Statistical Monitoring Procedures for High-Purity Manufacturing Processes

statistical monitoring procedures for high-purity manufacturing processes

caterina rizzo

statistical monitoring procedures for high-purity manufacturing processes caterina rizzo

Statistical Monitoring Procedures for High-Purity Manufacturing Processes

Caterina Rizzo

Statistical monitoring procedures for high-purity manufacturing processes.
A catalog record is available from the Eindhoven University of Technology Library.
ISBN: 978-90-386-5735-6
Document typeset with LaTeX.

Layout: Caterina Rizzo
Cover and printed by: Ridderprint | ridderprint.com

The research presented in this dissertation was sponsored by the Dow Chemical Company, within the collaboration agreement between Dow and the Eindhoven University of Technology number 255683.

Statistical Monitoring Procedures for High-Purity Manufacturing Processes

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr. S.K. Lenaerts, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op vrijdag 9 juni 2023 om 16:00 uur

door

Caterina Rizzo

geboren te Polistena, Italië

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof. dr. F.C.R. Spieksma
promotor:	dr. A. Di Bucchianico
copromotor:	prof. dr E.R. van den Heuvel
leden:	prof. dr. ir. C.F.J. den Doelder
	prof. dr. R. Göb (Universität Würzburg)
	prof. dr. S. Knoth (Helmut-Schmidt-Universität)
	prof. dr. W.H. Woodall (Virginia Tech)
adviseur:	dr. S.-T. Chin (The Dow Chemical Company)

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Abstract

Manufacturing processes in the chemical industry are intrinsically complex, but technological advances in the current Industry 4.0 era enable plant personnel to quantify and model chemical processes' inherent layers of complexity.

An increasing number of manufacturing systems are operated with digital devices that continuously monitor the number of non-conforming items to predict process quality and efficiency. While statistical process control techniques are still in vogue to monitor process variability and identify departures from in-control or expected conditions, traditional methodologies often in use are no longer adequate to track in more advanced scenarios. The consequent limitations might lead to failure in identifying shifts from the target of product quality characteristics, resulting in a potential increase of off-grade material that has both a financial and an environmental impact. This thesis explores several practical statistical challenges in monitoring high-purity processes arising from the complexity of enhanced production operations.

The dissertation builds upon the notions of process time and control chart time, which represent a distinctive feature of the high-purity monitoring framework. Based on these concepts, conditional performance metrics are adapted to allow fair evaluation and comparison of the newly proposed designs.

These measures are used to compare application-tailored control charts built on composite change-point models against their traditional competitors. These flexible charts are based on generalized likelihood ratios (GLR). The models proposed are industry-driven and represent two practical scenarios in the manufacturing industry: the indifference interval and the epidemic shift models. The first describes the situation when the monitored quality characteristic shifts within a specific interval not deemed of practical importance, while the latter describes a temporary failure mode typical of feedback controller systems. In both cases, the tailored approach outperforms the competitors in terms

of faster detection.

Another aspect often neglected in the literature is the influence a set of covariates exerts on the monitored quality variable of interest. Generalized linear models (GLM)-based control charts can detect contextual anomalies in those systems where the expected fluctuations of several process variables significantly influence the quality characteristic. Predictive residuals show excellent performance in detecting sustained shifts from the in-control conditions during real-time monitoring. In contrast, recursive residuals can be used at the early stages of the monitoring process or for short-run schemes. Finally, the use of the Hermite process for monitoring a bivariate homogeneous Poisson process on a continuous scale by monitoring the time between events extends the early-stage research on multivariate point process control chart. This procedure is relevant when the production process can be characterized by multiple correlated temporal point models with simple correlation structures.

Contents

List of Figures

List of Tables

Glossary

ACF autocorrelation function

ALI average length of inspection

ANOS average number of observations to signal

ARL average run length

ARMA autoregressive-moving-average

ATS average time to signal

CCC cumulative count of conforming

CL centre line

COM Conway-Maxwell Poisson

CQC cumulative quality control

CRL conforming run length

CUSUM cumulative sum

EWMA exponentially weighted moving average

GLM generalized linear model

GLR generalized likelihood ratio

GR group runs

HPP homogeneous Poisson process

LCL lower control limit

MLE maximum likelihood estimator

NHPP non-homogeneous Poisson process

PCA principal component analysis

SPC statistical process control

SQC statistical quality control

TBE time-between-events

UCL upper control limit

ZINB zero-inflated negative binomial

ZIP zero-inflated Poisson

Chapter 1

Introduction

In the current high-technology manufacturing era, in which customers are demanding more quality and competitors are providing more quality, companies are striving to enhance their processes to meet high-purity standards and sustain competitive advantage. Dow is no exception and processes are extensively optimized to improve productivity and quality. To achieve customer expectations, a process should be stable or repeatable within acceptable margins around the target of the product quality characteristics. Statistical process control (SPC) has been instrumental in the manufacturing industry, as it provides a collection of tools for process monitoring and quality improvement. Its popularity is due to its simplicity while guaranteeing significant impact. Control charts represent the primary technique in use in SPC and the most sophisticated one. It all comes down to acknowledging that process-inherent variation is always present due to random fluctuations and inconsistencies in the surrounding environments that cannot be fully controlled; this variation cannot be eliminated but it can be minimized, and, most importantly, it is predictable. When the process is around the target and only contains expected variation, it is considered 'in-control'. Any deviation of the model parameters from their in-control values is considered a shift to the 'out-of-control' state, which generally requires corrective actions. Control charts allow to plot each newly acquired observation or each group of observations over time, and each point plotted determines a decision on whether to take action or not. It is important to recall that control charts without a comprehensive action plan are ineffective for control purposes and might serve only as a silent surveillance agent. Surveillance is the task of monitoring without intervention; in industry, this case should not be the primary scope.
Despite the popularity of control charts, traditional methodologies commonly in use are not adequate

to track in advanced scenarios. They might lead to failure in identifying shifts from the target of product quality characteristics. This shortcoming is mainly due to the fact that, while the literature on control charts continues to be quite fertile, implementations of such advanced schemes in the manufacturing industry are lagging.

In spite of technological advances in sensor technology, many processes in the chemical industry are either intrinsically discrete or for which it is not economically feasible to obtain continuous data. A specific feature in manufacturing environments resulting from continuous improvement efforts, attracting interest from both academia and industry, is dealing with high-purity processes. These processes, also commonly referred to as high-yield in other contexts, are characterized by the fact that observations are only available as counts and that events are rare. Consequently, existing monitoring procedures must be adjusted to ensure meaningful monitoring in these non-traditional settings. In recent times, many efforts have been devoted to bridging the gap between advanced theory techniques, developed mainly by statisticians, and real-life practice, which relies on simplistic and often inappropriate procedures. This effort resulted in a series of more appropriate methodologies for monitoring manufacturing high-purity processes. The use of such charts can be easily extended to other fields, such as healthcare and reliability, where the process's features and the monitoring tasks are similar. This multi-faceted scope justifies the popularity of these charts and the fast-paced growth of related research. However, as Woodall and Faltin (2019) suggests, greater interactions between practitioners and research would lead to better, more practical, and meaningful implementations. This doctoral dissertation explores several practical statistical challenges in monitoring high-purity processes arising from the complexity of technology-enhanced production processes. The introductory chapter lays the dissertation's foundations, divided into four sections that all together provide the context for the thesis. Section 1.1 provides the reader with a high-level overview of the current state of the art in statistical process monitoring methodologies for high-purity processes. The discussion helps to identify gaps and research opportunities and introduces the notations and terminology relevant to the statistical process control framework. Then, the need for an comprehensive framework, which represents the goal of this research, is justified by the industrial motivation in 1.2. The main contributions that define the structure of the thesis are highlighted in Section 1.3. Finally, Section 1.4 reports the list of publications.

1.1 Background of research

Statistical process control has been widely used in industry to monitor mainly continuous variables describing a process's quality. Whereas many industrial quality features are continuous in nature, an increasing number of applications in the chemical industry involve monitoring discrete variables in the form of counts of defective or non-conforming items, i.e., items that do not satisfy technical specifications. It is widely known that the number of events can often be modeled by the binomial or the Poisson distributions because of their discrete nature. Therefore, traditional control charts for attribute data are modeled based on these distributions, which in most cases represent a simplistic approach as they do not account for over- and under-dispersion. This section highlights the limitations of traditional control charts for attribute data applied to high-purity processes and discusses the most common approaches to overcome these limitations in detail.

1.1.1 Traditional control charts for attribute data

Traditional control charts to monitor attribute data were first proposed by Shewhart (Shewhart (1926)). The *p*- and *np*- charts are intended to monitor the fraction of non-conforming products, while the *c*- and the *u*- chart are designed to monitor the number of non-conforming items. The charts mentioned above are widely used for monitoring industrial processes where the defect occurrence in a sample follows the binomial or the Poisson distributions, respectively. A relatively small sample size is required to detect changes when the shift amplitude is moderate or large. A comprehensive review and bibliography of attribute control charts by application is given by Woodall (1997). A more recent outline focusing on monitoring multinomial and multi-attribute processes is given by Topalidou and Psarakis (2009).

Limitations of traditional control charts When considering high-purity processes which exhibit low failure rates, the application of traditional process monitoring techniques is unsatisfactory mainly due to the normal approximation. Although they have been widely employed in industry to control the fraction and the number of non-conforming items, several pitfalls and practical problems are associated with p$-$, np$-$ and c$-$, u$-$ control charts (Xie et al. (2002b)).

- The 3-σ control limits are derived from the approximation of the binomial and the Poisson dis-

tributions by the normal distribution. The approximation holds for a large number of events, but it is inaccurate for low rates ($\lambda < 9$), and the approximation fails when the occurrence of defects rate of the process is very low. The 3-σ limits might also be justified by Chebychev's theorem, which guarantees a significant percentage coverage regardless of the underlying distribution. However, as indicated by Khakifirooz et al. (2021), this approach does not secure sufficient and practically relevant monitoring performance.

- Control limits can be unrealistic and meaningless. The lower control limit can be negative. In such circumstances, the limit is generally set to 0. This leads to unrealistic control limits, which make impractical the detection of process improvement, meaning a downward shift in the defect rate. Moreover, the upper control limit (UCL) can be smaller than 1 for np$-$ and c$-$ charts and that of the p$-$chart may be less than $\frac{1}{n}$. Therefore, the chart will raise a signal whenever there is a non-conforming item in the sample.

- The sample size heavily influences the frequency of raising an out-of-control signal. The frequency of signaling out-of-control depends on the choice of sample size. For a specific sample size value, there may be very frequent signals for out-of-control, while for a slightly larger sample size, there may be very infrequent signals or no signal at all. Chan et al. (2003) illustrate this problem with a practical example.

One can overcome these limitations by employing two main strategies. A more traditional approach is based on the transformation of the data into an approximately normal distribution. For example, Nelson (1994) proposed the transformation $X^{0.2777}$ coupled with a simple X-chart in Phase II. While being easy to implement, this method suffers from being difficult to interpret for practitioners, the transformation being not always accurate, and it poses a difficult challenge in fixing the desired in-control performance. Additionally, Santiago and Smith (2013) studied the effect of the normal transformation of exponentially distributed variables proposed by Nelson, demonstrating that the resulting chart holds a low power in detecting downward shifts, particularly when the decrease is small in magnitude. Therefore, rather than focusing on the counts themselves, it has been shown to be more promising to focus on cumulative quantities between non-conforming items.

Ali et al. (2016) and Woodall and Driscoll (2015) provide a thorough review of control charts for high-quality processes. Different names are used in the literature depending on the data type and distri-

bution assumed. In this thesis, time-between-events (TBE) and cumulative quantity control charts are used interchangeably. Monitoring Bernoulli/geometric and Poisson/exponential distributed variables represents the standard and the most effective approach to monitoring the occurrence of rare events.

1.1.2 Cumulative quantity control charts

Calvin (1983) was the first to propose a high-quality control chart based on the geometric process, while Goh (1987) studied its properties. This approach originates from the idea of plotting the cumulative number of conforming items between two successive non-conforming ones. This chart is known as the cumulative count of conforming (CCC) chart. Consider a sequence of independent Bernoulli random variables, $X, X_1, X_2, \ldots$ representing the status of the inspected items, classified as conforming or non-conforming. Let Y be the count of conforming items until a non-conforming one is observed. Then, the random variable Y follows the geometric distribution with probability mass function and cumulative distribution function f_g and $\mathcal{F}_G$, respectively

$$f_G(y) = p(1-p)^{y-1}, \tag{1.1}$$

$$\mathcal{F}_G(y) = 1-(1-p)^y \tag{1.2}$$

where p represents the probability of registering a non-conforming item. The cumulative number of items is plotted against the index of the sampling interval until a non-conforming item. The probability limits are typically set based on the in-control value of the non-conforming rate (p_0), posing $\mathcal{F}_G(LCL) = \alpha/2$ and $\mathcal{F}_G(UCL) = 1-\alpha/2$, so that

$$LCL_G = \ln(1-\alpha/2)/\ln(1-p_0), \tag{1.3}$$

$$UCL_G = \ln(\alpha/2)/\ln(1-p_0) \tag{1.4}$$

where α is the false alarm rate, specified a priori and LCL and UCL indicate the lower control and upper control limits, respectively. In the CCC chart, small values of Y indicate large values of p. Observations below the LCL therefore might indicate a deterioration of the process conditions ($p_1 > p_0$), while observations above the UCL might give an indication of process improvement ($p_1 < p_0$). Using prob-

ability limits instead of the traditional 3-σ approach is more appropriate due to the skewed nature of the distribution. Moreover, since the geometric distribution is discrete, the control limits can take only integer values and therefore it is difficult or almost impossible to obtain exactly the specified false alarm rate. The CCC chart can be easily extended to higher orders (CCC_r) by charting the number of conforming counts till the r^{th} successive non-conforming one. This extension was proposed by Xie et al. (2002a). Let Z be the count of conforming items until the r^{th} successive one. Then, Z is a negative binomial random variable with probability mass function $f_{NG}(z)$ as follows

$$f_{NG}(z) = \binom{z-1}{r-1} p^r (1-p)^{z-1}. \tag{1.5}$$

The parameterization used (1.5) expects z to be the counting number of trials, given r successes. Using high-order control charts reduces the variance at the cost of an increased waiting time to get a decision point. Other variations might be used depending on the definition of the counting variable. The probabilistic control limits can be computed numerically by solving the following equations

$$\sum_{k=r}^{LCL_{NB}} P(X_{r,p_0} = k) = \alpha \tag{1.6}$$

and

$$\sum_{k=r}^{UCL_{NB}} P(X_{r,p_0} = k) = 1 - \alpha. \tag{1.7}$$

Intuitively, increasing the order of the chart might cause a delay since r non-conforming items need to be recorded before plotting a new observation and therefore prompting a decision to the user. However, increasing the order might be appropriate for calibrating the sensitivity of the chart to the desired performance. General guidelines on choosing the value of r do not exist, which might represent a limitation for implementers and practitioners. As pointed out by Szarka and Woodall (2011), despite its simplicity, the CCC chart presents some technical issues when considering the two-sided case. These issues are the non-maximality and bias properties of the performance metrics, and its primary weakness resides in its inability to detect small to moderate shifts. Following the idea of Calvin (1983), the cumulative quality control (CQC) charts were first proposed by Chang and Gan (2001) as the continuous counterpart of the CCC chart. This chart can be used to determine whether or not the

defect rate is low, without shortcomings. The idea of Calvin is here generalized based on a homogeneous Poisson process. This chart is called t-chart. In a process in which the dynamic of counts can be described by a homogeneous Poisson process at a constant occurrence rate λ, for the particular case in which $r = 1$, the time between two consecutive events follows the exponential distribution with parameter λ with cumulative density function as follows

$$\mathcal{F}_{Exp}(x) = 1 - \exp(-\lambda x). \tag{1.8}$$

The probabilistic control limits can be then computed as follows

$$UCL_{Exp} = -\ln(\alpha/2)/\lambda_0, \tag{1.9}$$

$$LCL_{Exp} = -\ln(1-\alpha/2)/\lambda_0. \tag{1.10}$$

Zhang et al. (2013) extended the principle of the CQC chart to higher-order control charts by waiting until the specified number r of events before plotting the corresponding statistic. The time between the i^{th} and $(i+r)^{th}$ events is modelled by the gamma distribution with rate λ and integer shape r. Being the exponential and the gamma charts in the continuous scale, they offer the advantage of achieving the desired false alarm rate in the design phase of a control chart, simplifying the calibration procedure and overcoming the limitation of their discrete counterparts.

1.1.3 Exponentially weighted moving average and cumulative sum charts

Shewhart charts are the most familiar charts in use, with many desirable features like simplicity of interpretation and implementation; end-users can directly read the value of the variable of interest on the chart. However, a significant drawback of the Shewhart control chart is that it uses only the information about the process in the last sample observation – it does not have a history. This *memoryless* property makes it insensitive to small shifts since the information contained in the sequence of the last observations is neglected. Practitioners can use run or pattern rules in conjunction with a standard Shewhart set-up to improve the sensitivity of the monitoring scheme. These rules allow for quickly detecting specific patterns or deviations in the data at the cost of raising the rate of false alarms. This disadvantage can be avoided by adapting the control limits to the selected run rules

in the design phase. The cumulative sum (CUSUM) and the exponentially weighted moving average (EWMA) control charts represent alternatives to Shewhart charts with run rules. They possess long- and medium-term memory properties, respectively. These charts perform excellently, mainly when the assignable causes do not typically result in significant process deviation from the target, with the drawback of being difficult to interpret as they do not display the raw data value and are more difficult to construct.

The CUSUM control chart uses the sum of observations instead of the current one and plots the cumulative sum of deviations from a target value (Page (1961)). This chart is convenient for plotting individual observations. The CUSUM chart can be obtained by plotting the statistics S_i^+ and S_i^- defined as follows

$$S_i^+ = \max\{S_0, S_{i-1}^+ + (X_i - k)\}, \tag{1.11}$$

$$S_i^- = \min\{S_0, S_{i-1}^- + (X_i - k)\}. \tag{1.12}$$

The initialization value S_0 is generally the mean of historical data or can be set at 0, and k is the reference value and can be calculated as follows for the gamma distribution case

$$k = r\frac{\ln(\theta_1) - \ln(\theta_0)}{\theta_1 - \theta_0}. \tag{1.13}$$

The EWMA control chart shows comparable performance to the CUSUM chart and plots the moving average weighted by a constant value w, with $0 \leq w \leq 1$. This value indicates the weight to give to the current observation. When designing the control chart, the weight represent an additional parameter to choose from, typically in the range $[0.2, 0.3]$. Gan (1993) introduced the exponential EWMA chart, where the observations Z_i are the inter-arrival times. The EWMA statistic $(Q_1, Q_2, \ldots, Q_k)$ for the upper-sided and lower-sided settings are the following

$$Q_i = \max\{0, (1-w)Q_{i-1} + wZ_i\}, \tag{1.14}$$

$$q_i = \min\{0, (1-w)Q_{i-1} + wZ_i\} \tag{1.15}$$

respectively. To avoid delay in detection, a positive value instead of zero can be selected to speed up the detection limit. These charts can sometimes be used with non-transformed data. When the

monitoring task is to detect both process improvement and deterioration simultaneously, a two-sided version can be used as shown in (1.16)

$$Q_i = (1 - w)Q_{i-1} + wZ_i. \tag{1.16}$$

The control limits of memory-type control charts are not as easily computed as for the Shewhart chart because of the violation of the independence assumption. To compute the control limits, exact methods have been proposed in the literature (e.g., Vardeman and Ray (1985)). Alternatively, they can be computed by simulations. The above-mentioned control charts are advantageous in specific settings, and their features have been extensively compared, emphasizing their strengths and weaknesses under different scenarios.

1.2 Scope and motivation

The present dissertation aims to extend the current statistical process monitoring state of the art by developing a practical framework for monitoring high-purity processes, with the intent to harmonize the relationship between the academic literature and implementations in manufacturing environments. The research task is therefore motivated by the practical importance of statistical process monitoring in the manufacturing industry by meeting the needs of practitioners yet relying on the simplicity of the concepts so that non-statisticians and daily users can welcome the proposed designs. The main driver of this research stems from real-case operations at Dow, where some process quality characteristics have a discrete and often multivariate nature. A prime example of these applications is the production of plastic pellets, which needs to meet stringent technical specifications so that the product can be classified as conforming. Among the qualities subjected to strict requirements, pellet contamination plays an essential role as the presence of polluting agents worsens the desired optical properties. In these settings, defects are recorded as non-negative counts and are often categorized in classes, such as size or color, intrinsically correlated. A low occurrence of contamination defects distinguishes this process, and the defect rate is strongly influenced by the production volume or inspected weight and, possibly, by other process variables. Moreover, different products with different specifications are produced in the same line to meet customer demand closely. In most cases, the sys-

tem's complexity is streamlined by renouncing to monitor the process in real-time and by evaluating the performance at large time intervals in terms of batch key performance indicators, which contain no information about the stability of the process and hinder rapid corrective actions. The present thesis pursues to develop practical, relevant monitoring schemes using rigorous statistical analysis that are relatively unexplored in literature; these include tailored change-point, regression, and multivariate control charts in the sphere of high-purity processes. Generalized likelihood ratio (GLR) charts offer tailored solutions for complex change-point scenarios without requiring parameter tuning. Moreover, they can accommodate multi-product production lines with different specifications. However, GLR control charts are less explored in literature than traditional competitors, particularly for discrete data. Furthermore, most research on monitoring high-quality processes restricts its attention to response data. This assumption is far from realistic, as it is plausible to assume that the response variables are correlated and dependent on other process variables. Therefore, accounting for factors influencing the process response is of practical interest and represents an important research area in modern quality control. Finally, in many applications, the quality of a product is described by a set of variables ($m > 1$). The multivariate space for cumulative quantity control charts is relatively unexplored. Multivariate settings are used predominantly for hybrid solutions with the intent to monitor the frequency and amplitude of an event but not more than one type or class of event simultaneously.

1.3 Contributions

The main contributions shaping the structure of the thesis are the following.

- Chapter 1 briefly reviews the control chart methodologies widely used for attribute data and highlights their limitations when applied to high-purity processes. It presents the current state of the art in statistical process monitoring, and the discussion helps identify gaps and research opportunities.
- Chapter 2 lays the foundation of the thesis by introducing and discussing the asynchrony of process time and control-chart time, a distinctive and recurrent feature of the high-purity monitoring framework. As a result, appropriate performance metrics are adapted to allow fair evaluation and comparison of the several monitoring strategies proposed.

- Chapter 3 proposes tailored control charts for high-purity processes based on more complex change-point models. These charts are based on generalized likelihood ratios (GLR), which offer extra layers of flexibility, making the monitoring purpose more practical and industry-driven.
- Chapter 4 introduces generalized linear model (GLM)-based control charts with the intent to detect contextual anomalies. These charts are suitable when accounting for the influence that expected fluctuations of a set of co-variates exert on the variable(s) of interest. The novelty resides in applying these strategies to cumulative quantities control charts.
- Chapter 5 extends the research to multivariate point processes control charts, contributing to the existing literature. These charts are convenient and appropriate when the production process should be described by multiple correlated temporal point models with simple correlation structures.
- Chapter 6 contains the conclusions and the recommendations for future work.

1.4 List of publications

This doctoral dissertation has been partially published in the following peer-reviewed journals and conferences.

Journals

- Rizzo C., and Di Bucchianico, A., Generalized linear model-based control charts for high-purity processes based on cumulative quantities. *Quality and Reliability Engineering International. 2023. (Preview published online.)*
- Rizzo, C., Di Bucchianico, A., Generalized likelihood ratio control charts for high-purity (high-quality) processes. *Quality and Reliability Engineering International* 2023; 39: 523– 531.
- Rizzo, C., Chin, S-T., Van den Heuvel, E., and Di Bucchianico, A., Performance measures of time-between-events control charts. *Quality and Reliability Engineering International* 2020; 36: 2754–2768.

Conferences

- 2021, Generalized likelihood ratio control charts for high purity processes, Rizzo C., Di Bucchianico, European Network for Business and Industrial Statistics (ENBIS).
- 2020, Performance metrics of time between events control charts in high-purity processes, Rizzo, C., Di Bucchianico A., et al., INFORMS Annual Meeting.
- 2019, Performance metrics of time between events control charts, Rizzo, C., Di Bucchianico A., et al., European Network for Business and Industrial Statistics (ENBIS).

Chapter 2

Performance Metrics for High-Purity Monitoring Methods

Traditional statistical process control techniques like Shewhart $\overline{X}$ or individual charts are known to perform unsatisfactorily in processes with low-level defect rates. Dedicated methods, such as the cumulative quantities control charts, have been developed for discrete and continuous time, as discussed in Section 1.1.

During the design of a monitoring system, practitioners and implementers are first asked to choose the most appropriate type of control chart. This informed choice can be made by determining which method shows the best performance given a specific application by comparing different methodologies using suitable performance metrics. Performance evaluation of control charts is a crucial topic, and the connection of mathematical theory with the needs of practitioners is not receiving sufficient attention.

Performance metrics in SPC can be classified into two main categories. The first group is associated with the false-alarm rate expected when the process is 'in-control'. At the same time, the second group is concerned with the detection delay when the process has shifted to an 'out-of-control' state. Methods are often compared in terms of average run length (ARL). Nonetheless, the ARL suffers from several shortcomings. First, the ARL is a *zero-state* performance metric, i.e., it relies on the assumption that the change-point happens at the beginning of the monitoring. Moreover, run length distributions for cumulative quantity charts neglect the asynchrony of the two time scales; the number of items and

the number of control chart decisions differs, and it is more appropriate to consider inspection lengths instead. Finally, since run length distributions are typically skewed, means are not sufficient to fully describe their behavior, and additional summary statistics, like the standard deviation or quantiles, should be considered. The rest of the chapter is structured as follows. Section 2.1 discusses the zero-state and steady-state performance metrics proposed in the literature for standard control charts, while Section 2.2 adapts these measures to cumulative quantities control charts, with a strong focus on the asynchrony between process-time and control chart-time.

2.1 Performance evaluation of standard methods

Before discussing the case of control charts which involve data in an aggregated form, it is worthwhile to review the performance assessment of control charts in general. Statistical monitoring aims at quickly detecting process changes by using observations that arrive sequentially. Several performance measures have been introduced in the literature to quantify what 'quickly' means. The appropriate choice of performance metrics in comparing monitoring techniques is of crucial importance, particularly for low-level defect processes. A practically meaningful monitoring method is often the result of a trade-off between the tolerable false alarm rate and a fast shift detection, whereas, in more complex situations, other specific requirements inherent to the process and the application are more appropriate. In applied work, an evaluation conducted by using several measures is required. Commonly used metrics and optimality criteria have been extensively reviewed by Frisén (2009) and Kenett and Pollak (2012).

2.1.1 Steady-state versus zero-state performance

There are two assumptions under which control chart can be evaluated: *zero-state* and *steady-state*. In zero-state analysis, the process shift is assumed to occur immediately at the start of the monitoring procedure or when the chart is in its initial state. The terms initial state and zero-state can be used interchangeably. Traditionally, control charts are evaluated and compared based on zero-state performance metrics. This strong assumption tends to over-evaluate control chart performance. Therefore, one should use steady-state metrics instead, assuming that the parameters' shift occurs at a random

time after the monitoring has begun. This assumption is more realistic in most applications. Depending on this choice, one may obtain very different performance results. Therefore, a steady-state analysis should be performed when evaluating methods, particularly when the change-point τ plays an important role in the model. The steady-state approach is discussed in Szarka and Woodall (2011) for the discrete time and in Zhang et al. (2013) and Woodall and Driscoll (2015) for the continuous counterpart. These considerations should be taken into account when considering both in-control and out-of-control performance metrics, discussed in detail in the following sections.

2.1.2 In-control performance

When designing a control chart, a key parameter to control is the false alarm rate. The probability of a false alarm is the probability of having a signal when the process is in-control. Following this classification, the most common index is the average run length in-control (ARL_0), defined as the number of points plotted before a control chart signal is given when no change in the process has occurred. The strong point of the ARL criterion is its computational simplicity, but despite its popularity in practice, the use of this metric is controversial (Frisén and Wessman (1999)). One of the reasons can be attributed to the skewness of its distribution; the standard deviation is relatively large, and the average value alone may not suffice to describe the performance of a given monitoring system. Gan (1993) proposed using the median run length (MLE) as a more robust metric. Reporting the percentiles of the run-length distribution is also considered a valid alternative option. Additionally, an increasing number of monitoring applications limits its use, particularly when data aggregation at different levels is required for monitoring purposes over time. In such cases, the ARL becomes meaningless, and other performance metrics should be used instead (Woodall and Driscoll (2015)). Cumulative quantity control charts represent a relevant example of such applications. The limitations of the ARL when applied to cumulative quantity control charts are presented in Section 2.2. Regarding in-control metrics, it is also worthwhile to mention that the procedure to define α as the probability of Type I error and subsequently calculate the ARL, is still in vogue. As argued in Margavio et al. (1995), a direct relationship between a single alarm probability and the ARL exists only for basic Shewhart-type control charts. These charts rely on the assumption that all the occurrences are independent within and between samples, which allows electing a constant false alarms probability. This definition is unclear for more

advanced monitoring methods such as CUSUM, EWMA charts and Shewhart charts with run rules, since their statistics are not independent and therefore the definition of α is meaningless. Adams et al. (1992) describe this problem in detail. In the literature, the performance metrics of these inherently dependent statistics are sometimes evaluated assuming independence, leading to severe flaws in the design of a control chart, driven by inaccurate performance evaluation. Haq and Woodall (2022) identify several studies in which authors incorrectly approximate the signal event using independent Bernoulli random variables for EWMA charts. To overcome the zero-state nature of the ARL, Margavio et al. (1995) proposed that the alarm rate is defined as the probability of an alarm at sample i given no alarm prior to the i^{th} sample, i.e.,

$$r_i = P(N = i \mid N > i - 1), i = 1, 2, \ldots. \tag{2.1}$$

where N is the run length. It can be noticed that, in the stationary state, (2.1) is equal to $\frac{1}{ARL_0}$, with the advantage that it can be applied to any control chart. From a Bayesian point of view, where there is a prior on the change-point τ, the quantity of interest for describing the possibility of a false alarm is the probability of false alarm (PFA) = $P(T < \tau)$. The probability of false alarm is defined as the probability of an alarm no later than at time τ given that no change has occurred; it corresponds to the probability of erroneous rejection of the null hypothesis, at the level of significance chosen, but it is a function of the time τ.

2.1.3 Out-of-control performance

Any deviation of the model's parameters from their in-control values represents a shift to the out-of-control state, which generally requires an action. The performance of a method for monitoring sequential schemes depends on the time of the change τ. Precise definitions of τ are provided in Section 2.2. In some circumstances, it is appropriate to express the performance measure as a function of τ. Therefore, different approaches have been used to obtain evaluations independent of their value. The most widely used performance index is the ARL_1, which denotes the average run length until the detection of a true change. The SPC optimality criterion is to minimize ARL_1 when ARL_0 is fixed. As mentioned before, the ARL_1 is a zero-state performance metric. From the steady-state point of view, the delay from the first opportunity to detect a shift after the change-point denoted by τ, conditional

on not having raised a false alarm before is called conditional expected delay (CED) and it is defined as follows

$$\mathrm{CED}(\tau) = \mathrm{E}[\mathrm{N} - \tau + 1 \mid \mathrm{N} \geq \tau]. \tag{2.2}$$

Depending on the process requirements, it can be of interest to evaluate the response in the initial stage, where CED(1) is kept low. If, on the contrary, the change-point time is expected at higher τ, then the $\lim_{\tau\to\infty} \mathrm{CED}(\tau)$ is more appropriate. It can be easily noticed that $\mathrm{ARL}_1 = \mathrm{CED}(1)$. For most methods, the CED will converge to a constant value when the time of the change increases. The asymptotic value is called steady-state average delay time (SADT). To get a better understanding of the characteristics of a method, Frisén (2009) suggests calculating the CED for several values of τ. It must be noted that different formulation of (2.2) are available in literature, depending on the definition of the change-point τ and the first opportunity to detect a change. Kenett and Pollak (2012) agree that the CED in the deterministic framework and the PFA in the Bayesian framework represent more natural performance measures, stating that using of the CED could lead to a reassessment of procedures proposed in the literature, which typically rely on the widespread use of the ARL_1. However, ensuring the fastest response is not always sufficient. On some occasions in the manufacturing industry in particular, there is a limited window of opportunity to take action. In such cases, the probability of successful detection (PSD) is a convenient metric, as it measures the probability of detection within the constant time distance d from the time of change, conditional on not having raised a false alarm before. It can be expressed as follows

$$\mathrm{PSD}(\mathrm{d}, \tau) = \mathrm{P}(\mathrm{N} - \tau + 1 \leq \mathrm{d} \mid \mathrm{N} \geq \tau). \tag{2.3}$$

Different periods can be considered according to specific cases. For example, when the objective is to monitor a quick detection of major intensity, small values of d are important; when the long-term detection ability has more influence, larger values of d need to be evaluated.

2.2 Performance evaluation of high-purity processes methods

A distinctive feature of the monitoring framework on high-purity processes, particularly for cumulative quantity control charts, not present in standard setups for variable control charts, is that process time and control chart time scale might be asynchronous. Inevitably, the performance metrics must be adapted. As discussed in Section 2.1, the ARL is by far the most used metric in evaluating the performance of control charts for its simplicity and easy interpretation. However, when deploying aggregate charts, the ARL should not be used because of two main limitations:

- It expresses the performance in control chart-clock time and not in process-clock time;
- It cannot be used to compare control charts that involve aggregate data at different levels.

As depicted in Figure 2.1, given a Bernoulli process Y_i, conforming and non-conforming events are registered in process-clock time. However, when switching to a cumulative quantity control chart, the monitored random variable X_i represents the cumulative count or time until the r^{th} non-conforming event; the number of points plotted (decisions) depends on the non-conforming items and the aggregation level chosen so the control chart-clock time does not coincide with the process-clock time. Evaluating the performance in terms of non-conforming items is meaningless and the comparison of control charts at different aggregation levels is impractical. Since the change-point time has a crucial role in statistical process control, it is important to define and distinguish between process time and control chart time, both for model interpretability and fair performance evaluation and comparison. To properly define the time scales, the original (process) random variables are denoted by Y_i and the derived, aggregate random variables from the control chart, by X_j. Figure 2.1 illustrates the case for CCC_1 and CCC_2 control charts, i.e. the case that the Y_i's are Bernoulli random variables and the $X_j(1)$ and $X_j(2)$ are the number of conforming items until the first and the second event, respectively.
Note that the time until the r^{th} time follows a negative binomial distribution, and the geometric distribution for the special case $r = 1$.
There is a different time scale for the control chart, since it is related to the indices of the random variables X_j. Suppose that in the example in Figure 2.1, the change-point at the process time scale is set such that $\tau_\pi = 12$ and that the lower control limits are set at 5 and 10, for the $X_j(1)$ and $X_j(2)$

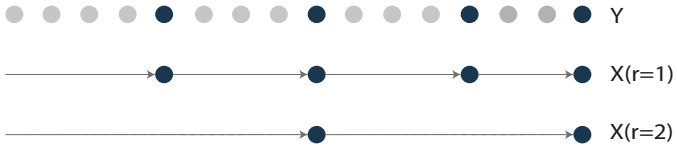


Figure 2.1: Number of points plotted for a Bernoulli process Y_i and the corresponding aggregate charts $X_j\,(r=1)$ and $X_j\,(r=2)$.

processes, respectively. For the $r=1$ case, the change-point occurs between $j=2$ and $j=3$, so that the control chart can only signal from $j=3$ onward and it will signal at $j=3$ since $X_3(1)=4$. For the case $r=2$, the change-point happens between $j=1$ and $j=2$, so that the control chart can only signal from $j=2$ onward and it will signal at $j=2$ since $X_2(2)=7$.

When the change-point at the process scale is zero, i.e., the process is out-of-control from the start, then $\tau=\tau_\pi=0$. If $\tau_\pi\geq 1$, then for the CCC_1 chart the change-point τ at control chart time scale in can be computed from τ_π via the following formula

$$\tau=\min\left\{\sum_{i=1}^{\eta}Y_i:\eta\geq\tau_\pi\right\}. \tag{2.4}$$

It follows from Figure 2.1 that for $r=1$ and $\tau_\pi=12$, then $Y_6=Y_{11}=Y_{15}=1$, so $\tau=3$. The random variable $X_3(1)$ has a mixed distribution and it is only starting with $X_4(1)$ that a 'pure' out-of-control situation occurs. For a CCC_2 chart, the formula in (2.4) needs to be adapted to incorporate that there is a control chart observation at every second value of $Y_i=1$

$$\tau=\min\left\{\left\lceil\sum_{i=1}^{\eta}Y_i/2\right\rceil:\eta\geq\tau_\pi\right\}. \tag{2.5}$$

It is evident in Figure 2.1 for $r=2$ and $\tau_\pi=12$ that $Y_6=Y_{11}=Y_{15}=Y_{18}=1$, so $\tau=2$. Also in this case, $X_1(2)$ has a mixed distribution and it is only starting with $X_2(2)$ that the 'pure' out-control situation occurs. In this case, the cumulative number of items until the r^{th} event follows the negative binomial distribution.

In the sequel, the essential probabilistic background for computing the performance of aggregate

control charts is presented and explicit formulas for the performance of the t_r-chart and the CCC_r chart in terms of length of inspection are derived. Additionally, steady-state alternatives to the length of inspection are adapted to the aggregate data framework.

2.2.1 Length of inspection

When the time interval between cumulative observations varies, the average length of inspection (ALI), often referred as the average number of observations to signal (ATS), is a more appropriate measure. The length of inspection (LI) is defined as follows

$$LI = \sum_{i}^{N} X_i, \tag{2.6}$$

where N is the run length. The ALI reflects the actual number of inspected items before the control chart raises a signal.

It is worth briefly mentioning that the ALI performance in the two-sided charts often presents the non-maximality problem due to the skewness of its distribution. In this case, the ALI is not maximum at the designed ALI_0 but peaks at some other value. This shortcoming is addressed in Xie et al. (1998) by calculating the control limit maximizing the performance metrics at the in-control parameter instead of using probabilistic limits. The stopping criterion of a control chart in online monitoring obviously depends on current and past observations. The random variable connected to it is thus an example of the mathematical concept of stopping time, defined as follows.

Definition 2.2.1. *Let* $X, X_1, X_2, \ldots, X_n$ *be an i.i.d. sequence of random variables. A random variable* N *is called a* stopping time *if for each* $n \geq 0$*, the event* $\{N = n\}$ *depends only on* $X_1, X_2, \ldots, X_n$.

Specific definitions of stopping times are considered for the control charts based on aggregate data for two-sided, upper-sided and lower-sided cases, respectively.

Definition 2.2.2. *Let* $X, X_1, X_2, \ldots, X_n$ *be an i.i.d. sequence of random variables with cumulative distribution function* $\mathcal{F}_X$*. The following stopping times are defined with respect to the sequence* $X_1, X_2, \ldots, X_n$.

TWO-SIDED CONTROL LIMITS

$$N_{\ell,u} = \min \{i \mid X_i < \ell \vee X_i > u\}$$

which follows a geometric distribution with parameter $p = 1 - (\mathcal{F}_X(u) - \mathcal{F}_X(\ell))$.

One-sided upper control limit

$$N_u = \min\{i \mid X_i > u\}$$

which follows a geometric distribution with parameter $p = 1 - \mathcal{F}_X(u)$.

One-sided lower control limit

$$N_\ell = \min\{i \mid X_i < \ell\}$$

which follows a geometric distribution with parameter $p = \mathcal{F}_X(\ell)$.

The following theorem due to Wald allows practitioners to easily compute the average inspection length given the process and the stopping time parameters. The original identity in Wald (1945) is valid for N independent from the observation sequence. Related results from Blackwell (1946) and Wolfowitz (1947) are considered.

Theorem 2.2.3 (Wald - Blackwell - Wolfowitz)**.** *Let* $X, X_1, X_2, \ldots X_n$ *be i.i.d. random variables with finite mean* μ *and finite variance* σ^2*. Let* N *be a stopping time with respect to* $X_1, X_2, \ldots, X_n$ *such that* $E(N) < \infty$*. Then*

$$E\left(\sum_{i=1}^{N} X_i\right) = \mu E(N), \tag{2.7}$$

$$E\left(\sum_{i=1}^{N} X_i - \mu N\right)^2 = \sigma^2 E(N) \tag{2.8}$$

As noted before, the inspection length distributions are typically skewed and other summary statistics than averages should be considered. In particular, the standard deviation is an useful addition. The first version of the standard deviation was obtained by Blackwell and Girshick (1946).

Theorem 2.2.4. [Blackwell-Girshick] *Let* $X, X_1, X_2, \ldots$ *be i.i.d. random variables with mean* $\mu < \infty$ *and variance* $\sigma^2 < \infty$*. If* N *is independent of* $X_1, X_2, \ldots$*, then*

$$\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) = E(N)\mathrm{Var}(X) + \mathrm{Var}(N)E^2(X). \tag{2.9}$$

Proof

$$\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) = \mathrm{E}\left(\sum_{i=1}^{N} X_i\right)^2 - \mathrm{E}^2\left(\sum_{i=1}^{N} X_i\right)$$

Conditioning on N the first term

$$\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{N} X_i\right)^2 &= \mathrm{E}\left[\mathrm{E}\left[\left(\sum_{i=1}^{N} X_i\right)^2 \mid N = k\right]\right] \\
&= \sum_{k=1}^{\infty} \mathrm{E}\left(\sum_{i=1}^{k} X_i\right)^2 \mathrm{P}(N = k) \\
&= \sum_{k=1}^{\infty} \left(k\mathrm{Var}(X) + k^2\mathrm{E}^2(X)\right)^2 \mathrm{P}(N = k) \\
&= \mathrm{E}(N)\mathrm{Var}(X) + \mathrm{E}(N^2)\mathrm{E}^2(X).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) &= \mathrm{E}(N)\mathrm{Var}(X) + \mathrm{E}(N^2)\mathrm{E}^2(X) - (\mathrm{E}(N)\mathrm{E}(X))^2 \\
&= \mathrm{E}(N)\mathrm{Var}(X) + [\mathrm{E}(N^2) - \mathrm{E}^2(N)]\mathrm{E}^2(X) \\
&= \mathrm{E}(N)\mathrm{Var}(X) + \mathrm{Var}(N)\mathrm{E}^2(X).
\end{aligned}$$

It must be noted that, unlike the Wald Theorem (2.2.3), the Blackwell-Girshick formula in (2.9) requires N to be independent from X_i. A simple counterexample shows that the formula is incorrect without that assumption. Let X_i be Bernoulli random variables with parameter p and let N be the first index for which $X_N = 1$. Then $\sum_{i=1}^{N} X_i = X_N = 1$ a.s., and thus the variance of that sum equals zero, while the right-hand side of the Blackwell-Girshick formula obviously does not equal zero. This fact seems to be overlooked in some of the control chart literature where the Blackwell-Girshick formula is applied in situations where it does not hold and thus yields incorrect results. A corrected formula for the variance was given as Theorem 2 in Di Bucchianico et al. (2005). The first part of the proof also holds for the following more general case.

Theorem 2.2.5. *Let* $X, X_1, X_2, \ldots, X_n$ *be i.i.d. random variables with mean* $\mu < \infty$ *and variance* $\sigma^2 < \infty$. *Let* N *be any of the stopping times in Definition 2.2.1 with respect to the sequence* $X_1, X_2, \ldots, X_n$ *such that*

$E(N) < \infty$. *Then*

$$\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) = \frac{\sigma^2}{p} - \mu^2\left(\frac{3-p}{p^2}\right) + 2\mu E\left(N\sum_{i=1}^{N} X_i\right) \tag{2.10}$$

where p *is the geometric distribution parameter of* N.

Proof. In general

$$E(X \mid X \geq X) = \frac{\int_0^\infty t f(t)dt}{1 - \mathcal{F}(0)}$$

Assuming N and X_i independent, with $X_i \sim \mathrm{Exp}(\lambda)$ and $N \sim \mathrm{Geom}(p)$

$$\begin{aligned}
P\left(\sum_{i=0}^{N} X_i \leq X\right) &= \sum_{k=0}^{\infty} P\left(\sum_{i=0}^{k} X_i \leq X \mid N = k\right) P(N = k) \\
&= \sum_{k=0}^{\infty} f_{\Gamma_{(k,\lambda)}}(t)(1-p)^{k-1}p \\
&= p\sum_{k=1}^{\infty} f_{\Gamma_{(k,\lambda)}}(t)(1-p)^{k-1} \\
&= p\sum_{k=1}^{\infty} \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}(1-p)^{k-1} \\
&= p\lambda e^{-\lambda x}\sum_{l=0}^{\infty} \frac{\lambda^l x^l}{l!}(1-p)^l \\
&= \lambda p e^{-\lambda x p} \sim \mathrm{Exp}(\lambda p).
\end{aligned}$$

Then,

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) &= E\left(\sum_{i=1}^{N} X_i - \mu E(N)\right)^2 = E\left(\sum_{i=1}^{N} X_i - \mu N + \mu N + \mu E(N)\right)^2 \\
&= E\left(\left(\sum_{i=1}^{N} X_i - \mu N\right)^2 + \mu^2(N + E(N))^2 + 2\mu\left(\sum_{i=1}^{N} X_i - \mu N\right)(N + E(N))\right) \\
&= \sigma^2 E(N) + \mu^2 E(N - E(N))^2 + 2\mu\left(E\left(N\sum_{i=1}^{N} X_i\right) - \mu E(N^2)\right) \\
&= \sigma^2 E(N) + \mu^2\left(E(N^2) - E^2(N)\right) + 2\mu E\left(N\sum_{i=1}^{N} X_i\right) - 2\mu^2 E(N^2).
\end{aligned}$$

Exploiting the geometric nature of the stopping time (N) for which $E[N] = 1/p$ and $E[N^2] = (2-p)/p^2$,

the equation simplifies to

$$\text{Var}\left(\sum_{i=1}^{N} X_i\right) = \frac{\sigma^2}{p} - \mu^2\left(\frac{3-p}{p^2}\right) + 2\mu E\left(N\sum_{i=1}^{N} X_i\right).$$

Conditioning on the values of N, for continuous X_i

$$\begin{aligned}
E\left(N\sum_{i=1}^{N} X_i\right) &= \sum_{k=1}^{\infty} E\left(N\sum_{i=1}^{N} X_i \mid N=k\right) P(N=k) = \sum_{k=1}^{\infty} k\sum_{i=1}^{k} E(X_i \mid N=k)P(N=k) \\
&= \sum_{k=1}^{\infty} k\left[(k-1)E\left(X \mid X>\ell \vee X<u\right) + E(X_k \mid X_k<\ell \vee X_k>u)\right] P(N=k) \\
&= pE(X \mid X>\ell \vee X<u)\sum_{k=1}^{\infty} k(k-1)(1-p)^{k-1} \\
&+ pE(X \mid X<\ell \vee X>u)\sum_{k=1}^{\infty} k(1-p)^{k-1} \\
&= \frac{2(1-p)}{p^2(1-p)}\int_{\ell}^{u} x f_X(x)dx + \frac{1}{p^2}\left(\int_0^{\ell} x f_X(x)dx + \int_u^{\infty} x f_X(x)dx\right) \\
&= \frac{2}{p^2}\int_{\ell}^{u} x f_X(x)dx + \frac{1}{p^2}\left(\int_0^{\ell} x f_X(x)dx + \int_u^{\infty} x f_X(x)dx\right) \\
&= \frac{1}{p^2}\left(\int_0^{\infty} x f_X(x)dx + \int_{\ell}^{u} x f_X(x)dx\right) \\
&= \frac{1}{p^2}\left(\mu + \int_{\ell}^{u} x f_X(x)dx\right).
\end{aligned}$$

Consider the continuous random variables V_i with common density function f_V. Then (2.2.5) yields for the two-sided control chart

$$\text{Var}\left(\sum_{i=1}^{N_{\ell,u}} V_i\right) = \frac{\sigma^2}{p} - \mu^2\left(\frac{1-p}{p^2}\right) + \frac{2\mu}{p^2}\int_{\ell}^{u} v f_V(v)dv. \tag{2.11}$$

Analogously, a general form for the variance of the length of inspection for the upper- and the lower-sided cases for continuous variables V_i can be written as follows

$$\text{Var}\left(\sum_{i=1}^{N_u} V_i\right) = \frac{\sigma^2}{p} - \mu^2\left(\frac{1-p}{p^2}\right) + \frac{2\mu}{p^2}\int_{0}^{u} v f_V(v)dv \tag{2.12}$$

and

$$\text{Var}\left(\sum_{i=1}^{N_\ell} V_i\right) = \frac{\sigma^2}{p} + \mu^2\left(\frac{1+p}{p^2}\right) - \frac{2\mu}{p^2}\int_0^\ell v f_V(v)dv. \tag{2.13}$$

The same calculations can be carried out considering the case in which Y_i follows a discrete distribution. Di Bucchianico et al. (2005) calculated the variance for the lower-sided CCC_r chart. The complete set of general formulas for discrete Y_i is as follows

$$\text{Var}\left(\sum_{i=1}^{N_{\ell,u}} Y_i\right) = \frac{\sigma^2}{p} + \mu^2\frac{(1+p)}{p^2} - \frac{2\mu}{p^2}\sum_{j=c}^{d} j\mathbb{P}(Y=j), \tag{2.14}$$

$$\text{Var}\left(\sum_{i=1}^{N_u} Y_i\right) = \frac{\sigma^2}{p} + \mu^2\frac{(1+p)}{p^2} - \frac{2\mu}{p^2}\sum_{j=0}^{c-1} j\mathbb{P}(Y=j), \tag{2.15}$$

and

$$\text{Var}\left(\sum_{i=1}^{N_\ell} Y_i\right) = \frac{\sigma^2}{p} - \mu^2\frac{(1-p)}{p^2} + \frac{2\mu}{p^2}\sum_{j=0}^{d} j\mathbb{P}(Y=j). \tag{2.16}$$

Thus far, the nature of X_i has not been exploited. As discussed in Section 1.1, the dynamics of low-rate counts can described by an homogeneous Poisson process (T_i) at a constant occurrence rate λ. Therefore, the time between r events is modelled by the gamma distribution with rate λ and integer shape r. The lower and upper control limits are denoted by ℓ and u, respectively. Following Theorem 2.2.3, the expected value of the length of inspection for the t_r-chart can be easily computed as the product of the expectations of T and N as follows

$$E\left(\sum_{i=1}^{N} T_i\right) = E(T)\cdot E(N) = \frac{r}{\lambda}\cdot\frac{1}{p}. \tag{2.17}$$

In the general case for $r > 1$, when different aggregation levels are considered, (2.11) yields to

$$\text{Var}\left(\sum_{i=1}^{N_{\ell,u}^{(r)}} T_i\right) = \frac{\sigma^2}{p} - \mu^2\left(\frac{1-p}{p^2}\right) + \frac{2\mu}{p^2}\int_\ell^u t\frac{\lambda^r t^{r-1}e^{-\lambda t}}{(r-1)!}dt \tag{2.18}$$

where

$$\int_{\ell}^{u} t\frac{\lambda^r t^{r-1}e^{-\lambda t}}{(r-1)!}dt = \frac{r}{\lambda}\left(F_{\Gamma(r+1,\lambda)}(u) - F_{\Gamma(r+1,\lambda)}(\ell)\right) \tag{2.19}$$

and, for the upper and lower one-sided cases

$$\text{Var}\left(\sum_{i=1}^{N_u^{(r)}} T_i\right) = \frac{\sigma^2}{p} + \mu^2\left(\frac{1+p}{p^2}\right) - \frac{2\mu}{p^2}\int_0^{\ell} t\frac{\lambda^r t^{r-1}e^{-\lambda t}}{(r-1)!}dt \tag{2.20}$$

and

$$\text{Var}\left(\sum_{i=1}^{N_\ell^{(r)}} T_i\right) = \frac{\sigma^2}{p} - \mu^2\left(\frac{1-p}{p^2}\right) + \frac{2\mu}{p^2}\int_0^{u} t\frac{\lambda^r t^{r-1}e^{-\lambda t}}{(r-1)!}dt \tag{2.21}$$

respectively. A closed form for the variance of the length of inspection can also be obtained for the particular case $r = 1$.

Design choice based on the length of inspection The performance used for evaluating the t_r-chart for lower-sided control charts can be computed numerically. The comparison between t_r-charts is carried out by specifying a priori the desired ALI_0. This ensures that the comparison among the different aggregate levels is fair. Moreover, the ALI_0 represents a more intuitive metric for practitioners as they are allowed to choose how many false alarms are able to tolerate within one time frame (e.g., once per operator shift or once per day). The procedure for choosing the optimal control chart order based on the ALI and the SDLI (i.e., standard deviation of the length of inspection) can be summarized as follows.

1. Calculate the control limits based on the estimated λ_0 and the chosen ALI_0
2. Determine the ALI_1 and $SDLI_1$ for the process mean shift of interest (λ_1)
3. Choose the control chart which minimize the ALI_1 and $SDLI_1$

where λ_0 is the Poisson parameter in-control and λ_1 is the Poisson parameter out-of-control, after the shift in rate. This approach assumes that sufficient historical data are available to estimate confidently the in-control parameters, condition often realistic in manufacturing environments, particularly when

on-line measurements are available. A numerical example is carried out considering $\lambda_0 = 0.05$ and $ALI_1 = 500$ for the lower-sided and the upper-sided t_r-chart and a set of λ_1. The set of parameters chosen reflects real-case scenarios within Dow. As shown in Table 2.1, for some values of λ_1, the two

	ALI_1				**$SDLI_1$**			
λ_1	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 1$	$r = 2$	$r = 3$	$r = 4$
0.05	500	500	500	500	519	523	522	518
0.075	224	172	149	**139**	237	185	159	**143**
0.1	127	83.6	69.9	**66.4**	137	92.5	74.4	**64.4**
0.125	82.5	49.3	**41.7**	41.8	89.7	55.4	43.0	**36.4**
0.15	57.8	32.7	**28.7**	30.6	63.8	37.0	28.2	**23.5**
0.175	42.9	23.6	**21.7**	24.5	47.9	26.7	20.0	**16.7**
0.2	33.2	18.0	**17.5**	20.7	37.5	20.2	15.0	**12.3**

Table 2.1: Lower-sided t_r-charts ALI_1 and $SDLI_1$ for $\lambda_1 = 0.05$ and $ALI_0 = 500$.

	ALI_1				**$SDLI_1$**			
λ_1	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 1$	$r = 2$	$r = 3$	$r = 4$
0.05	500	500	500	500	519	523	522	518
0.025	**200**	208	223	243	119	108	103	**102**
0.001	**190**	251	327	413	**108**	137	167	196
0.0075	**216**	307	417	540	**137**	183	226	264
0.005	**276**	428	609	803	**202**	277	343	399

Table 2.2: Upper-sided t_r-charts ALI_1 and $SDLI_1$ for $\lambda_1 = 0.05$ and $ALI_0 = 500$.

metrics point to a different chart order (i.e., from $\lambda_1 = 0.125$ to $\lambda_1 = 0.2$). In this situation, guidelines for choosing the optimal r are generally not available. Here two criteria are proposed based on which it is possible to determine the optimal order of the control chart based on both summary statistics.

1. Minimum ALI + kSDLI (e.g., $k = 0.5$);

2. ALI at most 10% larger than the min, then choose the one with the smallest SDLI.

For the case $\lambda_1 = 0.15$, both criteria would lead to choose the t_4-chart for this particular application by choosing $k = 0.5$ for the first criteria. The value of k can be modulated based on the importance

given to the measure of spread. The selection criteria proposed can be also applied for discrete and for two-sided and upper-sided settings. It is important to notice that for the lower-sided t_r-chart, higher-orders are preferred, showing faster responses to a shift of the rate parameter, particularly when its magnitude is relatively small. However, for upper-sided t_r-chart, lower order (i.e., $r = 1$) are preferred as shown in Table 2.2. This is justified by the fact that upper-sided t_r-chart are meant to detect improvements in the process, i.e., a defect rate decrease. In general, improvements are difficult to detect in these settings, mainly because of the nature of the process and the data aggregation set-up. A decision is prompted only when a defect is recorded, meaning that for improvement detection it might require a long time before a signal is raised. Higher-order control charts contribute to delay the out-of-control alarm.

Computing the exact values of the performance metrics also allows to quantify the error committed when using the Blackwell-Girshick formula in applications in which the strong assumption of N being independent from X_i does not hold. Table 2.3 and Figure 2.2 clearly show how serious the numerical error is when incorrectly applying the Blackwell-Girshick formula in this case. The absolute values of

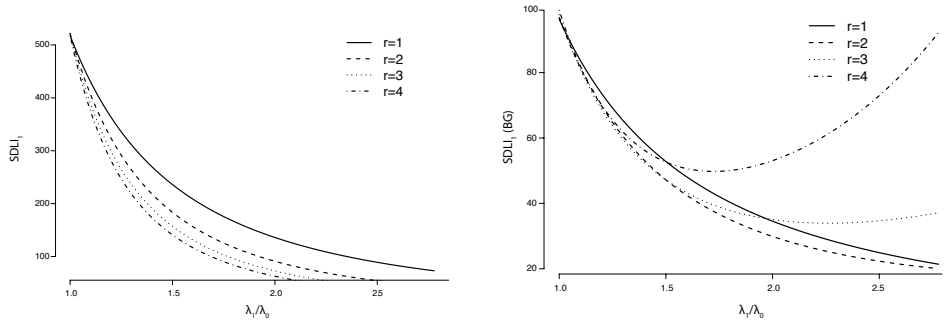
	SDLI$_1$				**SDLI$_1$(BG)**			
λ_1	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 1$	$r = 2$	$r = 3$	$r = 4$
0.05	519	523	522	518	100	101	102	104
0.1	137	92.5	74.4	**64.4**	36.0	**31.7**	38.6	60.9
0.15	63.8	37.0	28.2	**23.5**	**20.0**	20.6	46.9	138
0.2	37.5	20.2	15.0	**12.6**	**13.3**	19.3	79.6	422

Table 2.3: Lower-sided t_r-charts SDLI and SDLI(BG) comparison for $\lambda_1 = 0.05$, $ALI_1 = 500$ and for different values of r.

SDLI and SDLI(BG) differ of orders of magnitude and the trend showed by the SDLI(BG) is unexpected and it deviates from the distribution assumption of the length of inspection.

2.2.2 Conditional expected delay and probability of successful detection

The major weakness of the length of inspection is the assumption that the parameter shift happens at the beginning of the monitoring process, namely zero-state comparison, as discussed in Section 2.1.1. The conditional expected delay (CED) and the probability of successful detection (PSD) repre-



(a) $SDLI_1$ vs λ_1/λ_0 (b) $SDLI(BG)_1$ vs λ_1/λ_0

Figure 2.2: Lower-sided t_r-charts $SDLI_1$ and $SDLI_1(BG)$ versus λ_1/λ_0 for $\lambda_0 = 0.05$ and $ALI_0 = 500$.

sent alternative metrics to the length of inspection accounting the more realistic scenario in which the change-point happens some time after the monitoring has begun, i.e., steady-state comparison. The CED and the PSD adapted to the cumulative quantities control charts are as follows

$$CED(\tau) \quad = \quad E\left[\sum_{i}^{N} X_i - \tau + 1 \middle| \sum_{i}^{N} X_i \geq \tau\right] \tag{2.22}$$

and

$$PSD(d, \tau) \quad = \quad P\left(\sum_{i}^{N} X_i - \tau + 1 \leq d \middle| \sum_{i}^{N} X_i \geq \tau\right), \tag{2.23}$$

respectively. The right metric, or combination of different metrics, to use depends on what the practitioner wants to/or should optimize.

As an illustrative example, these measure are computed via Monte Carlo simulations for lower-sided t_r-charts with two additional run rules. The tests considered are the following.

- **Test 1**: 1 point below the lower control limit
- **Test 2**: 9 points in a row on the same side of the centre line
- **Test 3**: 6 points in a row, all increasing or all decreasing

The rules chosen aim at detecting three different shift modes: large shifts, small to moderate shifts and trends. Note that Shewhart charts without run rules are memory-less and therefore the concepts the average length of inspection and conditional expected delay coincide.

The behaviour of the CED and PSD is simulated for a lower-sided t_r-chart with λ_0 = 0.05, calibrated at ALI_0 = 500, considering a range of shift in the defect occurrence rate from 0.055 to 0.2.

Figures 2.4 and 2.5 show the kernel density plots of the distributions of the LI and CED since these coincide for $\tau = 0$. Negative values are an artifact of the kernel density estimator used. The CED optimality problem is more complex than the case for the LI, since the metric is expected to be sensitive to τ. In Figures 2.5 and 2.6 show the CED distributions and values, considering three different change-point scenarios (i.e., $\tau = 0, \tau = 175, \tau = 350$). The maximum τ is chosen based on the median length of inspection (MLI) for the desired in-control parameter λ_0. In the case of Shewhart charts without run

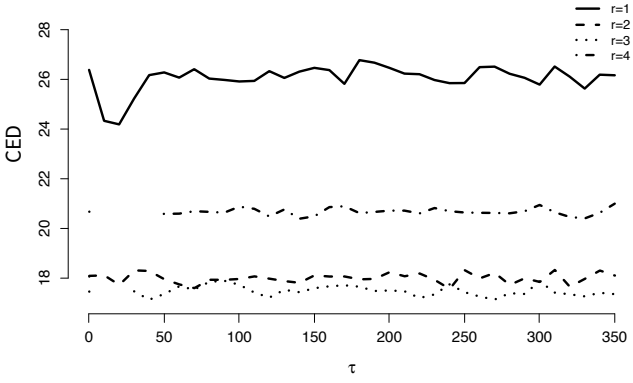
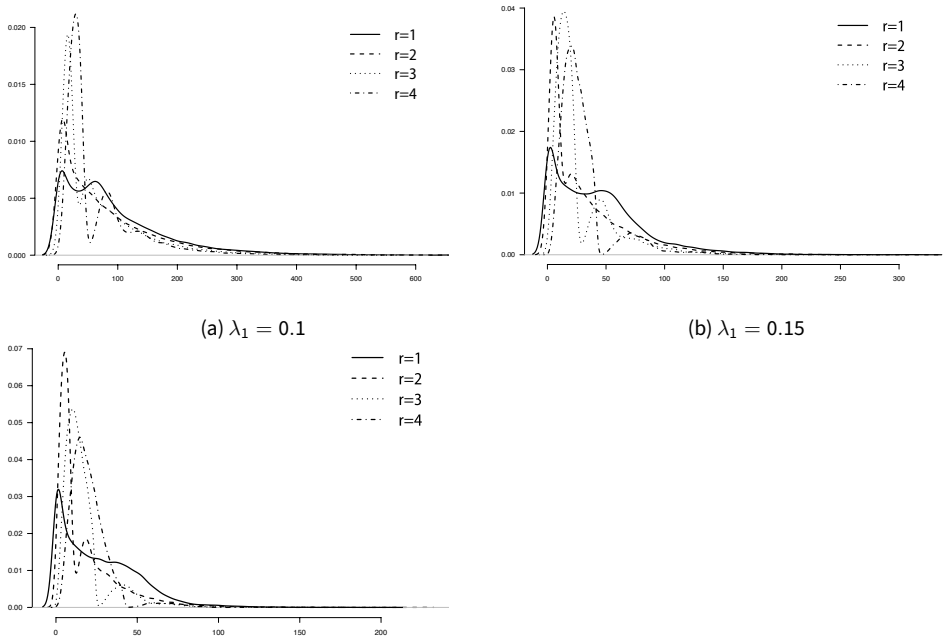


Figure 2.3: Lower-sided t_r-charts CED versus τ for $\lambda_0 = 0.05$, λ_1 = 0.2, $ARL_0 = 500$ and for different values of r.

rules for continuous data, the run length distribution is geometric and thus monotonic. In contrast, the figures show a multi-modal distribution behavior, most clearly for higher-orders. This might be explained by the fact that three (3) run rules are applied simultaneously with possibly different detection delay properties. For example, Test 3 signals rarely and at very long detection times; it does not allow the detection of trends, particularly when coupled with run rules with faster detection. The entire CED distribution can be therefore used for the screening of run rules. Conditional expected delay summary statistics are reported in Table 2.4 at different aggregation levels. Figure 2.3 shows the CED versus the change-point τ. The effect of τ is not observable within the selected range of values, while a strong effect of r is noticeable. Based on Table 2.4, for $\lambda_1 = 0.2$, the t_3-chart performs better than the other aggregation levels.



(a) $\lambda_1 = 0.1$

(b) $\lambda_1 = 0.15$

(c) $\lambda_1 = 0.2$

Figure 2.4: Lower-sided t_r-charts CED distributions for $\lambda_0 = 0.05$, $ALI_0 = 500$, $\tau = 0$ and for different values of λ_1 and r.

Figure 2.6 and 2.7 show the CED and the PSD against the ratio λ_1/λ_0, for different values of τ, r and d. The values of d are set at 60 and 120, corresponding to a window of 1 hour and 2 hours respectively to successfully detect a shift in λ, when λ is expressed in defects per minute. A moderate effect of the location of the change-point τ within the selected range of values is noticeable in this case as well, while the parameter d greatly affects the selection procedure.

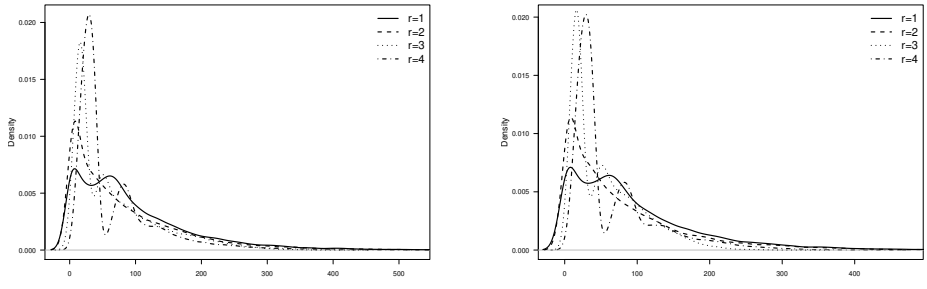
2.3 Conclusions

Cumulative quantity control charts represent an appropriate alternative to traditional monitoring strategies since they overcome the methodological challenges that standard control charts pose due to extreme parameter values. However, the asynchrony of process time and control chart time poses a severe limitation to the performance evaluation. Therefore, one should evaluate the control chart performance on the right time scale to allow fair comparison among different charts. This chapter derives

	$\tau = 0$				$\tau = 350$			
	r = 1	r = 2	r = 3	r = 4	r = 1	r = 2	r = 3	r = 4
Mean	26.5	18.1	**17.6**	20.6	25.9	18	**17.5**	20.7
Median	21.4	**8.3**	13.4	**18.3**	20.9	**8.5**	13.5	18.5
StDev	24	20.3	15.1	**12.4**	23.6	20	14.4	**12.5**
MAD	25.5	8.8	**7.9**	9.1	25.1	9.1	**8**	9.5
10%	0.5	2.6	5.5	8.7	0.5	2.7	5.5	8.7
90%	57.7	45	39.6	33.3	57.2	44.5	39.1	33.3

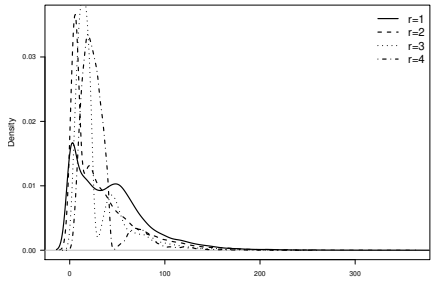
Table 2.4: Lower-sided t_r-charts CED summary statistics for $\lambda_0 = 0.05$, $\lambda_1 = 0.2$, $ARL_0 = 500$ and for different values of r and τ.

explicit formulas of the performance metrics of the t_r-chart and the CCC chart in terms of the inspection length. In addition, it quantifies the error's severity of the Blackwell-Girshick formula applied to situations where a strong independence assumption is required. The discussion leads to illustrating examples and proposes guidelines for the optimal choice of r for CCC_r and t_r- control charts to help practitioners choose the most appropriate aggregation level. Additionally, it extends the notions of the conditional expected delay and the probability of successful detection to aggregate data charts and thoroughly discusses the effects of their unique parameters. A remarkable phenomenon is the non-monotone behavior of the CED distribution; this aspect accentuates that relying simply on the expected value may be very misleading, particularly in industrial contexts.

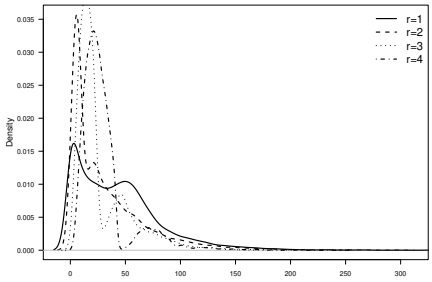


(a) $\tau = 175$, $\lambda_1 = 0.1$

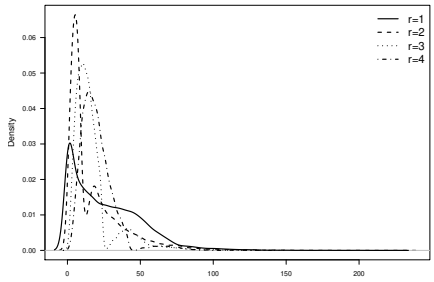
(b) $\tau = 350$, $\lambda_1 = 0.1$



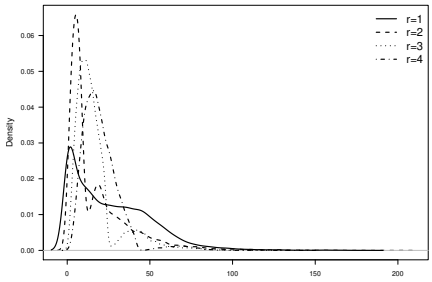
(c) $\tau = 175$, $\lambda_1 = 0.15$



(d) $\tau = 350$, $\lambda_1 = 0.15$

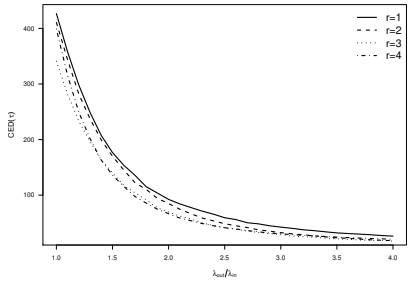


(e) $\tau = 175$, $\lambda_1 = 0.2$

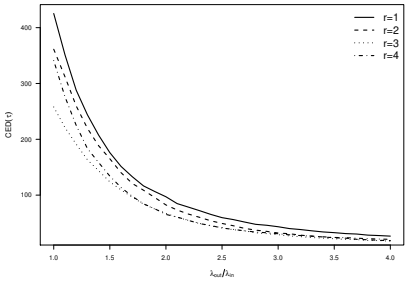


(f) $\tau = 350$, $\lambda_1 = 0.2$

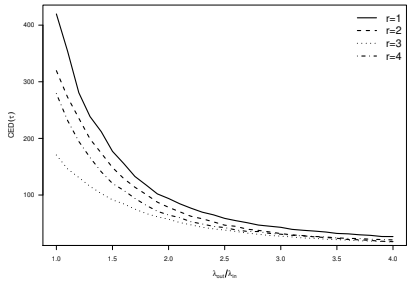
Figure 2.5: Lower-sided t_r-charts CED distributions for $\lambda_0 = 0.05$, $ARL_0 = 500$ and for different values of λ_1, r and τ.



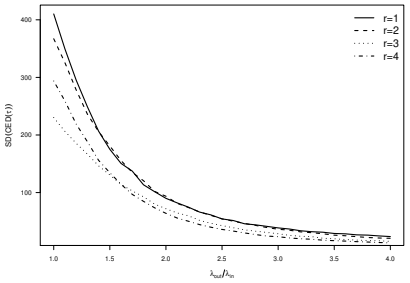
(a) CED, $\tau = 0$



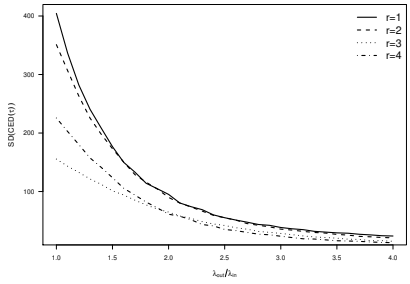
(b) CED, $\tau = 175$



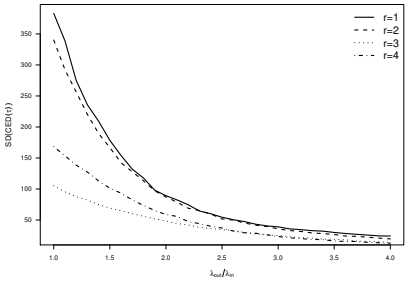
(c) CED, $\tau = 350$



(d) SD(CED), $\tau = 0$

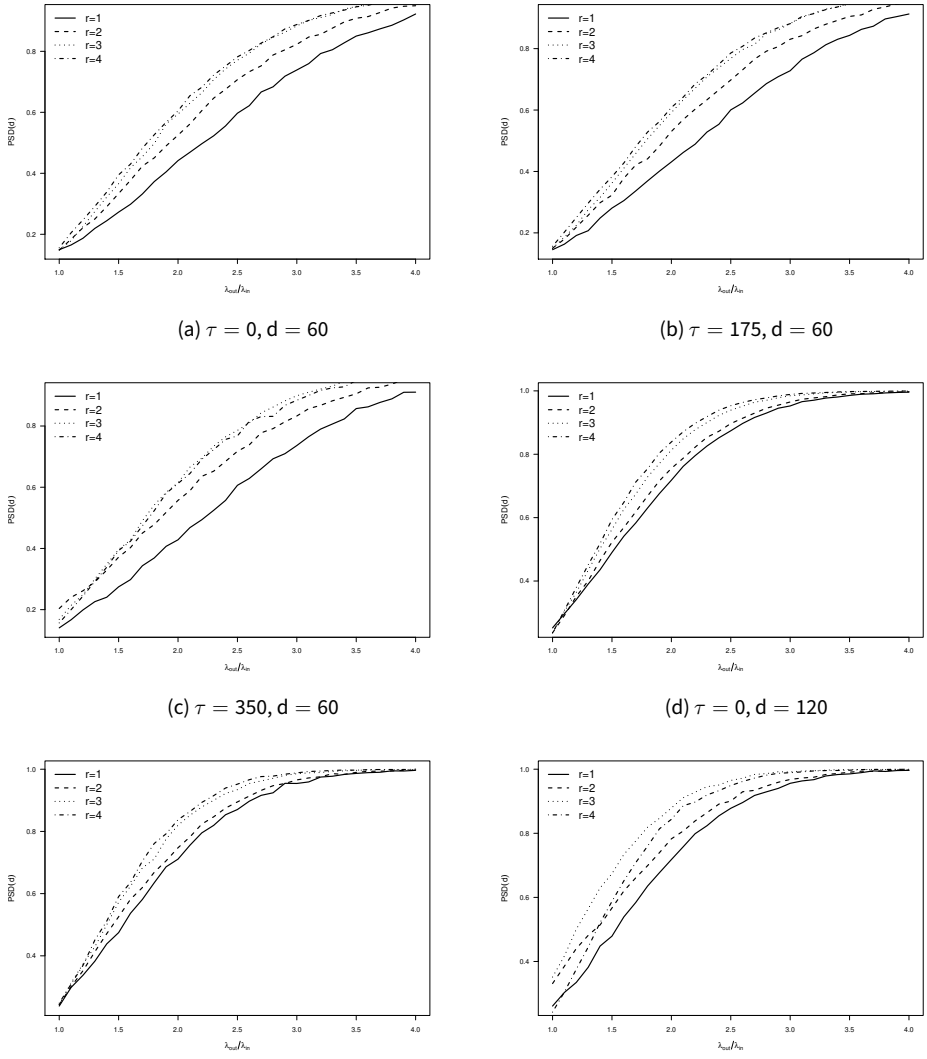


(e) SD(CED), $\tau = 175$



(f) SD(CED), $\tau = 350$

Figure 2.6: Lower-sided t_r-charts conditional expected delay for different values of r and τ.



(a) $\tau = 0$, d $= 60$

(b) $\tau = 175$, d $= 60$

(c) $\tau = 350$, d $= 60$

(d) $\tau = 0$, d $= 120$

(e) $\tau = 175$, d $= 120$

(f) $\tau = 350$, d $= 120$

Figure 2.7: Lower-sided t_r-chart probability of successful detection for different values of r, τ and d.

Chapter 3

Generalized Likelihood Ratio Monitoring Methods for High-Purity Processes

The successful implementation of monitoring procedures in the manufacturing industry requires tailored solutions to target specific out-of-control scenarios. Commonly, standard control charts are not suited for monitoring complex processes since they are built upon statistical models that are often too simplistic. Generalized likelihood ratio (GLR) control charts are a flexible tool for conveniently tailoring statistical monitoring systems to a specific application.

As described in Chapter 1, monitoring cumulative quantities between non-conforming items in the context of high-purity processes represents an adequate monitoring strategy to overcome the methodological challenges posed by standard approaches. Furthermore, as discussed in Chapter 2, the distinction between the process time and control chart time is fundamental when comparing different models.

This chapter discusses the design of generalized likelihood ratio-based charts for simple and composite hypotheses when applied to high-purity processes. Lee and Woodall (2018) point out and resolve a relevant technical issue in deriving generalized likelihood ratio control charts. Inspired by their work, this chapter proposes an improved chart for geometrically distributed data that removes the artificial bound introduced to address the statistic definition problem. Additionally, it stresses the importance of correctly estimating the model's interval-restricted parameters when using the maximum likelihood method. Finally, the control charts proposed are compared to their competitors by simulation

studies.

The rest of the chapter is organized as follows. Section 3.1 focuses on sequential change-point detection under simple hypotheses for cumulative quantity control charts. Section 3.2 proposes a new design for gamma GLR control charts and evaluates their performance. Section 3.3 extends the discussion to composite hypotheses models, followed by the performance comparison of simple and composite hypotheses schemes in Section 3.4. Finally, Section 3.5 contains the main conclusions of this chapter.

3.1 Sequential change-point detection under simple hypotheses

Statistical process control is concerned with the design and analysis of techniques for online quickest detection of a change in the state of a process, subjected to a tolerable limit on the false detection rate. The time instance at which the process changes is referred to as the change-point; estimating the change-point position and the magnitude of the change is crucial for post-signal diagnosis. Online detection conveniently fits in the framework of sequential hypothesis testing to properly assess its statistical performance. In the sequel, the random variables $Z_1, Z_2, \ldots$ are assumed to be independent and to follow the same probability distribution with possibly different parameters $\theta^{(i)}$. The most widely studied set of simple hypotheses represents an abrupt and unexpected persistent shift in the parameter as shown in (3.1), where τ is a fixed, unknown parameter, θ_0 is the fixed in-control parameter, and θ_1 can be an arbitrary value in a parameter set Θ.

$$\begin{aligned} H_0: &\quad \theta^{(i)} = \theta_0, \text{ for all } i \\ H_a(\tau): &\quad \theta^{(i)} = \begin{cases} \theta_0, \text{ for } i = 0, \ldots, \tau \\ \theta_1, \text{ for } i = \tau + 1, \ldots \end{cases} \end{aligned} \tag{3.1}$$

Once the change-point model has been chosen, the generalized log-likelihood ratio statistics can be defined for a specific density or probability mass function f_θ. For the set of hypotheses described in

(3.1), the generalized likelihood ratio can be written as follows

$$\begin{aligned}\log \Lambda_k(\theta_0, \theta_1; \mathbf{z}) &= \log \max_{0\le\tau\le k-1} \frac{\prod_{i=1}^{\tau} f_{\theta_0}(z_i) \sup_{\theta_1} \prod_{i=\tau+1}^{k} f_{\theta_1}(z_i)}{\prod_{i=1}^{k} f_{\theta_0}(x_i)} \\ &= \log \max_{0\le\tau\le k-1} \frac{\prod_{i=\tau+1}^{k} f_{\widehat{\theta_1}}(z_i)}{\prod_{i=\tau+1}^{k} f_{\theta_0}(z_i)}, \end{aligned} \tag{3.2}$$

where k represents each decision point. In this setting, $\tau = 0$ allows for the situation in which the change has already occurred before the monitoring has begun. By definition $\prod_{i=1}^{0}(\cdot) = 1$. As stated in the previous chapters, in the context of high-purity processes, parametric control charts which assume underlying gamma or negative binomial distributions represent the common choices when monitoring the cumulative quantities between events. Therefore, the generalized likelihood ratio statistics for both distributions are discussed and derived below.

Negative Binomial Consider the sequence of independent Bernoulli random variables, denoted by $Y_1, Y_2, \ldots$, representing the inspected items classified as conforming or non-conforming. Let $X_1, X_2, \ldots$ be the number of conforming items until the r^{th} non-conforming one is observed, then $X_1, X_2, \ldots$ follow the negative binomial distribution with parameters p and r, under the independence assumption and constancy of the parameter p over time. In the special case $r = 1$, e.g., the number of conforming items between two consecutive non-conforming ones, the random variables X_i follows the geometric distribution with parameter p. The likelihood ratio statistic for negative binomial distributed observations under the hypotheses in (3.1) is as follows

$$\log \Lambda_k(\theta_0, \theta_1; \mathbf{x}) = \max_{0\le\tau\le k-1} r(k-\tau) \left[\log\left(\frac{\widehat{\theta}_1}{\theta_0}\right) - \log\left(\frac{1-\widehat{\theta}_1}{1-\theta_0}\right) + \frac{1}{\widehat{\theta}_1} \log\left(\frac{1-\widehat{\theta}_1}{1-\theta_0}\right) \right]. \tag{3.3}$$

where $\widehat{\theta_1}$ is the maximum likelihood estimator (MLE) of the out-of-control parameter θ_1. One must note that this form is valid only for two-sided settings because the parameter estimate is unrestricted. For a short discussion on parameter restriction, refer to the end of this section. The statistic in (3.3) is undefined if there is at least one value of τ such that $\widehat{\theta}_1 = 1$, i.e., in the extreme case in which all

the observations from $\tau + 1$ to k are equal to r. KazemiNia et al. (2018) propose an artificial upper bound (θ_u) to ensure that $\theta_1 = 1$ does not occur. This artificial bound acts like a tuning parameter to the desired out-of-control scenario, a redundant specification in the design of likelihood ratio control charts. Following the method proposed in Lee and Woodall (2018), a slightly modified version of the negative binomial GLR chart is proposed to address the problem of the statistic being undefined if there are values of τ such $\widehat{\theta}_1 = 1$. Although the solution proposed by Lee and Woodall (2018) is correct for the negative binomial in its general form, the implementation for specific cases in that paper is imprecise. The problem is that the article gives formulas for the test statistic based on a condition on τ, which is impossible to check since the test statistic involves optimizing over τ. It follows from the definition of Λ_k in (3.2) that

$$\begin{aligned} \log \Lambda_k(\theta_0, \theta_1; \mathbf{x}) &= \log \max_{0 \le \tau \le k-1} \frac{\sup_{\theta_1} \prod_{i=\tau+1}^{k} \theta_1{}^r (1-\theta_1)^{x_i - r}}{\prod_{i=\tau+1}^{k} \theta_0^r (1-\theta_0)^{x_i - r}} \\ &= \max_{0 \le \tau \le k-1} \log \left(\frac{\sup_{\theta_1} \prod_{i=\tau+1}^{k} \theta_1^r (1-\theta_1)^{x_i - r}}{\prod_{i=\tau+1}^{k} \theta_0^r (1-\theta_0)^{x_i - r}} \right). \end{aligned} \tag{3.4}$$

If, for a given τ, it occurs that $x_j = r$ for $j \ge \tau + 1$, then the inner optimization in (3.4) can be performed as follows

$$\begin{aligned} \log \sup_{\theta_1} \left(\frac{\theta_1}{\theta_0} \right)^{r(k-\tau)} \left(\frac{1-\theta_1}{1-\theta_0} \right)^{-r(k-\tau)} \left(\frac{1-\theta_1}{1-\theta_0} \right)^{\sum_{j=\tau+1}^{k} x_j} &= \sup_{\theta_1} \log \left(\frac{\theta_1}{\theta_0} \right)^{r(k-\tau)} \\ &= \sup_{\theta_1} r(k-\tau) \log \left(\frac{\theta_1}{\theta_0} \right). \end{aligned} \tag{3.5}$$

The expression in (3.5) is maximized for $\theta_1 = 1$. If, for given τ, it does not hold that $x_j = r$ for $j \ge \tau + 1$, then $\widehat{\theta}_1 = \frac{r(k-\tau)}{\sum_{i=\tau+1}^{k} x_i}$. Therefore, the inner optimization of the GLR statistic, i.e., the optimization of the

value θ_1 for the alternative hypothesis for the negative binomial GLR statistic can be defined as

$$\log \sup_{\theta_1} \left(\frac{\prod_{i=\tau+1}^{k} \theta_1^r (1-\theta_1)^{x_i - r}}{\prod_{i=\tau+1}^{k} \theta_0^r (1-\theta_0)^{x_i - r}} \right) =$$

$$\begin{cases} -r(k-\tau)\log(\theta_0), \text{ if } x_i = r \text{ for } i = \tau+1, \ldots, k-1 \\ r(k-\tau)\left[\log\left(\frac{\widehat{\theta}_1}{\theta_0}\right) - \log\left(\frac{1-\widehat{\theta}_1}{1-\theta_0}\right) + \frac{1}{\widehat{\theta}_1}\log\left(\frac{1-\widehat{\theta}_1}{1-\theta_0}\right)\right], \text{otherwise.} \end{cases} \tag{3.6}$$

Gamma As previously stated, the dynamics of low-rate counts in continuous time could be described by a homogeneous Poisson process with rate $1/\theta$, so that the number of events up to time t follows a Poisson distribution with mean θt. Non-conforming items and their occurrence times are recorded. As described in (3.1), the inter-arrival times are used to determine whether an increase in the Poisson parameter θ has occurred from the in-control situation, which is assumed to be a known value θ_0, to an unknown out-of-control value, denoted by θ_1. The times between r events are then independent random variables Z_k, each of which is gamma distributed with rate $1/\theta$ and shape r. The likelihood function for the alternative hypothesis is

$$\mathcal{L}(\theta_0, \theta_1, r; \mathbf{z}) = \prod_{i=1}^{\tau} \frac{\theta_0^{-r}}{\Gamma(r)} z_i^{r-1} e^{-\frac{z_i}{\theta_0}} \times \prod_{i=\tau+1}^{k} \frac{\theta_1^{-r}}{\Gamma(r)} z_i^{r-1} e^{-\frac{z_i}{\theta_1}}, \tag{3.7}$$

where $\Gamma(r) = (r-1)!$ since the shape parameter $r \in \mathbb{N}$. For this application, the shape parameter r is fixed a priori and represents the order of the control chart. The unknown post-change mean is estimated using the maximum likelihood estimation method. Under the hypotheses defined in (3.1), the log-likelihood ratio statistic becomes as follows

$$\log \Psi_k(\theta_0, \theta_1, r; \mathbf{z}) = \log \max_{0 \le \tau \le k-1} \frac{\sup_{\theta_1} \prod_{i=\tau+1}^{k} \frac{1}{\theta_1^r \Gamma(r)} z_i^{r-1} e^{-\frac{z_i}{\theta_1}}}{\prod_{i=\tau+1}^{k} \frac{1}{\theta_0^r \Gamma(r)} z_i^{r-1} e^{-\frac{z_i}{\theta_0}}}. \tag{3.8}$$

For the two-sided setting the equation in (3.8) becomes

$$\log \Psi_k(\theta_0, \theta_1, r; \mathbf{z}) = \max_{0 \le \tau \le k-1} r(k-\tau)\left[\log\left(\frac{\theta_0}{\widehat{\theta}_1}\right) + \frac{\widehat{\theta}_1 - \theta_0}{\theta_0}\right], \tag{3.9}$$

where

$$\widehat{\theta_1} = \frac{\sum_{i=\tau+1}^{k} z_i}{r(k-\tau)}. \tag{3.10}$$

The MLE $\widehat{\theta_1}$ is undefined for $\sum_{i=\tau+1}^{k} z_i = 0$, i.e., all the observations from $\tau + 1$ to k are equal to zero. Nevertheless, this situation cannot occur in this specific context: the z_i are the inter-arrival times between r^{th} subsequent events, which is always a strictly positive quantity since for a Poisson distribution the probability or registering two or more events at the same time is zero. Therefore, $\log \Psi_k$ is always well-defined. Finally, the change-point τ can be estimated for the negative binomial and gamma distribution as follows

$$\widehat{\tau}_\Lambda = \underset{0\le\tau\le k-1}{\arg\max} \begin{cases} -r(k-\tau)\log(\theta_0), \text{ if } x_i = r \text{ for } i = \tau+1,\dots,k-1 \\ r(k-\tau)\left[\log\left(\frac{\widehat{\theta}_1}{\theta_0}\right) - \log\left(\frac{1-\widehat{\theta}_1}{1-\theta_0}\right) + \frac{1}{\widehat{\theta}_1}\log\left(\frac{1-\widehat{\theta}_1}{1-\theta_0}\right)\right], \text{otherwise} \end{cases} \tag{3.11}$$

and

$$\widehat{\tau}_\Psi = \underset{0\le\tau\le k-1}{\arg\max}\, r(k-\tau)\left[\log\left(\frac{\theta_0}{\widehat{\theta_1}}\right) + \frac{\widehat{\theta_1}-\theta_0}{\theta_0}\right], \tag{3.12}$$

respectively.

Parameter restriction in one-sided settings One must note that $\log \Lambda_k$ and $\log \Psi_k$ in (3.6) and (3.9) are correctly defined when aiming at detecting simultaneously upward and downward shifts from the in-control parameter, i.e., in two-sided settings, because the maximum likelihood estimator is unrestricted. This condition allows to further simplify the log-likelihood ratio statistics by direct substitution of the MLE. This is not valid for one-sided alternative hypotheses of the form $\theta_1 < \theta_0$ or $\theta_1 > \theta_0$, where the estimate is constrained, i.e., $\widehat{\theta_1} = \min(\theta_0, \text{MLE}_{\theta_1})$ and $\widehat{\theta_1} = \max(\theta_0, \text{MLE}_{\theta_1})$, respectively. Therefore, within each inner optimization, it is not guaranteed that $\widehat{\theta_1} = \text{MLE}_{\theta_1}$ and (3.6) and (3.9) would be incorrect. The inner optimization of the GLR statistic for negative binomial for

detecting only upward shifts in the mean parameter is

$$\log \sup_{\theta_1} = \begin{cases} -r(k-\tau)\log(\theta_0), \text{if } x_i = r \text{ for all } i = \tau+1,\dots,k-1 \\ (k-\tau)\left[\log\left(\frac{\widehat{\theta}_1}{\theta_0}\right) - \log\left(\frac{1-\widehat{\theta}_1}{1-\theta_0}\right)\right] + \sum_{i=\tau+1}^{k} x_i \log\left(\frac{1-\widehat{\theta}_1}{1-\theta_0}\right), \quad \text{otherwise.} \end{cases} \tag{3.13}$$

Similarly, the statistic for the gamma case one-sided upper case ($\log \Psi_k^u$) is as follows

$$\log \Psi_k^u(\theta_0, \theta_1, r; \mathbf{z}) = \max_{0\le\tau<k-1} r(k-\tau)\log\left(\frac{\theta_0}{\widehat{\theta}_1}\right) - \left(\frac{1}{\widehat{\theta}_1} - \frac{1}{\theta_0}\right)\sum_{i=\tau+1}^{k} z_i. \tag{3.14}$$

3.2 Simple hypotheses charts design and comparison

Once the practitioner chooses the most appropriate distribution and control chart type, the design of the control charts proposed in this study requires the determination of one parameter: the control limit h_{GLR}. One must note that GLR schemes can be computationally expensive, and window-limited charts have been proposed in the literature to overcome this limitation. In the following comparison, the effect of the window size is neglected. For window-limited schemes, the optimal window size m should also be specified. This issue is related to similar ones for MOSUM charts, which are window-limited versions of CUSUM charts (see e.g., Aue et al. (2012)). These charts were first introduced in the context of quality control by Bauer and Hackl (1978) and they are also known as Finite Moving Average Charts (Tartakovsky et al. (2015)). It is essential to stress that the importance of window-limited strategies is bound to permit the feasibility of thousands of simulation repetitions.

Control limit The selection of the control limit h_{GLR} should be determined based on the desired in-control performance. Woodall and Faltin (2019) provide an interesting overview of various aspects of choosing performance metrics that match industrial needs. In this chapter, the control chart design is solely based on the in-control average length of inspection (ALI) and the in-control process parameter. As discussed in Section 2.2, the average length of inspection is defined as the average sum of the monitored variable till the stopping time and it refers to the original data. In Table 3.1 control limits values are reported for different combinations of average length of inspection and in-control parameters. The control limits can be obtained via simulation. In practice, finding the control limit for

different combinations of desired performance and in-control parameter values can be further simplified. As empirically observed by Huang et al. (2012), the values of h_{GLR} are linearly related to the $\log_{10}(ALI_0 \times \theta_0)$. Figure 3.1 illustrates the linear relationship. The regression formula (3.15) is obtained

	θ_0		
$ALI_0 \times \theta_0$	0.001	0.005	0.01
20	2.77	2.77	2.78
30	3.26	3.23	3.28
40	3.67	3.61	3.64
50	3.98	3.98	3.91
60	4.31	4.31	4.14
70	4.71	4.71	4.32

Table 3.1: Control limit values (h_{GLR}) for several combinations of ALI_0 and θ_0.

from simulations and can be used to obtain either the h_{GLR} value directly or a good initial estimate for the iterative process of selecting the appropriate control limit.

$$h_{GLR} = -1.21 + 1.32 \log_{10}(ALI_0 \times \theta_0) \tag{3.15}$$

Non-homogeneous process simulations Several methods exist to generate pseudo-random observations from an NHPP. The algorithm described in the sequel is exact and based on the time scale transformation of a non-homogeneous Poisson process (NHPP). The main advantage of using a rigorous algorithm to generate NHPP is the possibility to exactly specify and fix the change-point time at the process scale τ_π without artificial corrections. It must be noted that if τ_π is fixed and therefore deterministic, while the control chart change-point is a random variable itself. The change-point models considered in this chapter originate from the assumption that the observations are generated from a non-homogeneous Poisson process. Non-homogeneous Poisson processes represent a more general form of the HPP, with the parameter $\lambda(t)$ as a function of time.

Definition 3.2.1. *Let* $\{N(t) : t \geq 0\}$ *be a counting process in the interval* $[0, \infty)$*. Then* $\{N(t) : t \geq 0\}$ *is a non-homogeneous Poisson process with intensity function* $\lambda(t)$ *if:*

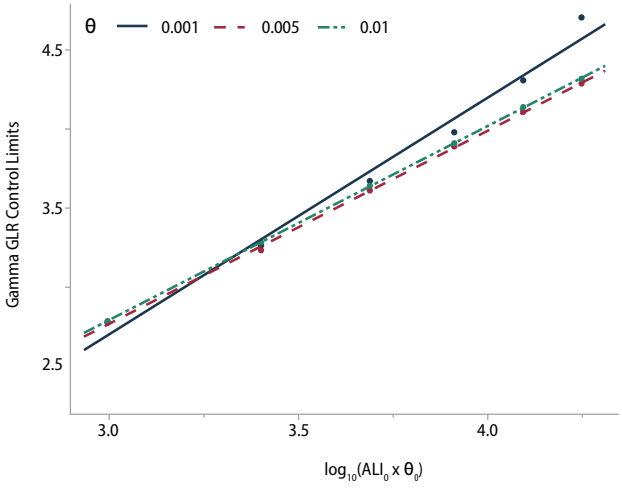


Figure 3.1: Linear relationship between the $\log_{10}(\text{ALI}_0 \times \theta_0)$ and h_{GLR} for different values of the out-of-control parameter θ_0.

- $N(0) = 0$;
- $N(t)$ *the process has independent increments;*
- $\lim_{s \downarrow 0} P[N(t+s) - N(t) \geq 2] = o(s)$;
- $\lim_{s \downarrow 0} P[N(t+s) - N(t) = 1] = \lambda(t)s + o(s)$.

The expected number of events by time t is equivalent to the cumulative intensity function at time t. The cumulative intensity function Λ completely specifies the Poisson process and it is as follows

$$E[N(t)] \equiv \Lambda(t) = \int_0^t \lambda(u)du. \tag{3.16}$$

For a non-homogeneous Poisson process with rate $\lambda(t)$, the number of arrivals in any interval is a Poisson random variable. However, its parameter can depend on the location of the interval. More specifically, NHPP processes describe more accurately change-point models in the high-purity processes field, but suffer from the constraining drawback of not inheriting the convenient proprieties of the simpler homogeneous case. For an NHPP process:

- The counts over intervals are still Poisson;

Algorithm 1 Time-scale transformation of NHPP

procedure
  $E_i^* \leftarrow \text{Exp}(1, n)$ ▷ Generate n inter-arrival times from a unit HPP
  $t_i^* \leftarrow t_{i-1}^* + E_i^*$ ▷ Get the unit HPP event times
  $t_i \leftarrow \Lambda^{-1}(t_i^*)$ ▷ Generate NHPP event times via the inverse of $\Lambda(t)$
  $E_i \leftarrow t_{i-1} - t_i$ ▷ Get the NHPP inter-arrival times
  return E_i

- The increments are still independent;
- Each increment has a Poisson distribution;
- The inter-arrival times are in general not independent, stationary nor exponentially distributed.

The latter is the main property on which the intuition behind the t-chart relies. Nevertheless, the design of the control chart is based on the null hypothesis assumption, under which the process can be modeled by a HPP with constant parameter λ. Deviations from this state will be considered out-of-control situations. For methods used to monitor non-homogeneous Poisson process, we refer to Richards et al. (2015). The following theorem can be used to simulate a non-homogeneous Poisson process.

Theorem 3.2.2. *(Çınlar (1975)) Let* $\{\Lambda(t), t \geq 0\}$ *be a positive, continuous, non-decreasing function. Then, the random variables* $T_1, T_2, \ldots$ *are event times corresponding to a NHPP with cumulative intensity function* $\Lambda(t)$ *is and only if* $\Lambda(T_1), \Lambda(T_2), \ldots$ *are the event times corresponding to an HPP with unit rate.*

The idea behind the algorithm is to use the relationship between a HPP with unit rate and NHPP via the inverse cumulative function as described in Theorem 3.2.2. The vector implementation of the algorithm is described in Algorithm 1.

Intensity functions for simple and composite hypotheses In this chapter, an unknown persistent shift from the mean target value is considered. In the sequel, θ is used to refer to the mean, the reciprocal of the intensity parameter λ. The intensity function $\theta(t)$ that describes how the mean changes over time is the following

$$\theta(t) = \theta_{ic} \mathbb{1}_{\{t \leq \tau_\pi\}} + \theta_{oc} \mathbb{1}_{\{t > \tau_\pi\}}, \tag{3.17}$$

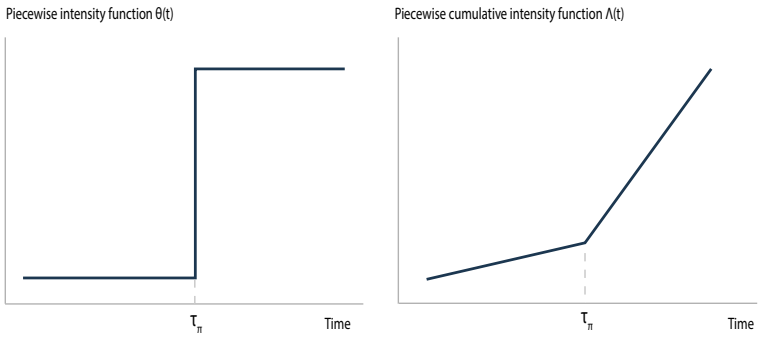


Figure 3.2: Piecewise intensity and cumulative intensity functions of a non-homogeneous Poisson process with change-point at τ_π.

where θ_{ic} is the in-control mean, θ_{oc} is the out-of-control mean and τ_π is the change-point in the process scale. Figure 3.2 shows the intensity function described in (3.17) and the corresponding cumulative intensity function. The explicit form of the inverse cumulative function of $\theta(\text{t})$ considered in this study is available; this simplifies the simulation implementation. If the explicit form of the inverse cumulative function is not available, the inversion can be solved numerically, but this approach can be time-consuming. For more complex intensity functions, the distribution of the inter-arrival times for a non-homogeneous Poisson process can be used as shown in the unpublished note by Yakovlev et al. (2005).

3.2.1 Performance evaluation and comparison

In this section, the gamma-based generalized likelihood control chart proposed in Section 3.2 is compared to the t-chart and other common memory-type methods introduced in Chapter 1. The comparison is performed based on the conditional expected delay metric (CED) considering a wide range of parameters for upper shifts in θ. This metric expresses the delay from the first opportunity to detect a shift after the change-point τ_π, conditional on not having raised a false alarm before. To allow a fair comparison, the control limits of each chart are calculated based on a fixed in-control average length of inspection (ALI_0), and the simulations are designed by fixing the change-point in the process scale τ_π. The performance metric values are calculated by simulations for all cases. It must be

noted that the simulation and computation of the generalized likelihood ratio statistics can be time-consuming. An optimized algorithm is developed to minimize the computational complexity of the simulation process. The algorithm receives as input the process simulation matrix and the vector of indexes which indicate the first index after the change, which is a random variable, and returns a matrix with the generalized likelihood statistics. The matrix is the result of two subsequent optimizations. Each k represents a decision point (i.e., a control chart point). From the process standpoint, at least a defective item has been registered at time k. At each decision point k, the GLR statistic is calculated and plotted in the control chart and represents the variable to be monitored. The assumption on which the simulation algorithm relies is that the actual change happens between two consecutive k, $\tau_\pi \in [\tau - 1, \tau]$. The statistic $\Lambda_k(\tau; x_i)$ is the result of the optimization over τ with $0 \leq \tau \leq k - 1$, where τ is the change-point on the control chart scale. Note that the statistic can be calculated only at the times in which a defective item has been registered, meaning only at the control chart decision points. Figure 3.3 illustrates the GLR algorithm built upon the simple hypotheses change-point model.

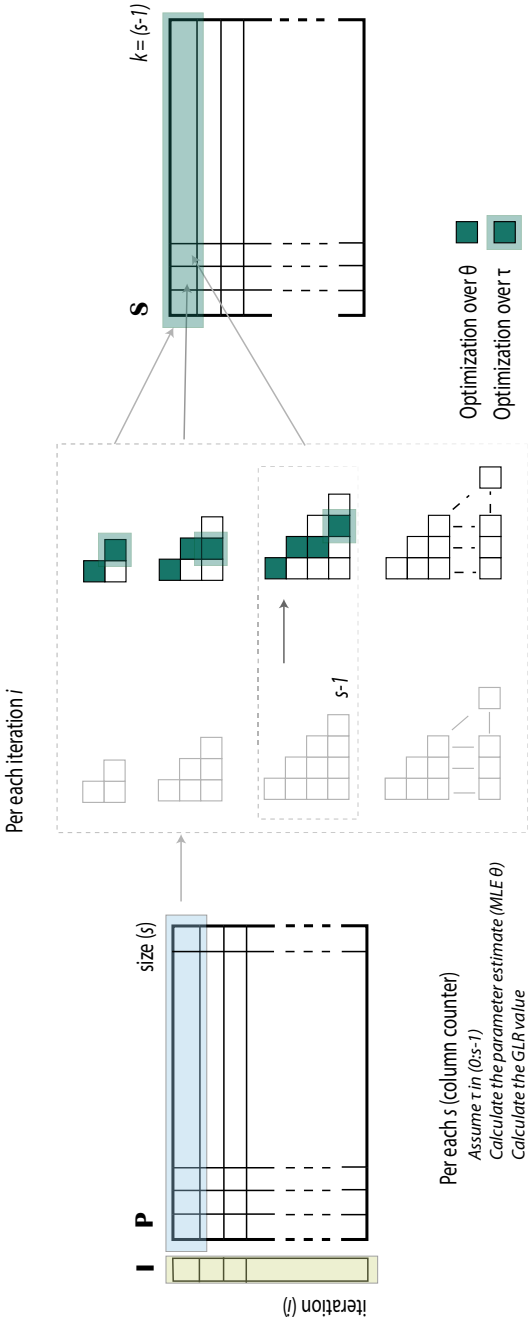


Figure 3.3: Schematic description of the generalized likelihood ratio algorithm for change-point models based on simple hypotheses.

Table 3.2 reports the CED$\times\theta_0$ values of the gamma GLR charts and their competitors, for a wide range of upward shifts from the in-control parameter $\theta_0 = 0.001$. The control limits for each chart are calculated based on the ALI_0 set at $60/\theta_0$. The simulation settings chosen are similar to the ones observed in real case-studies at Dow. Each iteration is set to cover a total time elapsed of $200/\theta_0$ time units, while the change-point τ_π is fixed at $15/\theta_0$ time units. The chosen length of the simulation ensures a good tail estimation of the performance metrics' distributions. The case r $= 1$ is considered for simplicity, but this evaluation can be easily extended to higher-order charts.

	t-chart	EWMA		CUSUM		Gamma GLR ($\ln \Psi_k$)
	r = 1	w = 0.3	w = 0.7	$\theta_1 = 0.002$	$\theta_1 = 0.02$	
h	16.08	0.369	0.129	2.34	0.155	4.31
θ_1						
0.0025	7.417	3.340	4.897	**3.103**	5.816	3.574
0.005	2.393	0.850	0.993	0.852	1.331	**0.759**
0.0075	1.165	0.445	0.818	0.437	0.593	**0.363**
0.01	0.623	0.297	0.300	0.328	0.332	**0.218**
0.02	0.159	0.128	0.101	0.161	0.114	**0.071**

Table 3.2: CED$\times\theta_0$ values for detecting upward shifts in θ, with $\theta_0 = 0.001$, $\text{ALI}_0 = 60/\theta_0$, $\tau_\pi = 15/\theta_0$ and r = 1.

It is shown that the generalized likelihood ratio control chart outperforms its competitors for almost all upwards parameter shifts, except for the CUSUM chart designed for $\theta_1 = 0.002$, which exhibits better performance than the proposed chart in detecting a very moderate shift. Two parameters have been chosen for the EWMA chart to cover a reasonable range of smoothing with more or less importance to the previous observations. For the CUSUM charts, the two out-of-control shifts chosen are the lowest and the highest, respectively, considered in the study (i.e., $\theta_1 = 0.002$ and $\theta_1 = 0.02$). Given the overall performance, the gamma GLR chart should be used when the goal of monitoring is to detect a wide range of parameter shifts.

One must note that each k represents a decision point. From the process point of view, at least one defective item is recorded at time k. The GLR statistic applied to cumulative quantities is evaluated only when a defective item is recorded, meaning only at the control chart times. As a result, detecting improvements in the process might be delayed.

3.3 Sequential change-point detection under composite hypotheses

The design and analysis of monitoring techniques based on sequential change-point detection offer a flexible and advantageous framework for detecting a wider range of shifts, while being simple to implement and interpret, as discussed in Section 3. The most common deviation type studied in statistical process control is represented by an unexpected and sustained shift from the in-control conditions. Nevertheless, most industrial applications require monitoring procedures that are tailored based on more appropriate statistical models and more effective at detecting realistic out-of-control scenarios. Composite hypotheses change-point models enhance GLR control charts by adding a new layer of flexibility. A composite hypothesis is any non-simple hypothesis. Simple hypotheses set a hard limit to a decision between one of two possible states of nature, while composite hypotheses cover a set of values from the parameter space. Composite hypotheses can help in representing the uncertainty inherent to manufacturing processes by specifying a collection of possible models instead of restricting the design based on a single parameter value. As Tartakovsky et al. (2015) report, there are two methods for dealing with composite hypothesis scenarios. The first one consists in weighting the likelihood

ratio statistics with respect to all possible values of the parameter θ using a weighting function, while the second approach consists in replacing the unknown parameters with their maximum likelihood estimates. In this chapter, the latter approach is used, with a strong emphasis on the correct definition of the maximum likelihood estimation under restriction. Composite hypotheses-based GLR models are used in conjunction with monitoring strategies based on cumulative quantities, which represents the standard choice in the high-purity processes framework. The following discussion considers two shift modes for that are practically relevant in the chemical manufacturing industry: the indifference interval and the epidemic shift model. The intensity function over time of the two models considered is shown in Figure 3.4.

Indifference interval model In more practical and realistic scenarios, one might be interested in detecting a change from a target value by allowing a margin, i.e., so that the process mean remains within a certain specified tolerance interval. Woodall and Faltin (2019) discuss in detail the practical importance of this model in industry. The set of hypotheses describing the model is reported in (3.18). It must be noted that the in-control parameter θ is not known beforehand, nor it is fixed.

$$\begin{aligned} \mathsf{H}_0 &: \quad \theta_{\mathsf{i}} = \theta \in [\theta_0 - \delta, \theta_0 + \delta], \text{for all i} \\ \mathsf{H}_{\mathsf{a}}(\tau) &: \begin{cases} \theta_{\mathsf{i}} = \eta \in [\theta_0 - \delta, \theta_0 + \delta], \text{for i} = 0, \dots, \tau \\ \theta_{\mathsf{i}} = \xi \notin [\theta_0 - \delta, \theta_0 + \delta], \text{for i} = \tau + 1, \dots, \end{cases} \end{aligned} \tag{3.18}$$

where $\delta > 0$ and θ_0 are fixed and chosen at the discretion of the business. A similar set of hypothesis can be found in Di Bucchianico et al. (2004). Assuming the change-point model in (3.18), the likelihood ratio function is given by the following

$$\Lambda_{\mathsf{k}}(\theta, \eta, \xi; \mathbf{x}) = \max_{0 \le \tau \le \mathsf{k}-1} \frac{\sup_{\eta \in \Theta_0} \prod_{\mathsf{i}=1}^{\tau} \mathsf{f}_{\eta}(\mathsf{x}_{\mathsf{i}}) \sup_{\xi \notin \Theta_0} \prod_{\mathsf{i}=\tau+1}^{\mathsf{k}} \mathsf{f}_{\xi}(\mathsf{x}_{\mathsf{i}})}{\sup_{\theta \in \Theta_0} \prod_{\mathsf{i}=1}^{\mathsf{k}} \mathsf{f}_{\theta}(\mathsf{x}_{\mathsf{i}})} = \max_{0 \le \tau \le \mathsf{k}-1} \frac{\prod_{\mathsf{i}=1}^{\tau} \mathsf{f}_{\widehat{\eta}}(\mathsf{x}_{\mathsf{i}}) \prod_{\mathsf{i}=\tau+1}^{\mathsf{k}} \mathsf{f}_{\widehat{\xi}}(\mathsf{x}_{\mathsf{i}})}{\prod_{\mathsf{i}=1}^{\mathsf{k}} \mathsf{f}_{\widehat{\theta}}(\mathsf{x}_{\mathsf{i}})}, \tag{3.19}$$

where $\widehat{\theta}$, $\widehat{\eta}$ and $\widehat{\xi}$ are the maximum likelihood estimates of the underlying distribution $\mathsf{f}(\cdot)$ based on the set of realizations x_{i}, under the null and the alternative hypotheses and given the interval constraints with $\Theta_0 = [\theta_0 - \delta, \theta_0 + \delta]$. It must be noted that when the null hypothesis' parameter need to be

estimated, the numerator and denominator do not cancel out since $\widehat{\theta} \neq \widehat{\eta}$. In this set up, $\tau = 0$ allows for the scenario in which the shift had happened before the monitoring started. Chang and Fricker (1999) studied a similar situation considering a persistent monotone threshold crossing for one-sided settings.

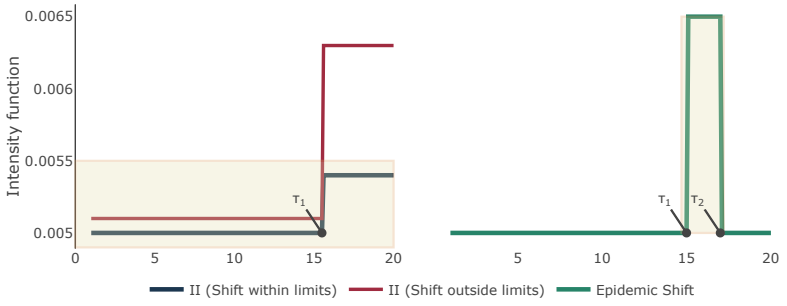


Figure 3.4: Intensity functions of the indifference interval and epidemic shift change-point models.

Epidemic shift model The epidemic shift model, also known as transient shift (Reynolds Jr and Lou (2010)), represents another relevant monitoring scenario in the chemical industry. The model describes a temporary change in the parameters, a typical situation where a feedback controller is active. The controller tries to compensate for changes in the processes, giving rise to a temporary deviation and bringing the process back to normal by adjusting several process variables. However, identifying the location of the change and investigating the root cause of the parameter shift can be of importance in many applications. Ramanayake and Gupta (2004) propose a generalized test for the exponential distribution family assuming the epidemic shift model. A set of hypotheses for the epidemic shift can be written follows

$$
\begin{aligned}
\mathsf{H}_0 : \quad & \theta_{\mathsf{i}} = \theta \leq \theta_{\mathsf{m}}, \text{all i} \\
\mathsf{H}_{\mathsf{a}}(\tau_1, \tau_2) : \quad & \begin{cases} \theta_{\mathsf{i}} = \eta \leq \theta_{\mathsf{m}}, \text{for i} = 0, \ldots, \tau_1 \\ \theta_{\mathsf{i}} = \xi > \theta_{\mathsf{m}}, \text{for i} = \tau_1 + 1, \ldots, \tau_2 \\ \theta_{\mathsf{i}} = \varphi \leq \theta_{\mathsf{m}}, \text{for i} = \tau_2 + 1, \ldots \end{cases}
\end{aligned}
\tag{3.20}
$$

where θ_m is a fixed parameter used to calculate and design the control chart to satisfy performance conditions at the boundary. This model represents a more generic version of the classical change-point problem since it involves multiple change-points with $0 \leq \tau_1 < \tau_2$. The likelihood ratio function for the epidemic shift model is as follows

$$\Psi_k(\theta, \eta, \xi, \varphi; \mathbf{x}) = \max_{\substack{0 \leq \tau_1 \leq \tau_2 - 1, \\ \tau_1 + 1 \leq \tau_2 \leq k-1}} \frac{\sup_{\eta \in \Theta_0} \prod_{i=1}^{\tau_1} f_\eta(x_i) \sup_{\xi \notin \Theta_0} \prod_{i=\tau_1+1}^{\tau_2} f_\xi(x_i) \sup_{\varphi \in \Theta_0} \prod_{i=\tau_2+1}^{k} f_\varphi(x_i)}{\sup_{\theta \in \Theta_0} \prod_{i=1}^{k} f_\theta(x_i)}, \tag{3.21}$$

where $\Theta_0 = [0, \theta_m]$.

In the sequel, the generalized likelihood ratio statistics for the proposed change-point models, assuming that the observations are either negative binomial or gamma distributed, are derived.

Negative Binomial Consider a sequence of inspected items classified as conforming or nonconforming. The mass probability function of the negative binomial distribution, as reported in Section, 1.1 is the following

$$f(r, \theta; y) \equiv P(Y = y) = \binom{y-1}{r-1} \theta^r (1-\theta)^{y-r}. \tag{3.22}$$

The parameter θ represents the probability of occurrence of non-conforming items. The statistic to be monitored under the composite hypotheses scenarios introduced above are obtained by substituting the generic density function $f(\cdot)$ in (3.19) and (3.21) with (3.22). By substituting the interval-restricted MLEs, the log-likelihood ratio statistic for the indifference interval model becomes

$$\log \Lambda_k^{NB}(\theta, \eta, \xi; \mathbf{y}) = \max_{0 \leq \tau \leq k-1} r(\tau \log(\widehat{\eta}) + (k - \tau) \log(\widehat{\xi}) - k \log(\widehat{\theta})). \tag{3.23}$$

Analogously, for the epidemic shift model the log-likelihood ratio statistic becomes

$$\begin{aligned} \log \Psi_k^{NB}(\theta, \eta, \xi, \phi; \mathbf{y}) &= \max_{\substack{0 \leq \tau_1 \leq \tau_2 - 1, \\ \tau_1 + 1 \leq \tau_2 \leq k-1}} r(\tau_1 \log(\widehat{\eta}) + (\tau_2 - \tau_1) \log(\widehat{\xi}) + (k - \tau_2) \log(\widehat{\varphi})) \\ &- r(k \log(\widehat{\theta})). \end{aligned} \tag{3.24}$$

Both statistics are undefined for $\widehat{\xi} = 0$ and/or $\widehat{\theta} = 0$. It is possible to completely define the statistic considering these extreme cases without introducing artificial bounds on the estimates similarly as presented in Section 3.1.

Gamma As discussed in the introductory chapter, the times between r events are independent random variables X_k, each distributed as a gamma distribution with rate $1/\theta$ and shape r, with the following probability density function

$$f(r, \theta; x) = \frac{1}{\theta^r \, \Gamma(r)} x^{r-1} e^{-\frac{x}{\theta}}. \tag{3.25}$$

Similarly to the negative binomial case, the two log-likelihood ratio statistics for the indifference interval and epidemic shift models described in (3.19) and (3.20), assuming gamma distributed observations, become

$$\log \Lambda_k^{\Gamma}(\theta, \eta, \xi; \mathbf{x}) = \max_{0 \le \tau \le k-1} r(k \log(\widehat{\theta}) - \tau \log(\widehat{\eta}) - (k - \tau) \log(\widehat{\xi})) \tag{3.26}$$

and

$$\begin{aligned} \log \Psi_k^{\Gamma}(\theta, \eta, \xi; \mathbf{x}) &= \max_{\substack{0 \le \tau_1 \le \tau_2 - 1, \\ \tau_1 + 1 \le \tau_2 \le k-1}} r(k \log(\widehat{\theta}) - \tau_1 \log(\widehat{\eta}) - (\tau_2 - \tau_1) \log(\widehat{\xi})) \\ &- r((k - \tau_2) \log(\widehat{\varphi})), \end{aligned} \tag{3.27}$$

respectively.

Estimation of the parameters The parameters in the composite hypotheses models are subjected to interval restrictions by definition. For both negative binomial and gamma distributions, for the indifference interval model, the MLEs of θ, η and ξ are the following

$$\widehat{\theta}(\tau, k; \mathbf{x}) = \min\left(\max\left(\widehat{\theta}^*, \widehat{\theta}_0 - \delta\right), \widehat{\theta}_0 + \delta\right); \tag{3.28}$$

$$\widehat{\eta}(\tau, k; \mathbf{x}) = \min\left(\max\left(\widehat{\eta}^*, \widehat{\theta}_0 - \delta\right), \widehat{\theta}_0 + \delta\right); \tag{3.29}$$

$$\widehat{\xi}(\tau, \mathsf{k}; \mathbf{x}) = \begin{cases} \widehat{\xi}^* & \text{if } \widehat{\xi}^* \notin \Theta_0 \\ \widehat{\theta_0} - \delta & \text{if } \mathcal{L}(\xi = \widehat{\theta_0} - \delta | \widehat{\eta}, \tau, \mathsf{k}, \mathbf{x}) > \mathcal{L}(\xi = \widehat{\theta_0} + \delta | \widehat{\eta}, \tau, \mathsf{k}, \mathbf{x}) \\ \widehat{\theta_0} + \delta & \text{otherwise,} \end{cases} \tag{3.30}$$

respectively, where $\widehat{\eta}^*$, $\widehat{\xi}^*$ and $\widehat{\theta}^*$ are the unrestricted MLE, considering that the likelihood function $\mathcal{L}(\eta, \xi, \theta, \tau, \mathsf{r}; \mathbf{x})$ has one maximum with respect to the individual parameters. Figure 3.5 shows how the out-of-control estimate $\hat{\xi}$ is moved at the boundaries of the indifference interval when the unrestricted maximum likelihood estimate is contained in it. Similarly to the previous change-point model,

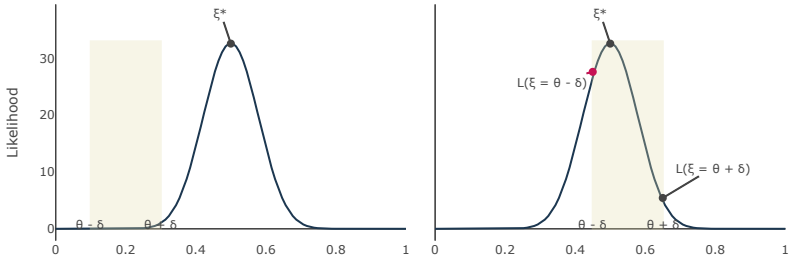


Figure 3.5: Estimate's correction of the alternative hypothesis parameter $\hat{\xi}$ under the indifference interval model.

the parameters for the epidemic shift can be estimated by imposing interval constraints on the unrestricted MLEs. In this case, since the objective is to detect a one-directional shift, the parameter restrictions for the negative binomial and gamma cases are slightly different. This is because the estimated parameters of the latter are expressed in terms of means rather than in rates. In Table 3.3 the parameter restricted are reported for detecting upward shifts in the rate or, equivalently, downwards shifts in mean. The parameters $\widehat{\eta}^*$, $\widehat{\xi}^*$, $\widehat{\varphi}^*$ and $\widehat{\theta}^*$ are the unrestricted MLEs of the assumed underlying distribution, following the parametrization choice in (3.22) and (3.25). Restricted intervals are not exclusive to composite hypotheses; GLR one-sided control charts assuming an unknown permanent shift model described in Section 3.1 are also subjected to interval restriction but this aspect is often neglected.

Change-point location As stated in the previous sections, one of the main advantages of change-point models is the possibility to estimate their location by taking the arguments that maximize the

Parameter	Negative Binomial	Gamma
$\widehat{\theta}(\tau;\mathbf{x})$	$\min\left(\widehat{\theta}^*,\widehat{\theta_{\mathsf{m}}}\right)$	$\max\left(\widehat{\theta}^*,\widehat{\theta_{\mathsf{m}}}\right)$
$\widehat{\eta}(\tau,\mathsf{k};\mathbf{x})$	$\min\left(\widehat{\eta}^*,\widehat{\theta_{\mathsf{m}}}\right)$	$\max\left(\widehat{\eta}^*,\widehat{\theta_{\mathsf{m}}}\right)$
$\widehat{\xi}(\tau,\mathsf{k};\mathbf{x})$	$\max\left(\widehat{\xi}^*,\widehat{\theta_{\mathsf{m}}}\right)$	$\min\left(\widehat{\xi}^*,\widehat{\theta_{\mathsf{m}}}\right)$
$\widehat{\varphi}(\tau;\mathbf{x})$	$\min\left(\widehat{\varphi}^*,\widehat{\theta_{\mathsf{m}}}\right)$	$\max\left(\widehat{\varphi}^*,\widehat{\theta_{\mathsf{m}}}\right)$

Table 3.3: Parameters' estimation for the the epidemic shift model.

log-likelihood ratios. For example, for the model based on the gamma distribution, the change-point position τ under the indifference interval alternative hypotheses is

$$\widehat{\tau} \;=\; \arg\max_{0\le\tau<\mathsf{k}-1} \mathsf{r}(\mathsf{k}\log(\widehat{\theta}) - \tau\log(\widehat{\eta}) - (\mathsf{k}-\tau)\log(\widehat{\xi})). \tag{3.31}$$

In a similar fashion, the two change-point positions τ_1 and τ_2 under the epidemic shift model can be estimated as follows

$$\begin{aligned}(\widehat{\tau}_1,\widehat{\tau}_2) \;&=\; \arg\max_{\substack{0\le\tau_1\le\tau_2-1,\\ \tau_1+1\le\tau_2\le\mathsf{k}-1}} \mathsf{r}(\mathsf{k}\log(\widehat{\theta}) - \tau_1\log(\widehat{\eta}) - (\tau_2-\tau_1)\log(\widehat{\xi}))\\ &-\; (\mathsf{k}-\tau_2)\log(\widehat{\varphi})).\end{aligned} \tag{3.32}$$

3.4 Composite hypotheses charts design and evaluation

This section discusses procedures and limitations in implementing the composite change-point models proposed in Section 3.3.

3.4.1 Implementation of the indifference interval monitoring scheme

When the process is in-control and no shift outside the indifference interval has occurred, the result of the outer maximization on the change-point expected is $\tau = \mathsf{k}-1$, which means that the hypothesized shift happened at the current observation k. In this case, $\log\Lambda_{\mathsf{k}} \approx 0$. For the indifference interval model described in (3.18), the outer maximization on the change-point might give as result $\tau = 0$ or $\tau = \mathsf{k}-1$ depending on the estimate of the out-of-control parameter ξ. In particular, the change-point maximization depends on which side the estimate of ξ is forced to, whether to the upper or

lower boundary. This effect is noticeable by looking at the extreme cases when $\tau = 0$ and $\tau = k - 1$. For $\tau = 0$, the likelihood in (3.19) becomes

$$k \log(\widehat{\theta}) - k \log(\widehat{\xi}) = k(\log(\widehat{\theta}) - \log(\widehat{\xi})), \tag{3.33}$$

and for $\tau = k - 1$, it becomes

$$k \log(\widehat{\theta}) - (k-1) \log(\widehat{\eta}) - \log(\widehat{\xi}) \approx \log(\widehat{\theta}) - \log(\widehat{\xi}). \tag{3.34}$$

When $\widehat{\xi}$ is pushed to the upper interval limit $\theta_0 + \delta$, the difference $\log(\widehat{\theta}) - \log(\widehat{\xi})$ is always negative, and vice versa. If $\log \Lambda < 0$, the maximization over τ will result in $\tau = k - 1$ since $\log(\widehat{\theta}) - \log(\widehat{\xi}) > k(\log(\widehat{\theta}) - \log(\widehat{\xi}))$. On the other hand, if $\log \Lambda > 0$, then the result of the maximization is $\tau \approx 0$ since $\log(\widehat{\theta}) - \log(\widehat{\xi}) < k(\log(\widehat{\theta}) - \log(\widehat{\xi}))$. The scaling factor k in (3.33) leads to instability of the statistics during in-control conditions. This is valid for the gamma distribution as well as for the negative binomial distribution, with the direction of the shift reversed. This effect gets stronger as the margin allowed increases, mainly due to the restrictions imposed on the maximum likelihood estimates. Additionally, the resulting instability is emphasized at the beginning of the monitoring, because of the high sensitivity of the estimates to the set of realizations of the random variable. This consequence is due to the skewness of the distribution and high variability, characteristic of low defect rate processes. This shortcoming might lead to quick departures while being in the in-control state. An alternative approach is to split the indifference model into two one-sided models. As an example, a one-sided case of (3.18) can be written as follows

$$\begin{aligned} H_0 &: \quad \theta_i = \theta < \theta_0 + \delta, \text{for all i} \\ H_a(\tau) &: \begin{cases} \theta_i = \eta < \theta_0 + \delta, \text{for i} = 0, \ldots, \tau \\ \theta_i = \xi \geq \theta_0 + \delta, \text{for i} = \tau + 1, \ldots \end{cases} \end{aligned} \tag{3.35}$$

The one-sided case solves the issue of instability due to parameter restriction. Figure 3.6 shows the one-sided indifference interval statistics over time for a series of out-of-control rates. The in-control parameter θ_0 and the critical value δ are both set at 0.01. The monitoring scheme only signals when the out-of-control rate exceeds the allowed margin and ignores shifts that are not of practical importance.

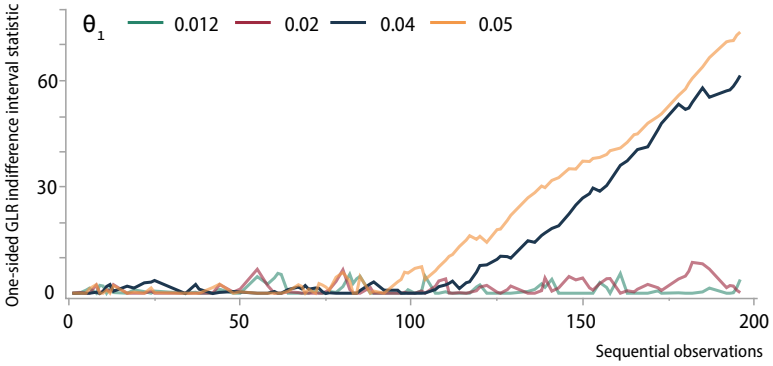


Figure 3.6: One-sided GLR indifference interval statistics over time for different shifts.

3.4.2 Implementation of the epidemic shift monitoring scheme

The design of the control chart proposed for the epidemic shift model requires the determination of two parameters: the control limit h and the window limit size m. The latter is not considered in the following discussion. This section evaluates the performance of the control chart based on the epidemic shift change-point model for the gamma distribution.

Control limits The selection of the control limit h is determined based on the desired in-control ALI and the in-control process parameter. The average inspection length refers to the original data. The appropriate h is obtained via Monte Carlo simulations. The control limits under an unknown permanent shift and the epidemic shift models are denoted as h and h_{ES}, respectively. The control limits for different combinations of performance and parameter values can be approximated via linear interpolation since it is linearly related to the natural logarithm of ALI_0 as shown in Section 3.2. It must be noted that the framework for high-purity processes relies on the assumption of constancy of the parameters over time (i.e., $\theta(t) = \theta$ and $\psi(t) = \psi$). While the process is assumed to be non-homogeneous, the assumption is valid for the null hypothesis, described by a homogeneous Poisson process.

	h			h_{ES}		
$ALI_m \times \theta_m$	0.001	0.005	0.01	0.001	0.005	0.01
20	2.81	2.78	2.80	3.13	2.94	2.96
30	3.30	3.29	3.31	3.71	3.57	3.59
40	3.68	3.67	3.70	4.12	4.03	4.05
50	4.02	4.02	4.04	4.50	4.39	4.42
60	4.36	4.38	4.38	4.88	4.78	4.79

Table 3.4: Control limits values (h, h_{ES}) for several combinations of ALI_0 and θ_m for $r = 1$.

3.4.3 Performance evaluation and comparison

This section compares the monitoring scheme proposed in (3.25) to the simple-hypotheses generalized likelihood ratio approach to detect an unknown, persistent shift, denoted by $\log \Phi_k^{\Gamma}$. It has been already shown in Section 3.2 that the simple-hypothesis GLR exhibits superior performance compared to other memory-type methods. The conditional expected delay (CED) metric is helpful to compare the charts in terms of the expected delay from the first opportunity to detect a shift, provided that the chart did not raise any false alarm before the change-point. Because of the conditional delay's skewed distribution, the standard deviation of the conditional expected delay SD(CED) is also reported. Both metrics can be used to select the chart order, as discussed in Section 2.2. For the epidemic shift, the function describing the intensity of the parameter $\psi(t)$ can be written as follows

$$\psi(t) = \psi_{ic}\mathbb{1}_{\{t\leq\tau_\pi\}} + \psi_{oc}\mathbb{1}_{\{\tau_{\pi_1}<t\leq\tau_{\pi_2}\}} + \psi_{ic}\mathbb{1}_{\{t>\tau_{\pi_2}\}}, \tag{3.36}$$

and the underlying process can be simulated using the algorithm described in Algorithm 1 in Section 3.1. The explicit form of the inverse cumulative function of $\psi(t)$ has an analytical form, simplifying the implementation procedure. The performance metrics for detecting upward shifts from the in-control situation are shown in Table 3.5. The results were obtained using the epidemic shift model in short-run settings, i.e., assuming that the in-control parameter is unknown and needs to be estimated. The simulation study consists of thousands of iterations of variable length. The simulation length is determined based on the desired expected total time elapsed from the beginning to the end of each run and it depends on the in-control and out-of-control parameters. The size is large enough to capture the behavior of the metrics away from the expected mean value and from the in-control average length

of inspection selected, while guaranteeing reasonable computational costs. The expected in-control parameter, the ALI_0 and the first-change point time (τ_{π_1}) are fixed at 0.001, $60/\theta_{\text{m}}$ and $20/\theta_{\text{m}}$ time-units, respectively. The duration of the epidemic shift ($\tau_{\pi\delta}$) varies and it is expressed as a function of the in-control parameter. The percentage of missing opportunities (% MO) to detect a change within the simulation length is also reported. It must be noted that the multiple change-point statistics must be reset after a shift is detected and the change-point locations are estimated.

GLR Type			Persistent Shift ($\log \Phi_k^{\Gamma}$)			Epidemic Shift ($\log \Psi_k^{\Gamma}$)		
Metric			CED	SD(CED)	%MO	CED	SD(CED)	%MO
r = 1	θ_1	$\tau_{\pi\delta} \times \theta_m$						
		3	20.8	18.9	0.55	**20.5**	**17.7**	**0.56**
	0.005	5	14.0	17.5	0.78	**13.8**	**15.9**	**0.79**
		10	**3.83**	10.3	0.97	4.84	**9.70**	0.97
		3	21.1	18.7	0.36	**19.9**	**17.7**	**0.37**
	0.01	5	14.0	18.2	0.55	**13.1**	**16.6**	0.55
		10	**2.23**	8.65	0.83	2.33	**7.11**	0.83
		3	21.8	19.0	0.19	**20.6**	**18.0**	**0.2**
	0.02	5	8.98	12.2	0.33	**7.63**	**10.5**	0.33
		10	1.1	4.9	0.58	1.1	**4.2**	0.58
r = 2	θ_1	$\tau_{\pi\delta} \times \theta_m$						
		3	23.9	17.3	0.14	**22.5**	**16.3**	0.14
	0.005	5	19.9	17.6	0.18	**17.2**	**15.8**	**0.2**
		10	7.14	13.9	0.41	**6.99**	**10.9**	0.42
		3	24.9	17.4	0.04	**22.9**	**16.5**	**0.04**
	0.01	5	23.2	17.5	0.06	**18.8**	**15.9**	0.06
		10	11.3	17.7	0.16	**6.59**	**11.9**	0.5
		3	25.1	16.9	0.01	**23.3**	**16.2**	0.01
	0.02	5	23.3	16.9	0.02	**18.8**	**15.3**	0.02
		10	12.3	17.7	0.05	**6.99**	**12.3**	0.05

Table 3.5: Epidemic shift performance metrics for detecting upward shifts in θ, with $\theta_m, = 0.001$, $\text{ALI}_0 = 60/\theta_m$, $\delta = \tau_{\pi\delta}$, for multiple chart orders (r).

Results show that the first-order ($r = 1$) chart tailored for an epidemic shift reacts slightly faster for any shift amplitude when the duration of the epidemic shift is very short. However, when the duration of the shift is longer, the traditional GLR performs better. The second-order chart ($r = 2$) performs consistently better than its competitor in terms of CED. It is also remarkable that the SD(CED) is consistently lower for the epidemic shift statistic for all combinations.

3.5 Conclusions

Generalized likelihood ratio control charts represent a convenient tool for tailoring monitoring schemes to industrial applications, but they are less developed for discrete processes. This chapter, inspired by the work of Lee and Woodall (2018), proposes an improved generalized likelihood ratio design that does not involve the specification of an artificial bound for geometrically distributed observations, while guaranteeing the full definition of the statistics. The notions of change-points and GLR charts is extended to observations that are aggregated over time, proposing the gamma GLR chart for monitoring homogeneous Poisson processes for simple null and alternative hypotheses. Inexact parameter estimations via the maximum likelihood method are corrected for interval restrictions, aspect often neglected in the literature. Via a simulation study, the proposed gamma GLR chart is compared to its traditional memory-type and memory-less competitors showing overall superior performance than any other chart for a wide range of shifts. Then, the generalized likelihood ratio framework applied to statistical process control is extended by considering composite null and alternative hypotheses. This approach offers an extra layer of flexibility, allowing to design monitoring schemes based on industry-tailored change-point models. Section 3.3 proposes the indifference interval and the epidemic shift for negative binomial and gamma-distributed observations. The indifference interval model suffers instability when the set-up allows for bidirectional shifts while the one-sided version proved effective at detecting only shifts of practical importance. The control chart based on the epidemic shift model showed better performance when the shift length is short, which is expected when a feedback-controller is active.

Chapter 4

Generalized Linear Model Monitoring Methods for High-Purity Processes

Traditional control charts aim at detecting process departures from normality by independently monitoring one or more quality variables. When the variables of interest exhibit a strong dependence on some covariates, out-of-control situations cannot be described in terms of the quality characteristics only. Instead, the monitoring scheme should be able to raise an alarm when the abnormal behavior is contextualized, i.e., to detect contextual anomalies. An example from the chemical industry is the production of plastic pellets. The monitored quality of interest is the number of defects that depends on the fluctuations in the weight of the inspected product. A valid method for detecting contextual anomalies is the application of regression-based control charts first introduced by Mandel (1969). The idea is to combine control charts with regression analysis by establishing the relationship between the variable of interest and its covariates and using the model residuals to detect out-of-control situations. The regression model contextualizes the process departure from the in-control state, and the use of the residuals allows for correcting for external factors. This setup has several applications in industrial settings, but it is also common in other fields. For example, this procedure is used in econometrics to detect structural changes (see e.g., Dufour (1982)). In medical monitoring applications, it is common to adjust observations by using health covariates of individual observations. Sachlas et al. (2019) provide an extensive overview of so-called risk-adjusted control charts. The simplest monitoring scheme using regression models is to fit and fix a regression model during Phase I, then monitor residuals with

respect to the fixed regression model, namely predictive residuals. The drawback of this procedure is that the residuals are correlated, making assessing the monitoring scheme's performance a difficult task. Brown et al. (1975) introduced a recursive approach for obtaining regression residuals over time, where a new regression model is fitted at the arrival of each new observation, after which the regression residuals are obtained. This approach fits within the area of self-starting control charts, which are practical in situations where Phase I is not feasible. Furthermore, recursive residuals are proved to be uncorrelated with zero mean and constant variance for normally distributed observations. Galpin and Hawkins (1984) discuss several applications of recursive residuals. Linear regression models with normally distributed error terms are by far the most common regression strategy in the field. Still, some quality response variables cannot be modeled by the normal distribution. Generalized linear models (GLM) offer a more flexible choice on the distribution of the monitored variables, especially for count data, as they extend it to a more comprehensive set of distributions belonging to the family of exponential dispersion models. Residuals for generalized linear models are more complex than for standard linear regression models. Several types of residuals are proposed for GLM models, with Pearson, deviance, and randomized quantile residuals being the most commonly used. The recursive residuals are introduced in the generalized linear model framework by McGilchrist and Matawie (1998). Several studies have been conducted for non-normal response variables by applying the regression framework. The most common choice for count data is to use the Poisson or negative binomial distributions and their variants, like the zero-inflated Poisson and the Conway-Maxwell (COM) distributions. It must be noted that the Conway-Maxwell and the Zero-inflated models can be used in regression settings, but they do not belong to the GLM framework. Skinner et al. (2003) introduced a GLM-based Shewhart chart with deviance residuals for Poisson distributed dependent observations and showed that it compared favorably to Shewhart charts based on OLS regression models. Skinner et al. (2004) extended the approach to the case in which the observations are assumed to follow an over-dispersed Poisson distribution that is modeled by a Poisson mixture model with an additional dispersion parameter, reaching similar conclusions. Park et al. (2018) introduced regression-based Shewhart charts for negative binomial and Conway-Maxwell (COM) Poisson distributions. They used principal component analysis (PCA) to reduce the number of covariates to address the problem of multicollinearity. It is shown that COM regression-based control charts perform better for under-dispersed count data. In contrast, the negative binomial regression-based control chart performs better in the

case of over-dispersed data. Zero-inflated regression models are commonly used to monitor high-purity count data affected by one or more covariates. Park et al. (2020) showed that for negative binomial and the Poisson distributions and their variants, as mentioned above, Shewhart control charts based on randomized quantile residuals perform better than their counterparts based on deviance residuals. Arshad et al. (2021) compared the performance of Shewhart, EWMA and CUSUM regression control charts residuals for the Poisson distribution and its variants, indicating which control chart performs best in specific cases. Mahmood (2020) showed that regression-based Shewhart charts for zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) distributions monitoring Pearson residuals performed better than ordinary Shewhart charts. The motivation behind this work is driven by applying monitoring schemes to high-purity processes characterized by low-rate count occurrence. Traditional count-based control charts are inappropriate for low defect rates as discussed in Section 1.1. A successful alternative approach is monitoring cumulative quantities over time. The observations can be modeled in these settings by a homogeneous Poisson process. Since the cumulative quantities are times until the r^{th} non-zero counts, the time-between-events observations follow a gamma distribution. GLM-based control charts for the gamma distribution have received less attention in the literature. Jearkpaporn et al. (2003) introduced a GLM-based Shewhart chart with deviance residuals and showed that GLM-based controls chart outperform traditional control charts and OLS-based control charts. Chimka (2009) presents the results of a case study related to influenza activity. The novelty of the discussion introduced in this chapter resides in applying these strategies to cumulative quantities control charts, namely time-between-events directly. This task is achieved by an in-depth analysis of the simulations to explore the performance of different residuals for Shewhart-based control charts for the gamma regression model. The rest of the chapter is structured as follows. Section 4.1 presents the theoretical background, including the null and alternative hypotheses statements and the introduction of the different types of residuals. Next, in Section 4.3 the results from the simulation study are shown and the schemes proposed are compared. Finally, Section 4.4 contains the conclusions and recommendation for future work.

4.1 The regression-based change-point model

When monitoring anomalies coupled with regression models (i.e., contextual anomalies) one is interested in testing the constancy of the regression parameters over time rather than the distributional parameters of the dependent variable. This assessment allows practitioners to take decisions that are independent of expected and unexpected fluctuations of the covariates. The default setting in discrete time is a generalized linear model with response observations Y_i with $i = 1, \dots,$ a link function $g(\cdot)$, and a set of covariates, given by the vector of regression parameters $\nu = (\nu^0, \dots, \nu^n)$ (see e.g, Dunn and Smyth (2018)). Assuming a persistent change in the regression parameters, the standard change-point detection model can be formalized in terms of null and alternative hypotheses as follows

$$\begin{aligned} H_0 : &\quad \nu_i = \nu_0, \text{for all } i \\ H_a(\tau) : &\quad \begin{cases} \nu_i = \nu_0, \text{for } i = 0, \dots, \tau \\ \nu_i = \nu_1, \text{for } i = \tau + 1, \dots \end{cases} \end{aligned} \tag{4.1}$$

where ν is the regression parameters vector ($\nu = (\nu^0, \dots, \nu^n)$) and $\nu_i = \nu_1$ when at least one of $\nu_i^0 \neq \nu_0^0, \dots, \nu_i^n \neq \nu_0^n$ holds. In the continuous-time space, in-control high-purity processes can be described by a homogeneous Poisson process with rate $1/\theta$ and their inter-arrival time can be used to determine whether an increase in the Poisson mean θ has occurred from the in-control situation. The times between r events are independent random variables X_k, each of which is distributed as a gamma distribution with rate $1/\theta$ and shape r, which is fixed beforehand and does not need to be estimated. Changing the parameter r allows for refining the calibration procedure. Small r values lead on average to more decision points, but larger variance which has a negative effect on detecting changes. By determining the performance of the monitoring schemes with different values of r, one can balance these two effects by choosing the most appropriate r for the monitoring goal at hand. The relationship between Y and its covariates is modeled via the logarithmic link function, to avoid the need for constraints on the linear predictor (i.e., $\theta_i > 0$), as shown in (4.3). For simplicity, it is assumed that the outcome variable is influenced by one covariate W. The covariate quantities are aggregate statistics over time and they can represent the aggregated quantities until the next event when this choice is relevant to the nature of the process (e.g., weight, volume) or other statistics as mean or median when

the covariate has no physical sense when aggregated over time (e.g., temperature, pressure, pH).

$$f_\Gamma(y; r, \theta_i) = \theta_i^{1-r}\Gamma(r)^{-1}y^{r-1}e^{-\frac{y}{\theta}} \tag{4.2}$$

$$E(Y) = r\theta = \exp(\nu^0 + \nu^1 W) \tag{4.3}$$

The regression coefficients can be estimated by maximizing the log-likelihood function, by means of the iterative weighted least squares (IWLS) algorithm during Phase I. The estimated regression coefficients are denoted by $\widehat{\nu}^0$ and $\widehat{\nu}^1$, respectively. As mentioned before, the shape parameter r of the distribution is fixed, since it represents a choice in the design of the control chart, and thus does not need to be estimated, in contrast to the general GLM case. The relationship in (4.3) can be used to predict the value of μ_i as follows

$$\widehat{\mu}_i = \exp(\widehat{\nu}^0 + \widehat{\nu}^1 w_i). \tag{4.4}$$

Per each coupled observation (y_i, w_i) and the corresponding mean $\widehat{\mu}_i$, the residuals or the regression coefficients can be calculated and monitored over time.

4.2 Monitoring residuals and regression coefficients

In this section, the types of residuals used in this chapter are described in detail.

Pearson Pearson residuals (r^P) are the most intuitive and straightforward to compute. Similarly to Pearson residuals in linear regression, the raw residuals are divided by the standard deviation, which is not constant in this case, as it depends on its expected value. They are expressed as follows

$$r^P = \frac{y - \widehat{\mu}}{\sqrt{V(\widehat{\mu})}}, \tag{4.5}$$

where $V(\mu)$ is the variance function of the selected distribution. It is widely known that the Pearson residuals are skewed in many GLM applications.

Deviance The deviance residuals (r^D) are defined as the signed square root of the unit deviance. The unit deviance is twice the difference in log-likelihood between the saturated model and the fitted model, multiplied by the dispersion parameter. The deviance represents a model performance metric, generalizing the concept of the residuals sum for normally distributed random variables (see, e.g., McCullagh and Nelder (1989)). The deviance residuals are expressed as follows

$$r^D = \text{sign}(y - \widehat{y})\sqrt{d(y, \widehat{y})}, \tag{4.6}$$

where $d(y, \widehat{y})$ is the unit deviance. The deviance statistic has an approximate χ^2 distribution when the saddle point approximation applies to the distribution of the responses Y, and they can be far from being normally distributed otherwise (Dunn and Smyth (2018)). Nevertheless, deviance residuals are more likely to be normally distributed compared to Pearson residuals, since the central limit theorem has a slower convergence rate then the saddle point approximation.

Quantile Quantile residuals (r^Q) were first proposed by Dunn and Smith (1996) to overcome the methodological drawbacks of Pearson and deviance residuals, particularly for discrete distributions. Quantile residuals have exactly a normal distribution. They are defined for the more general regression class of exponential dispersion models. Let $\mathcal{F}(y; \mu, \phi)$ be the cumulative distribution function of the random variable Y, where μ denotes the parameter vector and ϕ the dispersion parameter. If Y is a continuous random variable, the quantile residuals are given by

$$r^Q = \Phi^{-1}\{\mathcal{F}(y; \widehat{\mu}, \phi)\}, \tag{4.7}$$

where Φ is the cumulative function of the standard normal distribution. Deviation from normality is expected when the model does not show a good fit.

Recursive Recursive residuals (r^R) were introduced by Brown et al. (1975) for normal and non-normal space model diagnostics and thoroughly discussed by Dufour (1982) for linear regression models. Recursive residuals are defined from the one-step ahead predictive distribution as shown in (4.8). The idea behind is to re-estimate the regression model each time a new observation is recorded at time t, using the $t - 1$ previous observations, and the new model is used to calculate the residuals.

An advantage of the recursive method is that it can be used to shorten the initialization period of the monitoring procedure. The drawback is that persistent trends are learned by the model and the recursive method might fail to detect such trends since the model adapts to them. Recursive residuals in the GLM framework were introduced in McGilchrist and Matawie (1998) and they are as follows

$$r^R = \frac{y_i - E(\mu_i|y^{i-1})}{\sqrt{V(\mu_i|y^{i-1})}}, \tag{4.8}$$

where y^{i-1} is the observation at time $i-1$ and therefore $E(\mu_i|y^{i-1})$ and $V(\mu_i|y^{i-1})$ are the expected value and the variance from the model built with $i-1$ observations, respectively.

Regression coefficients Similarly to the recursive residuals, the regression model coefficients $(\widehat{\nu^0}, \widehat{\nu^1})$ can be re-estimated at every new observation and the estimates can be then monitored over time. Monitoring the constancy of the regression parameters is particularly popular in the econometric field when aiming at detecting structural changes governed by relationships between variables.

4.3 Performance evaluation and comparison

In this section, the process simulation set-up and the implementation of the residuals of the gamma regression model are discussed. Then, the predictive residuals are compared to each other in terms of faster detection and the performance of the recursive residuals as a monitoring tool for short runs are evaluated.

4.3.1 Process simulations

The time between events of a non-homogeneous Poisson process can be simulated by transforming a homogeneous Poisson process into a non-homogeneous one through the cumulative intensity function of the latter, similarly to the approach used in Section 3.2 but adjusted to accommodate the specification of the covariate W and the in-control regression parameters. By using this approach, it is possible to exactly specify and fix the change-point time at the process scale τ_π without artificial

corrections. The intensity function used in this study is a piece-wise function as follows

$$\theta(t) = \theta_1 \mathbb{1}_{\{t \leq \tau_{\pi_1}\}} + \cdots + \theta_m \mathbb{1}_{\{t > \tau_{\pi_m}\}}, \tag{4.9}$$

where $\theta_1, \ldots, \theta_m$ are the means of the probability distribution in (4.3), which vary with the set values of W. The random effects of the covariate W are neglected to isolate the different effects. However, in practice, combined effects are expected. A set of values of W are chosen and the response variable is drawn from the density probability function in (4.3) with $\theta_i = \exp(\nu_i^0 + \nu_i^1 w_i)$.

4.3.2 Implementation of the residuals

The Pearson and deviance residuals for the regression model in (4.3) for the new paired observation (y_i, w_i) and the correspondent mean $\widehat{\mu}_i$ are

$$r_i^P = \frac{y_i - \widehat{\mu}_i}{\widehat{\mu}_i} \tag{4.10}$$

since the variance function for the gamma distribution with fixed r is $V(\mu) = \mu^2 = (r\theta)^2$ and

$$r_i^D = \text{sign}(y_i - \widehat{\mu}_i)\sqrt{2\left\{-\log\left(\frac{y_i}{\widehat{\mu}_i}\right) + \frac{y_i - \widehat{\mu}_i}{\widehat{\mu}_i}\right\}}, \tag{4.11}$$

respectively. The deviance residuals in (4.11) are obtained by substituting the estimated expected mean according to the GLM model (4.4) in the general definition of the deviance residuals in (4.4). The quantile residuals can be defined by substituting the cumulative gamma distribution in (4.7) as follows

$$r_i^Q = \Phi^{-1}\{\mathcal{F}(y_i; \theta = \widehat{\mu}_i/r, r)\} \tag{4.12}$$

where $\mathcal{F}$ is the probability density function of the gamma distribution as follows.

$$\mathcal{F}(y; \theta, r) = \frac{1}{\theta^r \Gamma(r)} \int_0^y y^{r-1} \exp(-y/\theta)\, dy. \tag{4.13}$$

Pearson, deviance and quantile residuals are calculated assuming that the regression parameters are known and fixed in Phase I. Therefore, there is no need to re-fit a generalized linear model at every

new observation. When re-fitting the model one needs to consider the effect of the estimation of the parameters to compare the residuals performance. Note that when using a built-in GLM regression algorithm for the gamma distribution, the default setting is that the reciprocal of the shape parameter r is by definition the dispersion parameter and it is estimated. The estimation does not affect the regression parameter estimates, but it does affect the residuals since they are typically scaled by the dispersion parameter. In general, the dispersion parameter is seldom known except in cases in which it can be inferred from the context. In this study the dispersion parameter represents a user choice during the design of the control chart and it is therefore fixed as discussed in Section 4.2. Using the definition in (4.7), the recursive residuals for the chosen distribution at time i are as follows

$$r^R = \frac{y_i - \exp(\nu^0_{i-1} + \nu^1_{i-1} w_i)}{\exp(\nu^0_{i-1} + \nu^1_{i-1} w_i)}, \tag{4.14}$$

where ν^0_{i-1} and ν^1_{i-1} are the regression coefficient of the model estimated using the previous $i-1$ observations. To compute the regression coefficients, the first k observations are used to get an initial estimate of the vector ν_0 (burn-in period). The sample is then gradually enlarged one observation at a time and the model is re-estimated at each step. This method is more computationally expensive than the predictive counterpart. Since the recursive residuals are re-estimated at each time point, when the shift is moderate or the number of observations is large, the residuals smoothly adapt to the new conditions. Therefore, the number of observations in the burn-in period should be small enough to allow fast detection during the initial stage of the monitoring procedure, but large enough to be able to detect a shift in the parameter while not allowing the model to adjust to the new set of parameters. Intuitively, the intended use of recursive residuals is primarily for short runs or the initial stage of the monitoring procedure when the number of observations is not sufficient to reliably estimate the regression coefficients. Given the conditions tested in this study, it has been shown empirically that at least 40 observations should be obtained to confidently set the null hypothesis in terms of regression coefficients to initialize and design the control chart using the predictive residuals. The density plots of the residuals under the in-control conditions are shown in Figure 4.1. As expected, randomized quantile and deviance residuals follow approximately the standard normal distribution, while the Pearson and the recursive residuals distributions deviate from the normal behavior. Classical Box-Cox transformations could be used to shape the Pearson and recursive residuals into a sym-

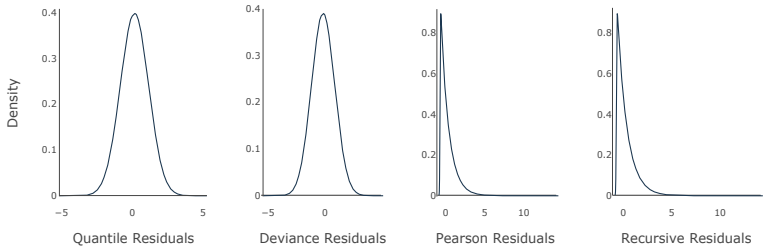


Figure 4.1: Quantile, deviance, Pearson and recursive residuals distributions.

metric distribution, but the presence of negative values and the dependency of the transformation shape from the offset chosen make this step unstable. The normality condition is not a prerequisite for using the Pearson and recursive residuals, but it might be useful in order to avoid the high sensitivity of the control limits over the performance measures when detecting downwards shifts, due to their pronounced skewness. The regression coefficients estimates are highly correlated to each other

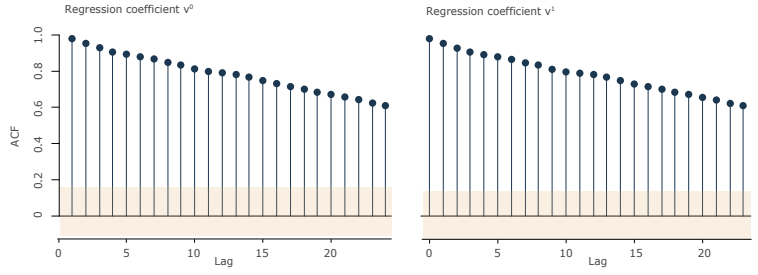


Figure 4.2: Autocorrelation plots of the regression coefficients ν^0 and ν^1.

and highly auto-correlated as shown in the auto-correlation plots in Figure 4.2. To monitor directly the regression coefficients, the dimension of the regression coefficient vector can be reduced using principal component analysis (PCA) coupled with an ARMA model to lose the auto-correlation pattern. However, this sequence of steps increases the complexity of the monitoring system. Since the information on the regression coefficients change over time is embedded in the recursive residuals, only the latter are evaluated further.

4.3.3 Performance evaluation and comparison

In this section, the control limits are calculated and the predictive residuals are compare to each other. Moreover, the recursive residuals are assessed for detecting departures from the in-control condition at the initial stages of the monitoring procedure.

Control limits The control limits are calculated based on the type of chart and the desired in-control average length of inspection (ALI), which always refers to the original data (i.e., the cumulative time to an event) and the chosen length of the simulations. This metric represents an intuitive measure that can be chosen by practitioners in relation to the technical specifications of each application as it expresses the average total time elapsed between two false alarms (e.g., one per shift, one per day). The control limits for upper and lower sided settings of the proposed residuals are reported in Table 4.1 for different values of ALI_0. The control limits for Pearson and recursive residuals are extremely sensitive to the in-control average length of inspection particularly for the lower-sided case because of the residuals' skewness.

	Upper-sided				**Lower-sided**			
ALI_0	h_{r^P}	h_{r^D}	h_{r^Q}	h_{r^R}	h_{r^P}	h_{r^D}	h_{r^Q}	h_{r^R}
30	2.42	1.54	1.84	2.42	-0.96	-2.20	-1.84	-0.97
40	2.73	1.68	1.98	2.74	-0.97	-2.35	-1.98	-0.98
50	3.01	1.81	2.10	3.02	-0.98	-2.46	-2.10	-0.98
60	3.29	1.92	2.20	3.29	-0.99	-2.58	-2.22	-0.99

Table 4.1: Predictive and recursive residuals' control limits for several ALI_0, $r = 1$ with regression parameters $\nu_0^0 = 2.9$ and $\nu_0^1 = 4$.

Predictive residuals comparison Pearson, deviance and quantile residuals are compared in one-sided settings in order to detect upwards and downwards shifts in at least one of the two regression parameters ν_0, ν_1 in terms of conditional expected delay (CED). The conditional expected delay expresses the delay of detection from the first opportunity to detect a change, conditional on not having raised an alarm prior to the change-point. Different change-points at the process time (τ_π) are considered. The τ_π values are chosen in a range between 0 (start of monitoring) and the calibration average length of inspection. It must be noted that, when W is kept constant, an increase in the parameters

corresponds to a decrease in the defect rate. Therefore, an increase of the regression coefficient, if positive, results in an improvement of the process and vice-versa. In most industrial applications, detecting deterioration of the process is more important for quality control. Since W in this study is kept constant and equal to 1, only changes in ν_1 are considered, which is equivalent to changes in ν_0 of the same intensity. As shown in Table 4.2, Pearson, deviance and quantile residuals show comparable delays of detection when the aim is to detect upward shifts in the regression parameters. Quantile residuals show slightly better performance when the shift is moderate, and the difference is less serious when considering larger deviations from the in-control conditions. No difference in performance is notable at different change-points. Table 4.3 reports the results from the simulation study considering a downward shift in the regression parameters. In this case, deviance and quantile residuals show comparable performance, while Pearson residuals show systematically longer delays of detection than their competitors. This is due to the high sensitivity of the Pearson residuals to the control chart limits on the lower-side of the distribution, which worsens their performance. Therefore, for detecting process deterioration, deviance or quantile residuals should be used. In this case, higher change-point values show slightly better performance, particularly at moderate shifts from the in-control condition and this trend is observable in all the three types of residuals.

		CED			**SD(CED)**		
Residual type		r^P	r^D	r^Q	r^P	r^D	r^Q
$\nu^1 = (\nu_0^1, \nu_1^1)$	τ_π						
$(2.9, 4.5)$	0	22.1	22.3	**21.9**	17.4	17.5	**17.2**
	10	22.4	22.4	**22.1**	17.5	17.7	**17.4**
	50	21.9	22.1	**21.8**	17.2	17.4	**17.0**
$(2.9, 5)$	0	13.1	13.1	**13.0**	7.71	7.76	**7.65**
	10	13.1	13.1	**13.0**	7.76	7.80	**7.69**
	50	13.1	13.1	**13.0**	7.73	7.77	**7.66**
$(2.9, 5.5)$	0	11.6	11.6	**11.5**	5.92	5.92	**5.89**
	10	11.6	11.6	**11.5**	**6.02**	6.04	6.05
	50	11.6	11.6	**11.5**	5.94	5.97	**5.92**

Table 4.2: Conditional expected delay of the GLM predictive residuals for detecting upwards shifts.

		CED			SD(CED)		
Residual type		r^P	r^D	r^Q	r^P	r^D	r^Q
$\nu^1 = (\nu_0^1, \nu_1^1)$	τ_π						
$(2.9, 2)$	0	1.88	**1.42**	1.43	2.00	1.55	1.55
	10	1.80	**1.42**	1.43	1.99	1.53	1.53
	50	1.88	1.42	1.42	2.01	**1.54**	1.55
$(2.9, 3.2)$	0	19.2	14.9	14.9	18.1	14.9	14.9
	10	18.8	**14.8**	14.9	17.5	**14.6**	14.7
	50	17.6	14.3	14.3	15.6	13.7	13.7
$(2.9, 3.5)$	0	32.1	**26.1**	26.2	28.2	24.6	**24.7**
	10	31.4	25.9	25.9	27.2	**24.1**	24.2
	50	28.1	24.9	**24.0**	23.0	**21.1**	21.2
$(2.9, 3.7)$	0	43.8	**36.7**	36.9	36.6	32.9	**32.0**
	10	43.8	**36.2**	36.4	35.3	32.1	32.1
	50	37.3	**32.7**	32.8	29.5	**27.6**	27.7

Table 4.3: Conditional expected delay of the GLM predictive residuals for detecting downwards shifts.

Recursive residuals evaluation The recursive residuals are also evaluated in terms of CED for different burn-in periods and process time change-points τ_π. The fraction of missed opportunities to detect a change within a limited time frame is also reported. The time frame corresponds to the chosen size of the iterations, set at the calibration average length of inspection for this study. The intent is to evaluate the recursive residuals as a tool to use for short runs or at the initial stage of the monitoring. It is therefore important that if a change occurs, it is detected within a reasonable time frame. The performance metrics of the recursive residuals are reported in Table 4.4. It can be noticed that the delay slightly decreases when more observations are available in the burn-in and the in-control period before the change point, at the expense of the probability of missing a shift, which is higher when the number of observations increases. This trend can be noticed for both downwards and upwards shifts. Choosing the burn-in period should be seen as a trade-off between fast detection and the probability of missing a change and it should be optimized based on the application requirements.

Shift	Burn-in period	τ_π	CED	SD(CED)	% MO
$\nu^1 = (\nu_0^1, \nu_1^1)$					
$(2.9, 3.5)$	2	0	20.3	13.6	0.48
	4	0	19.3	13.3	0.46
	10	0	18.5	13.0	0.46
	2	2	18.7	12.9	0.49
	4	2	18.5	12.8	0.49
	10	2	18.1	13.2	0.49
	2	10	16.5	11.8	0.54
	4	10	16.3	11.5	0.54
	10	10	14.7	10.4	0.57
$(2.9, 4.5)$	2	0	28.6	30.7	0.25
	4	0	29.8	29.7	0.24
	10	0	28.7	27.7	0.24
	2	2	30.2	30.6	0.43
	4	2	30.4	29.7	0.39
	10	2	28.4	27.1	0.38
	2	10	24.15	25.4	0.59
	4	10	22.15	23.5	0.61
	10	10	28.32	24.83	0.68

Table 4.4: Conditional expected delay and fraction of missing opportunity (%MO) of the GLM recursive residuals for detecting upwards and downwards shifts for different of burn-in and in-control lengths.

4.4 Conclusions

The use of regression models in statistical process control provides a contextualization of the process variables integrating real-time correction for external factors. The most common strategy is to monitor the residuals with respect to the fixed regression model. In this chapter, the generalized linear model framework is applied to model time between events observations from a high-purity process using the gamma distribution, considering the effect that one covariate exerts on the quality variable of interest. Different types of predictive residuals, namely Pearson, deviance, and quantile, are compared to each other via a simulation study. Pearson, deviance, and quantile residuals show comparable delays of detection when the aim is to detect upward shifts in the regression parameters. However,

when the monitoring aim is to detect downward shifts, deviance and quantile residuals show comparable performance, while Pearson residuals show systematically longer delays of detection than their competitors because of the skewness of their distribution. Therefore, deviance or quantile residuals should be used for monitoring purposes. The implementation of predictive residuals requires a Phase I calibration where a moderately large number of observations are collected in order to establish the in-control baseline. For short runs and the initial phase of long runs control charts, one could use recursive residuals. The regression coefficients used in the recursive residuals are updated at each new observation. Recursive residuals show shorter delays when more observations are available in the burn-in period, at the expense of the probability of missing a change. These two aspects can be balanced based on the application requirements and criticality.

Chapter 5

Multivariate Monitoring Procedures for High-Purity Manufacturing Processes

Multivariate control charts for monitoring correlated quality characteristics have become increasingly prevalent in recent times, since monitoring these variables independently can be misleading as quality variables can exhibit complex correlation patterns. A review of multivariate multinomial and multiattribute quality control charts can be found in Topalidou and Psarakis (2009), Saghir and Lin (2015), and Cozzucoli and Marozzi (2018). The development of monitoring schemes for multivariate count data faces several mathematical challenges.

- There is an inherent difficulty due to the discreteness already existing in the univariate case for obtaining exact control limits for discrete monitoring variables;
- There are no unique multivariate counterparts of univariate discrete distributions;
- Most control chart methods rely on the normality assumption and the approximation to the normal distribution is generally inadequate.

A common choice for monitoring multivariate count data, and for high-purity processes in particular, is modeling the random variables using a multivariate Poisson distribution and reducing the dimensionality to univariate monitoring by computing the sum, the maximum, or the difference between the attributes, but little focus has been invested when considering these methods for monitoring processes characterized by a low defect rate. Intuitively, the limitations that apply to the univariate case

described in Section 1.1, represent shortcomings also in multivariate settings. Cumulative quantity control charts have proved to be an appropriate solution for processes with a low defect rate, but few strategies are available to monitor correlated quality characteristics simultaneously. Zwetsloot et al. (2021) classify multivariate time-between-event monitoring into two fundamentally different scenarios: the vector-based and the multivariate point process scenarios. The first one applies when uncensored data are observed one vector at a time. This *complete* vector captures all the information of the multivariate cumulative quantity. A control chart decision is taken when at least one defect from all the classes (e.g., variables) has occurred and it represents the cumulative time from the first to the last defect. The main advantage is that it generates a simple, univariate vector of time or quantity differences, and well-known univariate strategies can be employed. On the other hand, the defect classes are treated equally and the requirement to wait until a defect has occurred in all classes inevitably results in a detection delay, especially for detecting changes in the components with relatively different rates. Examples of vector-based implementations are presented in Xie et al. (2011) and Flury and Quaglino (2018). The multivariate point process scenario involves several correlated temporal point processes. The cumulative quantities vectors are being monitored simultaneously. As the authors point out, one approach would be to monitor each process separately, taking into account the effect of the correlation structure on the performance metrics. In this chapter, a multivariate point process method is proposed to monitor a p-variate stochastic counting process based on the Hermite distribution. The theoretical background that lays the ground for the work is described in Section 5.1, in which a specific multivariate Poisson distribution is introduced and its basic properties are discussed. The corresponding stochastic counting process is introduced in Section 5.2. It is conjectured that the superposition of the two marginal processes is the Gauss-Poisson process, also known as the stuttering Poisson process, since it has a nonzero probability of two events occurring simultaneously. Finally, in Section 5.3 the conclusions and recommendations for future work are summarized.

5.1 Campbell's multivariate Poisson distribution

The assumption at the basis of this work is that the data can be described by either a multivariate Poisson distribution for equidistant time intervals or by multiple correlated Poisson processes for continuous time. There is no unique way to build a multivariate Poisson distribution. Inouye et al.

(2017) discuss three classes of multivariate distributions derived from the Poisson distribution that permit nontrivial dependencies between the variables. The first class considers the joint distribution of two Poisson univariate marginals. The second class represents a mixture of independent multivariate Poisson distributions and for the third class, it is assumed that the univariate conditional distributions are derived from the Poisson distribution. Given the nature of the processes that inspire this research, the models considered generalize the univariate Poisson to a multivariate distribution with the property that the marginal distributions of each variable are Poisson. As Inouye et al. (2017) states, several attempts have been made with the result of developing the same class of models, with different derivations. The following approach is based on a specific multivariate Poisson distribution introduced in Campbell (1934). Let be $X_1 = Y_1 + U$, $X_2 = Y_2 + U$, where Y_1, Y_2, and U are mutually independent Poisson random variables with means λ_1, λ_2 and θ_0, respectively. The joint probability mass function is given by

$$P(X_1 = x_1, X_2 = x_2) = e^{-(\lambda_1+\lambda_2+\theta_0)} \sum_{i=0}^{\min(x_1,x_2)} \frac{\lambda_1^{x_1-i}\,\lambda_2^{x_2-i}\,\theta_0^i}{(x_1-i)!(x_2-i)!i!}, \tag{5.1}$$

where λ_1, λ_2 and $\theta_0 > 0$ and $(x_1, x_2) \in \mathbb{N}^2$.

Proof. The distribution of (X_1, X_2) can be determined by the disjoint union of events $(U, Y_1, Y_2) = (i, x_1 - i, x_2 - i)$ for all i such that all components are non-negative, i.e., $0 \leq i \leq \min(x_1, x_2)$. Then

$$\begin{aligned}
P((X_1, X_2) = (x_1, x_2)) &= \sum_{i=0}^{\min(x_1,x_2)} P(U, Y_1, Y_2) = (i, x_1 - i, x_2 - i)) \\
&= \sum_{i=0}^{\min(x_1,x_2)} P(U = i)P(Y_1 = x_1 - i)P(Y_2 = x_2 - i) \\
&= \sum_{i=0}^{\min(x_1,x_2)} \left(e^{-\theta_0}\frac{\theta_0^i}{i!}\right)\left(e^{-\lambda_1}\frac{\lambda_1^{x_1-i}}{(x_1-i!)}\right)\left(e^{-\lambda_2}\frac{\lambda_1^{x_2-i}}{(x_2-i!)}\right) \\
&= e^{-(\theta_0+\lambda_1+\lambda_2)}\frac{\lambda_1^{x_1}}{x_1!}\frac{\lambda_2^{x_2}}{x_2!}\left(\sum_{i=0}^{\min(x_1,x_2)} \frac{\theta_0^i}{i!}\frac{\lambda_1^{x_1-i}}{(x_1-i!)}\frac{\lambda_1^{x_2-i}}{(x_2-i!)}\right) \\
&= e^{-(\theta_0+\lambda_1+\lambda_2)}\frac{\lambda_1^{x_1}}{x_1!}\frac{\lambda_2^{x_2}}{x_2!}\sum_{i=0}^{\min(x_1,x_2)} i!\binom{x_1}{i}\binom{x_2}{i}\left(\frac{\theta_0}{\lambda_1\lambda_2}\right)^i.
\end{aligned}$$

□

The random variable U appears in both X_1 and X_2, establishing the correlation between the variables. Therefore, the covariance is as follows

$$\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(Y_1 + U, Y_2 + U) = \mathrm{Cov}(U, U) = \mathrm{Var}(U) = \theta_0. \tag{5.2}$$

Since θ_0 represents a Poisson rate parameter, the Campbell model can accept only a positive correlation. If the parameters are unknown, they can be estimated via maximum likelihood estimation (MLE). The parameters can be written in terms of Poisson rates per each variable (r_i) and the correlation (ρ). For the bivariate case they are as follows

$$r_1 = \lambda_1 + \theta_0; \tag{5.3}$$

$$r_2 = \lambda_2 + \theta_0; \tag{5.4}$$

$$\rho = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} = \frac{\theta_0}{\sqrt{(\lambda_1 + \theta_0)(\lambda_2 + \theta_0)}}. \tag{5.5}$$

The joint probability mass function in (5.1) can be extended to a p-variate distribution as follows

$$P((X_1, \ldots, X_n) = (x_1, \ldots, x_n)) = e^{-(\theta_0 + \sum_{i=1}^{p} \lambda_i)} \mathrm{Prod}_{k=1}^{p} \frac{\lambda_k^{x_k}}{x_k!} \left(\sum_{j=0}^{\min_j x_j} j! \left(\prod_{k=1}^{p} \binom{x_k}{j} \frac{\theta_0}{\prod_{k=1}^{p} \lambda_k} \right)^j \right). \tag{5.6}$$

Proof. Following the derivation for the bivariate case, the p-variate Poisson distribution can be ob-

tained similarly. Let $X_i = Y_i + U$ where $Y_i \sim \text{Poi}(\lambda_i)_{i\in\{1,2,\dots\}}$ and $U \sim \text{Poi}(\theta_0)$. Then

$$\begin{aligned}
P((X_1,\dots,X_n) = (x_1,\dots,x_n)) &= \sum_{j=0}^{\min_j(x_j)} P(U, Y_1,\dots,Y_n) = (i, x_1 - j,\dots,x_n - j)) \\
&= \sum_{j=0}^{\min_j(x_j)} P(U = j)P(Y_1 = x_1 - j)\dots P(Y_n = x_n - j) \\
&= \sum_{j=0}^{\min_j(x_j)} \left(P(U = j) \prod_{k=1}^{p} P(Y_k = x_k - j) \right) \\
&= \sum_{j=0}^{\min_j(x_j)} \left(e^{-\theta_0} \frac{\theta_0^i}{i!} \prod_{k=1}^{p} e^{-\lambda_k} \left(\frac{\theta_0^i}{i!} \right) \frac{\lambda_k^{x_k - i}}{(x_k - i!)} \right) \\
&= e^{-(\theta_0 + \sum_{i=1}^{p} \lambda_i)} \sum_{j=0}^{\min_j(x_j)} \left(\frac{\theta_0^i}{i!} \prod_{k=1}^{p} \frac{\lambda_k^{x_k - i}}{(x_k - i!)} \right) \\
&= e^{-(\theta_0 + \sum_{i=1}^{p} \lambda_i)} \prod_{k=1}^{p} \frac{\lambda_k^{x_k}}{x_k!} \\
&\quad \cdot \sum_{j=0}^{\min_j(x_j)} j! \left(\prod_{k=1}^{p} \binom{x_k}{j} \left(\frac{\theta_0}{\prod_{k=1}^{p} \lambda_k} \right)^j \right).
\end{aligned}$$

□

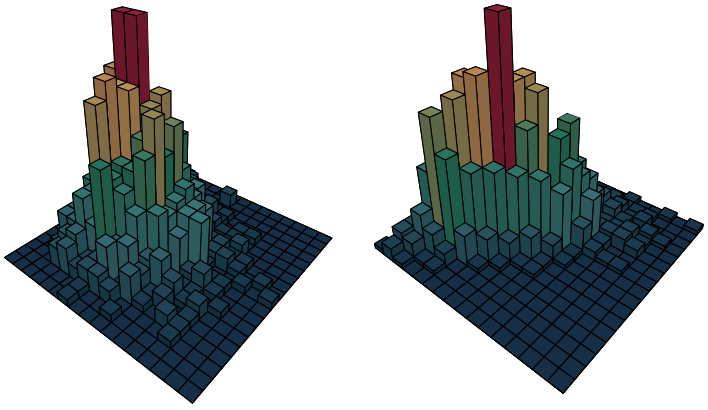


Figure 5.1: Campbell's bivariate Poisson histograms with independent Poisson marginals (left) and dependent Poisson marginals with a correlation of 0.8 (right).

Hermite distribution Consider the sum D of the count variables X_1 and X_2. Following the definition of Campbell's bivariate Poisson distribution, then the sum D is given by

$$D = X_1 + X_2 = (Y_1 + U) + (Y_2 + U) = (Y_1 + Y_2) + 2U. \tag{5.7}$$

The sum $Y_1 + Y_2$ is a Poisson variable with mean $\lambda_1 + \lambda_2$ and 2U is a Poisson doublet variable with mean $2\theta_0$. Kemp and Kemp (1965) shows that $Y_1 = Y_2 + 2U$ is distributed according to the so-called Hermite distribution. Let Z_1 and Z_2 be two independent Poisson variables with parameters γ_1 and γ_2, respectively. The probability distribution of the random variable $W = Z_1 + 2Z_2$ is the Hermite distribution. The probability mass function of W is given by

$$P(W = w) = e^{-(\gamma_1+\gamma_2)} \sum_{j=0}^{\lfloor n/2 \rfloor} \frac{(\gamma_1)^{n-2j}\gamma_2^j}{(n-2j)!j!}, \quad n = 0, 1, 2, \ldots \quad \gamma_1, \gamma_2 \geq 0, \tag{5.8}$$

where $\lfloor n/2 \rfloor$ denotes the floor function. It can be easily proved that $D \sim \text{Herm}(\lambda_1 + \lambda_2, \theta_0)$ with probability mass function as follows

$$P(D = d) = e^{-(\lambda_1+\lambda_2+\theta_0)} \sum_{j=0}^{\lfloor n/2 \rfloor} \frac{(\lambda_1 + \lambda_2)^{n-2j}\theta_0^j}{(n-2j)!j!}, \quad n = 0, 1, 2, \ldots \tag{5.9}$$

Proof. The probability generating function of the univariate Poisson distribution with rate λ is given by

$$P(s) = E[s^X] = e^{\lambda(1-s)}. \tag{5.10}$$

The probability generating function of the joint distribution of Y_1, Y_2, U is given by

$$\begin{aligned} P(t_1, t_2, t_{12}) &= E[t_1^{Y_1} t_2^{Y_2} t_{12}^{U}] && (5.11)\\ &= E[t_1^{Y_1}]E[t_2^{Y_2}]E[t_{12}^{U}] && (5.12)\\ &= e^{-\lambda_1(1-t_1)} e^{-\lambda_2(1-t_2)} e^{-\theta_0(1-t_{12})} && (5.13)\\ &= e^{-(\lambda_1+\lambda_2+\theta_0)+(\lambda_1 t_1+\lambda_2 t_2+\theta_0 t_{12})}. && (5.14) \end{aligned}$$

Recalling that $X_1 = Y_1 + U$, $X_2 = Y_2 + U$ and $t_1 = s_1$, $t_2 = s_2$ and $t_{12} = s_1 s_2$, then

$$\begin{aligned} P(s_1, s_2) &= E[s_1^{X_1} s_2^{X_2}] && (5.15)\\ &= E[s_1^{Y_1} s_2^{Y_2} (s_1 s_2)^{U}] && (5.16)\\ &= e^{-(\lambda_1+\lambda_2+\theta_0)+(\lambda_1 s_1+\lambda_2 s_2+\theta_0 s_1 s_2)}. && (5.17) \end{aligned}$$

In order to obtain the probability generating function of $X_1 + X_2$, let s be the equal to s_1 and s_2, yielding to

$$\begin{aligned} P(s) &= E[s^{X_1} s^{X_2}] && (5.18)\\ &= E[s^{X_1+X_2}] && (5.19)\\ &= e^{-(\lambda_1+\lambda_2+\theta_0)+(\lambda_1+\lambda_2)s+\theta_0 s^2}. && (5.20) \end{aligned}$$

which is the probability generating function of the Hermite distribution given in Kemp and Kemp (1965). Therefore, it can be concluded that $D \sim \text{Herm}(\lambda_1 + \lambda_2, \theta_0)$. If $(X_1, X_2) \sim \text{BivPois}(\lambda_1, \lambda_2, \theta_0)$ then $D = X_1 + X_2 \sim \text{Herm}(\lambda_1 + \lambda_2, \theta_0)$. □

Multivariate Poisson distributions using copulas As already mentioned, there is no general approach to define a multivariate discrete distribution. Copula models can be also used to form multivariate joint distributions, with the advantage to be extremely flexible in terms of dependency structure since, by Sklar's Theorem (Sklar (1959)), every multivariate probability distribution can be characterized by a copula function joining its marginal distributions. Control charts based on copulas have been studied in several papers, including Verdier (2013), Sukparungsee et al. (2017), Mühlig (2017), and Easton et al. (2022). A copula is a function that defines the dependency structure of multiple random variables when a more complex correlation structure is required to adequately describe the relationship between the variables. However, copula models paired with discrete marginal distribution are not unique due to identifiability problems. Genest and Nešlehová (2007) provide a detailed discussion on copulas for count distributions.

5.2 Hermite multiple Poisson process

Since the procedure aims at monitoring a change in the defect rate of different classes simultaneously, the behavior of independent and dependent multiple Poisson processes is of interest. The approach is to transform a multidimensional counting process into a one-dimensional one is studied. An option to achieve this transformation is to consider the sum of the arrivals of all the different classes of defects in the interval $[0, t]$. This process can be interpreted as the union or sum of the individual counting processes.

Multiple independent Poisson processes Consider two independent homogeneous Poisson processes, N_1 and N_2, with intensities γ_1 and γ_2, respectively. The inter-arrival times S_1 of N_1 are exponentially distributed with parameter γ_1, and the inter-arrival times S_2 of N_2 are exponentially distributed with parameter γ_2. The inter-arrival times of both processes are also independent of each other. The distribution of the total number of arrivals in $[0, t]$ of N_1 and N_2 is denoted by N and the inter-arrivals by S as the minimum of two independent exponential random variables. Let S be $S = \min(S_1, S_2)$, then

$$P(S > s) = P(\min(S_1, S_2) > s) = P(S_1 > s)P(S_2 > s) = e^{-s\gamma_1}e^{-s\gamma_2} = e^{-s(\gamma_1+\gamma_2)}. \tag{5.21}$$

It follows that S is also an exponential random variable, and $N(t)$ describes a Poisson process with intensity $\gamma_1 + \gamma_2$.

Multiple dependent Poisson processes Consider two dependent Poisson processes X_1 and X_2. Recall that the sum of two correlated Poisson random variables from Campbell's bivariate Poisson distribution described in (5.1) follows the Hermite distribution. The Poisson processes are composed of $Y_1 + U$ and $Y_2 + U$, respectively. The processes Y_1, Y_2, and U are mutually independent Poisson processes with intensities λ_1, λ_2, and θ_0, respectively. The total number of defects that have occurred in the interval $[0, t]$, can be described by the Hermite distribution, with parameters $((\lambda_1 + \lambda_2)t, \theta_0 t)$. Therefore we call the corresponding counting process, denoted by $D(t)$, the Hermite process[1]. The

[1] In the queuing literature a similar process is referred to as a *common Poisson shock process.*

probability mass function of $D(t)$ is given by

$$P(D(t) = n) = e^{-(\lambda_1+\lambda_2+\theta_0)t} \sum_{j=0}^{\lfloor n/2 \rfloor} \frac{((\lambda_1 + \lambda_2)t)^{n-2j}(\theta_0 t)^j}{(n-2j)!j!}, \; n = 0, 1, 2, \ldots \tag{5.22}$$

It follows from the previous section that the total number of arrivals of two correlated Poisson processes results in a Hermite process $D(t)$. We will see that the Hermite process differs from Poisson processes in the sense that it is possible to have two events at the same time, but the probability of having *more* than two events at the same is zero. In fact, the Hermite process turns out to be a compound Poisson process where a homogeneous Poisson process is compounded with a Bernoulli random variable.[2]

Theorem 5.2.1. *The Hermite process defined in (5.22) is a compound Poisson process of the form* $\sum_{i=1}^{N(t)} X_i$*, where* $N(t)$ *is a homogeneous Poisson process with intensity* λ*, the random variables* $X_i, i = 1, 2, \ldots$ *are iid (and independent of the* $N(t)$*) with* $P(X_i = 1) = \alpha_1, P(X_i = 2) = \alpha_2$*, and* $\alpha_1 + \alpha_2 = 1$*.*

Proof. Let $\widetilde{D}(t)$ be the process defined by

$$\widetilde{D}(t) = \sum_{i=1}^{N(t)} X_i,$$

where $N(t)$ is a homogeneous Poisson process with intensity λ, and the random variables $X_i, i = 1, 2, \ldots$ are iid (and independent of the $N(t)$) with $P(X_i = 1) = \alpha_1, P(X_i = 2) = \alpha_2$ and $\alpha_1 + \alpha_2 = 1$. We prove that $P\left(\widetilde{D}(t) = N\right) = P\left(D(t) = N\right)$. It is necessary to distinguish the case that n is even and

[2] The connection of the Hermite process with compound Poisson processes was conjectured in a private discussion by Marek Skarupski and later proved by him.

n is odd. Let n be an even number. Then

$$\begin{aligned}
P(\widetilde{D}(t) = n) &= P(N(t) = n, X_1 = 1, X_2 = 1, \dots X_n = 1) \\
&\quad + P(N(t) = n-1, \text{one of the } X_i\text{'s is 2, the rest are 1}) \\
&\quad + P(N(t) = n-2, \text{two of the } X_i\text{'s are 2, the rest are 1}) \\
&\quad + \dots \\
&\quad + P(N(t) = \frac{n}{2}, X_1 = 2, X_2 = 2, \dots, X_{\frac{n}{2}} = 2) \\
&= \sum_{j=0}^{\frac{n}{2}} P(N(t) = n-j) \binom{n-j}{j} \alpha_2^j \alpha_1^{n-j-j} \\
&= \sum_{j=0}^{\frac{n}{2}} e^{-\lambda t} \frac{(\lambda t)^{n-j}}{(n-j)!} \cdot \frac{(n-j)!}{j!(n-2j)!} \alpha_2^j \alpha_1^{n-2j} \\
&= e^{-(\alpha_1\lambda + \alpha_2\lambda)t} \sum_{j=0}^{\frac{n}{2}} \frac{(\lambda t)^{n-2j}(\lambda t)^j}{j!(n-2j)!} \alpha_2^j \alpha_1^{n-2j} \\
&= e^{-(\alpha_1\lambda + \alpha_2\lambda)t} \sum_{j=0}^{\frac{n}{2}} \frac{(\lambda\alpha_1 t)^{n-2j}(\lambda\alpha_2 t)^j}{j!(n-2j)!} \\
&= P\left(D(t) = N\right).
\end{aligned}$$

Let n be an odd number. Then

$$
\begin{aligned}
P\left(\widetilde{D}(t) = n\right) &= P(N(t) = n, X_1 = 1, X_2 = 1, \ldots, X_n = 1) \\
&\quad + P(N(t) = n-1, \text{one of the } X_i\text{'s is 2, the rest are 1}) \\
&\quad + P(N(t) = n-2, \text{two of the } X_i\text{'s are 2, the rest are 1}) \\
&\quad + \ldots \\
&\quad + P(N(t) = \frac{n-1}{2}, X_1 = 2, X_2 = 2, \ldots, X_{\frac{n}{2}} = 2) \\
&= \sum_{j=0}^{\frac{n-1}{2}} P(N(t) = n-j) \binom{n-j}{j} \alpha_2^j \alpha_1^{n-j-j} \\
&= \sum_{j=0}^{\frac{n-1}{2}} e^{-\lambda t} \frac{(\lambda t)^{n-j}}{(n-j)!} \cdot \frac{(n-j)!}{j!(n-2j)!} \alpha_2^j \alpha_1^{n-2j} \\
&= e^{-(\alpha_1\lambda+\alpha_2\lambda)t} \sum_{j=0}^{\frac{n-1}{2}} \frac{(\lambda t)^{n-2j}(\lambda t)^j}{j!(n-2j)!} \alpha_2^j \alpha_1^{n-2j} \\
&= e^{-(\alpha_1\lambda+\alpha_2\lambda)t} \sum_{j=0}^{\frac{n-1}{2}} \frac{(\lambda\alpha_1 t)^{n-2j}(\lambda\alpha_2 t)^j}{j!(n-2j)!} \\
&= P\left(D(t) = N\right).
\end{aligned}
$$

□

The compound Poisson process is a Hermite process with parameters $\lambda\alpha_1$ and $\lambda\alpha_2$. The Hermite process is a special case of the so-called Gauss-Poisson processes introduced by Newman (1970) through a probability generating functional as introduced in Moyal (1962). First, this concept is formally defined. Below, the treatment of Westcott (1972) is followed rather than the abstract treatment of Moyal (1962).

Definition 5.2.2. *Let* N *be a counting process on the real line. The probability generating function of* N *is defined as the following functional:*

$$
G(\xi) = E\left\{\int_0^\infty \log \xi(t) dN(t)\right\}, \tag{5.23}
$$

where ξ is any bounded complex-valued function on $\mathbb{R}$.

In general, the probability generating functional cannot be computed explicitly, but it is a useful tool

to derive properties. For non-homogeneous Poisson processes, it can be shown that

$$G(\xi) = \exp\left\{-\int_0^\infty (1-\xi(t))\Lambda(dt)\right\}. \tag{5.24}$$

In particular, for the homogeneous Poisson process it holds that[3]

$$G(\xi) = \exp\left\{-\int_0^\infty -\lambda(1-\xi(t))dt)\right\}, \tag{5.25}$$

which defines a Gauss-Poisson process through the probability generating functional specifying the following form (Milne and Westcott (1972), see also Example 6.2(c) of Daley and Vere-Jones (2003).

$$G(\xi) = \exp\left\{m\int_0^\infty (\xi(x)-1)dQ(x) + \frac{1}{2}m^2\int_0^\infty\int_0^\infty (\xi(t)-1)(\xi(u)-1)\rho(t,u)dH(t,u)\right\}. \tag{5.26}$$

It is not trivial to show that every such functional is indeed the probability generating functional of a counting process. It is obviously true if all the coefficients are non-negative, but there are cases in which this is also true when not all coefficients are non-negative (there is an example going back to Lévy (1937)). An alternative approach can be found in Milne and Westcott (1972), which is based on point process theory from Vere-Jones (1968), Vere-Jones (1970) and Westcott (1972). The Gauss-Poisson process differs from Poisson processes by the fact that it has a positive probability of having two events at the same time instead of one. The probability generating functional of this process is described in Milne and Westcott (1972) and it uniquely determines a point process.

Multivariate cumulative quantity control charts The Hermite process can be used in control chart settings to monitor the inter-arrival times of multiple, correlated, Poisson processes. This chart is called MTBE (i.e., Multivariate Time-Between-Events) Hermite chart. Let T_{ki} denote the time of the k^{th} arrival of the individual process i. From Theorem 5.2.1, since $N(t)$ is HPP with intensity $\lambda(t) = \lambda$, the process can be monitored using the cumulative quantities framework. The intensity parameter λ can be estimated from the Hermite process parameters through the relationship with the individual Poisson processes rates and their correlation (λ_1, λ_2 and θ_0). This chart is designed specifically for high-purity processes, with the limitation of accepting only a non-negative correlation parameter.

[3] from (Newman, 1970, p. 344).

5.3 Conclusions

This chapter explores a bivariate model for dependent count data with Poisson marginals. It is shown that the sum of these counts follows the Hermite distribution. By embedding these distributions into continuous-time through dependent Poisson processes and using the sums of the counts as control chart statistics, a Gauss-Poisson process is obtained. The Gauss-Poisson is a stochastic process that, unlike the Poisson process, has a positive probability of having two events co-occurring, which extends the implementation of such schemes to positively correlated Poisson processes. The study in this chapter is intended as a proof-of-concept on how established univariate procedures based on stochastic processes can be extended to multivariate settings for monitoring high-purity co-dependent variables.

Chapter 6

Conclusions and future perspectives

This last chapter summarizes the main contributions and conclusions, emphasizing directions for future work that would extend the current framework by providing other solutions to the challenges of monitoring high-purity processes.

6.1 Conclusions

Control charts represent the most popular technique in statistical process control and the most sophisticated one. Despite their popularity, traditional methodologies are inadequate to track in complex processes, particularly when these are intrinsically discrete.

Despite the fertile research on statistical process control, applications in the manufacturing industry are lagging, mainly because of the need for practical and targeted solutions. A specific feature in manufacturing environments resulting from continuous improvement efforts, attracting interest from academia and industry, is dealing with high-purity processes. These processes are characterized by the fact that observations are only available as counts and the occurrence rate of an event is low.

The research explores several practical and statistical challenges in monitoring high-purity processes arising from the complexity of technology-enhanced production processes, emphasizing some aspects that have practical importance in the industry but are relatively unexplored in the literature; these include tailored change-point and regression models, and multivariate control charts.

Chapter 1 provides the reader with a high-level overview of the current state of the art in statistical process monitoring for high-purity processes. Traditional attribute control charts reveal some limitations when applied to high-purity processes, such as meaningless control limits. Cumulative quantities control charts represent a valid alternative to conventional monitoring. The concept behind these charts is to consider the number of items or the time between two subsequent non-conforming events. Discrete and continuous versions are available based on the assumption that the non-conforming occurrence follows the geometric or Poisson distribution, respectively. This concept can be extended by waiting for r non-conforming items before a decision about the state of the process is prompted to the user. The aggregate statistics can then be used in Shewhart-type or memory-type settings.

Chapter 2 focuses on addressing the limitations regarding the performance evaluation of charts based on aggregate data. The main shortcoming is represented by the asynchronicity of the process and control chart time. Zero-state and steady-state metrics and their advantages are extensively discussed and adapted to the high-purity monitoring framework. Zero-state performance relies on the assumption that the shift has happened at the beginning of the monitoring procedure, while steady-state metrics abandon this unrealistic assumption. Explicit formulas for the performance of cumulative quantities control charts are given. It is shown that the error's severity of assuming the independence of the stopping time incorrectly from the observations is significant, and the trend is unexpected. To help practitioners, examples and guidelines for the optimal choice of the order of the control charts are given. Based on a simulation study, it is concluded that the behavior of several performance metrics is not always simple to describe. The entire distribution of the performance metric should be evaluated, particularly when run rules are applied.

Chapter 3 addresses the flexibility required to monitor complex processes than the one provided by simplistic models. Generalized likelihood ratio control charts represent a convenient tool for tailoring monitoring strategies to industrial purposes, but they are less developed for discrete processes. The generalized likelihood framework is applied to monitor aggregate data over time. In this chapter, several improvements of generalized likelihood ratio-based control charts are discussed. First, it is shown that for the geometric and negative binomial control charts, introducing an artificial bound-

ary is not required to guarantee that the statistic is fully defined. Since the artificial bound acts like a tuning parameter, this practice defies the purpose of using GLR charts. Additionally, it is noted that the correct parameter estimation is often overlooked in the literature. Generalized likelihood ratio control charts typically outperform competitors in detecting a wide range of parameter shifts.
Composite null and alternative hypotheses offer an extra layer of flexibility, allowing to design control charts based on industry-tailored change-point models. The indifference interval and the epidemic shift models are introduced, representing two expected deviation modes in the chemical manufacturing industry. The first allows a margin within which changes in the distributional parameters are not of practical importance; the latter describes the situation when a feedback controller is active and the deviation from the in-control state is corrected, but information about the change-point can be helpful for post-diagnostic. The indifference interval model suffers from instability when the set-up allows for bidirectional shifts, while the one-sided version proved to be effective at detecting only changes of practical importance. The control chart based on the epidemic shift model shows better performance when the length of the shift is short, which is expected when a feedback controller is active.

Chapter 4 addresses the need of correcting the monitoring characteristics for other variables. Using regression models in statistical process control provides a contextualization of the process deviations from normality by integrating real-time correction for external factors. The most common strategy is to monitor the residuals with respect to the fixed regression model. It is shown that different types of predictive residuals are effective at detecting such shifts, showing comparable performance. Nonetheless, implementing predictive residuals requires a Phase I calibration where a moderately large number of observations are collected to establish the in-control baseline. This requirement is often granted in practice, but this information is not always available for new products or short campaigns. One could use recursive residuals for short runs and the initial monitoring phase of long runs control charts. The regression coefficients used in the recursive residuals are updated at each new observation. Recursive residuals show shorter delays when more observations are available in the burn-in period at the expense of the probability of missing a change. These two aspects can be balanced based on the application requirements and criticality.

Chapter 5 In practice, most industrial manufacturing processes are characterized by several parameters exhibiting strong dependency. Such dependencies make designing monitoring schemes more challenging since one has to account for these often complex relationships to obtain reliable control limits. This chapter explores a bivariate model for dependent count data with Poisson marginals. It is demonstrated that the sum of these counts leads to the so-called Hermite distribution. It By embedding these distributions into continuous-time through dependent Poisson processes and using the sums of the counts as control chart statistics, a Gauss-Poisson process (also known as a *stuttering* Poisson process) is obtained. The Gauss-Poisson is a stochastic process that, unlike the Poisson process, has a positive probability of having two events co-occurring, extending the application of such schemes to case studies involving correlated counts.

6.2 Recommendations for future work

In the sequel, several ideas for future work are recommended that would extend the framework discussed in this dissertation.

- Other performance metrics than the average length of inspection in-control (ALI_0) could be explored for the task of calibrating the control chart. The average length of inspection is preferred in industrial contexts because it represents an intuitive metric to use as design parameter, which helps practitioners and non-statisticians taking informed and practical decisions about the monitoring capabilities of the control chart. Nonetheless, from a Bayesian point of view, the performance comparison could be based on the probability of false alarm $\text{PFA} = \text{P}(\text{N} < \tau)$, if there is a prior on the change-point τ. This approach is more consistent with the use of the conditional expected delay, being the PFA itself a steady-state metric.

- It is realistic to expect prior process information to be available, being high-purity processes generally well-documented and analyzed, particularly in the health sector. As future work, it is interesting to explore in detail how prior information can be utilized during the design phase of a monitoring scheme.

- It is well known that parameter estimation severely affects the performance of control charts. The severity of the effect of the estimation uncertainty and the model misspecification needs

to be studied more in detail for both the generalized likelihood ratio and the generalized linear model-based control charts.

- In the generalized linear model framework presented in Chapter 4, the random effects of the covariates are neglected. This might be a strong assumption and it might affect drastically the performance of such schemes. Therefore, the performance should be reassessed including the random effects of the covariates.
- The robustness of the generalized likelihood ratio and generalized linear model statistics to deviations from the binomial and Poisson distribution should be studied in comparison to more traditional monitoring approaches.

Bibliography

B. M. Adams, W. H. Woodall, and C. A. Lowry. The use (and misuse) of false alarm probabilities in control chart design. In Hans-Joachim Lenz, G. Barry Wetherill, and Peter-Theodor Wilrich, editors, *Frontiers in Statistical Quality Control 4*, pages 155–168. Physica-Verlag HD, 1992.

S. Ali, A. Pievatolo, and R. Gob. An overview of control charts for high-quality processes. *Quality and Reliability Engineering International*, 32:2171–2189, 2016.

J. Arshad, T. Mahmood, M. Riaz, and Hassan A. M. GLM-based flexible monitoring methods: An application to real-time highway safety surveillance. *Symmetry*, 13(2), 2021.

A. Aue, L. Horváth, M. Kühn, and J. Steinebach. On the reaction time of moving sum detectors. *Journal of Statistical Planning and Inference*, 142(8):2271–2288, 2012.

P. Bauer and P. Hackl. The use of mosums for quality control. *Technometrics*, 20(4):431–436, 1978.

D. Blackwell. On an equation of Wald. *The Annals of Mathematical Statistics*, 17(1):84–87, 1946.

D. Blackwell and M.A. Girshick. On functions of sequences of independent chance vectors with applications to the problem of the "random walk" in k dimensions. *The Annals of Mathematical Statistics*, 17:310–317, 1946.

R.L. Brown, J. Durbin, and J.M. Evans. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37(2):149–192, 1975.

T.W. Calvin. Quality control techniques for "zero defects". *IEEE Transactions on Components Hybrids and Manufacturing Technology*, 6(3):323–328, 1983.

J. T. Campbell. The Poisson correlation function. *Proceedings of the Edinburgh Mathematical Society*, 4(1):18–26, 1934.

L.Y. Chan, C.D. Lai, M. Xie, and T.N. Goh. A two-stage decision procedure for monitoring processes with low fraction nonconforming. *European Journal of Operational Research*, 150(2):420–436, 2003.

J.T. Chang and R.D. Fricker. Detecting when a monotonically increasing mean has crossed a threshold. *Journal of Quality Technology*, 31(2):217–234, 1999.

T.C. Chang and F.F. Gan. Cumulative sum charts for high yield processes. *Statistica Sinica*, 11(3):791–805, 2001.

J. R. Chimka. Gamma regressive individuals control charts for influenza activity. *Quality Engineering*, 21(2):182–189, 2009.

E. Çınlar. *Introduction to Stochastic Processes*. Prentice-Hall, 1975.

P.C. Cozzucoli and M. Marozzi. Monitoring multivariate Poisson processes: A review and some new results. *Quality Technology & Quantitative Management*, 15(1):53–68, 2018.

D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Springer, New York, 2003.

A. Di Bucchianico, M. Husková, P. Klástereckỳ, and W.R. van Zwet. Performance of control charts for specific alternative hypotheses. In J. Antoch, editor, *COMPSTAT 2004 Symposium*, pages 903–910, Heidelberg, 2004. Physica Verlag.

A. Di Bucchianico, G.D. Mooiweer, and E.J.G. Moonen. Monitoring infrequent failures of high-volume production processes. *Quality and Reliability Engineering International*, 21:521–528, 2005.

J-M. Dufour. Recursive stability analysis of linear regression relationships: An exploratory methodology. *Journal of Econometrics*, 19(1):31–76, 1982.

P.K. Dunn and G.K. Smyth. *Generalized Linear Models With Examples in R*. Springer Texts in Statistics. Springer New York, 2018. ISBN 9781441901187.

A. Easton, O. van Dalen, R. Göb, and A. Di Bucchianico. Bivariate copula monitoring. *Quality and Reliability Engineering International*, 38(3):1272–1288, 2022.

M. I. Flury and M. B. Quaglino. Multivariate EWMA control chart with highly asymmetric Gamma distributions. *Quality Technology & Quantitative Management*, 15(2):230–252, 2018.

M. Frisén. Optimal sequential surveillance for finance, public health, and other areas. *Sequential Analysis*, 28(3):310–337, 2009.

M. Frisén and P. Wessman. Evaluations of likelihood ratio methods for surveillance. *Communications in Statistics - Simulation and Computation*, 28(3):597–622, 1999.

J.S. Galpin and D.M. Hawkins. The use of recursive residuals in checking model fit in linear regression. *The American Statistician*, 38(2):94–105, may 1984.

F.F. Gan. An optimal design of EWMA control charts based on median run length. *Journal of Statistical Computation and Simulation*, 45(3-4):169–184, 1993.

C. Genest and J. Nešlehová. A primer on copulas for count data. *ASTIN Bulletin*, 37(2):475–515, 2007.

T.N. Goh. A control chart for very high yield processes. *Quality Assurance*, 13:18–22, 1987.

A. Haq and W. H. Woodall. A note on an average run length calculation for the ewma and other charts. *Quality and Reliability Engineering International*, 38(8):4351–4355, 2022.

W. Huang, M.R. Reynolds Jr., and S. Wang. A binomial GLR control chart for monitoring a proportion. *Journal of Quality Technology*, 44(3):192–208, 2012.

D. I. Inouye, E. Yang, G. I. Allen, and P. Ravikumar. A review of multivariate distributions for count data derived from the Poisson distribution. *WIREs Computational Statistics*, 9(3):e1398, 2017.

D. Jearkpaporn, D. C. Montgomery, G. C. Runger, and C. M. Borror. Process monitoring for correlated gamma-distributed data using generalized-linear-model-based control charts. *Quality and Reliability Engineering International*, 19(6):477–491, 2003.

A. KazemiNia, B.S. Gildeh, and Z. Abbasi Ganji. The design of geometric generalized likelihood ratio control chart. *Quality and Reliability Engineering International*, 34(5):953–965, 2018.

C. Kemp and A. Kemp. Some properties of the Hermite distribution. *Biometrika*, 52:381–394, 1965.

R. Kenett and M. Pollak. On assessing the performance of sequential procedures for detecting a change. *Quality and Reliability Engineering International*, 28(5):500–507, 2012.

M. Khakifirooz, V. G. Tercero-Gómez, and W. H. Woodall. The role of the normal distribution in statistical process monitoring. *Quality Engineering*, 33(3):497–510, 2021.

J. Lee and W. H. Woodall. A note on GLR charts for monitoring count processes. *Quality and Reliability Engineering International*, 34(6):1041–1044, 2018.

P. Lévy. Sur les exponentielles de polynomes. *Ann. École Normale Superieure*, 37:231–292, 1937.

T. Mahmood. Generalized linear model based monitoring methods for high-yield processes. *Quality and Reliability Engineering International*, 36(5):22, 2020.

B. J. Mandel. The regression control chart. *Journal of Quality Technology*, 1(1):1–9, 1969.

T.M. Margavio, M.D. Conerly, W.H. Woodall, and L.G. Drake. Alarm rates for quality control charts. *Statistics & Probability Letters*, 24(3):219 – 224, 1995.

P. McCullagh and J.A. Nelder. *Generalized Linear Models, Second Edition*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1989. ISBN 9780412317606.

C.A. McGilchrist and K.M. Matawie. Recursive residuals in generalised linear models. *Journal of Statistical Planning and Inference*, 70(2):335–344, 1998.

R.K. Milne and M. Westcott. Further results for Gauss-Poisson processes. *Advances in Applied Probability*, 4(1):151–176, 1972.

J.E. Moyal. The general theory of stochastic population processes. *Acta Mathematica*, 108:1–31, 1962.

B. Mühlig. Multivariate process monitoring based on copula structures. Master's thesis, Würzburg University, Germany, 2017.

L.S. Nelson. A control chart for parts-per-million nonconforming items. *Journal of Quality Technology*, 26(3):239–240, 1994.

D.S. Newman. A new family of point processes which are characterized by their second moment properties. *Journal of Applied Probability*, 7(2):338–358, 1970.

E. S. Page. Cumulative sum charts. *Technometrics*, 3(1):1–9, 1961.

K. Park, J-M. Kim, and D. Jung. GLM-based statistical control r-charts for dispersed count data with multicollinearity between input variables. *Quality and Reliability Engineering International*, 34(6): 1103–1109, 2018.

K. Park, Do. Jung, and J-M. Kim. Control charts based on randomized quantile residuals. *Applied Stochastic Models in Business and Industry*, 36(4):716–729, 2020.

A. Ramanayake and A.K. Gupta. Epidemic change model for the exponential family. *Communications in Statistics - Theory and Methods*, 33(9):2175–2198, 2004.

M.R. Reynolds Jr and J. Lou. An evaluation of a GLR control chart for monitoring the process mean. *Journal of Quality Technology*, 42(3):287–310, 2010.

S.C. Richards, W.H. Woodall, and G. Purdy. Surveillance of nonhomogeneous poisson processes. *Technometrics*, 57(3):388–394, 2015.

A. Sachlas, S. Bersimis, and S. Psarakis. Risk-adjusted control charts: theory, methods, and applications in health. *Statistics in Biosciences*, 11(3):630–658, 2019.

A. Saghir and Z. Lin. Control charts for dispersed count data: an overview. *Quality and Reliability Engineering International*, 31(5):725–739, 2015.

E. Santiago and J. Smith. Control charts based on the exponential distribution: Adapting runs rules for the t chart. *Quality Engineering*, 25(2):85–96, 2013.

W.A. Shewhart. Quality control charts. *Bell System Technical Journal*, 2:593–603, 1926.

K. Skinner, D.C. Montgomery, and G.C. Runger. Process monitoring for multiple count data using generalized linear model-based control charts. *International Journal of Production Research*, 41(6): 1167–1180, 2003.

K. Skinner, D. Montgomery, and G. Runger. Generalized linear model-based control charts for discrete semiconductor process data. *Quality and Reliability Engineering International*, 20:777 – 786, 12 2004.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8:229–231, 1959.

S. Sukparungsee, S. Kuvattana, P. Busababodhin, and Y. Areepong. Multivariate copulas on the MCUSUM control chart. *Cogent Mathematics*, 4(1):1342318, 2017.

J. Szarka and W.H. Woodall. A review and perspective on surveillance of high quality Bernoulli processes. *Quality and Reliability Engineering International*, 27:735–752, 10 2011.

A. Tartakovsky, I. Nikiforov, and M. Basseville. *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Taylor & Francis Group, Boca Raton, 2015.

E. Topalidou and S. Psarakis. Review of multinomial and multiattribute quality control charts. *Quality and Reliability Engineering International*, 25(7):773–804, 2009.

S. Vardeman and D. Ray. Average run lengths for cusum schemes when observations are exponentially distributed. *Technometrics*, 27(2):145–150, 1985.

G. Verdier. Application of copulas to multivariate control charts. *Journal of Statistical Planning and Inference*, 143(12):2151–2159, 2013.

D. Vere-Jones. Some applications of probability generating functionals to the study of input-output streams. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):321–333, 1968.

D. Vere-Jones. Stochastic models for earthquake occurrence. *Journal of the Royal Statistical Society: Series B (Methodological)*, 32(1):1–45, 1970.

A. Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 1945.

M. Westcott. The probability generating functional. *Journal of the Australian Mathematical Society*, 14 (4):448–466, 1972.

J. Wolfowitz. The efficiency of sequential estimates and Wald's equation for sequential processes. *The Annals of Mathematical Statistics*, 18(2):215–230, 1947.

W.H. Woodall. Control charts based on attribute data: Bibliography and review. *Journal of Quality Technology*, 29(2):172–183, 1997.

W.H. Woodall and A. Driscoll. Some recent results on monitoring the rate of a rare event. In S. Knoth and W. Schmid, editors, *Frontiers in Statistical Quality Control 11*, pages 15–27. Springer, 2015.

W.H. Woodall and F.W. Faltin. Rethinking control chart design and evaluation. *Quality Engineering*, 31 (4):596–605, 2019.

M. Xie, T. N. Goh, and X. S. Lu. A comparative study of CCC and CUSUM charts. *Quality and Reliability Engineering International*, 14(5):339–345, 1998.

M. Xie, T. Goh, and P. Ranjan. Some effective control chart procedures for reliability monitoring. *Reliab. Eng. Syst. Safety*, 77:143–150, 2002a.

M. Xie, T.N. Goh, and V. Kuralmani. *Statistical Models and Control Charts for High Quality Processes*. Kluwer, Boston, 2002b.

Y. Xie, M. Xie, and N. G. Thong. Two MEWMA charts for Gumbel's bivariate exponential distribution. *Journal of Quality Technology*, 43(1):50–65, 2011.

G. Yakovlev, J. B. Rundle, R. Shcherbakov, and D. L. Turcotte. Inter-arrival time distribution for the non-homogeneous Poisson process, 2005. URL https://arxiv.org/abs/cond-mat/0507657.

M. Zhang, Y. Peng, A. Schuh, F.M. Megahed, and W.H. Woodall. Geometric charts with estimated control limits. *Quality and Reliability Engineering International*, 29(2):209–223, 2013.

I. Zwetsloot, T. Mahmood, and W.H. Woodall. Multivariate time-between-events monitoring: An overview and some overlooked underlying complexities. *Quality Engineering*, 33(1):13–25, 2021.

Appendices

Appendix A

Statistical Monitoring Procedures for High-Purity Processes: A Hierarchical Simulation Structure in R

A.1 Introduction

The numerical results supporting the main conclusions of this thesis originate from an interconnected, hierarchical simulation structure for simplifying, automating, and efficiently reproducing the results. At the higher level, the order of the steps required to obtain any monitoring scheme's performance is the following

- Process simulation;
- (Optional) Statistic computation;
- Performance metrics estimation.

In the following sections, the simulation setup developed in R is reported per each item mentioned above, accompanied by a brief description of the files and functions.

A.2 Process Simulations

To streamline the simulation procedure of the underlying high-purity processes, the cumulative quantities, i.e., the total number of items or total time elapsed between two or more events, are simulated directly, circumventing the simulation of the original observations, modeled following the binomial or Poisson distributions, respectively.

In the sequel, the non-homogeneous Poisson process and the sequence of negative binomial observations simulation function are described. In general, the process simulation output consists of an $n \times m$ matrix **P** and a vector **I** of size n, where n is the number of iterations and m is the size of each simulation. Each row of the matrix **P** represents a sequence of observations X_i, i.e., an iteration. Therefore, the element X_{ij} represents the observation at time i, for the iteration j. Each element of the row vector **I** represents the time at which at change has occurred for the iteration j, i.e. the change-point. The matrix **P** can be readily passed as input to the functions for statistic computation or performance metric evaluation, while the vector **I** is needed for the calculation of the conditional metrics, for which the position of the change-point is important.

A.2.1 Non-homogeneous Poisson process simulation

Several methods exist to generate pseudo-random observations from a non-homogeneous Poisson process. The algorithm described in detail in Section 3.1, is exact and based on the time scale transformation of a non-homogeneous Poisson process (NHPP). The function `HPP_Simulation` requires in input the number of iterations, the length of each iteration, the order of the chart r, the fixed change-point in the process time scale, and the explicit inverse cumulative intensity function with its parameters. In addition, it returns the process simulation matrix **P**, while the function `index_NHPP` returns the index per each iteration at which the shift has happened, the vector **I**.

If the inverse cumulative intensity function is not available in an explicit form, the generic cumulative intensity and the inverse cumulative intensity functions can be used to solve the inversion numerically using the functions `Lambda.g` and `Lambda.inv.g`, respectively. The catalog also contains the intensity function for an abrupt change in the intensity parameter and for the epidemic shift described in Section 3.3.

```
##############################################################################
## High-Purity Process (HPP) Simulation
## Caterina Rizzo (crizzo@dow.com)
##############################################################################

##############################################################################
## Intensity functions catalog
## If the explicit inverse cumulative function is not available
## the user can select an intensity function and use the generic
## cumulative and inverse intensity functions to compute them
## numerically. This approach is more time-consuming.
##############################################################################

  # Generic (inverse and) cumulative intensity functions
  Lambda.g <- Vectorize(function(tupper, tau, lambda_in, lambda_out)
    integrate(function(x){lambda(x,tau,lambda_in, lambda_out)}, lower = 0, upper = tupper)$
        value)

  Lambda.inv.g <- Vectorize(function(y,tau,lambda_in, lambda_out){
    uniroot(function(x){Lambda.g(x,tau,lambda_in, lambda_out) - y}, c(0,100000))$root})

  # Explicit inverse cumulative intensity function (Unknown permanent shift)
  Lambda.inv.step <- function(s, tau, lambda_in, lambda_out){
    ifelse(s<=lambda_in*tau, s/lambda_in, (s-((lambda_in-lambda_out)*tau))/(lambda_out))
  }

  # Explicit inverse cumulative intensity function (Epidemic shift)
  Lambda.inv.step_EP <- function(s, tau1, tau2, lambda_in, lambda_out){
    if(s<=lambda_in*tau1){
      s/lambda_in
    } else if (s <= (lambda_out*tau2 - (lambda_out - lambda_in)*tau1)) {
      (s+(lambda_out-lambda_in)*tau1)/(lambda_out)
    } else {
      (s+((lambda_out-lambda_in)*tau1)+(lambda_in - lambda_out)*tau2)/(lambda_in)
    }
  }

##############################################################################
## Function: High-Purity Process (HPP) Simulation
## This function generate a matrix of simulations of specified size
## using the Time-Inverse transformation method.
## The output of the function is a list with the inter-arrival times matrix
```

```
## and a vector containing the first opportunity to detect the change.
#############################################################################

  HPP_Simulation <- function (iterations=100, size, r, shift, inv_func = Lamda.inv.step,
      ...){
    hpp_e <- matrix(rexp(size*iterations, rate = 1), nrow=iterations, byrow=T)
    hpp_t <- t(apply((hpp_e), 1, cumsum))
    nhpp_t <- t(apply(hpp_t, 1, function(x){sapply(x, function(y){inv_func(y, ...)})}))
    nhpp_tr <- t(apply(nhpp_t,1,function(x) x[seq(from = r, to=length(x), by = r)]))
    nhpp_e <- t(apply(nhpp_tr, 1, diff))
    return(nhpp_e)
  }

  index_NHPP <- function(data, shift){
    time <- t(sapply(1:nrow(data), function(x){cumsum(data[x,])}))
    index <- sapply(1:nrow(time),
                    function(x){ifelse(any(time[x,]>= shift), which(time[x,] >= shift)[1],
                        ncol(time))})
  }
```

Listing A.1: Non-homogeneous Poisson Process Simulations Supporting Functions.

A.2.2 Sequence of negative binomial observations simulation

Simulating negative binomial observations is easier than the continuous counterpart because of their discrete nature. The function NB_ProcessSimulation requires in input the number of iterations, the length of each iteration, the order of the chart, the change-point and the in-control and out-control parameters. The index of change is not random and fixed at the change-point in the control chart scale. The difference with the previous function is that the size of each iteration is not fixed nor depends on the realization of the random variable. Therefore, the output of the function is a list of iterations of different sizes, rather than a matrix.

```
#############################################################################
## Negative Binomial Process Simulation
## Caterina Rizzo (crizzo@dow.com)
#############################################################################

```

```
NB_ProcessSimulation <- function(iterations=100, size, r, shift, p_in, p_out){
  size_in <- size*shift
  size_out <- size - size_in
  sample_in <- matrix(rbinom(size_in*iterations, 1, p_in), nrow = iterations, byrow = T)
  sample_out<- matrix(rbinom(size_out*iterations, 1, p_out), nrow = iterations, byrow = T)
  sample <- cbind(sample_in, sample_out)
  events <- sapply(1:iterations, function(x){which(sample[x,] == 1)})
  sample_geom <- sapply(1:iterations, function(x){c(events[[x]][1], diff(events[[x]]))})
}
```

Listing A.2: Negative Binomial Process Simulations R Function.

A.3 Statistic computation

In most of the designs proposed in the thesis, the monitored variable is a transformation of the measured observations. Generally, the generic statistics computation function receives as input the raw observation matrix (**P**) and returns a new n $\times$ m matrix **G**. Each row of the matrix **G** represents a sequence of statistics Y_i.

A.3.1 The EWMA and CUSUM statistics

The `ewma.fun` and `cusum.fun` return the exponentially weighed moving average and the cumulative sum statistics, respectively. The `ewma.fun` function requires in input the matrix **P**, the weight w, the starting point z_0 and the type of chart (i.e., two-sided, one-sided lower, one-sided upper). The `cusum.fun` requires the matrix **P**, the tuning in-control and out-of-control parameters, the initialization value and the type of chart. The `cusum.fun` function is specifically designed for gamma-distributed data. The EWMA and CUSUM charts are described in detail in Section 1.1.

```
##############################################################################
## EWMA and CUSUM functions
## Caterina Rizzo (crizzo@dow.com)
##############################################################################


```

```
ewma.fun <- function(data, w, z0, type ="two-sided"){
  if(type == "two-sided"){
    z <- t(apply(data, 1, Reduce, f = function (v,y) {w*y+(1-w)*v}, init = z0, accumulate =
         TRUE))}
  if(type =="upper"){
    z <- t(apply(data, 1, Reduce, f = function(v,y) {max(z0, w*y+(1-w)*v)}, init = z0,
        accumulate = TRUE))}
  if(type == "lower"){
    z <- t(apply(data, 1, Reduce, f = function(v,y) {min(z0, w*y+(1-w)*v)}, init = z0,
        accumulate = TRUE))}
  return(z)
}


cusum.fun <- function(data, p_in, p_out, init, type ="two-sided"){

  k <- ((log(1/p_out)-log(1/p_in))/(1/p_out - 1/p_in))

  if(type == "two-sided"){
  s_pos <- t(apply(data, 1, Reduce, f = function(v,x) {max(0, v+(x-k))}, init = init,
      accumulate = TRUE))
  s_neg <- t(apply(data, 1, Reduce, f = function(v,x) {min(0, v+(x-k))}, init = init,
      accumulate = TRUE))
  return(list(s_pos, s_neg))}

  if(type =="upper"){s_pos <- t(apply(data, 1, Reduce, f = function(v,x) {max(0, v+(x-k))},
       init = init, accumulate = TRUE))
  return(s_pos)}

  if(type == "lower"){s_neg <- t(apply(data, 1, Reduce, f = function(v,x) {min(0, v+(x-k))
      }, init = init, accumulate = TRUE))
  return(s_neg)}
}


## Note: p_in and p_out are expressed in rates - not in means
```

Listing A.3: EWMA and CUSUM Functions R File.

A.3.2 Generalized likelihood ratio statistics

The R file `Gamma GLR Statistic Calculation Supporting Functions` contains the supporting functions for computing the generalized likelihood ratio statistics, for the simple-hypotheses, epidemic shift and indifference interval models, as discussed in Sections 3.1 and 3.3. The `GLR Stat Computation` function requires in input the process simulation matrix (**P**), the order of the chart, the in-control parameter and the change-point model type.

```
##############################################################################
## Gamma GLR statistic calculation supporting functions
## Caterina Rizzo (crizzo@dow.com)
##############################################################################

  ## Main function to select which type of GLR

  GLR_Stat_Computation <- function(sim, r, p0, m, d, type = "Unknown shift"){
    if(type == "Unknown shift"){
      return(GLR_Gamma(sim, r, p0, m))
    } else if (type == "Epidemic shift"){
      return(GLR_Gamma_ES(sim, r, p0, m))
    } else if(type == "Indifference Interval"){
      return(GLR_Gamma_II(sim,r, p0, m, d))
    }
  }

  ## Main function to calculate the GLR for gamma distributed variables

  GLR_Gamma <- function(sim, r, p0, m){
    if(m < ncol(sim)){
      sim <- sim[,(ncol(sim)-m):ncol(sim)]}
    matrix <- matrix(0, nrow = nrow(sim), ncol = ncol(sim))
    for (i in 1:nrow(sim)){
      k <- 1:length(sim[i,])
      matrix[i,] <- sapply(k, function(x){GLR_Fun_Gamma(sim[i,], x, r, p0)})}
    return(matrix)
  }

  GLR_Fun_Gamma <- function(sim, k, r, p0){
    cp <- 0:(k-1)
    p1 <- sapply(cp, function(x){MLE_Fun_Gamma(sim, k, x, r, p0)})
```

```
    glr <- sapply(cp, function(x){GLR_Stat_Gamma(sim, k, x, r, p1[x+1], p0)})
    return(max(glr))
  }

  ## MLE functions

  MLE_Fun_Gamma <- function(sim, k, cp, r, p0){
    mle <- (sum(sim[(cp+1):k], na.rm = TRUE))/(r*(k-cp))
    return(mle)
  }

  ## GLR Statistic function

  GLR_Stat_Gamma <- function(sim, k, cp, r, p1, p0){
    glr <- (r*(k-cp))*(log(p0/p1) + ((p1-p0)/p0))
    return(glr)
  }

##########################################################################
## Gamma GLR Statistic calculation supporting functions for the epidemic
## shift model
##########################################################################

  ### Main function, first and second change-point optimization

  GLR_Gamma_ES <- function(sim, r, pm, m){
    if(m < ncol(sim)){
      sim <- sim[,(ncol(sim)-m):ncol(sim)]}
    matrix <- matrix(0, nrow = nrow(sim), ncol = (ncol(sim)-1))
    for (i in 1:nrow(sim)){
      k <- 2:length(sim[i,])
      matrix[i,] <- sapply(k, function(x){GLR_Gamma_ES_T2(sim[i,], x, r, pm)})}
    return(matrix)
  }

  GLR_Gamma_ES_T2 <- function(sim, k, r, pm) {
    cp2 <- 1:(k-1)
    glr_mat <- sapply(cp2, function(x){GLR_Gamma_ES_Fun(sim, k, r, x, pm)})
    return(max(glr_mat))
  }


  GLR_Gamma_ES_Fun <- function(sim, k, r, cp2, pm){
```

```
76      cp <- 0:(cp2-1)
77      theta <- sapply(cp, function(x){MLE_GES_Theta(sim, k, r)})
78      eta <- sapply(cp, function(x){MLE_GES_Eta(sim, k, x, r, pm)})
79      xi <- sapply(cp, function(x){MLE_GES_Xi(sim, k, x, cp2, r, pm)})
80      psi <- sapply(cp, function(x){MLE_GES_Psi(sim, k, cp2, r, pm)})
81      glr <- sapply(cp, function(x){GLR_Stat_ES(sim, k, x, cp2, r, theta[x+1], eta[x+1], xi[x
            +1], psi[x+1])})
82      return(max(glr))
83    }
84
85    ### MLE ES functions
86
87    MLE_GES_Theta <- function(sim, k, r){
88      mle <- sum(sim[1:k], na.rm = TRUE)/(r*(k))
89    }
90
91    MLE_GES_Eta <- function(sim, k, cp1, r, pm){
92      ifelse(cp1 == 0, pm, max(sum(sim[0:cp1], na.rm = TRUE)/(r*(cp1)), pm))
93    }
94
95    MLE_GES_Xi <- function(sim, k, cp1, cp2, r, pm){
96      mle <- min(sum(sim[(cp1+1):cp2], na.rm = TRUE)/(r*(cp2-cp1)), pm)
97    }
98
99    MLE_GES_Psi <- function(sim, k, cp2, r, pm){
100     mle <- max(sum(sim[(cp2+1):k], na.rm = TRUE)/(r*(k-cp2)), pm)
101   }
102
103   ### GLR ES statistic function
104
105   GLR_Stat_ES <- function(sim, k, cp1, cp2, r, theta, eta, xi, psi){
106     glr <- r*(k*log(theta)-cp1*log(eta)-(cp2-cp1)*log(xi)-(k-cp2)*log(psi))
107     return(glr)
108   }
109
110
111 ##########################################################################
112 ## Gamma GLR Statistic calculation supporting functions for the
113 ## indifference interval model (one-sided case)
114 ##########################################################################
115
116   ### Main function and fist change-point optimization
117
```

```r
GLR_Gamma_II <- function(sim, r, p0, m, d){
  if(m < ncol(sim)){
    sim <- sim[,(ncol(sim)-m):ncol(sim)]}
  matrix <- matrix(0, nrow = nrow(sim), ncol = ncol(sim))
  for (i in 1:nrow(sim)){
    k <- 1:length(sim[i,])
    matrix[i,] <- sapply(k, function(x){GLR_Gamma_II_Fun(sim[i,], x, r, p0, d)})}
  return(matrix)
}


GLR_Gamma_II_Fun <- function(sim, k, r, p0, d){
  cp <- 0:(k-1)
  theta <- sapply(cp, function(x){MLE_GII_Theta(sim, k, r, p0, d)})
  eta <- sapply(cp, function(x){MLE_GII_Eta(sim, k, x, r, p0, d)})
  xi <- sapply(cp, function(x){MLE_GII_Xi(sim, k, x, r, p0, d)})
  glr <- sapply(cp, function(x){GLR_Stat_II(sim, k, x, r, theta[x+1], eta[x+1], xi[x+1])
      })
  return(max(glr))
}

### MLE II functions

MLE_GII_Theta <- function(sim, k, r, p0, d){
  mle <- min(((sum(sim[1:k], na.rm = TRUE))/r*(k)), p0-d)
}

MLE_GII_Eta <- function(sim, k, cp, r, p0, d){
  ifelse(cp == 0, p0, min((sum(sim[0:cp], na.rm = TRUE))/(r*(cp)), p0-d))
}

MLE_GII_Xi <- function(sim, k, cp, r, p0, d){
  mle <- min((sum(sim[(cp+1):k], na.rm = TRUE)/(r*(k-cp))), p0-d)
}

### GLR II statistic function

GLR_Stat_II <- function(sim, k, cp, r, theta, eta, xi){
  glr <- (r*(k*log(theta)-cp*log(eta)-(k-cp)*log(xi)))
  return(glr)
}
```

Listing A.4: Gamma GLR Statistic Calculation Supporting Functions R File.

A.3.3 Generalized linear model (residuals) statistics

The R file `Gamma GLM Residuals Calculation Supporting Functions` contains the supporting functions for the generalized linear model calculations, described in Section 4.2. It also includes a modified version of the non-homogeneous Poisson process simulation code to accommodate the change in the regression parameters and the co-variate (`Regression Simulation`). This function requires an additional list, `simul.param`, that contains the in-control and out-control regression parameters and co-variates values for a transient shift. The function returns a list with the co-variate matrix and the response variable matrix. The Pearson, deviance, and quantile residuals functions require the co-variate and response variable matrices (**X** and **Y**) and the list of model parameters. The recursive residuals also require the rolling regression coefficients estimates, which can be obtained using the `roll_glm` function.

```
##############################################################################
## Gamma GLM supporting functions
## Caterina Rizzo (crizzo@dow.com)
##############################################################################

##############################################################################
## Gamma GLM process simulation function neglecting the random effects
## of the explanatory variables
##############################################################################

##############################################################################
## Intensity functions catalog
##############################################################################

  # Explicit inverse cumulative intensity function

  Lambda.inv.step <- function(s, tau, lambda_in, lambda_out){
    ifelse(s<=lambda_in*tau, s/lambda_in, (s-((lambda_in-lambda_out)*tau))/(lambda_out))
  }

##############################################################################
## Function: High-Purity Process (HPP) Simulation
## This function generate a matrix of simulations of specified size using the Time inverse
    transformation method.
## The output of the function is a list with the inter-arrival times matrix and a vector
```

```
    containing the first opportunity to detect the change.
################################################################################

  Regression_Simulation <- function (iter, size, r, shift, sim.param){

    v0.in = sim.param[1]
    v1.in = sim.param[2]
    v0.out = sim.param[3]
    v1.out = sim.param[4]
    Xv1 = sim.param[5]
    Xv2 = sim.param[6]

    lambda_1 <- 1/exp(v0.in + v1.in*Xv1)
    lambda_2 <- 1/exp(v0.out + v1.out*Xv2)

    hpp_e <- matrix(rexp(size*iter, rate = 1), nrow=iter, byrow=T)
    hpp_t <- t(apply((hpp_e), 1, cumsum))
    nhpp_t <- t(apply(hpp_t, 1, function(x){inv_func(x, tau = shift, lambda_in = lambda_1,
        lambda_out = lambda_2)}))
    nhpp_tr <- t(apply(nhpp_t,1,function(x) x[seq(from = r, to=length(x), by = r)]))
    Y <- t(apply(nhpp_tr, 1, diff))
    time <- t(sapply(1:nrow(Y), function(x){cumsum(Y[x,])}))
    index <- sapply(1:nrow(time),
                    function(x){ifelse(any(time[x,]>= shift), which(time[x,] >= shift)[1],
                        ncol(time))})

    X = matrix(data = 0, nrow = iter, ncol = (size-1)/r)
    for (k in 1:nrow(Y)){
      X[k, 1:index[k]] = Xv1
      X[k, (index[k]+1):(size-1)/r] = Xv2}

    list(Y, X, index)
  }


################################################################################
## Gamma GLM residuals calculations
################################################################################

  # Pearson residuals
    pearson.residuals = function (X, Y, mod.param) {
      v0 = mod.param[1]
      v1 = mod.param[2]
```

```
65      r = mod.param[3]
66      Yp = exp(v0 + v1*X)
67      rp = (Y-Yp)/(Yp)
68      return(rp)
69    }
70
71  # Deviance residuals
72    deviance.residuals = function (X, Y, mod.param) {
73      v0 = mod.param[1]
74      v1 = mod.param[2]
75      r = mod.param[3]
76      Yp = exp(v0 + v1*X)
77      res = (Y-Yp)
78      d = 2*(-log(Y/Yp) + res/Yp)
79      rd = sign(res)*sqrt(d)
80      return (rd)
81    }
82
83  # Randomized Quantile residuals
84    quantile.residuals = function (X, Y, mod.param) {
85      v0 = mod.param[1]
86      v1 = mod.param[2]
87      r = mod.param[3]
88      Yp = exp(v0 + v1*X)
89      cdf = pgamma(Y, scale = Yp/r, shape = r)
90      rq = qnorm(cdf)
91      return(rq)
92    }
93
94  # Recursive Residuals
95    recursive.residuals = function (X, Y, r, RegressionCoefficients) {
96      v0 = RegressionCoefficients[[1]]
97      v1 = RegressionCoefficients[[2]]
98      Y = Y[,-c(1:stup)]
99      Yp = t(sapply(1:nrow(Y), function(y) {sapply(2:ncol(Y), function(x) exp(v0[y, x-1] +
            v1[y, x-1]*X[y, x])/r)}))
100     Y = Y[,-1]
101     rr = (Y-Yp)/(Yp)
102     return(rr)
103   }
104
105 # Regression Coefficients Base Functions
106
```

```
    roll_glm = function(X, Y, stup){
      roll_glm <- lapply(seq(stup, length(Y)), function(x){glm( Y[1:x] ~ X[1:x], family =
          Gamma(link = 'log'))})
    }

  # Regression Coefficients

    regression.coefficients = function(X_reg, Y_reg, stup){
      roll <- t(sapply(1:nrow(Y_reg), function(x){roll_glm(X_reg[x,], Y_reg[x,], stup)}))
      v0 <- matrix(unlist(lapply(1:length(roll), function(x){roll[[x]]$coefficients[1]})),
          nrow = iter, ncol = size)
      v1 <- matrix(unlist(lapply(1:length(roll), function(x){roll[[x]]$coefficients[2]})),
          nrow = iter, ncol = size)
      RegressionCoefficients <- list(v0, v1)
    }
```

Listing A.5: `Gamma GLM Residuals Calculations Supporting Function R File.`

A.4 Performance metrics estimation

The performance metric functions require as input the process simulation matrix **P** and, if applicable, the statistic matrix **G** and return the desired performance metrics or the control limits. The catalog of functions to calculate several performance metrics and/or their distribution for raw observation or derived statistics are contained in the R file `Performance measures supporting functions`.

```
##############################################################################
## Performance metrics supporting functions
## Caterina Rizzo (crizzo@dow.com)
##############################################################################

##############################################################################
## The h_solve and q_solve functions return the control limit(s) based on the
## ALI and the optimization function in input
##############################################################################

  h_solve <- function(y, pfun, SearchRange, ...){
    uniroot(function(x){pfun(x, ...) - y}, SearchRange, f.lower = -100000, f.upper =
```

```
        100000, extendInt = "yes")$root}

  q_solve <- function(y, pfun, SearchRange, ...){
    uniroot(function(x){pfun(x, ...) - y}, SearchRange, extendInt = "yes")$root}

##########################################################################
## Functions to calculate the ARL and ALI values and distributions
## The averageRunLength function is for all types of charts
## The inspectionRun_Stat function is for EWMA, CUSUM and GLR charts
## The inspectionRun is for Shewhart charts
##########################################################################

  averageRunLength <- function(origin, runRulesFun, ...){
    minIndex <- runRulesFun(data = origin, ...)
    runLength <- sapply(1:nrow(origin), function(x){ifelse(is.finite(minIndex[x]), minIndex
        [x], NA)})
    mean(runLength, na.rm = T)
  }

  inspectionRun <- function(type = "Shewhart", ...){
    if(type == "Shewhart"){
      inspectionRun <- inspectionRun_Shewhart(...)
    } else if (type == "Statistic"){
      inspectionRun <- inspectionRun_Stat(...)
    }
    return(inspectionRun)
  }

  inspectionRun_Shewhart <- function(origin, runRulesFun, ...){
    minIndex <- runRulesFun(data = origin, ...)
    inspection_length <- sapply(1:nrow(origin), function(x){ifelse(is.finite(minIndex[x]),
        sum(origin[x, 1:minIndex[x]]), NA)})
    n.na <- sum(is.na(inspection_length))
    mean(inspection_length, na.rm = T)
  }

  inspectionRun_Stat <- function(statistic, origin, runRulesFun, ...){
    minIndex <- runRulesFun(statistic, ...)
    inspection_length <- sapply(1:nrow(origin), function(x){ifelse(is.finite(minIndex[x]),
        sum(origin[x, 1:minIndex[x]]), NA)})
    n.na <- sum(is.na(inspection_length))
    return(mean(inspection_length, na.rm = T))
  }
```

```
52
53  inspectionRun_Shewhart_Distribution <- function(origin, runRulesFun, ...){
54    minIndex <- runRulesFun(origin, ...)
55    inspection_length <- sapply(1:nrow(origin), function(x){ifelse(is.finite(minIndex[x]),
          sum(origin[x, 1:minIndex[x]]), NA)})
56  }
57
58  inspectionRun_Stat_Distribution <- function(statistic, origin, runRulesFun, ...){
59    minIndex <- runRulesFun(statistic, ...)
60    inspection_length <- sapply(1:nrow(origin), function(x){ifelse(is.finite(minIndex[x]),
          sum(origin[x, 1:minIndex[x]]), NA)})
61  }
62
63 ##########################################################################
64 ## Functions to calculate the CED values and distributions
65 ## condExpDelay_Stat function is for EWMA, CUSUM and GLR charts
66 ## condExpDelay_Shewhart function is for Shewhart data
67 ## condExpDelay is the supporting function for both metrics
68 ##########################################################################
69
70  condExpDelay_fun <- function(type = "Shewhart", ...){
71    if(type == "Shewhart"){
72      condExpDelay <- condExpDelay_Shewhart(...)
73    } else if (type == "Statistic"){
74      condExpDelay <- condExpDelay_Stat(...)
75    }
76    return(condExpDelay)
77  }
78
79  condExpDelay_Shewhart <- function(origin, index, runRulesFun, ...){
80    minIndex <- runRulesFun(data = origin, ...)
81    expected_delay <- condExpDelay(minIndex, origin, index)
82    CED.na <- sum(is.na(expected_delay))
83    return(mean(expected_delay, na.rm = T))
84  }
85
86
87  condExpDelay_Stat <- function(statistic, origin, index, runRulesFun, ...){
88    minIndex <- runRulesFun(statistic, ...)
89    expected_delay <- condExpDelay(minIndex, origin, index)
90    CED.na <- sum(is.na(expected_delay))
91    return(list(mean(expected_delay, na.rm = T), sd(expected_delay, na.rm = T), CED.na))
92  }
```

```
93
94    condExpDelay_Shewhart_Dist <- function(origin, index, runRulesFun, ...){
95      minIndex <- runRulesFun(origin, ...)
96      expected_delay <- condExpDelay(minIndex, origin, index)
97    }
98
99    condExpDelay_Stat_Dist <- function(statistic, origin, index, runRulesFun, ...){
100     minIndex <- runRulesFun(statistic, ...)
101     expected_delay <- condExpDelay(minIndex, origin, index)
102   }
103
104   condExpDelay <- function(minIndex, data, index_array){
105     CED_array <- sapply(1:nrow(data),
106                         function(x){ifelse(minIndex[x] > index_array[x] & is.finite(
                                minIndex[x]), sum(data[x, (index_array[x]+1):minIndex[x]]), NA
                                 )})
107     return(CED_array)
108   }
109
110 ##########################################################################
111 ## Functions to calculate the PSD
112 ## The probDet_Metric function works for all types of charts
113 ## The probDet function is the supporting function
114 ##########################################################################
115
116   probDet_Metric <- function(origin, index, index_arrayd, runRulesFun, ...){
117     minIndex <- runRulesFun(data = origin, ...)
118     probDet_array <- probDet(minIndex, origin, index, index_arrayd)
119     return(mean(probDet_array))
120   }
121
122   probDet <- function(minIndex, data, index_array, index_arrayd){
123     probDet_Array <- sapply(1:nrow(data),
124                         function(x){ifelse(minIndex[x] > index_array[x] & minIndex[x] <= (
                                index_array[x]+ index_arrayd[x]) & is.finite(minIndex[x]), 1,
                                0)})
125     return(probDet_Array)
126   }
127
128 ##########################################################################
129 ## Supporting functions for quantile limits calculations
130 ##########################################################################
131
```

```
  propFunc <- function(data, type = "lower", L){
    if (type == "lower"){
      prop <- sum(data < L)/(length(data))}
    if (type == "upper"){
      prop <- sum(data > L)/(length(data))}
    return(mean(prop))
  }

#########################################################################
## Supporting functions that, given the chosen rule, return the first
## position at which a violation occurs
#########################################################################

  # Applied to the simulation matrix

  runRulesMinL <- function (data, L){
    LCL <- mean(data, na.rm = T) - L*sd(data, na.rm = T)
    UCL <- mean(data, na.rm = T) + L*sd(data, na.rm = T)
    return(sapply(1:nrow(data),
           function(x) {min(LowerShewhart(data[x, ], LCL), UpperShewhart(data[x, ], UCL))})
               )
  }

  runRulesMin <- function (data, LCL, UCL){
    sapply(1:nrow(data),
           function(x) {min(LowerShewhart(data[x, ], LCL), UpperShewhart(data[x, ], UCL))})
  }

  lowerRule <- function (data, LCL){
    sapply(1:nrow(data),
           function(x) (LowerShewhart(data[x, ], -LCL)))
  }

  upperRule <- function (data, UCL){
    sapply(1:nrow(data),
           function(x) (UpperShewhart(data[x, ], UCL)))
  }

  # Applied to one vector
  # Western Electric Rule 1 (1 point outside the control limits)

  UpperShewhart <- function(data, UCL){
    results <- data > UCL
```

```
174   ifelse(any(results),which.max(results), Inf) }
175
176 LowerShewhart <- function(data, LCL){
177   results <- data < LCL
178   ifelse(any(results), which.max(results), Inf) }
179
180 # Western Electric Rule 2 (if runlength = 9, 9 points in a row on same side of the centre
        line)
181 SameSideCL <- function(data,CL,runlength){min(InaRowBelowCL(data,CL,runlength),
      InaRowAboveCL(data,CL,runlength))
182 }
183
184 # Western Electric Rule 3 (all increasing, if runlength = 6, 6 points in a row all
      increasing)
185 AllIncreasing <- function(data,runlength){
186     pattern <- (diff(data) > 0)
187
188
189 results <- (rle(pattern)$lengths >= runlength-1)  & rle(pattern)$values
190
191 indexrle <- ifelse(any(results),which.max(results), Inf)
192
193 ifelse(indexrle == Inf, Inf, ifelse(indexrle==1,runlength,sum(rle(pattern)$lengths[1:(
      indexrle-1)])+runlength))
194 }
195
196 runRulesMin <- function (data, LCL, CL){
197   minIndex <- sapply(1:nrow(data),
198                      function(x) {min(LowerShewhart(data[x, ], LCL), SameSideCL(data[x,
                           ], CL, 9), AllIncreasing(data[x,], 6))})
199   return(minIndex)
200 }
```

Listing A.6: Performance Measures Supporting Functions R File.

Samenvatting

Productieprocessen in de chemische industrie zijn complex, maar de technologische vooruitgang in het huidige “Industry 4.0” tijdperk helpen fabriekspersoneel de inherente lagen van complexiteit van chemische processen te kwantificeren en doorgronden. Een groeiend aantal productiesystemen worden gemonitord door digitale apparaten die continu afwijkingen meten om een indicatie te geven van proceskwaliteit en efficiëntie. Waar “Statistical Process Control (SPC)” technieken nog wijdverspreid zijn in industrie om procesvariabiliteit te meten en afwijkingen ten opzichte van “in-control” of verwachte condities te meten, zijn de traditionele methodes die vaak gebruikt worden niet in staat om de meer geavanceerde scenario’s te volgen. De resulterende afwijkingen van de beoogde karakteristieken van productkwaliteit kunnen leiden tot off-grade materiaal, met financiële en ecologische gevolgen. Productieprocessen met een hoge mate van efficiëntie vormen een grotere uitdaging voor kwailteitscontrole door de extra lagen van complexiteit. Deze doctorale scriptie onderzoekt een aantal praktische statistische uitdagingen in het monitoren van complexe processen met een hoge productzuiverheid. De scriptie heeft als uitgangspunt de begrippen “process time” en “control chart time”, allebei kenmerkende eigenschappen van het monitoring kader voor processen met hoge productzuiverheid. Gebaseerd op deze concepten worden conditionele statistische prestatieparameters van procescontrole aangepast aan het monitoring kader voor processen met hoge productzuiverheid voor een eerlijke evaluatie en vergelijking met de nieuwe voorgestelde monitoringstrategiën. Deze acties worden gebruikt om toepassing-specifieke “control charts” gebouwd op “composite change-point”-modellen te kunnen vergelijken met de traditionele alternatieven. Deze flexibele “charts” zijn gebaseerd op “Generalized Likelihood Ratios (GLR)”. De voorgestelde modellen zijn ontwikkeld vanuit de industrie en representeren twee industriële scenario’s uit de praktijk: het “indifference interval” en “epidemic shift” modellen. Het eerste beschrijft de situatie waarin de gemonitorde prestatieparameter verandert met een hoeveelheid die als insignificant wordt beschouwd. De tweede beschrijven

een tijdelijke toestand van falen, kenmerkend voor feedback controlesystemen. In beide gevallen presteert de specialistische aanpak beter dan de alternatieven door een snellere detectie. Een ander aspect dat vaak ontbreekt in literatuur is de invloed die een covariaten-set heeft op de kwaliteitsparameter. “Generalized Linear Models (GLM)”-gebaseerde “control charts” kunnen worden gebruikt om contextuele afwijkingen te identificeren in systemen waarin de verwachte fluctuaties van een aantal procesvariabelen (zoals bijvoorbeeld gemeten gewicht of volume) een significante invloed hebben op de uitkomstvariabele. “Predictive residuals” laten uitstekende prestaties zien in het detecteren van aanhoudende veranderingen ten opzichte van de in-control condities tijdens “real-time” monitoring, terwijl “recursive residuals” gebruikt kunnen worden in de vroege stadia van het monitoring proces of voor “short run schemes”. Het relatief jonge onderzoeksveld van “multivariate point process control charts” wordt uitgebreid met monitoring van een bivariate Poisson proces beschreven door de Hermite distributie, op een continue schaal. Deze aanpak is adequaat wanneer het productieproces beschreven kan worden door meerdere gecorreleerde temporele puntmodellen met simpele correlatiestructuren.

Acknowledgements

First and foremost, my deepest gratitude goes to my supervisor, Alessandro, for his technical guidance, constant flow of insights and ideas, and an unpaired positive attitude; an example that goes beyond scientific supervision. I am thankful for his continuous encouragement and support; his trust made me feel less of an outsider in statistics and the academic world, a boost of confidence in my professional life.

Furthermore, I extend my sincere gratitude to the doctorate defense committee members for reviewing the dissertation and providing their valuable feedback. Their insights and comments definitely improved the thesis.

I acknowledge Dow's support and the people offering unique industrial insights and suggestions. This perspective helped contextualizing the research work and framing it into an industry-driven application; it was not always easy. Among those, I would like to mention Ruben 't Lam and Swee-Teng Chin for initiating this collaboration and May Roca for the continuous and ineffable support.

I want to thank my Dow Terneuzen colleagues who enriched my experience during these years, providing a stimulating working environment and supporting me in creative ways.

Lastly, I express my gratitude to family and friends who have accompanied me intellectually and emotionally, alleviating my academic journey and making it easier, worthwhile, and fulfilling.

About the author

Caterina Rizzo was born in Italy. She completed her secondary education at Liceo Classico Bruno Vinci in Nicotera before starting her studies in Chemical Engineering at the University of Calabria (IT). After obtaining her bachelor's degree, she continued her education by pursuing a master's degree in Chemical Engineering. During her master's studies, she spent an academic year at the University of Leeds (UK), at the Institute of Particle Science and Engineering in collaboration with Procter & Gamble. Before graduation, she engaged in an extracurricular internship at Dow, in a joint initiative between Plastics R&D and the Statistics group. She joined the Chemometrics, AI and Statistics group at Dow in 2018 as Technology Specialist. Later that year, she started her PhD journey at the department of Mathematics and Computer Science at the Eindhoven University of Technology (NL), under the supervision of prof. Alessandro Di Bucchianico with a focus on statistical process monitoring. In her role at Dow, Caterina supports manufacturing operations via data-driven domain expertise. She collaborates with Operations, Analytical, R&D, and other subject matter experts to identify the root causes and propose solutions to manufacturing problems via fundamental problem-solving methodology and data science.