products manufactured from the process. Then the optimum guardband is investigated for the MACCS scheme under both independent and autocorrelated processes involving measurement errors. The remainder of this paper is organized as follows. A brief review of the fractional nonconformance approach, guardbanding and acceptance control chart is provided in Section 4.3. The fractional nonconformance based modified acceptance control chart scheme is introduced in Section 4.4. Optimum guardband selection for the proposed scheme is presented in Section 4.5 under a risk based model, and Section 4.6 deals with a cost model. This is followed by a case study in Section 4.7 and conclusion at the end.

4.3 Review of fractional nonconformance, guardbanding and acceptance control chart

A brief review of fractional nonconformance measurement, guardbanding and acceptance control chart is provided in this section. Table 4.1 lists the notations used.

Table 4.1 Notations

Sign	Description
Y	Apparent measurement
Z	Measurement error, $Z \sim N(0, \sigma_Z^2)$
X	True measurement, $X = Y - Z$
N	Production length of short-run process
κ	Variance ratio, $\kappa = \sigma_Z^2/\sigma_Y^2$
USL	Upper specification limit, externally fixed
UGL	Upper guardbanded limit, adjustable
g	Guardband coefficient, g = UGL/USL, $g \leq 1$
AQL	Acceptance quality limit
RQL	Rejectable quality level
μ_0	In-control process mean
μ_1	Out-of-control process mean
$\hat{p}_i$	Fractional nonconformance of the i-th observation
$\bar{\hat{p}}_i$	Average fractional nonconformance
δ_I	Sum of fractional nonconformance, $\delta_I = \sum \hat{p}_i$
Ac	Fractional acceptance number
UCL_A	Upper control limit of the $\bar{\hat{p}}_i$ control chart
α_p	Probability of rejecting an in-control process
β_p	Probability of accepting an out-of-control process
α_l	Producer's risk of a sampling plan
β_l	Consumer's risk of a sampling plan
α_a	Type I error probability of the MACCS scheme
β_a	Type II error probability of the MACCS scheme
c_α	Cost of Type I error
c_β	Cost of Type II error
c	Cost ratio, $c = c_\beta/c_\alpha$
C_a	Total cost of Type I and II errors

4.3.1 Fractional nonconformance

Given an upper specification limit (USL), an individual measurement cannot be categorized with full certainty as conforming or not due to measurement errors. Measurement errors are seldom negligible, especially for analytical measurements, which are often indirect. Govindaraju and Jones [51] proposed a new fractional acceptance number sampling plan using fractional nonconformance probability measure for an observation given the measurement

error distribution. Let $Y = X + Z$ be the measurement error model, where X, Y, Z are the true measurement, the apparent measurement and the measurement error respectively. For a random sample of apparent measurements $(y_1, y_2, ..., y_n)$, fractional nonconformance of each individual observation is found as follows:

$$\hat{p}_i = P(x_i > \text{USL}) = P(z_i < y_i - \text{USL}) = \Phi\left(\frac{y_i - \text{USL}}{\sigma_z}\right) \tag{4.1}$$

Each observation is assigned a probability of nonconformance in the fractional nonconformance approach, rather than being classified as conforming or nonconforming. The batch is accepted if the overall nonconformance level, δ_I, is low.

4.3.2 Guardbanding

Guardbanding is an offset technique used to compensate for measurement and sampling uncertainty by using a new tightened limit instead of the original specification limit and thereby reducing the risk of nonconformance to specifications. The guardbanding technique was studied by many authors; see Easterling et al. [35], Healy et al. [65], and Pendrill [118]. A more stringent specification can decrease the probability of false acceptance at the cost of raising the false rejection rate. In other words, by tightening the specification artificially, the probability of Type II error and the associated costs are reduced, while the probability of Type I error increases as a matter of trade-off. Guardband width determination is expected to balance the Type I and Type II error probabilities, which can be controlled as required for risk management or based on economic considerations or both. Chou and Chen [25] developed a kernel density estimator after deconvolution of the density to adjust for the measurement error and then determined the optimal guardband limit that controls both producer's and consumer's risks. Williams and Hawkins [157] proposed a cost model under the presence of measurement errors and investigated the relationship between profitability, guardband placement and ratio of different costs. Kim et al. [92] introduced a cost model incorporating acceptance and rejection costs as well as measurement precision level so that the guardband determination is optimal.

The principle behind the guardbanding technique is illustrated in Figure 4.1. Process A represents an ordinary in-control process with observations subject to measurement errors. The guardband approach applies a tightened specification limit, called upper guardbanded limit (UGL), instead of USL as the new specification. By using the tightened specification UGL, some good quality produced from Process A may be wrongly rejected, which increases the probability of Type I error. At the same time, poor quality product will have a greater chance of detection and rejection, which lowers the probability of Type II error. The ratio of the guardband specification limit (UGL) to the original specification limit (USL) is defined as the guardband coefficient g:

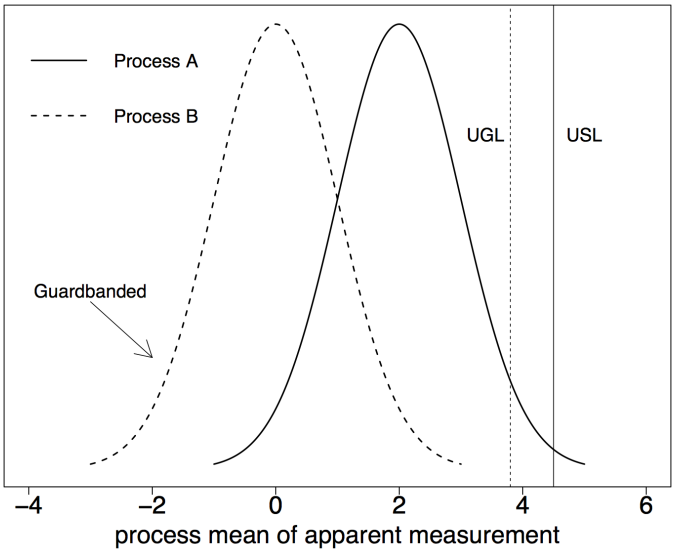


Fig. 4.1 Illustration of guardbanded process (one-sided)

$$g = \mathrm{UGL}/\mathrm{USL} \tag{4.2}$$

While the guardband limit reduces the effect of measurement errors, the guardband coefficient selection balances between Type I and Type II errors. In contrast to externally fixed USL, UGL is an adjustable limit to account for extra sampling-related variability. The optimum placement of UGL is required so that the process mean can be maintained at the optimum level (shown as Process B in Figure 4.1).

4.3.3 Acceptance control chart

A control chart is often used to ensure that the process mean is maintained at a constant level. In some cases, a drift in the process is tolerated as long as the process level is still far away from the specifications. One situation is that the natural dispersion of a process is much less than the specification spread (e.g. six-sigma process), so that the production lots are still acceptable even though the underlying process mean level is not constant. Another situation is that maintaining the process level in the middle of the specification range is expensive (for two-sided specifications) and the producer wishes to tolerate a certain amount of shift in the process, as long as the proportion nonconforming is still below the acceptable level. The acceptance control chart, originally proposed by Freund [43], is an approach based on the Shewhart control charting methodology aiming to limit the proportion nonconforming of products exceeding specifications.

The acceptance control chart can be used to monitor the proportion nonconforming at each test point, however, the overall nonconforming level of the lot (estimated from all test points) is also important, and can be evaluated by an acceptance sampling plan. Control charts mainly focus on monitoring the process levels or spread while acceptance sampling is an inspection procedure to assess the acceptability of the lots formed from a production process. We propose the MACCS scheme to serve both these objectives but without the indifference zone. We term the proposed acceptance control chart procedure as a modified acceptance control chart scheme mainly because the additional layer of indifference zone renders the scheme too complex to implement in practice. Under the proposed scheme, samples are taken directly from the production process rather than randomly selected from the final lot. The fractional nonconformance based MACCS scheme is introduced in the next section.

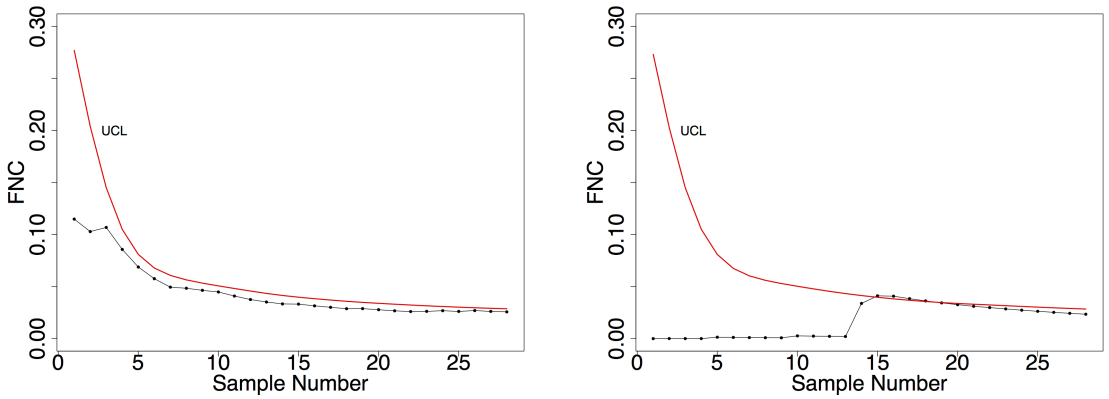
4.4 Fractional nonconformance based MACCS scheme

The proposed MACCS scheme is intended to monitor the in-control state of the process as well as dispose the products produced. Under the MACCS scheme, the batches formed are accepted only if the process is in-control and the overall proportion nonconforming level is low. Zhou et al. [175] found that the average fractional nonconformance statistic ($\bar{\hat{p}}_i$) is more sensitive than the individual fractional nonconformance statistic ($\hat{p}_i$) for detection of a sudden shift in the process. Hence, $\bar{\hat{p}}_i$ is used to monitor the stability of the process while $\hat{p}_i$ is chosen to evaluate the within process proportion nonconforming level because the product is split into lots at the end. The control chart rule for an in-control process and the rule for lot acceptance must be met at the same time. Therefore, the probability of acceptance of a short run production lot with size N under the MACCS scheme is defined as below:

$$P_a = \Pr\Big(\{\delta_I \leq N \times Ac\} \cap \{\max\{\bar{\hat{p}}_i\} < \mathrm{UCL}_A, \forall i\colon 1, ..., N\}\Big) \tag{4.3}$$

where Ac is the fractional (or broken) acceptance number, UCL_A is the upper control limit of the Shewhart $\bar{\hat{p}}_i$ control chart. $\{\delta_I \leq N \times Ac\}$ is the event that the lot or product quality does not go beyond the proportion nonconforming requirement while $\{\max\{\bar{\hat{p}}_i\} < \mathrm{UCL}_A, \forall i\colon 1, ..., N\}$ is the rule for declaring the process as in-control. Both criteria must be met at the same time to release the batch, in other words, the batch will be rejected if the sum of individual fractional nonconformance (δ_I) exceeds or any of average fractional nonconformance ($\bar{\hat{p}}_i$) at any stage breaches the control limit during the process.

Consider the following two datasets on milk cream fat composition from two short run production operations. Dataset 1: {0.5582, 0.5580, 0.5582, 0.5570, 0.5555, 0.5556, 0.5552, 0.5574, 0.5572, 0.5572, 0.5557, 0.5552, 0.5563, 0.5565, 0.5572, 0.5563, 0.5564, 0.5562, 0.5572, 0.5564, 0.5562, 0.5566, 0.5572, 0.5574, 0.5565, 0.5575, 0.5562, 0.5567} and Dataset 2: {0.5508,



(a) In-control process but δ_I is high (Dataset 1) (b) Out-of-control process with low δ_I (Dataset 2)

Fig. 4.2 Shewhart $\bar{\hat{p}}_i$ chart of two short run processes

0.5545, 0.5514, 0.5531, 0.5563, 0.5542, 0.5530, 0.5503, 0.5502, 0.5569, 0.5545, 0.5517, 0.5511, 0.5598, 0.5584, 0.5573, 0.5528, 0.5508, 0.5525, 0.5524, 0.5511, 0.556, 0.5522, 0.5520, 0.5503, 0.5502, 0.5529, 0.5545}. The upper specification of the fat content was set at 0.56. Let the measurement error $Z \sim N(0, 0.0015)$. Fractional nonconformance statistics were calculated for the two datasets, and then Shewhart $\bar{\hat{p}}_i$ charts of the two processes constructed as shown in Figure 4.2. Process (a) is well maintained at a constant level because all the plotted points are below the upper control limit. However, its overall fractional nonconformance is too high and exceeds the acceptance number Ac. Hence, the products produced from this process are not considered acceptable. Although the product quality coming from process (b) is acceptable ($\delta_I \leq N \times Ac$), a further investigation or sublotting is needed because a signal for lack of control in the process was observed, suggesting a potential shift in the process level. Based on the rules of the MACCS scheme, neither of the batches formed from the two processes is acceptable and a further final inspection is needed. The MACCS scheme proposed in this paper was designed and implemented in a real life scenario, where the overall proportion nonconforming level as well as the statistical control of individual proportion nonconforming levels within the production period was required to be kept at low levels. The MACCS scheme was specifically designed because only error-prone individual measurements were available.

4.5 Optimum guardband selection for the MACCS scheme

Guardbanding is implemented in the MACCS scheme to reduce the effect of measurement and sampling errors on the producer's and consumer's risks. The overall Type I error of the MACCS scheme (α_a) is due to Type I errors of the lot acceptance rule (α_l) and in-control process rule (α_p), which are determined by the choice of Ac and UCL_A respectively. Note that

α_l is the probability of rejecting acceptable quality product. In the control chart literature, the Type I error is usually defined as the probability of breaching the control limit for a single observation. For the sake of consistency, we define α_p as the overall probability of false alarm for an in-control process. Hence, the probability of getting a signal for each plotted point is calculated as $1-\sqrt[N]{1-\alpha_p}$, which corresponds to the value of UCL_A.

The guardband coefficient g is a function of both α_l and α_p. When g decreases from 1, a tightened limit UGL is adopted instead of USL in Equation 4.1, and as a result, both $\hat{p}_i$ and $\bar{\hat{p}}_i$ will increase. In other words, a small g leads to a high probability of Type I error in both lot acceptance and process control decisions. For a given α_a, numerous choices of α_l and α_p can be made; the same is true for the potential combinations of g, Ac and UCL_A. Hence we aim to find the optimum guardband coefficient for the MACCS scheme in terms of controlling both producer's and consumer's risks.

Monte Carlo simulation was applied since the distribution of fractional nonconformance is intractable algebraically. The simulation algorithm adopted to obtain the empirical distribution of the fractional nonconformance, the control statistic, and the optimum guardband coefficient is described below:

- Step 1. Generate a random sample of data from the in-control process, say $Y \sim N(\mu_0, \sigma_Y^2)$, for a given sample size N. Given the USL, compute the fractional nonconformance statistics $\hat{p}_i$ and $\bar{\hat{p}}_i$ for each individual observation using Equation 4.1.

- Step 2. For a given UCL_A, Type I errors α_a, α_l and α_p are obtained for various combinations of $g \in [0.5, 1]$ and $Ac \in [0.01, 0.6]$ by replicating Step 1 with large simulation runs such as 10,000. Find those g and Ac meeting the producer's risk related constraints, say $\alpha_a \leq 5\%$, $\alpha_l \geq 0.5\%$ and $\alpha_p \geq 0.5\%$.

- Step 3. Generate random samples of data from a shifted process, say $Y \sim N(\mu_1, \sigma_Y^2)$, and then calculate the probability of acceptance P_a and consumer's risk β_a for the various combinations of g and Ac obtained in Step 2. The minimum β_a and the corresponding g are then short-listed. We treat g corresponding to the minimum β_a as the optimum guardband coefficient.

4.5.1 Normal process

Assume that the observed quality characteristic Y and measurement error Z are normally distributed and process variance σ_Y^2 is known. Given USL and σ_Y^2, AQL corresponds to a target in-control process mean μ_0. In other words, if lots are formed from an in-control process, the proportion nonconforming is at the AQL level. Likewise, the RQL is also paired with a particular out-of-control process mean μ_1. For simplicity, let the in-control process $Y \sim N(0,1)$ and out-of-control process $Y \sim N(1,1)$, which correspond to AQL=3% and

RQL=18.9% respectively. Assume measurement error $Z \sim N(0, 0.3)$ when variance ratio $\kappa = \sigma_Z^2/\sigma_Y^2 = 0.09$ and the maximum production length is set at $N = 30$.

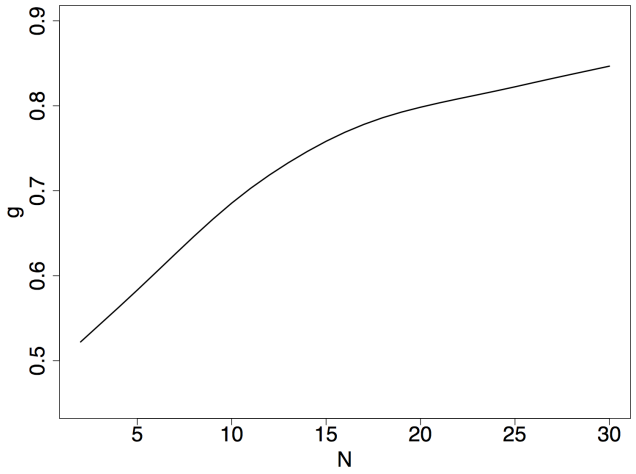
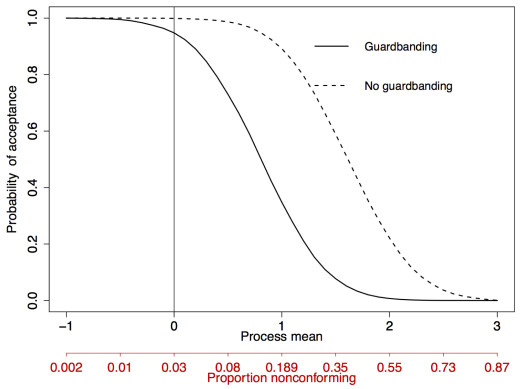


Fig. 4.3 Optimum guardband selection: $Y \sim N(0, 1)$

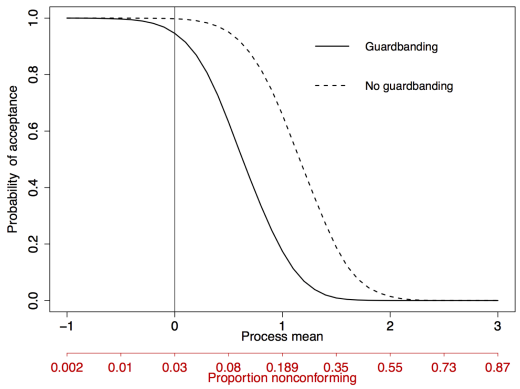
Monte Carlo UCL_A were found (shown in Table 4.2, Appendix 4.A), so that the false alarm rate of the process control plan is as small as $\alpha_p = 0.3\%$ when $g = 1$, but α_p goes up when g decreases. The optimum guardband coefficient g is obtained by following simulation algorithm. As shown in Figure 4.3, g varies for different production lengths N and the curve becomes flatter when $N \geq 20$. This implies that optimum g tends to be stable when the production length reaches 20.

After obtaining the optimum g and Ac, the probability of acceptance P_a for a batch under the MACCS scheme with guardbanding can be calculated at any proportion nonconforming level. The OC curves of the MACCS scheme with and without guardbanding are shown for different production lengths of $N = 5, 10, 20$ and 30 in Figure 4.4. It is clear that the proposed scheme can significantly reduce the consumer's risk without increasing the set producer's risk regardless of the production length, and the guardband approach is especially beneficial when the production length is very short such as 5. When N is small, there is lack of power to detect a shift in the (out-of-control) process. As a result, the consumer's risk can be extremely high. Hence, a more tightened g is adopted to control the consumer's risks.

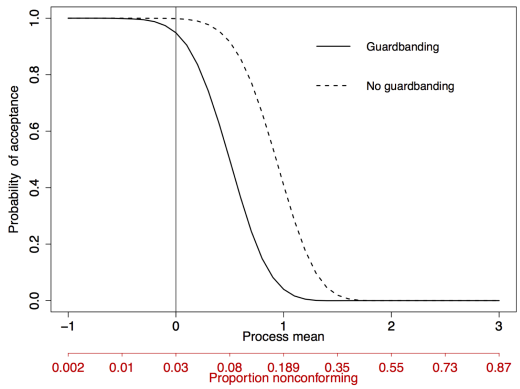
Another factor that may affect the optimum guardband selection is the process variance. Consider another in-control process $Y \sim N(0.25, 0.01)$ with smaller variance but with the same $\kappa = 0.09$ and measurement error distribution $Z \sim N(0, 0.003)$. For given AQL=3%, USL=0.268 and RQL=18.9% which corresponds to the out-of-control process $Y \sim N(0.26, 0.01)$. Figure



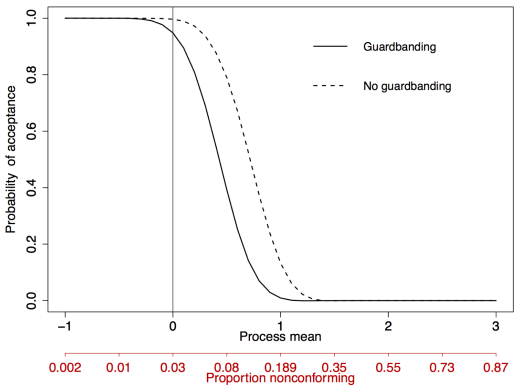
(a) $N = 5$, $g = 0.583$



(b) $N = 10$, $g = 0.685$



(c) $N = 20$, $g = 0.798$

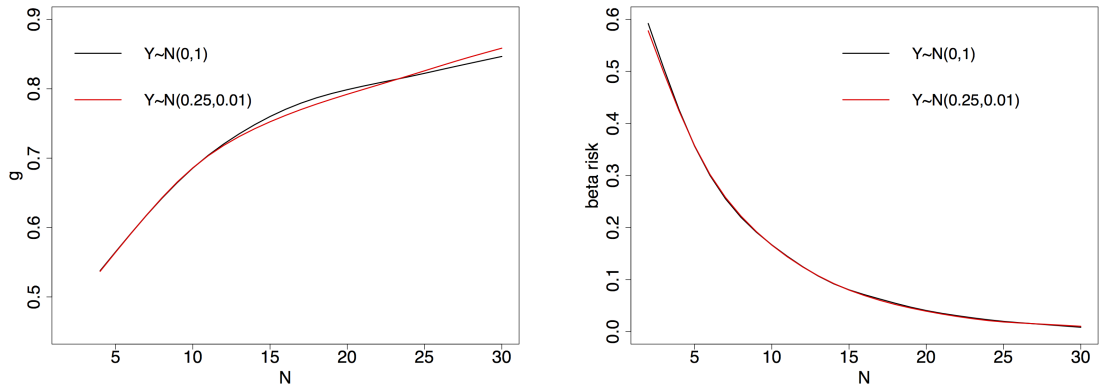


(d) $N = 30$, $g = 0.847$

Fig. 4.4 OC curves of the MACCS scheme

4.5 shows that the curve of optimum guardband coefficient g for the process $Y \sim N(0.25, 0.01)$ is quite close to the curve for the process $Y \sim N(0,1)$, and the corresponding beta risks are similar. This suggests that optimum guardband selection is somewhat robust to process variance as long as the variance ratio $\kappa = \sigma_Z^2/\sigma_Y^2$ is maintained. In other words, guardbanding is also helpful for processes with low process dispersion.

The choice of $\kappa = \sigma_Z^2/\sigma_Y^2$ might also affect optimum guardband selection. The ratio κ is commonly in the range between 0.01 and 0.25. We assumed $\kappa = 0.09$ in the previous cases and calculated the measurement error variance σ_Z^2 based on the assumed process variance σ_Y^2. However, in practice, σ_Z^2 is usually known from past calibration studies while σ_Y^2 may not be available. For any given σ_Z^2, the associated values of κ and σ_Y^2 are related: a small κ indicates a high process variance. For example, assuming $Z \sim N(0, 0.3)$, $\kappa = 0.04, 0.09$ and



(a) g vs N (b) β_a vs n

Fig. 4.5 Optimum guardband selection: $Y \sim N(0,1)$ vs $Y \sim N(0.25,0.01)$

0.16 correspond to three in-control processes with different process dispersions, which are $Y \sim N(0,1.5)$, $Y \sim N(0,1)$ and $Y \sim N(0,0.75)$ respectively. As shown in Figure 4.6, the curve of optimum g always becomes flatter when $N \geq 20$, no matter which κ is employed. However, the optimum g is more stringent when κ is small. This is not surprising because the guardbanding is expected to be severe for a process with large process dispersion when the measurement error variance σ_Z^2 is constant. Therefore, the optimum guardband selection depends on κ. For a large (small) κ, the associated optimum guardband coefficient g will also be large (small). We employed a moderate $\kappa = 0.09$ value in the discussion presented in further sections.

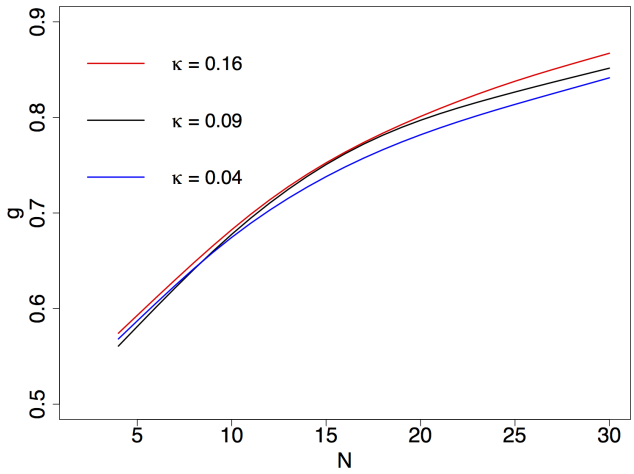
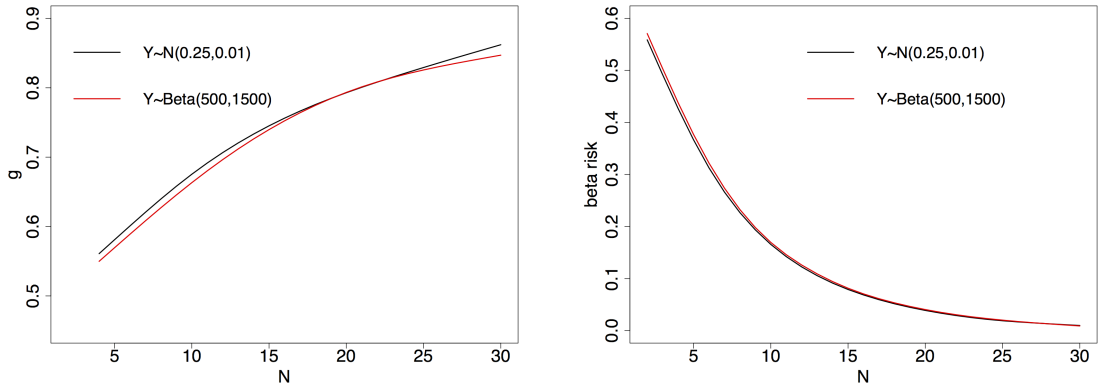


Fig. 4.6 Optimum guardband selection for different κ

4.5.2 Non-normal processes

The properties of the fractional nonconformance based control chart for non-normal processes, including beta and non-stationary processes, was studied by Zhou et al. [175]. We replicated this approach for the MACCS scheme and studied the effect of production length on optimum guardband selection under various models.

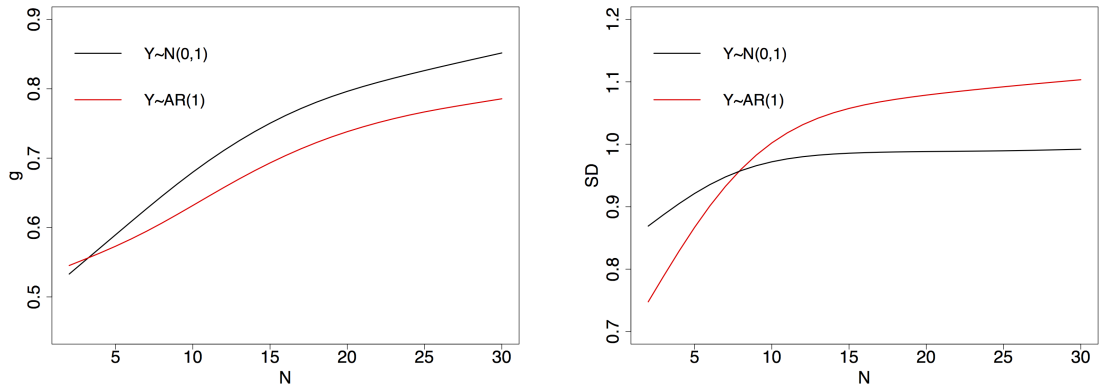
Consider the in-control process with apparent measurements $Y \sim \text{Beta}(500, 1500)$, which has similar process properties as the normal process $Y \sim N(0.25, 0.01)$. Let RQL=18.9%. This RQL corresponds to Beta(520, 1480), the out-of-control process with one standard deviation (SD) shift for given $\kappa = 0.09$ and measurement error $Z \sim N(0, 0.003)$. The optimum guardband coefficient g when $Y \sim \text{Beta}(500, 1500)$ can also be obtained by following the simulation steps illustrated in Section 4.5. The curves of optimum g and the corresponding consumer's risk for $Y \sim \text{Beta}(500, 1500)$ and $Y \sim N(0.25, 0.01)$ models, shown in Figure 4.7, are very similar. Moreover, the OC curves of the MACCS scheme with guardbanding under the beta model (Figure 4.15 in Appendix 4.B) are consistent with the OC curves under the normal model. Hence, we conclude that the optimum guardband selection does not heavily depend on the process distribution when the process mean and SD are matched.



(a) g vs N (b) β_a vs N

Fig. 4.7 Optimum guardband selection: $Y \sim \text{Beta}(500, 1500)$ vs $Y \sim N(0.25, 0.01)$

The assumption of independent and identically distributed may not be fully valid. Autocorrelated processes including stationary and non-stationary process are studied for the proposed scheme using an autoregressive integrated moving average (ARIMA) model; see Box et al. [12]. For simplicity, we only consider first order autocorrelated model AR(1): $y_t = 0.5y_{t-1} + \epsilon_t$ (stationary) and integrated moving average IMA(1,1): $y_t - y_{t-1} = \epsilon_t - 0.5\epsilon_{t-1}$ (non-stationary), where $\epsilon_t \sim N(0, 1)$ is white noise. Again, we used Monte Carlo simulations to obtain the optimum guardband coefficient g for the proposed scheme.

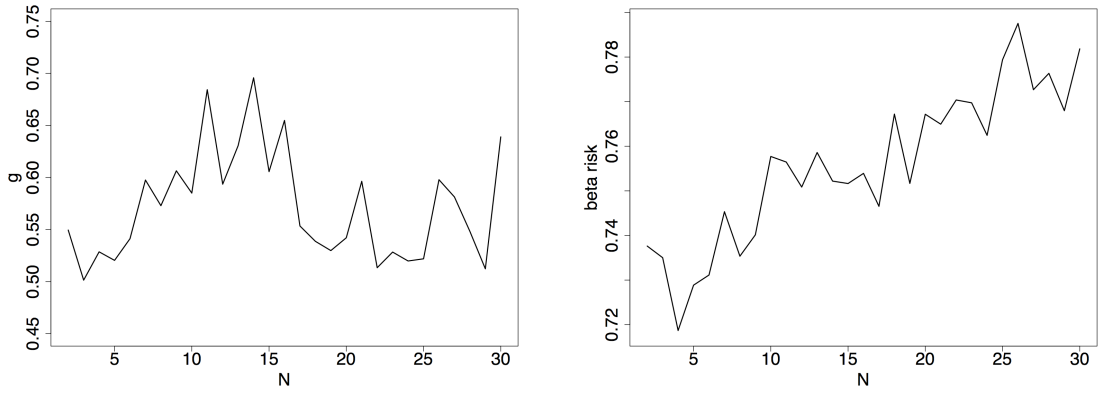


(a) g vs N (b) Empirical SD vs N

Fig. 4.8 Optimum guardband selection: $Y \sim N(0,1)$ vs $Y \sim \text{AR}(1)$

As demonstrated in Figure 4.8(a), the curve of optimum g for the stationary process $y_t = 0.5y_{t-1} + \epsilon_t$ also becomes stable when $N \geq 20$, which is similar to the curve under the normal model $Y \sim N(0,1)$. This is because the observation y_t under the stationary process also follows a normal distribution when ϵ_t is normal distributed. We also found that the optimum g is a bit more stringent for the normal process at the beginning of the process. However, when production length N increases, optimum g becomes more tightened for the stationary process. This is not surprising, because the process variance of the stationary process $y_t = 0.5y_{t-1} + \epsilon_t$ goes to $\sigma_Y^2 = \sigma_\epsilon^2/(1-\phi_1^2) = 1/(1-0.5^2) = 1.33$ when production length is long, which is larger than process variance of the normal process $Y \sim N(0,1)$. However, when production size is short, the empirical process variance of the stationary process is smaller than variance of the normal model; see Figure 4.8(b). This property agrees with our findings in Section 4.5.1: the larger the process dispersion, the tighter the guardband should be. Figure 4.16 in Appendix 4.B shows the OC curves of the MACCS scheme under stationary process AR(1), which is also congruent with the OC curves under the normal process.

Figure 4.9(a) shows that the optimum guardband coefficient g under the non-stationary process model wanders around $g = 0.6$, which suggests that optimum g selection does not depend on production length. The reason is that a non-stationary process model does not accumulate information from previous observations; hence increasing the production size does not improve the power of detection. This is consistent with the property that the β-risk curve of the MACCS scheme does not drop when N increases, shown in Figure 4.9(b). In contrast to independent and stationary processes, the OC curve of the MACCS scheme under a non-stationary process is not dependent on production length N; see Figure 4.17 in Appendix 4.B.



(a) g vs N (b) β_a vs N

Fig. 4.9 Optimum guardband selection: Non-stationary process

4.5.3 Unknown process variance

In the last two sections, we investigated the optimum guardband for the MACCS scheme assuming the process variance σ_Y^2 is known. Although this assumption can be validated on empirical grounds, we further examine guardband selection when the process dispersion is unknown.

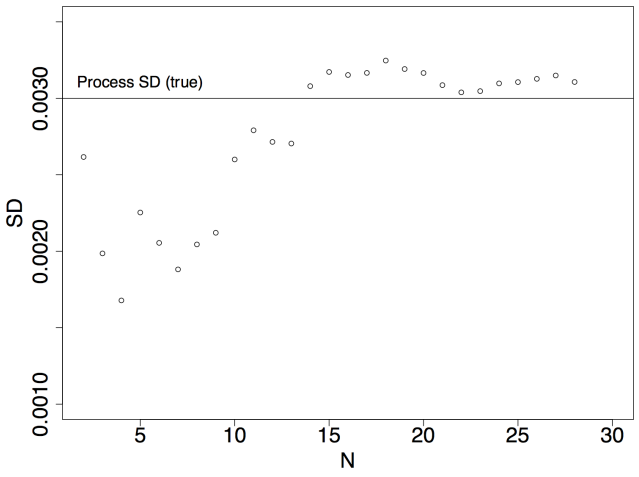
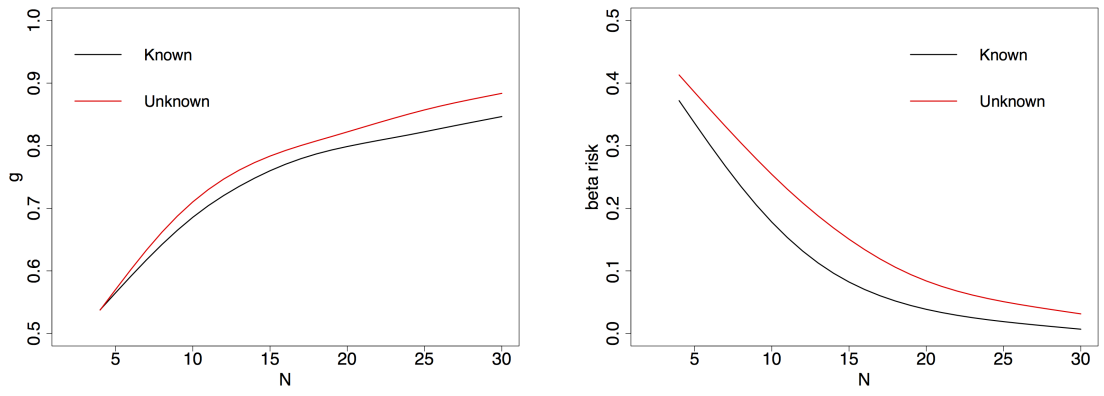


Fig. 4.10 Cumulative sample SD vs N

When the process variance is unknown, it can only be estimated using the samples collected from the process. As a result, the standard errors of the estimates will be large, especially at the beginning of production where the production length is too short for accurate estimation. Therefore, the unknown process variance case results in a higher consumer's risk

and hence a tighter guardband is needed. When production length increases, the estimated process variance becomes stable and will progressively converge to the true process variance. Consider the following example, when the known σ_Y is 0.003. Figure 4.10 shows that the estimated σ_Y becomes closer to the true value only when $N \geq 15$.

Recall the normal process $Y \sim N(0,1)$ in Section 4.5.1. Instead of using $\sigma_Y^2 = 1$, the sample variance can be dynamically updated and then optimization of the guardband coefficient g can be done. Such optimum g values and corresponding consumer's risks β_a are plotted in Figure 4.11 along with the known $\sigma_Y^2 = 1$ case. Although the optimum g for the unknown case is quite close to the known case at the beginning of the production, the unknown case always leads to higher consumer's risks. The guardband approach cannot completely control the consumer's risk because of the set constraint for the producer's risk namely $\alpha_a \leq 5\%$. In other words, although the consumer's risk is still very high for SD-unknown case, we are unable to guardband further without increasing the producer's risk. A cost model is helpful to balance the risks, and this approach is discussed in the next section.



(a) g vs N (b) β_a vs N

Fig. 4.11 Optimum guardband selection: SD-known vs SD-unknown process

4.6 Cost model for guardband selection

In Section 4.5, we choose the optimum guardband for the MACCS scheme based on risks, where the producer's risk was always maintained at a set level, say $\alpha_a \leq 5\%$ and the optimum g corresponds to the one which gives the minimum consumer's risk β_a. Due to the constraints of producer's risk, the consumer's risk of the guardbanded plan may still be very high in some cases, say when process variance is unknown. Although both Type I and Type II errors cause economic loss to manufacturers, the cost of a Type II error is generally much higher than the cost of a Type I error, especially in the food industry. A simple cost model is introduced in this section to balance the Type I and Type II errors, and optimum guardband selection is

based on the total costs of Type I and II errors. In contrast to risk based guardband selection, where Ac changes with g to ensure that the desired α_a is maintained, Ac is fixed in the cost model. Hence, given UCL_A and Ac, Type I error increases and Type II error decreases when g goes down.

The simple total cost function we employed depends on two parameters: c_β is the estimated loss associated with Type II error including warranty, customer dissatisfaction and brand damage etc., while c_α is the average loss for the manufacturer due to Type I error of rejecting acceptable quality product. The total cost (C_a) of the MACCS scheme is the sum of acceptance and rejection costs as below:

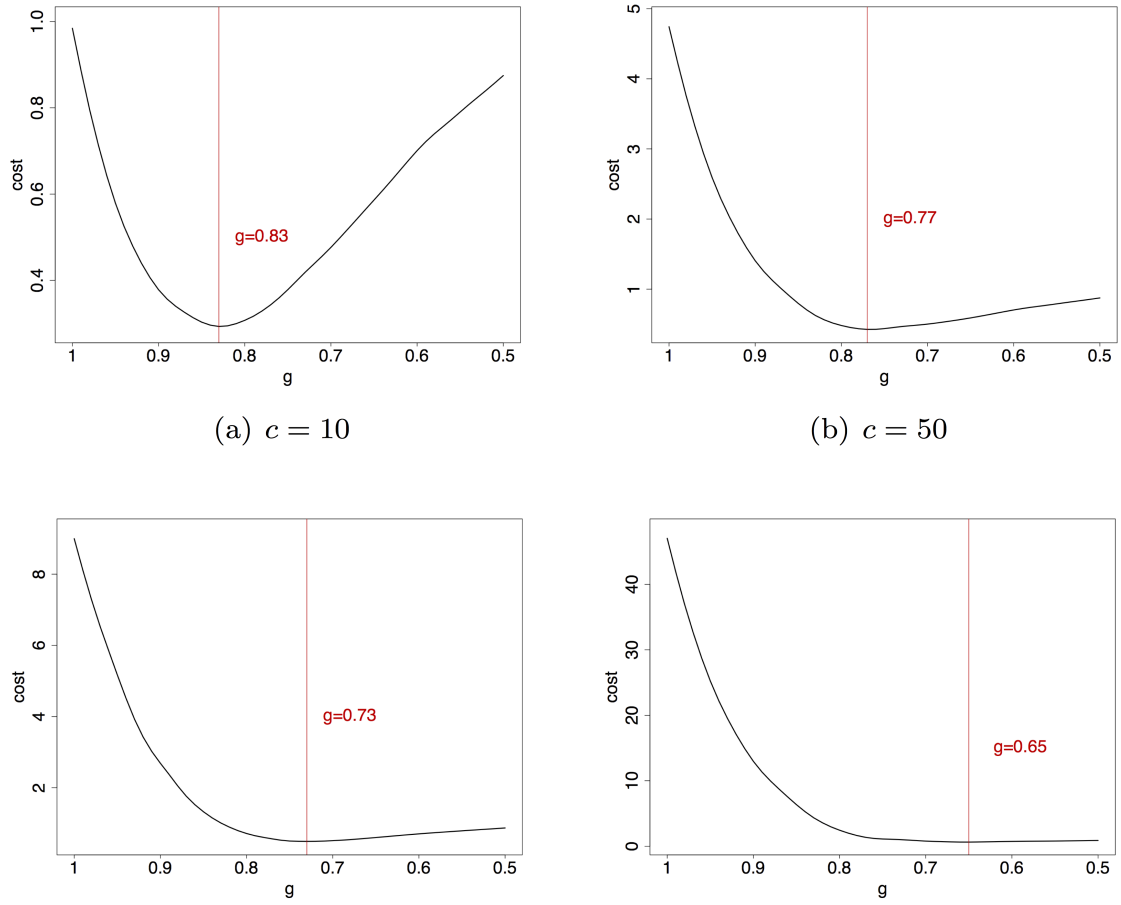
$$C_a = c_\alpha \alpha_a + c_\beta \beta_a \tag{4.4}$$

where α_a is the probability of rejection when the quality is at the AQL, and β_a is the probability of incorrect acceptance when the quality is at the RQL. The quantities $c_\alpha \alpha_a$ and $c_\beta \beta_a$ are the batch rejection and acceptance costs respectively. Different cost ratios $c = c_\beta / c_\alpha = 10, 50, 100, 500$ are considered for evaluating the proposed scheme.

Monte Carlo simulation was also adopted for cost based guardband selection as described below:

- Step 1. Calculate fractional nonconformance statistics $\hat{p}_i$ and $\bar{\hat{p}}_i$ for a random sample of data from normal process $Y \sim N(\mu_0, \sigma_Y^2)$.
- Step 2. For a given UCL_A, repeat Step 1 to find those Ac corresponding to the set producer's risk, say $\alpha_a \leq 5\%$. Generate random samples of data from a shifted process $Y \sim N(\mu_1, \sigma_Y^2)$ and select the unique Ac which results in the minimum β_a. Such matching combinations of Ac and UCL_A are used to evaluate the risks.
- Step 3. With UCL_A and Ac obtained in Step 2 and various values of $g \in [0.5, 1]$, calculate Type I error α_a when process is in-control and Type II error β_a when process is out-of-control.
- Step 4. Calculate the total cost C_α using Equation 4.4 with different cost ratios. The g which leads to the minimum C_α is considered as the optimum guardband coefficient.

As before, consider the in-control process $Y \sim N(0,1)$ and out-of-control process $Y \sim N(1,1)$ with measurement error $Z \sim N(0, 0.3)$. For a short run production with $N = 20$, the relationship between guardband coefficient g and total cost C_a is shown in Figure 4.12. The guardbanding approach is found to reduce the total cost no matter what the cost ratio is. However, guardbanding is more economical when the cost ratio c is high. For example,



(a) $c = 10$ (b) $c = 50$

(c) $c = 100$ (d) $c = 500$

Fig. 4.12 Total cost vs g for normal model $Y \sim N(0,1)$, $N = 20$

when the cost ratio c is small, say $c = 10$, the saving due to optimum guardbanding is only moderate, but when the cost ratio increases to $c = 50$ or more, the cost curve drops significantly allowing the total cost C_a to reach its minimum, which corresponds to the optimum guardband coefficient g. Moreover, the optimum g is also related to the set cost ratio. The higher the cost ratio of Type II and Type I errors, the tighter the guardband coefficient g should be.

Production length is also examined for guardband selection using the cost model. A more tightened g should be employed to control the consumer's risk for very short production lengths. This was also the case with the risk based optimum g selection discussed in Section 4.5. To investigate the effect of the production length N on the choice of guardband coefficient

in the cost model, we further explored different production lengths between $N = 2$ and $N = 30$ using the same cost model described earlier.

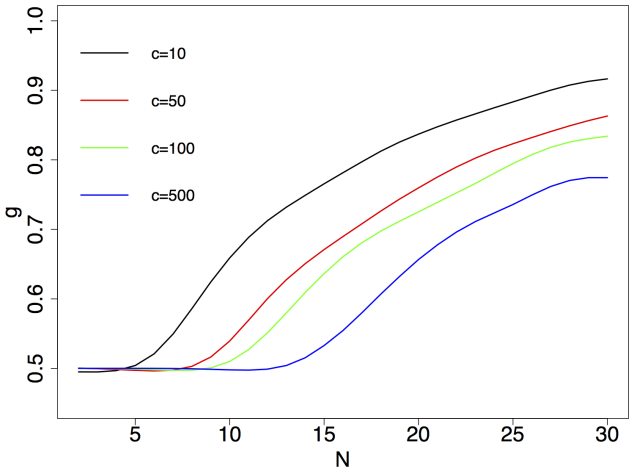
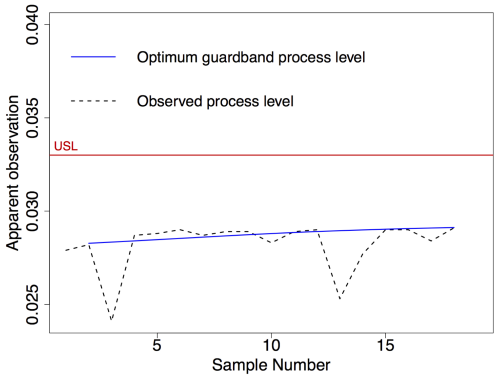


Fig. 4.13 Optimum g vs N for different cost ratios

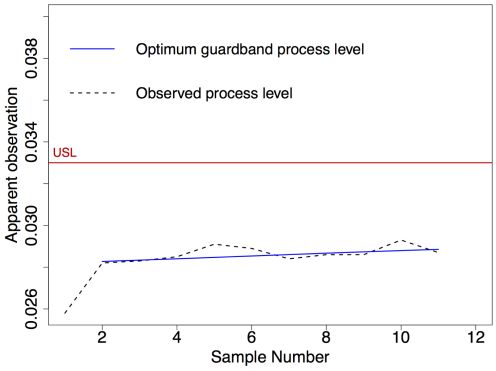
The relationship between total cost and optimum g for different production lengths and cost ratios is illustrated in Figure 4.13. When the cost ratio is small (such as $c = 10$) and the production length is greater than 25, the optimum guardband coefficient is more than 0.9, which means guardbanding is not really necessary when production length is long. This is not surprising for a small cost ratio because the saving from Type II error might be similar to the cost of Type I error. When the cost ratio is high, say $c \geq 50$, the optimum g is tighter when the production size is small ($N < 10$) and the optimum g becomes more stable when $N > 25$. This is because the consumer's risk is very high at the beginning of production and a strict g is more helpful to reduce consumer's risk at the expense of increasing producer's risk to some extent. Consumer's risk goes down as the production size is accumulated and when production length is longer, say $N > 25$, the consumer's risk is already smaller, so that the guardband approach will not decrease consumer's risk any further. Figures 4.13 also suggests that the optimum guardband coefficient g tends to become stable when production length N is greater than 25, no matter what the cost ratio c is. This result is consistent with what we observed under the risk model.

4.7 Case study

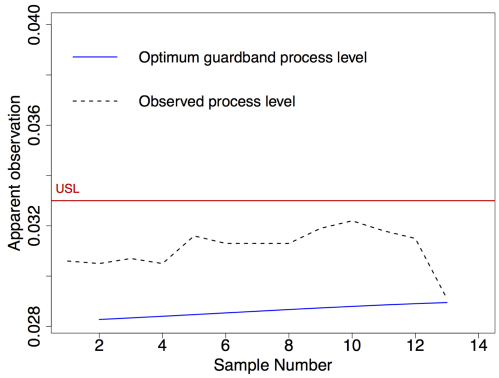
The modified acceptance control chart scheme was motivated by problems we encountered in a very high volume but short run dairy product manufacturing. We examined hundreds



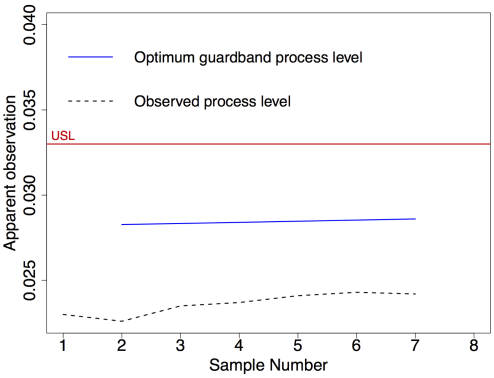
(a) Batch 21



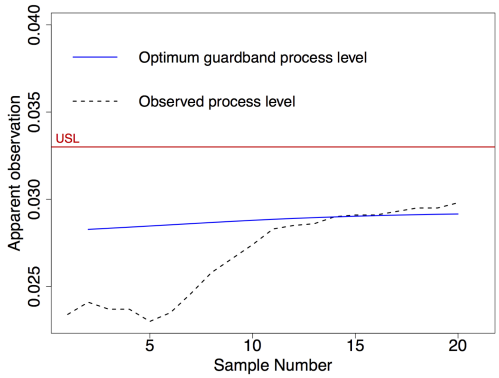
(b) Batch 24



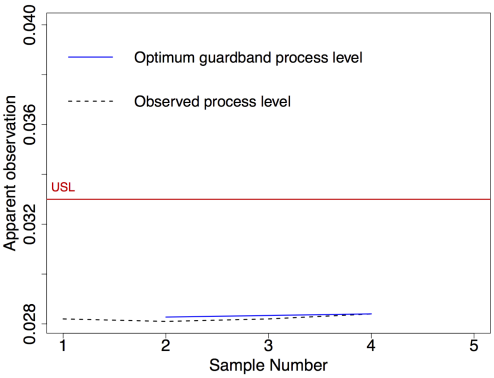
(c) Batch 18



(d) Batch 8



(e) Batch 9



(f) Batch 23

Fig. 4.14 Optimum guardbanding for various batches

batches of data from whole milk powder production ($N \leq 30$) using the proposed scheme; a few examples are presented in Figure 4.14. The process mean was maintained at the optimum guardband process level in Batches 21 and 24, while Batch 18 is obviously away from optimal because the observed process level is higher than the set nominal in-control process mean as well as the apparent observations being too close to the upper specification limit. The process level in Batch 8 is also sub-optimal as it is far away from the optimum guardband process level. The process mean in Batch 9 is too conservative at the beginning of the process and then becomes optimal during the later part of the production. We also show that the guardbanding approach works well for very short lengths, e.g. Batch 23. The implementation of the proposed guardbanding was done using a web based *Shiny* app written in R software language and this app hosted at https://zhouxin07.shinyapps.io/guardbanding/.

4.8 Conclusion

In this study, we developed a fractional nonconformance based modified acceptance control chart scheme to monitor short-run processes and simultaneously dispose the products manufactured by the process. Guardband selection for the proposed scheme was studied under various scenarios. The simulation results showed that the optimum guardband settings are consistent under both risk and cost based models.

Consumer's risk can be high when the production length is small. The guardbanding approach can reduce the consumer's risk significantly for small production lengths. When the production length becomes longer, the consumer's risk tends to be stable. Hence, the guardbanding approach is preferable for consumer protection when the production length is short.

We only implemented and demonstrated the guardbanding strategy for a one-sided modified acceptance control chart scheme. For short productions, it is difficult to detect quality improvement and our main focus was to exert acceptance control to reduce the consumer's risks.

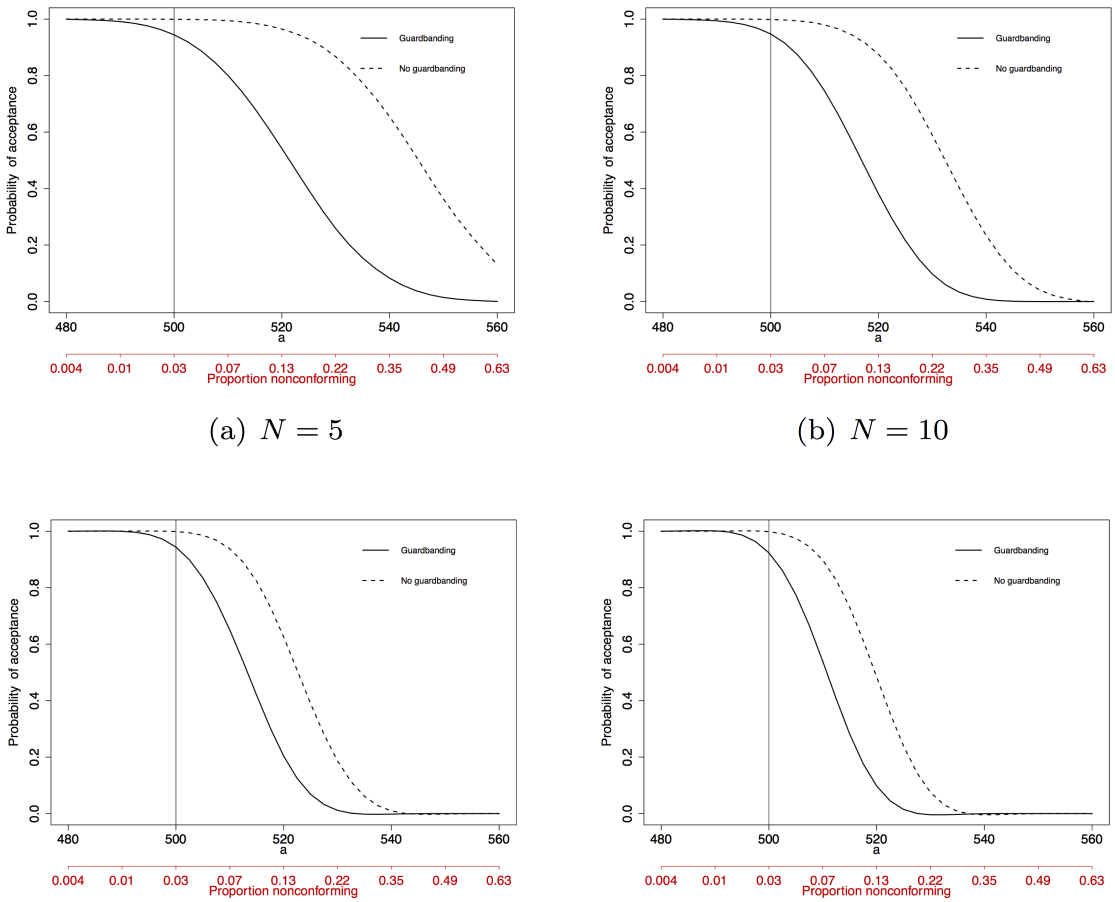
Even though the proposed scheme still requires knowledge of the error distribution, we performed sensitivity analysis to study the effect of unknown process variance etc. on the optimum guardband selection for the proposed modified acceptance control chart scheme.

Appendix 4.A Upper control limit of the MACCS scheme

Table 4.2 UCL_A for different processes

N	Normal	Beta	Stationary	Non-stationary
1	1	1	1	1
2	0.9594	0.9215	1	1
3	0.6624	0.6665	0.9994	1
4	0.5226	0.5261	0.9659	0.9999
5	0.4641	0.4825	0.8883	0.9991
6	0.3921	0.4366	0.8084	0.9980
7	0.3869	0.3691	0.7494	0.9948
8	0.3434	0.3285	0.6625	0.9932
9	0.3318	0.3209	0.6297	0.9907
10	0.2922	0.2990	0.5891	0.9890
11	0.2794	0.2719	0.5214	0.9836
12	0.2587	0.2634	0.5119	0.9821
13	0.2397	0.2439	0.4799	0.9804
14	0.2325	0.2381	0.4695	0.9796
15	0.2287	0.2256	0.4478	0.9768
16	0.2144	0.2200	0.4298	0.9750
17	0.2029	0.2129	0.4045	0.9733
18	0.2017	0.2045	0.3989	0.9731
19	0.1909	0.1952	0.3698	0.9702
20	0.1900	0.1952	0.3698	0.9665
21	0.1851	0.1882	0.3498	0.9636
22	0.1790	0.1820	0.3390	0.9622
23	0.1780	0.1747	0.3376	0.9619
24	0.1677	0.1674	0.3227	0.9608
25	0.1646	0.1674	0.3114	0.9608
26	0.1643	0.1615	0.3061	0.9608
27	0.1615	0.1580	0.2956	0.9608
28	0.1561	0.1580	0.2841	0.9608
29	0.1532	0.1541	0.2821	0.9608
30	0.1520	0.1518	0.2821	0.9608

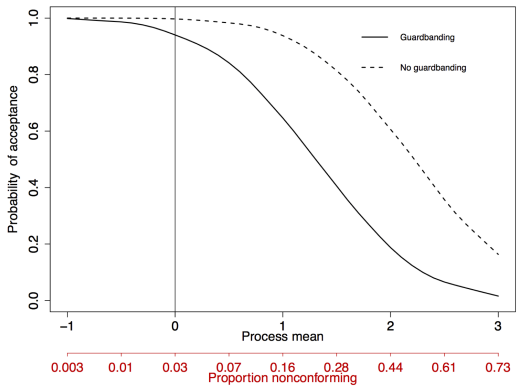
Appendix 4.B OC curves of the MACCS scheme



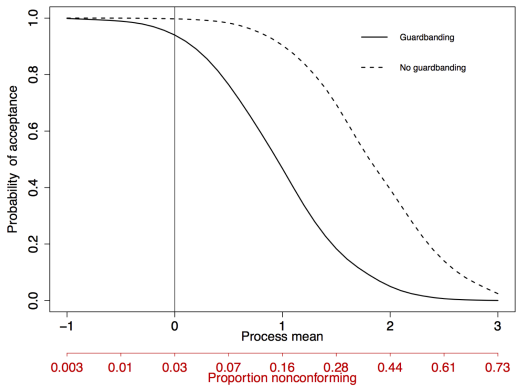
(a) $N = 5$ (b) $N = 10$

(c) $N = 20$ (d) $N = 30$

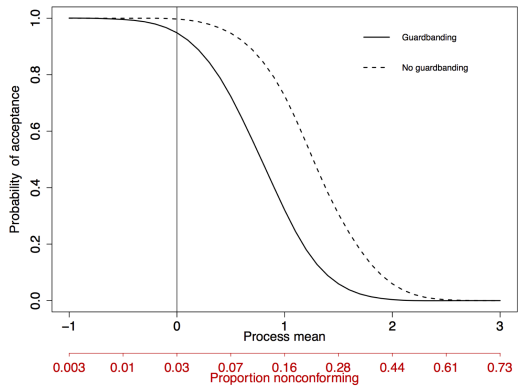
Fig. 4.15 OC curves of the MACCS scheme (Beta model)



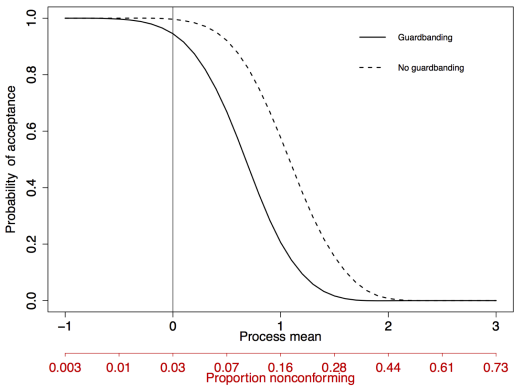
(a) $N = 5$



(b) $N = 10$

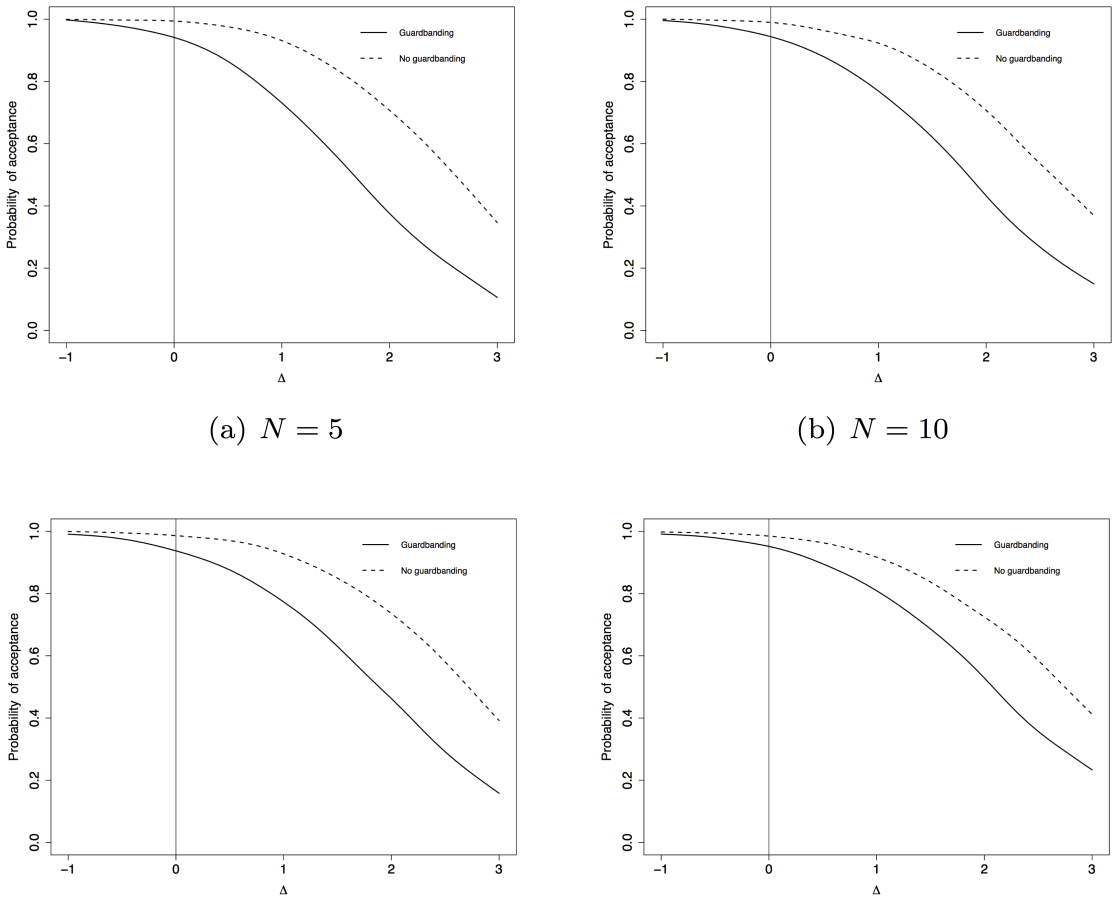


(c) $N = 20$



(d) $N = 30$

Fig. 4.16 OC curves of the MACCS scheme (Stationary process)



(a) $N = 5$ (b) $N = 10$

(c) $N = 20$ (d) $N = 30$

Fig. 4.17 OC curves of the MACCS scheme (Non-stationary process)

Appendix 4.C *Shiny* app user guide

The following step-by-step are followed to implement the MACCS scheme with *Shiny* R software app hosted at https://zhouxin07.shinyapps.io/guardbanding/.

- Step 1. **Upload data.** Click “Browse" button and upload single column txt/csv file; the first row of the dataset should be the header.
- Step 2. **Select specifications.** Select upper or lower specification and enter the value.
- Step 3. **Select mode.** Select the distribution of the quality characteristic Y namely “Normal" or “Beta”. Optimum guardband width will be automatically calculated based on total cost if “Minimum cost” is selected.
- Step 4. **Select ratio.** Select the error ratio (Error SD/Process SD) for the Normal/Beta model or the cost ratio (Type II/Type I) for the cost model.
- Step 5. **View results.** Click “View data and plot" button and the results will be displayed on the right side panels; see Figure 4.18.

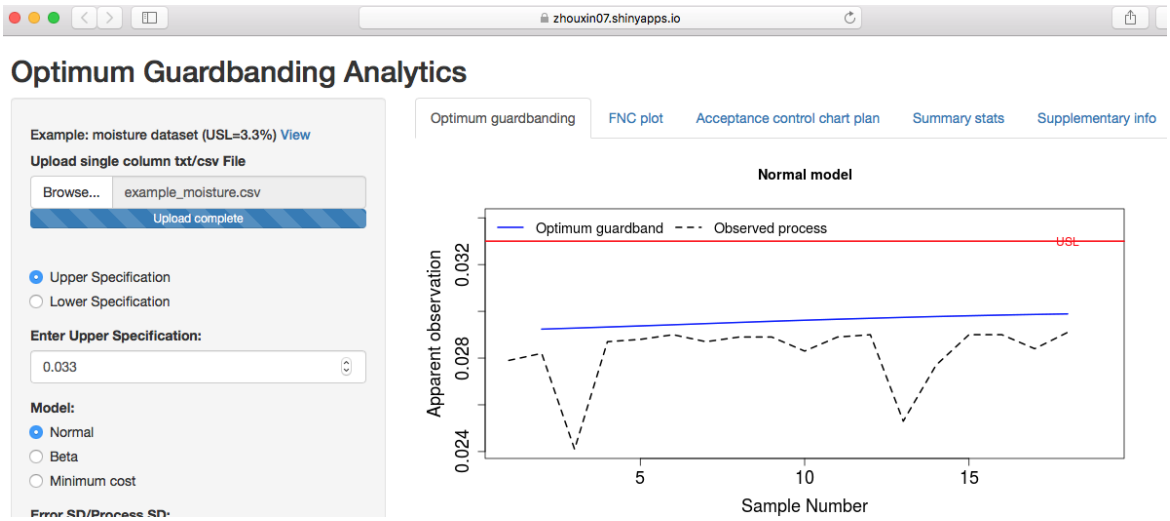


Fig. 4.18 Optimum guardbanding analytics

Chapter 5

Fractional nonconformance based conformity testing

Xin Zhou, Kondaswamy Govindaraju, Geoff Jones
Computers & Industrial Engineering, 2019
https://www.sciencedirect.com/science/article/pii/S0360835219303432

5.1 Abstract

Conformity testing, also known as evaluation of conformity or compliance testing, is exercised to assure that an entity meets a specific requirement and/or regulatory standard. Measurement and sampling uncertainties, including metrological traceability, become crucial for the declaration of conformity, especially when the measured value is close to the set limiting value. The fractional nonconformance (FNC) approach was recently introduced to deal with measurement errors for acceptance sampling purposes. This approach can be extended to develop new conformity evaluation protocols. In this study, we develop a new fractional nonconformance based conformity testing procedure, and compare its performance with the conformity testing procedure prescribed in the ISO Standard 10576. The proposed procedure explicitly distinguishes the limiting value from a specification by linking the false alarm rate and the limiting value using a baseline distribution. Our research shows that the two-stage conformity testing procedure using the fractional nonconformance statistic reduces the probability of incorrect declaration of conformity or inconclusive result for nonconforming entities when the number of test samples is larger than one. This means that the new FNC based conformity testing is more powerful than the current procedure set by the ISO standard.

Keywords

conformity testing; fractional nonconformance; limiting value; measurement uncertainty; specification; statistics

5.2 Introduction

Conformity testing is used to systematically examine whether an entity conforms to specified standards or criteria. The entity could be a process, product, service or some personnel measure. The specification for the quantifiable characteristic, such as the maximum allowable concentration of a drug or trace element in blood for normal people, is called as a limiting value (LV) in the conformity testing protocols. The LV could be understood as either a minimum value (Lower limit or L_{SL}) or a maximum value (Upper limit or U_{SL}), or both. The interval containing all permissible values of the characteristic is called the *region of permissible values*. A conformity testing protocol provides assurance of conformity by checking whether the measurement of interest falls within the region of permissible values or not. The conformity testing procedure is often employed to check whether the entity meets specific requirements and/or regulatory standards. The ISO 10576-1 International Standard [79] (Guidelines for the evaluation of conformity with specified requirements) gives examples of conformity testing such as a test of the concentration of some trace elements in the blood of employees for their health evaluation. In sports, the analysis of an athlete's urine to detect abuse of xenobiotic anabolic steroids, testosterone and doping etc is often based on the conformity testing methodology. In bulk product manufacturing applications, conformity testing is implemented to ensure that the quality characteristic such as percentage fat falls within the permissible range, which is often set for export products. For further novel applications; see Källgren et al. [88], Theodorou and Zannikos [146] and Widlowski et al. [156].

According to the ISO/IEC Guide 98-3:2008 [81], measurement results can only be compared with references values when the measurement uncertainty (MU) is clearly indicated. Some common models for measurement uncertainty were presented in Dieck [32]. Measurement uncertainty, including both sampling variation and metrological traceability, is inevitable in practice and it may have a significant impact on conformity test results and subsequent decision making, especially when the measurement is close to LV. For example, a conforming product can be incorrectly declared nonconforming due to measurement uncertainty, which is a risk to the producer. Consumer's risk occurs when a nonconforming entity is declared as conforming. Pendrill [117] discussed different costs and revenue in the context of conformity testing from a producer's point of view and optimized the uncertainty related costs by balancing test costs against consequential costs. Forbes [42] applied Bayesian decision-making approaches to quantify and minimize the cost of wrong decisions using a loss

function. The impact and risks of MU as well as its mitigation, have been studied by many researchers; for example Hinrichs [67], Wu and Govindaraju [166] and Maleki et al. [103].

Measurement uncertainty is usually reported as an uncertainty interval, given in the form of a (95%) confidence interval. The practice in conformity testing procedures is to compare the MU interval around the measurement result with the region of permissible value; see Pendrill [119], King [93], Ellison et al. [38], Kessel [91] and many others. Conformity can be declared if and only if the whole uncertainty interval is located within the region of permissible values. When the MU interval overlaps the region of permissible values, Hibbert [66] advocated in favour of making a statement on the probability of compliance in addition to reporting the uncertainty interval. On the other hand, Holst et al. [69] proposed the two-stage conformity testing procedure to increase the probability of declaring conformity for those entities with permissible values of the quantity of interest close to the LV, without unduly increasing the probability of declaring conformity for nonconforming entities. They also claimed that the probability of obtaining an inconclusive result is reduced for the two-stage procedure. The two-stage procedure of Holst et al. [69] forms the basis of the ISO 10576-1 International Standard [79]. Desimoni and Brunetti [31] presented a thorough review of measurement uncertainty and conformity testing.

According to a report from the World Health Organization [164], the methods of deriving LVs are different from one country to another. In some cases, the value of LV is laid down by legislation or regulatory authorities, which is analogous to the specification limit in acceptance sampling. Limiting values and specification limits are interchangeable in the ISO 10576-1 International Standard [79]. However, for conformity testing, especially in the area of legal metrology, a limiting value must be distinguished from a specification limit. A specification limit is often understood as an engineering requirement rather than a legal requirement, whereas LV is intrinsically connected to a common cause (baseline) model for the measured characteristic. For example, the LV for the epitestosterone ratio to declare xenobiotic anabolic steroid intake is based on the probability distribution of testosterone for known drug-free athletes. The set LV for the epitestosterone ratio is different from a specification limit and must be viewed as an upper tail quantile of the baseline distribution, depending on the significance level set. Much of the published work on conformity testing mixes up LV with specification limit and does not consider the false positive error associated with LV.

Acceptance sampling procedures are sometimes applied in conformity testing, however the principles of the two activities are different. Acceptance sampling uses a moderate size sample to determine whether to accept or reject a batch (population). In contrast, conformity testing infers the true value of the sample tested. In other words, the conformity testing procedure is a decision making procedure on a particular sample rather than the population.

Govindaraju and Jones [51] proposed a new acceptance sampling plan based on the concept of *fractional nonconformance* (FNC), which was initially introduced to adjust for

measurement errors in individual measurements when the measurement error distribution is known. This approach can be extended to develop new conformity evaluation protocols. In this study, we develop a fractional nonconformance based two-stage conformity testing procedure, and compare its performance with the conformity testing procedure prescribed in ISO 10576-1 International Standard [79]. The results show that the proposed testing procedure performs better than the ISO standard procedure when the sample size is greater than one. This paper stresses the need to distinguish the LV from an arbitrarily imposed specification limit because LV pertains to a common cause model. This study explicitly considers a false alarm rate for the set LV under the common cause model, and then investigates its impact on the performance of conformity testing. Lastly, we develop a web based *Shiny* app to implement the FNC based conformity testing in practice.

The remainder of this article is organized as follows. Section 5.3 presents a brief review of the current ISO standard for conformity testing and the fractional nonconformance approach. Fractional nonconformance based conformity testing is newly introduced in Section 5.4 and its performance is investigated in Section 5.5. Application of the proposed procedure is demonstrated in Section 5.6. A brief discussion is provided at the end.

5.3 Review of International Standard for conformity testing and FNC approach

A brief review of current international guidelines for the evaluation of conformity with specified requirements and the fractional nonconformance approach is provided in this section. Table 5.1 lists the definition and notations used in this study.

5.3.1 ISO 10576-1 International Standard

The objective of the ISO 10576-1 International Standard [79] (the ISO standard hereafter) is to assure conformity in the form of a supplier's declaration, or a certification from a third party. According to this ISO standard, conformity testing is defined as a systematic examination of the extent to which an entity conforms to a specified criterion. The ISO standard provides guidelines for the evaluation of conformity with specified requirements; some basic terms and definitions are listed in Table 5.1. The ISO standard requires that the entity and the quantifiable characteristic of the entity shall be unambiguously specified and defined. The requirements for definition and reporting of limiting values are also provided with several illustrated examples including double limits and single limit. Given a measurement Y, assumed to be $N(\mu, \sigma_Y^2)$, the uncertainty of the measurement result (σ_Y) should be considered from all stages in the measurement procedure and reported as an uncertainty interval, together with the confidence level (1-α_m), where α_m is the significance level. The uncertainty interval of the measurement result is compared with the interval of all permissible/non-permissible

Table 5.1 Definitions and notations

Term	Description
Y	Apparent measurement result of the entity
Z	Measurement error, $Z \sim N(0, \sigma_Z^2)$
X	True measurement, $Y = X + Z$, the measurement error model
μ	Population mean of the entity
U	Upper specification
n_1	Sample size for stage one
n_2	Sample size for stage two
n	Total sample size, $n = n_1 + n_2$
k	Variance ratio, $k = \sigma_Z^2/\sigma_Y^2$
LV	Limiting value, a general term to refer to the upper and/or lower bounds of the permissible values
L_{SL}	Lower bound of the permissible values of the characteristic
U_{SL}	Upper bound of the permissible values of the characteristic
$\hat{p}_i$	Fractional nonconformance of the i-th observation
$\hat{p}_{c_i}$	Conditional fractional nonconformance of the i-th observation
δ	Sum of conditional fractional nonconformance, $\delta = \sum_{i=1}^{n} \hat{p}_{c_i}$
$\bar{\delta}$	Overall nonconformance level, $\bar{\delta} = \delta/n$
α_l	Significance level related to the limiting values
α_m	Significance level related to the measurement results
α_f	Significance level related to the fractional nonconformance statistic
$[B_L,\ B_H]$	Confidence interval of the measurement results
l_c	Decision limit for declaring conformity
l_n	Decision limit for declaring non-conformity
P_a	Probability of declaring conformity

values of the characteristic, which is defined as the region of permissible/non-permissible values. The rules for asserting conformity or non-conformity are:

- **Assurance of conformity**: *The uncertainty interval is inside the region of permissible values.*
- **Assurance of non-conformity**: *The uncertainty interval is included in the region of non-permissible values.*
- **Inconclusive result**: *The uncertainty interval includes LV.*

The ISO standard recommends performing the conformity test as a two-stage procedure, as initially proposed by Holst et al. [69]. In stage one, n_1 samples are tested and the uncertainty interval of the measurement result is calculated as a confidence interval $[B_L,\ B_H]$, with the degree of confidence 1-α_m. Conformity/non-conformity is declared if and only if both B_L and B_H are in the region of permissible/non-permissible values. Otherwise, an inconclusive result is declared in the first stage, and the second stage of the test with another n_2 samples

is performed. A new confidence interval based on the n_1+n_2 measurements from both stages is calculated and compared with the region of permissible/non-permissible values to evaluate conformity/non-conformity.

Examples of ISO two-stage conformity testing

Several illustrative examples of two-stage conformity testing procedures are provided in Annex B of the ISO standard, two of which are outlined below:

Example I: $n_1 > 1, n_2 > 1$, σ_Y is estimated from sample (Annex B.5 of the ISO standard)

A lot of Scandinavian dolomite was selected for conformity testing of asbestos content. The upper specification limit of mass fraction of asbestos in Scandinavian dolomite is 0.1%, so LV=0.1% was set. A two-stage procedure was adopted with five primary increments ($n_1 = 5$) in the first stage and four primary increments ($n_2 = 4$) in the second stage. The measurement results (in %) are presented as below:

Stage one: 0.152, 0.0704, 0.0772, 0.0731, 0.0551

Stage two: 0.0828, 0.0671, 0.0743, 0.0561

In the first stage, the mean and variance of the measurement results are calculated as $\hat{\mu} = 0.0856$ and $\hat{\sigma}_Y = 0.0381$. Given $\alpha_m = 5\%$, the uncertainty interval of the measurement results is calculated in the form of a 95% confidence interval as: $0.0856\% \pm (2.776 \times 0.0381\%)/\sqrt{(5)} = (0.038\%, 0.0133\%)$. Because LV is located within the uncertainty interval, the test is inconclusive at the given confidence level. Given the data from the second stage, the confidence interval of the measurement result becomes $0.0787\% \pm (2.306 \times 0.0290\%)/\sqrt{(9)} = (0.056\%, 0.101\%)$. Although the confidence interval becomes shorter after including more samples from stage two, the uncertainty interval still contains LV. Hence, conformity cannot be demonstrated at the given confidence level.

Example II: $n_1 = n_2 = 1$, σ_Y is known (Annex B.3 of the ISO standard)

The concentration of Pb in blood for individuals should be less than 0.97 μmol/l, hence LV=0.97 is defined. The measurement results are obtained using a standard measurement procedure with known uncertainty $\sigma_Y = 0.048$. The concentration of Pb in the blood for a particular individual is given as $Y_1 = 1.06$ for the first subsample ($n_1 = 1$) and the corresponding 95% confidence interval is calculated as (0.96, 1.15), which contains LV. The test is inconclusive and the second subsample ($n_2 = 1$) is measured as $Y_2 = 1.00$. The combined measurement results is (1.06+1.00)/2=1.03 with 95% confidence interval of (0.96, 1.10), which still includes LV. Thus, an inconclusive result is declared from the two-stage conformity testing.

Evaluation of two-stage conformity testing procedure

Holst et al. [69] evaluated the performance of the two-stage testing procedure using the following functions, which are simply the probabilities of declaring conformity at each stage.

$$\text{First/Single-stage: } P_a(\mu) = P(B_{H1} \leq LV) = \Phi\left(\frac{LV-\mu}{\sigma_Y/\sqrt{n_1}} - u_{1-\alpha_m/2}\right) \tag{5.1}$$

$$\text{Two-stage: } P_a(\mu) = P(B_{H1} \leq LV) + P\{(B_{L1} \leq LV < B_{H1}) \cap (B_{H2} \leq LV)\} \tag{5.2}$$

where the measurement Y is assumed to be $N(\mu, \sigma_Y^2)$ and $u_{1-\alpha_m/2}$ is the $1-\alpha_m/2$ quantile of the standard normal distribution.

The conformity testing procedure has three possible outcomes namely declaration of conformity, non-conformity or inconclusive result. We are more interested to know the risks when the conformity testing fails to assure conformity (non-conformity) for conforming (nonconforming) entities. Hence, we consider the following two types of *classification* errors:

- Type A classification error: Probability of incorrect declaration of non-conformity or inconclusive result when the entities are conforming.
- Type B classification error: Probability of incorrect declaration of conformity or inconclusive result when the entities are nonconforming.

It is obvious that the two-stage conformity testing increases the P_a, no matter whether the entities are conforming or nonconforming. Holst et al. [69] found that the two-stage procedure reduces Type A classification error of the conformity test significantly at the cost of slightly increased Type B classification error. This property remains when the test is performed for non-conformity. The decision could be inconclusive when both tests for conformity as well as non-conformity are performed. This is true for both one-stage and two-stage procedures. Holst et al. [69] illustrated in their example that the probability of obtaining an inconclusive result for the two-stage procedure is reduced. In other words, it is less probable to obtain an inconclusive result when a two-stage procedure is applied. This demonstrates the advantage of the two-stage procedure in declaring a more informative result. Due to the above advantages, the two-stage conformity testing procedure is adopted in the ISO standard. Although Holst et al. [69] described the objectives of the acceptance sampling plan and conformity testing, the principal differences of the two activities were not fully discussed. In addition, limiting values in the conformity testing were not distinguished from specifications. Moreover, the effect of the limiting value on the performance of conformity testing and the efficiency of the two-stage design were not investigated.

5.3.2 Fractional nonconformance

Given an upper or lower specification, an observation can be classified as conforming or nonconforming with full confidence only when measurement errors are absent or negligible. However, measurement error cannot be ignored for bulk materials because the analytical measurements are often indirect leading to large errors in measurement. To alleviate the effect of the measurement errors, Govindaraju and Jones [51] introduced the fractional nonconformance or FNC approach, which assigns a probability of nonconformance to *each* observed measurement when the measurement error distribution is known.

The measurement error model is assumed to be $Y = X + Z$, where X and Y are the true and measured values respectively, and Z is the measurement error, assumed to follow $N(0, \sigma_Z^2)$. For a random sample $(y_1, y_2, ..., y_n)$ with a given upper specification (U), fractional nonconformance of each apparent measurement is defined as:

$$\hat{p}_i = P(x_i > U) = P(z_i < y_i - U) = \Phi\left(\frac{y_i - U}{\sigma_Z}\right) \tag{5.3}$$

The sum of $\hat{p}_i$ is an inverse measure of quality and was initially proposed by Govindaraju and Jones [51] for acceptance sampling inspection. Besides $\hat{p}_i$, they also defined the fractional nonconformance conditional on the given apparent measurement Y as below:

$$\hat{p}_{c_i} = P(x_i > U|Y) = \Phi\left(\frac{(y_i - U) - k(y_i - \bar{y}_i)}{\sigma_Z\sqrt{1-k}}\right) \tag{5.4}$$

where $k = \sigma_Z^2/\sigma_Y^2$ is the variance ratio which can be large, say up to 0.5 for bulk material. The conditional fractional nonconformance statistic $\hat{p}_{c_i}$ incorporates the information from other observations through its distance from the sample mean. Hence, $\hat{p}_{c_i}$ is more sensitive to identify nonconformance. More details of fractional nonconformance and its applications in acceptance sampling, short-run production process control and guardbanding can be found in Govindaraju and Jones [51] and Zhou et al. [175, 176].

5.4 Fractional nonconformance approach for conformity testing

We introduce fractional nonconformance based conformity testing in this section and then compare its performance with the confidence interval based procedure prescribed in the ISO standard in the next section. The case with a single upper LV is discussed and can be adapted to two-sided LVs. Much of the published work in conformity testing relies on a confidence interval for the measurement result, comparing it with LV to declare conformity or nonconformity. In contrast, fractional nonconformance quantifies the measurement uncertainty as the probability of true value corresponding to an observed measurement exceeding the LV. In other words, the confidence interval approach for conformity testing is avoided. Govindaraju

and Jones [51] found that the conditional fractional nonconformance $\hat{p}_{c_i}$ is more powerful than $\hat{p}_i$ in acceptance sampling inspection, hence $\hat{p}_{c_i}$ is adopted in this study. The statistics $\bar{\delta}_1 = \frac{1}{n_1}\sum_1^{n_1}\hat{p}_{c_i}$ and $\bar{\delta}_2 = \frac{1}{n_1+n_2}\sum_1^{n_1+n_2}\hat{p}_{c_i}$ are used to evaluate the overall nonconformance level of the characteristic of interest in the first and second stages respectively. Figure 5.1 displays a flow diagram for the fractional nonconformance based two-stage conformity testing procedure.

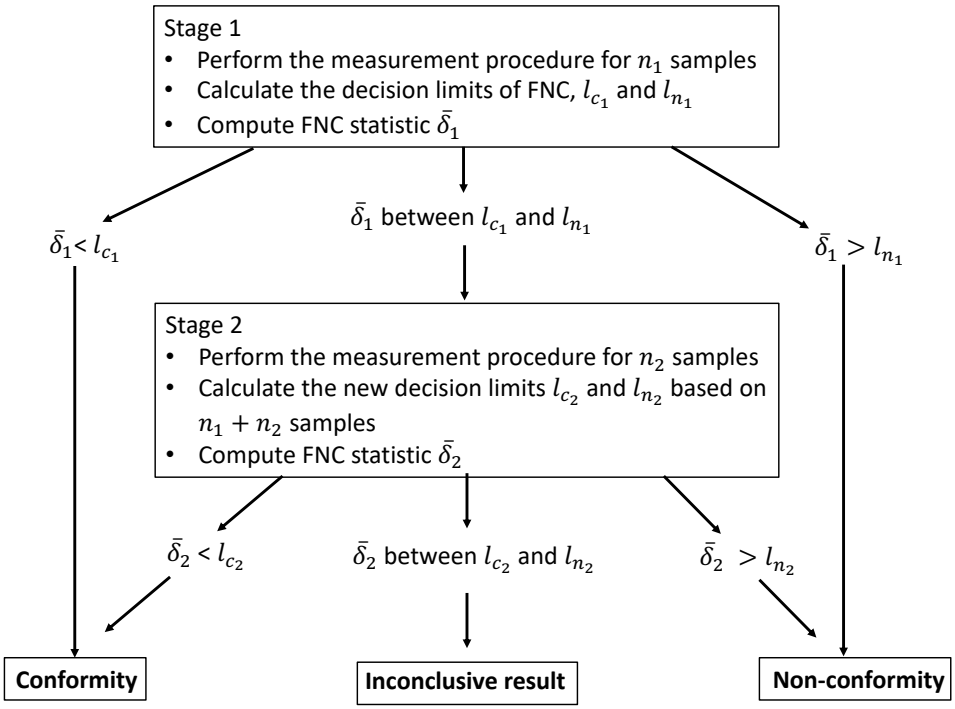


Fig. 5.1 Flow diagram for the FNC based two-stage conformity testing

The rules for asserting conformity under the fractional nonconformance based conformity testing are stated below:

- **Assurance of conformity**: $\bar{\delta}$ *is less than* l_c
- **Assurance of non-conformity**: $\bar{\delta}$ *is greater than* l_n
- **Inconclusive result**: $\bar{\delta}$ *is between* l_c *and* l_n

where l_c is the decision limit for declaring conformity, which depends on the significance level for the FNC statistic α_f. The smaller the α_f, the larger the l_c. l_n is the decision limit for declaring non-conformity, which is the extreme value of the FNC statistic under the baseline distribution.

To be consistent with the statistical measure used in Holst et al. [69], the probability of declaring conformity (P_a) is compared for the two procedures. The probability of conformity declaration, P_a, for the first and second stages of FNC based conformity testing are given as below:

$$P_a = P(\bar{\delta}_1 \leq l_{c_1}) \tag{5.5}$$

$$P_a = P(\bar{\delta}_1 \leq l_{c_1}) + P\{(l_{c_1} \leq \bar{\delta}_1 < l_{n_1}) \cap (\bar{\delta}_2 \leq l_{c_2})\} \tag{5.6}$$

Monte Carlo simulations can be used to compute P_a values for the two stages, since the distribution of $\hat{p}_{c_i}$ is intractable algebraically. The simulation algorithms adopted to compare the FNC based conformity testing and the ISO standard procedure are described below:

- Step 1. Generate a random sample from the characteristic of interest of the conforming entities (baseline distribution), say $Y \sim N(\mu, \sigma_Y^2)$. Given the baseline model, set LV for a given α_l. In other words, LV corresponds to the quantile (1-α_l) of the common cause distribution.

- Step 2. By replacing U with LV, compute the conditional fractional nonconformance statistic $\hat{p}_{c_i}$ for each single measurement using Equation 5.4. Calculate the overall nonconformance $\bar{\delta} = \frac{1}{n}\sum_1^n \hat{p}_{c_i}$.

- Step 3. Obtain the empirical distribution of $\bar{\delta}$ by repeating Step 2 (100,000 simulations). The decision limit for declaring conformity l_c corresponds to a given significance level α_f. The decision limit for declaring non-conformity l_n is the extreme quantile of $\bar{\delta}$.

- Step 4. Generate n_1 random samples in stage one. Estimate the mean and standard deviation from the first n_1 measurements and calculate confidence interval for given α_m. Claim conformity if the confidence interval is inside the region of permissible values. If the confidence interval contains LV, generate another n_2 samples in stage two. Calculate the confidence interval of the combined measurement results from stage one and two. Declare conformity if the confidence interval is inside the region of permissible values.

- Step 5. Calculate $\hat{p}_{c_i}$ and $\bar{\delta}$ for the first n_1 random values and compare $\bar{\delta}_1$ with l_{c_1} and l_{n_1}. Declare conformity if $\bar{\delta}_1 < l_{c_1}$ or non-conformity if $\bar{\delta}_1 > l_{n_1}$, otherwise generate another n_2 samples and calculate $\bar{\delta}_2$ for the $n = n_1 + n_2$ samples. Declare conformity if $\bar{\delta}_2 < l_{c_2}$.

- Step 6. Repeat Step 4 and Step 5 for 10,000 times and compute the proportion of cases of declaring conformity (P_a) for both approaches.

Besides the significance level for the measurement results (α_m) and FNC statistic (α_f), we also evaluate the effect of the significance level at the limiting values (α_l) in the proposed conformity testing, which has not been addressed in the ISO standard. Different Type A classification error rates for the limiting value are considered.

Let $Y \sim N(\mu, \sigma_Y^2)$ be the apparent measurement distribution of the conforming entities (baseline distribution) with population mean of μ. For simplicity, let $Y \sim N(0,1)$. The variance ratio $k = 0.25$ is adopted, which means 25% of the total variance is due to the measurement errors, so that measurement error $Z \sim N(0, 0.25)$. The single upper limiting value $LV = 2.576$ is set, which corresponds to an upper tail quantile ($\alpha_l = 0.005$) of the baseline distribution.

Given the empirical distribution of the FNC statistic (when $\mu = 0$), the decision limit for declaring conformity l_c under the fractional nonconformance based conformity testing is uniquely defined by α_f and the probability of declaring conformity P_a for conforming entities is only associated with α_f for any given LV. A significance level (or Type A classification error probability) for the FNC statistic of $\alpha_f = 0.01$ is assumed, which means the probability of not breaching LV is 99% for the fractional nonconformance statistic. In contrast, as the uncertainty interval is given in the form of a two-sided (1-α_m) confidence interval for the measurement results in the ISO standard, the level of significance for the measurement is $\alpha_m/2$. Hence $\alpha_m = 0.02$ is set (to correspond to $\alpha_f = 0.01$), so that we are 99% 'confident' that the measurement result fall below LV.

5.5 Comparison of FNC and ISO conformity testing procedures

We compare the performance of the proposed approach with the current procedure prescribed in the ISO standard in this section. According to Equations 5.1 and 5.2, when the entities are conforming, P_a, probability of declaring conformity, increases with n for the ISO procedure, as the confidence interval of the measurement results becomes shorter with larger sample sizes. For the special case when the sample size is one, the confidence interval of the measurement requires prior knowledge of σ_Y. Hence, we separate the cases $n_1 > 1, n_2 > 1$ and $n_1 = n_2 = 1$ in our comparison.

5.5.1 Case I: $n_1 > 1, n_2 > 1$, σ_Y is estimated from sample

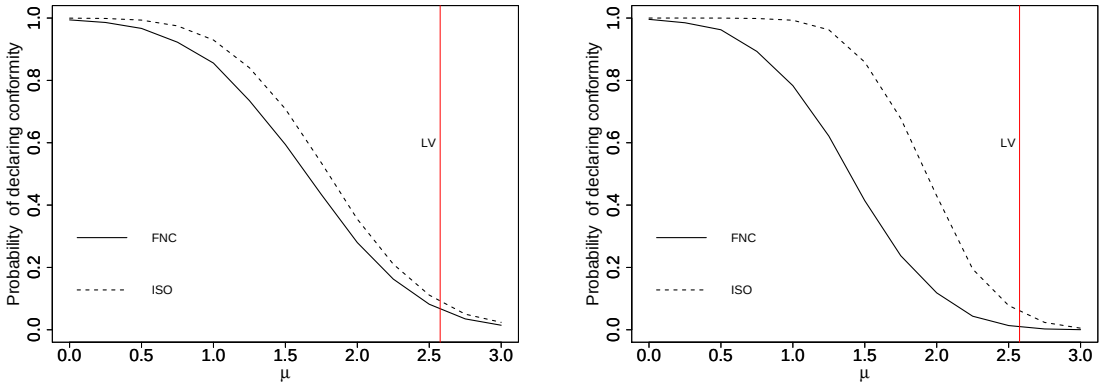
When the sample size is greater than one, both mean and variance of the sample can be estimated from the sample. n_1 samples are examined in the first stage, and conformity declared if $\bar{\delta} < l_c$ or non-conformity if $\bar{\delta} > l_n$, otherwise another n_2 samples will be tested if an inconclusive result is obtained in the first stage. The empirical distribution of the FNC statistic was obtained by a large scale simulation (10^6) and the decision limits for the

fractional nonconformance based conformity testing are listed in Table 5.2 for five different scenarios. In the first scenario of $n_1 = n_2 = 3$, both FNC and ISO approaches have the same P_a when the entities are conforming. In other words, P_a is matched for the baseline parameters.

Table 5.2 Decision limits for fractional nonconformance based conformity testing

Scenario	Sample size		Significance level			Stage 1		Stage 2	
	n_1	n_2	α_l	α_m	α_f	l_c	l_n	l_c	l_n
I	3	3	0.005	0.02	0.01	0.086	0.44	0.051	0.17
II	5	5	0.005	0.02	0.01	0.064	0.23	0.038	0.12
III	3	3	0.001	0.02	0.01	0.011	0.25	0.007	0.12
IV	3	3	0.005	0.10	0.05	0.011	0.34	0.010	0.17
V	1	1	0.005	0.10	0.05	0.031	0.99	0.012	0.51

Following the steps in the simulation algorithm, the operating characteristics (OC) curves of the FNC and ISO two-stage conformity testing procedures are shown in Figure 5.2(a), where the probability of declaring conformity (P_a) is plotted against the population mean of the entity (μ). As illustrated, the fractional nonconformance based conformity testing results in the same P_a for conforming entities ($\mu = 0$) but much lower P_a when entities are nonconforming ($\mu > 0$). In other words, the two-stage conformity testing procedure using the fractional nonconformance statistic reduces the probability of incorrect declaration of conformity or inconclusive result for nonconforming entities (Type B classification error). Therefore, we conclude that the fractional nonconformance based two-stage conformity testing is more discriminatory, especially when the entities are nonconforming.

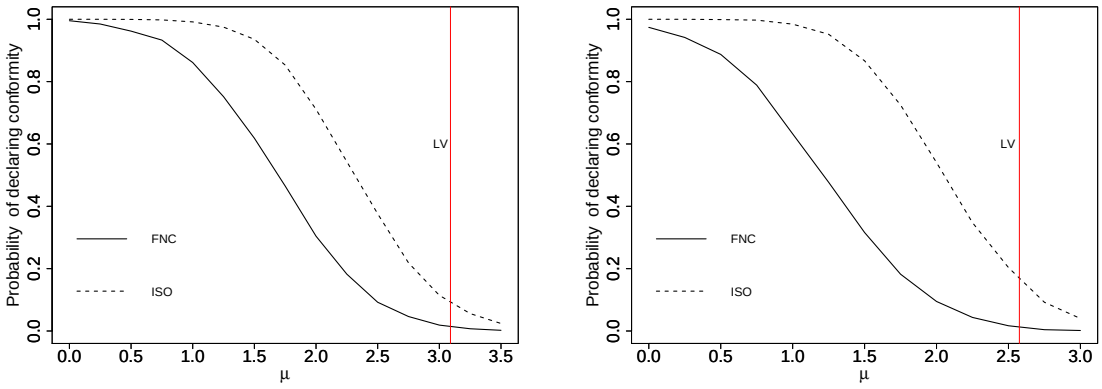


(a) $n_1 = n_2 = 3$ (Scenario I) (b) $n_1 = n_2 = 5$ (Scenario II)

Fig. 5.2 OC curves of two conformity tests ($\alpha_l = 0.005$, $\alpha_m = 0.02$, $\alpha_f = 0.01$)

As shown in Figure 5.2(b), the OC curve of the ISO two-stage procedure is inflated when the sample size increases to $n_1 = n_2 = 5$. According to Equation 5.1, $PW(\mu)$ of the ISO

conformity testing increases with sample size for given LV. In other words, P_a increases with n in the ISO two-stage procedure. As a result, conformity is wrongly declared even for a large value of μ, which means that the ISO two-stage procedure becomes less sensitive when the entity is nonconforming for large sample sizes. In contrast, for the fractional nonconformance approach, P_a for nonconforming entities decreases when the sample size increases to $n_1 = n_2 = 5$. This is because the conditional fractional nonconformance is more sensitive in identifying nonconformance; see Zhou et al. [175]. When the observations accumulate, the probability of the conditional FNC approach not detecting the nonconformance becomes lower. Hence, the FNC two-stage approach sacrifices a little when entities are conforming but becomes more powerful for identifying the nonconforming entities when sample size increases. In other words, the superiority of the FNC two-stage procedure over the ISO two-stage procedure is that of considerably lower probability of incorrect declaration of conformity or inconclusive result when the entities are nonconforming (Type B classification error), but with a marginally higher probability of declaring non-conformity or inconclusive results when the entities are conforming (Type A classification error). We also find that this superiority does not depend on the choice of k when we repeat the test with different variance ratios $k = 0.36$ and $k = 0.16$; see Figure 5.8 in Appendix 5.A.

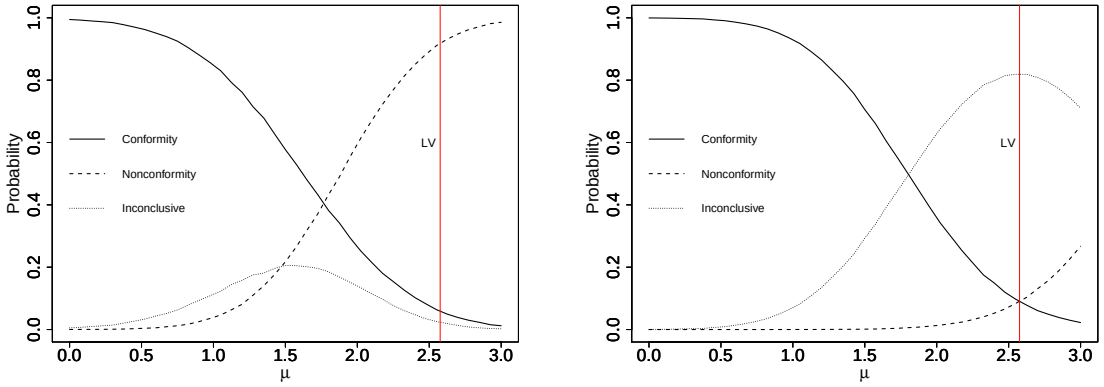


(a) $\alpha_l = 0.001$, $\alpha_m = 0.02$, $\alpha_f = 0.01$ (b) $\alpha_l = 0.005$, $\alpha_m = 0.10$, $\alpha_f = 0.05$

Fig. 5.3 OC curves of two conformity tests (Scenario III and Scenario IV)

When a more extreme LV, say $\alpha_l = 0.001$ quantile of the baseline distribution is adopted, the OC curves of the two conformity tests have similar patterns. In other words, the advantage of the FNC procedure does not depend on the selection of LV or α_l too much; see Figure 5.3(a). We also investigate the effect of significance level of measurement results (α_m) and FNC statistic (α_f) on the performance of the two approaches. When a larger significance level, say $\alpha_m = 10\%$ is employed, the confidence interval of the measurement results becomes more compact, hence a higher P_a is expected. As shown in Figure 5.3(b), the FNC procedure has a more discriminatory OC curve than the ISO procedure. In other words, the FNC

two-stage conformity testing results in a much lower Type B classification error with the price of a little higher Type A classification error, compared to the ISO two-stage approach.



(a) FNC two-stage procedure

(b) ISO two-stage procedure

Fig. 5.4 OC curves of two conformity tests ($n_1 = n_2 = 3, \alpha_l = 0.005, \alpha_m = 0.02, \alpha_f = 0.01$)

Another advantage of the two-stage conformity testing in Holst et al. [69] is that the two-stage procedure reduces the probability of an inconclusive result. We plotted the probabilities of obtaining various outcomes, including conformity, non-conformity and inconclusive result, of the FNC and ISO two-stage conformity testing in Figure 5.4. The solid curve represents the test for conformity, which gives the probability of declaring conformity on the vertical axis. The dashed curve gives the probability of declaring non-conformity when the test is performed for non-conformity. The dotted line is the probability of obtaining an inconclusive result, which is much lower when the FNC two-stage procedure is used, especially when μ is approaching LV. In other words, the FNC two-stage conformity testing is less likely to deliver an inconclusive result. This is not surprising as LV is very likely to be enclosed by the confidence interval of the measurement results when the underlying distribution mean is closed to LV. In contrast, the FNC approach transforms the uncertainty into a probability, thereby avoiding the construction of the confidence interval. This advantage of the FNC procedure holds for different sample sizes and significance levels; see Figures 5.9 and 5.10 in Appendix 5.B.

5.5.2 Case II: $n_1 = n_2 = 1$, σ_Y is known

In some situations, only a single measurement may be made in the first stage of the conformity testing. For example, in drug testing, a single blood sample of an individual is taken and tested. The sample will be retested only if an inconclusive result is obtained in the first stage. Another example is doping testing: a urine sample is taken from the participant, which will be divided into two subsamples; the second subsample will be measured if conformity or

non-conformity cannot be asserted in stage one. The confidence interval cannot be determined with a sample size of one, hence the ISO two-stage conformity testing protocol can only be invoked with the assumption of known σ_Y. For FNC two-stage conformity testing, the measurement from the first stage will be calculated as the unconditional FNC statistic ($\hat{p}_i$) and the measurement in the second stage can be calculated conditionally on the single result obtained in stage one.

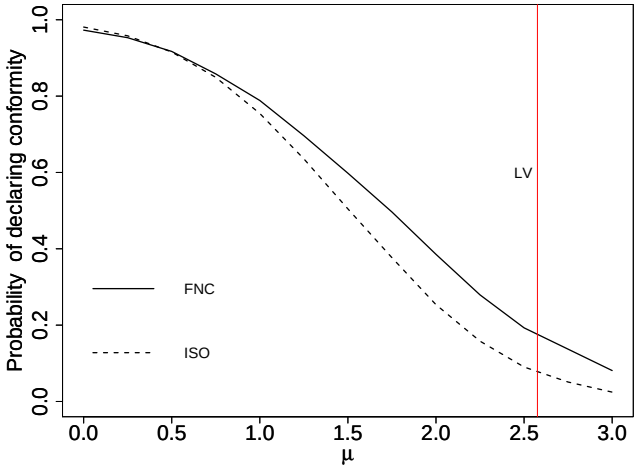
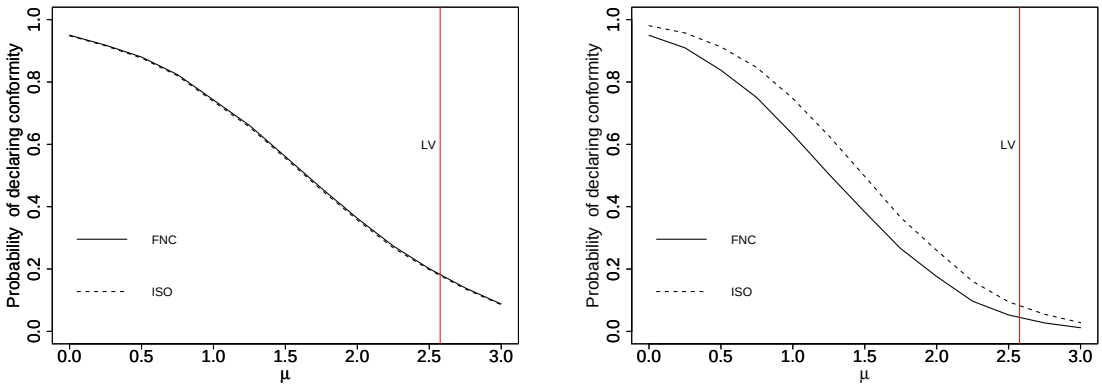


Fig. 5.5 OC curves of two conformity tests (Scenario V)

As before, assume $Y \sim N(0,1)$ and $\alpha_l = 0.005$ corresponding to $LV = 2.576$. Given the sample size $n_1 = n_2 = 1$, $\alpha_m = 0.10$ and $\alpha_f = 0.05$ are selected for the ISO and FNC two-stage procedure respectively. The OC curves of the two procedures are plotted in Figure 5.5. The results show that the ISO two-stage procedure is superior to the FNC approach with lower Type B classification error when the entities are nonconforming. The main reason is that the conditional approach cannot be applied in the first stage when $n_1 = 1$. The unconditional FNC statistic is not as sensitive for nonconformance as the conditional one, which leads to a high P_a in the first stage, and this will not be lowered even though the conditional FNC is adopted in the second stage. The inferiority of the FNC two-stage procedure for the $n_1 = 1$ case is due to lack of power. In other words, the conditional FNC approach cannot improve the power for small sample sizes and has no advantage over the ISO approach.

Instead, if the one-stage procedure is applied when $n_1 = 1$, Figure 5.6(a) shows that the OC curves of the one-stage FNC and ISO procedures are overlapping when Type A classification error probability is matched, which indicates similar performance. An alternative option for the FNC procedure to deal with the $n_1 = n_2 = 1$ case is to skip the decision making in the first stage and only compare $\hat{p}_{c_i}$ with the decision limit in the second stage using two samples in stages two. This improves the performance of the FNC procedure and the OC curves of the two procedures are almost parallel; see Figure 5.6(b).



(a) One-stage ($n_1 = 1$) (b) $n_1 = n_2 = 1$

Fig. 5.6 OC curves of two conformity tests

5.5.3 Effect of limiting value

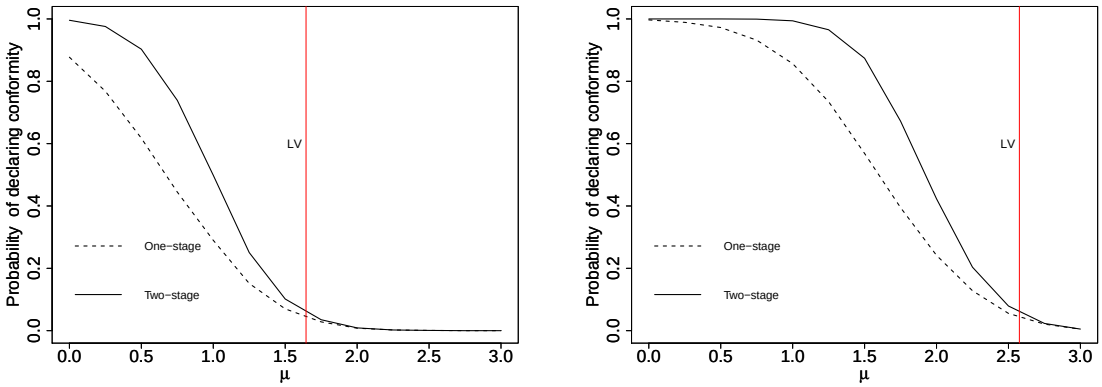
In the previous sections, we compared the performance of the FNC and ISO two-stage conformity testing procedures and found that, when $n_1 > 1, n_2 > 1$, the FNC two-stage procedure can considerably lower the probability of declaring conformity or inconclusive result for nonconforming entities (Type B classification error), but with a slightly higher probability of declaring non-conformity or inconclusive result for conforming entities. The classification accuracy of the FNC procedure becomes much more pronounced when n increases. The proposed procedure does not depend on the significance level or LV too much and hence is robust. Table 5.3 presents a summary of how well the FNC and ISO procedures performed in the five scenarios considered.

Table 5.3 Summary of performance for the two conformity tests

Scenario	Classification error	
	Type A	Type B
I	Similar	Higher for ISO
II	Little higher for FNC	Much higher for ISO
III	Similar	Higher for ISO
IV	Little higher for FNC	Much higher for ISO
V	Similar	Higher for FNC

According to the ISO standard, the advantage of the two-stage procedure over the one-stage procedure is the considerably higher P_a for entities close to LV, but with a slightly higher P_a for nonconforming entities. This property was established fixing the limiting value. In this section, we investigate the effect of limiting value on the performance of the conformity testing procedures in the ISO standard.

As before, assume $Y \sim N(0,1)$ is the baseline distribution and LV=1.645 corresponds to a significant level $\alpha_l = 0.05$; the superiority of the two-stage procedure is demonstrated in Figure 5.7(a). This result is not surprising and was discussed in Holst et al. [69]. However, the impact of LV selection on the performance of the conformity testing was not evaluated. As shown in Figure 5.7(b), the gains of the two-stage procedure are moderate when a higher LV=2.576 (corresponding to $\alpha_l = 0.005$) is set, as P_a for conforming entities increases only in the middle part. The reason is that both LV and n are positively associated with P_a when the test for conformity is performed. When LV is defined as an extreme quantile, say corresponding to $\alpha_l = 0.005$, P_a for conforming entities is quite large in the first stage and increases only moderately in the second stage. In other words, the two-stage design is not efficient in this situation. Hence, the superiority of the two-stage procedure in the ISO standard depends on the selection of LV, and the two-stage procedure should be properly designed when LV is set very high; otherwise the second stage might not be necessary.



(a) $LV = 1.645$ ($\alpha_l = 0.05$)

(b) $LV = 2.576$ ($\alpha_l = 0.005$)

Fig. 5.7 OC curves of ISO conformity testing ($n_1 = n_2 = 5$)

5.6 Case study

In the previous section, we demonstrated that the FNC two-stage procedure is more powerful than the ISO two-stage procedure when more than one sample is taken for conformity testing. A web based *Shiny* app is given at https://zhouxin07.shinyapps.io/conformity/ to visualize and compare the two conformity testing procedures with different parameters. This app is also available at http://shiny.massey.ac.nz/kgovinda/conformity/. In this section, we provide a step by step example for practitioners to implement the fractional nonconformance based conformity testing.

Recall the examples illustrated in Section 5.3.1 where inconclusive results are obtained with the ISO two-stage conformity testing procedure. We apply the FNC two-stage conformity testing to Example I with the following steps:

- Step 1. $\sigma_Y^2 = 0.038\%$ is estimated based on the samples from the first stage. Given LV=0.1%, $\sigma_Y^2 = 0.038\%$ and assuming $\alpha_l = 0.005$, the empirical mean of the conforming entities $\mu_0 = 0.002\%$ is calculated.

- Step 2. The baseline distribution is assumed to be normally distributed with $\mu_0 = 0.002\%$ and $\sigma_Y^2 = 0.0038\%$, which was estimated in Step 1. Obtain the empirical distribution of $\bar{\delta}$ by simulation. The significance level $\alpha_f = 2.5\%$ (corresponding to $\alpha_m = 5\%$) The decision limits for declaring conformity and non-conformity are $l_{c_1} = 0.025\%$ and $l_{n_1} = 0.202\%$ respectively.

- Step 3. $\sigma_Z = 0.019\%$ is calculated when the variance ratio $k = 0.25$ is assumed, $\hat{p}_{c_i}$ for each single measurement from the first stage ($n_1 = 5$) are calculated. The overall nonconformance $\bar{\delta}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} \hat{p}_{c_i} = 0.247\%$.

Since $\bar{\delta}_1$ is greater than l_{n_1}, the second stage test is not necessary and non-conformity is declared based on the rules under fractional nonconformance based conformity testing. We also revisit Example II and non-conformity is asserted with the proposed method. These examples confirm the superiority of the fractional nonconformance based conformity test in obtaining a conclusive result.

The conformity testing procedure, which provides benefits for both manufacturers and consumers, is also applicable in manufacturing contexts. Källgren et al. [88] provided two examples showing the importance of conformity testing in industry. Here is another example to illustrate the usefulness of the proposed approach in the dairy industry. The manufacturer aims to evaluate the conformity of protein composition (in percent) for a small whole milk powder lot using five sample test results: {24.45, 24.16, 24.15, 24.22, 24.26}. The lower limiting value for percentage protein is set at 24.1, which corresponds to $\alpha_l = 0.005$. The 95% confidence interval is: $24.248 \pm (2.776 \times 0.122)/\sqrt{(5)} = (24.097, 24.399)$. The test is inconclusive because the LV is located within the uncertainty interval, hence five more samples are tested in the second stage: {24.13, 24.10, 24.11, 24.70, 24.08}. After pooling the observations from the two stages, the 95% confidence interval becomes $(24.096, 24.376)$, which still contains the LV. Hence, an inconclusive result is obtained under the ISO standard conformity testing procedure. Alternatively, we apply the FNC two-stage testing procedure to evaluate conformity of the same data. The overall nonconformance in stage one is calculated as $\bar{\delta}_1 = 0.0295$, which is between $l_{c_1} = 0.025$ and $l_{n_1} = 0.202$. After combining with observations from the second stage, the overall nonconformance becomes $\bar{\delta}_2 = 0.18$, which is greater than $l_{n_2} = 0.103$. Hence non-conformity is declared with the proposed procedure.

5.7 Discussion

In this study, we have developed a two-stage conformity testing procedure using the fractional nonconformance statistic, and compare its performance with the conformity testing procedure prescribed in the international guidelines for the evaluation of conformity. Much of the published work in the conformity testing literature assumes that the LV is externally imposed as a specification limit. For conformity evaluation in the area of legal metrology, the set LV must be distinguished from a specification limit. The proposed procedure links LV to a common cause model for the measurement results.

Our research shows that the FNC two-stage conformity testing procedure reduces the probability of incorrect declaration of conformity or inconclusive result for nonconforming entities (Type B classification error) when the number of test samples is greater than one, and this superiority becomes more significant when n increases. In addition, the proposed approach substantially reduces the probability of inconclusive decisions. These advantages are demonstrated using examples, which showed that the FNC based conformity testing is desirable. Our study found that the fractional nonconformance based procedure is preferable when more than one sample is tested for conformity evaluation. When only one sample is tested, the current two-stage procedure in the ISO Standard is recommended due its simplicity.

We also considered different false alarm rates for the set LV and investigated the effect of LV on conformity testing. The results show that the FNC two-stage conformity testing is robust, as it does not depend on the selection of LV. In our future research, we intend to study the efficiency of the FNC two-stage conformity test with unequal sample sizes and provide an optimum design for the fractional nonconformance based conformity test when the total sample size is fixed. FNC based two-sided conformity testing research work is also in progress.

Appendix 5.A OC curves of two conformity tests with different k

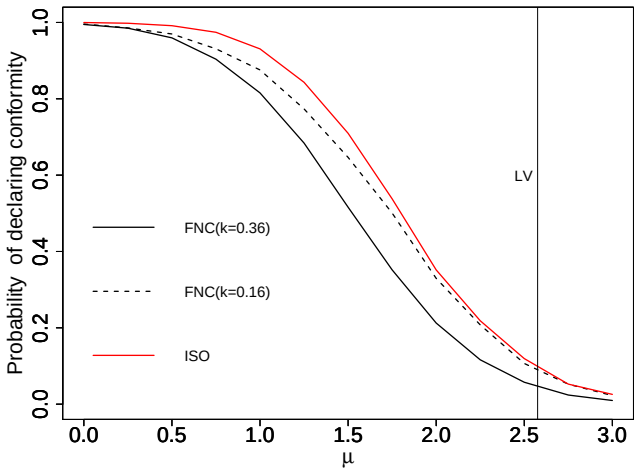
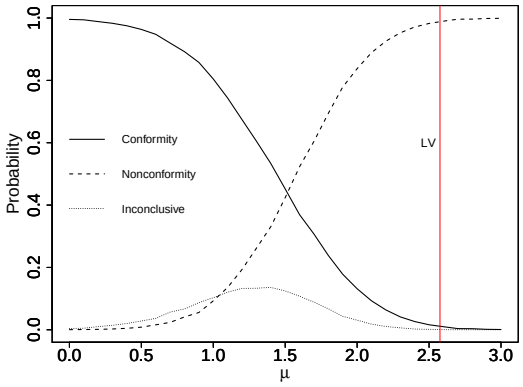
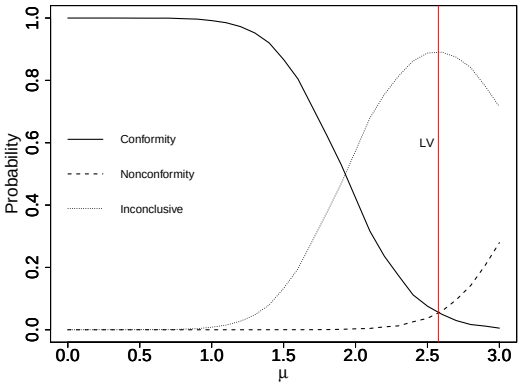


Fig. 5.8 $n_1 = n_2 = 3, \alpha_l = 0.005, \alpha_m = 0.02, \alpha_f = 0.01$

Appendix 5.B OC curves of two conformity tests

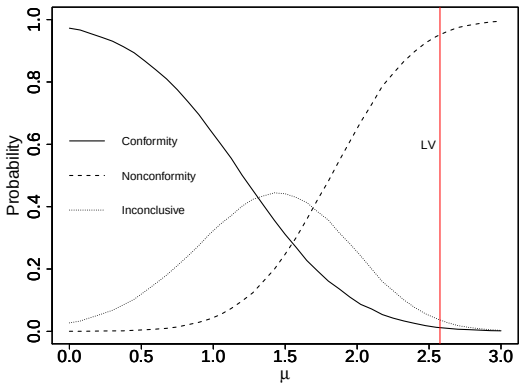


(a) FNC two-stage procedure

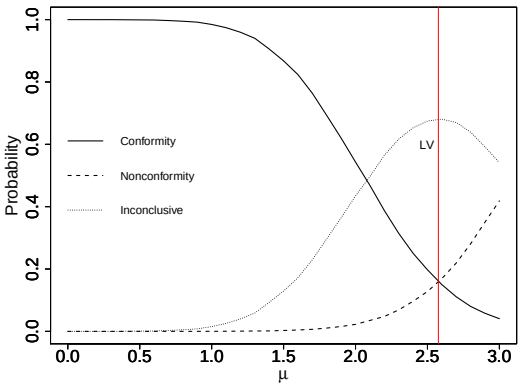


(b) ISO two-stage procedure

Fig. 5.9 $n_1 = n_2 = 5, \alpha_l = 0.005, \alpha_m = 0.02, \alpha_f = 0.01$



(a) FNC two-stage procedure



(b) ISO two-stage procedure

Fig. 5.10 $n_1 = n_2 = 3, \alpha_l = 0.005, \alpha_m = 0.10, \alpha_f = 0.05$

Chapter 6

Modified FNC based conformity testing using subject-specific limiting value

6.1 Abstract

Conformity testing procedures are used to systematically examine whether an entity conforms to a specific requirement or regulatory standard. In legal metrology, the concept of limiting value (LV) is intrinsically linked to a common cause model of the population. The population-based LV is widely used in conformity testing, such as doping screening test, health evaluation etc. However, the generic population-based LV does not consider individual heterogeneous factors, and hence it may result in lower specificity. In this Chapter, we develop a modified FNC conformity testing procedure using subject-specific LV in conjunction with the population-based LV. The proposed two-stage approach mitigates the effect of measurement errors in addition to accommodating individual differences. Our research shows that the modified conformity testing procedure reduces the probability of incorrect declaration of non-conformity and inconclusive result for conforming subjects whose mean level is naturally high.

Keywords

Bayesian inference; conformity testing; fractional nonconformance; population-based LV; subject-specific LV

6.2 Introduction

Conformity testing is used to ensure the compliance of an entity to a specified requirement and/or regulatory standard. Applications of conformity testing include the analysis of a sportsman's urine for control of doping, the examination of alcohol level in driver's blood, the test for a minimum/maximum limit for fat in milk powder and many others. ISO 10576-1 International Standard [79] (Guidelines for the evaluation of conformity with specified requirements) illustrates the conformity testing procedure to test the concentration of some trace metal in blood. In this conformity testing protocol, the characteristic of interest is compared to the set limiting value (LV), which refers to the upper and/or lower bounds of the permissible values.

Measurement uncertainty (MU) is unavoidable in conformity testing resulting in the risk of erroneous decisions; see Weißensee et al. [153]. The effect of MU on decision making and risks has been discussed in many studies, such as Pendrill [117], Forbes [42], Wu and Govindaraju [166] and Maleki et al. [103]. As suggested in ISO/IEC Guide 98-3:2008 [81], measurement uncertainty is explicitly defined and expressed in the form of a confidence interval. Much of the publish work, i.e. Pendrill [119], Hibbert [66], Holst et al. [69] compare the MU interval with LV and declare conformity/non-conformity only when MU interval is inside the region of permissible/non-permissible values. However, this approach is ineffective when the underlying mean level is closer to LV. In contrast, Govindaraju and Jones [51] proposed the *fractional nonconformance* (FNC) approach, which transforms the uncertainty into a probability and avoids construction of the confidence interval. Zhou et al. [177] extended this approach to conformity testing and developed an FNC based two-stage conformity testing procedure, which is shown to be superior to the current two-stage procedure stipulated in the ISO 10576-1 International Standard [79].

LV is defined as the specified value of the characteristic giving upper and/or lower bounds of the permissible value in ISO 10576-1 International Standard [79]. Zhou et al. [177] distinguished the LV from a specification limit and linked LV to a common cause model of the population. In brief, LV for declaring conformity should be based on the probability distribution of conforming entities in the population so that a false positive error is associated with LV. The population-based LV is widely adopted in a screening test for doping. For example, an athlete is considered suspicious if the test urinary testosterone/epitestosterone (T/E) ratio is greater than 4, where 4 is a population-based LV corresponding to a false positive rate at 3.8%; see Sottas et al. [140]. However, the unique and non-specific population-based LV is criticized by many researchers as it does not consider individual heterogeneous factors, such as gender, ethnicity and genotype. Ellis and Nyborg [37] pointed out the existence of racial variations in male testosterone levels. A deletion mutation in UGT2B17 gene is shown to be associated with testosterone level in Jakobsson et al. [82]. Łagowska and Kapczuk [95] found that nutritional factors, energy availability, frequency and age at the

beginning of training are associated with testosterone levels in the blood when comparing female athletes and ballet dancers with menstrual disorders. Cook et al. [27] showed further subgrouping of elite and non-elite athletes comparing the baseline levels of free testosterone and cortisol concentrations.

The World Anti-Doping Agency (WADA) are implementing the Athlete Biological Passport (ABP) to monitor an individual's biomarkers of doping, where individual-thresholds can be established by longitudinal monitoring; see the latest ABP operating guidelines published by WADA [151]. The ABP testing paradigm has been implemented widely in anti-doping, and discussed by many researchers, such as Saugy et al. [131], Sottas et al. [141], Robinson et al. [127] and Vernec [149]. However, this approach is not applicable when a subject-specific reference range is not available. Sharpe et al. [137] proposed a subject-specific threshold to improve the detection of doping. An individual's baseline can be established with at least one subject-specific measurement and then a z score is derived by assuming a universal variance for all subjects. Sottas et al. [140] proposed to combine population-based limits with individual-based data and developed individual-specific thresholds with a Bayesian method. This approach outperformed four other tests in a longitudinal study for detecting testosterone abuse; see Sottas et al. [143]. Robinson et al. [128] and Strahm et al. [145] applied this approach in the detection of abnormal hematological values, and examining the variability of steroid profiles in soccer players from different countries respectively. Schulze et al. [135] combined this approach with genotype information to enhance the sensitivity of the Bayesian analysis.

Subject-specific LV is being used in literature for drug testing and anti-doping but none of them considered the effect of measurement error. Hence the logical extension is to apply the conformity testing procedures, which take measurement uncertainty into consideration, for drug testing. However, the subject-specific LV is rarely discussed in conformity testing literature. In this Chapter, we establish a subject-specific threshold using a Bayesian approach and develop a modified FNC based conformity testing procedure using subject-specific LV. The proposed conformity testing procedure explicitly incorporates the effect of measurement uncertainty, as well as individual heterogeneity. This new method has not been reported in conformity testing or drug testing literature to the best of our knowledge. The analysis shows that the modified conformity testing improves the performance of the traditional FNC conformity testing developed in Zhou et al. [177].

The remainder of this Chapter is organized as follows. We present a brief review of the current FNC based conformity testing procedure and the Bayesian approach to conformity testing in Section 6.3. The modified FNC based conformity testing is introduced in Section 6.4 and its performance is investigated in Section 6.5. A case study is provide in Section 6.6 to demonstrate the proposed approach for drug testing. The last section gives a brief discussion.

6.3 Review of FNC based conformity testing and Bayesian approach

This section provides a brief review of the FNC based conformity testing procedure and the Bayesian approach to conformity testing. The definitions and notations are listed in the Table 6.1.

Table 6.1 Definitions and notations

Term	Description
Y	Apparent measurement result of the entity
Z	Measurement error, $Z \sim N(0, \sigma_Z^2)$
X	True measurement in the measurement error model $Y = X + Z$
μ, σ_Y	Population mean and SD
μ_i, σ_{Y_i}	Mean and SD of the subject i
Δ_i	Shift size or incrementation in mean after drug intake for subject i
U	Upper specification
n_1	Sample size for Stage I
n_2	Sample size for Stage II
n	Total sample size, $n = n_1 + n_2$
k_1	Error/response variance ratio, $k_1 = \sigma_Z^2/\sigma_Y^2$
k_2	Intra/inter variance ratio, $k_2 = \sigma_{Y_i}^2/\sigma_Y^2$
LV_p	Population-based limiting value
LV_s	Subject-specific limiting value
$\hat{p}_j$	Fractional nonconformance of the j-th observation
$\hat{p}_{c_j}$	Conditional fractional nonconformance of the j-th observation
$\bar{\delta}$	Overall nonconformance level, $\bar{\delta} = \frac{1}{n}\sum_{j=1}^{n} \hat{p}_{c_j}$
α_l	False alarm rate of LV
α_f	False alarm rate of FNC statistic
l_c	Decision limit for declaring conformity (negative or not suspicious)
l_n	Decision limit for declaring non-conformity (positive or no start)
P_a	Probability of declaring conformity
P_r	Probability of declaring non-conformity
se	Sensitivity, the test's ability to detect the true positives
sp	Specificity, the test's ability to detect the true negativities

6.3.1 Fractional nonconformance based conformity testing

Conformity testing is defined as a systematic examination of the extent to which an entity conforms to a specified criterion in ISO 10576-1 International Standard [79], which provides guidelines for the evaluation of conformity with specified requirements. In line with the practice in the conformity testing literature, measurement uncertainty is reported as an uncertainty interval of the measurement result and compared to LV, which is often considered as an externally imposed specification limit. Conformity/non-conformity is only assured if

the uncertainty interval is entirely located in the region of permissible/non-permissible values, otherwise an inconclusive result is declared if LV is inside the uncertainty interval. When sample size is small or those entities with permissible values are close to LV, it becomes more likely that the uncertainty interval overlaps with LV. As a result, the probability of obtaining an inconclusive result becomes high.

Govindaraju and Jones [51] proposed fractional nonconformance (FNC) approach, which assigns a probability of nonconformance to each observed measurement when the measurement error distribution is known. FNC quantifies the measurement uncertainty in probabilistic terms rather than constructing a confidence interval. The FNC statistic for a given apparent measurement $y_j, j = 1,2,...n$ is defined as:

$$\hat{p}_j = P(x_j > U) = P(z_j < y_j - U) = \Phi\left(\frac{y_j - U}{\sigma_Z}\right) \tag{6.1}$$

where x_j and z_j are the true measurement and measurement error respectively under the measurement error model $Y = X + Z$. Alternatively, the conditional FNC statistic $\hat{p}_{c_j}$ which is shown to be more sensitive to identify nonconformance in Zhou et al. [175], is defined as below:

$$\hat{p}_{c_j} = P(x_j > U|Y) = \Phi\left(\frac{(y_j - U) - k(y_j - \bar{y_j})}{\sigma_Z\sqrt{1-k_1}}\right) \tag{6.2}$$

The FNC approach has been implemented for acceptance sampling, short-run process monitoring and guardbanding. Zhou et al. [177] extended this approach to conformity testing and developed an FNC based two-stage conformity testing procedure with following rules:

- **Assurance of conformity**: $\bar{\delta}$ *is less than* l_c
- **Assurance of non-conformity**: $\bar{\delta}$ *is greater than* l_n
- **Inconclusive result**: $\bar{\delta}$ *is between* l_c *and* l_n

where $\bar{\delta} = \frac{1}{n}\sum\hat{p}_{c_j}$ is the overall fractional nonconformance level, and l_c and l_n are the decision limits for declaring conformity and non-conformity respectively, which depend on the false alarm rate α_f set for the FNC statistic.

Confidence interval calculation is averted in the FNC based conformity testing procedure, and the probability of an inconclusive decision is substantially reduced, when compared to the conformity testing procedure prescribed in the ISO standard. Zhou et al. [177] also distinguished LV from an arbitrary specification limit in legal metrology and linked it to a common cause model for the measurement results.

6.3.2 Bayesian approach

The Bayesian approach is a method of statistical inference using Bayes theorem to update the prior information when new information becomes available. The idea behind the Bayesian

approach is to integrate prior knowledge about unknown parameters with the information from observed data and summarize the results in the form of a posterior distribution. Bayesian techniques have been widely used in the literature, including conformity testing, i.e. Carobbi and Pennecchi [18] and Weißensee et al. [153]. See Gelman et al. [46] for a thorough introduction to Bayesian statistics. Bayesian methods have become increasingly popular in sports lately; see a recent comprehensive review given by Santos-Fernandez et al. [130].

Bayesian methods fit well within the FNC two-stage conformity testing framework. The subject-specific LV_s is established using the Bayesian approach in the first stage, and it will replace the population-based LV_p in the second stage of the conformity testing. In Bayesian inference, the prior knowledge often comes from previous studies. For example, an informative prior could be adopted if some historical data are available for a subject. Otherwise noninformative priors are used when there is insufficient information, and the research output (posterior distribution) could be used as informative prior for future studies.

6.4 Modified FNC conformity testing using subject-specific LV

In traditional FNC conformity testing, LV is intrinsically connected to a common cause model, which means that the population-based LV is defined as an upper tail quantile of the population baseline distribution. For example, given a random sample $y_1, y_2, ..., y_i, ..., y_n$ from the population, the baseline distribution is assumed to be $Y \sim N(\mu_0, \sigma_Y^2)$, where both μ_0 and σ_Y^2 are known, so that LV_p is defined as the population-based limiting value, corresponding to a false alarm rate (α_l). Similarly, for a random sample $y_{i_1}, y_{i_2}, ..., y_{i_n}$ coming from a particular subject i, the subject-specific LV_s can be defined in the same manner when the subject-specific distribution, $N(\mu_i, \sigma_{Y_i}^2)$, is known or can be estimated. The subject-specific distributions may not be identical; the true value of subject mean μ_i, which may vary from the population μ_0, is often unknown except when historical data is available. The intra-individual variability $\sigma_{Y_i}^2$ may be smaller than the inter-individual variability σ_Y^2. We employ the Bayesian method to estimate the subject's baseline mean μ_i and then set up a subject-specific limiting value for subject i assuming known $\sigma_{Y_i}^2$.

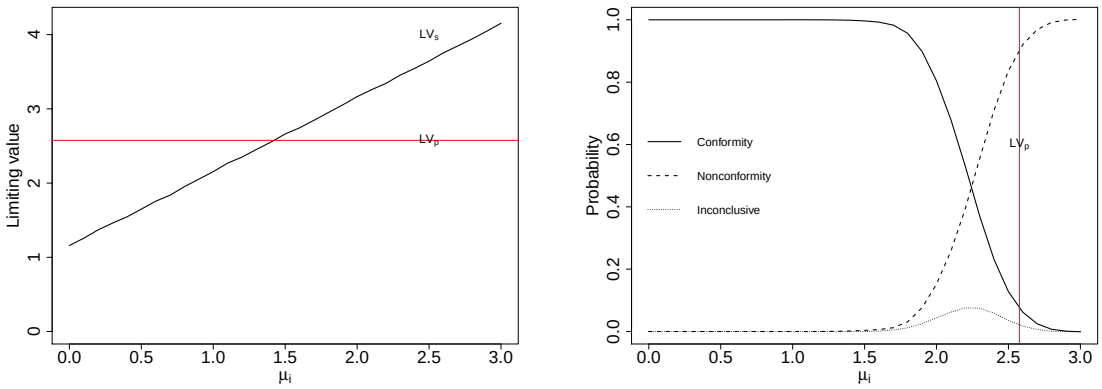
For simplicity, let $Y \sim N(0,1)$ be the (known) population baseline distribution and $Y \sim N(\mu_i, \sigma_{Y_i}^2)$ be the subject-specific baseline distribution. The intra/inter individual variance ratio $k_2 = \sigma_{Y_i}^2/\sigma_Y^2$ is assumed to be known for any subject so that $\sigma_{Y_i}^2$ is also known. For a normal prior $N(\mu_{i_0}, \sigma_{i_0}^2)$ for μ_i and normal likelihood $N(\mu_i, \sigma_{Y_i}^2)$ with known $\sigma_{Y_i}^2$, the posterior distribution of μ_i can be calculated algebraically as $N(\hat{\mu}_i, \hat{\sigma}_{\mu_i}^2)$ given the observations $y_{i_1}, y_{i_2}, ..., y_{i_n}$ from subject i, where:

$$\hat{\mu}_i = \hat{\sigma}_{\mu_i}(\mu_{i_0}\sigma_{i_0}^{-2} + \sum_{j=1}^{n} y_{i_j}\sigma_{Y_i}^{-2}) \tag{6.3}$$

$$\hat{\sigma}^2_{\mu_i} = (\sigma^{-2}_{i_0} + n\sigma^{-2}_{Y_i})^{-1} \tag{6.4}$$

see Gelman et al. [46] for mathematical derivation.

Alternatively, the problem can be solved by simulation, for example using the package *rjags* developed by Plummer et al. [120]; see JAGS models in Appendix 6.A. This enables models of greater complexity to be handled, for example when $\sigma^2_{Y_i}$ is unknown but a prior distribution for it is available. Given $\sigma^2_{Y_i} = 0.2$ and the posterior mean of μ_i, the subject-specific LV_s for subject i can be calculated for a given false alarm rate, say $\alpha_l = 0.5\%$.



(a) LV_p vs LV_s (b) Conforming subject with different μ_i

Fig. 6.1 Traditional FNC conformity testing using population-based LV_p

Figure 6.1(a) illustrates the population-based $LV_p = 2.576$ versus subject-specific LV_s for conforming subjects with mean $\mu_i \in [0,3]$. The subject-specific LV_s is more stringent when $\mu_i < 1.5$ due to the smaller intra-individual variability $\sigma^2_{Y_i}$ and becomes more relaxed when the subject mean μ_i is closer to the population-based LV_p. The operating characteristics (OC) curves of the traditional FNC conformity testing using population-based LV_p are plotted in Figure 6.1(b). When the mean of conforming subject is small, say $\mu_i < 1.5$, the traditional FNC conformity testing works well, with 100% specificity and no false positives. However, when the conforming subject's mean is approaching LV_p, the traditional FNC conformity testing using non-specific population-based LV_p leads to a large probability of false declaration of non-conformity or an inconclusive result. Therefore, the population-based LV_p is not suitable when $\mu_i > 1.5$. We propose a modified FNC two-stage conformity testing procedure, which uses the subject-specific LV_s in Stage II of the conformity testing. The proposed approach accommodates individual differences, and is suitable for subjects with higher μ_i. The flow diagram of modified FNC conformity testing procedure is shown in Figure 6.2. The following step-by-step guide describes the modified FNC conformity testing procedure:

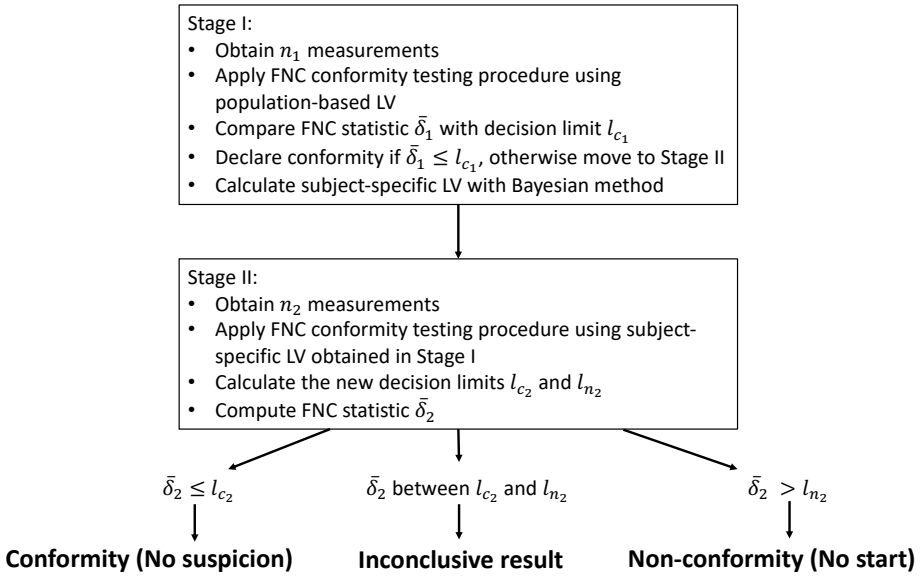


Fig. 6.2 Flow diagram for the modified FNC conformity testing

Step 1. Apply the traditional FNC conformity testing in Stage I.

An initial sample of size n_1 samples are tested by traditional FNC conformity testing procedure using population-based LV_p. Assuming the population baseline distribution, say $Y \sim N(0,1)$, the population-based $LV_p = 2.576$ is set for given $\alpha_l = 0.5\%$. The error/measurement variance ratio $k_1 = \sigma_Z^2/\sigma_Y^2$ is assumed to be known, say $k_1 = 0.25$, so that measurement error is $Z \sim N(0, 0.25)$. Given LV_p and Z, the FNC statistics of the n_1 samples are calculated. If conformity cannot be assured, proceed to the second stage.

Step 2. Establish the subject-specific LV_s from the given subject-specific distribution.

Given the prior information and n_1 samples from Stage I, subject mean μ_i is estimated using the Bayesian approach. Subject variance $\sigma_{Y_i}^2$ is also available when intra/inter individual variance ratio $k_2 = \sigma_{Y_i}^2/\sigma_Y^2$ is known. Given the estimated subject-specific baseline distribution $Y \sim N(\hat{\mu}_i, \sigma_{Y_i}^2)$ and $\alpha_l = 0.5\%$, the subject-specific LV_s can be calculated.

Step 3. Apply the FNC conformity testing using LV_s in Stage II.

Another n_2 samples are obtained in Stage II and assessed for the FNC conformity testing, using the subject-specific LV_s calculated in Step 2. In contrast to Stage I, where normal distribution $Y \sim N(0,1)$ is adopted, the n_2 samples in Stage II are evaluated with the subject-specific distribution $Y \sim N(\hat{\mu}_i, \sigma_{Y_i}^2)$, which is $Y \sim N(\hat{\mu}_i, 0.2)$ given $k_2 = 0.2$. Note that the measurement error becomes $Z \sim N(0, 0.05)$ when the constant $k_1 = 0.25$ is adopted. Decision limits l_c and l_n correspond to upper tail quantiles, say $\alpha_f = 1\%$ and $\alpha_f = 0.1\%$ of empirical

distribution of the FNC statistics for declaring conformity and non-conformity respectively in Stage II, are also recalculated based on the subject-specific distribution.

The traditional FNC conformity testing procedure does not consider individual-specific differences, since the population-based LV_p is used for all subjects and universal decision limits for the FNC statistics are adopted in both stages. Instead of using a multilevel model, the modified FNC based two-stage conformity testing procedure combines two single level models via population-based LV_p in Stage I and subject-specific LV_s in Stage II. The main reason for this approach is that single level models are easy to understand and widely used in conformity testing and doping literature; see Sharpe et al. [137], Sottas et al. [140], Holst et al. [69] and others. Moreover, a multilevel model fits well to a two-class decision making procedure (conforms or not) whereas drug testing protocols need a three-class decision making procedure. In the proposed approach, decisions are made in two stages. Stage II is implemented only when the subject fails under the population-based LV_p in Stage I. Conformity testing using subject-specific LV_s is then adopted in the second stage, which accommodates individual differences. Non-conformity is only declared in the second stage of the modified FNC conformity testing, which reduces the probability of misclassification for a conforming subject whose mean level is naturally higher than others.

6.5 Performance of the modified FNC conformity testing

In this section, we compare the performance of the modified FNC conformity testing procedure using subject-specific LV_s with the traditional FNC approach considering various scenarios. The population-based limiting value is calculated as $LV_p = 2.576$, corresponding to $\alpha_l = 0.5\%$ for the population baseline distribution $N(0,1)$. The error/measurement variance ratio $k_1 = 0.25$ and intra/inter variance ratio $k_2 = 0.2$ are assumed. A non-informative normal prior $N(0,100)$ is assigned to μ_i and $n_1 = n_2 = 3$ sample measurements are randomly generated for the two stages. We only consider a conforming subject with large mean because traditional FNC approach performs well when $\mu_i < 1.5$, and small μ_i is also not practically important for dopping or drug testing.

6.5.1 Monte Carlo simulations

The probability of declaring conformity P_a and non-conformity P_r in the modified FNC conformity testing are given as below:

$$P_a = P(\bar{\delta}_1 \leq l_{c_1}) + P\{(\bar{\delta}_1 > l_{c_1}) \cap (\bar{\delta}_2 \leq l_{c_2})\} \tag{6.5}$$

$$P_r = P\{(\bar{\delta}_1 > l_{c_1}) \cap (\bar{\delta}_2 > l_{n_2})\} \tag{6.6}$$

Monte Carlo simulations are used to compute P_a and P_r, since the distribution of $\hat{p}_{c_i}$ is intractable algebraically. The simulation algorithms adopted for the traditional and modified FNC based conformity testing procedures are described below:

Step 1. Traditional FNC conformity testing

- Step 1.1 Generate a random sample from the baseline distribution, say $Y \sim N(0,1)$. Given the baseline model, set population-based LV_p for a given α_l.
- Step 1.2 Compute the conditional fractional nonconformance statistic $\hat{p}_{c_i}$ for each single measurement using Equation 6.2 and calculate the overall nonconformance $\bar{\delta} = \frac{1}{n}\sum_{j=1}^{n}\hat{p}_{c_j}$.
- Step 1.3 Obtain the empirical distribution of $\bar{\delta}$ by repeating Step 1.2 (100,000 simulations). The decision limits for declaring conformity (l_c) and non-conformity (l_n) are defined, to correspond to $\alpha_f = 1\%$ and $\alpha_f = 0.1\%$ respectively.
- Step 1.4 Calculate $\bar{\delta}_1$ for the first n_1 random samples from subject i in Stage I and compare $\bar{\delta}_1$ with l_c and l_n. Declare conformity if $\bar{\delta}_1 \leq l_c$ or non-conformity if $\bar{\delta}_1 > l_n$, otherwise generate another n_2 samples in Stage II and calculate $\bar{\delta}_2$. Declare conformity if $\bar{\delta}_2 \leq l_c$ or non-conformity if $\bar{\delta}_2 > l_n$.

Step 2. Modified FNC conformity testing

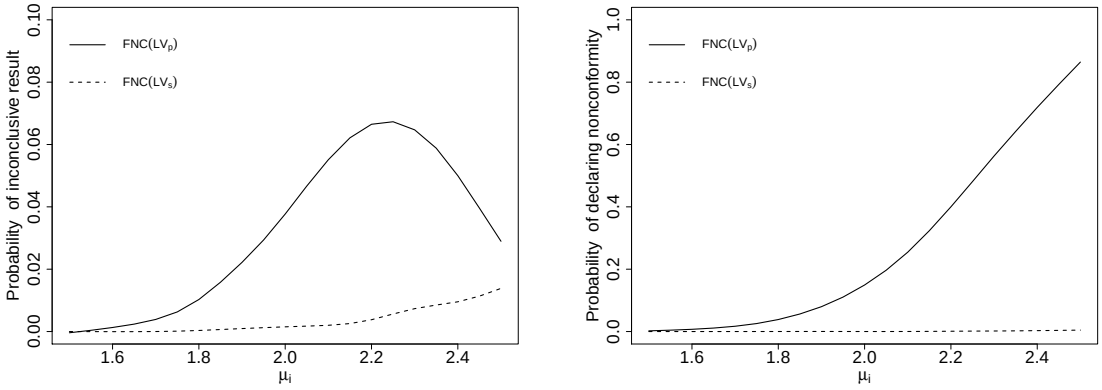
- Step 2.1 Estimate subject mean μ_i from the n_1 samples in Stage I using the Bayesian method. Given $\sigma^2_{Y_i}$ and posterior mean of the subject, calculate subject-specific LV_s for the same α_l adopted in Step 1.1.
- Step 2.2 Repeat Steps 1.2 - 1.3 by replacing population-based LV_p with subject-specific LV_s and calculate new decision limits l_{c_2} and l_{n_2}.
- Step 2.3 In Stage I, only declare conformity if $\bar{\delta}_1 \leq l_c$, otherwise move on to Stage II. Declare conformity if $\bar{\delta}_2 \leq l_{c_2}$ or non-conformity if $\bar{\delta}_2 > l_{n_2}$.

Step 3. Repeat Steps 1 and 2 for 2,000 times and compute the proportion of cases declaring conformity (P_a) and non-conformity (P_r) for each approach.

6.5.2 Scenario I: Subject is conforming

As discussed in Section 6.4, the traditional FNC conformity testing does not work well for a conforming subject with large mean, which results in a high probability of incorrect declaration of non-conformity or inconclusive result. Following the simulation algorithm, the

OC curves of the two FNC conformity testing procedures were obtained and are shown in Figure 6.3, where the probabilities of an inconclusive result and declaring non-conformity are plotted against the conforming subject mean $\mu_i \in [1.5, 3]$ respectively. The results show that the modified FNC conformity testing procedure, which uses subject-specific LV_s, significantly reduces the probability of declaring non-conformity or an inconclusive result when μ_i is high. This is not surprising because the population-based LV_p is too stringent for a subject with large μ_i, but the subject is declared as conforming using a more relaxed LV_s in the second stage. Hence, the modified FNC approach improves the *specificity* of the conformity testing procedure, especially when the conforming subject's mean is close to the population-based limiting value.



(a) Probability of inconclusive result

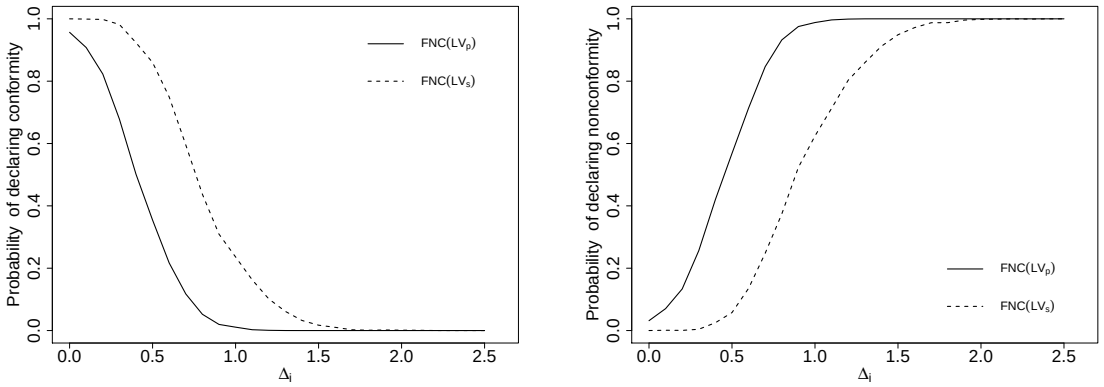
(b) Probability of declaring non-conformity

Fig. 6.3 OC curves of traditional and modified FNC conformity tests for conforming subjects

The error/measurement variance ratio k_1 will vary for different laboratory instruments and/or methods, and so may the intra/inter variance ratio k_2. Variance ratios $k_1 = 0.01$ (rarely measurement error), $k_1 = 0.5$ (large measurement error), $k_2 = 0.01$ (small intra-subject variance) and $k_2 = 0.5$ (large intra-subject variance) are also considered for comparison. In addition, we also evaluate the performance of the two FNC conformity testing procedures when the sample size increases to $n_1 = n_2 = 10$ or a more extreme $LV_p = 3.09$ ($\alpha_l = 0.1\%$) is applied. The OC curves of traditional and modified FNC conformity testing procedures shown in Figures 6.6 - 6.11 (Appendix 6.B) have the same pattern. Hence we conclude that the performance of the modified FNC based conformity test for a conforming subject with large mean is somewhat robust and does not depend on the selection of k_1, k_2, LV_p or sample size too much.

6.5.3 Scenario II: Subject is nonconforming

The modified FNC approach, which uses subject-specific LV_s, is demonstrated to possess better specificity when the subject is conforming. In this section, we compare the *sensitivity* of the two FNC conformity testing procedures when the subject becomes nonconforming. An artificial shift $\Delta_i \in [0,3]$ is added into the conforming subject with naturally high baseline mean (say $\mu_i = 1.8$) before the first sample is taken, to mimic the incrementation after drug intake.



(a) Probability of declaring conformity (b) Probability of declaring non-conformity

Fig. 6.4 OC curves of traditional and modified FNC conformity tests for subject with $\mu_i = 1.8$

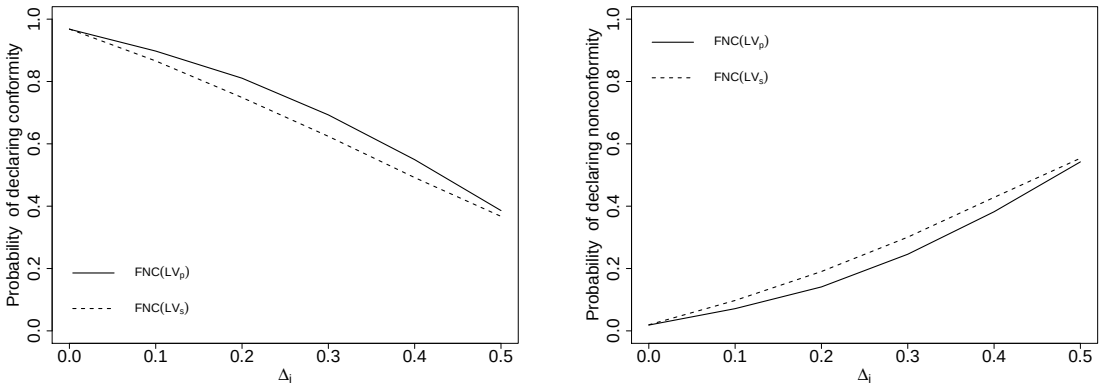
Figure 6.4 shows that the modified FNC conformity testing results in a higher P_a, compared to the traditional FNC conformity testing procedure when $\Delta_i = 0$. This is not surprising and was discussed in Scenario I. However, the modified approach is not as sensitive as the traditional FNC conformity testing for detecting non-conformity when the subject becomes nonconforming. When shift size is large, say $\Delta_i > 1.5$, the performances of the traditional and modified FNC approaches are similar.

6.5.4 Scenario III: Optimum design for conforming subject with large μ_i

In Scenario II, we fixed α_f for the two FNC conformity testing procedures: $\alpha_f = 0.1\%$ as the cut-off for declaring non-conformity and $\alpha_f = 1\%$ as the cut-off for declaring conformity. For conforming subjects with higher mean, say $\mu_i = 1.8$, the results show that the modified FNC conformity testing has a higher P_a for conforming subject but lower P_r for detecting non-conformity when a shift in the mean level occurs. In other words, no universal advantage exists exclusively to either of the two conformity testing procedures. In this section, we provide an optimum design of the modified FNC approach for a conforming subject with higher mean.

According to Equation 6.5, P_a is determined only by the decision limits of the FNC statistic, which depend on false alarm rate α_f. In other words, P_a of the modified FNC conformity testing can be adjusted by selecting a range of values for α_f. Instead of adopting the same $\alpha_f = 1\%$, we aim to optimize the sensitivity of the modified FNC conformity testing by using different α_f values, given that P_a for the two FNC procedures are matched when the subject is conforming.

For the traditional FNC conformity testing, the decision limits and cut-offs for declaring conformity and non-conformity remain the same, so that the P_a and P_r performance do not change. In order to maintain the same level of P_a for a conforming subject, an optimum decision limit can be established using numerical methods for the modified FNC conformity testing procedure. After a grid search, $\alpha_f = 14\%$ was found so that the corresponding decision limit l_c for declaring conformity ensures the same $P_a = 0.96$ for the two approaches when $\Delta_i = 0$. Similarly, the cut-off $\alpha_f = 9\%$ for declaring non-conformity, is chosen so that $P_r = 0.025$ is fixed for both approaches.



(a) Probability of declaring conformity (b) Probability of declaring non-conformity

Fig. 6.5 OC curves of traditional and modified FNC conformity tests (optimum design)

The OC curves of the optimally designed modified FNC conformity testing approaches are plotted in Figure 6.5. Only small shift size was considered, since the performance of the two FNC approaches was shown to be similar for detecting large shifts in Scenario II. Figure 6.5 shows that the modified FNC conformity procedure has the same P_a and P_r when the entity is conforming but is more sensitive to declare non-conformity when the entity is nonconforming. In other words, the modified FNC conformity testing increases the probability of declaring non-conformity for nonconforming entities without sacrificing its specificity for conforming entities. Therefore, the modified FNC conformity testing with optimum design is superior to the traditional approach for conforming subjects with large means.

In this section, we have evaluated the performance of the modified FNC conformity testing procedure using a subject-specific limiting value under various scenarios. A summary

Table 6.2 Summary of performance for the modified FNC conformity testing procedure

Scenario	Conforming mean (μ_i)	Shift (Δ_i)	Modified FNC approach
I	small ($\mu_i \leq 1.5$)	0	Similar to traditional approach
	large ($\mu_i > 1.5$)	0	Reduces probability of inconclusive result & false declaring non-conformity
II	large ($\mu_i = 1.8$)	small	Increases specificity but decrease sensitivity
		large	Similar to traditional approach
III	large ($\mu_i = 1.8$)	small	Increases sensitivity with similar specificity

is presented in Table 6.2. Specificity is vitally important in from legal perspective, and is preferred over having higher sensitivity in doping tests; see Sottas et al. [142] and Delanghe et al. [30]. Our results show that the modified FNC conformity testing works better, in terms of specificity, especially for conforming subjects with large means.

6.6 Case study

As illustrated in the ISO 10576-1 International Standard [79], the conformity testing procedure is used for health evaluation, and this approach can be extended to drug testing. It is up to the anti-doping authorities to determine the cut-offs used for declaring drug abuse. For example, McHugh et al. [107] suggested a false alarm rate of 1/10,000 for the detection of growth hormone abuse in sport. Sottas et al. [143] used 1/1,000 and 1/30,000 as the cut-offs for suspicious and positive result, respectively, for T/E ratio. Sharpe et al. [137] proposed two thresholds for detecting erythropoietin abuse in athletes: 1 in 100 cut-off for *suspicious* and 1 in 1,000 cut-off for *no start*. When the FNC conformity testing procedure is applied, non-conformity is declared if his/her FNC level exceeds the decision limit for declaring non-conformity l_n, which corresponds to a given false alarm rate, say $\alpha_f = 0.01\%$ or $\alpha_f = 0.1\%$.

As an application of the modified FNC conformity testing procedure to detect testosterone abuse in elite sports, a web-based *Shiny* app written in R has been developed. The app is hosted at https://zhouxin07.shinyapps.io/conformity2/. The default subject-specific LV is calculated based on subject-specific mean and SD, where the subject-specific mean is estimated using the Bayesian method, which combines the prior information and observed data, while the SD is kept the same for all subjects by assuming a fixed variance ratio k_2. In practice, the subject-specific mean and SD might be available from historical data. Hence the *Shiny* app allows the user to define subject-specific LV arbitrarily based on historical data.

A case study will now be discussed. The relevant data are from many anti-doping laboratories, with details of the material used and methods for data collection described in Sottas et al. [143]. According to a recent study reported in Ahrens and Butch [4], the range of urinary T/E ratios for 17,813 athletes is between 0.5 and 1.5. The population-based $LV_p = 4$ is selected by WADA, which corresponds to a false positive rate of $\alpha_l = 3.8\%$.

The dataset {4.6, 5.6, 5.9, 6.5} shows the T/E values from an athlete which are all greater than the population-based $LV_p = 4$. For the modified FNC conformity testing, the subject-specific $LV_s = 6.16$ is obtained. Consider the test results of {4.6, 5.6, 5.9} forming the subject-specific baseline in Stage I. Given $LV_s = 6.16$, the FNC statistic for the new test result of 6.5 is calculated as 0.937, which is less than decision limit for declaring conformity $l_c = 0.996$. Hence, conformity is declared with the modified FNC conformity testing. This conclusion matches with the negative result from the Bayesian test in Sottas et al. [143].

6.7 Discussion

The universal population-based LVs are widely adopted in conformity testing procedures. However, this traditional approach does not consider the different individual means that occur naturally and hence results in lower specificity. In this Chapter, we developed a modified FNC conformity testing procedure, which uses subject-specific LV in combination with population-based LV. The newly proposed two-stage conformity testing procedure mitigates the effect of measurement errors, and then accommodates subject-specific differences by using the subject-specific LV in Stage II. We compared the performance of the proposed approach with the traditional FNC conformity testing under various scenarios. The simulation results showed that the modified FNC approach reduced the probabilities of incorrectly declaring non-conformity or an inconclusive result for the conforming subject whose true mean is close to population-based LV. Moreover, the modified FNC approach with optimum design increases sensitivity for declaring non-conformity when the subject turns nonconforming after taking drugs. The proposed approach is also demonstrated using clinical data.

The current conformity testing procedures usually have three possible outcomes: conformity, inconclusive result and non-conformity, which corresponds in drug testing to negative, suspicious and positive respectively. However, inconclusive result is also an artifact of the small sample size.

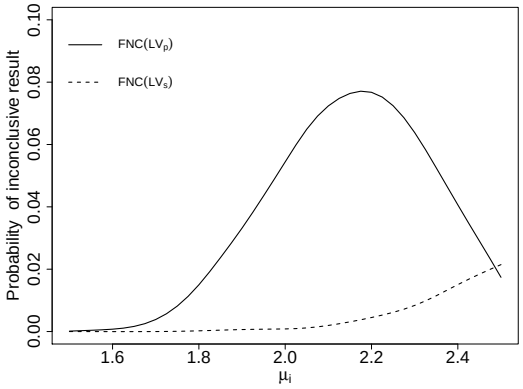
The intra-individual variability is assumed to be constant for each subject. This assumption may not be realistic for some biomarkers in drug testing. An extended study involving subject-specific variation using the Bayesian approach is deferred for further research. In addition, this Chapter considered only a normal distribution for the population and a normal prior for the subject. Extensions to other possible distributional assumptions are also left for future research.

Appendix 6.A Models in JAGS

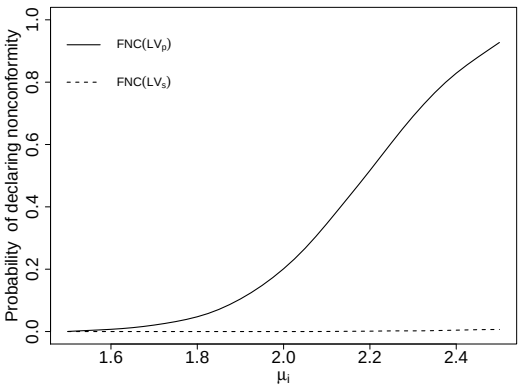
A.1 R codes to obtain the posterior distribution of μ_i

```
model {
mu_i ~ dnorm(mu0,1/sd0^2)        #prior

for (i in 1:length(y)){
y [i]~ dnorm(mu_i,1/sd_yi^2)     #likelihood
}
}

#data
y = c(y1,y2,...yn)
mu0 = 0
sd0 = 100
k2 = 0.2
sd_yi = sqrt(k2)
```

Appendix 6.B OC curves of traditional and modified FNC conformity tests with different variance ratios k_1, k_2, sample sizes n and population-based LV_p

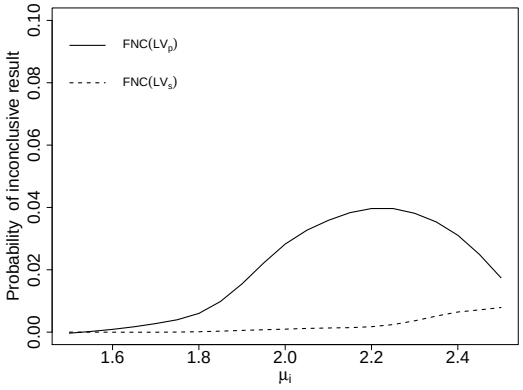


(a) Probability of inconclusive result

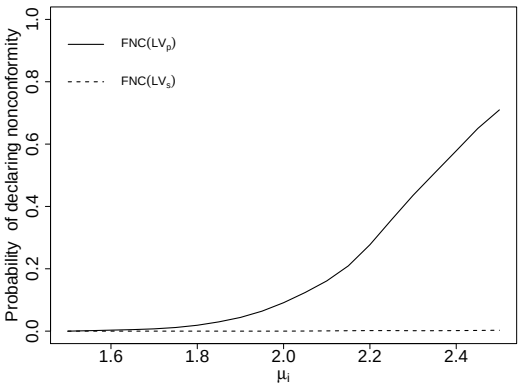


(b) Probability of declaring non-conformity

Fig. 6.6 OC curves of traditional and modified FNC conformity tests ($k_1 = 0.5$)

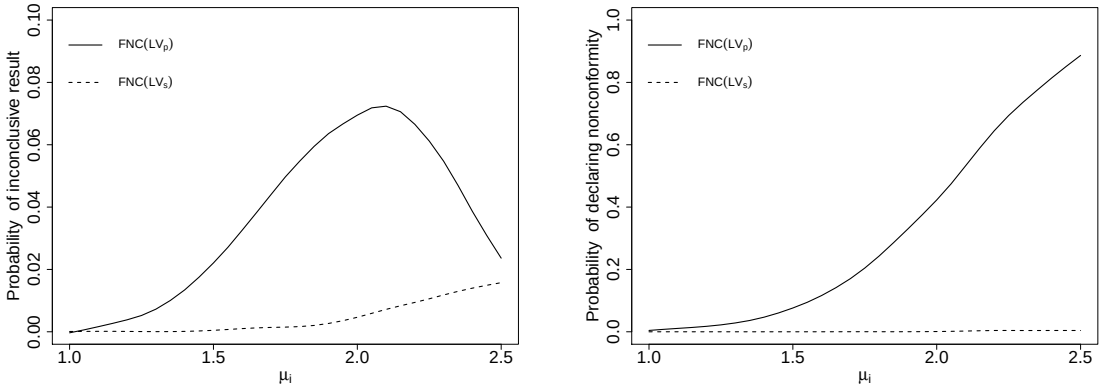


(a) Probability of inconclusive result



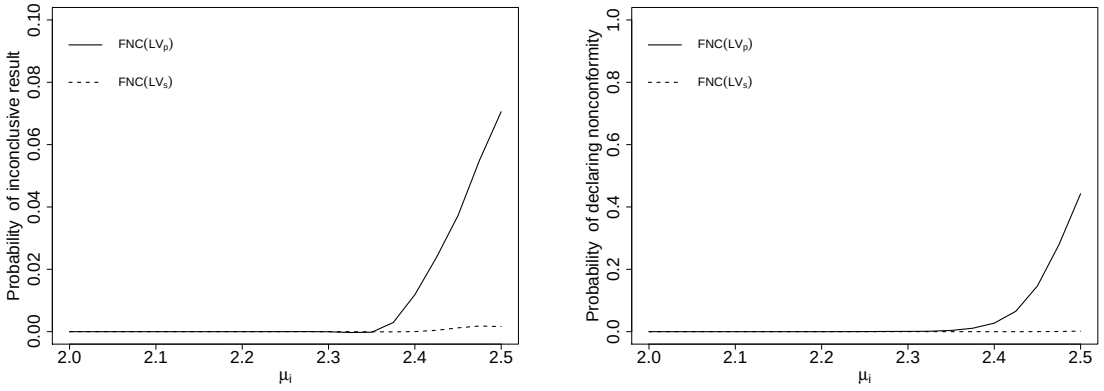
(b) Probability of declaring non-conformity

Fig. 6.7 OC curves of traditional and modified FNC conformity tests ($k_1 = 0.01$)



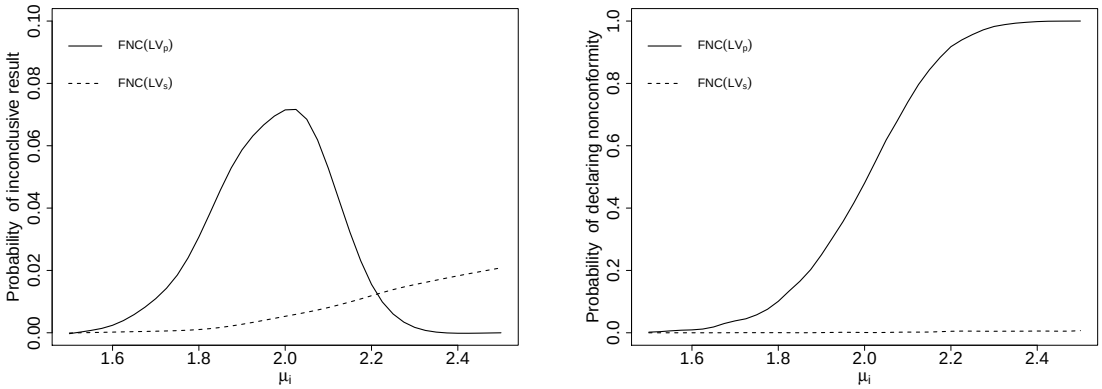
(a) Probability of inconclusive result (b) Probability of declaring non-conformity

Fig. 6.8 OC curves of traditional and modified FNC conformity tests ($k_2 = 0.5$)



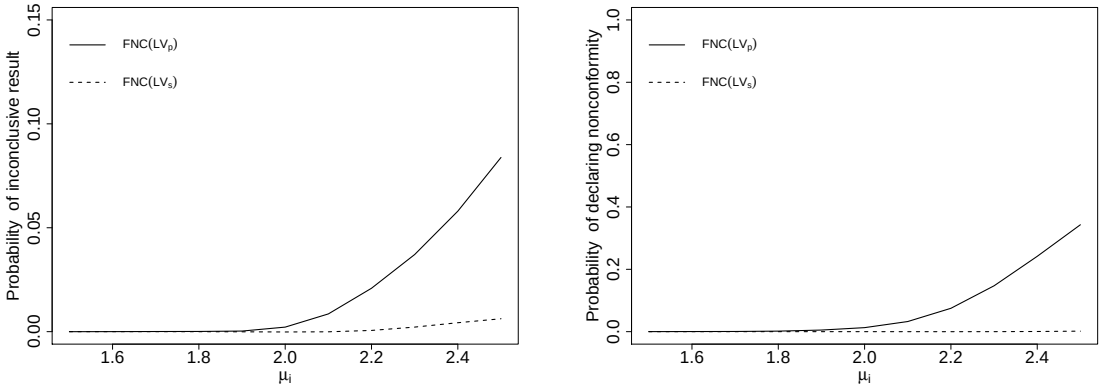
(a) Probability of inconclusive result (b) Probability of declaring non-conformity

Fig. 6.9 OC curves of traditional and modified FNC conformity tests ($k_2 = 0.01$)



(a) Probability of inconclusive result (b) Probability of declaring non-conformity

Fig. 6.10 OC curves of traditional and modified FNC conformity tests ($n_1 = n_2 = 10$)



(a) Probability of inconclusive result (b) Probability of declaring non-conformity

Fig. 6.11 OC curves of traditional and modified FNC conformity tests ($LV_p = 3.09$)

Chapter 7

Conclusions and future work

7.1 Conclusions

New Zealand is one of the leading milk product exporting countries in the world. Dairying makes a significant contribution to the New Zealand economy. According to the Ministry of Primary Industries, the dairy sector is New Zealand's biggest export earner. Quality and safety assurance play a key role for the export market dominance of the New Zealand dairy sector. Measurement of various compositional qualities in milk products using near infra-red (NIR) instruments is not error-free. Measurement errors have an impact on the performance of quality and safety monitoring schemes. The fractional nonconformance or FNC statistic was developed for the real-life problem of inspecting the compositional quality of manufactured milk products. The FNC approach takes into account the measurement error in an individual measurement, and quantifies the probability of nonconformance. The main assumption made is that the distribution of the measurement errors is known. For the application of FNC, the metrological process must be kept in a good state of statistical control with negligible bias in the instrument calibrations.

The FNC approach has been implemented for acceptance sampling inspection in a New Zealand leading dairy manufacturer to evaluate the batch-wise fraction nonconforming levels. The assessment of individual fraction nonconforming at the time of sampling is also important for dairy product manufacturing. In this thesis, I expanded the FNC approach to four FNC statistics and developed FNC based monitoring schemes to monitor the individual fraction nonconforming in the dairy production process. The existing methods in the process control literature are inadequate to evaluate the true fraction nonconforming under measurement uncertainty for short-run production processes. The proposed FNC control charting procedures fill this gap in the literature.

Consumer's risk is the priority in the food industry due to the nature of the product. Consumer's risk is high for short-run production, especially at the beginning of the process. For better protection for the consumers, the guardband technique has been incorporated in

FNC based acceptance control chart scheme in this thesis. The proposed MACCS scheme reduces the consumer's risk significantly for short-run production and a large guardband is recommended at the initial stage of the production.

The fractional nonconformance assessment approach has also been extended beyond quality control and industrial applications in this thesis, such as screening for doping. Doping in sport is a challenging issue for all sportsmen and their countries, which can lead to the demonization and exclusion of a country from major international sport events. The FNC approach has been extended to conformity testing for doping screening. The proposed FNC based conformity testing takes the effect of both measurement errors and individual heterogeneous factors into consideration, and hence improves the accuracy and specificity of the traditional testing method.

Five pieces of work for fractional nonconformance assessment, applicable for food quality assurance and other applications, are presented in this thesis. The newly developed methods are illustrated with real data from dairy product manufacturing and anti-doping laboratories. In addition, interactive web-based *Shiny* applications, with step-by-step guidance to implement FNC analytics, are developed for practitioners.

To be specific, the FNC approach was implemented for short-run process monitoring of individual measurements in Chapter 2. Four different monitoring schemes of fractional nonconformance were evaluated in the form of Shewhart charts under various process models. Based on the scenarios examined, the choice of the monitoring scheme does not seem to heavily depend on the distribution of the quality characteristic of interest.

The FNC monitoring scheme was extended to EWMA and CUSUM charts in Chapter 3. The performance of these types of control charting was examined for short run monitoring of FNC. It was found that EWMA and CUSUM charting methods are useful for detection of both small and large shifts and are preferable to Shewhart control charting when the set false alarm rate is small.

In Chapter 4, we proposed a modified acceptance control chart scheme to monitor a short-run process and to simultaneously dispose the product manufactured from the process. The guardband technique was also incorporated in the proposed scheme to reduce the impact of measurement errors. The optimum process level can be set based on a risk-based model or alternatively a cost-based model can be employed.

The FNC approach was extended to conformity assessment in Chapter 5. A two-stage conformity testing procedure using the FNC statistic is shown to outperform the ISO standard procedure for conformity testing. Finally, the FNC approach was applied in drug testing problems in Chapter 6. Bayesian inference was used to establish subject-specific thresholds for detection of abnormal values in doping.

7.2 Future work

The success of the FNC depends on the correctness of the error distribution. The Gaussian error model was assumed for much of the work presented in this thesis but the work can be extended to other non-normal distributions. A violation of assumption that can occur in practice is the lack of independence in measurement errors. For example, the instruments may exhibit *hysteresis*: consecutive errors may not be independent. In such cases, the FNC statistic will show auto-correlation while the infrequently sampled process data are truly independent. In a future study, the performance of the FNC statistic will be studied for such situations.

It was found that FNC control using acceptance sampling plans and control charts perform well in detecting mean level shifts and controlling the consumer's risks. This thesis did not attempt to derive a closed form approximation to the distribution of the FNC statistic, applicability of Central Limit Theorem (CLT), unbiasedness and other estimation properties. The main reason for not adopting the bias correction is that the same amount of bias can be added to the set limit for FNC. In other words, FNC $<$ Limit is the same as FNC+bias $<$ Limit+bias.

Both EWMA and CUSUM individual FNC charts and Shewhart average FNC chart accumulate information from past observations. However, it has been established that these charts are more sensitive than Shewhart $\hat{p}_{iu}$ chart for detecting both small and large shifts (Chapters 2 and 3). It is planned to compare the performance of EWMA and CUSUM $\hat{p}_{iu}$ charts with the Shewhart $\hat{p}_{au}$ chart in future research.

For the short-run monitoring, the lack of Phase I data forced the use of an AQL value (a figure of 3% was used in the examples but the *Shiny* apps allow this value to be changed). While this level of AQL is acceptable for the milk product industry, it may not be universally valid for all types of products. For example, the AQL can be very low for electronic products. In such cases, the number of simulations needed to get the optimum plans or chart constants can be very large and time consuming. In such cases, alternative methods of designing for short-run production may be useful. Since the FNC statistic measures nonconformance, the AQL is a natural choice for short-run control. Other options such as using the Average Outgoing Quality Limit (AOQL) can be studied in the future.

For correlated quality characteristics, multivariate quality control methods are appropriate. However, the correlations among the observed variables were found to be weak for compositional variables under measurement uncertainty. Therefore, this thesis focussed only on univariate process monitoring. Our FNC approach can be extended to Multivariate Statistical Process Control (MSPC) and Principal Component Analysis (PCA), to deal with correlated variables. The challenge is to set up appropriate control limits with only small samples drawn from a short-run production process, which needs to be investigated in future research.

The current FNC based conformity testing procedures deal with three outcomes: conformity, inconclusive result, and non-conformity. These three outcomes were roughly matched to *negative*, *suspicious* and *positive* decisions respectively when the conformity testing protocol is applied to drug testing. An inconclusive result in drug testing is often obtained due to limited number of tests done for drug screening. It is planned to develop a new conformity testing procedure with four outcomes *negative*, *inconclusive*, *suspicious* and *positive* for drug screening problems.

The FNC statistic for two-sided specifications can be implemented as the sum of two one-sided FNC statistics, but simulations to examine this in depth have not been done. Similar to double-limit variables plans, FNC double limit plans can be developed for cases where the specification spread is too short compared to the process spread so that both limits can be simultaneously breached. For milk product applications, the possibility of breaching both specifications is rare and hence this case was not examined in depth.

There are many other future research potentials such as extending the FNC approach to double, multiple, sequential sampling etc. This thesis was largely driven by the practical needs of milk product manufacturing, and hence a large number of other possible extensions have not been considered and are left for future research.

References

[1] Abbas, N., Riaz, M., and Does, R. (2011). Enhancing the performance of EWMA charts. *Quality and Reliability Engineering International*, 27(6):821–833.

[2] Abbasi, S. A. (2010). On the performance of EWMA chart in the presence of two-component measurement error. *Quality Engineering*, 22(3):199–213.

[3] Abraham, B. (1977). Control charts and measurement error. In *Annual Technical Conference of the American Society for Quality Control*, volume 31, pages 370–374.

[4] Ahrens, B. D. and Butch, A. W. (2019). Performance enhancing drugs in sports. In *Critical Issues in Alcohol and Drugs of Abuse Testing*, pages 495–510. Elsevier.

[5] Alwan, L. C. (1995). The problem of misplaced control limits. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44(3):269–278.

[6] Alwan, L. C. and Roberts, H. V. (1988). Time-series modeling for statistical process control. *Journal of Business & Economic Statistics*, 6(1):87–95.

[7] Aven, T. (2010). On the need for restricting the probabilistic analysis in risk assessments to variability. *Risk Analysis*, 30:354–360.

[8] Balamurali, S. and Jun, C.-H. (2006). Repetitive group sampling procedure for variables inspection. *Journal of Applied Statistics*, 33(3):327–338.

[9] Basnet, C. and Case, K. E. (1992). The effect of measurement error on accept/reject probabilities for homogeneous products. *Quality Engineering*, 4(3):383–397.

[10] Bennett, C. A. (1954). Effect of measurement error on chemical process control. *Industrial Quality Control*, 10(4):17–20.

[11] Bennett, G. K., Case, K. E., and Schmidt, J. (1974). The economic effects of inspector error on attribute sampling plans. *Naval Research Logistics Quarterly*, 21(3):431–443.

[12] Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

[13] CAC (1993a). Guidelines for the Application of the Hazard Analysis Critical Control Point (HACCP) System. Codex Alimentarius Commission.

[14] CAC (1993b). Procedural Manual. Codex Alimentarius Commission.

[15] CAC (1997). Principles for the Establishment and Application of Microbiological Criteria for Foods. Codex Alimentarius Commission.

[16] CAC (2004). General Guidelines on Sampling. Codex Alimentarius Commission.

[17] Capizzi, G. and Masarotto, G. (2003). An adaptive exponentially weighted moving average control chart. *Technometrics*, 45(3):199–207.

[18] Carobbi, C. and Pennecchi, F. (2016). Bayesian conformity assessment in presence of systematic measurement errors. *Metrologia*, 53(2):S74.

[19] Castillo, E. D. and Montgomery, D. C. (1994). Short-run statistical process control: Q-chart enhancements and alternative methods. *Quality and Reliability Engineering International*, 10(2):87–97.

[20] Celano, G., Castagliola, P., Fichera, S., and Nenes, G. (2013). Performance of t control charts in short runs with unknown shift sizes. *Computers & Industrial Engineering*, 64(1):56–68.

[21] Celano, G., Castagliola, P., Trovato, E., and Fichera, S. (2011). Shewhart and EWMA t control charts for short production runs. *Quality and Reliability Engineering International*, 27(3):313–326.

[22] Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2017). *Shiny: Web Application Framework for R*. R package version 1.0.3.

[23] Cheng, X.-B. and Wang, F.-K. (2018a). The performance of EWMA median and CUSUM median control charts for a normal process with measurement errors. *Quality and Reliability Engineering International*, 34(2):203–213.

[24] Cheng, X.-B. and Wang, F.-K. (2018b). VSSI median control chart with estimated parameters and measurement errors. *Quality and Reliability Engineering International*, 34(5):867–881.

[25] Chou, Y. and Chen, K. (2005). Determination of optimal measurement guardbands. *Quality Technology & Quantitative Management*, 2(1):65–75.

[26] Chun, Y. H. and Rinks, D. B. (1998). Three types of producer's and consumer's risks in the single sampling plan. *Journal of Quality Technology*, 30(3):254–268.

[27] Cook, C. J., Crewther, B. T., and Smith, A. A. (2012). Comparison of baseline free testosterone and cortisol concentrations between elite and non-elite female athletes. *American Journal of Human Biology*, 24(6):856–858.

[28] Costa, A. F. and Castagliola, P. (2011). Effect of measurement error and autocorrelation on the $\bar{X}$ chart. *Journal of Applied Statistics*, 38(4):661–673.

[29] Crowder, S. V. (1987). A simple method for studying run–length distributions of exponentially weighted moving average charts. *Technometrics*, 29(4):401–407.

[30] Delanghe, J., Maenhout, T., Speeckaert, M., and De Buyzere, M. (2014). Detecting doping use: more than an analytical problem. *Acta Clinica Belgica*, 69(1):25–29.

[31] Desimoni, E. and Brunetti, B. (2011). Uncertainty of measurement and conformity assessment: a review. *Analytical and Bioanalytical Chemistry*, 400(6):1729–1741.

[32] Dieck, R. H. (1997). Measurement uncertainty models. *ISA Transactions*, 36(1):29–35.

[33] Duncan, A. J. (1986). *Quality Control and Industrial Statistics, 5th edn.* Richard D Irwin Inc., Homewood, IL.

[34] Eagle, A. (1954). A method for handling errors in testing and measuring. *Industrial Quality Control*, 10(3):10–15.

[35] Easterling, R., Johnson, M., Bement, T., and Nachtsheim, C. (1991). Statistical tolerancing based on consumer's risk considerations. *Journal of Quality Technology*, 23(1):1–11.

[36] Elam, M. E. (2008). Control charts for short production runs. In *Encyclopedia of Statistics in Quality and Reliability*. John Wiley.

[37] Ellis, L. and Nyborg, H. (1992). Racial/ethnic variations in male testosterone levels: a probable contributor to group differences in health. *Steroids*, 57(2):72–75.

[38] Ellison, S., Wegscheider, W., and Williams, A. (1997). Measurement uncertainty. *Analytical Chemistry*, 69(19):607A–613A.

[39] Ewan, W. D. (1963). When and how to use CUSUM charts. *Technometrics*, 5(1):1–22.

[40] Fang, X. and Zhang, Y. (1995). Adjusting plans of acceptance sampling based on sample mean for undesired measurement conditions. *Quality Engineering*, 8(1):57–73.

[41] Farnum, N. R. (1992). Control charts for short runs: Nonconstant process and measurement error. *Quality Control and Applied Statistics*, 37(12):639–640.

[42] Forbes, A. B. (2006). Measurement uncertainty and optimized conformance assessment. *Measurement*, 39(9):808–814.

[43] Freund, R. (1957). Acceptance control charts. *Industrial Quality Control*, 14(4):13–23.

[44] Freund, R. (1960). A reconsideration of the variables control chart with special reference to the chemical industries. *Industrial Quality Control*, 16(11):35–41.

[45] Gan, F. F. (1991). An optimal design of CUSUM quality control charts. *Journal of Quality Technology*, 23(4):279–286.

[46] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.

[47] Gilbert, J. (1999). Sampling of raw materials and processed foods for the presence of GMOs. *Food Control*, 10(6):363–365.

[48] Goldsmith, P. and Whitfield, H. (1961). Average run lengths in cumulative chart quality control schemes. *Technometrics*, 3(1):11–20.

[49] Gonzales-Barron, U. and Butler, F. (2011). A comparison between the discrete Poisson-gamma and Poisson-lognormal distributions to characterise microbial counts in foods. *Food Control*, 22(8):1279–1286.

[50] Govindaraju, K. (2005). Statistical performance of control charts. *Economic Quality Control*, 20(1):5–20.

[51] Govindaraju, K. and Jones, G. (2015). Fractional acceptance numbers for lot quality assurance. In Knoth, S. and Schmid, W., editors, *Frontiers in Statistical Quality Control*, volume 11, pages 271–286. Springer.

[52] Govindaraju, K. and Kissling, R. (2016). Sampling plans for beta distributed compositional fractions. *Chemometrics and Intelligent Laboratory Systems*, 151:103–107.

[53] Graves, S. B., Murphy, D. C., and Ringuest, J. L. (1996). Reevaluating producer's and consumer's risks in acceptance sampling. *Computers & Industrial Engineering*, 30(2):171–184.

[54] Grigg, N. P. (1998). Statistical process control in UK food production: an overview. *International Journal of Quality & Reliability Management*, 15(2):223–238.

[55] Grubbs, F. and Coon, H. (1954). On setting test limits relative to specification limits. *Industrial Quality Control*, 10(5):15–20.

[56] Hald, A. (1968). Bayesian single sampling attribute plans for continuous prior distributions. *Technometrics*, 10(4):667–683.

[57] Haq, A., Brown, J., Moltchanova, E., and Al-Omari, A. I. (2015). Effect of measurement error on exponentially weighted moving average control charts under ranked set sampling schemes. *Journal of Statistical Computation and Simulation*, 85(6):1224–1246.

[58] Harris, T. J. and Ross, W. H. (1991). Statistical process control procedures for correlated observations. *The Canadian Journal of Chemical Engineering*, 69(1):48–57.

[59] Hawkins, D. M. (1987). Self-starting CUSUM charts for location and scale. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 36(4):299–316.

[60] Hawkins, D. M. and Olwell, D. H. (2012). *Cumulative sum Charts and Charting for Quality Improvement*. Springer Science & Business Media.

[61] Hawkins, D. M. and Wu, Q. (2014). The CUSUM and the EWMA head-to-head. *Quality Engineering*, 26(2):215–222.

[62] Hazelton, M. L. and Turlach, B. A. (2009). Nonparametric density deconvolution by weighted kernel estimators. *Statistics and Computing*, 19(3):217–228.

[63] Hazelton, M. L. and Turlach, B. A. (2010). Semiparametric density deconvolution. *Scandinavian Journal of Statistics*, 37(1):91–108.

[64] He, F., Jiang, W., and Shu, L. (2008). Improved self-starting control charts for short runs. *Quality Technology & Quantitative Management*, 5(3):289–308.

[65] Healy, S., Wallace, M., and Murphy, E. (2009). Mathematical modelling of test limits and guardbands. *Quality and Reliability Engineering International*, 25(6):717–730.

[66] Hibbert, D. B. (2001). Compliance of analytical results with regulatory or specification limits: a probabilistic approach. *Accreditation and Quality Assurance*, 6(8):346–351.

[67] Hinrichs, W. (2010). The impact of measurement uncertainty on the producer's and user's risks, on classification and conformity assessment: An example based on tests on some construction products. *Accreditation and Quality Assurance*, 15(5):289–296.

[68] Hoelzer, K. and Pouillot, R. (2013). Practical considerations for the interpretation of microbial testing results based on small numbers of samples. *Foodborne Pathogens and Disease*, 10(11):907–915.

[69] Holst, E., Thyregod, P., and Wilrich, P.-T. (2001). On conformity testing and the use of two stage procedures. *International Statistical Review*, 69(3):419–432.

[70] Hryniewicz, O. (2008). Statistics with fuzzy data in statistical quality control. *Soft Computing*, 12(3):229–234.

[71] Hu, X., Castagliola, P., Sun, J., and Khoo, M. (2015). The effect of measurement errors on the synthetic $\bar{X}$ chart. *Quality and Reliability Engineering International*, 31(8):1769–1778.

[72] Hu, X., Castagliola, P., Sun, J., and Khoo, M. (2016). Economic design of the upper-sided synthetic S^2 chart with measurement errors. *International Journal of Production Research*, 54(19):5651–5670.

[73] Hubbard, M. R. (2012). *Statistical Quality Control for the Food Industry*. Springer Science & Business Media.

[74] ICMSF (1986). Microorganisms in Foods 2. Sampling for Microbiological Analysis: Principles and Specific Applications. Blackwell Scientific Publications.

[75] ICMSF (2002). Microorganisms in Foods 7. Microbiological Testing in Food Safety Management. International Commission on Microbiological Specifications for Foods. Kluwer Academic/Plenum Publishers, New York.

[76] ICMSF (2011). Microorganisms in Foods 8. Use of Data for Assessing Process Control and Product Acceptance. International Commission on Microbiological Specifications for Foods, volume 8. Springer.

[77] IFST (1991). Food and Drink Good Manufacturing Practice: A Guide to its Responsible Management. Institute of Food Science and Technology (UK).

[78] ISO (2005). ISO and conformity assessment. International Standards Organization, Geneva, Switzerland.

[79] ISO 10576-1 International Standard (2003). ISO 10576-1:2003 Statistical Methods – Guidelines for the Evaluation of Conformity with Specified Requirements, Part 1: General Principles. International Standards Organization, Geneva, Switzerland.

[80] ISO 5725-1 International Standard (1994). Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 1: General Principles and Definitions. International Standards Organization, Geneva, Switzerland.

[81] ISO/IEC Guide 98-3:2008 (2008). Uncertainty of Measurement–Part 3: Guide to the Expression of Uncertainty in Measurement. International Standards Organization/International Electrotechnical Commission, Geneva, Switzerland.

[82] Jakobsson, J., Ekström, L., Inotsume, N., Garle, M., Lorentzon, M., Ohlsson, C., Roh, H.-K., Carlström, K., and Rane, A. (2006). Large differences in testosterone excretion in Korean and Swedish men are strongly associated with a UDP-glucuronosyl transferase 2B17 polymorphism. *The Journal of Clinical Endocrinology & Metabolism*, 91(2):687–693.

[83] Jamkhaneh, E. B., Sadeghpour-Gildeh, B., and Yari, G. (2011). Inspection error and its effects on single sampling plans with fuzzy parameters. *Structural and Multidisciplinary Optimization*, 43(4):555–560.

[84] Jarvis, B. (2007). On the compositing of samples for qualitative microbiological testing. *Letters in Applied Microbiology*, 45(6):592–598.

[85] Johnson, N. L., Kotz, S., and Wu, X.-Z. (1991). *Inspection Errors for Attributes in Quality Control*, volume 44. CRC Press.

[86] Jones, L. A., Champ, C. W., and Rigdon, S. E. (2001). The performance of exponentially weighted moving average charts with estimated parameters. *Technometrics*, 43(2):156–167.

[87] Jones, L. A., Champ, C. W., and Rigdon, S. E. (2004). The run length distribution of the CUSUM with estimated parameters. *Journal of Quality Technology*, 36(1):95–108.

[88] Källgren, H., Lauwaars, M., Magnusson, B., Pendrill, L., and Taylor, P. (2003). Role of measurement uncertainty in conformity assessment in legal metrology and trade. *Accreditation and Quality Assurance*, 8(12):541–547.

[89] Kanazuka, T. (1986). The effect of measurement error on the power of $\bar{X}$-R charts. *Journal of Quality Technology*, 18(2):91–95.

[90] Kemp, K. W. (1961). The average run length of the cumulative sum chart when a V-mask is used. *Journal of the Royal Statistical Society: Series B (Methodological)*, 23(1):149–153.

[91] Kessel, W. (2002). Measurement uncertainty according to ISO/BIPM-GUM. *Thermochimica Acta*, 382(1-2):1–16.

[92] Kim, Y., Cho, B., and Kim, N. (2007). Economic design of inspection procedures using guard band when measurement errors are present. *Applied Mathematical Modelling*, 31(5):805–816.

[93] King, B. (1999). Assessment of compliance of analytical results with regulatory or specification limits. *Accreditation and Quality Assurance*, 4(1-2):27–30.

[94] Knoth, S. and Schmid, W. (2002). Monitoring the mean and the variance of a stationary process. *Statistica Neerlandica*, 56(1):77–100.

[95] Łagowska, K. and Kapczuk, K. (2016). Testosterone concentrations in female athletes and ballet dancers with menstrual disorders. *European Journal of Sport Science*, 16(4):490–497.

[96] Lavin, M. (1946). Inspection efficiency and sampling inspection plans. *Journal of the American Statistical Association*, 41(236):432–438.

[97] Li, Y., Liu, Y., Zou, C., and Jiang, W. (2014). A self-starting control chart for high-dimensional short-run processes. *International Journal of Production Research*, 52(2):445–461.

[98] Lieberman, G. J. and Resnikoff, G. J. (1955). Sampling plans for inspection by variables. *Journal of the American Statistical Association*, 50(270):457–516.

[99] Linna, K. W. and Woodall, W. H. (2001). Effect of measurement error on Shewhart control charts. *Journal of Quality Technology*, 33(2):213–222.

[100] Liu, F. and Cui, L. (2016). A design of attributes double sampling plans for three-class products. *Communications in Statistics-Simulation and Computation*, 45(3):1054–1071.

[101] Liu, L., Zhang, J., and Zi, X. (2015). Dual nonparametric CUSUM control chart based on ranks. *Communications in Statistics-Simulation and Computation*, 44(3):756–772.

[102] Lucas, J. M. and Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, 32(1):1–12.

[103] Maleki, M. R., Amiri, A., and Castagliola, P. (2017). Measurement errors in statistical process monitoring: A literature review. *Computers & Industrial Engineering*, 103:316–329.

[104] Maravelakis, P. E. (2012). Measurement error effect on the CUSUM control chart. *Journal of Applied Statistics*, 39(2):323–336.

[105] Maravelakis, P. E., Panaretos, J., and Psarakis, S. (2004). EWMA chart and measurement error. *Journal of Applied Statistics*, 31(4):445–455.

[106] McHugh, A. J., Feehily, C., Hill, C., and Cotter, P. D. (2017). Detection and enumeration of spore-forming bacteria in powdered dairy products. *Frontiers in Microbiology*, 8:109.

[107] McHugh, C. M., Park, R. T., Sönksen, P. H., and Holt, R. I. (2005). Challenges in detecting the abuse of growth hormone in sport. *Clinical Chemistry*, 51(9):1587–1593.

[108] Mei, W.-H., Case, K. E., and Schmidt, J. (1975). Bias and imprecision in variables acceptance sampling effects and compensation. *International Journal of Production Research*, 13(4):327–340.

[109] Mittag, H.-J. and Stemann, D. (1998). Gauge imprecision effect on the performance of the $\bar{X}$-S control chart. *Journal of Applied Statistics*, 25(3):307–317.

[110] Mizuno, S. (1961). Problems of measurement errors in process control. *Bulletin of the International Statistical Institute*, 38:405–415.

[111] Montgomery, D. (2013). *Introduction to Statistical Quality Control (7th edn)*. John Wiley & Sons.

[112] Nenes, G. and Tagaras, G. (2005). The CUSUM chart for monitoring short production runs. In *Proceedings of 5th International Conference on Analysis of Manufacturing Systems–Production Management. Zakynthos Island, Greece*, pages 43–50.

[113] Nenes, G. and Tagaras, G. (2010). Evaluation of CUSUM charts for finite-horizon processes. *Communications in Statistics - Simulation and Computation*, 39(3):578–597.

[114] Noorossana, R. and Zerehsaz, Y. (2015). Effect of measurement error on phase II monitoring of simple linear profiles. *The International Journal of Advanced Manufacturing Technology*, 79(9-12):2031–2040.

[115] Owen, D. and Chou, Y.-M. (1983). Effect of measurement error and instrument bias on operating characteristics for variables sampling plans. *Journal of Quality Technology*, 15(3):107–117.

[116] Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115.

[117] Pendrill, L. R. (2007). Optimised measurement uncertainty and decision-making in conformity assessment. *NCSLi Measure*, 2(2):76–86.

[118] Pendrill, L. R. (2009). An optimised uncertainty approach to guard-banding in global conformity assessment. In *Advanced Mathematical and Computational Tools in Metrology and Testing: AMCTM VIII*, pages 256–261. World Scientific.

[119] Pendrill, L. R. (2014). Using measurement uncertainty in decision-making and conformity assessment. *Metrologia*, 51(4):S206.

[120] Plummer, M., Stukalov, A., and Denwood, M. (2017). rjags: Bayesian graphical models using MCMC, 2016. *URL http://CRAN. R-project. org/package= rjags. R package version*, 2:0–4.

[121] Powell, M. R. (2014). Optimal food safety sampling under a budget constraint. *Risk Analysis*, 34(1):93–100.

[122] Quesenberry, C. P. (1991). SPC Q charts for start-up processes and short or long runs. *Journal of Quality Technology*, 23(3):213–224.

[123] Quesenberry, C. P. (1993). The effect of sample size on estimated limits for $\bar{X}$ and $\bar{X}$ control charts. *Journal of Quality Technology*, 25(4):237–247.

[124] Quesenberry, C. P. (1995). On properties of Q charts for variables. *Journal of Quality Technology*, 27(3):184–203.

[125] Reynolds, M. R. (1975). Approximations to the average run length in cumulative sum control charts. *Technometrics*, 17(1):65–71.

[126] Roberts, S. (1959). Control chart tests based on geometric moving averages. *Technometrics*, 1(3):239–250.

[127] Robinson, N., Saugy, M., Vernec, A., and Sottas, P.-E. (2011). The athlete biological passport: An effective tool in the fight against doping. *Clinical Chemistry*, 57(6):830–832.

[128] Robinson, N., Sottas, P.-E., Mangin, P., and Saugy, M. (2007). Bayesian detection of abnormal hematological values to introduce a no-start rule for heterogeneous populations of athletes. *Haematologica*, 92(8):1143–1144.

[129] Santos-Fernández, E., Govindaraju, K., and Jones, G. (2016). Compressed limit sampling inspection plans for food safety. *Applied Stochastic Models in Business and Industry*, 32(4):469–484.

[130] Santos-Fernandez, E., Wu, P., and Mengersen, K. L. (2019). Bayesian statistics meets sports: a comprehensive review. *Journal of Quantitative Analysis in Sports*.

[131] Saugy, M., Lundby, C., and Robinson, N. (2014). Monitoring of biological markers indicative of doping: the athlete biological passport. *British Journal of Sports Medicine*, 48(10):827–832.

[132] Schilling, E. G. (1985). The role of acceptance sampling in modern quality control. *Communications in Statistics-Theory and Methods*, 14(11):2769–2783.

[133] Schilling, E. G. (1990). Acceptance control in a modern quality program. *Quality Engineering*, 3(2):181–191.

[134] Schilling, E. G. and Neubauer, D. V. (2009). *Acceptance Sampling in Quality Control*. Chapman and Hall/CRC.

[135] Schulze, J. J., Lundmark, J., Garle, M., Ekström, L., Sottas, P.-E., and Rane, A. (2009). Substantial advantage of a combined Bayesian and genotyping approach in testosterone doping tests. *Steroids*, 74(3):365–368.

[136] Separovic, L. and Lourenço, F. R. (2019). Measurement uncertainty and risk of false conformity decision in the performance evaluation of liquid chromatography analytical procedures. *Journal of Pharmaceutical and Biomedical Analysis*, 171:73–80.

[137] Sharpe, K., Ashenden, M. J., and Schumacher, Y. O. (2006). A third generation approach to detect erythropoietin abuse in athletes. *Haematologica*, 91(3):356–363.

[138] Shewhart, W. A. (1924). Some applications of statistical methods to the analysis of physical and engineering data. *Bell System Technical Journal*, 3(1):43–87.

[139] Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*. ASQ Quality Press.

[140] Sottas, P.-E., Baume, N., Saudan, C., Schweizer, C., Kamber, M., and Saugy, M. (2006). Bayesian detection of abnormal values in longitudinal biomarkers with an application to T/E ratio. *Biostatistics*, 8(2):285–296.

[141] Sottas, P.-E., Robinson, N., Rabin, O., and Saugy, M. (2011). The athlete biological passport. *Clinical Chemistry*, 57(7):969–976.

[142] Sottas, P.-E., Robinson, N., and Saugy, M. (2010). The athlete's biological passport and indirect markers of blood doping. In *Doping in Sports: Biochemical Principles, Effects and Analysis*, pages 305–326. Springer.

[143] Sottas, P.-E., Saudan, C., Schweizer, C., Baume, N., Mangin, P., and Saugy, M. (2008). From population- to subject-based limits of T/E ratio to detect testosterone abuse in elite sports. *Forensic Science International*, 174(2-3):166–172.

[144] Starbird, S. A. (2005). Moral hazard, inspection policy, and food safety. *American Journal of Agricultural Economics*, 87(1):15–27.

[145] Strahm, E., Sottas, P.-E., Schweizer, C., Saugy, M., Dvorak, J., and Saudan, C. (2009). Steroid profiles of professional soccer players: an international comparative study. *British Journal of Sports Medicine*, 43(14):1126–1130.

[146] Theodorou, D. and Zannikos, F. (2014). The use of measurement uncertainty and precision data in conformity assessment of automotive fuel products. *Measurement*, 50:141–151.

[147] Tran, K., Castagliola, P., and Celano, G. (2016). The performance of the Shewhart-RZ control chart in the presence of measurement error. *International Journal of Production Research*, 54(24):7504–7522.

[148] Vasilopoulos, A. and Stamboulis, A. (1978). Modification of control chart limits in the presence of data correlation. *Journal of Quality Technology*, 10(1):20–30.

[149] Vernec, A. R. (2014). The athlete biological passport: an integral element of innovative strategies in antidoping. *British Journal of Sports Medicine*, 48(10):817–819.

[150] Vining, G. (2009). Technical advice: Phase I and phase II control charts. *Quality Engineering*, 21(4):478–479.

[151] WADA (2019). *Athlete Biological Passport Operating Guidelines, version 7.1*. World Anti-Doping Agency.

[152] Wardell, D., Moskowitz, H., and Plante, R. (1994). Run-length distributions of special-cause control charts for correlated processes. *Technometrics*, 36(1):3–17.

[153] Weißensee, K., Kühn, O., Linß, G., and Sommer, K.-D. (2008). Risk of erroneously deciding conformity of measuring instruments. *Accreditation and Quality Assurance*, 13(11):663–669.

[154] Whitaker, T. (2006). Sampling foods for mycotoxins. *Food Additives and Contaminants*, 23(1):50–61.

[155] Whiting, R., Rainosek, A., Buchanan, R., Miliotis, M., LaBarre, D., Long, W., Ruple, A., and Schaub, S. (2006). Determining the microbiological criteria for lot rejection from the performance objective or food safety objective. *International Journal of Food Microbiology*, 110(3):263–267.

[156] Widlowski, J., Mio, C., Disney, M., Andredakis, I., Atzberger, C., Brennan, J., Busetto, L., Chelle, M., Ceccherini, G., Colombo, R., et al. (2015). The fourth phase of the radiative transfer model intercomparison (RAMI) exercise: Actual canopy scenarios and conformity testing. *Remote Sensing of Environment*, 169:418–437.

[157] Williams, R. H. and Hawkins, C. F. (1993a). The economics of guardband placement. In *Proceedings of IEEE International Test Conference*, pages 218–225. IEEE.

[158] Williams, R. H. and Hawkins, C. F. (1993b). The effect of guardbands on errors in production testing. In *Proceedings ETC 93 Third European Test Conference*, pages 2–7. IEEE.

[159] Wilrich, P.-T. (2000). Single sampling plans for inspection by variables in the presence of measurement error. *AStA Advances in Statistical Analysis*, 3(84):239–250.

[160] Woodall, W. H. (1985). The statistical design of quality control charts. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 34(2):155–160.

[161] Woodall, W. H. (2000). Controversies and contradictions in statistical process control. *Journal of Quality Technology*, 32(4):341–350.

[162] Woodall, W. H. and Faltin, F. W. (1993). Autocorrelated data and SPC. *ASQC Statistics Division Newsletter*, 13(4):18–21.

[163] Woodall, W. H. and Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology*, 31(4):376–386.

[164] World Health Organization (1977). Methods Used in Establishing Permissible Levels in Occupational Exposure to Harmful Agents: Report of a WHO Expert Committee with the Participation of ILO. World Health Organization.

[165] Wortham, A. and Baker, R. (1976). Multiple deferred state sampling inspection. *International Journal of Production Research*, 14(6):719–731.

[166] Wu, H. and Govindaraju, K. (2014). Computer-aided variables sampling inspection plans for compositional proportions and measurement error adjustment. *Computers & Industrial Engineering*, 72:239–246.

[167] Wu, Z. (1998). An adaptive acceptance control chart for tool wear. *International Journal of Production Research*, 36(6):1571–1586.

[168] Xie, M., Lu, X., Goh, T., and Chan, L. (1999). A quality monitoring and decision-making scheme for automated production processes. *International Journal of Quality & Reliability Management*, 16(2):148–157.

[169] Yang, S.-F., Ho, H.-W., and Rahim, M. A. (2007). Effects of measurement error on controlling two dependent process steps. *Economic Quality Control*, 22(1):127–139.

[170] Yang, S.-F. and Yang, C.-M. (2005). Effects of imprecise measurement on the two dependent processes control for the autocorrelated observations. *The International Journal of Advanced Manufacturing Technology*, 26(5-6):623–630.

[171] Zaidi, F., Castagliola, P., Tran, K., and Khoo, M. (2019). Performance of the Hotelling T^2 control chart for compositional data in the presence of measurement errors. *Journal of Applied Statistics*, 46(14):2583–2602.

[172] Zantek, P. (2005). Run-length distributions of Q-chart schemes. *IIE transactions*, 37(11):1037–1045.

[173] Zhang, C. W., Xie, M., and Jin, T. (2012). An improved self-starting cumulative count of conforming chart for monitoring high-quality processes under group inspection. *International Journal of Production Research*, 50(23):7026–7043.

[174] Zhang, L., Chen, G., and Castagliola, P. (2009). On t and EWMA t charts for monitoring changes in the process mean. *Quality and Reliability Engineering International*, 25(8):933–945.

[175] Zhou, X., Govindaraju, K., and Jones, G. (2018). Monitoring fractional nonconformance for short-run production. *Quality Engineering*, 30(3):498–510.

[176] Zhou, X., Govindaraju, K., and Jones, G. (2019a). Acceptance control and guardbanding for error-prone individual measurements. *Quality and Reliability Engineering International*, 35(2):517–534.

[177] Zhou, X., Govindaraju, K., and Jones, G. (2019b). Fractional nonconformance based conformity testing. *Computers & Industrial Engineering*, 135:402–411.

[178] Zwietering, M. H. (2009). Quantitative risk assessment: is more complex always better?: Simple is not stupid and complex is not always more correct. *International Journal of Food Microbiology*, 134(1-2):57–62.

Appendix A

Contributions to publications

DRC 16

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Xin Zhou
Name/title of Primary Supervisor:	Dr. K. Govindaraju
Name of Research Output and full reference:	
Zhou, X., Govindaraju, K., & Jones, G. (2018). Monitoring fractional nonconformance for short-run production. Quality Engineering, 30, 498–510.	
In which Chapter is the Manuscript /Published work:	2
Please indicate:	
• The percentage of the manuscript/Published Work that was contributed by the candidate:	80
and	
• Describe the contribution that the candidate has made to the Manuscript/Published Work:	
The candidate did the necessary research, statistical modelling, code writing and preparation of the manuscript.	
For manuscripts intended for publication please indicate target journal:	
Candidate's Signature:	Zhou Xin
Date:	14/10/2019
Primary Supervisor's Signature:	Govindaraju
Date:	14/10/2019

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)

DRC 16

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Xin Zhou
Name/title of Primary Supervisor:	Dr. K. Govindaraju
Name of Research Output and full reference:	
FNC control charts for bulk material	
In which Chapter is the Manuscript /Published work:	3
Please indicate:	
• The percentage of the manuscript/Published Work that was contributed by the candidate:	80
and	
• Describe the contribution that the candidate has made to the Manuscript/Published Work:	
The candidate did the necessary research, statistical modelling, code writing and preparation of the manuscript.	
For manuscripts intended for publication please indicate target journal:	
Journal of Statistical Computation and Simulation	
Candidate's Signature:	Zhou
Date:	14/10/2019
Primary Supervisor's Signature:	Ghovindaraju
Date:	14/10/2019

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)

DRC 16

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Xin Zhou
Name/title of Primary Supervisor:	Dr. K. Govindaraju
Name of Research Output and full reference:	
Zhou, X., Govindaraju, K., & Jones, G. (2019). Acceptance control and guardbanding for error-prone individual measurements. Quality and Reliability Engineering International, 35, 517–534.	
In which Chapter is the Manuscript /Published work:	4
Please indicate:	
• The percentage of the manuscript/Published Work that was contributed by the candidate:	80
and	
• Describe the contribution that the candidate has made to the Manuscript/Published Work:	
The candidate did the necessary research, statistical modelling, code writing and preparation of the manuscript.	
For manuscripts intended for publication please indicate target journal:	
Candidate's Signature:	Zhou Xin
Date:	14/10/2019
Primary Supervisor's Signature:	K Govindaraju
Date:	14/10/2019

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)

DRC 16

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Xin Zhou
Name/title of Primary Supervisor:	Dr. K. Govindaraju
Name of Research Output and full reference:	
Zhou, X., Govindaraju, K., & Jones, G. (2019). Fractional nonconformance based conformity testing. Computers & Industrial Engineering, 135, 402– 411.	
In which Chapter is the Manuscript /Published work:	5
Please indicate:	
• The percentage of the manuscript/Published Work that was contributed by the candidate:	80
and	
• Describe the contribution that the candidate has made to the Manuscript/Published Work:	
The candidate did the necessary research, statistical modelling, code writing and preparation of the manuscript.	
For manuscripts intended for publication please indicate target journal:	
Candidate's Signature:	Zhou
Date:	14/10/2019
Primary Supervisor's Signature:	Kthoundaraju
Date:	14/10/2019

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)

DRC 16

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Xin Zhou
Name/title of Primary Supervisor:	Dr. K. Govindaraju
Name of Research Output and full reference:	
Modified FNC based conformity testing using subject-specific limiting value	
In which Chapter is the Manuscript /Published work:	6
Please indicate:	
• The percentage of the manuscript/Published Work that was contributed by the candidate:	80
and	
• Describe the contribution that the candidate has made to the Manuscript/Published Work:	
The candidate did the necessary research, statistical modelling, code writing and preparation of the manuscript.	
For manuscripts intended for publication please indicate target journal:	
Journal of Quantitative Analysis in Sports	
Candidate's Signature:	Zhou Xin
Date:	14/10/2019
Primary Supervisor's Signature:	Govindaraju
Date:	14/10/2019

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)