Sufficient reduction methods for multivariate health surveillance

Sawaporn Siripanthana

Thesis submitted for the degree of
Doctor of Philosophy

School of Mathematics and Statistics
The University of Sheffield
September 2013

Abstract

Surveillance systems aim to detect sudden changes or aberrations in data series which might signal the possibility of disease outbreaks. Early detection with a low false alarm rate (FAR) is the main aim of outbreak detection as used in health surveillance or in regard to bioterrorism. Multivariate surveillance is preferable to univariate surveillance since correlation between series (CBS) is recognized and incorporated and so small but consistent shifts are more likely to be detected. In this thesis, sufficient reduction (SR) methods are considered. These are dimensionality reduction tools for multivariate surveillance which have proved promising for handling CBS, and lag between change points (LCP), but have not previously been used when correlation within series (CWS) is present.

We develop SR methods for reducing a p-dimensional multivariate series to a univariate series of statistics shown to be sufficient for monitoring a sudden, but persistent, shift in a multivariate process of normal or Poisson data. CBS, CWS and LCP are all taken into account, as health data typically exhibit these forms of association. Different types of change point and shift sizes are investigated. A standard one-sided EWMA chart is used as a detection tool. Due to the nature of health data, the one-sided EWMA chart is modified for independent Poisson data and to allow for CWS in normal and Poisson processes. The performance of the proposed method is compared with existing SR and parallel methods. A simulation study shows that the proposed method is superior, giving a shorter delay and a lower FAR than other methods which have high FARs when CWS is clearly present. Although their high FAR can be improved by using a suitably modified EWMA chart, the proposed method still gives shorter delays than the others. The implementation of the proposed methods is illustrated with four real data sets.

Acknowledgements

I would like to express my appreciation and gratitude to people for their contributions to this thesis. First and foremost, I offer my sincerest gratitude to my supervisor, Dr Eleanor Stillman, who has supported me kindly. This thesis would not have been possible without her great help, valuable guidance and insightful comments. It has been a honor to work with and learn from her. I would also like to thank Dr Nick Fieller, Dr Kostas Triantafyllopoulos and the staff at the School of Mathematics & Statistics, the University of Sheffield, for their useful support and guidance during my great time studying the PhD.

I am very thankful to the Royal Thai Government and the Ministry of Science and Technology, Thailand, for their financial support throughout my study. I would also like to thank the Faculty of Science at Naresuan University, Thailand, for giving me such a great opportunity to study in the UK.

I would like to acknowledge the Swedish Radiation Safety Authority and Dr Eric Järpe for the use of the Swedish radiation data, the Swedish Meteorological and Hydrological Institute for the snow depth data, the Health Protection Agency North West Regional Epidemiology Unit and Dr Paul Clearly for the Scarlet fever notification data, and Dr Sotiris Bersimis for the Greek Pollution data. Without their help, suggestions and useful information, this thesis would not have been completed. I am also indebted to people, who have not been named here, for their friendly, sincere help and useful suggestions for my PhD work.

Last of all my special thanks go to my parents, Mr Songkram and Mrs Khemthong Siripanthana, and my sisters for their infinite love, support and encouragement throughout my study. I also would like to thank my friends both in Thailand and the UK for their help and support. It has been a memorable journey.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

Public health surveillance has been defined by the Centers for Disease Control and Prevention (CDC) as 'the ongoing, systematic collection, analysis, and interpretation of data (e.g., regarding agent/hazard, risk factor, exposure, health event) essential to the planning, implementation, and evaluation of public health practice, closely integrated with, the timely dissemination of these data to those responsible for prevention and control (CDC, 2009). Public health organizations have been emphasizing focus on public health surveillance since the exposure of anthrax in the USA in 2001 and subsequent outbreaks of other communicable diseases such as severe acute respiratory syndrome (SARS) in 2002, avian influenza (bird flu) in 2003 and swine flu in 2009 (Goldenberg *et al.*, 2002; Fienberg and Shmueli, 2005; Shmueli and Fienberg, 2006; Rolka *et al.*, 2007). Another purpose in developing such systems is the possibility of detecting evidence of bioterrorism, since it is not easy to know when an attack has occurred until huge numbers of people become sick or obvious symptoms are clearly presented (Stoto *et al.*, 2004; Fienberg and Shmueli, 2005; Rolka *et al.*, 2007; Shmueli and Burkom, 2010). In this usage, the monitoring is referred to "biosurveillance" (Burkom *et al.*, 2004; Fienberg and Shmueli, 2005; Shmueli and Fienberg, 2006; Fricker *et al.*, 2007; Rolka *et al.*, 2007).

The main purposes of surveillance systems are to monitor and be able to detect an occurrence of an outbreak as soon as possible after it has started. The sooner the public health organization can detect occurrence of a disease outbreak, the more people will be safe from disease transmission. The performance of such system relies on how quickly it can signal the disease outbreak and how accurate it is (i.e. a short delay in detection with low false alarm rate is required) (Fienberg and Shmueli, 2005; Frisén, 2005; Shmueli and Fienberg, 2006; Rolka *et al.*, 2007). Several statistical techniques, such as statistical process control methods and statistical modelling techniques have been implemented in practice. Research in this area has focused on improving the performance of outbreak detection which is practically applicable for public health surveillance (Sonesson and Bock, 2003; Frisén, 2005; Woodall, 2006). However, no gold standard methods or criteria for public health surveillance have been agreed. Due to the nature of disease data, which are noisy and vary over time, the statistical methods proposed for a particular

disease or under particular circumstances might not be applicable in another due to the violation of assumptions or changed circumstances. (Strat, 2005; Fienberg and Shmueli, 2005; Rolka *et al.*, 2007; Shmueli and Fienberg, 2006; Shmueli and Burkom, 2010; Unkel *et al.*, 2012)

Public health surveillance can be categorized depending on the purposes of use and the numbers of series monitored (see more detail in section 2.1.1). Univariate surveillance monitors a univariate series or several series separately, while multivariate surveillance monitors the p dimensional multivariate series together. The advantages of multivariate over univariate surveillance is that the correlation between series is taken into account, thus small shifts in series are more likely to be detected (MacGregor, 1995; Mohtashemi *et al.*, 2007; Montgomery, 2009; Shmueli and Burkom, 2010; Unkel *et al.*, 2012). Having reviewed the advantages and disadvantages of various statistical methods used in multivariate surveillance (chapter 2) and considered the limitations of health data which are dependent over time, this thesis proposes sufficient reduction methods, which are based on the dimensionality reduction and the sufficiency principle, for detecting a mean shift in a multivariate process of autocorrelated data.

According to the sufficiency property, the p dimensional multivariate series is reduced to a univariate series of likelihood ratio statistics summarizing all relevant information from the original series. The likelihood ratio statistics are derived from the likelihood ratio between in control and out of control stages defined according to a change point of the process. By the factorization theorem, a univariate series of likelihood ratio statistics has been proved to be sufficient for monitoring a shift in the multivariate process (Wessman, 1998; Frisén *et al.*, 2011). As a result of the sufficient reduction, the univariate series can be easily monitored with a univariate control chart in order to avoid the complexity of using multivariate control charts (see more discussion of multivariate control chart in chapter 2). Sufficient reduction methods have been proposed under various assumptions and the main aim of the thesis is to weaken these assumptions and so to broaden the applicability of sufficient reduction methods in public health surveillance. More details of sufficient reduction methods are provided in chapter 4. The structure of the thesis is summarized briefly as follows.

1.2 Thesis structure

Chapter 2 gives the background to public health surveillance and how it can be categorized according to the purpose of use and the data used. The use of real data and simulated data in system evaluation is also discussed. Statistical methods used for univariate and multivariate surveillance are reviewed with some limitations discussed. The conclusion summarizes the limitations of statistical methods used for multivariate surveillance and the rationale as to why sufficient reduction methods would be preferable for detecting a shift in a multivariate process.

Since rapid and correct identification is desired, chapter 3 reviews the measures used to evaluate system performance in public health surveillance. Since there are many sta-

tistical methods used and proposed for health surveillance, possible optimality criteria are also considered in order to choose the optimal methods for the surveillance. Having reviewed the measures, those used to evaluate the performance of sufficient reduction methods proposed in this thesis are defined.

In chapter 4, the background of the sufficiency property and the existing sufficient reduction (SR) methods proposed by Wessman (1998) and Frisén *et al.* (2011) are reviewed. The limitations of the existing SR methods proposed for detecting a shift in a multivariate process of independent observations are discussed and an extension permitting such methods to detect a shift in a process of (normally distributed) dependent observations in public health surveillance is presented. Correlation within series, correlation between series and various types of change point are incorporated. System evaluation is conducted by simulation study, using a standard one-sided EWMA chart as a detection tool. The performance of the proposed SR methods is compared with those of the existing SR methods.

According to the results from chapter 4, the existing SR methods produce high false alarm for detecting a shift in an autocorrelated process when a standard one-sided EWMA chart is used. Therefore, in chapter 5, the one-sided EWMA chart is modified to account for the autocorrelation in a process in order to see the improvement in the detection performance of the existing SR methods for detecting a shift in the autocorrelated process. Apart from this, the robustness of the SR methods against misspecified parameters used in the sufficient reduction, such as correlation within series, correlation between series, types of change points and shift size we want to detect, is conducted using a simulation study.

In chapter 6, due to the nature of sparse disease data which are low counts, the SR methods have been extended for detecting a shift in a process of Poisson data. However, the assumptions made for the SR methods for normal data cannot be applied in the Poisson case (see the discussion in chapter 6). One-sided EWMA charts for detecting a shift in a process of both independent and dependent Poisson data are modified. The detection performance of the SR methods proposed for Poisson data are evaluated via simulation study.

Chapter 7 provides four case studies implementing SR methods with four real data sets (influenza mortality data (USA), scarlet fever notifications (UK), Greek pollution data (Greece) and Swedish radiation data (Sweden)). Due to the assumptions made for the SR methods, some data sets cannot be used directly and some preliminary analysis is needed. The limitations of using real data for system evaluation is also discussed.

Chapter 8 discusses the performance of SR methods for detecting a mean shift in different scenarios. Their limitations are also discussed with suggestions for the further study and how the SR methods can be improved for public health surveillance.

Chapter 2

Statistical methods used in health surveillance

This chapter gives an overview of public health surveillance and reviews the statistical methods used in the area. According to the definition given in chapter 1, public heath surveillance can be categorized in different ways depending on the purposes of the surveillance and the type of health data. The details are provided in section 2.1 with some considerations in using health data in public health surveillance. Section 2.2 provides an overview of statistical methods used for univariate and multivariate surveillance with details of the practical methods in use. More details of commonly used statistical methods or those proposed in research studies are given along with some of their limitations. In this chapter, the methods used were roughly categorized into four groups according to the calculation procedure and implementation. The four groups are statistical process control (SPC) methods, those based on likelihood ratio methods, disease modelling (time series techniques and statistical modelling) and other methods (spatio-temporal statistics and scan statistics, etc.). The details of each method are provided separately in the subsequent sections (sections 2.3 - 2.6). Section 2.7 summarizes the uses of statistical methods for detecting a shift in a multivariate process with the considerations of how they can be developed for the purpose of outbreak detection in public health surveillance.

2.1 Overview of public health surveillance

2.1.1 Types of public health surveillance

2.1.1.1 Disease surveillance and syndromic surveillance

Traditional or disease surveillance has been used by public health organizations for monitoring aberrations which might give a possible signal of an outbreak. The numbers of confirmed cases or diagnosed cases, mortality or morbidity rates recorded weekly or monthly from hospitals or health centres are monitored and compared regularly with thresholds or alarm limits calculated from historical data (Stroup *et al.*, 2004; Meynard *et al.*, 2008). With regard to the use of diagnosed cases in disease surveillance, the performance of outbreak detection of such system depends on the accuracy of reported

cases and reporting delays of such data. Normally, reported cases are diagnosed by general practitioners in different hospitals or health centres, the different definitions of case diagnosis may cause under or over reported cases. The delays in confirming results by laboratories also causes a delay in outbreak detection (Fienberg and Shmueli, 2005).

As an alternative, syndromic surveillance has been developed in order to avoid reporting delays when using numbers of confirmed cases or diagnosed cases in disease surveillance. Syndromic surveillance is also an on-going system used to monitor aberrations for a possibility of a disease outbreak. Rather than using numbers of reported cases as in disease surveillance, syndromic surveillance uses pre-diagnosis data and health-related data including use of health care services for monitoring aberrations as a clue or indication of a possibility of a disease outbreak (Burkom, 2003; Elliot, 2010; Shmueli and Burkom, 2010; Unkel *et al.*, 2012).

Although the timeliness of syndromic surveillance is likely to be better than disease surveillance, as the (variable) delays in awaiting confirmed or diagnosed cases is avoided, the performance of such a system depends on the reliability of health-related data and pre-diagnosis data which may contain lots of noise and might not be representative of the public health.

2.1.1.2 Retrospective and prospective surveillance

Public health surveillance can also be categorized into two different groups according to the purpose of the surveillance. Retrospective surveillance uses health data to explain the course of a disease outbreak and how the disease has been transmitted. The aims of retrospective surveillance are to know and be aware of the variation and the development of the disease nature, which can be used by the organization to develop health policy for future disease or outbreak prevention (Sonesson and Bock, 2003). Several statistical methods such as time series and statistical modelling are used to describe the variation in disease nature, whereas spatial and scan statistics are used to investigate the pattern of disease clustering and disease mapping (Lawson and Kleinman, 2005).

Unlike retrospective surveillance, the main aim of prospective surveillance is to monitor aberrations for outbreak detection purposes. In prospective surveillance, health data are used for capturing the variation in disease nature, and then some fitted model is used for predicting expected numbers of cases for future periods. Thresholds or alarm limits are calculated against which to compare observed counts in the future. The system will signal an alarm if the observed count exceeds an alarm limit. To detect aberrations during a period of time, statistical process control (SPC) methods are widely used in public health surveillance (Strat, 2005; Woodall, 2006). Also, time series and statistical modelling techniques, used to forecast and calculate the threshold for observed counts in the future, are either implemented directly or applied with SPC methods (Höhle *et al.*, 2009; Paul, 2010).

2.1.2 Data used in public health surveillance

2.1.2.1 Data used for disease surveillance and syndromic surveillance

Obviously, the numbers of confirmed or diagnosed cases have been used in disease surveillance. Even though the delay in reporting cases from hospitals or health centres may cause a delay in outbreak detection, such data is more accurate and reliable than pre-diagnosis data used in syndromic surveillance since at least the reported cases have been confirmed by medical staff or laboratories. In syndromic surveillance, health-related data and pre-diagnosis data are used for monitoring an occurrence of disease. These kinds of data are used as an indicator which might give a possible signal when an outbreak has occurred since it has been thought that people will first treat themselves with medication sold at pharmacies, or phone health service centres, before going to hospital (Stoto *et al.*, 2004; Fienberg and Shmueli, 2005). Example of health-related data are information from hospitals and health centres on numbers of patients symptoms, numbers of chief complaints and numbers of emergency department visits (Jackson *et al.*, 2007; Meynard *et al.*, 2008), while pre-diagnosis data is the data collected from the first stages of illness, such as number of phone calls to health centres, over-the-counter pharmacy sales, and work and school absenteeism (Burkom, 2003; Buehler *et al.*, 2004; Fienberg and Shmueli, 2005; Cooper *et al.*, 2006). Recently, there have been uses of search engines on the internet. For example, the number of searches for the word "flu" were observed, aggregated and used to monitor influenza activity in several countries (Google, 2010).

The advantages of syndromic surveillance are firstly that it is more likely to give an early detection; secondly that some kinds of such data, for example drug sales in pharmacies, are collected routinely and made available for public users and researchers without any ethical issues concerning use of patient information from hospitals and thirdly that some sources of data, such as medication sales, provide additional useful information for investigations in public health surveillance (e.g. postcodes and specific details of customers) (Fienberg and Shmueli, 2005). However, there are some limitations in using health-related and pre-diagnosis data. Such data contain a lot of noise, so it is hard to know whether it can be used as truly representative of underlying health. For example, increasing medication sales might be due to either a disease outbreak, or unusual events such as sales promotion or long holidays (Fienberg and Shmueli, 2005). The results from the use of health-related and pre-diagnosis data in syndromic surveillance should therefore be interpreted with care.

2.1.2.2 Univariate data and multivariate data

Typically, public health surveillance has used univariate data, a single series, for monitoring occurrence of disease at a local level (i.e. for one particular region). Even though univariate series monitoring is easy to implement, the interpretation of the result from outbreak detection is limited since it can be interpreted only locally. Additionally, there might be other variables or characteristics which might be better used for detecting a disease outbreak. In order to incorporate other variables which might improve the performance of outbreak detection, multivariate data have been introduced in public health surveillance for outbreak detection. Multivariate data for public health surveillance can

be either one variable or characteristic measured in several regions (e.g. number of chief complaints from five hospitals) or different variables or characteristics measured in a particular region (e.g. number of chief complaints, phone calls and emergency department visits to a specific hospital), or a combination of the two (multiple variables in multiple regions). There are several statistical techniques proposed and used for monitoring multivariate data, such as dimensionality reduction, parallel surveillance, scalar accumulation and vector accumulation. The details of these methods are provided in section 2.2.2.

The advantages of using multivariate data in public health surveillance is that the covariance and correlation between series may be taken into account. For example, if there is a correlation pattern between numbers of phone calls in five hospitals, monitoring data from each hospital separately might not be able to detect a disease outbreak, but instead, monitoring by using multivariate methods might give a possibility for detecting an occurrence of outbreak more rapidly after it has started (MacGregor, 1995; Mohtashemi *et al.*, 2007; Shmueli and Burkom, 2010; Unkel *et al.*, 2012). Apart from the direct correlation between series mentioned previously, due to the nature of a disease and disease transmission, the cross-correlation, expressing the relationship with time lag between series, should also be considered. Incorporating the cross-correlation between series properly by monitoring realigned multivariate data according to the time lag between series might be able to detect small shifts between series which are not large enough to detect by monitoring such data in a ordinary way. The idea of incorporating time lag between series was found in Frisén *et al.* (2011), though data are assumed independent (i.e. no correlation between series). This is discussed in more details in section 4.2.4.

Even though the multivariate surveillance has advantage of taking correlation between series into account, it faces the problem of signal interpretation since sometime it is hard to know which variable contributed to the signal (MacGregor, 1995; Bersimis *et al.*, 2007). The advantages and disadvantages of univariate surveillance and multivariate surveillance are described in section 2.2.

2.1.2.3 Real health data and simulated data

To assess the performance of outbreak detection in public health surveillance, statistical methods have been evaluated with either real health data or simulated data or both. Using real health data to develop such system gives external validity as the result might be used and interpreted in practice for detecting a real disease outbreak. However, it might not be possible to evaluate the timeliness, which measures the delay of outbreak detection, since the date of disease exposure is rarely exactly known (as is whether or not a disease outbreak has truly occurred) (Buckeridge, 2007; Meynard *et al.*, 2008).

To make a sensible evaluation, simulated data have been used in system development in order to avoid the problem of unknown date of disease exposure. It can also be used to investigate the robustness of proposed statistical methods under different outbreak conditions or circumstances (Mandl *et al.*, 2004; Hutwagner *et al.*, 2005). For system evaluation, statistical simulation is used to simulate either background data or outbreak

signals or both. Outbreak signals are injected into background data and, therefore, timeliness can be evaluated precisely as the date of exposure is known according to the time of the injected signals. However, simulated data gives only internal validity. Although the results from simulation show effective performance of outbreak detection, the system might not be useful for real outbreak situations (Meynard *et al.*, 2008).

Several studies have developed tools to simulate data used in syndromic surveillance or biosurveillance for purposes of system evaluation. Simulated data are generated by incorporating some additional concerns for public health surveillance such as seasonality, holidays, day of the week effects, and spatial information, while different characteristics of the outbreak signal can also be generated for investigating performance of outbreak detection (Lotze *et al.*, 2010; Höhle, 2007; Maciejewski *et al.*, 2009). In addition, to permit assessment of both internal and external validity, real, or pseudo-real, health data might be used as background along with simulated outbreak signals to evaluate system performance for outbreak detection (Lotze *et al.*, 2010; Frisén, 2007).

Due to the ethical issue and a rare event of spare disease or bio-terrorism attack, heath or biosurveillance data is rarely available for researchers. Several studies proposed the methods to simulate health data in different circumstances in order to use the data for evaluating the statistical techniques proposed for heath surveillance (Höhle, 2007; Maciejewski *et al.*, 2009; Lotze *et al.*, 2010; Yahav and Shmueli, 2012b). The details of these methods are provided in section 2.6.5.

2.1.3 Data related issues

Some consideration should be given to issues concerning data used in public health surveillance. Firstly, health data are normally autocorrelated over time. Secondly, the data used for on-line monitoring, such as numbers of diagnosed cases, numbers of emergency department visits and numbers of phone calls to health centres, are non-negative count data which are more likely to be Poisson or negative binomial data rather than data from a normal distribution.

In order to use statistical techniques to develop tools for public health surveillance, one needs to consider these limitations. For example, control charts, which are one of statistical techniques widely used in this research area, are constructed from the basis of normality assumptions, where data are assumed to be continuous, independent and normally distributed. Applying control chart directly to health data might violate normality assumptions and lead to wrong conclusions. Moreover, autocorrelation with past observations should be considered and taken into account for developing tools for the surveillance. Several studies have used generalized linear models for fitting health data and modified control charts for monitoring health data which are assumed Poisson or negative binomial distributed (Farrington *et al.*, 1996; Held *et al.*, 2005; Höhle and Paul, 2007; Paul, 2010).

Thirdly, due to non-negative count data monitoring, the main focus of the surveillance is to detect a sudden change in a positive direction, where a change in a negative direction

is less important for the surveillance since it might imply a better health scenario or a non-epidemic period. Therefore, some statistical methods, such as control charts, will need to be modified for detecting a shift in positive directions (i.e. only upper control limits will be used).

Finally, in order to detect a change which might give a possible signal for a disease outbreak, one should make a distinction between non-epidemic (or endemic) and epidemic periods to make sure that a change from non-epidemic periods causes a possible signal for epidemic periods. Non-epidemic or endemic periods represent normal situations where health data fluctuate around normal level due to the natural variation in health data. Epidemic periods represent an outbreak period, where a disease spread and results in higher numbers of cases than expected normal level of heath data in endemic periods. For this reason, background or baseline data representing non-epidemic periods should remain at the 'normal' level and not contain the data from epidemic periods in order to distinguish a change from baseline data when an outbreak occurs (Fienberg and Shmueli, 2005). Due to the natural variation in real health data, it might be difficult to discriminate epidemic periods from non-epidemic periods since the date of disease exposure is rarely known and the strong evidence of substantially increasing numbers of cases might not be presented clearly. In this case, judgments from epidemiologists might be needed. For simulation study, simulated baseline data should represent non-epidemic periods, so that deviation from the baseline data results from injected simulated outbreak signals (Fienberg and Shmueli, 2005; Hutwagner *et al.*, 2005).

2.2 Overview of methods used in public health surveillance

The type of surveillance can be categorized by number of series or variables (p) monitored. Univariate surveillance is used to monitor univariate data ($p = 1$), whereas multivariate surveillance is used for monitor multivariate series ($p > 1$). The statistical techniques used for univariate surveillance are quite straightforward, while those for multivariate surveillance are more complicated. Statistical techniques used for univariate and multivariate surveillance are summarized as follow. More details of specific methods are given in following sections.

2.2.1 Statistical methods for univariate surveillance

Statistical techniques used for monitoring univariate data can be roughly grouped into four groups. The first group is Statistical Process Control (SPC) techniques which includes Shewhart charts (e.g. p-chart and u-chart), cumulative sum (CUSUM) charts and exponentially weighted moving average (EWMA) charts. Control charts are used to monitor an aberration which might occur during a period of time. A test statistic is calculated and monitored continuously by comparing with the corresponding control limits. Once the test statistic goes beyond the control limits, the alarm will be signalled.

Secondly, several statistical techniques used for univariate surveillance are based on statistical modelling. The purpose of statistical modelling in public health surveillance is to

explain the variation of disease nature by modelling observed data with time or other explanatory variables which might affect the variation observed . Once the model is fitted, a prediction is made in order to calculate thresholds for monitoring future observations. The alarm will be flagged, if the observed data is greater than the thresholds. There are various techniques based on statistical modelling such as the Serfling model (Serfling, 1963), the Stroup approach (Stroup *et al.*, 1993), the Farrington approach (Farrington *et al.*, 1996), generalized linear models (Farrington and Andrews, 2004; Jackson *et al.*, 2007; Kleiman *et al.*, 2004) and branching process models (Held *et al.*, 2005). Similarly, time series techniques have also been used in order to capture the time dependency in observed data. One-step ahead forecasting is used to predict and calculate thresholds for the next observation. There are several types of time series models used for this purpose, such as autoregressive (AR) models, moving average (MA) models, autoregressive integrated moving average (ARIMA) models, seasonal autoregressive integrated moving average (SARIMA) models (Pérez *et al.*, 1998; Reis and Mandl, 2003) and dynamic linear (DLM) models (Cowling *et al.*, 2006).

Thirdly, several proposed statistical techniques are based on likelihood ratio (LR) method. The ratio is defined between the distributions of an out of control process and an in control process (Frisén, 2003; Zhou *et al.*, 2010). The ratios are monitored continuously with specified alarm limits. In order to calculate the ratio, the distributions and parameters need to be specified. It has been suggested that the likelihood ratio method fulfills the criteria of optimal surveillance described in Frisén (2003). However, sometimes it is hard to implement since the distribution of in control and out of control processes are unknown (Frisén, 2003, 2009; Sonesson and Frisén, 2005). Several studies have used likelihood ratio methods for parallel surveillance (i.e. multiple univariate surveillance, see below) (Andersson, 2008) and implemented likelihood ratio methods with other statistical methods such as CUSUM charta or generalized likelihood ratio (Höhle and Paul, 2007; Zhou *et al.*, 2010). Apart from these, a fourth, non-parametric, approach for detecting an increasing incidence in public health surveillance was proposed by Frisén *et al.* (2007).

2.2.2 Statistical methods for multivariate surveillance

Due to the availability of multivariate health data, multivariate surveillance has been considered since it might give a better chance of outbreak detection than univariate surveillance. Statistical techniques used for multivariate surveillance are categorized into four groups depending on the methods used to calculate test statistics. The methods are dimensionality reduction (principal components and sufficient reduction), parallel surveillance, scalar accumulation (Hotelling's T^2 charts) and vector accumulation (MCUSUM and MEWMA charts) (Sonesson and Frisén, 2005; Frisén *et al.*, 2011). Details of each method are summarized as follows.

2.2.2.1 Dimensionality reduction

Because of the high dimension of multivariate data, a dimensionality reduction approach is proposed in order to reduce the number of variables to be monitored. Grouping or

reducing dimensionality can be done by using principal component analysis where covariance and correlation between variables are taken into account (Scranton *et al.*, 1996; Mohtashemi *et al.*, 2007) and sufficient reduction methods (Frisén *et al.*, 2011). The test statistics, which can be either the scores of major principal components or likelihood ratio statistics, are monitored continuously by the statistical techniques used in univariate surveillance (e.g. likelihood ratio statistics monitored via EWMA charts (Frisén *et al.*, 2011)) or multivariate surveillance (e.g. scores from first-five principal components used in a parallel surveillance (Mohtashemi *et al.*, 2007)). Additionally, loadings from major principal components can also be used for reducing numbers of variables used to be monitored by selecting and monitoring those original variables which contribute to large values of loadings, while variables which have small values of loadings might be removed. This might save cost and time for data collection since only fewer variables will be collected. However, as well as loss of some information, another drawback of dimensionality reduction is that when an alarm is flagged, it is sometimes difficult to interpret which variable causes the alarm (Bersimis *et al.*, 2007).

2.2.2.2 Parallel surveillance

Unlike dimensionality reduction, parallel surveillance uses a univariate surveillance technique to monitor each variable of a multivariate collection in parallel (Sonesson and Frisén, 2005). For example, numbers of diagnosed cases from five hospitals are monitored by using five EWMA charts separately. The alarm will be flagged if any one of the charts gives an out of control signal. Obviously, the signal resulting from a parallel surveillance indicates which variable causes a disease outbreak since the variables are independently monitored. However, if correlation between series is present and the evidence of high incidence is not strong, monitoring each variable separately may not detect the signal of an outbreak since the correlation between series has been ignored (MacGregor, 1995). Additionally, monitoring and testing many variables at the same time leads to multiplicity problems, which increase the type I error from using statistical test repeatedly (though, it might be adjusted by a Bonferroni correction or the union-intersection principle (Frisén, 2003; Andersson, 2007; Marshall *et al.*, 2004)) .

2.2.2.3 Scalar accumulation

Scalar accumulation is another approach for multivariate surveillance where the components of the multivariate series at each time point are summarized into a scalar statistic. This series of scalar statistics are monitored by univariate surveillance (Sonesson and Frisén, 2005). A Hotelling's T^2 chart is one example of scalar accumulation where T^2 is calculated from multivariate series and monitored with a Shewhart scheme (Frisén, 2007; Andersson, 2007) (the details of the Hotelling's T^2 chart are provided in section 2.3.2.1). Even though the covariance of the multivariate series is incorporated, the scalar accumulation approach still faces the difficulty of signal interpretation.

2.2.2.4 Vector accumulation

Correspondingly, vector accumulation uses accumulated information from each component of multivariate series by transforming the vector of component-wise alarm statistics

into a scalar statistic. This scalar statistic is again monitored with a threshold or alarm limit. The system will flag an alarm if the scalar statistic is greater than this limit (Sonesson and Frisén, 2005). There are several statistical methods based on vector accumulation approach such as multivariate cumulative sum (MCUSUM) charts (Crosier, 1988; Pignatiello and Runger, 1990), and multivariate exponentially weighted moving average (MEWMA) charts (Lowry *et al.*, 1992).

2.2.3 Statistical methods in current use

There are several statistical techniques which have been used for public health surveillance in practice. The mostly common used are statistical process control and statistical modelling methods (Burkom *et al.*, 2004; Shmueli and Burkom, 2010) such as CUSUM and EWMA charts and the methods based on the Farrington approach (Farrington *et al.*, 1996). Several public health surveillance systems have been implemented and used in practice in the United States of America such as ESSENCE, EARS and RODS.

Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE), a web-based syndromic surveillance system, has been developed for monitoring the health status of military health care beneficiaries since 1999 (Pavlin, 2003). ESSENCE uses non-traditional data sources, such as emergency department chief complaints, over-the-counter medication sales and absenteeism rates in order to give an earlier signal for outbreak detection (ESSENCE, 2010). Tools for ESSENCE are based on SPC methods. Raw data are monitored with a EWMA chart and a Shewhart I-chart is applied to residuals from a fitted linear regression model with time, day of week and holiday effects as explanatory variables.

In 2001, the Centers of Disease Control and Prevention (CDC) developed an Early Aberration Reporting System (EARS) as a pioneering method for monitoring bioterrorism since the terrorist attacks using anthrax in September 11, 2001 (EARS, 2010). This system consists of three modified CUSUM charts called C1, C2 and C3. C1 and C2 are modified CUSUM charts with different sliding window baselines, while C3 is a truncated CUSUM chart used for accumulating deviation from the past three years (EARS, 2010; Shmueli and Burkom, 2010). Likewise, SPC has also been used for public health surveillance in RODS (Real-time Outbreak and Disease Surveillance) (RODS, 2010). Several statistical methods including control charts have been implemented and used as tools for outbreak detection such as an MA chart with sliding window baseline of 120 days; the nonstandard combination of CUSUM and EWMA charts where the CUSUM chart is used to monitor the predicted next day count from a EWMA chart; a recursive least squares (RLS) algorithm where parameters in an autoregressive model are continuously updated with minimized prediction error; and a Shewhart-I chart with the de-trended count data from wavelet decomposition (Shmueli and Burkom, 2010).

In the UK, several public health organizations have used surveillance systems for monitoring disease outbreaks such as NHS (National Health Service) Direct, HPA (Health Protection Agency), HPS (Health Protection Scotland) and CDSC (the Public Health Wales Communicable Disease Surveillance Centre). Practical public health surveillances

have been developed by using statistical modelling based on the Farrington approach (Farrington *et al.*, 1996) (see detail in section 2.5.2) (Unkel *et al.*, 2012). NHS Direct and the HPA have used symptoms based on daily the phone call data to monitor the occurrence of influenza and other seasonal communicable diseases (Andrews, 2010; Elliot, 2010). The HPA uses the Farrington approach, an overdispersed Poisson linear regression model with time as an explanatory variable, to fit the data (Andrews, 2010), while HPS uses a slightly different Farrington approach by including the effects of season, public holiday, weekday and weekend in the model (Robertson, 2010; Kavanagh *et al.*, 2012). Daily calls to NHS Direct categorized by syndrome, call outcome and post code are monitored routinely by control charts (Elliot, 2010). Another system developed from the Farrington approach is an ERS (Exception Reporting System), which has been implemented to health surveillance in Scotland using NHS24 data (NHS24, 2012). This system is used for monitoring influenza-like syndromes in Scotland. Data from the past 28 consecutive days before the current count are used as a baseline and fitted by an overdispersed Poisson linear regression model with the effects of days of the week and holidays (Kavanagh *et al.*, 2010; Unkel *et al.*, 2012; Kavanagh *et al.*, 2012).

In European countries, the Basic Surveillance Network (BSN) has started for gathering and sharing useful information about infectious communicable disease in European community since 2000 (Ternhag *et al.*, 2004; Hulth, 2010). This network provides a database for surveillance data, for example date of onset, age and sex, of over 40 communicable diseases. Similar to the UK, public health organizations in several European countries have implemented statistical methods in public health surveillance. For example, the Robert Koch Institute (RKI), which is an organization responsible for disease control and prevention in Germany, uses the Farrington approach (Farrington *et al.*, 1996), the Stroup approach (Stroup *et al.*, 1989) (see detail in section 2.5.2) and SatScan (SatScan, 2010) for disease monitoring. Likewise, in Denmark and the Netherlands, the Farrington approach and the Stroup approach have also been used, while in Sweden, apart from those methods, the semi-parametric method (Frisén *et al.*, 2007) (see detail in section 2.4.1) has been implemented (Hulth, 2010).

2.3 Statistical process control

Statistical process control (SPC) was developed for controlling product quality in manufacturing processes but has since been used practically in health surveillance (Hutwagner *et al.*, 2005; Fricker and Rolka, 2006; Woodall, 2006; Shmueli and Burkom, 2010; Unkel *et al.*, 2012). Control charts, one tool of SPC, are widely used for monitoring an aberration in a process for the purpose of a process improvement in industry or outbreak detection in health surveillance. Generally, a control chart is a line graph displaying ongoing data or statistics collected or calculated. The values are plotted along with an lower control limit (LCL) and upper control limit (UCL) calculated from the distribution of the process characteristic. If points lie between these limit, the process is called 'in control' (i.e. no aberration or anomaly in the process). On the other hand, if any point goes beyond either the lower or upper control limit, the process is called 'out of control' (or is in an out of control stage; i.e. there might be a shift or aberration occurring in

the process). The process is then examined to find the cause of the deviation.

Normally, the use of control chart is defined into two phases. Phase I (or the training set) is for retrospective analysis and is where the control chart is constructed. The lower and upper control limits are calculated according to the parameters defined, or the estimates estimated from the data, in this phase. Phase II (or the test set) is for prospective surveillance where the control limits calculated from phase I are used to monitor data or statistics in a process (Oakland, 1990; Bersimis *et al.*, 2007). The issue of defining phases I and II is discussed in section 2.3.3. The performance in detecting a change in a process is commonly evaluated through the average run length. The average run length (ARL) is the average of number of points plotted on the control chart before a point falls outside the control limits. The in control average run length (ARL_0) is the average run length when there is no change or abberation during the process (i.e. process is in control). In other word, ARL_0 the average run length until a false alarm occurs (see more details in section 3.2). The out of control average run length (ARL_1) is the average run length when a true change or abberation has occurred in a process (Oakland, 1990; Montgomery, 2009). Ideally, large ARL_0 is preferred if the process is still in control, while small ARL_1 is required since we want to detect a change in the process as soon as possible when the process is out of control. In this thesis SPC tools for health surveillance are categorized into two groups: SPC for univariate surveillance and SPC for multivariate surveillance. The development of SPC tools used in both are summarized as follows.

2.3.1 SPC for univariate surveillance

Several univariate control charts have been used for health surveillance. In order to detect a small increasing incidence which might be an evidence of a possible outbreak, CUSUM charts and EWMA charts are commonly used rather than Shewhart charts which perform better for detecting a large shift (Bersimis *et al.*, 2007; Montgomery, 2009). The details of each method are provided in the sections below.

2.3.1.1 Shewhart chart

The Shewhart chart was proposed by Walter Shewhart in the 1920s (Montgomery, 2009). Theoretically, Shewhart charts were set up under normality assumptions, so (univariate) data are assumed to be independently normally distributed. Data used in an online-process are monitored with control limits which are confidence intervals of process characteristics (mean, standard deviation and proportion, etc.) calculated according to a specific probability of false alarm or type I error (α). An alarm will be signalled if the current observation goes beyond the control limits. Types of Shewhart charts depend on characteristics monitored and whether they are from continuous or count data. In many cases, health data are count data, p-charts are used to monitor mortality and morbidity rates (Shephard, 2006; Fox, 2007), while u-chart have been used to monitor the number of diagnosed cases in units or regions (Hanslik *et al.*, 2001).

However, Shewhart charts perform best in detecting large shifts which is not the aim of disease surveillance where we want to detect early a small shift (Frisén, 2007). Addition-

ally, comparing the current observation (or a statistics based in the current observation) with control limits to identify an alarm may not be sensible for health surveillance since health data normally are dependent, i.e. time-correlated. So due to the dependence in health data, the previous information should be taken in the consideration. Also, using fixed thresholds, calculated from parameters estimated in phase I, to decide on an alarm may lead to wrong decisions if gradual changes are present (Frisén *et al.*, 2007; Frisén, 2007; Marshall *et al.*, 2004; Höhle, 2008).

2.3.1.2 Cumulative Sum (CUSUM) charts

Unlike Shewhart charts that use only the current observation to decide on an alarm, CUSUM charts use all information in the process to date to decide whether an aberration has occurred. CUSUM charts perform better for detecting a small shift than Shewhart chart (Bersimis *et al.*, 2007; MacGregor, 1995; Montgomery, 2009) since they monitor the accumulation of deviations between observations and a target value of the process. Therefore, an increasing small shift might be represented in such deviations. The recommended value for the control limit, or a decision bound, is five times the process standard deviation of phase I (Montgomery, 2009). The charts will flag an alarm if the cumulative sum of the deviations goes beyond these decision bounds. For the purpose of outbreak detection, a one-sided upper CUSUM chart is used to detect a positive shift in prevalence.

Some developments of CUSUM charts are C1, C2 and C3 charts. These are the modified CUSUM charts with different sliding window baseline (i.e. the baseline data in phase I, which are used for estimating parameters in a process, are updated and moved consecutively over time relative to a current observation) (Shmueli and Burkom, 2010).

CUSUM charts were developed for manufacturing processes, so the data and the deviation are assumed to be continuous. Therefore, using CUSUM charts without adjustment may be not suitable as health data are usuaslly count data. In order to monitor count data, several studies have modified CUSUM charts by transforming Poisson data to normality, so that an ordinary CUSUM chart can be applied (Rossi *et al.*, 1999; Meyer *et al.*, 2008), and by adjusting parameters used to construct CUSUM charts for time-varying count data and the count data with the association of geographical regions (Rogerson and Yamada, 2004).

2.3.1.3 Exponentially Weighted Moving Average (EWMA) charts

Like CUSUM charts, EWMA charts are useful for detecting a small shift in a process. All information in the past is used to calculate test statistics and control limits by weighting all past observations. The smoothing constant ($\lambda; 0 < \lambda \leq 1$) gives more weight to the current observation and less to past observations. The range of smoothing constant is between 0 and 1, where $\lambda = 1$ gives all weight to the current observation and ignores the rest, so this is similar to Shewhart chart (Montgomery, 2009). The choice of λ is important since an early change may be detected by EWMA chart with small λ, while a later change may be detected with a larger value of λ (Frisén *et al.*, 2007). The test

or EWMA statistic used to construct EWMA chart can be calculated from

$$z_t = \lambda y_t + (1 - \lambda) z_{t-1},\ z_0 = \mu_0 \tag{2.1}$$

where z_t is the EWMA statistic at time t (Montgomery, 2009) and μ_0 is the process target. The lower and upper control limits (LCL and UCL) can be calculated from

$$LCL = \mu_0 - L\sqrt{\sigma_{z_t}^2} \tag{2.2}$$

$$UCL = \mu_0 + L\sqrt{\sigma_{z_t}^2} \tag{2.3}$$

where $\sigma_{z_t}^2$ is the variance of EWMA statistics and L is the width of the control limits. The variance can be either the exact process variance or the asymptotic variance. As t gets larger, the asymptotic variance might be used (Montgomery, 2009). More details of EWMA charts used in this thesis are provided in section 4.2.5 (chapter 4). EWMA charts have been implemented in practical health surveillance systems such as ESSENCE (Shmueli and Burkom, 2010). They have also been used in some research studies in public health surveillance as a tool used for monitoring the performance of proposed statistical methods for both univariate surveillance and parallel surveillance for multivariate data (Andersson, 2007; Frisén *et al.*, 2011; Frisén, 2009). More detail concerning EWMA charts is given in chapter 4.

2.3.2 SPC for multivariate surveillance

For multivariate surveillance, the univariate control charts used for univariate surveillance can be applied in parallel surveillance and to monitor the univariate statistics from the dimensionality reduction methods (e.g. the score from the first principal component or the likelihood ratio statistics from the sufficient reduction methods). Apart from this, multivariate control charts have been developed for the purpose of detecting directly aberrations in multivariate processes or series where the covariance and correlation between series are taken in account. There are several statistical techniques proposed for multivariate surveillance such as Chi square and Hotelling's T^2 charts, Multivariate Cumulative Sums (MCUSUM) charts and Multivariate Exponentially Weighted Moving Average (MEWMA) charts. Details of each method are described below.

2.3.2.1 Chi square chart and Hotelling's T^2 chart

Chi square charts and Hotelling's T^2 charts use the same procedure to construct a multivariate control chart which is based on a scalar accumulation approach. The former uses the known covariance ($\mathbf{\Sigma}_0$) to calculate test statistics, while the latter uses the estimated covariance (S) instead. The Chi square chart and Hotelling's T^2 charts use the idea of multivariate analysis (multivariate χ^2 and Hotelling's T^2) applied in a Shewhart-like SPC scheme (Hotelling, 1974; MacGregor, 1995). At each time point, the test statistic (χ^2 or T^2), which is a Mahalanobis distance between the observation and the in control state mean (μ_0), is calculated and monitored against an alarm limit. An alarm will be raised if the test statistic exceeds an alarm limit (MacGregor, 1995; Bersimis *et al.*, 2007). Since these charts use a Shewhart-like SPC scheme, they face the

same problem as Shewhart chart in that the decision is made only on the basis of the current observation, without involving previous information (Mason and Young, 2008). Also it is shown in several research studies that it performs best in detecting a large shift which is not typically the purpose of public health surveillance (MacGregor, 1995; Frisén, 2007; Andersson, 2008).

2.3.2.2 Multivariate Cumulative Sum (MCUSUM) charts

Based on the univariate CUSUM scheme, several types of MCUSUM were proposed in order to monitor a small shift in a process. Crosier (1988) proposed two MCUSUM charts. The first is based on a scalar accumulation where the square root of T^2 from Hotelling's T^2 method is monitored with a univariate CUSUM chart. The second is based on a vector accumulation approach where the scalar statistic that would be monitored by a univariate CUSUM scheme is replaced by a vector of similar components, each relating to a component of the multivariate series. Having specified the in control average run length (ARL_0), an alarm limit can be calculated by a Markov chain approach (Brook and Evans, 1972; Crosier, 1988). The results from simulation studies showed that MCUSUM based on vector accumulation performs better than the one based on scalar accumulation since it gave a faster detection and also indicates direction the mean shift (Crosier, 1988).

Similarly, Pignatiello and Runger (1990) proposed another two MCUSUM charts based on scalar accumulation and vector accumulation (Pignatiello and Runger, 1990). The purposes of these two MCUSUM charts are the same as the MCUSUM charts proposed by Crosier (1988) but the calculation procedures are slightly different. The results from their study showed that performance of both sets of charts were similar (Crosier, 1988; Bersimis *et al.*, 2007). MCUSUM charts proposed by Crosier (1988) and Pignatiello and Runger (1990) have been widely used and implemented in many research studied for both process or quality control and also in public health surveillance (Marshall *et al.*, 2004; Fricker, 2007; Fricker *et al.*, 2007; Bodnar and Schmid, 2007).

2.3.2.3 Multivariate Exponentially Weighted Moving Average (MEWMA) charts

Lowry *et al.* (1992) proposed a MEWMA chart derived from a standard univariate EWMA chart. At each time point, t, the p dimensional vector observation is used to calculate a test statistic using smoothing parameters, $\boldsymbol{\lambda}$, defined as a $p \times p$ diagonal matrix with entries representing smoothing parameter λ_i used to weight the past observations for each series. Note that if $\boldsymbol{\lambda}$ is an identity matrix ($\boldsymbol{I}$), a MEWMA chart will be similar to Hotelling's T^2 chart. Like univariate EWMA charts, MEWMA chart can be constructed by using either an exact or asymptotic variance. An alarm will be sounded if the test statistic exceeds than an alarm limit determined by simulation for a specified ARL_0 (Lowry *et al.*, 1992; Bersimis *et al.*, 2007).

Generally, the proposed multivariate control charts (i.e. MCUSUM and MEWMA chart) are directionally invariant meaning that they are designed to detect a mean shift in all directions which is not the interest of public health surveillance (Fricker, 2007). Therefore,

several authors have modified multivariate control charts in order to focus only detecting a positive shift rather than negative shift, such as directionally sensitive Hotelling's T^2 chart (Fricker, 2007), modified MCUSUM charts (Testik and Runger, 2006; Fricker, 2007; Fricker *et al.*, 2007) and modified MEWMA charts (Testik and Runger, 2006; Fricker, 2007; Fricker *et al.*, 2007; Joner *et al.*, 2008). However due to the complexity, the implementation of these methods are rarely in practice use.

With regard to the use of control charts in multivariate surveillance (parallel surveillance, scalar accumulation and vector accumulation), the results from several studies have shown that the vector accumulation approach is the best since it gives the smallest ARL for detecting changes in process (Pignatiello and Runger, 1990), whereas the parallel approach performs slightly better if a shift is clearly presented in one region or series (Rogerson and Yamada, 2004). Also the MCUSUM chart based on vector accumulation is superior to MCUSUM based on scalar accumulation (Crosier, 1988).

2.3.3 Limitations

Originally, the tools of statistical process control (SPC) were developed for product quality control in industrial processes but have subsequently been implemented for monitoring for possible signals of disease outbreaks in public health surveillance. However, there are some limitations of the use of SPC applications in public health. Theoretically, control charts were developed from the assumption of normal distributions, and the data much more strongly are assumed to be independent as well as normally distributed. Therefore, applying control chart directly to health data might violate these assumptions, since health data are usually dependent counts, often correlated over time (Sonesson and Bock, 2003; Shmueli and Burkom, 2010). Consequently, Shewhart, CUSUM and EWMA charts which are widely used in public health surveillance must be considerably adjusted before being used for monitoring health data.

In order to use control chart to monitor aberrations in health surveillance, there are some other points to be considered. Firstly, the baseline or background data (phase I), used to estimate and calculate control or alarm limits in phase II, should represent non-epidemic period so observations in a period of outbreak should be removed (Shmueli and Burkom, 2010). However, due to the substantial natural variation in health data, it is often hard to know either whether or when an outbreak has occurred. Therefore, defining a non-epidemic period is not easy, and the judgement from an epidemiologist may be needed. Secondly, in public health surveillance, only a positive shift is of interest, and control charts, especially for multivariate surveillance, should be modified carefully for detecting positive increasing incidences or a shift in a positive direction (Testik and Runger, 2006; Fricker, 2007; Fricker *et al.*, 2007; Joner *et al.*, 2008). Moreover, using a fixed threshold or alarm limit based on parameters in phase I to monitor aberrations in phase II (e.g. Shewhart charts and CUSUM charts) may not be proper since the variation of health data may change over time.

In addition, unlike univariate control chart, the interpretation of an alarm from multivariate control chart is not easily made, since it is difficult to know which variable has

contributed a signal (MacGregor, 1995; Bersimis *et al.*, 2007), especially for multivariate control chart developed by dimensionality reduction (i.e. principal components), scalar accumulation (Hotelling's T^2 chart) and vector accumulation (MCUSUM and MEWMA charts). Even though a parallel surveillance, where each variable is monitored separately in parallel, gives a clear interpretation of an alarm, it faces a multiplicity problem regarding the multiple testing at the same time (Sonesson and Frisén, 2005; Frisén, 2009; Shmueli and Burkom, 2010). An alarm from the multivariate surveillance is considered primarily as a warning of a possible signal of an occurrence of disease outbreak.

The performance of a control chart for detecting aberration during a process is evaluated by the average run length (ARL). Good performance of outbreak detection relies on a small ARL_1 for fixed ARL_0. In order to compare control chart performances, ARL_0 for each method needs to be fixed, then ARL_1 can be sensibly compared.

Several studies have suggested ways to improve the use of SPC in the outbreak detection over several issues such as dependent data, the direction of shifts and non-parametric for multivariate control charts (Woodall, 2006). The key problem of dependent data (i.e. autocorrelation within series) may be handled by studying the robustness of ordinary methods; modifying ordinary methods with wider alarm limits based on the correct variance; or autocorrelation and monitoring residuals from time series models (Frisén *et al.*, 2007; Höhle, 2010).

2.4 Statistical methods based on LR method

Several statistical methods used for detecting a change in a process are developed from the principle of the likelihood ratio (LR) methods. These methods derive from the likelihood ratio between two distributions in the process, i.e. in control and out of control stages. It has been proved that the methods based on the LR method are optimal under the detection probability and minimal expected delay criteria (Shiryaev, 1963; Frisén and de Maré, 1991; Frisén, 2005) (see more detail in section 3.3). Methods based on LR methods have been proposed and used in several studies such as Shiryaev-Robert (SIR) method (Shiryaev, 1963; Roberts, 1966), the generalized likelihood ratio (GLR) method (Lai, 1995; Höhle and Paul, 2007; Höhle, 2010; Zhou *et al.*, 2010), sufficient reduction methods (Wessman (1998); Frisén *et al.* (2011)), the LRpar and LRjoint methods (Andersson, 2007) and semi-nonparametric method (Frisén *et al.*, 2007).

Most of the methods were proposed and used for univariate surveillance using a univariate control chart as a detection tool (e.g. SIR and GLR and semi-nonparametric methods). However, these methods can also be used in parallel for multivariate surveillance. LRpar was also proposed as a parallel method, while LRjoint and sufficient reduction methods were proposed for multivariate surveillance where the p dimensional multivariate series are incorporated in the likelihood ratio. Brief details of some of these methods are summarized in the following sections.

2.4.1 Semi-parametric method

In order to avoid problem in finding a proper baseline in parametric approaches where parameters have to be estimated from baseline data, Frisén *et al.* (2007) proposed a semi-parametric method for detecting monotonic increasing incidences without requiring historical baseline data. This method is based on a likelihood ratio test between the out of control stages, where the mean starts increasing monotonically after an in control stage, where the mean of the process is constant. The likelihood ratio statistic is calculated sequentially from the ratio between the current observation at time t, y_t, and the average of observations since surveillance started if y_t is greater than y_{t-1}; otherwise an average of observations since surveillance has started is used instead of y_t. A decision as to whether or not there is a shift in the process is made sequentially. The evaluation of such a system was made with real data and simulated data from Poisson distributions by comparing sensitivity, specificity and timeliness (Frisén *et al.*, 2007).

2.4.2 Sufficient reduction method

The sufficient reduction method is one of methods able to perform multivariate surveillance based on dimensionality reduction. The dimensionality reduction is developed with regard to sufficiency properties (i.e. no information from the p dimensions is lost). Wessman (1998) first proposed the sufficient reduction method for detecting a simultaneous shift of a parameter in a multivariate process of independent observations while incorporating the correlation between series. The derived sequence from the sufficient reduction was proved to be sufficient for monitoring a mean shift in a multivariate process (Wessman, 1998). Later Frisén *et al.* (2011) developed the method for detecting changes with time lags in the multivariate process. The detection performance of such methods is evaluated by monitoring the derived univariate sequences with EWMA charts. It shows that they perform better (with shorter delay) than parallel surveillance or monitoring the original series with MEWMA charts (Frisén *et al.*, 2011). More details and the limitations of sufficient reduction methods proposed by Wessman (1998) and Frisén *et al.* (2011) are provided in sections 4.2.3 and 4.2.4, respectively.

2.4.3 LRpar and LRjoint methods

Andersson (2007) proposed two statistical techniques based on the likelihood ratio (LR) called LRpar and LRjoint. LRpar is a parallel method between in and out of control stages for multiple series (i.e. each marginal density is monitored separately), whereas LRjoint is derived from the full likelihood ratio of the joint density between in and out of control stages. The derivation of LRpar and LRjoint statistics can be found in Andersson (2007), illustrated with bivariate normal distributions. The calculation of LRjoint statistics is complicated as all possible joint densities between in and out of control stages are considered. An alarm will be flagged, if the test statistics exceed an alarm limit derived from the distribution of change points. The detection performance of such methods are compared with monitoring the original series with Hotelling's T^2 chart. Hotelling's T^2 chart performs better in detecting simultaneous changes when there is no correlation between series, while LRjoint performs better in detecting changes with

time lags with correlation between series present. LRpar outperforms Hotelling's T^2 for detecting changes with time lags between two independent series (Andersson, 2007).

2.4.4 Limitations

Even though the statistical methods based on the likelihood ratio are optimal under the criterion, mentioned earlier in section 2.2.1, it is difficult to implement the LR methods in practice since the distributions of the process both in control stage or out of control stages may be unknown or difficult to identify. However, they do have advantages over other methods for detecting shifts in multivariate processes, especially for sufficient reduction methods where the relevant information from the original data is taken into account while achieving dimensionality reduction. In order to use sufficient reduction methods in public health surveillance, there are some further limitations to be considered. For example, such methods have so far been developed under the specific assumption that the data are independent over time and so may be unsuitable for health data, which are typically dependent over time. The semi-parametric approach has only been illustrated for univariate surveillance, hence it can only be used in a parallel mode. Also, the results from the method should be interpreted carefully since a monotonic increasing incidence might not signal a true outbreak since it can occur naturally as the beginning of an annual disease season (Frisén *et al.*, 2007).

2.5 Disease modelling

Disease modelling has been used for both retrospective and prospective surveillance. It has been used to describe and explain the variation in disease outbreaks using historical data for retrospective surveillance, while fitting a model and predicting an expected count and an alarm limit for monitoring aberrations in the future is the aim of prospective surveillance. One reason why disease modelling has been widely used is because of the nature of health data. Health or biosurveillance data are normally dependent over time or locations. Therefore, the use of statistical process control may not be applicable since the usual assumption of independent observations is violated.

A more indirect use has been to fit models to explain structural variation in health data and then leave residuals, assumed independent, to be assessed via some standard surveillance techniques. There have been uses of many types of statistical modelling methods in public health surveillance, such as time series techniques, regression models, generalized linear models and branching process models. In this study, statistical techniques used in disease modelling are categorized into two groups: time series techniques and statistical modelling (regression model, generalized linear model and branching process model). The details of each type are discussed below.

2.5.1 Time series techniques

Time series techniques such as AR, MA, ARIMA models have been used for both retrospective and prospective surveillance. In prospective surveillance, time series techniques

are used to model the variation of health data in order to make a prediction as a threshold for the next count. Typically, autoregressive terms are used to explain the variation and the dependence on the past observations during a period of time using ARIMA models (Pérez *et al.*, 1998; Reis and Mandl, 2003). However, due to the large natural variation of health data and its diversity across series, the data may not be stationary over time, therefore using time series models with fixed estimated parameters may not be appropriate. Wang *et al.* (2005) proposed a model incorporating sine and cosine function for capturing seasonality and also allowing the autoregressive term to change over time. Likewise, state space models (dynamic linear models) which allow parameters changing over time have been recently used in several studies (Cowling *et al.*, 2006; Diggle and Stanton, 2008). Time series models have frequently been implemented in univariate surveillance (Pérez *et al.*, 1998; Reis and Mandl, 2003; Wang *et al.*, 2005; Cowling *et al.*, 2006), while multivariate time series, generally more complicated, are rarely used (Bodnar and Schmid, 2007).

2.5.2 Statistical modelling techniques

Apart from time series techniques, several other forms of statistical modelling have been used for both univariate surveillance and multivariate surveillance. Obviously, univariate surveillance uses statistical modelling to fit a single series, while multivariate series can be either fitted separately under parallel surveillance or fitted by incorporating correlation between series or neighbourhoods using a multivariate model of some form. Some techniques used in public health surveillance are summarized below.

2.5.2.1 Serfling model

Proposed by Serfling (1963), the Serfling model has been implemented and widely used as a basis for modelling baseline data for 'non-epidemic' or 'endemic' periods (i.e. health data fluctuating around the 'normal' level due to the natural variation) in many studies (Wang *et al.*, 2005; Held *et al.*, 2005; Höhle *et al.*, 2009; Pelat *et al.*, 2007). The approach is based on three parts: a trend, the variation due to seasonality, and the deviations between expected and observed counts for calculating an alarm limit. Let y_t be the observation at time t, then the Serfling model can be written as

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_{2t} \sin\theta_t + \hat{\beta}_{3t} \cos\theta_t$$

where θ is a linear function of t, $\theta_t = (2\pi t/52)$ for weekly data, and $\hat{\beta}$ is the estimated parameter of a trend (Serfling, 1963). The parameters in the model can be estimated by least squares. An alarm limit is set at the upper limit of 90% prediction interval where the variance is calculated from the sum of squares of the differences between fitted values and observed values for the non-epidemic periods. Since it has only been used for univariate surveillance or parallel surveillance, the correlation between series is not incorporated. In order to capture seasonality, a long historical baseline data may be required. Also, since seasonality and specific term are incorporated the in model, the data under study should be relatively regular over time (Strat, 2005).

2.5.2.2 Stroup approach

Stroup *et al.* (1989) proposed a surveillance method specifically for identifying aberrations in public health data. This aims to monitor whether the number of cases currently observed is different from the number of cases in the past. Therefore, to assess the significance of an observation in current week (y_0), the corresponding baseline data for y_0 consists of the corresponding weeks and the 'surrounding' weeks (the preceding and following weeks) in each of five previous years (15 observations), denoted by $y_1, y_2, ..., y_{15}$. Using a classical parametric method based on normality assumptions, the 95% prediction intervals for the expected value calculated from the corresponding baseline are used as alarm limits for outbreak detection purposes. The result from this approach is typically presented graphically in a horizontal bar chart on a log scale of the ratio between the current observation and the mean of the corresponding baseline data. A vertical line on x-axis at $x = 1$ indicates where the two are equal, while lengths of bars from this vertical to the left and right sides represent how far the incidences are lower or higher than the reference level, respectively (Stroup *et al.*, 1989). This method, called a Current and Past Experience Graph (CPEG), has been implemented in several studies (Cox, 2004; Farrington and Andrews, 2004; Meynard *et al.*, 2008).

It has been suggested that the calculations based on normality assumptions may not be appropriate, since the mean is sensitive to outliers. Using the mean as a denominator of the ratio may influence the capability of the method for outbreak detection since small increases are not likely to be detected. The median of the corresponding baseline data might therefore be an alternative choice (Stroup *et al.*, 1989). However, a problem which is not so easily avoided is that even though selecting the corresponding baseline data might handle the seasonality in the data, it ignores longer scale autocorrelation between time points as only data from three consecutive weeks of each year are selected (Lawson and Kleinman, 2005). A trend is not also incorporated (Strat, 2005). Moreover, due to the nature of health data, which are typically small positive counts, a variant on the method based on the Poisson distribution should be applied (Stroup *et al.*, 1989, 1993; Cox, 2004; Meynard *et al.*, 2008).

2.5.2.3 Farrington approach

Farrington *et al.* (1996) proposed a Poisson generalized linear model adjusted for overdispesion for modelling count data in public health surveillance. A linear trend term in time t is incorporated in the model, and the seasonal variation is taken into account by selecting a corresponding baseline from past years for monitoring the current week. The baseline set of similar form to that proposed by Stroup *et al.* (1989) has been implemented in the Farrington approach by defining parameters w and b. w is a window half width (i.e. $x - w$ to $x + w$ for the current week (x)) and b is number of the years back. So, if x is the observation in the current week of year h, the corresponding baseline data are the data for week $x - w$ to week $x + w$ of year from $h - b$ to $h - 1$. Let y_i be the observation at time i,($i = 1, 2, , t$), the Farrington model can be written as

$$\log \mu_i = \alpha + \beta t_i$$

where $E(Y_i) = \mu_i$ and $V(Y_i) = \phi\mu_i$. Thresholds can be calculated as an upper prediction

interval for $\hat{y}_i$ (Farrington *et al.*, 1996). The authors also proposed the calculation of a threshold for detecting a change in small counts and correcting for past outbreaks in the baseline data. The Farrington approach has been widely used in univariate and parallel surveillance in practical use and research studies (Meyer *et al.*, 2008; Höhle, 2008; Andrews, 2010; Elliot, 2010; Robertson, 2010; Unkel *et al.*, 2012; Kavanagh *et al.*, 2012) (see detail in section 2.2.3). With regard to the use of corresponding baseline data, as with the Stroup approach, the Farrington approach also has the limitation of not incorporating autocorrelation of consecutive past observations.

2.5.2.4 Branching process model

The model proposed by Held *et al.* (2005) was derived from a Poisson branching process model consisting of both parameter-driven and observation-driven components. A parameter-driven component is used to describe the endemic variation including trend and seasonality, while an observation-driven component is used for explaining epidemic patterns with an auto-regression on the observation at the previous time point. A Poisson branching process model can be expressed as

$$\begin{gathered} Z_t = X_t + Y_t,\ t = 1, 2, , t \text{ with} \\ X_t \sim \text{Poisson}(\nu), \text{ for endemic term, and} \\ Y_t | Y_{t-1} \sim \text{Poisson}(\lambda y_{t-1}), \text{ for epidemic term.} \end{gathered}$$

For univariate surveillance, the series is modelled with a generalized linear model (GLM) with Poisson observation and identity link. The GLM model with endemic term (ν) and epidemic terms (λy_{t-1}) can be written as follows.

$$\mu_t = \nu_t + \lambda y_{t-1}$$

where $\nu_t = \exp\left(\alpha + \beta t + \sum_{t=1}^{s}(y_t \sin(2\pi t/52) + y_t \cos(2\pi t/52))\right)$, for weekly data, s is the number of harmonics to include and ϕ is an autoregressive parameter.

For multivariate surveillance, multivariate series can be modelled either by using parallel surveillance or by incorporating correlation between series or neighbourhoods into the model. Parallel surveillance can be carried out by using the model described above, where the autoregressive term of each series is defined separately. In order to incorporate the correlation between series, the subscripts of parameters in both endemic and epidemic components are added according to the units within the multivariate series. The model based on GLM with identity link can be found in Held *et al.* (2005); Höhle *et al.* (2009). A number of extensions have subsequently been considered. Instead of using a fixed autoregressive term (λ) during the monitoring period, a Bayesian change point model was used in order to allow λ_i to change over time (Held *et al.*, 2006). Later, Paul *et al.* (2008) investigated the dependence between different series (i.e. pathogens) and between different or neighbouring locations. The data were modelled by a GLM with negative binomial observations. The autoregressive effects were investigated with different weights for past observations and different time lags (Paul *et al.*, 2008). Moreover, the model proposed by Held *et al.* (2005) was modified by allowing the autoregressive parameter (λ) to change over time, however, the model was illustrated only for univariate surveillance (Paul, 2010).

2.5.3 Limitations

The main limitation of using time series techniques in public health surveillance is the non-stationarity of health data. Although many non-stationary patterns might be transformed to be stationary by differencing, the variation due to unexplainable patterns and the diversity across the series may still remain. Due to the non-stationary data, the estimated parameters in the model need to be sequentially updated. Therefore, ARIMA model might be suitable for retrospective surveillance rather than prospective surveillance (Shmueli and Burkom, 2010). Moreover, the performance of systems developed by time series techniques might depend on the characteristics of outbreaks we want to detect. Incorporating other explanatory variables might improve the detection performance (Wang *et al.*, 2005). The results of Wang *et al.* (2005) suggest that a small outbreak is unlikely to be detected by a model with updated parameters and the results of comparing three methods (CUSUM chart, regression model and dynamic linear model (DLM)) by Cowling *et al.* (2006) indicate that a CUSUM chart is preferred for detecting small shifts and DLM performs better than a simple regression.

Obviously, univariate statistical modelling is commonly used, while multivariate modelling is rarely in practical use. Even though the statistical modelling can be implemented in a parallel surveillance for multivariate series, the multiplicity problem is still of concern along with the issues of neglecting correlation between series. Some proposed statistical modelling methods, such as Stroup and Farrington approaches, can only be applied for univariate surveillance or parallel surveillance. According to the Stroup approach, the corresponding baseline data were chosen to handle the seasonality in data. However, it ignores the pattern of a trend and autocorrelation between past consecutive observations. Also, the calculation method based on normality assumption may not be appropriate for small positive counts (Strat, 2005). The Farrington approach has been implemented in multivariate surveillance, but the implementation is just a parallel comparison in which, again, the autoregressive effect and the dependence between series were not taken into account (Höhle *et al.*, 2009). Regarding the branching process model, in order to incorporate the effect of time and spatial dependency of multivariate series, one is reliant on techniques for identifying those effects (Held *et al.*, 2005).

2.6 Other methods

There have been a variety of other statistical methods used in public health surveillance in different conditions and circumstances. Some studies have used one statistical method directly, while some have used combinations of methods for the purpose of outbreak detection such as the use of disease clustering statistics with CUSUM charts. Brief details of some of these other techniques are summarized below.

2.6.1 Spatial analysis, spatial-temporal analysis and scan statistics

Apart from statistical modelling which can incorporate the dependency of spatial locations or neighbourhoods in the model (Held *et al.*, 2005), spatial and spatio-temporal analysis and scan statistics use the information on case locations if available for disease

mapping, disease clustering and ecological analysis (Lawson and Kleinman, 2005). Spatial analysis or point process analysis uses points, or case locations, to examine whether or not there is a pattern of randomness or disease clustering. Spatial patterns can be examined by using the $K(s)$ function, calculating the expected number of further events within distance s of an arbitrary event. $K(s)$ is likely to be large at small distances if there is evidence of disease clustering (Diggle, 2007). In order to monitor the spatial pattern with time varying, Diggle (1995) extended $K(s)$ to $K(s,t)$ which is the expected number of further events within distance s and time t of an arbitrary event of the process (Diggle, 2007).

Likewise, scan statistics proposed by Kulldorff (1997) are also used for investigating the degree of spatial randomness. Circles with varying radii are randomly created across the area, then the likelihood ratio between the observed and the expected numbers of cases within them are calculated. Evidence of disease clustering is reflected by high likelihood ratios (Kulldorff, 1997). Later, spatio-temporal scan statistics have been proposed and used to investigate whether or not cases are randomly distributed over space and time. The method extended the use of circles to cylinders, where the heights of the cylinders represent time (Kulldorff, 2001). Many studies have used spatial or spatio-temporal analysis in public health surveillance. However, spatial analysis is generally suitable for a retrospective surveillance, to examine how a disease has spread, while spatio-temporal analysis is more suitable for prospective surveillance where changes in spatial distribution are monitored over time (Kulldorff, 2001).

Several studies have applied spatial analysis (to monitor space) in conjunction with SPC (to monitor time) for investigating spatial patterns in prospective surveillance. For example, Tango statistics and Knox statistics, which indicate the degree of disease clustering, are monitored with CUSUM charts (Rogerson, 1997, 2001). In the case of multivariate surveillance, different data sets such as number of phone calls, chief complaint and emergency department visits in the same area can be monitored for spatial pattern by using multivariate scan statistics (Kulldorff *et al.*, 2007).

2.6.2 Methods developed by Bayesian approaches

Some of the methods mentioned above have been modified to incorporate prior knowledge in a Bayesian framework. For example, from the scan statistics and multivariate scan statistics proposed by Kulldorff (1997 and 2007), Bayesian spatial scan statistics and multivariate Bayesian scan statistics have been developed (Neill *et al.*, 2005, 2007). More generally, Bayesian approaches have been implemented with Markov models in public health surveillance in order to detect a shift between endemic and epidemic states such as a Bayesian Markov switching model (Martínez-Beneito *et al.*, 2008), a Bayesian hierarchical model (Spencer *et al.*, 2011) and a Hidden Markov model where the state is unobserved (Le Strat and Carrat, 1999). Additionally, Perry and Allen (2005) used the probability from a posterior distribution to flag an alarm if a high unusually posterior probability of outbreak is presented (Perry and Allen, 2005). The main difference of these methods is how the parameters are estimated. These models are used for univariate surveillance and also can be extended for the multivariate case. However, due to the

computation time, these methods might present difficulties for prospective surveillance since the model needs to be refitted or updated over time (Unkel *et al.*, 2012).

2.6.3 Principal component analysis

As mentioned in section 2.2.2, there have also been the uses of the scores of the first principal component monitored by several control charts such Shewhart chart, CUSUM and EWMA charts. The results showed that the use of dimensionality reduction (i.e. the first principal component) improve the process capability as the sensitivity of detection increased (Kullaa, 2003) and the ARL (Average Run Length) from such approach is smaller than the ARL from the ordinary MEWMA chart (Scranton *et al.*, 1996). As well as its use in dimensionality reduction in multivariate surveillance, principal component analysis can also be used for detecting an outlier during a process which might possibly signal when an outbreak has occurred. The standardized scores of major and minor principal components are used to investigate outliers for outbreak detection purposes (Mohtashemi *et al.*, 2007). However, users might need to use more than one component in order to get sufficient information from the original data from the principal component analysis. Therefore, with regard to the dimensionality reduction, a univariate series of statistics summarizing relevant information from the sufficient reduction method might be preferable than the principal component analysis (Wessman, 1998).

2.6.4 Poisson regression chart

The Poisson regression chart or the generalized likelihood ratio (GLR) test chart are implemented in several studies (Höhle and Paul, 2007; Höhle, 2010; Zhou *et al.*, 2010). Höhle and Paul (2007) proposed a (seasonal) Poisson regression chart based on a generalized linear model used for monitoring aberrations arising in disease surveillance. Parameters are estimated by using GLM and then the likelihood ratio (LR) detector and the generalized likelihood ratio (GLR) detector between in and out-of-control stages is recursively computed and monitored using a generalization of the CUSUM scheme (Höhle and Paul, 2007). Later, the LR CUSUM detector was modified for online monitoring for a change point in categorical time series (Höhle, 2010). Similarly, the GLR statistics are monitored with a EWMA chart in Zhou *et al.* (2010). Like the methods based on the likelihood ratio, the LR detector is derived from the ratio between in control and out of control stages, which sometimes might be unknown or hard to identify.

2.6.5 Simulated health data

Since health or biosurveillance data are rarely available for researchers, several studies have proposed techniques or functions for simulating health data and outbreak signals for the purpose of system development and evaluation. Lotze *et al.* (2010) proposed methods for simulating multivariate health data and mimicking real health data incorporating additional effects such as seasonality, holiday and day of week effects (Lotze *et al.*, 2010). These simulations are based on the multivariate normal distribution, and thus the simulated data may not be sensible for health surveillance. Later, due to the complexity of the multivariate Poisson distribution, Yahav and Shmueli (2012b) proposed methods for generating multivariate Poisson series used in management science application (Yahav

and Shmueli, 2012b). Höhle (2007) has developed a function in the *surveillance* package in *R* used to simulate health data based on his proposed branching process model (Held *et al.*, 2005; Höhle, 2007). Maciejewski *et al.* (2009) proposed SYDOVAT (Synthetic Syndromic Surveillance Data Creation Toolkit) used for generating syndromic surveillance data for system evaluation. Generated data consist of locations, detail of illness (e.g. symptoms) and background information of patient in monitored areas. Outbreak signals can also be generated and injected to background data for the purpose of system evaluation (Maciejewski *et al.*, 2009).

2.6.6 Other methods

Due to the nature of health data, some methods such as wavelet decomposition (Shmueli and Fienberg, 2006) and Fourier transforms (Goldenberg *et al.*, 2002) have been used to remove noise from data before monitoring with other statistical methods.

2.7 Conclusions

Having reviewed the statistical methods used in public health surveillance, none of the methods perform uniformly better than the others. Each shows advantages over the others in some conditions, but always with some limitations. For example, several types of control charts developed under the normality assumptions might not be directly used to monitor health data which are non-normally distributed and correlated over time. Even though the time series and disease modelling might capture the dependence in the health data, they are preferred to used for retrospective rather than prospective surveillance since fitting the proper models and predicting an alarm limit for the next count might be difficult due to the nature and non-stationarity of health data. Also, there have been several statistical methods for public health surveillance proposed under certain assumptions but used in different applications which might be problematic because of violation of assumptions (Unkel *et al.*, 2012).

Considering the statistical methods used for multivariate surveillance, parallel surveillance, which is easy to implement and practical, suffers from a problem of multiplicity, since it relies on multiple tests, while the others (dimensionality reduction, scalar and vector accumulation) have the problem of signal interpretation since once the dimension is reduced to one, it is difficult to know which variable (series) caused a signal. However, unlike using control charts in industrial or manufacturing process, where the cause of a signal needs to be identified for a process capability, a signal in health surveillance is used primarily as a warning of a possible disease outbreak. Therefore, signal interpretation is not our main concern in this thesis. Due to the complexity, multiple univariate control charts are more commonly used in heath surveillance rather than multivariate control charts. Hotelling's T^2, MCUSUM and MEWMA charts, which are directionally invariant, might not be directly appropriate to use in health surveillance since only a positive shift is of interest. Although the adjustment of these charts for detecting a particular positive shift has been proposed in several studies, they are rarely used in practice (Pignatiello and Runger, 1990; Fricker, 2007; Joner *et al.*, 2008; Yahav and Shmueli, 2012a).

Having considered the methods currently available for multivariate surveillance, in this thesis we aim to develop the sufficient reduction method for detecting a mean shift in public health surveillance. Compared to the principal component analysis where some information is lost from using only the first component (Scranton *et al.*, 1996; Kullaa, 2003), sufficient reduction methods can be thought of as an ideal dimension reduction method. The idea of the sufficient reduction methods is to appeal to the sufficiency principle (see more details in section 4.2.1) to identify a lower dimensional function of the original series which retains all relevant information to detect a shift in the multivariate process. Even though the sufficient reduction methods have a limitation over signal interpretation, it is still preferable than the others for health surveillance since its derived univariate series can be easily monitored with univariate control charts in order to avoid the complexity of multivariate charts. As mentioned earlier, in order to use sufficient reduction methods in public health surveillance, such methods need to be modified. More details of such methods and how they can be developed for public health surveillance are provided in chapter 4.

Chapter 3

System evaluation

3.1 System evaluation

Many studies have focused on the development of detection performance of public health surveillance systems. A system's performance depends on how quickly it can detect a possible signal for a disease outbreak and how accurate it is. Due to the advantages and drawbacks of using real data and simulated data in health surveillance discussed in section 2.1.2.3, several kinds of health data have been used in the system evaluation: real data (real background data with real shift), simulated data (simulated background data and simulated outbreak signals) or mixed (real background data with simulated outbreak signals) (Burkom, 2003; Hutwagner *et al.*, 2005; Fienberg and Shmueli, 2005; Watkins *et al.*, 2006; Jackson *et al.*, 2007; Buckeridge, 2007). The use of real and simulated data in system evaluation are summarized in section 2.1.2.3.

Various types of measures have been used in the system evaluation such as sensitivity, specificity, timeliness, average run length, false alarm rate, probability of successful detection and predictive value. More details of measures used in the system evaluation are provided in section 3.2, followed by the optimality of surveillance methods in section 3.3. Having reviewed the measures used to evaluate detection performance, those chosen for use in this thesis are defined in section 3.4.

3.2 Measures for system evaluation

For the purpose of detecting a disease outbreak in public health surveillance, a decision as to whether or not there is a shift of aberration or anomaly in a process is made sequentially at each time t, $t = 1, 2, ...$ Let τ be the change point of a process and t_A the time of an alarm. The system will flag an alarm at time t_A if a test statistic exceed a threshold or alarm limit. A summary of measures used for system evaluation is given in Table 3.1 (Frisén, 2003; Sonesson and Bock, 2003; Farrington and Andrews, 2004; Hutwagner *et al.*, 2005; Buckeridge, 2007) and the discussion is provided afterward.

Table 3.1: A summary of measures used for system evaluation

Measure	Definition
Sensitivity	Proportion of outbreak signals correctly identified by system.
Specificity	Proportion of non-outbreak signals correctly identified by system.
ROC curve	A receiver operating characteristic curve is a plot of sensitivity against false alarm rate (1-specificity) used for assessing the precision of prediction of outbreak detection.
Timeliness	Timeliness or delay in detection can be measured by ED (Expected Delay) is the expected value of a delay between the change time (τ) and the time of alarm (t_A). $$\text{ED}(t) = E(max(0, t_A - t)|\tau = t)$$ ED typically tends to zero as t increases. Therefore, the delay conditional on no alarm before τ was proposed (Frisén, 2003): CED (Conditional Expected Delay) is the expected delay for a specific time change, τ, conditional on there being no alarm before τ. $$\text{CED}(t) = E(t_A - \tau)|t_A \geq \tau = t)$$ ARL_1 (Average Run Length for an out of control stage) is the average run length until an alarm is signalled when the true change has occurred (see section 2.3). $$\text{ARL}_1 = E(t_A|\tau \leq t)$$
False alarm	False alarm or false detection can be measured by ARL_0, PFA and FAR.

Continued on next page

Table 3.1: A summary of measures used for system evaluation (continued)

Measure	Definition
False alarm	ARL_0 (Average Run Length for an in control stage) is the average run length until an alarm is signalled where there is no outbreak or change during the surveillance period (see section 2.3). $$\mathrm{ARL}_0 = E(t_A \mid \tau > t) \text{ or } \mathrm{ARL}_0 = E(t_A \mid \tau = \infty)$$ PFA (Probability of False Alarm) is the probability that alarm was trigged before the true change has occurred. $$\mathrm{PFA} = P(t_A > \tau)$$ FAR (False Alarm Rate) is the proportion that alarm was trigged before the true change has occurred. $$\mathrm{FAR} = P(t_A > \tau)$$
PSD	Probability of successful detection is the probability that a delay in outbreak detection is no longer than d, where d is a specified period for detecting a particular disease. The usefulness of PSD is that it measures the ability for outbreak detection in a limited time after change has occurred. $$\mathrm{PSD}(d,t) = P(t_A - \tau < d \mid t_A \geq \tau = t)$$
PV	Predictive value is the the probability that a true change has occurred when an alarm is signalled (Frisén, 1992). $$\mathrm{PV}(t) = P(t_A \leq \tau \mid t_A = t)$$

An efficient surveillance system requires high sensitivity and specificity with low false alarm rate and a short delay. Probability of successful detection and predictive value are also expected to be high, while the average run length should be large for the in control stage (ARL_0) and small for the out of control state (ARL_1) (i.e. the system

should flag an alarm as soon as possible when a change has occurred). The proposed statistical techniques used in public health surveillance were evaluated by using various kinds of criteria mentioned above. However, there seems no gold standard surveillance system or even standard evaluation methods used in the field. The proposed methods are evaluated with one or more measures which sometimes are more or less suitable for the particular situation. The choice of evaluation measure in public health surveillance, for both retrospective and prospective surveillance, is given further consideration below.

Sensitivity, specificity and ROC curve have been widely used in medical diagnosis tests and also applied for assessing the performance of outbreak detection in public health surveillance in many studies (Farrington *et al.*, 1996; Dafni *et al.*, 2004; Buehler *et al.*, 2004; Fienberg and Shmueli, 2005; Hutwagner *et al.*, 2005; Cowling *et al.*, 2006; Kleinman and Abrams, 2006; Frisén, 2007; Paul, 2010). Since, the calculations of such measures are based on the proportion of outbreak detections of a fixed sample in a known time period, they are more appropriate to use for retrospective surveillance, where the times of surveillance are known and fixed, rather than in on-line or perspective surveillance (Frisén, 2005). Therefore, for retrospective surveillance, false alarms can be measured by specificity, while sensitivity is used to evaluate detection ability. In addition, ROC curves have been implemented under the assumption of a stationary process over the entire series, and so might not be applicable in public health surveillance where the stationarity assumption might be violated. Moreover, sensitivity, specificity and ROC curves can only be used for evaluating the precision of detection, not for delay in detection. Therefore, an AMOC (Activity Monitoring Operating Characteristic) curve, which is a plot of timeliness against false positive rate, might be preferable (Shmueli and Burkom, 2010).

For prospective surveillance, the false alarm rate can be measured by ARL_0 (average run length in the in control state), while the timeliness or the delay of detection can be measured by ED (expected delay), CED (conditional expected delay) and ARL_1 (average run length in out of control state). ARL_1 is the measure used for assessing a delay in detection in many research studies (Frisén, 2005), but rarely used in biosurveillance literature and practice (Shmueli and Burkom, 2010). There are some concerns over using ARL_1 in system evaluation. Strictly, ARL_1 is the out of control average run length measuring the run length until the system signals an alarm assuming that a change occurred when the surveillance started. Therefore, ARL_1 might not be suitable for assessing the timeliness in prospective surveillance since a true change in health data is not likely to occur at the same time as the surveillance starts (Sonesson and Bock, 2003; Frisén, 2003). Apart from these, PV (predictive value) and PSD (probability of successful detection) have been suggested for evaluating the system performance for prospective disease surveillance (Frisén *et al.*, 2007).

Additionally, some surveillance tools developed by statistical modelling, such as time series techniques and branching process model, have used MAPE (Mean Absolute Percentage Error), MSPE (Mean Squared Prediction Error) and RMSE (the Root Mean Squared Error) for system evaluation (Reis and Mandl, 2003; Held *et al.*, 2005; Burkom *et al.*, 2007). However, these criteria measure how well the models are fitted rather than assessing the performance of outbreak detection of the systems developed by such meth-

ods. The assessment of system performance for prospective surveillance should depend on how accurate the model prediction will be for detecting an aberration which might signal an occurrence of disease. Some studies have suggested and used proper scoring rules to assess model prediction (Held *et al.*, 2005; Czado *et al.*, 2009; Höhle, 2010; Held *et al.*, 2010). A proper scoring rule is used to assess the calibration and sharpness of probabilistic forecasts by assigning a numerical score based on the predictive distribution and the event that occurs. A larger score implies a better probabilistic prediction (Gneiting and Raftery, 2007; Held *et al.*, 2010). Therefore, a proper scoring rule might be another criterion used to evaluate the performance of the model's predictions, showing whether or not there is a high possibility of detecting a true signal of a disease outbreak.

3.3 Optimality of surveillance methods

Having identified a number of performance measures, it is useful to consider whether any proposed detection method can be proven to be 'best' in terms of optimizing certain of these criteria. Obviously, a short delay in detection with a low false alarm rate plays an important role in describing the performance of outbreak detection in public health surveillance. Various measurements have been used to evaluate the statistical methods for surveillance. In order to choose an optimal surveillance method, the optimality of statistical surveillance methods has been considered. Frisén and de Maré (1991); Frisén (2003) studied the optimality properties of statistical surveillance methods for detecting a mean shift in a univariate process under the assumptions that a shift in mean level at time τ indicates a change in a process shifting from one distribution to another. It is also required that the data before change and after change need to be independently and identically distributed. The optimality criteria considered under this assumptions are maximal detection probability, minimal expected delay, minimal ARL_1 for fixed ARL_0 and minimax optimality.

The maximal detection probability for a fixed false alarm rate is considered as one of the optimality criteria. The higher the detection probability is, the more accurate is detection. Also, a system that gives minimum expected delay for a fixed false alarm rate would be preferred. Consider the average run length normally used in the SPC, if a process has shifted, then minimum ARL_1 for fixed ARL_0 is required. On the other hand, to be pessimistic, the worst case of the conditional expected delay after a change point is also considered. In the worst case, having fixed the false alarm rate and calculated maximum CED after a change point, a system that give a minimal maximum CED would be preferred (i.e. minimizes the maximum value of CED).

With respect to the optimality criteria mentioned above, the likelihood ratio (LR) method (see detail in section 2.4) is optimal under maximal detection probability and minimal expected delay criteria. Shewhart charts are optimal under the same criteria for detecting a large mean shift. EWMA charts with small values of smoothing parameter are optimal under maximal detection probability criteria, whereas CUSUM charts are optimal under minimax optimality criteria. More details can be found in (Frisén and de Maré, 1991; Frisén, 2003; Sonesson and Bock, 2003).

However, due to the nature of health data which are more variable, autocorrelated and non-stationary, these optimality criteria might not easily be applied with the statistical methods used in public health surveillance due to the assumptions defining in each criterion (see more details in Frisén (2003, 2005)). In view of this limitation, empirical measures are suggested for use in the system evaluation in public health surveillance (Unkel *et al.*, 2012). Also, in multivariate surveillance, the optimality criteria might be difficult to specify due to the different aspects of multivariate problems in surveillance (Frisén, 2003). Even though the optimality of the multivariate statistical methods cannot be proved, Frisén *et al.* (2011) showed that sufficient reduction methods for detecting changes with time lags (see section 4.2.4) perform well for the condition that the optimality might not hold.

3.4 Measures for system evaluation used in this thesis

Having reviewed the measures used for system evaluation, the detection performance of statistical methods proposed and used in this thesis is evaluated by considering the timeliness, and correct and false detection rate of the system. For the purpose of outbreak detection in public health surveillance, in this thesis, we aim to detect a mean shift in a process within 7 time points (e.g. days or weeks) of a shift at time τ.

From Table 3.1, the timeliness is measured by conditional expected delay (CED) which is defined as the expected delay for a specific change point given that there is no false alarm before a true change occurs. False detection is measured by the false alarm rate (FAR) defined as the proportion of false alarms, i.e. alarms raised before a true change occurs ($t_A < \tau$). We define correct and false detection rate by reference to PSD in Table 3.1 and our requirement to detect a shift within 7 time points (an arbitrary but sensible period), i.e. $d = 7$. The correct detection is measured by the true alarm rate (TAR), defined as the proportion of correct identifications where a system gives a signal within 7 time points after a true change occurs ($t_A - \tau \leq 6 | t_A \geq \tau$). On the other hand, non-detection rate (NDR), the proportion of cases in which the system fails to detect any change in the process within 7 time points when a shift truly occurs ($t_A - \tau \geq 7 | t_A \geq \tau$), is used to measure the failure detection of the system. Short CED, high TAR, with low FAR and NDR are desired. The definitions of CED, TAR, FAR and NDR are given below.

$$\begin{aligned} \text{CED} &= E(t_A - \tau | t_A \geq \tau) \\ \text{TAR} &= P(t_A - \tau \leq 6 | t_A \geq \tau) \\ \text{FAR} &= P(t_A < \tau) \\ \text{NDR} &= P(t_A - \tau \geq 7 | t_A \geq \tau) = 1 - \text{TAR} - \text{FAR} \end{aligned}$$

Ideally, we construct our detection scheme to provide practical and reasonable values for TAR, FAR and NDR, denoted by TAR_0, FAR_0 and NDR_0. As considered in several studies (Burkom *et al.*, 2008; Watkins *et al.*, 2009; Kuang *et al.*, 2012) typical

requirements are

$$\begin{aligned} \text{TAR}_0 &= 0.90 \\ \text{FAR}_0 &= 0.05 \\ \text{NDR}_0 &= 0.05. \end{aligned}$$

Note that not all proposed schemes meet these target in practice.

The methods proposed or used in this thesis are evaluated by use of a simulation study to estimate CED, FAR, TAR and NDR. The results are based on 10,000 simulations. Let $\widehat{\text{CED}}$, $\widehat{\text{TAR}}$, $\widehat{\text{FAR}}$ and $\widehat{\text{NDR}}$ be the estimators of CED, TAR, FAR and NDR, respectively. According to the simulations, the sampling distribution of each estimator is considered. With 10,000 simulations, the standard error of each estimator is sufficiently small to investigate differences in the measure between the methods. More details of sampling distribution and the standard error of each measure considered in the thesis are given in Section 3.5.

3.5 Performance comparison

For the purpose of comparison, the following sections detail the sampling distribution of each measure and the sampling distribution of the difference in the measure between two methods. Even though the comparison of detection performance between the methods are conducted by using the same data set generated from 10,000 simulations, we assume that the estimates for each method are approximately independent and normally distributed with mean and variance defined in each section separately.

A comparison of detection performance between the methods is conducted by performing a two-sided hypothesis test at significance level 0.05. The difference of the measure between two methods is compared against the critical value (i.e. 1.96 times the standard error of the difference of measure between two methods).

Let D be the delay, before which the system can detect a shift, up to a maximum of 7 time points after the process has shifted (i.e. $D = 0, 1, 2, ..., 6$). Thus, D has a discrete Uniform distribution

$$D \sim U(0, 6),$$

with the mean and variance

$$\begin{aligned} E(D) &= (a + b)/2 = 3 = \mu \\ V(D) &= (m^2 - 1)/12 = 4 = \sigma^2 \end{aligned}$$

where $a = 0$, $b = 6$ and $m = b - a + 1$.

Let n be the number of simulations ($n = 10,000$). Following the Central Limit Theorem, the sampling distribution of sample mean, $\bar{D}$, is

$$\bar{D} \sim N(\mu, \sigma^2/n).$$

Let TA, FA and ND be a true alarm, false alarm and non-detection defined below

$$\begin{aligned} \text{TA} &= \text{I}[0 \leq D \leq 6], \\ \text{FA} &= \text{I}[D < 0], \\ \text{ND} &= \text{I}[D \geq 7], \end{aligned}$$

where I[P] is an indicator function of statement P having value 1 if the statement P is true. Having defined the notation above, the details of the sampling distribution and the critical value for each measure are defined as follows.

3.5.1 TAR

The TAR is estimated by

$$\widehat{\text{TAR}} = \frac{1}{n}\sum_{i=1}^{n} \text{TA}.$$

Thus, the sampling distribution of $\widehat{\text{TAR}}$ follows normal distribution with mean and variance defined below

$$\widehat{\text{TAR}} \sim N(\mu_{\widehat{\text{TAR}}}, \sigma^2_{\widehat{\text{TAR}}}),$$

where

$$\mu_{\widehat{\text{TAR}}} = \text{TAR}$$

and

$$\sigma^2_{\widehat{\text{TAR}}} = \frac{\text{TAR}(1-\text{TAR})}{n}.$$

Let $\widehat{\text{TAR}}_i$ and $\widehat{\text{TAR}}_j$ be the estimators of TAR for methods i and j ($i \neq j$). We compare the difference of TAR between two methods by considering

$$\widehat{\text{TAR}}_i - \widehat{\text{TAR}}_j \sim N(\mu_{\widehat{\text{TAR}}_i - \widehat{\text{TAR}}_j}, \sigma^2_{\widehat{\text{TAR}}_i - \widehat{\text{TAR}}_j}),$$

where $\mu_{\widehat{\text{TAR}}_i - \widehat{\text{TAR}}_j} = \text{TAR}_i - \text{TAR}_j$ and $\sigma^2_{\widehat{\text{TAR}}_i - \widehat{\text{TAR}}_j} = \sigma^2_{\widehat{\text{TAR}}_i} + \sigma^2_{\widehat{\text{TAR}}_j}$ (i.e. two samples are independent). The standard error of the difference of TAR between two methods is

$$S.E.(\widehat{\text{TAR}}_i - \widehat{\text{TAR}}_j) = \sqrt{\frac{\text{TAR}_i(1-\text{TAR}_i)}{n} + \frac{\text{TAR}_j(1-\text{TAR}_j)}{n}}.$$

The TARs of two methods are significantly different at the level 0.05, if

$$|\widehat{\text{tar}}_i - \widehat{\text{tar}}_j| > (1.96 \times S.E.(\text{TAR}_i - \text{TAR}_j)),$$

where $\widehat{\text{tar}}_i$ is the estimate, or observed value, of $\widehat{\text{TAR}}_i$ of method i from 10,000 simulations.

The sampling distributions of $\widehat{\text{FAR}}$ and $\widehat{\text{NDR}}$ and the sampling distributions of the differences of these measures between the methods can be defined in a similar way.

3.5.2 CED

From the definition of CED, we estimate CED from

$$\widehat{\text{CED}} = \frac{\bar{D}}{1 - \text{FAR}}.$$

Note that even though FAR can be estimated from the simulations, the sampling distribution of $\widehat{\text{CED}}$ might be difficult to identify. Therefore, we consider the sampling distribution of $\widehat{\text{CED}}$ assuming that FAR is a constant value equal to $\text{FAR}_0 = 0.05$ which is the value the scheme was designed to achieve. Thus, the sampling distribution of $\widehat{\text{CED}}$ follows normal distribution with mean and variance defined below

$$\widehat{\text{CED}} \sim N\left(\mu_{\widehat{\text{CED}}}, \sigma^2_{\widehat{\text{CED}}}\right),$$

where

$$\mu_{\widehat{\text{CED}}} = \frac{\mu}{1 - \text{FAR}} \tag{3.1}$$

and

$$\sigma^2_{\widehat{\text{CED}}} = \frac{\sigma^2}{n(1 - \text{FAR})^2}.$$

Let $\widehat{\text{CED}}_i$ and $\widehat{\text{CED}}_j$ be the estimators of CED for methods i and j $(i \neq j)$. We compare the difference of CED between two methods by considering

$$\widehat{\text{CED}}_i - \widehat{\text{CED}}_j \sim N(\mu_{\widehat{\text{CED}}_i - \widehat{\text{CED}}_j}, \sigma^2_{\widehat{\text{CED}}_i - \widehat{\text{CED}}_j}),$$

where $\mu_{\widehat{\text{CED}}_i - \widehat{\text{CED}}_j} = \mu_{\widehat{\text{CED}}_i} - \mu_{\widehat{\text{CED}}_j}$ and $\sigma^2_{\widehat{\text{CED}}_i - \widehat{\text{CED}}_j} = \sigma^2_{\widehat{\text{CED}}_i} + \sigma^2_{\widehat{\text{CED}}_j}$ (because we assume that the two sample measures are approximately independent). The standard error of the difference of $\widehat{\text{CED}}$ between two methods is

$$S.E.(\widehat{\text{CED}}_i - \widehat{\text{CED}}_j) = \sqrt{\frac{\sigma^2}{n(1 - \text{FAR})^2} + \frac{\sigma^2}{n(1 - \text{FAR})^2}}.$$

The CEDs of two methods is significantly different at the level 0.05, if

$$|\widehat{\text{ced}}_i - \widehat{\text{ced}}_j| > (1.96 \times S.E.(\widehat{\text{CED}}_i - \widehat{\text{CED}}_j)),$$

where $\widehat{\text{ced}}_i$ is the estimate, or observed value, of $\widehat{\text{CED}}_i$ of method i from 10,000 simulations.

3.6 Critical values and result format

Although the standard error of the estimators can be estimated directly from the simulations, we prefer to calculate them according to their sampling distributions by using the practical values used as targets in construction of the detection scheme in Section 3.4. This decision is considered further below. Thus, the critical values (1.96 times the

standard error of the difference of measure between two methods) are calculated by using TAR_0, FAR_0 and NDR_0. For example, the critical value for the CED comparison is

$$1.96 \times S.E.(\widehat{\mathrm{CED}}_i - \widehat{\mathrm{CED}}_j) = 1.96 \times \sqrt{\frac{4}{n(1-0.05)^2} + \frac{4}{n(1-0.05)^2}} = 0.058.$$

We calculate the critical values for each comparison and list them in Table 3.2.

Table 3.2: Critical values for performance comparison

Comparison	Critical value
CED_i - CED_j	$1.96 \times S.E.(\widehat{\mathrm{CED}}_i$ - $\widehat{\mathrm{CED}}_j) = 0.058$
TAR_i - TAR_j	$1.96 \times S.E.(\widehat{\mathrm{TAR}}_i$ - $\widehat{\mathrm{TAR}}_j = 0.008$ or 0.8%
FAR_i - FAR_j	$1.96 \times S.E.(\widehat{\mathrm{FAR}}_i$ - $\widehat{\mathrm{FAR}}_j) = 0.006$ or 0.6%
NDR_i - NDR_j	$1.96 \times S.E.(\widehat{\mathrm{NDR}}_i$ - $\widehat{\mathrm{NDR}}_j) = 0.006$ or 0.6%

The differences in detection performance of potential methods are investigated by comparing the differences in estimated performance statistics (estimated from the simulations) against the critical values defined in Table 3.2.

Before proceeding, we make a brief check on the advisability of our decision to calculate the critical values according to the requirement of the practical scheme in terms of acceptable TAR_0, FAR_0 and NDR_0 by comparing them also with versions calculated from the the standard error of the measure estimated directly by its variability in the 10,000 simulations. Tables A.7 and A.8 in Appendix A are examples of the critical values calculated from the simulations for Section 4.6.1. P, F, W and C1 in the tables stand for Parallel, Frisén, Wessman and our proposed method (Case 1), respectively. Overall, the critical values estimated from the simulations are slightly lower than the critical values calculated from the defined scheme requirements. Therefore, for convenience, we regard the critical values defined in Table 3.2 as "upper bounds" and use them to investigate the significance of differences in measures between the methods in Chapters 4, 5 and 6.

3.7 Conclusions

This chapter provides details of how statistical methods used for health surveillance are evaluated. Due to the varied nature of potential data, several kinds of data have been used in the system evaluation. Various measures have also been used in practice and research studies. However, they should be used with care, since some measures are more appropriate for evaluating retrospective rather than prospective surveillance system.

In public health surveillance, where a rapid detection with low false alarm rate is required, statistical methods which can be shown to be optimal under various desirable optimality criteria are preferable. However, again the complex nature of available data

means that in many situations it is not possible to identify a practical, optimal method. In this thesis, therefore, we consider performance across a platform of four measures as our performance profile. The critical values which will be used in the performance comparison between the methods are defined.

Chapter 4

Sufficient reduction methods for normal data

4.1 Introduction

In this chapter, we will focus on the sufficient reduction (SR) method used for reducing the dimension of a p dimensional multivariate series to a univariate series of likelihood ratio statistics. The idea of sufficient reduction methods has been implemented in several studies. Wessman (1998) and Frisén *et al.* (2011) proposed SR methods for detection of a sudden but persistent shift in multivariate series. Wessman aimed to detect a simultaneous change, where shifts in p series occur at the same time, while Frisén *et al.* (2011) allow for detecting changes with time lag between series. Both methods are proposed by assuming that observations in each series are independent over time. While Wessman (1998) allows for correlation between series (CBS), Frisén *et al.* (2011) does not (see more details in Section 4.2.3 and 4.2.4). Consideration of such assumptions suggests that the methods proposed by Wessman (1998) and Frisén *et al.* (2011) can be used for an industrial or manufacturing context as such data are assumed to be independent over time. However, in order to use sufficient reduction method in public health surveillance, correlation within series (CWS) should be taken into account since health data are usually dependent over time.

To handle the presence of an autocorrelation in a process, several studies have suggested and used the model-based approach. The model residuals are subjected to the univariate surveillance methods used for detecting a mean shift in a process of independent data such as control charts (e.g. Alwan and Roberts (1988); Montgomery and Mastrangelo (1991); Harris and Ross (1991); Kramer and Schmid (1997); Schmid (1997); MacCarthy and Wasusri (2001); Frisén (2003)). However, the model-based approach might be more suitable for retrospective surveillance rather than prospective surveillance, since in the prospective case, a decision whether or not there is a shift in the process is made sequentially, and so repeated refitting model is required. Sometimes it is hard to identify the state of a process as in control or out of control. A change in a process might not be clearly seen from this approach since the parameters are re-estimated and updated over time. Using the model residuals from refitting the model at each time point in

prospective surveillance might therefore be inappropriate. Alternatively, instead of re-fitting the model over time, data in the in control stage are fitted and then the residuals are calculated from the difference between the observed values and the predictions from the fitted model. However, making predictions far ahead might not be sensible, since the independence of residuals might not hold (Box *et al.*, 1994). Noting these limitations, the residuals from model-based approach might not be suitable for prospective health surveillance.

Apart from CWS, taking CBS into account in the multivariate surveillance should also be considered since small shifts are more likely to be detected when CBS is present and incorporated properly. Also, if data are correlated with lag between series, incorporating time lag with corresponding cross-correlation between series might improve the detection performance. In this chapter, we aim to develop the sufficient reduction method to detect a mean shift in a p dimensional multivariate series by taking both CWS and CBS into account as well as the possibility of lagged shifts. The background to sufficient reduction methods and the sufficient reduction methods proposed by Wessman (1998) and Frisén *et al.* (2011) are given in Section 4.2. In Section 4.3, the proposed methods for use in public health surveillance are summarized, with the detailed derivation for independent and dependent observations in sections 4.4 and 4.5, respectively. System evaluation and concluding remarks are provided in sections 4.6 and 4.7, respectively.

4.2 Background to sufficient reduction methods

4.2.1 Sufficiency

Sufficiency is a key property for parameter estimation in which all relevant information contained in data regarding the parameter is collected and summarized succinctly. An estimator or a statistic which has this property is called a *sufficient statistic*.

Following Cox and Hinkley (1974), we define the definition of a sufficient statistic and the factorization theorem used for deriving such statistics as follows.

Definition 4.1 A statistic T is defined to be a *sufficient statistic* for a family of distributions $\mathcal{F}$ if and only if the conditional distribution $f_{X|T}(x|t)$ is the same for all distributions in $\mathcal{F}$. In the parametric case this means that $f_{X|T}(x|t;\theta)$ is the same for all values of θ and so does not depend on $\theta, \theta \in \Omega$. Additionally, T is said to be *minimal sufficient* if T is a function of all other sufficient statistics that can be found.

Factorization theorem Let $L(x;\theta)$ be the likelihood function of X. A statistic T is sufficient for θ in the family $\mathcal{F}$ if there exist functions $h(x)$ and $g(T;\theta)$ such that for all $\theta \in \Omega$,

$$L(x;\theta) = h(x)g(T;\theta)$$

where $h(x)$ is a function that depends on observations but does not depend on the parameter θ and $g(T;\theta)$ is a function that depends on θ and depends on observations only through the value of T.

Since a decision as to whether there is a shift in a process will be made sequentially, we also consider the definition of *sufficiency of a sequence of statistics* which is

Definition 4.2 Let $\boldsymbol{x}_t$ be a p dimensional vector of observations at time $t, t = 1, 2, ..., s$. A sequence $T_1(\boldsymbol{x}_1), T_2(\boldsymbol{x}_2), ...$ of statistics is a sufficient sequence of statistics for families $\mathcal{F}_1, \mathcal{F}_2, ...$ of distributions if for all s, $T_s(\boldsymbol{x}_s)$ is a sufficient statistic for family $\mathcal{F}_s$ (Arnold, 1988).

Both definitions are used to develop SR methods for detecting a shift in the distribution of a multivariate process. Details of SR methods and their implementation are provided in subsequent sections.

4.2.2 SR methods

In a similar vein, sufficient reduction methods use the sufficiency principle to reduce the dimension of a p dimensional multivariate series to a univariate series of likelihood ratio statistics summarising all relevant information from the original series (Wessman, 1998; Frisén *et al.*, 2011). In comparison with other statistical methods used for multivariate surveillance, SR methods automatically handle the multiplicity problem of parallel surveillance and also the loss of some information from the original series by other methods used for dimensionality reduction such as principal component analysis where only the first, or the first few, component(s) is (are) used (Kullaa, 2003). Therefore, in taking all relevant information from the original series by use of sufficient statistics, SR methods should give a better result for detecting a shift in the parameter of the multivariate process.

In multivariate surveillance we define $\boldsymbol{x}_t$ as a p dimensional vector representing the observation made on a p dimensional multivariate series at time t.

$$\boldsymbol{x}_t = \begin{pmatrix} x_{1,t} \\ \vdots \\ x_{p,t} \end{pmatrix},$$

where $x_{i,t}$ is an observation of a random variable X_i observed at time t, $t = 1, 2, ..., s$. Let τ_i be the time of a change point occurring in series i $(i = 1, 2, ..., p)$. We consider two types of change points: simultaneous changes $(\tau_1 = \tau_2 = ... = \tau_p = \tau)$ and changes with time lag (i.e. $\tau_i = \tau_j + l_{ij}$, where l_{ij} is the lag between series i and series j and $i < j$). Assume that changes occur sequentially $(\tau_1 < \tau_2 < ... < \tau_p)$, we regard the earliest change point in one of the p series, denoted by $\tau = min(\tau_1, \tau_2, ..., \tau_p)$, as a change or shift in the process.

For simplicity, the background of SR methods in this section is illustrated for only the case of detecting a simultaneous change in a process of independent observations (no CWS). Detecting changes with time lags is considered as an extension case of the method and detailed in Section 4.2.4. Both ideas are also used to develop SR methods for detecting changes in a process of dependent observations (CWS) which are detailed

in Section 4.5.

In order to detect a simultaneous change in a process, indicated by a shift in the consistent distributions, we define two stages of the process according to the change point τ. An 'in control' stage (I) is where there has been no shift in the distributions, while an 'out of control' stage (O) is where the distributions of the process have shifted at the change point τ. Let F^I and F^O be the two completely specified distributions of in control and out of control stages defined as follows.

$$\mathbf{X}_t \sim F_t = \begin{cases} F^I & t < \tau \quad \text{in control stage} \\ F^O & t \geq \tau \quad \text{out of control stage} \end{cases}$$

At each decision time s $(s = 1, 2, ...)$, the decision will be made as to whether the process is in control, $I(s) = \{\tau > s\}$, or out of control, $O(s) = \{\tau \leq s\}$. According to the change point τ, the available information at time s is $X_s = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_s)$ which has (because of the independence assumption) a distribution from the family

$$X_s \sim \mathcal{F}_s = \begin{cases} \mathcal{F}^{I(s)} = \{F^I\}^s & s < \tau \quad \text{in control stage} \\ \mathcal{F}^{O(s)} = \{\bigcup_{\tau<s} (F^I)^{\tau-1} \times (F^O)^{\tau-s-1}\} & s \geq \tau \quad \text{out of control stage} \end{cases}$$

(Wessman, 1998).

Having assumed that data are continuous and independent over time, let $f^I(\boldsymbol{x}_t)$ and $f^O(\boldsymbol{x}_t)$ be the probability density function (pdf) of in control and out of control stages, respectively. We define a ratio

$$lr(\boldsymbol{x}_t) = \frac{f^O(\boldsymbol{x}_t)}{f^I(\boldsymbol{x}_t)}, \; t = 1, 2, ..., s$$

and a sequence

$$\mathbf{lr}_t(\mathbf{x}_s) = \{lr(\boldsymbol{x}_t), lr(\boldsymbol{x}_{t+1}), ..., lr(\boldsymbol{x}_s)\}, \; s = 1, 2, ...$$

Let $L(\boldsymbol{x}_t|\tau)$ be the likelihood function of the p dimensional multivariate series given that a change occurs at time τ. The sequence of statistics used for monitoring a shift in a process of independent observations, for any s, can be derived from

$$\begin{aligned} L(\boldsymbol{x}_t|\tau) &= \prod_{t=1}^{\tau-1} f^I(\boldsymbol{x}_t) \prod_{t=\tau}^{s} f^O(\boldsymbol{x}_t) \\ &= \prod_{t=1}^{s} f^I(\boldsymbol{x}_t) \prod_{t=\tau}^{s} \frac{f^O(\boldsymbol{x}_t)}{f^I(\boldsymbol{x}_t)} \\ &= h(\mathbf{x}_s) \prod_{t=\tau}^{s} lr(\boldsymbol{x}_t) = h(\mathbf{x}_s) k(\mathbf{lr}_\tau(\mathbf{x}_s)) \end{aligned} \tag{4.1}$$

where h and k are two real valued functions. By the factorization theorem, the vector $\mathbf{lr}_\tau(\mathbf{x}_s) = \{lr(\boldsymbol{x}_\tau), lr(\boldsymbol{x}_{\tau+1}), ..., lr(\boldsymbol{x}_s)\}$, which is a sequence of likelihood ratio statistics, is sufficient for the distribution family of X_s, $\{\mathcal{F}^{I(s)}, \mathcal{F}^{O(s)}\}$, defined by a change point

τ. Equivalently, since $\mathbf{lr}_\tau(\mathbf{x}_s)$ is sufficient for any decision time s, by the definition 4.2, the sequence $\{\mathbf{lr}_\tau(\mathbf{x}_s);\ s = 1, 2, ...\}$ is a sufficient sequence of statistics for the sequence of families $\{\mathcal{F}^{I(s)}, \mathcal{F}^{O(s)};\ s = 1, 2, ...\}$ (Wessman, 1998).

In order to implement the SR method to detect a shift in a process from $\mathcal{F}^{I(s)}$ to $\mathcal{F}^{O(s)}$, the vector $\mathbf{lr}_1(\mathbf{x}_s)$ is used for detecting a shift for any decision time s since the process started at time $t = 1$. Consequently, this also makes the sequence $\{\mathbf{lr}_1(\mathbf{x}_s);\ s = 1, 2, ...\}$ a sufficient sequence of statistics for the sequence of families $\{\mathcal{F}^{I(s)}, \mathcal{F}^{O(s)};\ s = 1, 2, ...\}$. The idea of the sufficient reduction can therefore be used to derive a sequence of likelihood ratio statistics which is sufficient to detect a shift in a parameter in a process, such as for detecting a mean shift (Wessman, 1998; Frisén *et al.*, 2011) or a variance shift (Wessman, 1998). In this thesis the shift in the distributions is considered and investigated as a shift in the mean vector between in control and out of control stages because it is the most sensible and practical use in public health surveillance. The SR methods developed for detecting the mean shift proposed by Wessman (1998) (simultaneous change) and Frisén *et al.* (2011) (changes with time lags) are reviewed in sections 4.2.3 and 4.2.4, respectively.

Since the dimension of the multivariate series is reduced to one by the sufficient reduction, the optimality properties (see Section 3.3) of univariate surveillance used for detecting a change between two distributions (in control and out of control stages) can be applied. Frisén and de Maré (1991) showed that the likelihood ratio method, which is the likelihood ratio between two distributions in the process (i.e. in control and out of control stage according to the change point τ), has properties of minimal expected delay and maximal detection probability. This can also be applied to SR methods since the SR methods derive a sequence of likelihood ratio statistics based on the likelihood ratio between out of control and in control stage given that a simultaneous change of a process occurs at $t = \tau$. Data in both stages are still independent and identically distributed over time. Thus, in the case of a simultaneous change, where the process is clearly shifted between two distributions, the SR method is optimal under minimal expected delay and maximal detection probability properties. However, these properties might not hold in the case of detecting changes with time lag since the distributions of the process has shifted more than once (see more detail in Section 4.2.4).

4.2.3 SR method of Wessman (1998)

Wessman proposed SR methods to detect a shift in distributions of a multivariate process of independent observations when the change point τ is fixed and known, and when τ is stochastic from known distribution (e.g. geometric distribution with intensity ν). The use of SR methods to detect the shift in the parameter vector of a distribution from the exponential family was given with specific examples of detecting a shift in mean or variance of a multivariate normal distribution. Note that the examples used all assumed the distributional form was unchanged though this is not a requirement of the methods. The proposed likelihood ratio statistics derived for detecting mean or variance shift can be found in the original paper (Wessman, 1998). In this study we will consider a mean shift only. The assumptions underlying his method are summarized as follows:

- aim to detect a sudden, but persistent, shift in mean a process of multivariate normal distributed observations.
- observations are independent (no correlation within series (CWS)).
- Correlation between series (CBS) is taken into account.
- consider simultaneous changes (i.e. $\tau_1 = \tau_2 = ... = \tau_p = \tau$).
- a change point τ of the process is known.
- the distributional form and the variance of the process are unchanged over time.

The idea of deriving a sequence which is sufficient for detecting a shift in distribution in Section 4.2.2 is used here, but a shift in a mean vector, $\boldsymbol{\mu}$, is considered specifically. Assuming that a simultaneous change occurs at time τ, the in control stage (I) is where $\boldsymbol{\mu}$ is unchanged and the out of control stage (O) is where mean shifts have occurred in p series at time τ. The mean vector of the process for in control and out of control stages can be defined as

$$\boldsymbol{\mu}_t = \begin{cases} \boldsymbol{\mu}^I = (\mu_1^I, ..., \mu_p^I)' & t < \tau \quad \text{in control stage} \\ \boldsymbol{\mu}^O = (\mu_1^O, ..., \mu_p^O)' & t \geq \tau \quad \text{out of control stage} \end{cases}$$

and

$$F(\boldsymbol{x}_t; \boldsymbol{\mu}) \sim \begin{cases} F^I(\boldsymbol{x}_t; \boldsymbol{\mu}^I) & t < \tau \quad \text{in control stage} \\ F^O(\boldsymbol{x}_t; \boldsymbol{\mu}^O) & t \geq \tau \quad \text{out of control stage} \end{cases}$$

where $t = 1, 2, ..., s$. The mean of in control and out of control stages of series i are denoted by μ_i^I and μ_i^O, respectively, and $\boldsymbol{\Sigma}$ is a covariance matrix which is unchanged over time. Having assumed that observations are independent, we define

$$lr(\boldsymbol{x}_t; \boldsymbol{\mu}) = \frac{f^O(\boldsymbol{x}_t; \boldsymbol{\mu}^O)}{f^I(\boldsymbol{x}_t; \boldsymbol{\mu}^I)}, \; t = 1, 2, ..., s$$

For convenience, we assume that the shift in $\boldsymbol{\mu}$ does not change the distributional form of $f^I(\boldsymbol{x}_t; \boldsymbol{\mu}^I)$ and $f^O(\boldsymbol{x}_t; \boldsymbol{\mu}^O)$ which come from the same exponential family distribution.

Since a shift $\boldsymbol{\mu}$ is of interest, we define $\boldsymbol{\mu}^O - \boldsymbol{\mu}^I$ as a natural parameter for this case. We also assume, for definiteness that data are multivariate normally distributed. The likelihood function at decision time s given that change occurs at τ is

$$\begin{aligned} L(\boldsymbol{x}_t; \boldsymbol{\mu}|\tau) &= \prod_{t=1}^{s} f^I(\boldsymbol{x}_t; \boldsymbol{\mu}^I) \prod_{t=\tau}^{s} \frac{f^O(\boldsymbol{x}_t; \boldsymbol{\mu}^O)}{f^I(\boldsymbol{x}_t; \boldsymbol{\mu}^I)} \\ &= h(\mathbf{x}_s) \prod_{t=\tau}^{s} lr(\boldsymbol{x}_t; \boldsymbol{\mu}^O - \boldsymbol{\mu}^I) \\ &= h(\mathbf{x}_s) \prod_{t=\tau}^{s} (\boldsymbol{\mu}^O - \boldsymbol{\mu}^I)' M_t(\boldsymbol{x}_t) \end{aligned} \tag{4.2}$$

where $M_t(\boldsymbol{x}_t)$ is a minimal sufficient statistics for $\boldsymbol{\mu}^O - \boldsymbol{\mu}^I$ at time t. According to the factorization theorem, the likelihood ratio statistics at time t, $T_t(\boldsymbol{x}_t)$, used for detecting a mean shift at time t is

$$T_t(\boldsymbol{x}_t) = (\boldsymbol{\mu}^O - \boldsymbol{\mu}^I)' M_t(\boldsymbol{x}_t) = (\boldsymbol{\mu}^O - \boldsymbol{\mu}^I)' \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_t,\ t = 1, 2, ..., s$$

From Section 4.2.2, it follows that $\mathbf{lr}_\tau(\mathbf{x}_s; \boldsymbol{\mu}^O - \boldsymbol{\mu}^I) = \{lr(\boldsymbol{x}_\tau), ..., lr(\boldsymbol{x}_s)\}$ is sufficient for monitoring a shift in distribution, which in this case is the mean shift $\boldsymbol{\mu}^O - \boldsymbol{\mu}^I$ for any decision time s. Since the vector $\mathbf{lr}_t(\mathbf{x}_s; \boldsymbol{\mu}^O - \boldsymbol{\mu}^I)$ is used to monitor a process from its start, it can be simply written as

$$\mathbf{lr}_1(\mathbf{x}_s) = \{T_1(\boldsymbol{x}_1), ..., T_s(\boldsymbol{x}_s)\},\ t = 1, 2, ..., s$$

From definition, it follows that the sequence $\{\mathbf{lr}_1(\mathbf{x}_s),\ s = 1, 2, ...\}$ is a sufficient sequence of statistics for detecting a mean shift in a multivariate normal process.

Let $\mathbf{c}$ be a $p \times 1$ matrix which is the difference between the in control and out of control means, i.e. $\mathbf{c} = \boldsymbol{\mu}^O - \boldsymbol{\mu}^I$, $\boldsymbol{\mu}^O > \boldsymbol{\mu}^I$. In order to use the Wessman method to detect a desired shift size in a process, $\mathbf{c}$ is pre-specified. The use of such method might be implemented for detecting a change in public health surveillance, since a minimum shift in health data indicating an epidemic period can be specified. Due to the nature of a disease outbreak, a minimum shift which we want to detect can be defined in different ways. More details of how the shift size, $\mathbf{c}$, is defined in the public health surveillance context are given in Section 4.3.2.

According to the optimality of SR methods mentioned in Section 4.2.2, the SR method proposed by Wessman for detecting a simultaneous mean shift in a process of independent observations is also optimal since the process has shifted between two fully-specified distributions. In this case, despite the shift in the mean vector, the data are still independently and identically normally distributed before and after the change point.

4.2.4 SR method of Frisén *et al.* (2011)

Frisén *et al.* (2011) developed the SR method proposed by Wessman (1998) by considering time lags between series and two types of change point (i.e. simultaneous changes or changes with time lag). A sequence of likelihood ratio statistics is derived in a similar manner, but there are some differences in the assumptions they make. Most assumptions are similar to those of Wessman, ***except***

- the p series are independent (no CBS)
- there are known time lags between change points (LCP) of the different series. The degenerate case of zero lag, simultaneous change, is considered as a special case.

A sequence of statistics for detection of shifts in the parameter vector of a distribution in the exponential family when changes occur with known time lag can be derived in a similar manner to Section 4.2.2. The derivation of this method can be found in the original paper (Frisén *et al.*, 2011). For simplicity Frisén *et al.* illustrate such a method

for detecting a shift in a bivariate normal process ($p = 2$) with unit covariance matrix.

In the bivariate case, the shift size that we want to detect in series i is c_i, $i = 1, 2$. Like the Wessman method, c_i need to be pre-specified. Assume that two series are independent (CBS = 0), the in control and out of control of two series, X_1 and X_2, in the bivariate normal case are defined as follows.

$$X_{1,t} \sim \begin{cases} N(\mu_1^I, 1) & t < \tau_1 \quad \text{in control stage} \\ N(\mu_1^O, 1) & t \geq \tau_1 \quad \text{out of control stage} \end{cases}$$

$$X_{2,t} \sim \begin{cases} N(\mu_2^I, 1) & t < \tau_2 \quad \text{in control stage} \\ N(\mu_2^O, 1) & t \geq \tau_2 \quad \text{out of control stage} \end{cases}$$

where $t = 1, 2, ..., \tau_1, ..., \tau_2, ..., s$ and $\mu_{i,1} = \mu_{i,0} + c_i$. The covariance matrix between X_1 and X_2, which is unchanged over time, is denoted by $\mathbf{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Let l be the known time lag between change points in two series, X_1 and X_2, (i.e. $l = \tau_2 - \tau_1$, labelling the series so that $\tau_2 > \tau_1$) and $x_{i,t}$ be an observation at time t in series i. Since a sudden, but persistent, shift of a parameter vector is of interest, we assume that the first change occurs in X_1 at time $t = \tau_1$, where μ_1^I shifts to μ_1^O, and then l time points later the second change occurs in X_2 at time $t = \tau_2 = \tau_1 + l$ (i.e. μ_2^I shifts to μ_2^O), $t = 1, 2, ..., \tau_1, ..., \tau_2, ..., s$. So the change point of the process in this case is denoted by $\tau = \tau_1 = min(\tau_1, \tau_2)$.

Instead of using $\boldsymbol{x}_t = \begin{pmatrix} x_{1,t} & x_{2,t} \end{pmatrix}'$ in the sufficient reduction like the Wessman method, it turns out that for efficient monitoring it is sensible to realign the series so that they are at the same point in their evolution. If LCP is known, we examine $x_{1,t}$ alongside $x_{2,t+l}$, which is later defined as

$$\boldsymbol{y}_t = \begin{pmatrix} x_{1,t} & x_{2,t+l} \end{pmatrix}'.$$

Due to the availability of data, at each s, $x_{1,t}$ and $x_{2,t+l}$ are used in the sufficient reduction for $t = 1, 2, ..., s - l$. For $t = s - l + 1, ..., s$, data from X_1 only, $x_{1,t}$, are used, since $x_{2,t+l}$ from X_2 is not yet available (Frisén *et al.*, 2011). Frisén *et al.* derived a sequence of likelihood ratio statistics used for monitoring a change with time lag from the likelihood function of bivariate normal series according to a change point τ, a known time lag l and the availability of data.

Let $f^I(x_{i,t})$ and $f^O(x_{i,t})$ be the pdf of in control and out of control stages, respectively, for series i. We have

$$
\begin{aligned}
L(\boldsymbol{x}_t;\boldsymbol{\mu}|\tau;\boldsymbol{\mu}) &= \prod_{t=1}^{\tau-1} f^I(x_{1,t};\mu_1^I) f^I(x_{2,t};\mu_2^I) \prod_{t=\tau}^{\tau+l-1} f^O(x_{1,t};\mu_1^O) f^I(x_{2,t};\mu_2^I) \\
&\quad \times \prod_{t=\tau+l}^{s} f^O(x_{1,t};\mu_1^O) f^O(x_{2,t};\mu_2^O) \\
&= \prod_{t=1}^{s} f^I(x_{1,t};\mu_1^I) f^I(x_{2,t};\mu_2^I) \prod_{t=\tau}^{\tau+l-1} \frac{f^O(x_{1,t};\mu_1^O) f^I(x_{2,t};\mu_2^I)}{f^I(x_{1,t};\mu_1^I) f^I(x_{2,t};\mu_2^I)} \\
&\quad \times \prod_{t=\tau+l}^{s} \frac{f^O(x_{1,t};\mu_1^O) f^O(x_{2,t};\mu_1^O)}{f^I(x_{1,t};\mu_1^I) f^I(x_{2,t};\mu_2^I)} \\
&= h(\mathbf{x}_s) \prod_{t=\tau}^{s-l} \frac{f^O(x_{1,t};\mu_1^O) f^O(x_{2,t+l};\mu_1^O)}{f^I(x_{1,t};\mu_1^I) f^I(x_{2,t+l};\mu_2^I)} \prod_{t=s-l+1}^{s} \frac{f^O(x_{1,t};\mu_1^O)}{f^I(x_{1,t};\mu_1^I)} \\
&= h(\mathbf{x}_s) \prod_{t=\tau}^{s-l} \frac{f^O(\boldsymbol{y}_t;\boldsymbol{\mu}^O)}{f^I(\boldsymbol{y}_t;\boldsymbol{\mu}^I)} \prod_{t=s-l+1}^{s} \frac{f^O(x_{1,t};\mu_1^O)}{f^I(x_{1,t};\mu_1^I)}, \qquad (4.3)
\end{aligned}
$$

where $\boldsymbol{y}_t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu}^I = \begin{pmatrix} \mu_1^I & \mu_2^I \end{pmatrix}'$, $\boldsymbol{\mu}^O = \begin{pmatrix} \mu_1^O & \mu_2^O \end{pmatrix}'$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Similarly to the Wessman method, by the factorization theorem, the vector $\mathbf{lr}_t(\mathbf{x}_s)$, $t = 1, 2, ..., s$, which is sufficient for monitoring a change with known time lag in a process for any decision time s, can be written as a set of likelihood ratio statistics, $T_t(\boldsymbol{x}_t)$, with

$$\mathbf{lr}_1(\mathbf{x}_s) = \{T_1(\boldsymbol{x}_1), ..., T_s(\boldsymbol{x}_s)\}$$

where

$$T_t(\boldsymbol{x}_t) = \begin{cases} c_1\, x_{1,t} + c_2\, x_{2,t+l} & t = 1, 2, ..., s-l \\ c_1\, x_{1,t} & t = s-l+1, ..., s \end{cases} \qquad (4.4)$$

for $c_i = \mu_i^O - \mu_i^I$ the shift size that we want to detect in series i. Consequently, the sequence $\{\mathbf{lr}_1(\mathbf{x}_s),\ s = 1, 2, ...\}$ is also a sufficient sequence of statistics for a change with time lag in bivariate normal process. This idea can be extended for the p dimensional multivariate case.

If LCP = 0 and CBS = 0, the SR method proposed by Frisén *et al.* is similar to the SR method by Wessman where the mean shifts at τ. In the case of simultaneous change, the SR method proposed by Frisén *et al.* is still optimal since the process has shifted between two distributions at change point τ and data are still independently and identically normally distributed over time. However, in the case of changes with time lags, the optimality cannot be guaranteed since the requirement for identical distributions might not be satisfied as the process has shifted more than once. For example, in the case of bivariate series, there are three stages to consider; 1) X_1 and X_2 both in control, 2) X_1 is out of control but X_2 is still in control, 3) both are out of control. Even though, the SR method is not optimal for detecting changes with time lags, Frisén *et al.* gives examples from a simulation study showing that the SR methods proposed for detecting

changes with time lags perform well compared to other methods used in multivariate surveillance such as parallel EWMA charts and MEWMA chart (Frisén *et al.*, 2011).

4.2.5 Detection tools for SR methods

After the dimensionality reduction, the derived (univariate) sequence of statistics is monitored by a simple univariate control chart such as a Shewhart chart (Wessman, 1998) or EWMA chart (Frisén *et al.*, 2011). A shift in parameter is indicated by an alarm signalled by the control chart. However, there are some limitations in using control charts for outbreak detection purposes. These limitations are discussed in Section 2.3.3.

Generally, a common (two-sided) EWMA chart is a plot of EWMA statistics along with upper and lower control limits calculated from the estimated parameter from an in control stage. The system will flag an alarm if any points of the statistics go beyond either upper or lower control limits (Montgomery, 2009). Let z_t be a EWMA statistic at time t, the two-sided EWMA statistic can be calculated from

$$z_t = \lambda x_t + (1-\lambda) z_{t-1},\ z_0 = \mu^I$$

where μ^I is the process target. The mean and the exact variance of Z_t are

$$E(Z_t) = \mu^I$$

$$V(Z_t) = \sigma^2 \frac{\lambda}{2-\lambda}(1-(1-\lambda)^{2t}), \tag{4.5}$$

where σ^2 is the variance of the process. For large t, $t \to \infty$, from equation (4.5) the asymptotic variance of Z_t is

$$V(Z_t) = \sigma^2 \frac{\lambda}{2-\lambda}. \tag{4.6}$$

These two variances give different UCLs. It is suggested one use the exact control limit if the values of t is small (Montgomery, 2009), however, due to its simplicity, the asymptotic control limit is more commonly used (Borror *et al.*, 1998; Frisén *et al.*, 2011; Shiau and Hsu, 2005).

Since detecting a positive shift is the main interest in public health surveillance, a one-sided EWMA chart is preferred over a two-sided EWMA chart. Focusing on detecting the positive shift, the one-sided EWMA statistics is considered instead of two-sided statistics. In order to avoid inertia problem, for a quicker detection, the one-sided EWMA statistics is calculated by resetting z_t to the target process, μ^I, if $z_t < \mu^I$. Having assumed that data are normally distributed and independent over time, the one-sided EWMA statistics can be calculated from

$$z_t = \max(\mu^I, \lambda x_t + (1-\lambda) z_{t-1}),\ z_0 = \mu^I$$

(Harris and Ross, 1991; Schmid, 1997; Morais and Pacheco, 2001; Hu *et al.*, 2011). The one-sided EWMA statistics are monitored with upper control limits adjusted for the desired ARL_0. ARL_0 (Average Run Length; for an in control stage), is the average run

length until an alarm is signalled where there is no change during the surveillance period (see more detail of ARL_0 in Section 3.2).

Given our adjustment to produce a one-sided EWMA chart, the exact variance for the two-sided version stated in equation (4.5) might not hold. Obviously, due to the use of the maximum function in the one-sided EWMA statistics, the variance of the one-sided chart will be less than or equal to the variance of the two-sided chart. However, several studies have used the asymptotic variance of the two-sided chart (equation (6.44)) as an upper bound for the variance of the one-sided chart due to its simplicity (Morais and Pacheco, 2001; Hu *et al.*, 2011). Thus, in this thesis, we construct a one-sided EWMA chart based on the asymptotic variance. Let L be the width of a control chart, the upper control limit is calculated from

$$UCL = \mu^I + L\sqrt{\sigma^2 \frac{\lambda}{2-\lambda}}$$

(Robinson and Ho, 1978; Crowder, 1989; Shu *et al.*, 2007).

To construct upper control limits for a one-sided EWMA chart, the values of L and λ need to be specified in order to meet the desired the average run length (ARL_0) (Lucas and Saccucci, 1990; Montgomery, 2009). It has been suggested that values of λ in the range $0.05 < \lambda < 0.25$ (Montgomery, 2009) or $0.2 < \lambda < 0.3$ (Hunter, 1986) work well. Frisén *et al.* (2011) used $\lambda = 0.35$ in their simulation study. In this study, we arbitrarily chose $\lambda = 0.3$, the value in between 0.25 and 0.35 giving a little more weight on the current observation. To make the sensible comparison between the SR methods and the parallel methods, we define $\text{ARL}_0 = 370$ for the Frisén *et al.*, Wessman and later our own methods and $\text{ARL}_0 = 741$ for parallel method adjusted for multiplicity (see more detail in Section 4.2.6). According to the defined λ and desired ARL_0, $L = 2.815$ for $\text{ARL}_0 = 370$ and $L = 3.046$ for $\text{ARL}_0 = 741$. The choices of L and λ corresponding to desired ARLs for a standard one-sided EWMA chart can be obtained from package '*spc*' in *R* (Knoth, 2012).

4.2.6 Parallel method

The parallel method is introduced in this section since it is used as a baseline to compare the performance of SR methods against (Section 4.6). Due to its simplicity, the parallel method is commonly used in multivariate surveillance. Each single series from the p dimensional multivariate series is monitored separately, by the same univariate statistical methods, in parallel. The earliest change from any one of the p series is regarded as the change of the multivariate process. Even though this method is easy for implementation in practice, the problem of multiplicity from multiple hypothesis testing and ignoring CBS are major concerns (see Section 2.2.2).

Due to the limitations mentioned above, SR methods have the advantage over the parallel methods by taking CBS into account (in Wessman case) and reducing the dimension of multivariate series to a univariate series so that it can be monitored by using a univariate statistical method avoiding the multiplicity problem.

A one-sided EWMA chart is used as a detection tool for both SR and parallel methods. For a sensible comparison, the parallel method is adjusted for the multiplicity problem by using the Bonferroni correction by adjusting the desired ARL_0. For a one-sided EWMA chart used for the Frisén and Wessman and later our proposed methods, we define $\mathrm{ARL}_0 = 370$. Let q be the probability that any point exceeds the upper control limit (UCL), ARL_0 can be expressed as

$$\mathrm{ARL}_0 = \frac{1}{q} = \frac{1}{P(z_t > UCL|\mu = \mu^I)} = 370,$$

so $q = 0.0027$ (Borror *et al.*, 1998; Montgomery, 2009). In order to use the parallel method for monitoring a mean shift in a bivariate process where the data in each series are monitored separately in parallel, the probability q defined above is divided by 2 according to the Bonferroni correction to address the multiplicity from monitoring two series in parallel. Therefore, the ARL_0 for the parallel method is

$$\mathrm{ARL}_0 = \frac{1}{q} = \frac{1}{0.00135} = 740.7 \simeq 741$$

4.3 Proposed extensions to SR methods

In order to develop the SR methods proposed by Wessman (1998) and Frisén *et al.* (2011), we further consider their limitations and aim to broaden their applicability. Wessman and Frisén *et al.* proposed the SR methods based on the assumption of independent observations (no CWS). While the former allows CBS, the latter does not. Using such methods for outbreak detection might not be appropriate as health data are usually dependent over time. Moreover, types of change points (LCP) between series are also another issue for Wessman (1998) as he considered only simultaneous changes. Therefore, Frisén's method might have advantages over Wessman's method if LCP is clearly present.

In order to bridge these gaps, we aim to develop SR method by taking both CBS and CWS along with LCP between series into account. Again, in this study, deriving likelihood ratio statistics for detecting a mean shift in the multivariate process is our main interest, while a shift in variance might be considered later on.

According to whether CWS, CBS and LCP between series are considered, we categorize methods for sufficient reduction. Table 4.1 below summarizes the methods proposed by Wessman (1998) (W) and Frisén *et al.* (2011) (F) along with five new cases proposed in this chapter in order to make the technique more suitable for outbreak detection.
For independent observations (no CWS), we develop the SR method for detecting a mean shift when CBS and LCP are present (Case 1) (see Section 4.4). This will be a case of developing the SR method proposed by Wessman (1998) to allow for LCP between series. In Section 4.5, we focus on developing SR methods for detecting a mean shift in a process of dependent observations (CWS) under the different conditions when either CBS or LCP or both are present (Cases 2-5). Details of the forms of CWS, CBS and LCP considered in this study are provided in the following sections.

Table 4.1: Summary of existing SR methods and proposed SR methods

		CWS			
		No, i.e. indep. obs.		Yes, i.e. dep. obs.	
		CBS		CBS	
		No	Yes	No	Yes
Time lag	No	W and F	W	Case 2	Case 3
	Yes	F	Case 1	Case 4	Case 5

4.3.1 Correlation within series (CWS)

Due to the nature of health data, we aim to derive a SR method for monitoring a mean shift in a p dimensional multivariate series when CWS is present. The correlation within series will be represented by using the normal autoregressive model of order 1 (AR(1) model), where the current time point, x_t, is conditional on the previous time point, x_{t-1}. The simple AR(1) model is defined as

$$x_t = \phi x_{t-1} + \epsilon_t,$$

where $t = 1, 2, ..., s$, ϕ is an autoregressive coefficient $(-1 < \phi < 1)$ and ϵ_t are random white noise with $N(0, \sigma_\epsilon^2)$.

Consider the bivariate case, so the model for each series can be written as

$$\boldsymbol{x}_t = \boldsymbol{\phi}\boldsymbol{x}_{t-1} + \boldsymbol{\epsilon}_t,$$

where $\boldsymbol{x}_t = \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix}$, $\boldsymbol{\phi} = diag(\phi_1 \quad \phi_2) = \begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{pmatrix}$, $(-1 < \phi_i < 1,\ i = 1, 2)$, and $\boldsymbol{\epsilon}_t = \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$.

Note that the innovations $\epsilon_{1,t}$ and $\epsilon_{2,t}$ follow the assumptions below.

1 $E(\epsilon_{i,t}) = 0$

2 $E(\epsilon_{1,t}, \epsilon_{2,t}) = \gamma$

3 $E(\epsilon_{i,t}, \epsilon_{i,t-l}) = 0$ (i.e. no autocorrelation in the innovation series)

Therefore, the covariance matrix, $\boldsymbol{\Sigma}_\epsilon$, of innovations $\epsilon_{1,t}$ and $\epsilon_{2,t}$ is

$$\boldsymbol{\Sigma}_\epsilon = \begin{pmatrix} \sigma_{1\epsilon}^2 & \gamma \\ \gamma & \sigma_{1\epsilon}^2 \end{pmatrix},$$

where γ, the covariance between $\epsilon_{1,t}$ and $\epsilon_{2,t}$ at time t, is a constant. If two series are independent, $\gamma = 0$.

Because of the CWS, the likelihood function for $\boldsymbol{\mu}$ is defined as below. Note that the independent case (no CWS, $\phi_i = 0$) is a special case of this case when $\boldsymbol{x}_t$ does not depend on $\boldsymbol{x}_{t-1}$.

$$L(\boldsymbol{x}_t; \boldsymbol{\mu}) = f(\boldsymbol{x}_1; \boldsymbol{\mu}) \prod_{t=2}^{s} f(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}; \boldsymbol{\mu}),$$

where $t = 1, 2, ..., s$.

With regard to a stationary AR(1) process, the mean and variance of the process are defined below

$$\boldsymbol{\mu} = E(\mathbf{X}_t) = \mathbf{0}$$
$$\boldsymbol{\Sigma} = V(\mathbf{X}_t) = (1 - \boldsymbol{\phi})^{-1} \boldsymbol{\Sigma}_\epsilon (1 - \boldsymbol{\phi})^{-1}$$

Thus, for the initial value x_t, $t = 1$, we define

$$\boldsymbol{X}_t \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

The probability density function of the bivariate series observed at time $t = 1$ is

$$f(\boldsymbol{x}_1; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{2\pi |\boldsymbol{\Sigma}|^{1/2}} \exp\{-\frac{1}{2} (\boldsymbol{x}_1 - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_1 - \boldsymbol{\mu})\} \tag{4.7}$$

where $\boldsymbol{x}_1 = (x_{1,1} \quad x_{2,1})'$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$.

The conditional mean and conditional variance of the observation at time $t = 2, 3, ..., s$ is

$$\boldsymbol{\mu} = E(\boldsymbol{X}_t | \boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1}) = \boldsymbol{\phi} \boldsymbol{x}_{t-1}$$
$$\boldsymbol{\Sigma}_\epsilon = V(\boldsymbol{X}_t | \boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1}) = V(\boldsymbol{\epsilon}_t) = \boldsymbol{\Sigma}_\epsilon \tag{4.8}$$

where $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Sigma} - \boldsymbol{\phi} \boldsymbol{\Sigma} \boldsymbol{\phi}'$ (Reinsel, 1993; Lotze *et al.*, 2010). We will use this approach to calculate the conditional variance based on the prior estimate of Σ, although it could also be estimated by direct maximum likelihood approach, the method commonly used to estimate parameters in AR(1) model (Lütkepohl, 1993; Brockwell and Davis, 2002).

Thus, for $t = 2, 3, ..., s$,

$$\boldsymbol{X}_t | \boldsymbol{X}_{t-1} \sim N_2(\boldsymbol{\phi} \boldsymbol{x}_{t-1}, \boldsymbol{\Sigma}_\epsilon)$$

The conditional pdf's needed at times $t = 2, 3, ..., s$ are

$$f(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}; \boldsymbol{\mu}, \boldsymbol{\Sigma}_\epsilon) = \frac{1}{2\pi |\boldsymbol{\Sigma}_\epsilon|^{1/2}} \exp\{-\frac{1}{2} (\boldsymbol{x}_t - \boldsymbol{\phi} \boldsymbol{x}_{t-1})' \boldsymbol{\Sigma}_\epsilon^{-1} (\boldsymbol{x}_t - \boldsymbol{\phi} \boldsymbol{x}_{t-1})\}. \tag{4.9}$$

The pdf and conditional pdf will be used to derive a sequence of statistics according to the types of change points. SR methods developed for dependent data are provided in Section 4.5.

4.3.2 Correlation between series (CBS)

SR method have the advantage over parallel methods of taking CBS into account, which might improve detection capability. The relationships between the mean shifts of the series modelled through CBS can be different types, defined, for example, according to patient behaviours and disease natures. Let $\mathbf{c} = (c_1, c_2, ..., c_p)'$ be a vector of shift size we want to detect (i.e. c_i is the shift size in series i). Two situations for the relations between mean shifts in health data are described below.

Consider patient behaviour concerning how they react or treat themselves when they are infected. When people become ill, some of them may go to a hospital to receive medical treatment and the others may treat themselves by taking medicine from pharmacists. So the number of people infected might relate to values of other variables. For example, a shift in number of people infected might be reflected in a shift in number of phone calls to health centres, emergency department visits, pharmacy sales and work or school absenteeism, etc. Therefore, in this case we assume that a mean shift in one series might be reflected in a mean shift in the rest of the p dimensional multivariate series. The relations between mean shifts of the series are determined by the correlations between series 1. If ρ_{1i} is the correlation between series 1 (where the initial shift occurs; in this case the number infected) and series i ($i \neq 1$), the shift size in series i, c_i ($i \neq 1$), is determined by ρ_{1i} and the shift size in series 1, c_1. Therefore,

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{pmatrix} = \begin{pmatrix} c_1 \\ \rho_{12}\, c_1 \\ \vdots \\ \rho_{1p}\, c_1 \end{pmatrix}$$

If we consider the nature of a disease, the vector of shift size $\mathbf{c}$ may be defined in a different way. In this case, we assume that the shifts in series i occur sequentially and, therefore, the relations of mean shifts between the series also depend on CBS in the corresponding order. An example might be the meaning of how symptoms develop after disease exposure. For example, in the case of hand, foot and mouth disease, the early symptoms developed are fever and small red spots in the mouth. One or two days afterwards, the later symptoms are mouth lesions and finally a skin rash a further one or two days later (NHS, 2012). Therefore, if there is an increase in the number of people having a fever with red spots, it might reflected in a shift in the numbers of subsequent symptoms later on.

In this case, with suitable labelling of the series, the mean shift in series i is $c_i = \rho_{(i-1)i}\, c_{i-1}$. The vector of shift sizes is thus

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_p \end{pmatrix} = \begin{pmatrix} c_1 \\ \rho_{12}\, c_1 \\ \rho_{23}\, c_2 \\ \vdots \\ \rho_{(p-1)p}\, c_{p-1} \end{pmatrix}$$

Of course, in the bivariate case, the two scenarios are not distinguishable.

Figures 4.1 and 4.2 illustrate the simultaneous mean shifts in a bivariate process of independent and dependent observations, respectively. If there is no CBS (Figure 4.1 (a) and 4.2 (a)), the mean shift in the first series does not affect the mean shift in the second. On the other hand, the shift in the first series will be reflected in the second series if CBS is clearly present (Figure 4.1 (b) and 4.2 (b)). Here, c_2 depends on c_1 and the strength of CBS. In the case of non-simultaneous changes, the order of changes might need to be specified, however, the similar arguments to those above can be applied.

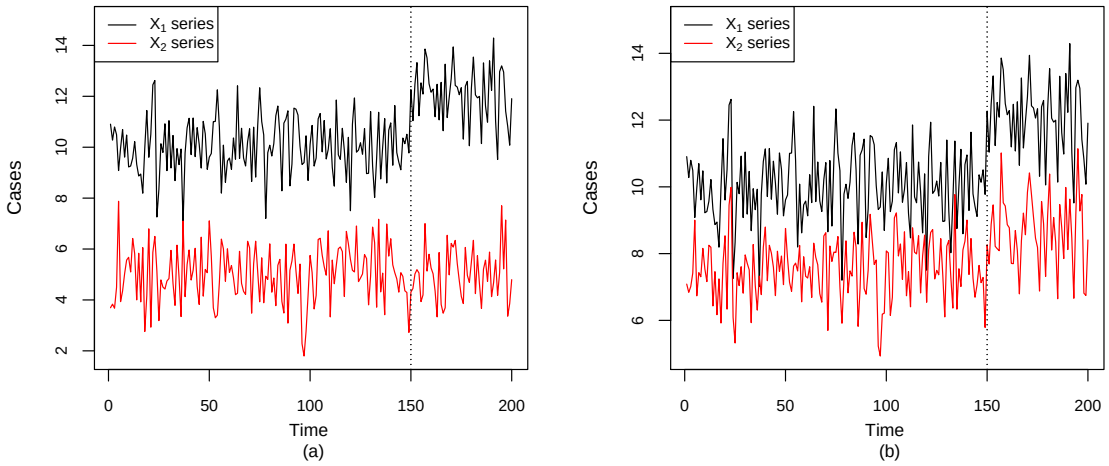

Figure 4.1: The illustration of relation of mean shift between series when there is no CWS ($\phi = 0$) with a) no CBS ($\rho = 0$) and b) CBS ($\rho = 0.6$). Shift occurs at time = 150

4.3.3 Types of changes

Two types of mean shifts in a multivariate process considered in this study are simultaneous changes and changes with time lags. In case of simultaneous changes, where the mean of series i $(i = 1, 2, .., p)$ has shifted from μ_i^I to μ_i^O at time τ (i.e. $\tau_1 = \tau_2 = ... = \tau_p = \tau$), the mean vector, $\boldsymbol{\mu}$, of in control and out of control stages is

$$\boldsymbol{\mu}_t = \begin{cases} \boldsymbol{\mu}^I = (\mu_1^I, ..., \mu_p^I)' & t < \tau \quad \text{in control stage} \\ \boldsymbol{\mu}^O = (\mu_1^O, ..., \mu_p^O)' & t \geq \tau \quad \text{out of control stage} \end{cases}$$

In case of changes with time lags, with suitable labelling of the series, we assume that changes occur sequentially $(\tau_1 < \tau_2 < ... < \tau_p)$ with known time lags l_i where $l_i = \tau_{i+1} - \tau_i$. The earliest change point is regarded as a change point a process denoted by $\tau = \tau_1 = min(\tau_1, \tau_2, ..., \tau_p)$. The mean vectors of in control stage, $\boldsymbol{\mu}^I$, and out of control

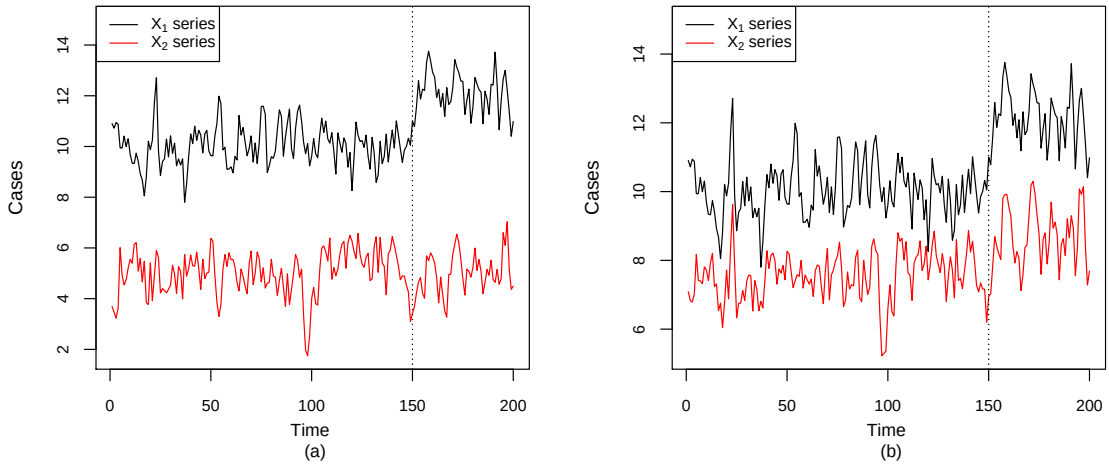

Figure 4.2: The illustration of relation of mean shift between series when there is CWS ($\phi = 0.6$) with a) no CBS ($\rho = 0$) and b) CBS ($\rho = 0.6$). Shift occurs at time = 150.

stage, $\boldsymbol{\mu}^O$, are

$$\boldsymbol{\mu}^I = (\mu_1^I, ..., \mu_p^I)' \qquad t < \tau_1$$

and

$$\boldsymbol{\mu}^O = \begin{cases} (\mu_1^O, \mu_2^I, \mu_3^I, ..., \mu_p^I)' & \tau_1 \leq t < \tau_2 \\ (\mu_1^O, \mu_2^O, \mu_3^I, ..., \mu_p^I)' & \tau_2 \leq t < \tau_3 \\ \cdots & \cdots \\ (\mu_1^O, \mu_2^O, ..., \mu_{p-1}^O, \mu_p^O)' & t \geq \tau_p \end{cases}$$

where $t = 1, 2, ..., s$.

4.3.4 Assumptions of proposed SR methods

In this study, we propose SR methods for detecting a mean shift in multivariate process by incorporating CWS, CBS and types of change points in sufficient reduction. For simplicity, the SR methods developed in this study are illustrated by bivariate normal series ($p = 2$) but most extend naturally (see the discussion in Chapter 8). The assumptions of the sufficient reduction methods proposed in this study are

- aim to detect a sudden, but persistent, shift in mean of bivariate process;
- data are bivariate normally distributed;
- allow for CWS and CBS;
- decision regarding detection of a mean shift will be made at each decision time point ($t = 1, 2, ..., s$);

- two types of change points are investigated. Simultaneous changes ($\tau_1 = \tau_2 = \tau$) and changes with time lags ($l = \tau_2 - \tau_1, \tau_2 > \tau_1,\ t = 1, 2, ..., \tau_1, ..., \tau_2, ..., s$);

- the mean shift size we aim to detect in series i is denoted by c_i which is pre-specified (i.e. the process mean of series i has shifted from μ_i^I to μ_i^O where $\mu_i^O = \mu_i^I + c_i$);

- due to the CBS denoted by ρ, the mean shift in one series will be reflected in another if CBS, ρ, is present (i.e. $c_2 = \rho c_1$);

- the variance of the process is unchanged over time;

- the sequence of statistics derived for detecting a mean shift for any s will be monitored by a one-sided EWMA chart with $\lambda = 0.3$ and $\text{ARL}_0 = 370$ (Fricker *et al.*, 2007; Joner *et al.*, 2008; Montgomery, 2009);

- the performance of detection will be measured by the conditional expected delay (CED), true alarm rate (TAR), false alarm rate (FAR) and non-detection rate (NDR) (see more details in Section 3.4).

The SR method developed for detecting a mean shift in a process of independent observations (no CWS) when CBS and LCP are present (case 1) is detailed in Section 4.4, while SR methods developed for a process of dependent observations (CWS) when either CBS or LCP or both are present (cases 2-5, see Table 4.1) are provided in Section 4.5.

4.4 SR methods for independent observations

As mentioned previously Wessman (1998) and Frisén *et al.* (2011) derived SR methods based on the assumption that observations in each series are independent (no CWS). However, the former allows for CBS but not for LCP, while the latter does not allow for CBS but does for LCP. In this section we provide a comprehensive analysis of all combinations from the previous studies. In the case of independent observations, we define the in control and out of control of series X_1 and X_2 in the bivariate normal case as follows.

$$X_{1,t} \sim \begin{cases} N(\mu_1^I, \sigma_1^2) & t < \tau_1 \quad \text{in control stage} \\ N(\mu_1^O, \sigma_1^2) & t \geq \tau_1 \quad \text{out of control stage} \end{cases}$$

$$X_{1,t} \sim \begin{cases} N(\mu_2^I, \sigma_2^2) & t < \tau_2 \quad \text{in control state} \\ N(\mu_2^O, \sigma_2^2) & t \geq \tau_2 \quad \text{out of control stat} \end{cases}$$

where $t = 1, 2, ..., \tau_1, ..., \tau_2, ..., s$ ($\tau_2 = \tau_1 + l$) and $\mu_i^O = \mu_i^I + c_i$. The covariance matrix between X_1 and X_2, which is unchanged over time, is denoted by $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$.

4.4.1 Case 1 (No CWS but CBS and LCP)

Here we develop a SR method for detecting a mean shift in bivariate normal series when CBS and LCP between two series, X_1 and X_2, are present. LCP is defined as the known time lag of change points between two series ($l = \tau_2 - \tau_1$, where $\tau_2 > \tau_1$). Of course, putting $l = 0$ give the 'CBS but no lag' case.

Regarding the LCP between series, like the Frisén method, $x_{1,t}$ and $x_{2,t+l}$ are used in SR method instead of $x_{1,t}$ and $x_{2,t}$. For convenience we then define

$$\boldsymbol{y}_t = \begin{pmatrix} x_{1,t} \\ x_{2,t+l} \end{pmatrix}$$

Incorporating the known time lag l by using $\boldsymbol{y}_t$ instead of $\boldsymbol{x}_t$ makes detecting changes with the known time lag equivalent to the case of detecting simultaneous change, since both changes (at τ_1 and $\tau_2 = \tau_{1+l}$) occur at the same time at time τ_1, if $\boldsymbol{y}_t$ is used. Therefore in the case of LCP, τ_1 is defined as a change point of the process.

Due to the data available in this case (see Section 4.2.4), the pdf of in control and out of control states according the change points, τ_1, are defined as follow.

$$f(\boldsymbol{y}_t; \boldsymbol{\mu}) = \begin{cases} f(\boldsymbol{y}_t; \boldsymbol{\mu}^I) & t = 1, ..., \tau_1 - 1 \quad \text{in control stage} \\ f(\boldsymbol{y}_t; \boldsymbol{\mu}^O) & t = \tau_1, ..., s - l \quad \text{out of control stage} \\ f(x_{1,t}; \mu_1^O) & t = s - l + 1, ..., s \quad \text{out of control stage} \end{cases}$$

where $\boldsymbol{\mu}^O = \boldsymbol{\mu}^I + \mathbf{c}$, $\boldsymbol{\mu}^I = \begin{pmatrix} \mu_1^I \\ \mu_2^I \end{pmatrix}$, $\boldsymbol{\mu}^O = \begin{pmatrix} \mu_1^O \\ \mu_2^O \end{pmatrix}$ and $\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$. The covariance matrix, $\boldsymbol{\Sigma}_y$, is unchanged over time, $\boldsymbol{\Sigma}_y = \begin{pmatrix} \sigma_1^2 & \sigma_y \\ \sigma_y & \sigma_2^2 \end{pmatrix}$, where σ_y is the covariance between $x_{1,t}$ and $x_{2,t+l}$.

The pdf of the bivariate process defined according to the change points are given in the subsequent sections as follows.

4.4.1.1 Stage 1: in control stage

At time $t = 1, 2, ..., \tau_1 - 1$, X_1 and X_2 are both in control.

$$\boldsymbol{Y}_t \sim N_2(\boldsymbol{\mu}^I, \boldsymbol{\Sigma}_y)$$

The pdf of bivariate process can be written as

$$f(\boldsymbol{y}_t; \boldsymbol{\mu}^I, \boldsymbol{\Sigma}_y) = \frac{1}{2\pi |\boldsymbol{\Sigma}_y|^{1/2}} \exp\{-\frac{1}{2} \left(\boldsymbol{y}_t - \boldsymbol{\mu}^I\right)' \boldsymbol{\Sigma}_y^{-1} \left(\boldsymbol{y}_t - \boldsymbol{\mu}^I\right)\}$$

where $t = 1, 2, ..., \tau_1 - 1$.

4.4.1.2 Stage 2: out of control stage while $x_{2,t+l}$ is still available

At time $t = \tau_1, ..., \tau_{1+l}, ..., s-l$, X_1 and X_2 are both out of control.

$$\boldsymbol{Y}_t \sim N_2(\boldsymbol{\mu}^O, \boldsymbol{\Sigma}_y)$$

At this stage the mean of the X_1 series has shifted from μ_1^I to μ_1^O ($\mu_1^O = \mu_1^I + c_1$) at time $t = \tau_1$. The mean of the X_2 series has shifted from μ_2^I to μ_2^O ($\mu_2^O = \mu_2^I + c_2,\ c_2 = \rho c_1$) at time $\tau_2 = \tau_1 + l$. The pdf of the bivariate process can be written as

$$f(\boldsymbol{y}_t; \boldsymbol{\mu}^O, \boldsymbol{\Sigma}_y) = \frac{1}{2\pi|\boldsymbol{\Sigma}_y|^{1/2}} \exp\{-\frac{1}{2}\left(\boldsymbol{y}_t - \boldsymbol{\mu}^O\right)' \boldsymbol{\Sigma}_y^{-1} \left(\boldsymbol{y}_t - \boldsymbol{\mu}^O\right)\}$$

where $t = \tau_1, \tau_1 + 1, ..., \tau_1 + l, ..., s - l$.

4.4.1.3 Stage 3: out of control stage while only $x_{1,t}$ is available

At time $t = s - l + 1, ..., s$, at this stage X_1 is still out of control, but data from series X_2, $x_{2,t+l}$, are not yet available. Therefore data from X_1 only will be used, having univariate normal distribution.

$$X_{1,t} \sim N(\mu_1^O, \sigma_1^2)$$

where $\mu_1^O = \mu_1^I + c_1$. The pdf of univariate normal distribution can be written as

$$f(x_{1,t}; \mu_1^O, \sigma_1^2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\{-\frac{1}{2\sigma_1^2}(x_{1,t} - \mu_1^O)^2\}$$

where $t = s - l + 1, ..., s$.

From equation (4.1), the likelihood function of process according to the change points and LCP between series can be illustrated as

$$\begin{aligned}
L(\boldsymbol{y}_t; \boldsymbol{\mu}|\tau_1) &= \prod_{t=1}^{\tau_1 - 1} f^I(\boldsymbol{y}_t; \boldsymbol{\mu}^I) \prod_{t=\tau_1}^{s-l} f^O(\boldsymbol{y}_t; \boldsymbol{\mu}^O) \prod_{t=s-l+1}^{s} f^O(x_{1,t}; \mu_1^O) \\
&= \prod_{t=1}^{s-l} f^I(\boldsymbol{y}_t; \boldsymbol{\mu}^I) \prod_{t=s-l+1}^{s} f^I(x_{1,t}; \mu_1^I) \prod_{t=\tau_1}^{s-l} \frac{f^O(\boldsymbol{y}_t; \boldsymbol{\mu}^O)}{f^I(\boldsymbol{y}_t; \boldsymbol{\mu}^I)} \prod_{t=s-l+1}^{s} \frac{f^O(x_{1,t}; \mu_1^O)}{f^I(x_{1,t}; \mu_1^I)} \\
&= h(\mathbf{x}_s) \prod_{t=\tau_1}^{s-l} \frac{f^O(\boldsymbol{y}_t; \boldsymbol{\mu}^O)}{f^I(\boldsymbol{y}_t; \boldsymbol{\mu}^I)} \prod_{t=s-l+1}^{s} \frac{f^O(x_{1,t}; \mu_1^O)}{f^I(x_{1,t}; \mu_1^I)} \\
&= h(\mathbf{x}_s)\ \mathbf{lr}_{\tau_1}(\mathbf{x}_s; \boldsymbol{\mu}^O - \boldsymbol{\mu}^I) \qquad (4.10)
\end{aligned}$$

where $t = 1, 2, ..., \tau_1, ..., \tau_1 + l, ..., s - l, ..., s$ (i.e. $\tau_2 = \tau_1 + l$).

Similarly to the Wessman argument, $\mathbf{lr}_{\tau_1}(\mathbf{x}_s)$ is a sufficient sequence for monitoring a mean shift with known time lag in a process for any decision time s. By the factorization theorem, $T_t(\boldsymbol{x}_t)$ is a likelihood ratio statistic for detecting a mean shift in process at time

t. Therefore, the sufficient sequence $\mathbf{lr}_1(\mathbf{x}_s)$ for monitoring a mean shift with time lag in the process is the sequence

$$\mathbf{lr}_1(\mathbf{x}_s) = \{T_1(\boldsymbol{x}_1), ..., T_s(\boldsymbol{x}_s), \}$$

where

$$T_t(\boldsymbol{x}_t) = \begin{cases} \mathbf{c}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_t & t = 1, 2, ..., s-l \\ c_1 x_{1,t}/\sigma_1^2 & t = s-l+1, ..., s \end{cases} \tag{4.11}$$

where $c_2 = \rho c_1$. The derivation is detailed in Section A.1 in Appendix A.

4.5 SR methods for dependent observations (CWS)

In order to investigate a mean shift in the bivariate time series, in and out of control stages are defined according to a change point in each series (τ_1 and τ_2). During the in control state, the conditional mean for series i is $\mu_i^I = \phi_i x_{i,t-1}$, while during the out of control state, the conditional mean for series i is $\mu_i^O = \phi_i x_{i,t-1} + c_i$, where $i = 1, 2$ and c_i is the constant shift size we aim to detect in series i. Thus

$$X_{1,t}|X_{1,t-1} \sim \begin{cases} N(\mu_1^I, \sigma_{1\epsilon}^2) & t < \tau_1 \quad \text{in control state} \\ N(\mu_1^O, \sigma_{1\epsilon}^2) & t \geq \tau_1 \quad \text{out of control state} \end{cases}$$

$$X_{2,t}|X_{2,t-1} \sim \begin{cases} N(\mu_2^I, \sigma_{2\epsilon}^2) & t < \tau_2 \quad \text{in control state} \\ N(\mu_2^O, \sigma_{2\epsilon}^2) & t \geq \tau_2 \quad \text{out of control state} \end{cases}$$

where $t = 1, 2, ..., \tau_1, ..., \tau_2, ..., s$ and $\tau_1 \leq \tau_2$). The covariance matrix, $\boldsymbol{\Sigma}$, for $t = 1$ and the conditional covariance matrix, $\boldsymbol{\Sigma}_\epsilon$, for $t = 2, 3, ..., s$ are unchanged over time.

The sufficient reduction methods proposed for detecting a mean shift in bivariate time series when either LCP or CBS or both are present (cases 2, 3, 4 and 5; see Table 4.1) will be detailed in the following sections.

4.5.1 Case 2 (CWS but no CBS or LCP)

Assuming that there is no CBS or LCP between series, then changes occur at the same time τ ($\tau_1 = \tau_2 = \tau$) and $\tau > 1$. To derive a sequence of likelihood ratio statistics, we define the pdf of the bivariate series ($t = 1$) and the conditional pdf of bivariate series ($t = 2, ..., s$) for in and out of control stages according to change point τ as follows

$$f(\boldsymbol{x}_t; \boldsymbol{\mu}) = \begin{cases} f^I(\boldsymbol{x}_1; \mu^I) & t = 1 \quad \text{in control state} \\ f^I(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \boldsymbol{\mu}^I) & t < \tau \quad \text{in control state} \\ f^O(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \boldsymbol{\mu}^O) & t \geq \tau \quad \text{out control state} \end{cases}$$

4.5.1.1 Stage 1: the initial in control stage

At time $t = 1$, an in control stage, X_1 and X_2 are both in control.

$$\mathbf{X}_t \sim N_2(\boldsymbol{\mu}^I, \boldsymbol{\Sigma})$$

The pdf of the series is similar to the pdf defined in equation (4.7) with mean $\boldsymbol{\mu}^I = \begin{pmatrix} \mu_1^I & \mu_2^I \end{pmatrix}'$ and variance $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$ since there is no CBS.

4.5.1.2 Stage 2: in control stage

At time $t = 2, 3, ..., \tau - 1$, the process is still in control.

$$\mathbf{X}_t|\mathbf{X}_{t-1} \sim N(\boldsymbol{\mu}^I, \boldsymbol{\Sigma}_\epsilon)$$

The conditional pdf of the process is similar to the conditional pdf defined in equation (4.9), where the condition mean and variance are $\boldsymbol{\mu}^I = E(\boldsymbol{X}_t|\boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1}) = \boldsymbol{\phi}\boldsymbol{x}_{t-1}$ and $\boldsymbol{\Sigma}_\epsilon = V(\boldsymbol{X}_t|\boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1}) = V(\boldsymbol{\epsilon}_t) = \boldsymbol{\Sigma} - \boldsymbol{\phi}\boldsymbol{\Sigma}\boldsymbol{\phi}'$, respectively.

4.5.1.3 Stage 3: out of control stage

At time $t = \tau, \tau + 1, ..., s$,

$$\mathbf{X}_t|\mathbf{X}_{t-1} \sim N(\boldsymbol{\mu}^O, \boldsymbol{\Sigma}_\epsilon)$$

The process is out of control, the conditional mean changes from $\boldsymbol{\mu}^I$ to $\boldsymbol{\mu}^O = E(\boldsymbol{X}_t|\boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1}) = \boldsymbol{\phi}\boldsymbol{x}_{t-1} + \mathbf{c}$, though the conditional covariance matrix, $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Sigma} - \boldsymbol{\phi}\boldsymbol{\Sigma}\boldsymbol{\phi}'$, is constant over time. Assume that $\boldsymbol{c} = \begin{pmatrix} c_1 & c_2 \end{pmatrix}'$ is the vector of constant shift sizes in the means of the X_1 and X_2 series which we aim to detect.

From the bivariate normal distribution defined at each stage, the likelihood function of of bivariate series given that a mean shift occurs at τ can be defined as

$$\begin{aligned} L(\boldsymbol{x}_t; \boldsymbol{\mu}|\tau) &= f^I(\boldsymbol{x}_1; \boldsymbol{\mu}_0) \prod_{t=2}^{\tau-1} f^I(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \boldsymbol{\mu}^I) \prod_{t=\tau}^{s} f^O(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \boldsymbol{\mu}^O) \\ &= f^I(\boldsymbol{x}_1; \boldsymbol{\mu}_0) \prod_{t=2}^{s} f^I(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \boldsymbol{\mu}^I) \prod_{t=\tau}^{s} \frac{f^O(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \boldsymbol{\mu}^O)}{f^I(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \boldsymbol{\mu}^I)} \\ &= h(\mathbf{x}_s) \prod_{t=\tau}^{s} lr(\boldsymbol{x}_t; \boldsymbol{\mu}^O - \boldsymbol{\mu}^I) \\ &= h(\mathbf{x}_s)\ k(\mathbf{lr}_\tau(\mathbf{x}_s; \boldsymbol{\mu}^O - \boldsymbol{\mu}^I)), \end{aligned} \tag{4.12}$$

where $t = 1, 2, ..., \tau, ..., s$.

From Section 4.2.3, we know a sufficient sequence $\mathbf{lr}_1(\mathbf{x}_s)$ for monitoring a mean shift in a process when CWS is present is

$$\mathbf{lr}_1(\mathbf{x}_s) = \{T_1(x_1), ..., T_s(x_s)\}$$

and

$$T_t(\boldsymbol{x}_t) = \mathbf{c}'\boldsymbol{\Sigma}_\epsilon^{-1} \begin{pmatrix} x_{1,t} - \phi_1 x_{1,t-1} \\ x_{2,t} - \phi_2 x_{1,t-1} \end{pmatrix},$$

where $t = 1, 2, ..., s$ and $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Sigma} - \boldsymbol{\phi}\boldsymbol{\Sigma}\boldsymbol{\phi}'$. The full derivation for this case is provided in Section A.2 in Appendix A.

4.5.2 Case 3 (CWS and CBS but no LCP)

If CBS is present, the likelihood ratio statistic, $T_t(\boldsymbol{x}_t)$, used for detecting a simultaneous change (no LCP) in the bivariate series is a generalization of case 2 when CBS is not zero (i.e. $\sigma_{12} \neq 0$) but this has no effect on the argument. Therefore, the likelihood ratio statistics derived for this case are

$$T_t(\boldsymbol{x}_t) \;=\; \mathbf{c}'\boldsymbol{\Sigma}_\epsilon^{-1} \begin{pmatrix} x_{1,t} - \phi_1 x_{1,t-1} \\ x_{2,t} - \phi_2 x_{1,t-1} \end{pmatrix} \tag{4.13}$$

where $t = 1, 2, ..., s$, $c_2 = \rho c_1$ and $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Sigma} - \boldsymbol{\phi}\boldsymbol{\Sigma}\boldsymbol{\phi}'$, $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$.

4.5.3 Case 4 (CWS, LCP but no CBS)

Like case 1, assuming that time lag between change points is known ($l = \tau_2 - \tau_1$), we then monitor $\boldsymbol{y}_t$ rather than $\boldsymbol{x}_t$. Also, due to the known time lag and the lack of availability of the data from the later parts of series X_2 (see Section 4.2.4), the initial pdf for $t = 1$ and the conditional pdf for $t = 2, 3, ..., s$ of bivariate series defined for in control and out of control stages given that the change point of the process is $\tau_1 = min(\tau_1, \tau_2)$ and LCP between series are summarized as follows.

$$f(\boldsymbol{y}_t; \boldsymbol{\mu}) = \begin{cases} f^I(\boldsymbol{y}_t; \boldsymbol{\mu}^I) & t = 1 & \text{in control state} \\ f^I(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}; \boldsymbol{\mu}^I) & t = 2, ..., \tau_1 - 1 & \text{in control state} \\ f^O(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}; \boldsymbol{\mu}^O) & t = \tau_1, ..., s - l & \text{out of control state} \\ f^O(x_{1,t} | x_{1,t-1}; \mu_1^O) & t = s - l + 1, ..., s & \text{out of control state} \end{cases}$$

where $\boldsymbol{\mu}^O = \boldsymbol{\mu}^I + \mathbf{c}$ and $t = 1, 2, ..., s$.

4.5.3.1 Stage 1: the initial in control stage

At time $t = 1$, the pdf of the bivariate series can be written in a similar form defined in equation (4.7), where $\boldsymbol{y}_t = \begin{pmatrix} x_{1,t} & x_{2,t+l} \end{pmatrix}'$, $\boldsymbol{\mu}^I = \begin{pmatrix} \mu_1^I & \mu_2^I \end{pmatrix}'$ and due to no CBS, $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$.

4.5.3.2 Stage 2: in control stage

At time $t = 2, 3, ..., \tau_1 - 1$, X_1 and X_2 are both in control.

$$\mathbf{Y}_t | \mathbf{Y}_{t-l} \sim N(\boldsymbol{\mu}^I, \boldsymbol{\Sigma}_\epsilon)$$

The conditional mean and variance of the bivariate series are

$$\boldsymbol{\mu}^I = E(\, \boldsymbol{Y}_t |\, \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}) = \boldsymbol{\phi}\boldsymbol{y}_{t-1}$$

$$\boldsymbol{\Sigma}_\epsilon = V(\, \boldsymbol{Y}_t |\, \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}) = \boldsymbol{\Sigma} - \boldsymbol{\phi}\boldsymbol{\Sigma}_y\boldsymbol{\phi}''$$

The conditional pdf of the series is similar to that in equation (4.9).

4.5.3.3 Stage 3: out of control stage while $x_{2,t+l}$ is still available

At time $t = \tau_1, ..., \tau_1 + l, ..., s - l$, both X_1 and X_2 are out of control. Thus

$$\mathbf{Y}_t | \mathbf{Y}_{t-l} \sim N(\boldsymbol{\mu}^O, \boldsymbol{\Sigma}_\epsilon).$$

The conditional mean and variance are

$$\boldsymbol{\mu}^O = E(\boldsymbol{Y}_t | \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}) = \boldsymbol{\phi}\boldsymbol{y}_{t-1} + \mathbf{c}$$

$$\boldsymbol{\Sigma}_\epsilon = V(\boldsymbol{Y}_t | \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}) = \boldsymbol{\Sigma} - \boldsymbol{\phi}\boldsymbol{\Sigma}_y\boldsymbol{\phi}'$$

With the parameters defined above, the conditional pdf of the series is similar to that in equation (4.9).

4.5.3.4 Stage 4: out of control stage while only $x_{1,t}$ is available

At time $t = s - l + 1, ..., s$, at this stage X_1 is still out of control, while data from X_2 series, $x_{2,t+l}$, are not yet available. We have

$$X_{1,t} | X_{1,t-1} \sim N(\mu_1^O, \sigma_{1\epsilon}^2)$$

The conditional mean and variance of $x_{1,t}$ at this stage will be

$$\mu_1^O = E(X_{1,t} | X_{1,t-1}) = \phi_1 x_{1,t-1} + c_1$$
$$\sigma_{1\epsilon}^2 = Var(X_{1,t} | X_{1,t-1}) = Var(\epsilon_{1,t})$$

The conditional pdf of X_1 series at time $t = s - l + 1, ..., s$ can be written as

$$f(x_{1,t} | x_{1,t-1}; \mu_1, \sigma_{1\epsilon}^2) = \frac{1}{\sqrt{2\pi\sigma_{1\epsilon}^2}} \exp\{-\frac{1}{2\sigma_{1\epsilon}^2}(x_{1,t} - (\phi_1 x_{1,t-1} + c_1))^2\}$$

where $t = s - l + 1, ..., s$.

From the pdfs defined above, the likelihood function of the series when there is LCP between series can be written as

$$\begin{aligned} L(\boldsymbol{y}_t; \boldsymbol{\mu} | \tau_1) &= f^I(\boldsymbol{y}_1; \boldsymbol{\mu}^I) \prod_{t=2}^{\tau_1 - 1} f^I(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}; \boldsymbol{\mu}^I) \prod_{t=\tau_1}^{s-l} f^O(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}; \boldsymbol{\mu}^O) \prod_{t=s-l+1}^{s} f^O(x_{1,t} | x_{1,t-1}; \mu_1^O) \\ &= h(\mathbf{x}_s) \prod_{t=\tau_1}^{s-l} \frac{f^O(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}; \boldsymbol{\mu}^O)}{f^I(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}; \boldsymbol{\mu}^I)} \prod_{t=s-l+1}^{s} \frac{f^O(x_{1,t} | x_{1,t-1}; \mu_1^O)}{f^I(x_{1,t} | x_{1,t-1}; \mu_1^I)} \\ &= h(\mathbf{x}_s) \, \mathbf{lr}_{\tau_1}(\mathbf{x}_s; \boldsymbol{\mu}^O - \boldsymbol{\mu}^I) \end{aligned} \tag{4.14}$$

From Section 4.2.4 and 4.4.1, a sequence $\mathbf{lr}_{\tau_1}(\mathbf{x}_s; \boldsymbol{\mu}^O - \boldsymbol{\mu}^I)$ is sufficient for detecting a mean shift with known time lag l. By the factorization theorem, the sufficient sequence for monitoring a mean shift in a process when CWS and LCP are both present is

$$\mathbf{lr}_1(\mathbf{x}_s; \boldsymbol{\mu}^O - \boldsymbol{\mu}^I) = \{T_1(\boldsymbol{x}_1), ..., T_s(\boldsymbol{x}_s)\}$$

and

$$T_t(\boldsymbol{x}_t) = \begin{cases} \mathbf{c}'\boldsymbol{\Sigma}_\epsilon^{-1}(\mathbf{y}_t - \boldsymbol{\phi}\mathbf{y}_{t-1}) & t = 1, ..., s-l \\ \\ c_1(x_{1,t} - \phi x_{1,t-1})/\sigma_{1\epsilon}^2 & t = s-l+1, ..., s \end{cases}$$

where $t = 1, 2, ..., \tau_1, ..., s-l, ..., s$, $c_2 = \rho c_1$ and $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Sigma} - \boldsymbol{\phi}\boldsymbol{\Sigma}_y\boldsymbol{\phi}'$, $\boldsymbol{\Sigma}_y = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$. The derivation of sufficient statistics for case 4 is also provided in Section A.3 in Appendix A.

4.5.4 Case 5 (CWS, LCP and CBS)

The sequence of likelihood ratio statistics for monitoring a change with time lag in bivariate time series when both CWS and CBS are present is again generalization of case 4 when CBS is not zero. Again, the generalization does not affect the sufficiency argument, therefore, the sequence derived for detecting a mean shift in bivariate series, for any s, in this case is

$$\mathbf{c}'\boldsymbol{\Sigma}_\epsilon^{-1}(\mathbf{y}_t - \boldsymbol{\phi}\mathbf{y}_{t-1}) \qquad t = 1, ..., s-l$$

$$c_1(x_{1,t} - \phi x_{1,t-1})/\sigma_{1\epsilon}^2 \qquad t = s-l+1, ..., s$$

where $t = 1, 2, ..., s-l, ..., s$. $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Sigma} - \boldsymbol{\phi}\boldsymbol{\Sigma}_y\boldsymbol{\phi}'$ where $\boldsymbol{\Sigma}_y = \begin{pmatrix} \sigma_1^2 & \sigma_y \\ \sigma_y & \sigma_2^2 \end{pmatrix}$.

4.6 System evaluation

The evaluation of the SR methods proposed in this chapter is conducted using a simulation study. Bivariate series are generated from the bivariate normal distribution with mean and variance defined below.

$$\mathbf{X}_t \sim N_2\left(\begin{pmatrix} 10 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix}\right)$$

Since the variance in each series is one, the correlation between series (CBS) is equal to the covariance, i.e. $\rho = \sigma_{12}$. Data are also generated with different values of CWS and CBS and types of change points between series in order to investigated the detection performances under different conditions.

- CBS: $\rho = 0, 0.2, 0.4$ or 0.6
- CWS: $\phi_i = 0, 0.2, 0.4$ or 0.6 (i.e. for simplicity, each series has the same ϕ_i)
- LCP: simultaneous change ($l = 0$) and change with time lag ($l = 5$)

The mean shift size we aim to detect in this study is set at either two and three times the standard deviation of original series in the in control stage. Therefore, in case of mean shift size two,

$$\mathbf{c} = \begin{pmatrix} c_1 & \rho c_1 \end{pmatrix}' = \begin{pmatrix} 2 & 2\rho \end{pmatrix}'$$

A shift of size $\mathbf{c}$ is added to the data at and after time τ. In the case of lag between series (l), a shift is added to the first series at time τ and to the second series at time $\tau + l$. Note that, in this chapter, we aim to evaluate the detection performance of SR methods for detecting a pre-specified shift size (i.e. 2 or 3 time standard deviation of the original series in the in control stage). Therefore, a shift size, $\mathbf{c}$, is pre-specified and so the data with the added signals are generated according to $\mathbf{c}$. The performance of such methods for detecting a shift size which might be smaller or or greater than $\mathbf{c}$ will be investigated in Chapter 5.

The performance of detection of our proposed methods are compared against three other methods.

- Parallel method adjusted for multiplicity (see Section 4.2.6)
- Frisén (LCP but no CWS and no CBS) (Frisén *et al.*, 2011)
- Wessman (CBS but no CWS or LCP) (Wessman, 1998)

A one-sided EWMA chart is used as a tool for detecting a shift in the process (see Section 4.2.5). Several statistics are used to evaluate and compare the detection performance of each method, namely conditional expected delay (CED), true alarm rate (TAR), false alarm rate (FAR) and non-detection rate (NDR). Let t_A be the time of an alarm, then define:

$$\begin{aligned} \text{CED} &= E(t_A - \tau | t_A \geq \tau) \\ \text{FAR} &= P(t_A < \tau) \\ \text{TAR} &= P(t_A - \tau \leq 6 | t_A \geq \tau) \\ \text{NDR} &= P(t_A - \tau \geq 7 | t_A \geq \tau) = 1 - \text{FAR} - \text{TAR} \end{aligned}$$

More details of measurements for system evaluation are provided in Section 3.4. The detection performance between the methods are investigated by comparing the differences of these statistics estimated from the simulations against the critical values defined in Section 3.6.

The results are from 10,000 simulations. The system evaluation for three different scenarios are provided in subsequent sections with some illustrations from the simulation study. Both types of change points will be considered in each scenario. The three scenarios are

- scenario 1: CBS but no CWS
- scenario 2: CWS but no CBS
- scenario 3: CWS and CBS

In each scenario, examples of the EWMA charts are illustrated. The vertical dotted line indicates the change point of the process and the horizontal dashed line is the upper control limit. The comparison of the detection performance between four methods is illustrated in bar charts, where P, F, W and Ck stand for Parallel, Frisén, Wessman and our proposed method Case k, respectively. The original results from the simulation study are given in Tables A.1 - A.6 in Appendix A.

4.6.1 Scenario 1 (CBS but no CWS)

We assume that observations in each series are independent (no CWS). Three levels of CBS (0.2, 0.4 and 0.6) and two types of change points (simultaneous change ($l = 0$) and changes with time lag ($l = 5$)) are considered. This scenario is used to investigate the effect of not taking CBS into account in the Frisén method and not taking LCP into account in the Wessman method. The bar charts comparing the detection performance for detecting simultaneous changes and changes with time lag are shown in Figures 4.3 and 4.4, respectively, while the original results from the simulation are given in Tables A.1 and A.2, respectively, in Appendix A.

Major differences are only apparent in CED; the reader is referred to Tables A.1 and A.2 (in Appendix A) to observe slight differences in $\widehat{\text{far}}$ and $\widehat{\text{ndr}}$. In the case of simultaneous change, if CBS is low, there is no difference in the performance between Frisén and Wessman methods. However, if CBS is moderate or high, Wessman method perform significantly differently from Frisén method by giving shorter $\widehat{\text{ced}}$, while there is no significant difference in TAR, FAR and NDR between the methods.

On the other hand, in the case of change with time lag, there is no difference in all measures between Frisén and our method (case 1) since both are accounted for the time lag. The CEDs of Frisén and our methods are significantly different from the others. Ignoring the LCP between series in Wessman method gives a longer $\widehat{\text{ced}}$ compared to Frisén and our methods (Table A.2). Figure 4.5 shows the EWMA charts of the four methods for detecting changes with time lag. The parallel method (Figure 4.5 (b-c)) and Wessman method (Figure 4.5 (e)) give longer detection than the others.

For both cases, the parallel method performs significantly differently from other methods by giving a longer delay, although generally a lower $\widehat{\text{far}}$, compared to the others. This improved $\widehat{\text{far}}$ is because the parallel method is adjusted for multiplicity, so its upper control limit is slightly higher than those of other methods. Also, unsurprisingly, the larger shift size is more likely to be detected than smaller shift size; all methods perform better in detecting a shift of 3 times the standard deviation of the original data than one of size 2.

4.6.2 Scenario 2 (CWS but no CBS)

We investigate the performance of detection when CWS is clearly present in each series, but with no CBS. The examples of the EWMA chart for detecting simultaneous changes and changes with time lag are illustrated in Figures 4.6 and 4.7, respectively. Since evidence of CWS is still present in the sequence of statistics derived by older methods, an alarm is flagged before a true change occurs which is regarded as a false alarm. Both Figures are examples of false alarms produced by the parallel, Frisén and Wessman methods, while the proposed methods (cases 2 and 4) have less chance of producing false alarms as CWS is accounted for in the sufficient reduction. This effect can be seen from the plots of autocorrelation function of the likelihood ratio statistics of the four methods used for detecting the simultaneous changes (Figure A.1) and changes with time lag

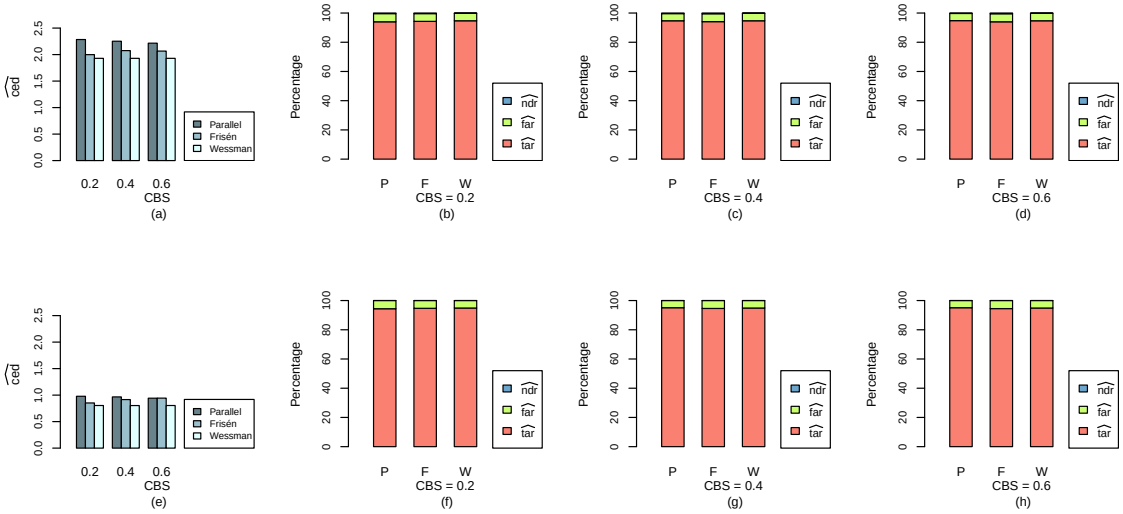

Figure 4.3: A bar chart comparing the performance in detecting simultaneous changes in scenario 1: (a) - (d) shift size = 2 and (e) - (h) shift size = 3.

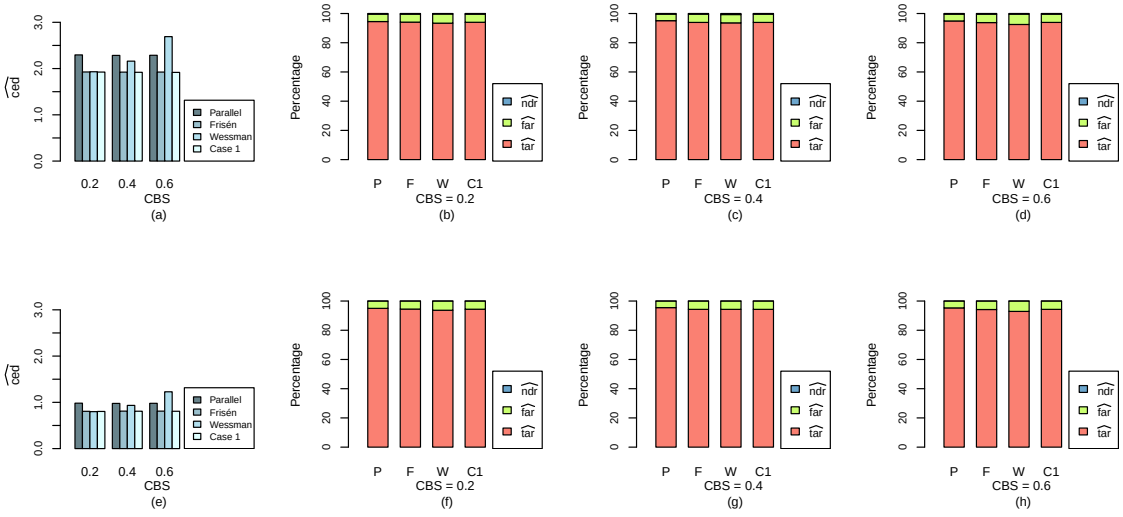

Figure 4.4: A bar chart comparing the performance in detecting changes with time lag in scenario 1: (a) - (d) shift size = 2 and (e) - (h) shift size = 3.

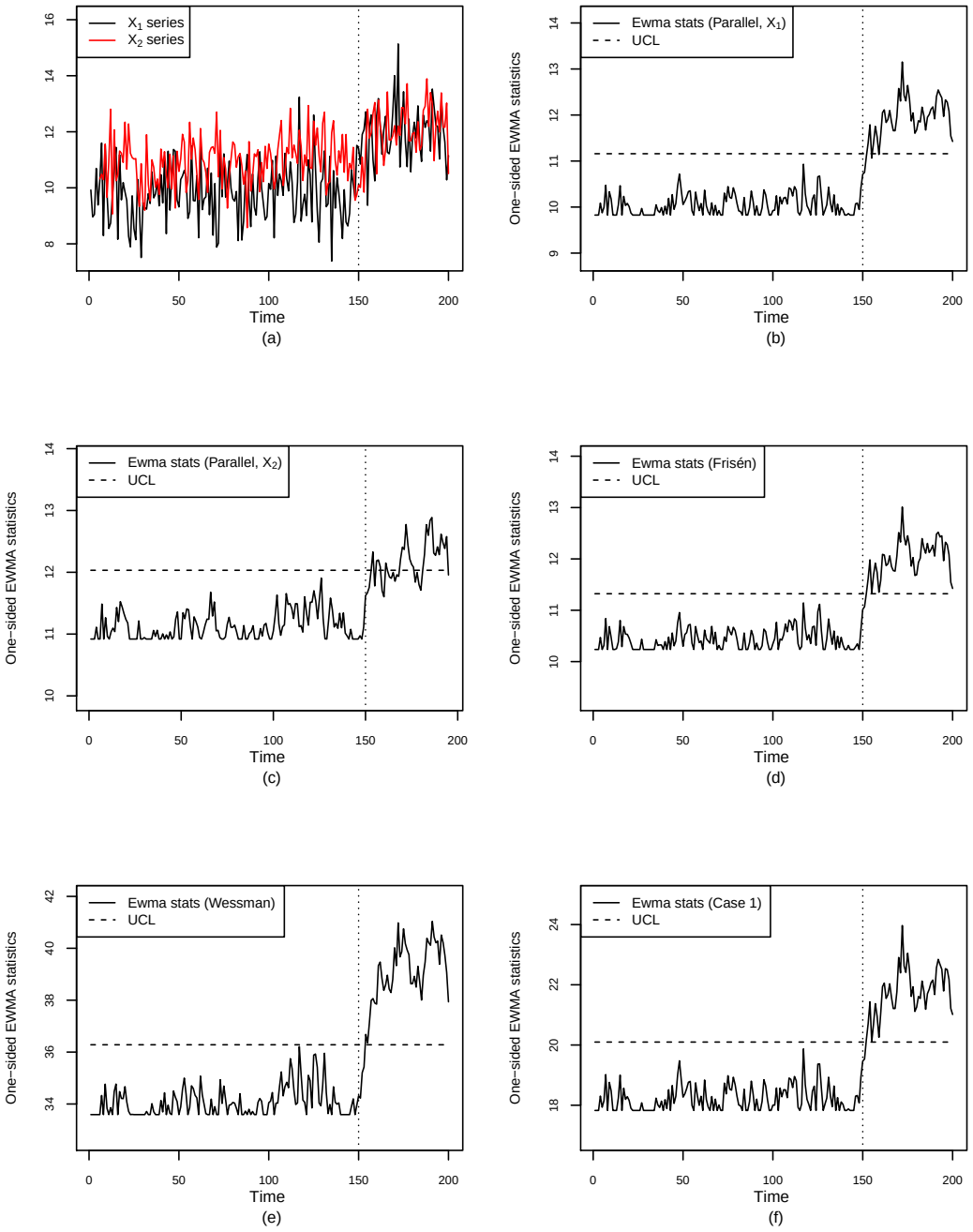

Figure 4.5: An example of scenario 1 with changes with time lag (shift size 2, no CWS and CBS = 0.6): (a) plot of the original series, (b) - (c) one-sided EWMA charts of parallel methods, (d), (e) and (f) one-sided EWMA charts of Frisén, Wessman and proposed method (case 1), respectively.

(Figure A.2) in Appendix A.

The bar charts comparing the detection performance for detecting simultaneous changes and changes with time lag are shown in Figures 4.8 and 4.9, respectively, with the full results from the simulation in Tables A.3 and A.4, respectively, in Appendix A. The proposed SR methods developed for detecting simultaneous change (case 2), and developed for changes with time lag (case 4), perform significantly differently from other methods by giving shorter $\widehat{\text{ced}}$ and lower $\widehat{\text{far}}$. The same pattern is repeated when the shift size is large, though a larger shift size is, of course, generally detected more easily than a smaller one. Due to the high CWS resulting in high variation in data (CWS = 0.6) (i.e. a small shift size is hardly to be detected), the NDR of our method is significantly different from the others for detecting a small shift size. From Figures 4.8(d) and 4.9(d), it seems as though our method gives higher $\widehat{\text{ndr}}$ than the others for detecting a small shift size when CWS is large. However, while our method failed to detect a small shift size, other methods give high false alarm instead due to the effect of CWS. The $\widehat{\text{tar}}$ and $\widehat{\text{ndr}}$ is substantially improved when shift size is large, especially for the SR methods (cases 2 and 4). Additionally, higher CWS, higher $\widehat{\text{far}}$s are produced by the parallel, Frisén and Wessman methods for both types of change point. This is because the SR methods take CWS into account, while the others do not. Also ignoring LCP by using the ordinary Wessman method to detect changes with time lag gives a longer delay with high $\widehat{\text{far}}$.

4.6.3 Scenario 3 (CWS and CBS)

Having fixed CWS ($\phi = 0.6$), the detection performance is investigated with different values of CBS ($\rho = 0.2, 0.4, 0.6$). As in scenario 2 where observations are also dependent over time (CWS = 0.6), the adjusted parallel, Frisén and Wessman methods produce high numbers of false alarms since the effect of CWS is still present in the derived sequence of statistics and this often produces a false alarm before a true change occurs (see Figures A.3 and A.4 in Appendix A). This effect can also be observed from the plot of EWMA charts in Figures 4.10 and 4.11. The pattern of results is the same for both simultaneous change and change with time lag.

The bar charts comparing the detection performance for simultaneous changes and changes with time lag are shown in Figures 4.12 and 4.13, respectively, and full results are given in Tables A.5 and A.6, respectively, in Appendix A. Again, whether or not there is LCP between series, our method (cases 3 and 5) performs significantly differently from other methods by giving shorter $\widehat{\text{ced}}$ and lower $\widehat{\text{far}}$, especially when CBS is large. Unsurprisingly, the pattern is repeated when shift size is large; also $\widehat{\text{tar}}$ and $\widehat{\text{ndr}}$ are improved, with shorter $\widehat{\text{ced}}$. Additionally, ignoring LCP in the Wessman method for detecting changes with time lag (Table A.6) gives slightly worse result, for detecting a small shift size, compared to the Frisén method, which does not take CBS into account.

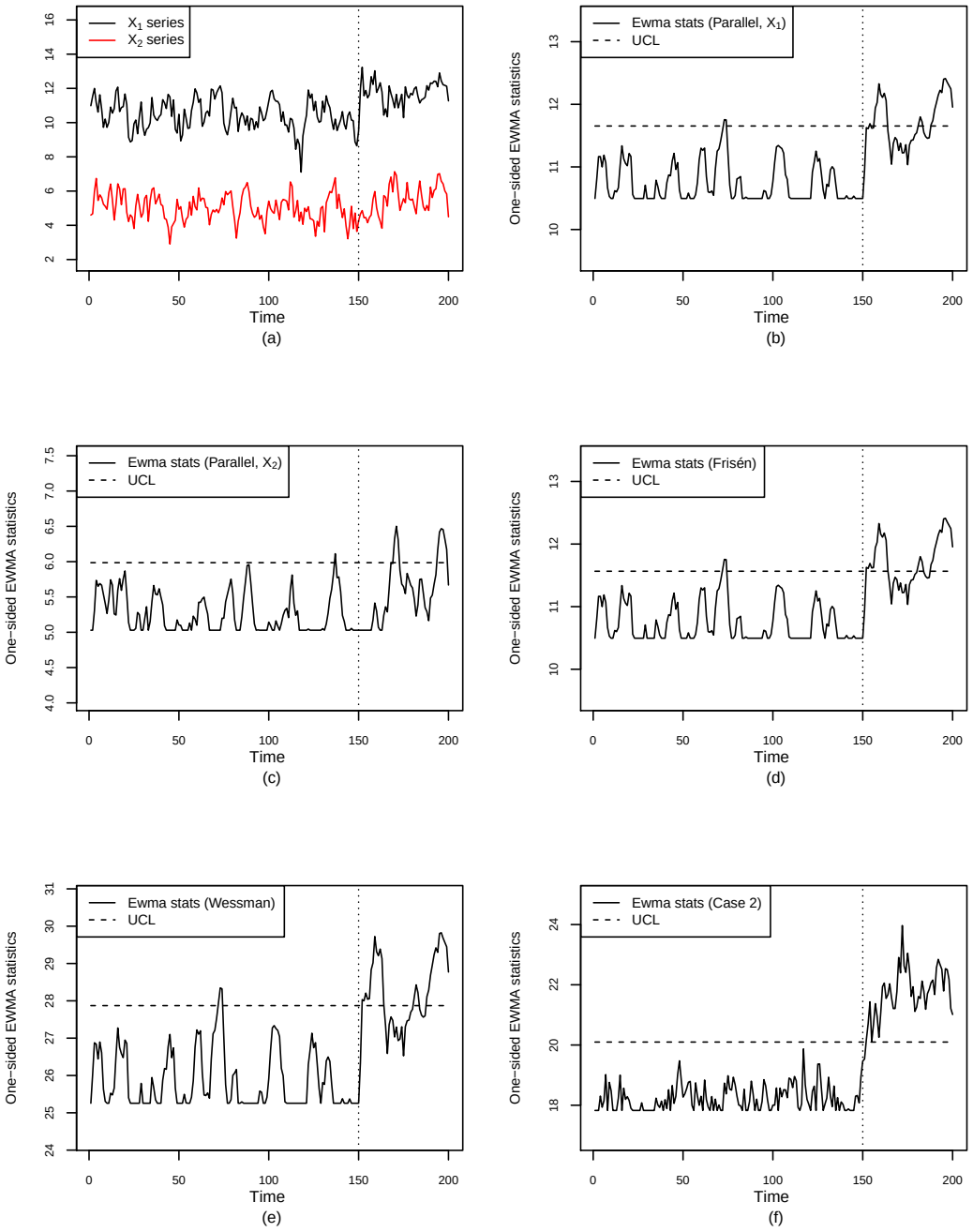

Figure 4.6: An example of scenario 2 (simultaneous changes, shift size 2, CWS = 0.6 and no CBS): (a) plot of the original series, (b) - (c) one-sided EWMA charts of parallel methods, (d), (e) and (f) one-sided EWMA charts of Frisén, Wessman and proposed method (case 2), respectively.

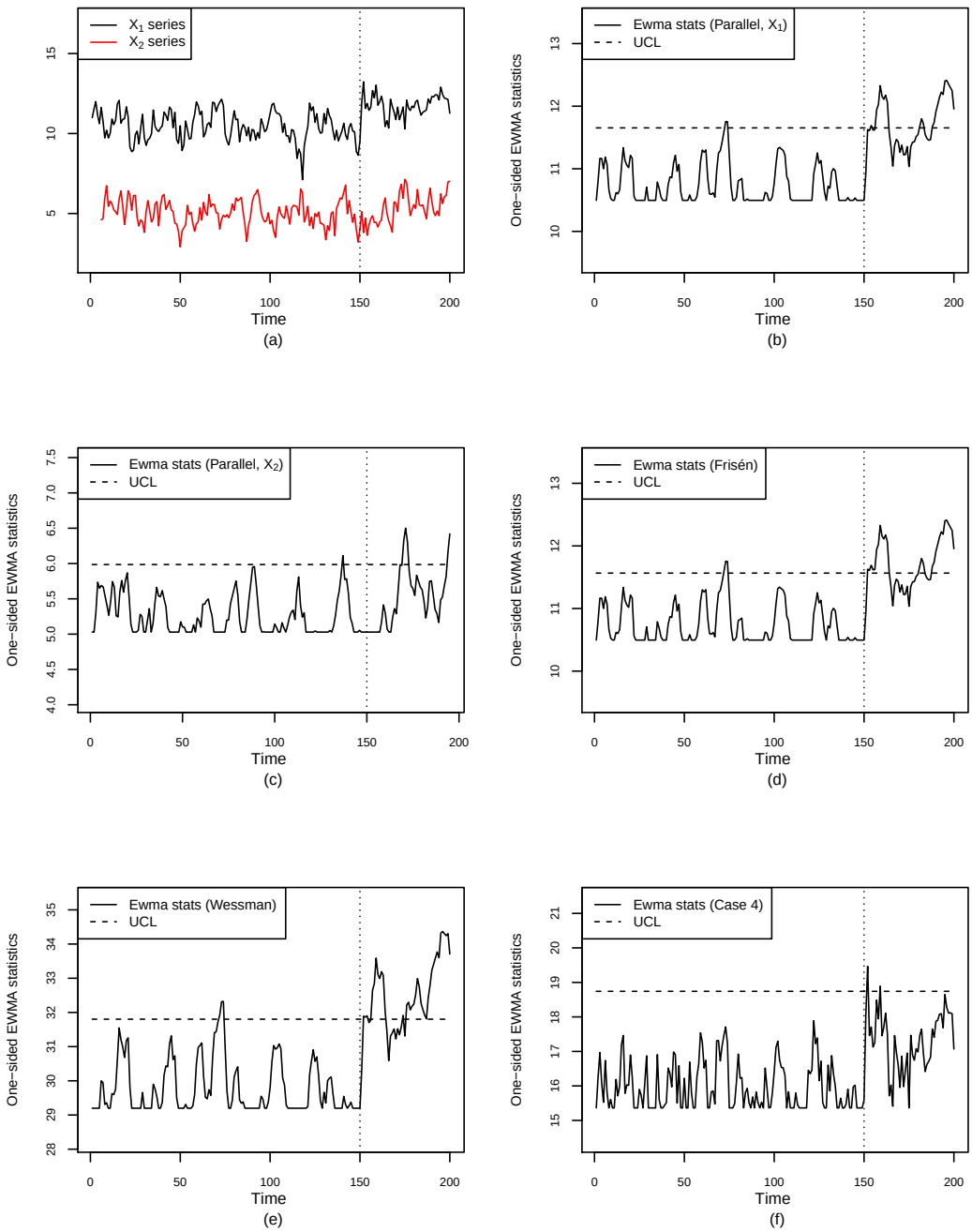

Figure 4.7: An example of scenario 2 (changes with time lag, shift size 2, CWS = 0.6 and no CBS): (a) plot of the original series, (b) - (c) one-sided EWMA charts of parallel methods, (d), (e) and (f) one-sided EWMA charts of Frisén, Wessman and proposed method (case 4), respectively.

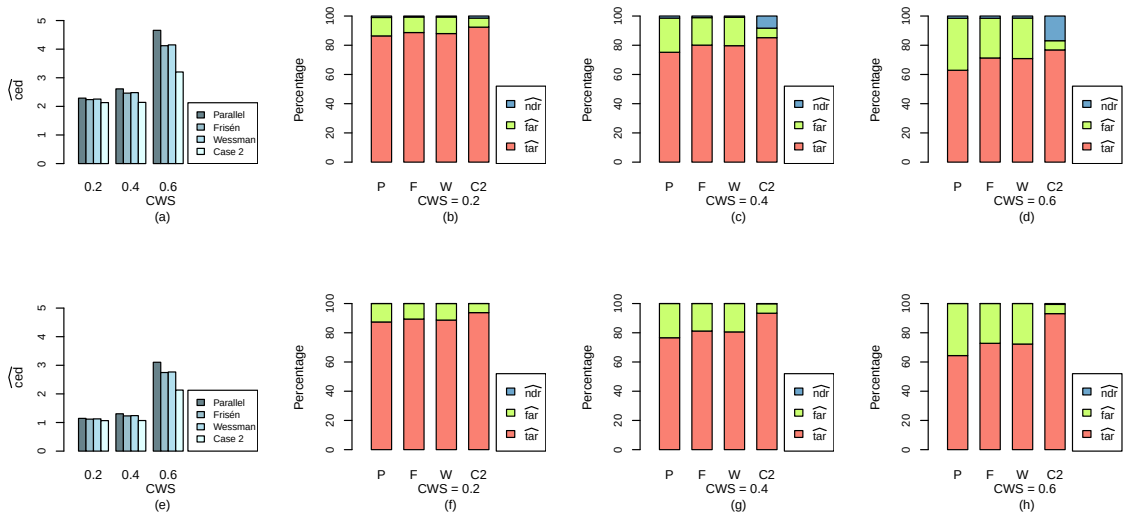

Figure 4.8: A bar chart comparing the performance in detecting simultaneous changes in scenario 2: (a) - (d) shift size = 2 and (e) - (h) shift size = 3.

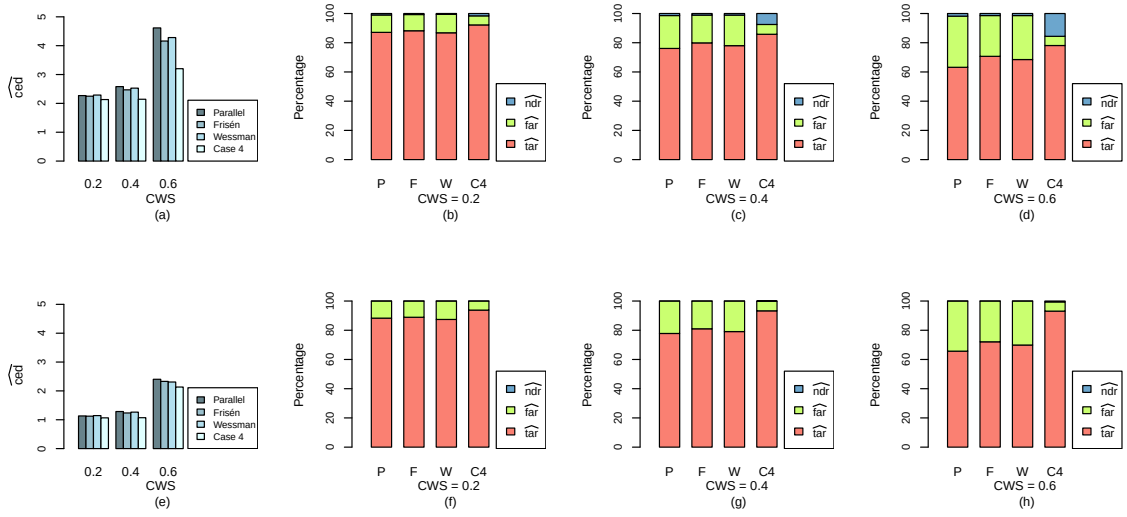

Figure 4.9: A bar chart comparing the performance in detecting changes with time lag in scenario 2: (a) - (d) shift size = 2 and (e) - (h) shift size = 3.

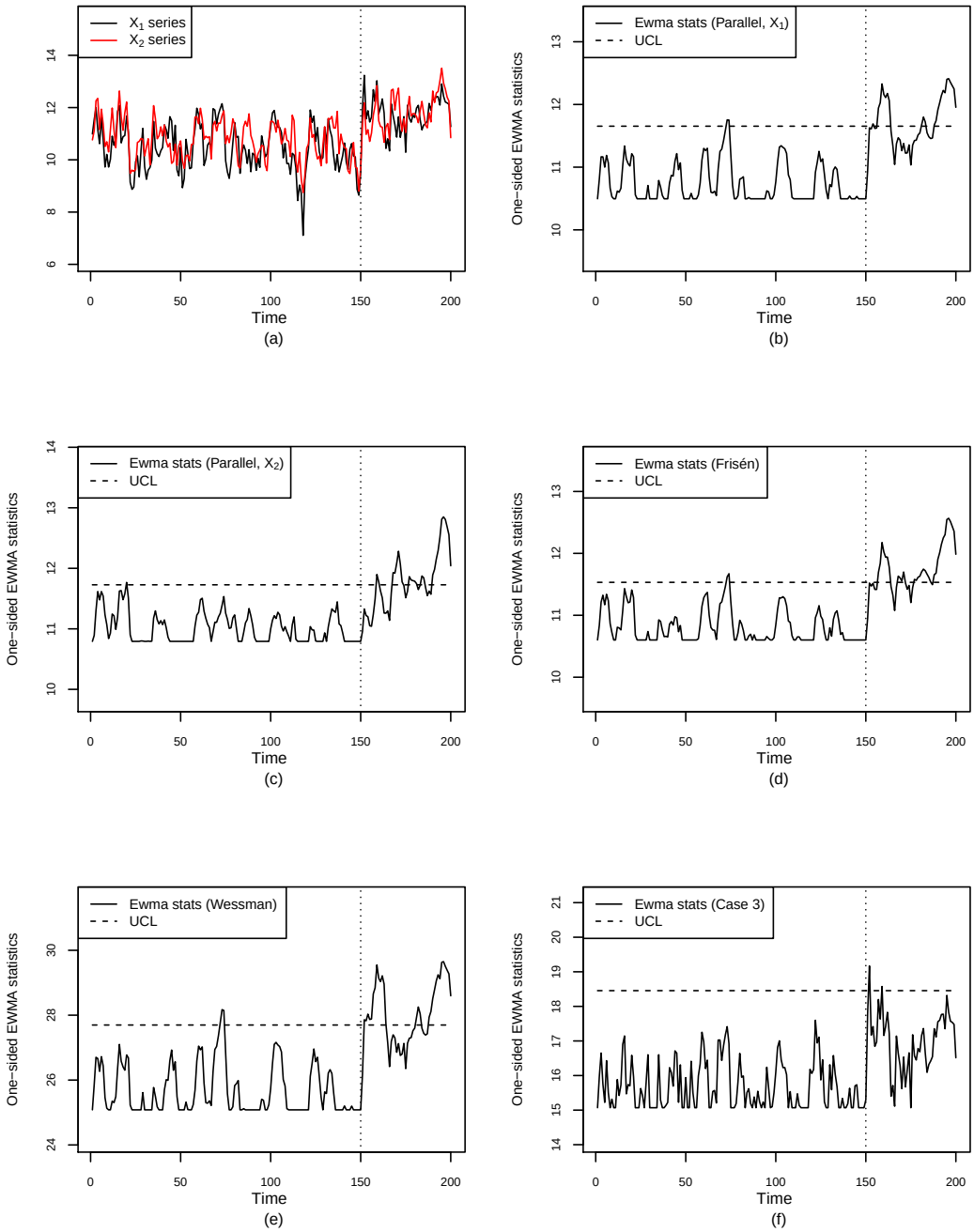

Figure 4.10: An example of scenario 3 (simultaneous changes, shift size 2, CWS = 0.6 and CBS = 0.6): (a) plot of the original series, (b) - (c) one-sided EWMA charts of parallel methods, (d), (e) and (f) one-sided EWMA charts of Frisén, Wessman and proposed method (case 3), respectively.

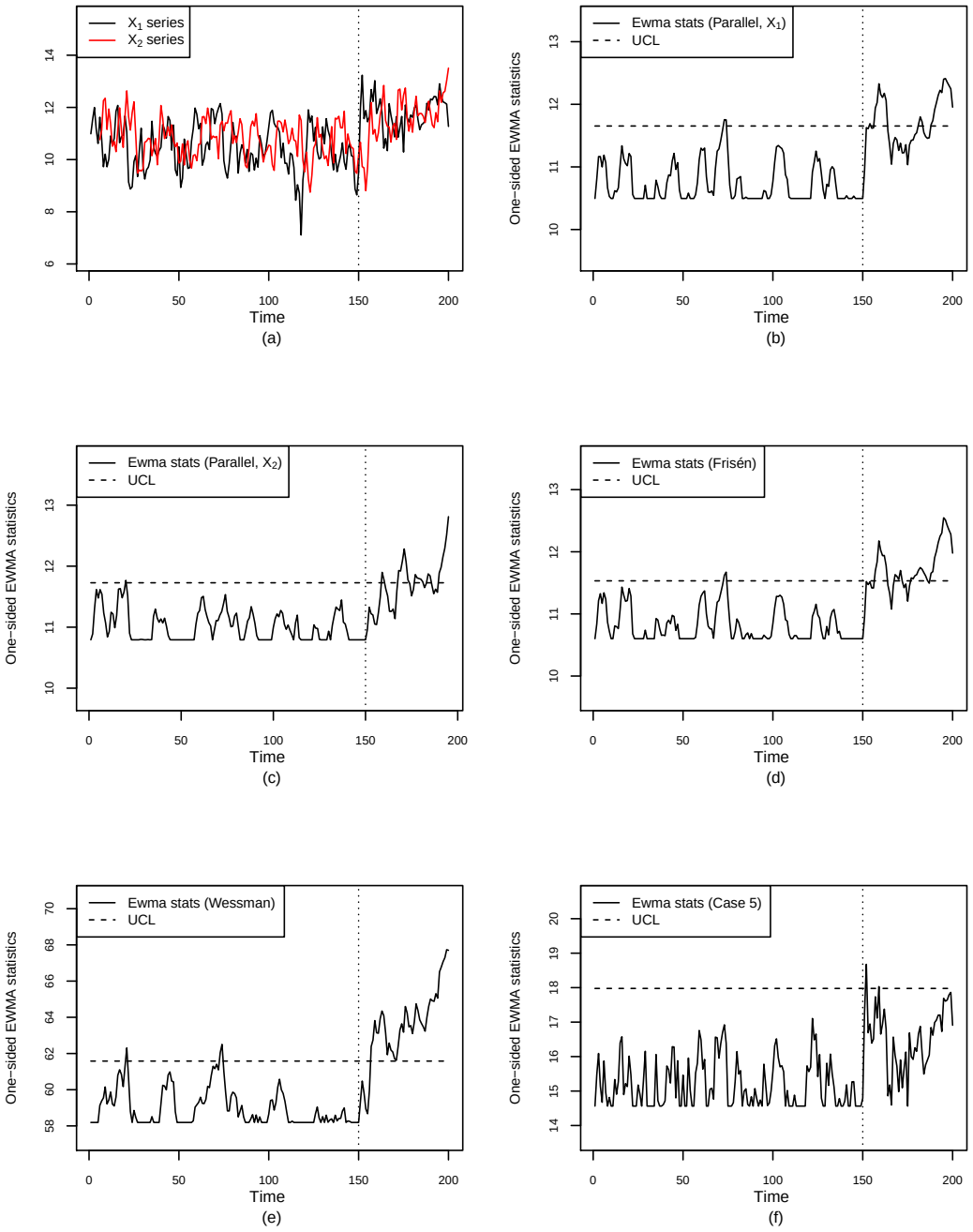

Figure 4.11: An example of scenario 3 (change with time lag, shift size 2, CWS = 0.6 and CBS = 0.6): (a) plot of the original series, (b) - (c) one-sided EWMA charts of parallel methods, (d), (e) and (f) one-sided EWMA charts of Frisén, Wessman and proposed method (case 5), respectively.

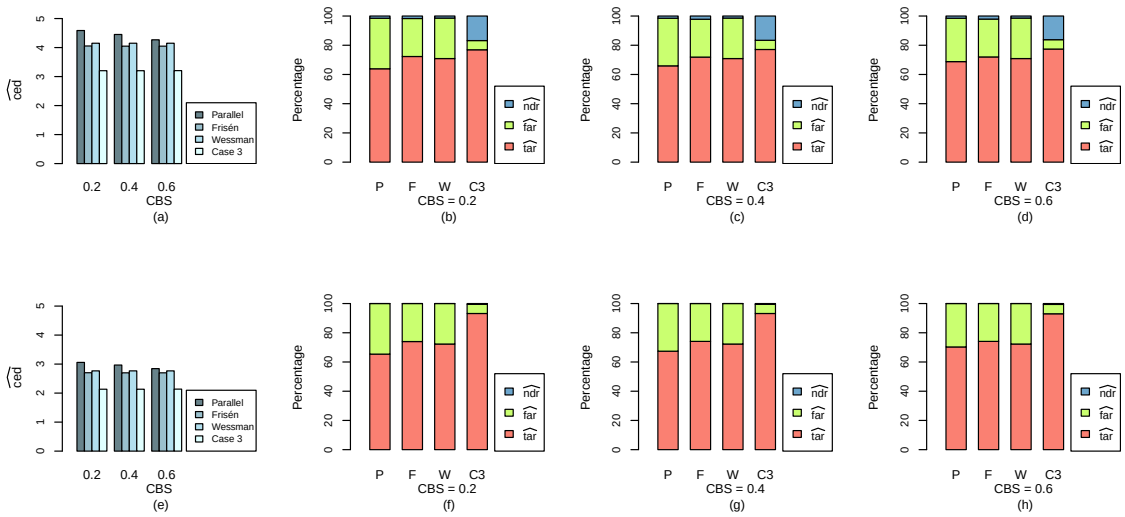

Figure 4.12: A bar chart comparing the performance in detecting simultaneous changes in scenario 3: (a) - (d) shift size = 2 and (e) - (h) shift size = 3.

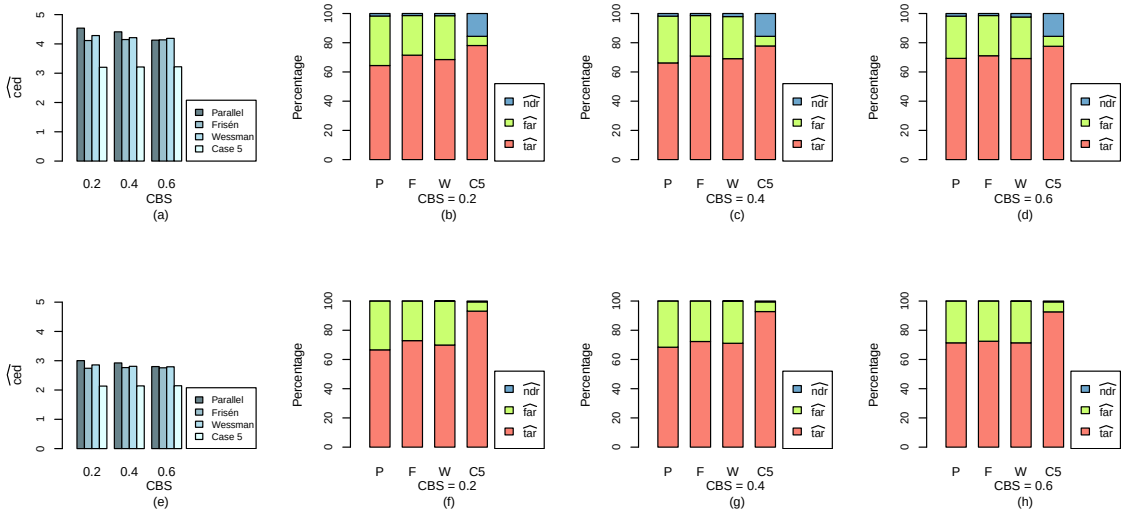

Figure 4.13: A bar chart comparing the performance in detecting changes with time lag in scenario 3: (a) - (d) shift size = 2 and (e) - (h) shift size = 3.

4.7 Conclusions

In this study sufficient reduction methods are proposed for detecting a persistent mean shift in multivariate series under different conditions (CWS, CBS and LCP). Our proposed methods perform better than adjusted parallel, Frisén and Wessman methods when applied under conditions in which these effects occur. If the true LCP can be identified, sufficient reduction methods proposed for simultaneous changes and changes with time lag give similar results (i.e. detecting changes with time lag is the generalization case of detecting simultaneous change if time lag is known).

Whether or not there is LCP between series, if CWS is clearly present, our proposed method gives better results, with shorter CED and lower FAR, than the others. Ignoring CWS (i.e. in the adjusted parallel, Frisén and Wessman methods) delays detection and produces a high FAR since the effect of CWS is still present in the derived likelihood ratio statistics. Therefore, monitoring such series using the one-sided EWMA chart produces a higher FAR because it violates the assumption of the EWMA chart where data need to be independent over time. Also, recognizing CBS gives a chance of detecting even a small shift in the process mean. Thus, Wessman's method performs slightly better than Frisén's in detecting a simultaneous change when CBS is clearly present with no CWS. However, it perform worse than Frisén's method for detecting changes with time lag since LCP is not incorporated.

Since the effect of CWS is still present in the derived sequence of statistics, using a one-sided EWMA chart developed for independent data as a detection tool produces high false alarm rates. Thus, a univariate detection tool for monitoring a change in the process of autocorrelated data might be needed. It can be clearly seen that ignoring or mis-specifying the parameters (CWS, CBS and LCP) in the sufficient reduction leads to the wrong conclusion. The effects of, possibly imprecise, mis-specification and estimation of CWS, CBS and LCP also need to be considered in order to improve the capability for outbreak detection. Apart from this, data and the added signals used in the system evaluation are generated according to the pre-specified shift size aimed to detect. The detection performance for detecting a shift which is smaller or greater than the the pre-specified shift size might be considered. These aspects are investigated further in Chapter 5.

Chapter 5

Extended sufficient reduction methods (normal case)

5.1 Background

In the previous chapter, sufficient reduction (SR) methods used for detecting a mean shift in a process of dependent observations are proposed and compared with existing methods. We found that if data are autocorrelated, the SR methods proposed for detecting a mean shift in a process of independent data (e.g. the Frisén method (Frisén *et al.*, 2011) and Wessman method (Wessman, 1998)) produce high false alarms when we monitor the derived sequence of likelihood ratio statistics with a standard one-sided EWMA chart. Since both methods do not take correlation within series (CWS) into account, the effect of CWS is still present in the derived sequence of likelihood ratio statistics resulting in high FAR when monitored with a standard one-sided EWMA chart for independent data.

Furthermore, we can see that SR methods perform ineffectively if the parameters, such as CWS, CBS and LCP, are not specified appropriately in the sufficient reduction. For example, if there is a correlation between series (CBS), the Wessman method gives better results than Frisén method since it takes CBS into account. However, it performs poorly compared to Frisén for detecting a change with time lag. As mentioned previously, both also perform poorly, by producing high FAR, for detecting change in an autocorrelated process. Due to the variation in health data, noticing and appropriately specifying the parameters used in SR methods might improve the detection performance.

Additionally, the data with the shifted mean used in the system evaluation in the previous chapter are generated according to the pre-specified shift size that we want to detect by the SR methods. In this chapter, having defined a pre-specified shift size, we also aim to investigate the performance of the methods for detecting a shift in the process which might be smaller or greater than the pre-specified size. Obviously, a larger shift is more likely to be detected than a smaller one, however, we still want to investigate how SR methods perform in this case. The performance of such methods are compared with the parallel methods which basically do not require the prior specification of the shift size to be detected.

Noting these limitations, we extend the SR methods for normal data (Chapter 4) by modifying a standard one-sided EWMA chart for monitoring a positive shift in a process of dependent observations and by studying the robustness of the SR methods for the mis-specification of parameters and specification of shift size in the sufficient reduction. We believe that using a one-sided EWMA chart modified for autocorrelated data and specifying parameters appropriately in the sufficient reduction will improve the detection performance and reduce false alarm rates. The modification of one-sided EWMA chart is detailed in Section 5.2, while the robustness study of SR methods are given Section 5.3.

5.2 The modification of a one-sided EWMA chart to accommodate autocorrelated data

As mentioned previously, ignoring CWS by monitoring dependent data with a standard EWMA chart gives a high false alarm rate. This feature has been noticed previously with possible modification proposed. Examples of the violation of independence in control charts can be found in Alwan and Roberts (1988); Harris and Ross (1991); English *et al.* (2000); Mohtashemi *et al.* (2007); Weiß (2011). To modify control charts for autocorrelated process, two possible ways are to calculate control limits adjusted for autocorrelation or to apply model residuals to a standard control chart, called a residual chart (Alwan and Roberts, 1988) (the model-based approach). To handle the autocorrelation in a process, many studies have suggested and used model residuals, which are assumed identically and independently normal distributed, with the standard control charts (e.g. Alwan and Roberts (1988); Montgomery and Mastrangelo (1991); Harris and Ross (1991); Kramer and Schmid (1997); Schmid (1997); MacCarthy and Wasusri (2001); Frisén (2003)).

However, there are some limitations of using model residuals with control charts for independent data. It was found that applying model residuals to standard control charts is less effective when a process is highly positively autocorrelated (Harris and Ross, 1991). Even though residuals, which are the difference between observed values and one-step ahead predictions, are independent, this might not be true for predictions further ahead (Box *et al.*, 1994). Fitting a perfect model in order to get *iid* residuals might also be difficult in practice and the estimation error of model parameters can also affect the detection performance, with shorter ARL_0 and longer ARL_1 (Adams and Tseng, 1998).

One-sided EWMA charts for detecting a positive shift have been developed in several studies (e.g. Robinson and Ho (1978); Crowder (1989); Harris and Ross (1991); Schmid (1997); Shu *et al.* (2007); Hu *et al.* (2011)), however, most of them are developed only for a process of independent observations. Even though EWMA charts for monitoring autocorrelated data have been modified in several studies, most of them are based on model-based approach (MacCarthy and Wasusri, 2001; English *et al.*, 2000; Shiau and Hsu, 2005) and only few are based on adjusting the control limits (Kramer and Schmid, 1997; Schmid and Schöne, 1997; Shiau and Hsu, 2005). These authors have all consid-

ered a standard two-sided chart, so the modification for detecting only a positive shift is not readily available.

In this chapter, we aim to modify the one-sided EWMA chart for detecting a positive shift in an autocorrelated process by adjusting the upper control limits accounted for autocorrelation in an AR(1) process by using the Monte Carlo simulation proposed for modifying a two-sided EWMA chart for positively autocorrelated process in Shiau and Hsu (2005). Shiau and colleagues modified a two-sided EWMA chart for an AR(1) process with positive autocorrelation ($0 < \phi < 1$). The restriction to the case of positive autocorrelation does not cause practical problems, since real data often exhibit this feature.

Let x_t be an observed value of the AR(1) model defined below.

$$x_t = \phi x_{t-1} + \epsilon_t, \tag{5.1}$$

where $t = 1, 2, ..., s$, ϕ is an autoregressive coefficient ($0 < \phi < 1$) and $\epsilon_t \sim N(0, 1)$. A two-sided EWMA statistic, z_t, can be calculated from

$$z_t = \lambda x_t + (1 - \lambda) z_{t-1},\ z_0 = \mu^I, \tag{5.2}$$

where λ is a smoothing parameter and μ^I is a target value or the mean of an in control stage. The correlation within series (CWS) is taken into account by considering the the autoregressive coefficient from the AR(1) process. Regarding the CWS, the mean and the asymptotic variance used to construct the EWMA chart are

$$E(Z_t) = \mu^I \tag{5.3}$$

$$V(Z_t) = \frac{1}{1 - \phi^2} \left(\frac{\lambda}{2 - \lambda} \right) \frac{(1 + \phi(1 - \lambda))}{(1 - \phi(1 - \lambda))} \tag{5.4}$$

(Wieringa, 1999; Shiau and Hsu, 2005). The asymptotic variance is thus derived by taking the autoregressive coefficient of the AR(1) model into account. The derivation of the variance can be found in Wieringa (1999). The lower and upper control limits (LCL and UCL) for a two-sided EWMA chart for autocorrelated process can be calculated from

$$LCL = \mu^I - L\sqrt{V(Z_t)} \tag{5.5}$$

$$UCL = \mu^I + L\sqrt{V(Z_t)} \tag{5.6}$$

where L is the width of the control limits chosen to achieve the desired ARL_0. Note that the control limits of EWMA chart for autocorrelated process are wider than those for in independent process due to the increase in the variance of the autocorrelated process (Wieringa, 1999). The limit L can be obtained by using Monte Carlo simulation. The values of the limit, L, for various values of ϕ and λ for $\mathrm{ARL}_0 = 370$ can be found in Shiau and Hsu (2005).

To modify the one-sided EWMA chart for for an AR(1) process with positive autocorrelation, a one-sided EWMA statistic is calculated instead of a two-sided EWMA statistic

and monitored with the upper control limit only. Like Section 4.2.5, Chapter 4, the one-sided EWMA statistic can be calculated from

$$z_t = \max(\mu^I, \lambda\, x_t + (1-\lambda) z_{t-1}),\ t > 0, \tag{5.7}$$

whereas the upper control limit is calculated from

$$UCL = \mu^I + L\sqrt{\frac{1}{1-\phi^2}\left(\frac{\lambda}{2-\lambda}\right)\frac{(1+\phi(1-\lambda))}{(1-\phi(1-\lambda))}}. \tag{5.8}$$

Let N be a number of iterations, ϕ be a parameter of AR(1), $0 < \phi < 1$, process and RL_0 be an run length. The width L can be obtained from the ARL calculation by using the Monte Carlo simulation defined as follows.

step 1 specify initial values of L and λ and calculate UCL from equation (5.8)

step 2 set $t = 1$

step 3 generate x_t which follow from AR(1) model with $\epsilon_t \sim N(0,1)$ and calculate one-sided EWMA statistics from equation (5.7)

step 4 if $z_t > UCL$, process is out of control and $\mathrm{RL}_0 = t$; otherwise, set $t = t+1$ and go back to step 3

step 5 repeat steps 2 - 4 N times ($N = 100,000$)

step 6 average RL_0 from N iterations to get the in control average run length (ARL_0)

Having defined a desired $\mathrm{ARL}_0 = 370$ (or 741 for parallel method adjusted for multiplicity), ϕ and λ, according to the procedure defined above, the limit L is chosen to archive the desired ARL_0. Having tried the different values of limit L in the Monte Carlo simulation, the limits L suitable for the desired ARL_0 (370 or 741) and $\lambda = 0.3$ are summarized in Table 5.1.

According to the results of scenarios 2 and 3 in Chapter 4, ignoring CWS in the autocorrelated process produce a high false alarm rate. The one-sided EWMA chart modified for the autocorrelated process was then used in Section 5.2.1 to re-evaluate the detection performance of the parallel, Frisén, Wessman and our proposed methods for detecting a mean shift in the autocorrelated process in order to see how the false alarm rate is improved.

5.2.1 Re-evaluation of SR methods

From the previous chapter, it can be seen that the parallel, Frisén and Wessman methods give high false alarm rate (FAR) when CWS is present. In this section, we aim to investigate how the false alarm rates of such methods can be improved if CWS is taken into account properly in the one-sided EWMA chart modified for autocorrelated data. To see whether the modified one-sided EWMA chart can improve FAR, we evaluate the detection performance of such methods by conducting a simulation study and using a

modified one-sided EWMA chart defined in Section 5.2 as a detection tool. In order to see the improvement of the FAR, the results in this section are compared to those in Chapter 4, where a standard one-sided EWMA chart for independent data is used as a detection tool. In the previous chapter, even though our proposed methods produced lower FAR since the CWS has been taken out, they are also included principally in the simulation in order to investigate whether the modification of one-sided EWMA chart is necessarily needed for our methods.

In order to see the improvement and make a sensible comparison, the data, parameters, shift sizes and types of change points defined and used in scenarios 2 and 3 in Chapter 4 are also used in this chapter. Here, however, the sequences of statistics derived from the four methods are monitored with the modified one-sided EWMA chart with $\lambda = 0.3$ and $\text{ARL}_0 = 370$ (or $\text{ARL}_0 = 741$ for the parallel method adjusted for multiplicity). The detection performance is evaluated by estimating CED, TAR, FAR, and NDR from the simulations (see more details in Section 3.4). The detection performance between the methods are investigated by comparing the differences of these statistics estimated from the simulations against the critical values defined in Table 3.2.

The results for each scenario are summarized with the examples of the EWMA charts. The vertical dotted line indicates the change point of the process and the horizontal dashed line is the upper control limit. The comparison of the detection performance is illustrated in bar charts, where P, F, W and Ck stand for Parallel, Frisén, Wessman and our proposed method Case k, respectively. The full results from the simulation study are given in Appendix B.

5.2.2 Scenario 2: CWS and no CBS

Assume that two series are independent (CBS = 0), three levels of CWS (0.2, 0.4 and 0.6) are investigated. Figures 5.1 and 5.2 show an example of using standard and modified charts for detecting simultaneous changes and changes with time lag, respectively. It can be seen that the control limits of the modified chart of parallel, Frisén and Wessman methods are higher than those of the standard charts since they are accounting for the effect of CWS in the parallel data and the derived sequence of statistics. However control limits of our methods are only slightly different since the CWS has already been removed during the sufficient reduction and so the likelihood ratio statistics are (nearly) independent. This makes the modification of one-sided EWMA chart for autocorrelated

Table 5.1: The limits L for the modified one-sided EWMA charts for an AR(1) process with positive autocorrelation ($\text{ARL}_0 = 370, 741$) when $\epsilon_t \sim N(0, 1)$

ARL_0	λ	Autoregressive coefficient (ϕ)								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
370	0.3	2.817	2.779	2.737	2.691	2.639	2.576	2.495	2.387	2.190
741	0.3	3.124	3.108	2.981	2.906	2.883	2.817	2.782	2.685	2.497

data essentially unnecessary for our methods.

With the modified one-sided EWMA chart, the detection performance of the four methods for detecting simultaneous changes and changes with time lag are illustrated in Figures 5.3 and 5.4, and the full results from the simulation are summarized in Tables B.1 and B.2 in Appendix B. Compared to the results of scenario 2 with the standard one-sided EWMA charts (Figures 4.8 and 4.9 in Chapter 4), the performances of parallel, Frisén and Wessman methods with the modified charts are significantly different from those with the standard charts by substantially reducing $\widehat{\text{far}}$ and increasing $\widehat{\text{tar}}$. The $\widehat{\text{far}}$ and $\widehat{\text{tar}}$ of our methods (case 2 and case 4) are slightly improved despite the CWS already having been removed in the sufficient reduction.

Due to the higher control limits adjusted for CWS in the modified charts, four methods give higher $\widehat{\text{ndr}}$ for detecting a small shift size compared to monitoring with the standard charts, especially when CWS is strong. However, the $\widehat{\text{ndr}}$ is improved when the shift size is large. Whether standard or modified chart is used, the CED of our proposed method is significantly different from other methods by giving shorter $\widehat{\text{ced}}$ for detecting a small shift size when CBS = 0.4 and 0.6.

5.2.3 Scenario 3: CWS and CBS

Having fixed CWS ($\phi = 0.6$), three levels of CBS (0.2, 0.4 and 0.6) are investigated. Figures 5.5 and 5.6 show an example of using standard and modified one-sided EWMA charts for detecting simultaneous changes and changes with time lag, respectively. Like scenario 2, the control limits of modified charts of parallel, Frisén and Wessman methods are higher than those of standard charts due to the the adjustment of CWS in the process, while there is little difference in both limits in our methods.

Using a modified one-sided EWMA chart, the detection performance of the four methods for detecting simultaneous changes and changes with time lag are illustrated in Figures 5.7 and 5.8 and the full original results are summarized in Tables B.3 and B.4 in Appendix B. To see the improvement, the results in this section are compared to the results of scenario 3 with the standard charts (Figures 4.12 and 4.13) in Chapter 4.

As with the results of scenario 2 comparing monitoring with standard and modified charts, the same pattern is repeated here. Overall, the detection performances of the parallel, Frisén and Wessman methods with the modified charts are significantly different from the results with the standard charts by giving lower $\widehat{\text{far}}$ and higher $\widehat{\text{tar}}$, while there is only slight improvement in our methods (case 3 and case 5) because the method inherently allows for the CWS.

Irrespective of whatever the standard or modified chart is used, our proposed methods still perform significantly differently from other methods by giving shorter $\widehat{\text{ced}}$ for detecting a small shift size. However, when shift size is large, our methods perform slightly different to Wessman method to detect the simultaneous changes and to Frisén method to detect changes with time lags.

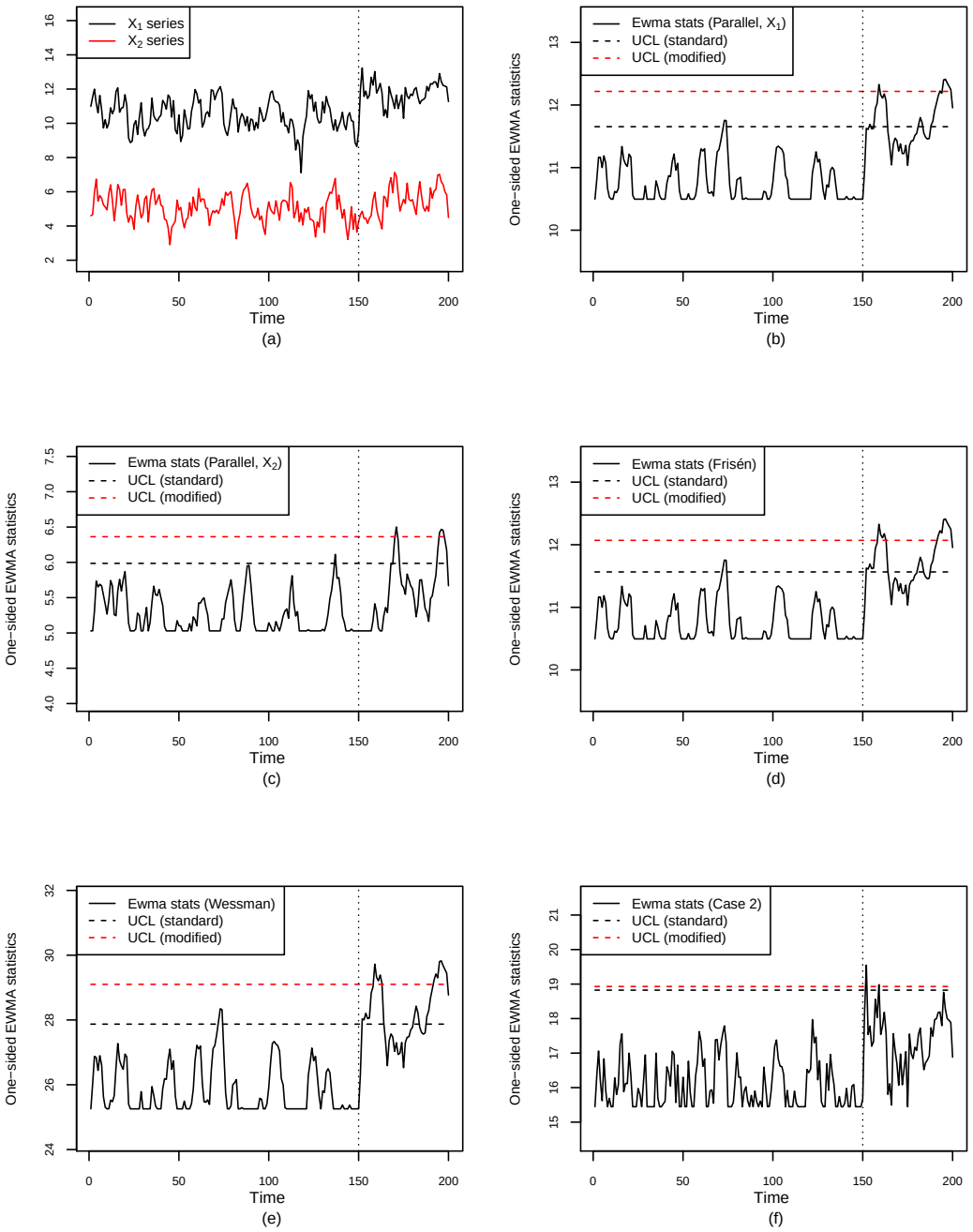

Figure 5.1: An example of scenario 2 (simultaneous changes, shift size 2, CWS = 0.6 and no CBS): (a) plot of the original series, (b) - (c) one-sided EWMA charts of parallel methods, (d), (e) and (f) one-sided EWMA charts of Frisén, Wessman and our methods (case 2), respectively.

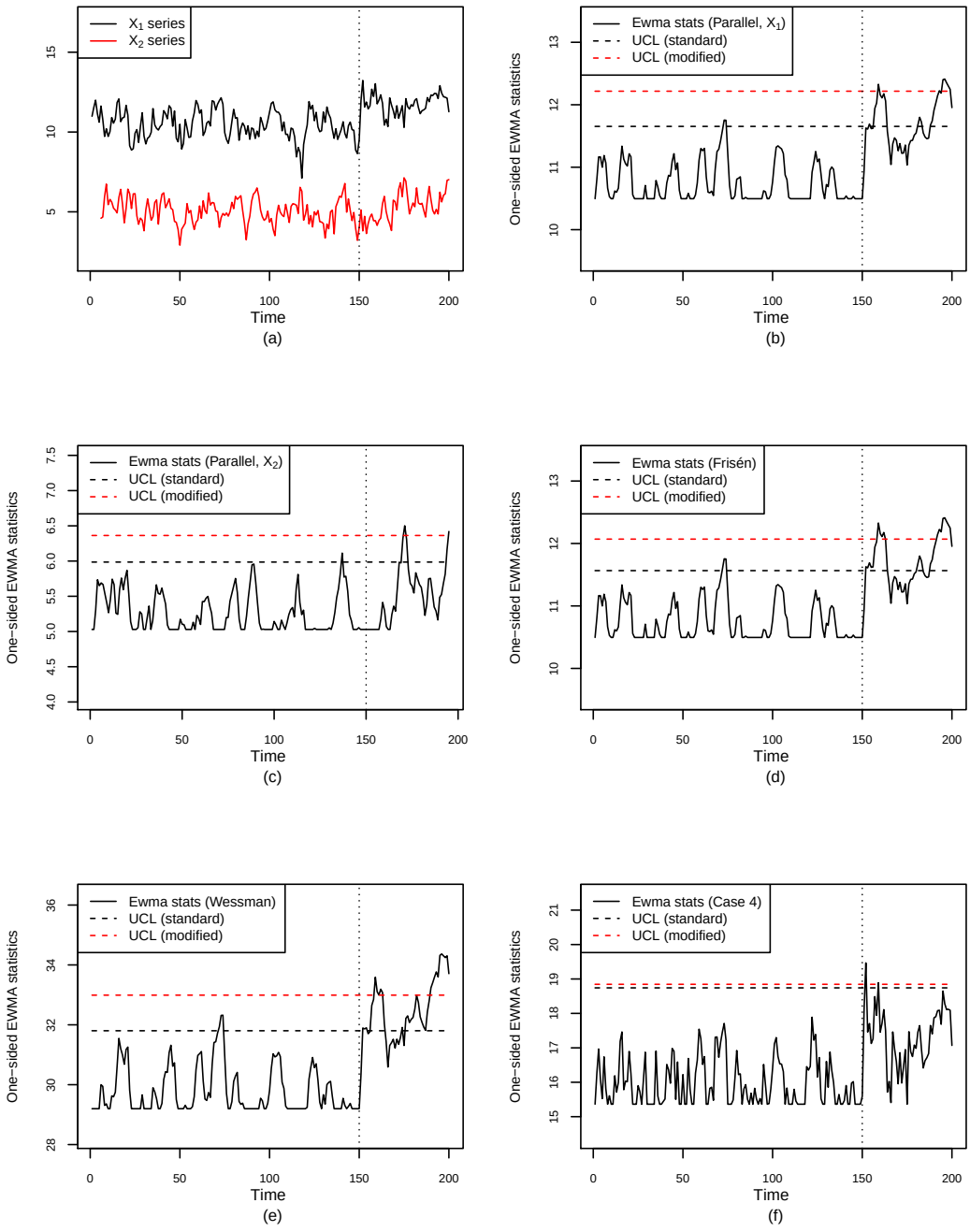

Figure 5.2: An example of scenario 2 (changes with time lag, shift size 2, CWS = 0.6 and no CBS): (a) plot of the original series, (b) - (c) one-sided EWMA charts of parallel methods, (d), (e) and (f) one-sided EWMA charts of Frisén, Wessman and our methods (case 4), respectively.

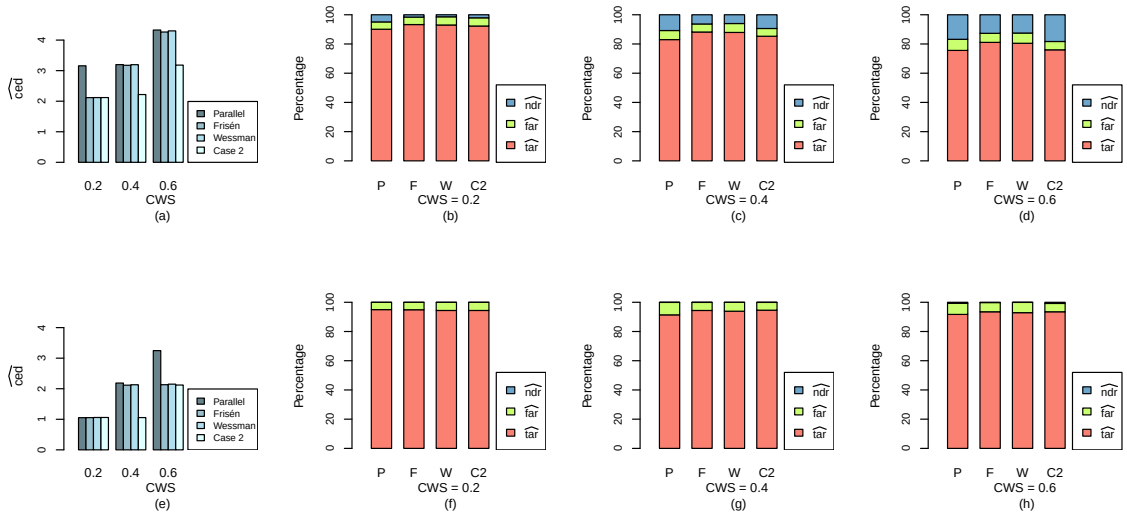

Figure 5.3: A bar chart comparing the performance in detecting simultaneous changes in scenario 2: (a) - (d) shift size = 2 and (e) - (h) shift size = 3.

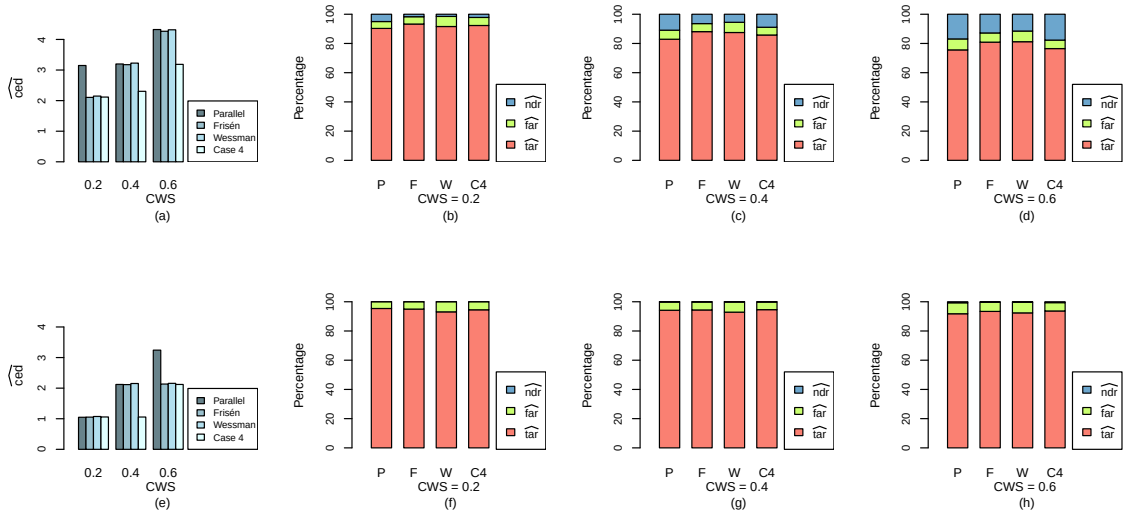

Figure 5.4: A bar chart comparing the performance in detecting changes with time lag in scenario 2: (a) - (d) shift size = 2 and (e) - (h) shift size = 3.

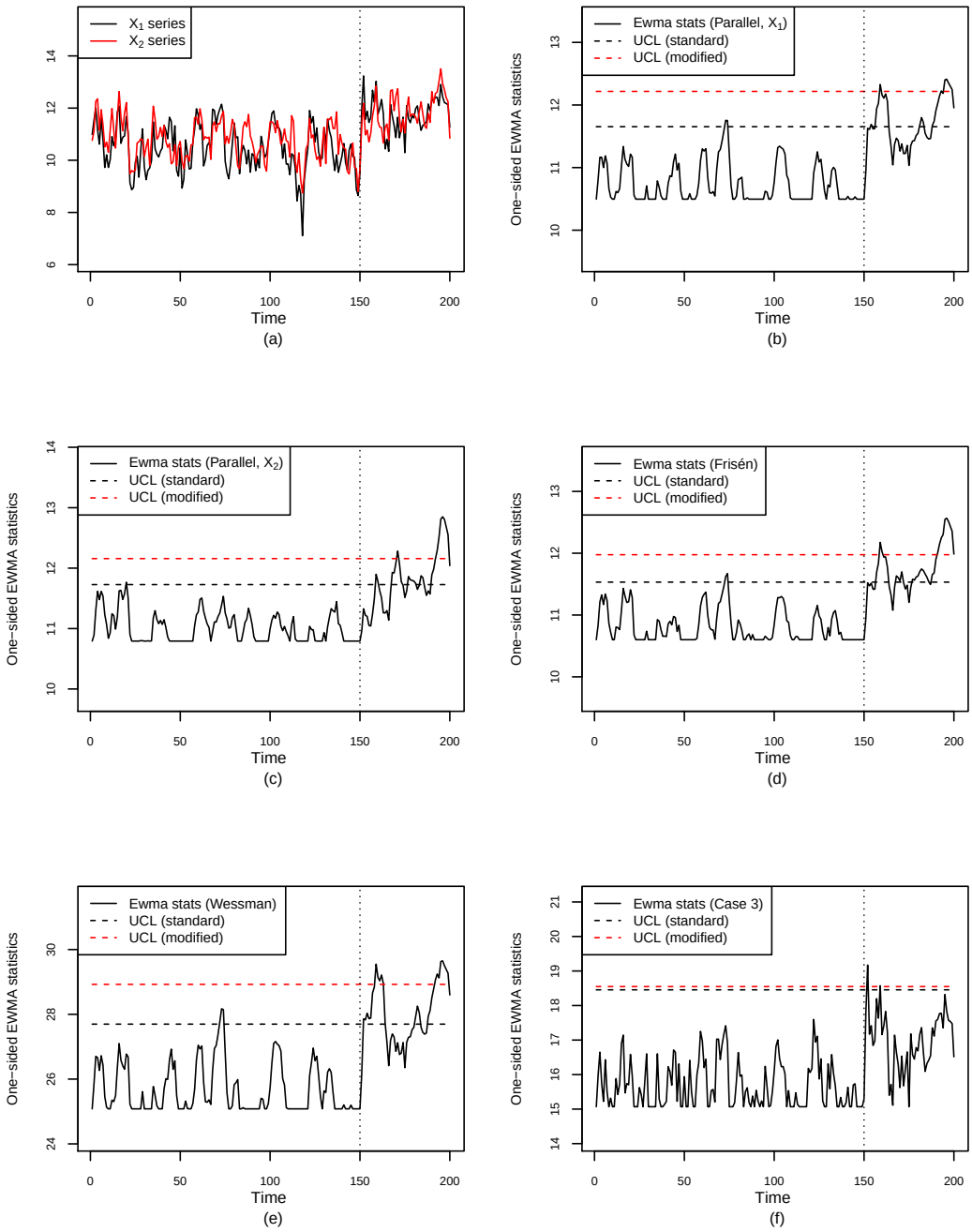

Figure 5.5: An example of scenario 3 (simultaneous changes, shift size 2, CWS = 0.6 and CBS = 0.6): (a) plot of the original series, (b) - (c) one-sided EWMA charts of parallel methods, (d), (e) and (f) one-sided EWMA charts of Frisén, Wessman and our methods (case 3), respectively.

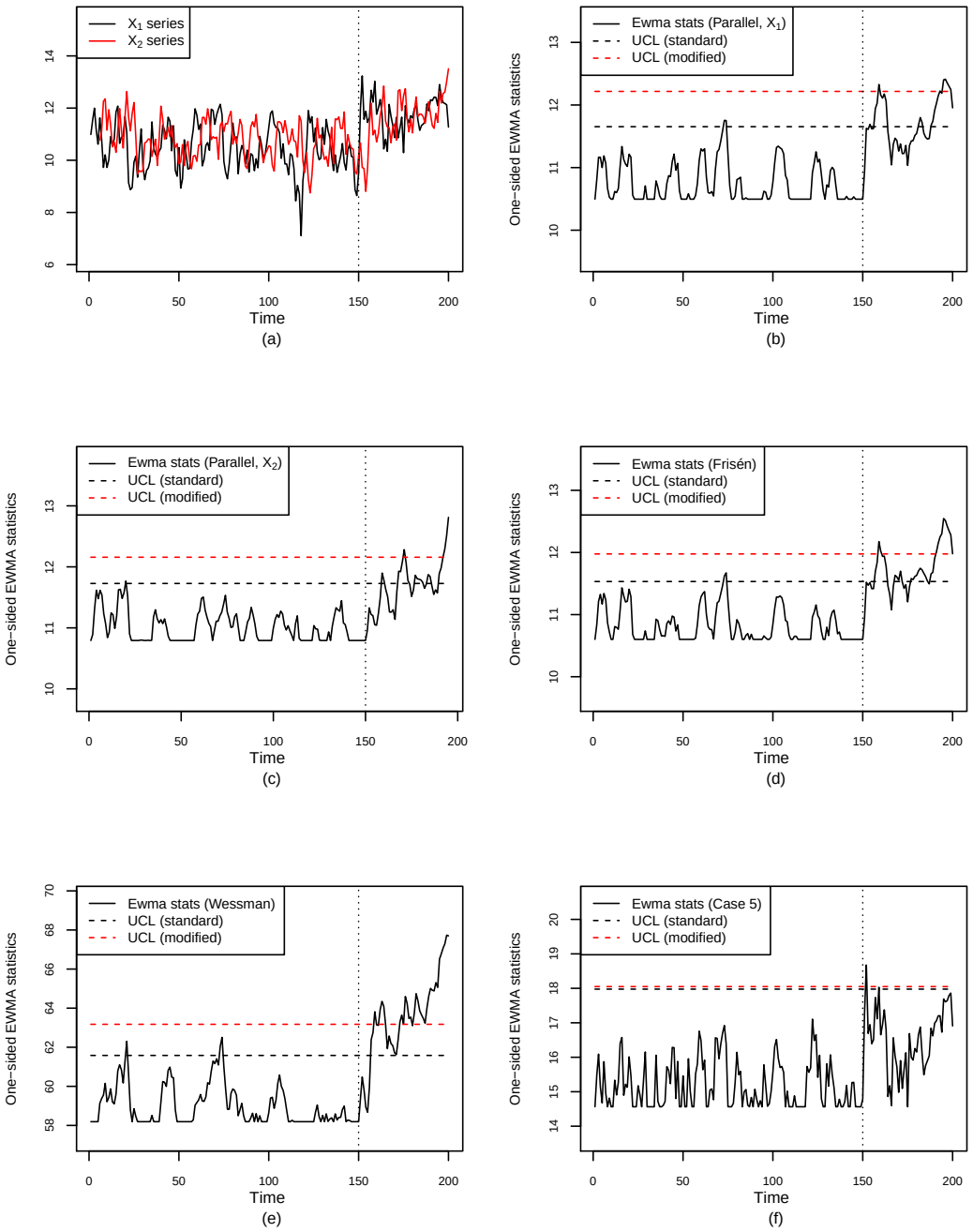

Figure 5.6: An example of scenario 3 (change with time lag, shift size 2, CWS = 0.6 and CBS = 0.6): (a) plot of the original series, (b) - (c) one-sided EWMA charts of parallel methods, (d), (e) and (f) one-sided EWMA charts of Frisén, Wessman and our methods (case 5), respectively.

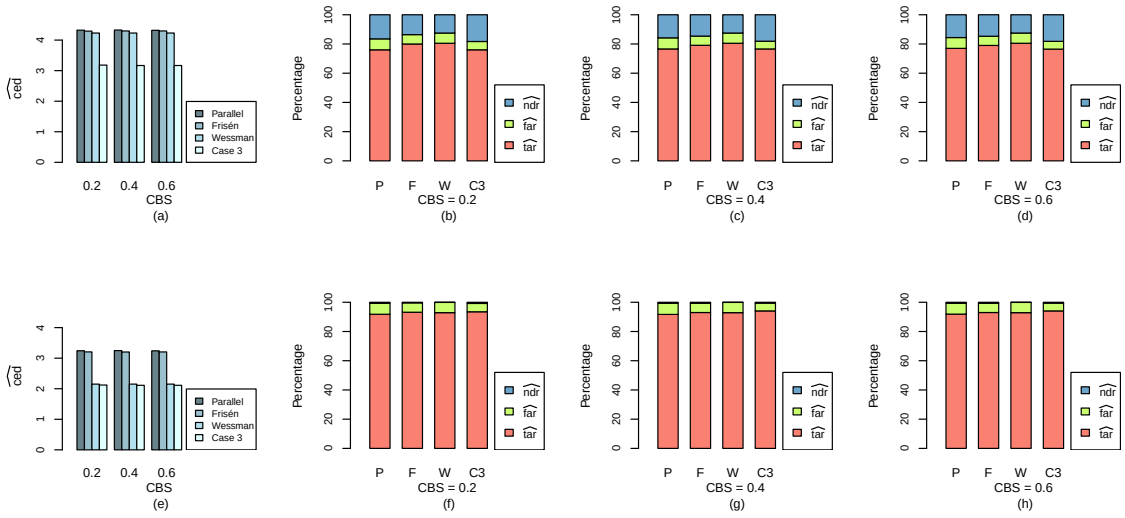

Figure 5.7: A bar chart comparing the performance in detecting simultaneous changes in scenario 3: (a) - (d) shift size = 2 and (e) - (h) shift size = 3.

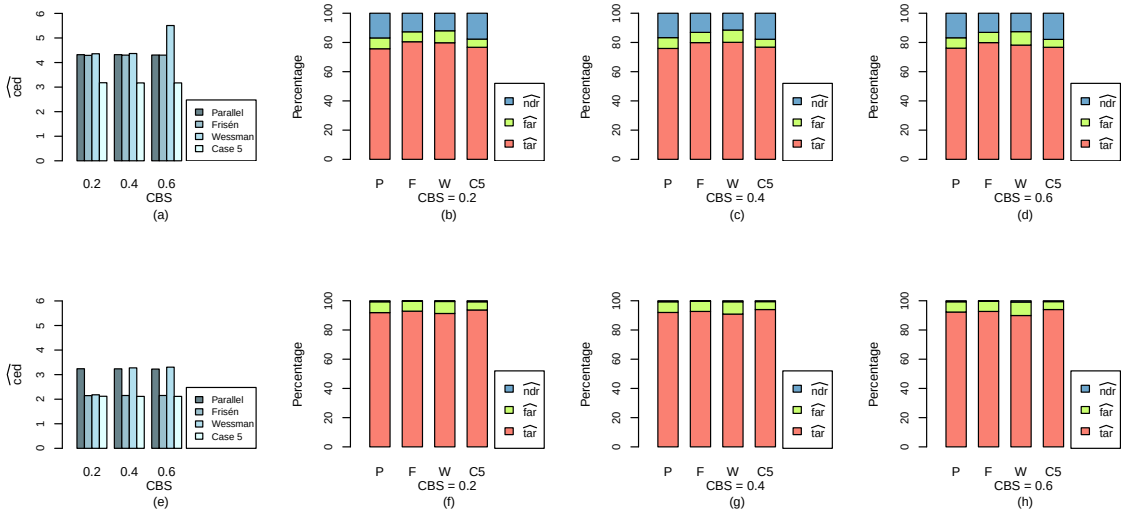

Figure 5.8: A bar chart comparing the performance in detecting changes with time lag in scenario 3: (a) - (d) shift size = 2 and (e) - (h) shift size = 3.

5.3 Robustness study

In this section we aim to investigate the robustness of SR methods against mis-specified parameters used in sufficient reduction, such as CWS, CBS and LCP, and specification of shift size aimed to detect. The effect of mis-specification of parameters and specification of shift size are investigated separately with different possible scenarios. The robustness study is conducted by using simulation. For simplicity, data are generated from the bivariate normal distribution. The performance of detection is evaluated by estimating CED, TAR, FAR and NDR defined in Section 3.4. Since SR methods are proposed under different assumptions, we investigate the robustness of SR methods by choosing a suitable SR method with the corresponding one-sided EWMA chart for each scenario. Section 5.3.1 provides the details the robustness of SR methods against mis-specified parameters and the results from simulations, while the study of the specification of shift size aimed to detect by the SR methods are provided in Section 5.3.2

5.3.1 Mis-specification of parameters

For simplicity, we conduct the simulation study by using the bivariate normal series ($p = 2$) generated according to the parameters correctly specified in each scenario. To investigate the robustness against mis-specified parameters in the SR methods, two different shift sizes are investigated. In this section, the pre-specified shift size, c, aimed to be detected by the SR methods is a shift of size either 2 or 3 (i.e. 2 or 3 times standard deviation of original data). Since there are three main parameters used in SR methods (CWS, CBS and LCP), the effect of mis-specifying each parameter is investigated separately in different scenarios. There are ten possible scenarios where one of these parameters are mis-specified. We investigate the robustness of mis-specifying one parameter by varying the value of that parameter and fixing the values of the other two. The robustness of the SR methods is investigated by comparing the detection performance between using the correctly specified parameters (i.e. those used to generate the data), estimates of the parameters estimated from the in control stage and several mis-specified values of the parameters in the sufficient reduction (see below). Additionally, both standard and modified EWMA charts are considered as detection tools.

Let N be the number of simulations and c be the pre-specified shift size to be detected by the SR methods, the procedure of the robustness study of mis-specification of parameter for each scenario can be described as follows.

step 1 a bivariate series is generated according to the correctly specified parameters defined in each scenario. We consider scenarios covering possibilities of each of CWS and CBS being specified as 0 and 0.6.

step 2 a change point of a process, τ, is randomly selected ($\tau > 150$; i.e. to make sure that the in control baseline is long enough for parameter estimation). A shift of size c is added to the data at and after time τ. In the case of lag between series (l), a shift is added to the first series at time τ and to the second series at time $\tau + l$.

step 3 estimates of the correlation parameters (CBS and CWS) are estimated from the in control baseline.

step 4 the data are used to derive a sequence of likelihood ratio statistics by using the correctly specified parameter in the sufficient reduction. The sequence is then monitored with the corresponding one-sided EWMA chart defined in each scenario. The delay in detection is recorded (i.e. $t_A - \tau$, where t_A is a time of an alarm). If $t_A - \tau \geq 7$, NA (non-applicable) is reported (i.e. system fails to detect a shift within 7 time points).

step 5 repeat step 4 but the estimated parameter is used in the sufficient reduction instead

step 6 repeat step 4 but the mis-specified value of parameter (i.e. 0, 0.2, 0.4 or 0.8) is used instead. The step is repeated until all mis-specified values are used. The delay in detection for each value is recorded separately.

step 7 repeat steps 1-6 N times ($N = 10,000$). The CED, TAR, FAR and NDR are estimated from the results in each step (NB there are four sets of results from step 6).

Thus, in each scenario, six results of detection performance from using six values of the parameter (i.e. correctly specified parameter, the estimate of the parameter and four mis-specified values of the parameter) in the sufficient reduction are compared. Each result is from 10,000 simulations. In the case of mis-specification of LCP (Section 5.3.1.3), five mis-specified values of LCP are investigated, therefore, the detection performances are compared between seven results, each from 10,000 simulations. The difference in performances of the SR methods using different sets of parameters are investigated by comparing the differences of the statistics estimated from the simulations against the critical values defined in Table 3.2. The comparison of the detection performance for each scenario is illustrated in a bar chart, where 'Est.' represents the results from using the estimated parameter in the sufficient reduction. More details of ten possible scenarios with correctly specified parameter, the values of mis-specified parameters, a suitable SR method and the corresponding one-sided EWMA chart are given in Sections 5.3.1.1 - 5.3.1.3. The estimation method for each parameter is given in Section 5.3.1.4.

5.3.1.1 Mis-specification of CWS

Having assumed that data are dependent over time, in this section we aim to investigate how SR methods are robust when CWS is mis-specified. Four possible scenarios with the values of correctly specified parameters and mis-specified parameters along with a suitable SR method and the corresponding one-sided EWMA chart for each scenario are listed in Table 5.2. Data used in each scenario are generated according to correctly specified parameters defined in the Table 5.2. Four values of mis-specified CWS (0, 0.2, 0.4 and 0.8) are investigated. Since data are autocorrelated over time (CWS = 0.6), our methods proposed for detecting a shift in a process of dependent observations (cases 2 - 5, Chapter 4) are used in these scenarios. Also, since the CWS has been removed by our proposed methods, there is little difference whether a standard or modified EWMA chart is used, thus, we use the standard chart for all scenarios in this section.

Table 5.2: Four scenarios of mis-specification of CWS

Scenario	Correctly specified parameters			Mis-specified CWS	Method	Chart
	CWS	CBS	LCP			
R1	0.6	0	0	0, 0.2, 0.4, 0.8	Case 2	standard
R2	0.6	0	5	0, 0.2, 0.4, 0.8	Case 4	standard
R3	0.6	0.6	0	0, 0.2, 0.4, 0.8	Case 3	standard
R4	0.6	0.6	5	0, 0.2, 0.4, 0.8	Case 5	standard

5.3.1.2 Mis-specification of CBS

In this case, we assume that two series are dependent with CBS = 0.6. Four scenarios with a suitable SR method and the corresponding one-sided EWMA chart for each scenario are detailed in Table 5.3. Data used in each scenario are generated according to the correctly specified parameters defined in the Table 5.3. Four values of mis-specified CBS (0, 0.2, 0.4 and 0.8) are investigated. Wessman's method and our method (case 1, Chapter 4) are used to detect the mean shift in the process of independent data (CWS = 0), while our methods (cases 3 and 5, Chapter 4) are used for detecting the mean shift in the autocorrelated process (CWS = 0.6).

Table 5.3: Four scenarios of mis-specification of CBS

Scenario	Correctly specified parameters			Mis-specified CBS	Method	Chart
	CBS	CWS	LCP			
R5	0.6	0	0	0, 0.2, 0.4, 0.8	Wessman	standard
R6	0.6	0	5	0, 0.2, 0.4, 0.8	Case 1	standard
R7	0.6	0.6	0	0, 0.2, 0.4, 0.8	Case 3	standard
R8	0.6	0.6	5	0, 0.2, 0.4, 0.8	Case 5	standard

5.3.1.3 Mis-specification of LCP

According to the SR assumptions defined in Chapter 4, the relation of shift sizes between series is determined by CBS. If two series are independent, we investigate the shift size from one series only. Therefore, a lag between series is not necessarily considered if two series are independent. In this section we aim to investigate the effect of mis-specifying lag between series by assuming that the two series are correlated with LCP = 5 and CBS = 0.6. Two possible scenarios with a suitable SR method and the corresponding one-sided EWMA chart for each scenario are detailed in Table 5.4. Data used in each scenario are generated according to the correctly specified parameters defined in Table 5.4. Five values of mis-specified lag (0, 3, 4, 6 and 7) are investigated. SR methods proposed for detecting changes with time lag in the process of independent data (case

1) and the autocorrelated process (case 5) are used.

Table 5.4: Two scenarios of mis-specification of LCP

Scenario	Correctly specified parameters			Mis-specified LCP	Method	Chart
	LCP	CBS	CWS			
R9	5	0.6	0	0, 3, 4, 6, 7	Case 1	standard
R10	5	0.6	0.6	0, 3, 4, 6, 7	Case 5	standard

5.3.1.4 Parameter estimation

The second general form of mis-specification of parameters we shall investigate is that arising when the values of parameter are not assumed known but are estimated from the in control stage of the process. There are many ways of estimating the relevant parameters (ϕ, ρ and l (lag between series)), but in the following section we outline the standard methods that we have selected.

5.3.1.4.1 Correlation within series

In this study we assume that data are autocorrelated in a form of an autoregressive model of order 1 where the current observation is conditional on the previous observation. There are three way to estimate the autoregressive coefficient; maximum likelihood, least squares and solving the Yule-Walker equations. The maximum likelihood is preferable to the others since it provided a unified, and entirely practicable, procedure for estimating parameters in the autoregressive model which can be obtained by a numerical maximization (Diggle, 1990; Triantafyllopoulos, 2009). The maximum likelihood estimator can be obtained by maximizing the log likelihood function of the AR(1) process with respect to the parameter in the model (Diggle, 1990; Brockwell and Davis, 2002; Triantafyllopoulos, 2009). The maximum likelihood approach is widely used in practice and also available in packages. For example, the maximum likelihood estimators can be obtained by using function `arima` in `R` (Cowpertwait and Metcalfe, 2009).

5.3.1.4.2 Correlation between series

Let ρ_l be the correlation between X_1 and X_2 at lag l $(l = 0, 1, ...)$ which is defined as

$$\rho_l = Cor(X_{1,t}, X_{2,t+l}) = \frac{Cov(X_{1,t}, X_{2,t+l})}{\sqrt{Var(X_{1,t})Var(X_{2,t+l})}} = \frac{E[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})]}{\sqrt{Var(X_{1,t})Var(X_{2,t+l})}}$$

For detecting a simultaneous change ($l = 0$), the estimate of CBS, r_0, can be calculated from

$$r_0 = \frac{\sum_{t=1}^{n}(X_{1,t} - \bar{X}_1)(X_{2,t} - \bar{X}_2)}{\sqrt{\sum_{t=1}^{n}(X_{1,t} - \bar{X}_1)^2 \sum_{t=1}^{n}(X_{2,t} - \bar{X}_2)^2}}, \tag{5.9}$$

where $\bar{X}_i = \sum_{i=1}^{n} X_i/n,\ i = 1, 2.$

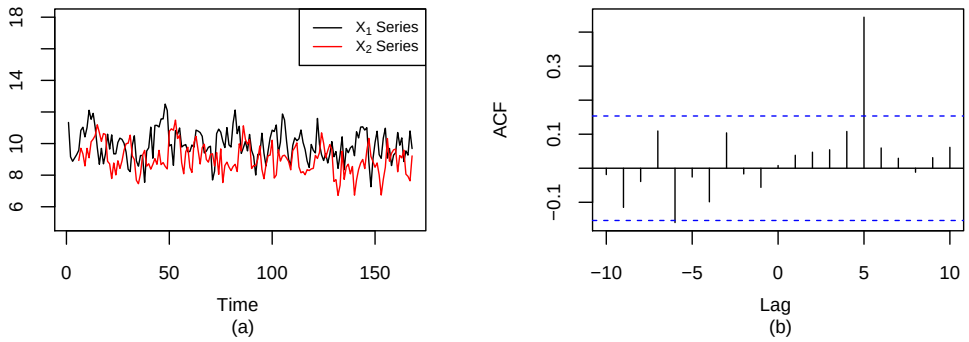

Figure 5.9: The illustration of lag estimation between two series.

In case of changes with time lags, the cross correlation at lag l $(l > 0)$, ρ_l, is considered instead. The estimate of the cross correlation, r_l, is

$$r_l = \frac{\sum_{t=1}^{n-l}(X_{1,t} - \bar{X}_1)(X_{2,t+l} - \bar{X}_2)}{\sqrt{\sum_{t=1}^{n-l}(X_{1,t} - \bar{X}_1)^2 \sum_{t=1}^{n-l}(X_{2,t+l} - \bar{X}_2)^2}} \tag{5.10}$$

where $\bar{X}_i = \sum_{i=1}^{n-l} x_i/n - l,\ i = 1, 2.$

5.3.1.4.3 Lag between change points

LCP is estimated from cross correlation between X_1 and X_2 (equation (5.10)). The cross correlation also can be obtained by using the `ccf` function in `R` which estimates the correlation between X_{t+l} and Y_t (Cowpertwait and Metcalfe, 2009). In our case, we estimate the cross correlation between $X_{1,t}$ and $X_{2,t+l}$ by using $ccf(X_{2,t+l}, X_{1,t})$. The lag that gives the highest cross correlation is regarded as the lag between series. However if data are autocorrelated over time, calculating cross correlation from the original data might be misleading. Alternatively, it is suggested to assess cross correlation from model residuals or using pre-whitening approach. In this study we chose to estimate the cross correlation from the model residuals. The illustration of how LCP is estimated is shown in Figure 5.9. In figure 5.9 (a), X_1 and X_2 series are generated with CWS = 0.4 and CBS = 0.4 where X_2 follows X_1 at lag 5. The plot of the autocorrelation function of the residuals, from fitting the AR(1) model, in 5.9 (b) indicates that the estimated LCP is 5 since it gives the highest correlation at lag 5.

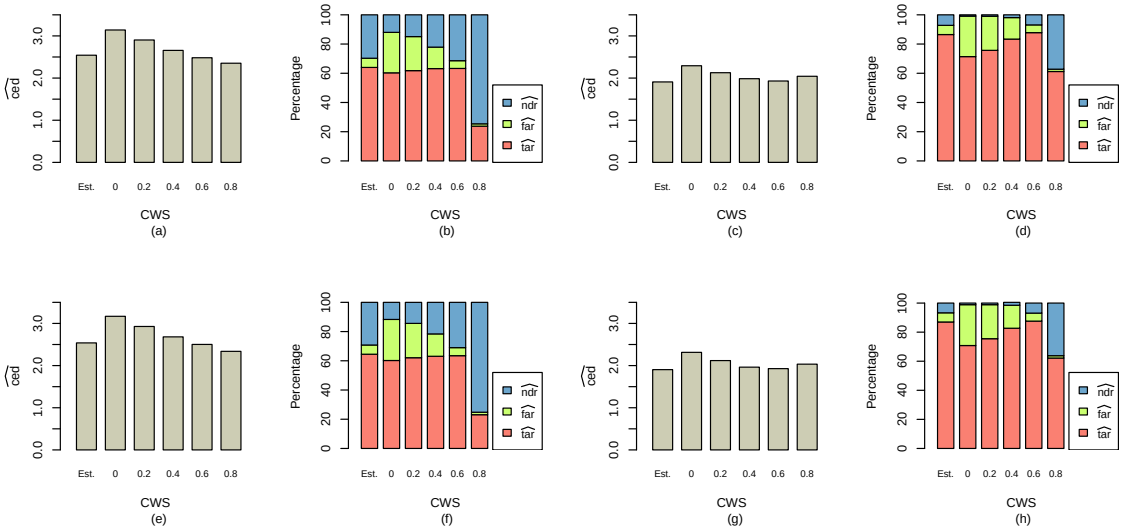

Figure 5.10: Scenarios R1 and R2 (Mis-specification of CWS): (a-b) and (c-d) scenario R1 for shift sizes 2 and 3, and (e-f) and (g-h) scenario R2 for shift sizes 2 and 3.

5.3.1.5 Results

5.3.1.5.1 Mis-specification of CWS

The detection performances of scenarios R1 and R2 are illustrated in Figure 5.10, while those for scenarios R3 and R4 are in Figure 5.11. In both Figures the correctly specified CWS is 0.6 and the estimated CWS is denoted by Est. The results from the simulation study are provided in Tables B.5 - B.8, in Appendix B. In the case of two independent series (scenarios R1 and R2), there is no difference in the results from using correctly specified CWS and estimated CWS, while the results from using mis-specified values of CWS are significantly different by showing longer $\widehat{\text{ced}}$, lower $\widehat{\text{tar}}$ and higher $\widehat{\text{far}}$. The larger the difference between mis-specified CWS and correctly specified CWS, the longer $\widehat{\text{ced}}$, lower $\widehat{\text{tar}}$ and higher $\widehat{\text{far}}$ produced. Even though using mis-specified CWS = 0.8 produces less $\widehat{\text{far}}$, it gives the lowest $\widehat{\text{tar}}$ with the highest $\widehat{\text{ndr}}$. However, the $\widehat{\text{tar}}$ and $\widehat{\text{ndr}}$ is improved when the shift size is large. In the case of two dependent series (scenarios R3 and R4), a similar pattern is repeated, since our methods (cases 3 and 5) take CBS into account, resulting in only slight differences in detection performance. Even though our proposed methods (cases 2, 3, 4 and 5) account for CWS, mis-specifying CWS in the sufficient reduction delays the detection and gives high $\widehat{\text{far}}$.

5.3.1.5.2 Mis-specification of CBS

The detection performances of four scenarios (R5 - R8) are illustrated in Figures 5.12 and 5.13. In both Figures the correctly specified CBS is 0.6 and the estimated CBS is denoted by Est. The results from the simulation study are provided in Tables B.9 -

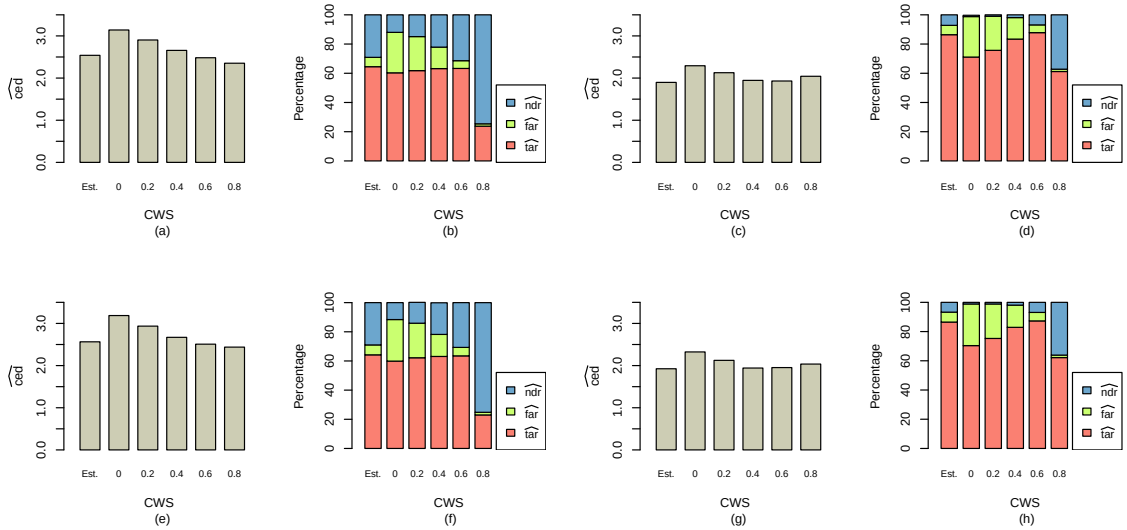

Figure 5.11: Scenarios R3 and R4 (Mis-specification of CWS): (a-b) and (c-d) scenario R3 for shift sizes 2 and 3, and (e-f) and (g-h) scenario R4 for shift sizes 2 and 3.

B.12 in Appendix B. As expected, there is no difference in all measures between using correctly specified CBS and estimated CBS in the sufficient reduction for both types of change point.

In the independent process, mis-specifying CBS in detecting a simultaneous changes (scenario R5, Figure 5.12 (a-d)) gives worse results with longer $\widehat{\text{ced}}$ and slightly lower $\widehat{\text{tar}}$, higher $\widehat{\text{far}}$ and $\widehat{\text{ndr}}$ than those for detecting changes with time lag (scenario R6, Figure 5.12 (e-f)). The larger the difference between mis-specified CBS and correctly specified CBS, the longer the $\widehat{\text{ced}}$ produced. The CEDs of using mis-specified CBS are significantly different from using correctly specified CBS for detecting simultaneous change (scenario R5), while there is no difference in CED for detecting changes with time lag (scenario R6). Also, there is no difference in FAR and NDR for all shift sizes. The same pattern is repeated for detecting mean shifts in the autocorrelated process in scenarios R7 and R8 (Figure 5.13), although the $\widehat{\text{ndr}}$ is quite high due to the CWS. However, $\widehat{\text{ndr}}$ is improved when shift size is large.

It can be seen that the effect of mis-specifying CBS is more obvious in detecting simultaneous changes rather than changes with time lags. This is because due to the LCP, at the decision time t, $t = s - l + 1, ..., s$, the likelihood ratio statistics is derived based on the information in one series only (i.e. the information of another series is not available yet) and so CBS is not incorporated at time t, $t = s - l + 1, ..., s$. This makes the effect of mis-specifying CBS less for detecting changes with time lags.

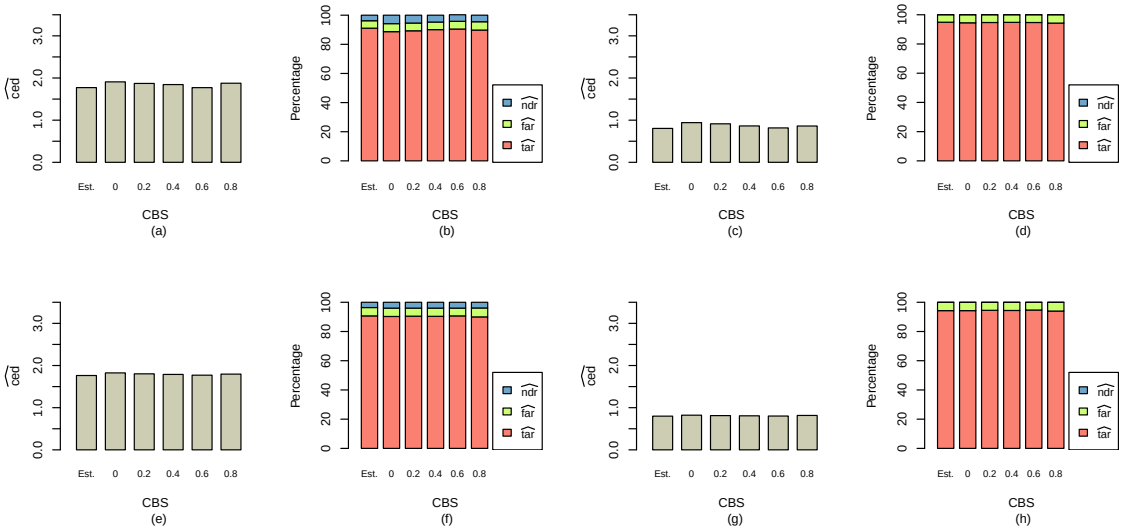

Figure 5.12: Scenarios R5 and R6 (Mis-specification of CBS): (a-b) and (c-d) scenario R5 for shift sizes 2 and 3, and (e-f) and (g-h) scenario R6 for shift sizes 2 and 3.

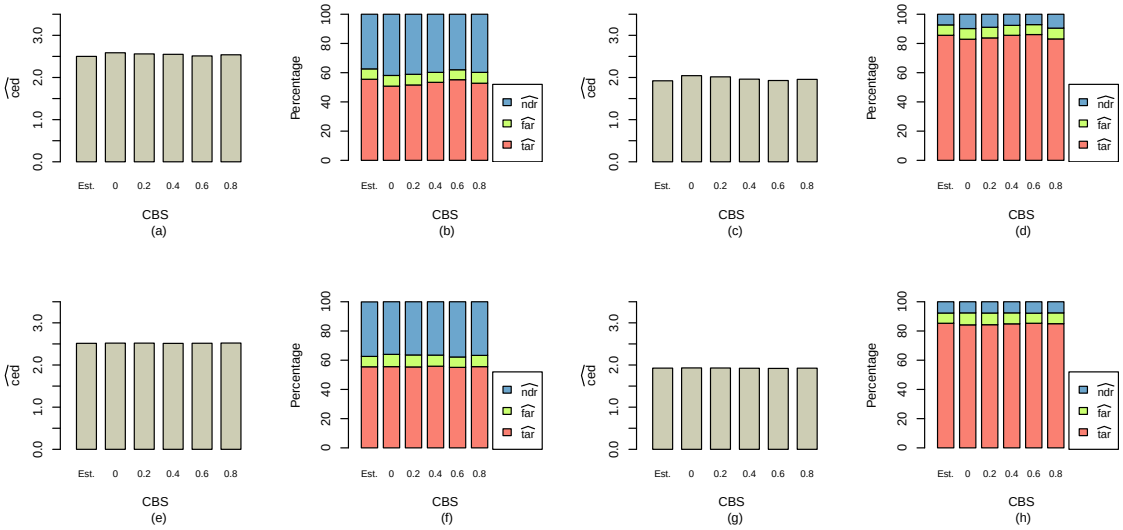

Figure 5.13: Scenarios R7 and R8 (Mis-specification of CBS): (a-b) and (c-d) scenario R7 for shift sizes 2 and 3, and (e-f) and (g-h) scenario R8 for shift sizes 2 and 3.

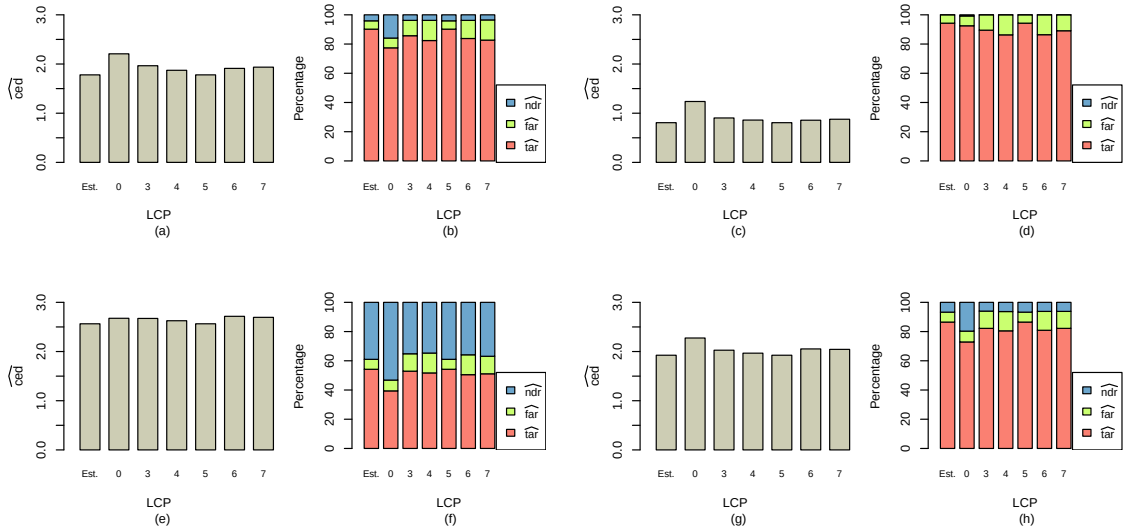

Figure 5.14: Scenarios R9 and R10 (Mis-specification of LCP): (a-b) and (c-d) scenario R9 for shift sizes 2 and 3, and (e-f) and (g-h) scenario R10 for shift sizes 2 and 3.

5.3.1.5.3 Mis-specification of LCP

Having assumed that there is lag between series (LCP = 5) which is related to correlation between series (CBS $\neq$ 0), the detection performances of scenarios R9 and R10 are illustrated in Figure 5.14 (a-b) and (c-d), respectively. In both Figures the correctly specified LCP is 5 and the estimated LCP is denoted by Est. The full results from the simulation study are provided in Tables B.13 and B.14 in Appendix B. Assuming that the observations are independent (scenario R9), mis-specifying LCP delays detection with longer $\widehat{\text{ced}}$ at all shift sizes. There is no difference in the performances between using correctly specified LCP and estimated LCP, while the more LCP is mis-specified, the longer $\widehat{\text{ced}}$ with lower $\widehat{\text{tar}}$ and higher $\widehat{\text{far}}$ produced. The same pattern is repeated for the autocorrelated process (Scenario R10). Due to the CWS, $\widehat{\text{ndr}}$ is quite high compared to scenario R9. However, again, the $\widehat{\text{ndr}}$ is substantially improved when the shift size is large.

5.3.2 Specification of shift size

Since a shift size that we want to detect by the SR methods needs to be pre-specified, in this section we aim to investigate the detection performance of the SR methods if the shift in a process is smaller or greater than that pre-specified. To investigate the detection performance, we pre-specify a shift of size 2 as a shift size that we aim to detect by the SR methods. We investigate the performance in detecting various shift sizes for simultaneous changes. Changes with time lag, which are considered as the generalization case of the simultaneous change if time lag is known, could be investigated

in the similar manner but are not presented here. The detection performance of SR methods are compared against those from the parallel method since pre-specified shift shift is not a requirement of the parallel method. However, we anticipate that due to the adjustment for the multiplicity in the parallel methods, it will still perform less well than the SR methods.

Four possible scenarios, with details of a suitable SR method and the corresponding one-sided EWMA chart in each scenario, are summarized in Table 5.5. Data are generated according to the parameters defined in each scenario. We define $c = 2$ as the pre-specified shift size which SR methods are designed to detect and d as a shift size in the actual process simulated. Five different shift sizes (d = 0.5, 1, 2, 3, 4) are investigated. In scenarios R11 and R12, the parallel method is monitored with a standard one-sided EWMA chart, while in scenarios R13 and R14 (for the autocorrelated process, CWS = 0.6), the one-sided EWMA chart modified for autocorrelated data (see Section 5.2) is used instead, in order to avoid a high false alarm rate.

Let N be the number of simulations. The robustness study for specification of shift size can be described as follows.

step 1 a bivariate series is generated according to parameters (CWS, CBS and LCP) defined in each scenario.

step 2 a change point of a process, τ, is randomly selected ($\tau > 150$; i.e. to make sure that the in control baseline is long enough for the parameter estimation). A shift of size d is added to the data at time τ and thereafter.

step 3 data (parallel method) and a derived sequence of statistics (from SR methods designed to detect a specified shift size of $c = 2$) are monitored with the appropriate one-sided EWMA chart. The delay in detection for each method is recorded (i.e. $t_A - \tau$, where t_A is the time of an alarm). If $t_A - \tau \geq 7$, NA (non-applicable) is reported, i.e. system fails to detect a shift within 7 time points.

step 4 repeat steps 1-3 N times ($N = 10,000$). The CED, TAR, FAR and NDR for each method are estimated.

In each scenario, this entire procedure is repeated for five values of shift size in a process d, $d = 0.5, 1, 2, 3, 4$. The performances of the SR methods for detecting different shift sizes are investigated by comparing the differences of the statistics estimated from the simulations against the critical values defined in Table 3.2.

5.3.2.1 Results

Having pre-specified a shift of size 2 as that which the SR method is designed to detect, five different values of actual shift in the process are investigated. The comparisons of the detection performance of the parallel and the SR methods are illustrated in Figures 5.15 (scenarios R11 and R12) and 5.16 (scenarios R13 and R14). The full results from the simulation study are summarized in Tables B.15 - B.18, respectively, in Appendix B. Overall, the CED, TAR and NDR of the methods for detecting different shift sizes

Table 5.5: Four scenarios of specification of shift sizes

Scenario	Parameters			Pre-specified shift size	Shift size in a process	Method	Chart
	CWS	CBS	LCP				
R11	0	0	0	2	0.5, 1, 2, 3, 4	Frisén	standard
R12	0	0.6	0	2	0.5, 1, 2, 3, 4	Wessman	standard
R13	0.6	0	0	2	0.5, 1, 2, 3, 4	Case 2	standard
R14	0.6	0.6	0	2	0.5, 1, 2, 3, 4	Case 4	standard

are significantly different. However, there is no difference in FAR for all cases. Unsurprisingly, the larger shift size, the shorter delay with higher $\widehat{\text{tar}}$ and lower $\widehat{\text{ndr}}$ produced.

Consider process of independent data (scenarios R11 and R12), the Frisén and Wessman methods perform significantly differently from the parallel method by giving shorter $\widehat{\text{ced}}$, higher $\widehat{\text{tar}}$ and lower $\widehat{\text{far}}$. As expected, larger shift size are generally easier to detect even if they differ from the pre-specified value. In the case of autocorrelated process (scenarios R13 and R14), the same pattern is repeated. Since a small shift size ($d = 0.5$) is unlikely to be detected in the autocorrelated process (CWS = 0.6), the $\widehat{\text{tar}}$ is quite low (10% or less). The $\widehat{\text{ced}}$, which is calculated based on the true detection, of this shift size might be misleading since the calculations are only from the true detection (i.e. 10% or fewer of the simulations). Overall, even though the shift size we aim to detect by the SR methods needs to be pre-specified, such methods still perform well for detecting the shift size in a process which might be smaller or greater than the pre-specified value. Due to the adjustment for the multiplicity, the parallel method, for which the specification of shift size is not required, performs significantly differently from the SR methods by showing longer $\widehat{\text{ced}}$, lower $\widehat{\text{tar}}$ and higher $\widehat{\text{ndr}}$.

5.4 Conclusions

Since not incorporating CWS in the sufficient reduction delays the detection and produces many false alarms in Parallel, Frisén and Wessman methods, a one-sided EWMA chart modified for detecting a positive shift in an autocorrelated process is proposed. The comparison between using standard and modified one-sided EWMA chart is carried out by using those charts to monitor the parallel data and the derived sequence of statistics from three methods (Frisén and Wessman and our methods). The FAR of Parallel, Frisén and Wessman methods is improved substantially if the modified chart is used, even though the CED is slightly higher due to the higher control limits of the modified chart adjusted for the CWS. There is little difference in the detection performance between charts in our methods, since the CWS has been removed during the sufficient reduction. This makes the modification of one-sided EWMA chart for autocorrelated process unnecessary for our methods. Taking CWS into account by using either SR methods proposed for autocorrelated data (our methods) or using modified one-sided EWMA chart for autocorrelated data should improve the performance of detection and

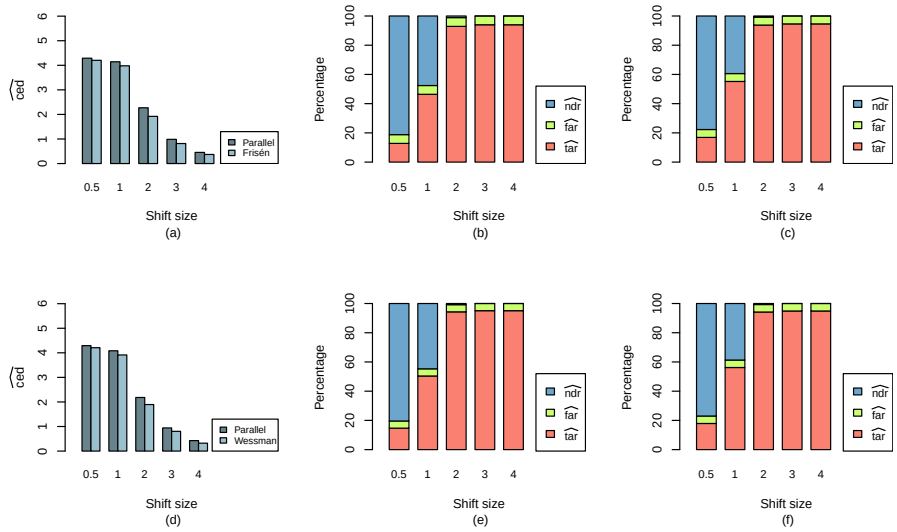

Figure 5.15: Scenarios 11 and 12 (Specification of shift size): (a-c) results for parallel and Frisén methods, scenario R11 and (d-f) results for parallel and Wessman methods, scenario R12.

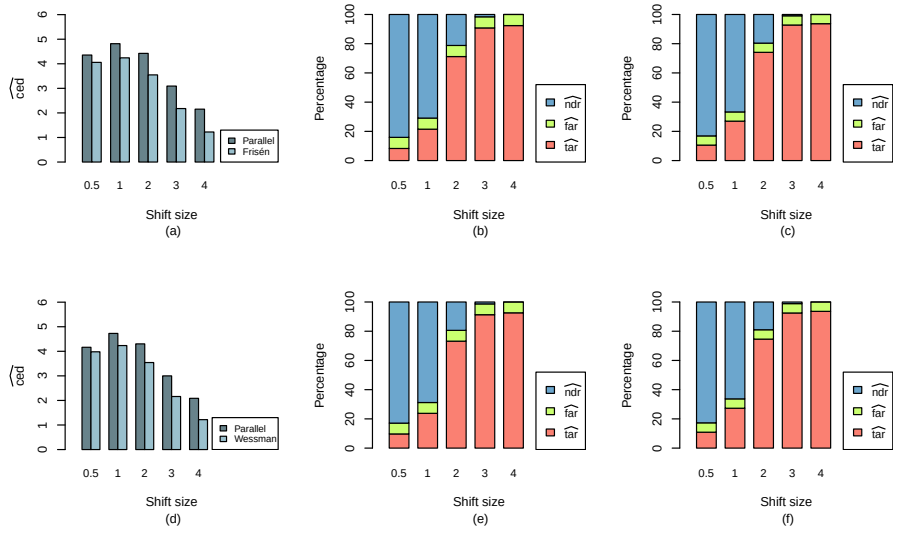

Figure 5.16: Scenarios 13 and 14 (Specification of shift size): (a-c) results for parallel and case 2 methods, scenario R13 and (d-f) results for parallel and case 4 methods, scenario R14.

reduce false alarm rates. Therefore, due to the nature of original data and the dependency of the derived likelihood ratio statistics, the one-sided EWMA chart should be chosen carefully since the result will be misleading if an inappropriate detection tool is used. Overall, comparing the detection performances of scenarios 2 and 3 between using a standard EWMA chart (Chapter 4) and a modified EWMA chart (Chapter 5), our proposed methods with the standard chart give shorter delay than other methods with the modified chart due to the higher control limits adjusted for CWS in the modified chart.

Apart from not taking CWS into account, mis-specifying parameters used in the sufficient reduction is another concern. A simulation study was set up to investigate how SR methods perform if the parameter (CWS, CBS or LCP) is mis-specified in some way. The detection performance using the correctly specified parameter, an estimated parameter (estimated using a standard method) and mis-specified values of parameter are compared. As expected, the results of using estimated parameter are quite similar to those of using the correctly specified parameter (i.e. that used to generate the data), while using mis-specified values of the parameter gave poor detection performance. Mis-specifying CWS gave the worst scenarios by giving longer CED and higher FAR for detecting simultaneous changes and changes with time lags. Mis-specifying CBS gave longer delays in detecting simultaneous change, but no appreciable difference in detecting change with time lags. Even though the effect of mis-specifying CBS is not quite as strong compared to those of mis-specifying CWS, the detection performance is improved if CBS is taken into account appropriately. Also, if LCP is truly present, mis-specifying LCP also gives worse results with longer CED. In this case, the results might be misleading because not only is a mis-specified LCP used, but this also affects the cross correlation between series (CBS) which will be mis-calculated due to mis-specified LCP.

Since a shift size, which the method aims to detect, must be pre-specified for SR methods, data with different shift sizes in the process are used to investigate how the SR methods perform when the shift size in the process is less or greater than that pre-specified. The results from the simulation shows that the SR methods still perform better than the parallel method, which does not require the specification of shift size. A large shift size is more likely to be detected than a small shift size even if it differs from the value pre-specified.

The SR methods considered in this thesis are proposed to detect a step change in the mean of a process. However, according to the nature of diseases and the limiting assumptions of the SR methods, the robustness study of the SR methods might be extended in several respects. For example, in public health surveillance, changes in health data might be of several shapes (e.g. single spike, linear, exponential shapes or epidemic curves) depending on types of disease (Goldenberg *et al.*, 2002; Mandl *et al.*, 2004; Wang *et al.*, 2005; Jackson *et al.*, 2007). Thus a robustness study of how SR methods perform in detecting a mean shift within different shapes might be of interest but is not presented here.

Chapter 6

Sufficient reduction methods for Poisson data

6.1 Background

In previous chapters, we proposed sufficient reduction (SR) methods for detecting a shift in public health surveillance by assuming that data are normal distributed. However, incidences for some diseases, such as emerging or sparse diseases, might be small counts. Such data are preferably considered as coming from a Poisson distribution rather than a normal distribution. In this chapter we aim to develop sufficient reduction methods for detecting a mean shift in a process of Poisson data. One important thing to be noted is that unlike detecting a mean shift in normal data, where the variance is assumed unchanged over time, mean and variance of Poisson distribution are not separable. If there is a shift in mean level of Poisson data, of course, the variance will shift as well. Therefore, the SR methods for Poisson data are developed under this consideration.

Before we develop SR methods for Poisson data, we note that an alternative (and perhaps simpler) way to monitor a mean shift in a process of Poisson data is to monitor residuals from a Poisson regression model. It is assumed that (Pearson) residuals are independent and normal distributed. Thus they can be handled by the SR methods proposed for normal data in Chapter 4. Details of fitting Poisson regression models are provided in Section 6.2. Even though using model residuals instead of actual Poisson data might be an alternative way for monitoring mean shift, we still want to develop SR methods for Poisson data in order to investigate whether or not monitoring actual Poisson data by SR methods has advantages over the alternative method mentioned above. Bivariate Poisson series and the relations between series considered in this study are defined in Section 6.3.

SR methods for Poisson data are developed under four different scenarios: either correlation within series (CWS) or correlation between series (CBS) or neither or both is (are) present. Detail of each scenario is given in Section 6.4. The detection performance of SR methods are compared with parallel and Frisén methods. With regard to the distribution of derived likelihood ratio statistics, the one-sided EWMA chart, which is used as a detection tool in this study, should be chosen carefully for each method. It

was found later (see Section 6.6) that the sequences of likelihood ratio statistics derived from SR methods proposed for Poisson data look more normal than Poisson. Therefore, the derived sequences from SR methods are monitored with a one-sided EWMA chart for normal data. On the other hand, the actual data in the parallel method, which are Poisson distributed, are monitored with a one-sided EWMA chart for Poisson data. In this chapter, one-sided EWMA charts for Poisson data are modified for monitoring a mean shift in the parallel method. Details of the modification of the one-sided EWMA chart for independent and dependent Poisson data are given in Sections 6.5.1 and 6.5.2, respectively. The comparison of detection performances between the three methods is conducted via a simulation study. The results are summarized in Section 6.6 with tables and figures followed by conclusions in Section 6.7.

6.2 Poisson regression model

Since health data are normally count data, the Poisson regression model has been used in public health surveillance, either retrospective or prospective, in several studies. Count data are fitted with a Poisson regression model by using a log link function, $g(\mu_t)$, which links the mean of response variable (Y_t), μ_t, to a linear predictor, η_t, and so to explanatory variables $(X_{1,t}, ..., X_{k,t})$. In general, a Poisson regression model can be written as

$$g(\mu_t) = \eta_t = \beta_0 + \beta_1 x_{1,t} + ... + \beta_k x_{k,t}$$

where $\beta_0, ..., \beta_k$ are coefficients and $x_{k,t}$ is the observed value of explanatory variable k at time t. The log link, which is the natural logarithm of the mean of the response variable, is

$$g(\mu_t) = log(\mu_t) = \beta_0 + \beta_1 x_{1,t} + ... + \beta_k x_{k,t}$$

Therefore,

$$E(Y_t) = \mu_t = exp(\beta_0 + \beta_1 x_{1,t} + ... + \beta_k x_{k,t})$$

(Dobson and Barnett, 2008).

With regard to Poisson data, where the variance is equal to, or a function of the mean, we investigate how well the model fits by considering Pearson residuals, which are the difference between y_i and $\hat{\mu}_t$ standardized by the estimated Poisson standard deviation. We consider the Pearson residuals because they are commonly used to assess the generalized linear models due to their direct analogy with linear models. If the model fits well, the residuals fluctuate around zero and are approximately standard normal when μ_t is large and the sample size is not too small. Pearson residuals at time t can be calculated from

$$e_t = \frac{y_t - \hat{\mu}_t}{\sqrt{V(\hat{\mu}_t)}} = \frac{y_t - \hat{\mu}_t}{\sqrt{\hat{\mu}_t}}$$

where $\hat{\mu}_t$ is an estimate of μ_t (Agresti, 2007).

Before developing SR methods for Poisson data, we want to investigate whether or not there are the advantages in using Pearson residuals, which are assumed normally distributed, with SR methods proposed for normal data in chapter 4. If it works well,

the extension of SR methods for Poisson data, which is more complicated, might not be necessary. For simplicity as a toy example, we consider two independent Poisson series (i.e. both are without CWS and CBS). Data in each series are generated from Poisson distribution with parameter $\mu = 5$. In order to evaluate the detection performance, let τ be a change point which divides the process into two stages, in control $(t = 1, ..., \tau - 1)$ and out of control $(t = \tau, ..., s)$. A mean shift of size 2 is added to series i at time $t = \tau, ..., s$. Data for in control and out of control stages are defined below.

$$\begin{array}{ll} X_i \sim Pois(5),\ t = 1, ..., \tau - 1 & \text{(in control stage)} \\ X_i \sim Pois(7),\ t = \tau, ..., s & \text{(out of control stage)} \end{array}$$

where $i = 1, 2$. Due to the data generated with parameter μ_i, we fit Poisson regression model with only an intercept as a constant mean level to each series separately. Data and the model for each series are defined as follows.

$$X_i \sim Pois(\mu_i)$$

$$E(X_{i,t}) = \mu_{i,t} = exp(\beta_0)$$

There are two different ways of fitting models considered in this study. The first approach, called M1, is to fit Poisson regression model defined above to the in control stage data $(t = 1, ..., \tau - 1)$ and then make $s - \tau - 1$ steps ahead forecasts for the rest. The residuals from the first part are from fitting the model $(t = 1, ..., \tau - 1$, while the rest are from subtracting the predictions from observed data $(t = \tau, ..., s)$. The M1 model can be written as

$$E(X_{i,t}) = \mu_{i,t} = exp(\beta_0),\ t = 1, 2, ..., \tau - 1$$

The second approach, called M2, is to refit the model defined above at each time t using data available from $t = 1, 2, ..., t$. The M2 model can be written as

$$E(X_{i,t}) = \mu_{i,t} = exp(\beta_t),\ t = 1, 2, ..., s$$

Residuals from each model are monitored separately in the parallel method and used as the data in SR methods proposed for normal data since they are approximately normally distributed. Detection performance is evaluated by monitoring the parallel residual series or derived sequences from SR methods with a EWMA chart. Since residuals are assumed to be independent and normally distributed with mean zero, a standard one-sided EWMA chart for independent normal data is used.

Even though using the model residuals, which are approximately normally distributed, is the common method for handling autocorrelated and non-normal data, in the case of prospective surveillance, where a decision is made sequentially, it should be used with care. Using residuals from model M1 is commonly used, since a step change in mean might be seen by fitting the model and making a prediction according to the constant mean level observed in the in control stage. However, Box *et al.* (1994) suggested that residuals from making predictions far ahead from the in control stage might not be independent (Box *et al.*, 1994). Refitting time series at each time point in model M2 is computationally time consuming and might oversmooth a step change which is more obviously seen from the M1 model approach. Therefore, one might be open to other

methods, such as modified sufficient reduction methods (see Section 6.3). We compare the methods below, but note specifically that in order to use the model residuals with the sufficient reduction methods proposed for normal data, one should check particularly carefully whether or not the model fits well so that the residuals are approximately normally distributed.

6.3 Bivariate Poisson series

In this study we aim to develop SR methods for Poisson data. For simplicity we consider only bivariate Poisson (BP) series (two dimensions, $p = 2$) with no lag between series (i.e. only simultaneous change is considered), while the extension to multivariate Poisson series might be considered as a further study. As in the normal case, SR methods proposed for Poisson data are aimed to take both CWS and CBS into account. For the BP series considered in this study, we define $\boldsymbol{x}_t$ as a two dimensional vector representing the observation made on a BP series at time t, $t = 1, ..., s$

$$\boldsymbol{x}_t = \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix}$$

Regarding CWS, the definitions of BP series for independent and dependent observations are given in Sections 6.3.1 and 6.3.2, respectively. Since mean and variance of Poisson distribution are not separable, the relation of mean shifts between series is defined differently from the normal case and is detailed in Section 6.3.3.

6.3.1 Independent observations

Assume that data in each series are independent (CWS = 0), BP series (X_1 and X_2) can be defined according to the CBS. If two series are independent (CBS = 0), X_1 and X_2 are independent random variables from Poisson distributions with parameters μ_1 and μ_2, respectively. The joint probability function is

$$f(\boldsymbol{x}_t; \mu_1, \mu_2) = e^{-(\mu_1+\mu_2)} \frac{\mu_1^{x_1}}{x_1!} \frac{\mu_2^{x_2}}{x_2!} \tag{6.1}$$

The mean and variance of each series are

$$E(X_i) = V(X_i) = \mu_{X_i} \tag{6.2}$$

where $Cov(X_1, X_2) = 0$.

If two series are dependent (CBS $\neq$ 0), BP series can be generated by the trivariate reduction method (Holgate, 1964; Johnson *et al.*, 1997; Karlis and Ntzoufras, 2005; Yahav and Shmueli, 2012b). The idea is to construct X_1 and X_2 from three mutually independent Poisson random variables. Let Z_1, Z_2 and Z_3 be mutually independent random variables from Poisson distributions with parameters μ_1, μ_2 and μ_3, respectively. The random variables X_1 and X_2 follow a bivariate Poisson distribution, $X_1, X_2 \sim BP(\mu_1, \mu_2, \mu_3)$, if

$$X_1 = Z_1 + Z_3 \quad \text{and} \quad X_2 = Z_2 + Z_3$$

Marginally, X_1 and X_2 follow Poisson distributions with means (and variances), covariance and correlation between series defined below.

$$\begin{aligned} E(X_i) = V(X_i) &= \mu_i + \mu_3 = \mu_{X_i} \\ Cov(X_1, X_2) &= \mu_3 \\ Cor(X_1, X_2) &= \frac{\mu_3}{\sqrt{(\mu_1 + \mu_3)(\mu_2 + \mu_3)}} \end{aligned} \tag{6.3}$$

The probability function (pdf) of BP distribution can be written as

$$f(\boldsymbol{x}_t; \mu_1, \mu_2, \mu_3) = e^{-(\mu_1+\mu_2+\mu_3)} \frac{\mu_1^{x_1}}{x_1!} \frac{\mu_2^{x_2}}{x_2!} \sum_{j=0}^{m} \begin{pmatrix} x_1 \\ j \end{pmatrix} \begin{pmatrix} x_2 \\ j \end{pmatrix} j! \left(\frac{\mu_3}{\mu_1 \mu_2}\right)^j \tag{6.4}$$

where $m = min(x_1, x_2)$ (Johnson *et al.*, 1997; Karlis and Ntzoufras, 2005). The dependence between series is measured by μ_3. If $\mu_3 = 0$, X_1 and X_2 are independent and so the pdf will be similar to equation (6.1).

6.3.2 Dependent observations

In the case of dependent observations, we consider bivariate Poisson series in which each series has an autoregressive model of order 1 (AR(1)) which is generally defined by

$$x_t = \phi x_{t-1} + \epsilon_t, \tag{6.5}$$

where ϕ is an autoregressive coefficient, $0 < \phi < 1$, and ϵ_t is Gaussian random noise with $N(0, \sigma_\epsilon^2)$.

With regard to positive count data, an autoregressive model for Poisson data, a Poisson integer-valued autoregressive model (Poisson INAR model), was introduced by McKenzie (1985, 1988), Al-Osh and Alzaid (1987) and Alzaid and Al-Osh (1990). In addition to the use of Poisson innovation terms, the use of binomial thinning was introduced, replacing scalar multiplication in the original AR(1) model defined above by a binomial thinning operator denoted by $*$. Let X_t be a positive discrete random variable and Y_i a sequence of independent Bernoulli trials, $B(1, \phi)$, such that $P(Y_i = 1) = \phi = 1 - P(Y_i = 0)$. $\phi * X$ is a binomial random variable defined from

$$\phi * x = \sum_{i=1}^{x} Y_i$$

(Steutel and Van Harn, 1979). Thus, $\phi * x$ is the number of success from x independent trials with probability of success ϕ for each trial (McKenzie, 1988; Cardinal *et al.*, 1999).

To develop the Poisson INAR(1) model, the scalar multiplication in AR(1) model (equation (6.5)) is replaced by the binomial thinning operator. Let R_t be a random variable

following a Poisson distribution with parameter μ. A Poisson integer-valued autoregressive model of degree 1 (Poisson INAR(1) model) can be written as

$$x_t = \phi * x_{t-1} + r_t. \tag{6.6}$$

Thus, x_t is a stationary process of a Poisson random variable of mean $\mu/(1-\phi)$ and Gaussian AR(1)-like properties (McKenzie, 1985, 1988). The conditional mean, conditional variance and the autoregressive function of degree 1 can be calculated from

$$E(X_t|X_{t-1}) = \phi x_{t-1} + \mu \tag{6.7}$$
$$V(X_t|X_{t-1}) = \phi(1-\phi)x_{t-1} + \mu \tag{6.8}$$
$$Cor(X_t, X_{t+1}) = \phi$$

(McKenzie, 1985, 1988). The conditional probability function is given by

$$f(x_t|x_{t-1}) = \sum_{j=0}^{g} \left(\begin{array}{c} x_{t-1} \\ j \end{array} \right) \phi^j (1-\phi)^{x_{t-1}-j} \frac{e^{-\mu}\mu^{x_t-j}}{(x_t-j)!}, \tag{6.9}$$

where $g = min(x_t, x_{t-1})$. The idea of using Poisson INAR(1) in public health surveillance can be explained in the sense that X_t is the number of patients admitted to hospital on day t, R_t is the number of new patients on day t and $\phi * X_{t-1}$ is the number of previous patients still in the hospital (Al-Osh and Alzaid, 1987; Strat, 2005).

In the case of bivariate series, if the two series are independent (CBS = 0), X_1 and X_2 are independent and follow Poisson INAR(1) process with $R_{1,t} \sim P(\mu_1)$ and $R_{2,t} \sim P(\mu_2)$, respectively. According to equation (6.9), the conditional pdf of independent bivariate Poisson INAR(1) process can be written as

$$\begin{aligned} f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1};\phi_1,\phi_2,\mu_1,\mu_2) &= \sum_{j=0}^{g_1} \left(\begin{array}{c} x_{1,t-1} \\ j \end{array} \right) \phi_1^j (1-\phi_1)^{x_{1,t-1}-j} \cdot \frac{e^{-\mu_1}\mu_1^{x_{1,t}-j}}{(x_{1,t}-j)!} \cdot \\ & \sum_{j=0}^{g_2} \left(\begin{array}{c} x_{2,t-1} \\ j \end{array} \right) \phi_2^j (1-\phi_2)^{x_{2,t-1}-j} \cdot \frac{e^{-\mu_2}\mu_2^{x_{2,t}-j}}{(x_{2,t}-j)!} \end{aligned} \tag{6.10}$$

where $g_1 = min(x_{1,t}, x_{1,t-1})$ and $g_2 = min(x_{2,t}, x_{2,t-1})$. The mean and variance of series i is

$$E(X_i) = V(X_i) = \mu_i/(1-\phi_i), \tag{6.11}$$

while the conditional mean and variance of each series are as follows.

$$E(X_{i,t}|X_{i,t-1}) = \phi_i x_{t-1} + \mu_i \tag{6.12}$$
$$V(X_{i,t}|X_{i,t-1}) = \phi_i(1-\phi_i)x_{t-1} + \mu_i \tag{6.13}$$
$$Cor(X_{i,t}, X_{i,t+1}) = \phi_i$$

If the two series are not independent (CBS $\neq$ 0), Pedeli and Karlis (2005) have extended the idea of the Poisson INAR(1) model and BP distribution defined in Section 6.3.1 to propose a model for two dependent Poisson INAR(1) processes called a Poisson Bivariate Integer-Valued Autoregressive Model of degree 1 (Poisson BINAR(1) model). Let X_1

and X_2 be a positive integer random variable and $R_1, R_2 \sim BP(\mu_1, \mu_2, \mu_3)$, the Poisson BINAR(1) model can be written as

$$x_{1,t} = \phi_1 * x_{1,t-1} + r_{1,t},$$
$$x_{2,t} = \phi_2 * x_{2,t-1} + r_{2,t},$$

where ϕ_i is an autoregressive coefficient of series i, $i = 1, 2$, respectively. Since $R_1, R_2 \sim BP(\mu_1, \mu_2, \mu_3)$, $E(R_i) = \mu_i + \mu_3$ and $Cov(R_1, R_2) = \mu_3$. The mean, variance, covariance and correlation between series can be calculated from

$$E(X_i) = V(X_i) = \frac{\mu_i + \mu_3}{1 - \phi_i} = \mu_{X_i} \tag{6.14}$$
$$Cov(X_1, X_2) = \frac{\mu_3}{1 - \phi_1\phi_2}$$
$$Cor(X_1, X_2) = \frac{\mu_3\sqrt{(1-\phi_1)(1-\phi_2)}}{(1-\phi_1\phi_2)\sqrt{(\mu_1+\mu_3)(\mu_2+\mu_3)}}$$
$$Cor(X_{i,t}, X_{i,t+1}) = \phi_i$$

Due to the correlation within series of lag 1, the conditional mean, variance and covariance can be calculated from

$$E(X_{i,t}|X_{i,t-1}) = \phi_i x_{i,t-1} + \mu_i + \mu_3 \tag{6.15}$$
$$V(X_{i,t}|X_{i,t-1}) = \phi_i(1-\phi_i)x_{i,t-1} + \mu_i + \mu_3 \tag{6.16}$$
$$Cov(X_{1,t}, X_{2,t}|X_{1,t-1}, X_{2,t-1}) = \mu_3$$

(Pedeli and Karlis, 2005).

The conditional probability function of Poisson BINAR(1) model can be derived from the convolution of two binomial distributions and the bivariate Poisson distribution defined below.

$$X_{1,t} \sim Bin(X_{1,t-1}, \phi_1)$$
$$X_{2,t} \sim Bin(X_{2,t-1}, \phi_2)$$
$$R_{1,t}, R_{2,t} \sim BP(\mu_1, \mu_2, \mu_3)$$

Thus,

$$\begin{aligned} f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \phi_1, \phi_2, \mu_1, \mu_2, \mu_3) &= \sum_{r_1=0}^{g_1}\sum_{r_2=0}^{g_2} f_1(x_{1,t}-r_1) f_2(x_{2,t}-r_2) f_3(r_1, r_2) \\ &= \sum_{r_1=0}^{g_1}\sum_{r_2=0}^{g_2} \Bigg\{ \binom{x_{1,t-1}}{x_{1,t}-r_1} \phi_1^{x_{1,t}-r_1}(1-\phi_1)^{x_{1,t-1}-x_{1,t}-r_1} \cdot \\ &\qquad \binom{x_{2,t-1}}{x_{2,t}-r_2} \phi_2^{x_{2,t}-r_2}(1-\phi_2)^{x_{2,t-1}-x_{2,t}-r_2} \cdot \\ &\qquad e^{-(\mu_1+\mu_2+\mu_3)} \frac{\mu_1^{r_1}}{r_1!} \frac{\mu_2^{r_2}}{r_2!} \cdot \\ &\qquad \sum_{j=0}^{m} \binom{r_1}{j} \binom{r_2}{j} j! \left(\frac{\mu_3}{\mu_1\mu_2}\right)^j \Bigg\}, \end{aligned} \tag{6.17}$$

where $g_1 = min(x_{1,t}, x_{1,t-1})$, $g_2 = min(x_{2,t}, x_{2,t-1})$ and $m = \min(r_1, r_2)$.

6.3.3 Relation of mean shift between series

The relation of mean shifts between BP series is considered in a different way from the normal case. For normal data, we assumed that the variance is unchanged over time, while the mean has shifted from one to another level (see more detail in Section 4.3.2). Unfortunately, the mean and variance of Poisson distribution cannot be separated, so the variance will change according to a shift in mean. This has implications for the type of real data which can be described by Poisson models. However, in many instances, it is not unreasonable that if true disease incidence rises, counts of infected patients will become more variable at a similar rate, so the coupling of mean and variance is not unrealistic.

With regard to the BP distribution defined above, means (and also variances) of X_1 and X_2 are related to the covariance between series (μ_3). If the covariance changes, it will affect the mean levels in both series. Therefore, in Poisson data, we investigate a mean shift in Poisson process considered as a consequence of a shift in covariance. A shift in covariance is reflected in a mean shift in both series if X_1 and X_2 are not independent ($\mu_3 \neq 0$ and CBS $\neq 0$). On the other hand, if two series are independent, a shift in any one series can occur independently. In this study we assume that if CBS $= 0$, a mean shift occurs in X_1 series only.

Let c be the constant mean shift size we want to detect. In the case of two independent series, $X_1 \sim P(\mu_1)$ and $X_2 \sim P(\mu_2)$ are independent Poisson random variables. If there is a mean shift of size c in a process, $X_1 \sim P(\mu_1 + c)$. The mean and variance after the process has shifted ($t = \tau, ..., s$) are

$$E(X_1) = V(X_1) = \mu_1 + c \tag{6.18}$$

$$E(X_2) = V(X_2) = \mu_2 \tag{6.19}$$

On the other hand, if X_1 and X_2 are correlated with $BP(\mu_1, \mu_2, \mu_3)$, we investigate the mean shift of size c in both series from a shift of size c in covariance parameter ($Z_3 \sim P(\mu_3 + c)$). After the process has shifted X_1 and X_2 follow $BP(\mu_1, \mu_2, \mu_3 + c)$ with mean, variance, covariance and correlation between series defined as follows.

$$\begin{aligned} E(X_i) = V(X_i) &= \mu_i + \mu_3 + c \\ Cov(X_1, X_2) &= \mu_3 + c \\ Cor(X_1, X_2) &= \frac{\mu_3 + c}{\sqrt{(\mu_1 + \mu_3 + c)(\mu_2 + \mu_3 + c)}} \end{aligned} \tag{6.20}$$

Note that, in the case of independent observations (CWS $= 0$), the mean shift size, c, in both series is the same as the covariance shift size due to the trivariate reduction method used for BP distribution.

For a process of dependent observations (CWS $\neq 0$), the relation of mean shifts between series can be defined in a similar manner. We investigate the relation of mean shift from

a shift in covariance between series (μ_3) defined in $R_{1,t}, R_{2,t} \sim BP(\mu_1, \mu_2, \mu_3)$, while $X_{1,t} \sim Bin(X_{1,t-1}, \phi_1)$ and $X_{2,t} \sim Bin(X_{2,t-1}, \phi_2)$ are the same for both stages. If two series are independent (CBS = 0; $\mu_3 = 0$), the mean shift occurs in the X_1 series only. The mean and variance of Poisson INAR(1) process after the process has shifted are

$$E(X_1) = V(X_1) = \frac{\mu_1 + c}{1 - \phi_1} \tag{6.21}$$

$$E(X_2) = V(X_2) = \frac{\mu_2}{1 - \phi_2} \tag{6.22}$$

and the conditional mean and variance of Poisson INAR(1) process are

$$E(X_{1,t}|X_{1,t-1}) = \phi_1 x_{1,t-1} + \mu_1 + c \tag{6.23}$$

$$V(X_{1,t}|X_{1,t-1}) = \phi_1(1 - \phi_1)x_{1,t-1} + \mu_1 + c \tag{6.24}$$

$$E(X_{2,t}|X_{2,t-1}) = \phi_2 x_{2,t-1} + \mu_2 \tag{6.25}$$

$$V(X_{2,t}|X_{2,t-1}) = \phi_2(1 - \phi_2)x_{2,t-1} + \mu_2. \tag{6.26}$$

On the other hand, if two series are dependent (CBS $\neq 0$; $\mu_3 \neq 0$), the mean shift occurs in both series. The shift in each series depends on the shift of size c in the covariance and ϕ_i. The mean and variance of Poisson INAR(1) process after the process has shifted are

$$E(X_i) = V(X_i) = \frac{\mu_i + \mu_3 + c}{1 - \phi_i} \tag{6.27}$$

$$Cov(X_1, X_2) = \frac{\mu_3 + c}{1 - \phi_1\phi_2}$$

$$Cor(X_1, X_2) = \frac{(\mu_3 + c)\sqrt{(1 - \phi_1)(1 - \phi_2)}}{(1 - \phi_1\phi_2)\sqrt{(\mu_1 + \mu_3 + c)(\mu_2 + \mu_3 + c)}}$$

and the conditional mean and variance of the process are

$$E(X_{i,t}|X_{i,t-1}) = \phi_i x_{i,t-1} + \mu_i + \mu_3 + c \tag{6.28}$$

$$V(X_{i,t}|X_{i,t-1}) = \phi_i(1 - \phi_i)x_{i,t-1} + \mu_i + \mu_3 + c \tag{6.29}$$

$$Cov(X_{1,t}, X_{2,t}|X_{1,t-1}, X_{2,t-1}) = \mu_3 + c$$

where $i = 1, 2$ (Pedeli and Karlis, 2005).

6.4 SR methods for Poisson data

Wessman and Frisén proposed SR methods for detecting a mean shift in any exponential family distribution, including the Poisson distribution, however, their work proceeded under the assumption that observations are independent (CWS = 0) (Wessman, 1998; Frisén *et al.*, 2011) which is not typically assumption in health surveillance. To bridge the gap, in this section we aim to develop SR methods for detecting a mean shift in a process of Poisson count data. As mentioned previously, detecting mean shift in a Poisson process is not quite straightforward as mean and variance are not separable. A shift in mean results in an increase in variance of a process as well. For simplicity, we develop SR methods for Poisson data for detecting a mean shift in a BP series. The assumptions made in this study are

- aim to detect a sudden, but persistent, mean shift in bivariate Poisson series (X_1 and X_2)
- allow for CWS and CBS
- simultaneous change, denoted by τ, only is considered
- investigate the relation of mean shifts via the covariance between series
- if CBS $\neq 0$, mean shifts occur in both series at time τ; otherwise, a mean shift occurs in one series, X_1, at time τ if CBS $= 0$ (i.e. covariance is zero)

SR methods for Poisson data can be derived in a similar manner to Chapter 4 (SR methods for Normal data). Let $X_s = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_s)$ be the available information at time s, $t = 1, 2, ..., s$. A sequence of likelihood ratio statistics for detecting a mean shift in a Poisson process, $\mathbf{lr}_t(\mathbf{x}_s)$, can be derived from the likelihood ratio between out of control and in control stages (see more details in Section 4.2.2).

In this chapter, we define $l(\boldsymbol{x}_t; \boldsymbol{\theta}|\tau)$ as the log likelihood function of bivariate Poisson series given that a change occurs at time τ where $\boldsymbol{\theta}$ is a set of parameters, $\boldsymbol{\theta} = (\mu_1, \mu_2, \mu_3)$ for a simple BP distribution and $\boldsymbol{\theta} = (\phi_1, \phi_2, \mu_1, \mu_2, \mu_3)$ for a Poisson BINAR(1) process. In order to derive a likelihood ratio statistic, we also define the ratio

$$lr(\boldsymbol{x}_t; \boldsymbol{\theta}) = ln \frac{f(\boldsymbol{x}_t; \boldsymbol{\theta}^O)}{f(\boldsymbol{x}_t; \boldsymbol{\theta}^I)} \tag{6.30}$$

where $f(\boldsymbol{x}_t; \boldsymbol{\theta}^I)$ and $f(\boldsymbol{x}_t; \boldsymbol{\theta}^O)$ is the probability function (pdf) of in control (I) and out of control (O) stages, respectively, and a sequence

$$\mathbf{lr}_t(\mathbf{x}_s) = \{lr(\boldsymbol{x}_t), lr(\boldsymbol{x}_{t+1}), ..., lr(\boldsymbol{x}_s)\}, \ s = 1, 2, ... \tag{6.31}$$

Similar to the SR methods proposed in Chapter 4, firstly assume that data are independent (CWS $= 0$). The sequence of likelihood ratio statistics for monitoring a mean shift in BP process can be derived from

$$\begin{aligned} l(\boldsymbol{x}_t; \boldsymbol{\theta}|\tau) &= \sum_{t=1}^{\tau-1} ln\, f(\boldsymbol{x}_t; \boldsymbol{\theta}^I) + \sum_{t=\tau}^{s} ln\, f(\boldsymbol{x}_t; \boldsymbol{\theta}^O) \\ &= \sum_{t=1}^{s} ln\, f(\boldsymbol{x}_t; \boldsymbol{\theta}^I) + \sum_{t=\tau}^{s} ln\, \frac{f(\boldsymbol{x}_t; \boldsymbol{\theta}^O)}{f(\boldsymbol{x}_t; \boldsymbol{\theta}^I)} \\ &= h(\mathbf{x}_s) + \sum_{t=\tau}^{s} lr(\boldsymbol{x}_t; \boldsymbol{\theta}^O - \boldsymbol{\theta}^I) \\ &= h(\mathbf{x}_s) + k(\mathbf{lr}_\tau(\mathbf{x}_s; \boldsymbol{\theta}^O - \boldsymbol{\theta}^I)) \end{aligned} \tag{6.32}$$

where h and k are two real functions and $\mathbf{x}_s = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_s)$ is the information available at time s.

By the factorization theorem, it was proved that a sequence $\mathbf{lr}_t(\mathbf{x}_s; \boldsymbol{\theta}^O - \boldsymbol{\theta}^I))$ is sufficient to monitor a shift in parameters, $(\boldsymbol{\theta}^O - \boldsymbol{\theta}^I))$, in a bivariate process (Wessman, 1998; Frisén *et al.*, 2011). A sequence $\mathbf{lr}_t(\mathbf{x}_s; \boldsymbol{\theta}^O - \boldsymbol{\theta}^I))$ also can be written in a form of likelihood ratio statistics

$$\mathbf{lr}_t(\mathbf{x}_s) = \{T_t(\boldsymbol{x}_t), ..., T_s(\boldsymbol{x}_s)\},\ t = 1, 2, ..., s \tag{6.33}$$

where $T_t(\boldsymbol{x}_t)$ is likelihood ratio statistic derived from the ratio, $lr(\boldsymbol{x}_t; \boldsymbol{\theta})$, at time t (see more detail in Section 4.2.2).

In this chapter SR methods for Poisson data are proposed for four different cases listed below.

- Case 1: no CWS or CBS
- Case 2: no CWS but CBS
- Case 3: CWS but no CBS
- Case 4: CWS and CBS

The distribution of X_1 and X_2 defined for in control and out of control stages and the derivation of SR method for each case are given in each section separately.

6.4.1 Case 1 (no CWS or CBS)

In this case, we assume that observations in each series are independent (CWS = 0) and the two series are independent. Since CBS = 0, a mean shift of size c occurs in X_1 series only at time τ. X_1 and X_2 follow Poisson distributions with parameters μ_1 and μ_2 defined for in control and out of control stages as below.

$$\left.\begin{array}{l} X_{1,t} \sim P(\mu_1^I) \\ X_{2,t} \sim P(\mu_2^I) \end{array}\right\} \quad t < \tau \quad \text{in control stage}$$

$$\left.\begin{array}{l} X_{1,t} \sim P(\mu_1^O) \\ X_{2,t} \sim P(\mu_2^I) \end{array}\right\} \quad t \geq \tau \quad \text{out control stage}$$

where μ_i^I and μ_i^O are the parameters of in control and out of control stages of series i and $\mu_1^O = \mu_1^I + c$. The mean and variance of in control and out of control stage are defined in equations (6.2) and (6.18), respectively, while the joint pdf of the BP distribution, of in control stage, $f(\boldsymbol{x}_t; \mu_1^I, \mu_2^I)$, and out of control stage, $f(\boldsymbol{x}_t; \mu_1^O, \mu_2^I)$, with corresponding parameters can be written in the form of equation (6.1).

According to the principle of SR methods defined in equation (6.32), a sequence of

likelihood ratio statistics can be derived from

$$\begin{aligned} l(\boldsymbol{x}_t; \mu_1, \mu_2|\tau) &= \sum_{t=1}^{\tau-1} ln\, f(\boldsymbol{x}_t; \mu_1^I, \mu_2^I) + \sum_{t=\tau}^{s} ln\, f(\boldsymbol{x}_t; \mu_1^O, \mu_2^I) \\ &= \sum_{t=1}^{s} ln\, f(\boldsymbol{x}_t; \mu_1^I, \mu_2^I) + \sum_{t=\tau}^{s} ln\, \frac{f(\boldsymbol{x}_t; \mu_1^O, \mu_2^I)}{f(\boldsymbol{x}_t; \mu_1^I, \mu_2^I)} \\ &= h(\mathbf{x}_s) + \sum_{t=\tau}^{s} lr(\boldsymbol{x}_t; \mu_1^O - \mu_1^I) \\ &= h(\mathbf{x}_s) + k(\mathbf{lr}_\tau(\mathbf{x}_s; \mu_1^O - \mu_1^I)) \end{aligned} \tag{6.34}$$

Let $T_t(\boldsymbol{x}_t)$ be a likelihood ratio statistic for detecting mean shift of size c in a process at time t. From equation (6.34), by the factorization theorem, a sequence $\mathbf{lr}_t(\mathbf{x}_s)$ which is sufficient for monitoring a mean shift in BP process when CWS = 0 and CBS = 0 is

$$\mathbf{lr}_1(\mathbf{x}_s; \mu_1^O - \mu_1^I) = \{T_1(\boldsymbol{x}_1), ..., T_s(\boldsymbol{x}_s)\}$$

and

$$T_t(\boldsymbol{x}_t) = x_{1,t}\, ln\left(\frac{\mu_1^O}{\mu_1^I}\right) \tag{6.35}$$

where $\mu_1^O = \mu_1^I + c$ and $t = 1, 2, ..., s$. Since there is no correlation between series, a shift occurs only in the X_1 series, and the likelihood ratio statistics are based on X_1 only. The derivation for this case is provided in Section C.1 in Appendix C.

6.4.2 Case 2 (no CWS but CBS)

In this case we still assume that observations in each series are independent (CWS = 0), but X_1 and X_2 are correlated with the covariance μ_3. Due to no CWS, X_1 and X_2 follow a BP distribution with parameters, μ_1, μ_2, μ_3. A mean shift of size c in both series is reflected from a change in covariance $(\mu_3^O - \mu_3^I)$. The distributions of in control and out of control stages are defined below.

$$X_{1,t}, X_{2,t} \sim \begin{cases} BP(\mu_1^I, \mu_2^I, \mu_3^I) & t < \tau \quad \text{in control stage} \\ BP(\mu_1^I, \mu_2^I, \mu_3^O) & t \geq \tau \quad \text{out of control stage} \end{cases}$$

where μ_3^I and μ_3^O are the covariances of in control and out of control stages and $\mu_3^O = \mu_3^I + c$. Means and variances of in control and out of control stages can be calculated from equations (6.3) and (6.20), respectively. The pdf of the BP distribution, of in control stage, $f(\boldsymbol{x}_t; \mu_1^I, \mu_2^I, \mu_3^I)$, and out of control stage, $f(\boldsymbol{x}_t; \mu_1^I, \mu_2^I, \mu_3^O)$, with corresponding parameters can be written in the form set out in equation (6.4).

In this case, we aim to detect mean shifts in both series which shift from $\mu_i + \mu_3^I$ to $\mu_i + \mu_3^O$, $i = 1, 2$. From equation (6.32), a sequence of likelihood ratio statistics for the

shift can be derived from

$$\begin{aligned} l(\boldsymbol{x}_t;\mu_1,\mu_2,\mu_3|\tau) &= \sum_{t=1}^{\tau-1} ln\, f(\boldsymbol{x}_t;\mu_1^I,\mu_2^I,\mu_3^I) + \sum_{t=\tau}^{s} ln\, f(\boldsymbol{x}_t;\mu_1^I,\mu_2^I,\mu_3^O) \\ &= \sum_{t=1}^{s} ln\, f(\boldsymbol{x}_t;\mu_1^I,\mu_2^I,\mu_3^I) + \sum_{t=\tau}^{s} ln\, \frac{f(\boldsymbol{x}_t;\mu_1^I,\mu_2^I,\mu_3^O)}{f(\boldsymbol{x}_t;\mu_1^I,\mu_2^I,\mu_3^I)} \\ &= h(\mathbf{x}_s) + \sum_{t=\tau}^{s} lr(\boldsymbol{x}_t;\mu_3^O-\mu_3^I) \\ &= h(\mathbf{x}_s) + k(\mathbf{lr}_\tau(\mathbf{x}_s;\mu_3^O-\mu_3^I)) \end{aligned} \tag{6.36}$$

By the factorization theorem, a sequence of statistics, $T_t(\boldsymbol{x}_t)$, which is sufficient for monitoring a mean shift in BP series when CBS $\neq 0$, is

$$\mathbf{lr}_1(\mathbf{x}_s;\mu_3^O-\mu_3^I) = \{T_1(\boldsymbol{x}_1),...,T_s(\boldsymbol{x}_s)\}$$

and

$$T_t(\boldsymbol{x}_t) = ln\left(\frac{\sum_{j=0}^{m}\binom{x_{1,t}}{j}\binom{x_{2,t}}{j} j!\left(\frac{\mu_3^O}{\mu_1^I\mu_2^I}\right)^j}{\sum_{j=0}^{m}\binom{x_{1,t}}{j}\binom{x_{2,t}}{j} j!\left(\frac{\mu_3^I}{\mu_1^I\mu_2^I}\right)^j}\right) \tag{6.37}$$

where $\mu_3^O = \mu_3^I + c$ and $m = min(x_{1,t}, x_{2,t})$. The likelihood ratio statistic is derived according to the shift in covariance $(\mu_3^O - \mu_3^I)$ which is reflected in mean shifts in both series.

Even though X_1 and X_2 can be theoretically generated from a BP distribution according to the trivariate reduction method, it might be difficult to know the combination of μ_1, μ_2 and μ_3 used to construct the X_1 and X_2 series in practice. Based on the trivariate reduction method, these three parameters can be estimated from the mean levels and the covariance between series during the in control stage. Let $\hat{\mu}_j$ be the estimate of parameter μ_j, $j = 1, 2, 3$, and $\hat{\mu}_{X_i}^I$ be the estimated in control mean level of series i. The estimated parameters can be calculated from

$$\hat{\mu}_3^I = \widehat{Cov}(X_1, X_2)$$

$$\hat{\mu}_i^I = \hat{\mu}_{X,i}^I - \hat{\mu}_3^I$$

Thus, the SR calculation for case 2 is based on the estimated parameters calculated as above. The derivation of likelihood ratio statistics in this case is provided in Section C.2 in Appendix C.

6.4.3 Case 3 (CWS but no CBS)

In this case we assume that observations within each series are autocorrelated, but X_1 and X_2 are independent. X_1 and X_2 are two independent Poisson INAR(1) processes with a Poisson marginal distribution defined in Section 6.3.1. Since CBS = 0, we aim to detect a mean shift in X_1 only, while the mean level of X_2 is the same. Let $R_{1,t}$ and $R_{2,t}$ be mutually independent random variables from Poisson distributions with parameters

μ_1 and μ_2. In this case we investigate a mean shift from a shift in parameter of $R_{1,t}$ in X_1 series. The distributions of in control and out of control stages are define below.

$$\left.\begin{array}{c} X_{1,t} \sim Bin(X_{1,t-1}, \phi_1) \\ X_{2,t} \sim Bin(X_{2,t-1}, \phi_2) \\ R_{1,t} \sim P(\mu_1^I) \\ R_{2,t} \sim P(\mu_2^I) \end{array}\right\} \quad t < \tau \quad \text{in control stage}$$

$$\left.\begin{array}{c} X_{1,t} \sim Bin(X_{1,t-1}, \phi_1) \\ X_{2,t} \sim Bin(X_{2,t-1}, \phi_2) \\ R_{1,t} \sim P(\mu_1^O) \\ R_{2,t} \sim P(\mu_2^I) \end{array}\right\} \quad t \geq \tau \quad \text{out control stage}$$

where $\mu_1^O = \mu_1^I + c$.

In this case we aim to detect a mean shift of size c in X_1 only. The mean and variance and the conditional mean and conditional variance of the in control stage are defined in equations (6.11) and (6.12), respectively, while those defined for the out of control stage are in equations (6.21) and (6.23), respectively.

The conditional pdf of X_1 and X_2 of in control and out of control stages with the corresponding parameters can be written in the form defined in equation (6.10). Due to CWS, a sequence of likelihood ratio statistics can be derived from the log likelihood function of pdf $(t = 1)$ and the conditional pdf $(t = 2, ..., s)$ of BP series between out of control and in control stages, which is

$$\begin{aligned} l(\boldsymbol{x}_t; \phi_1, \phi_2, \mu_1, \mu_2 | \tau) &= ln\, f(\boldsymbol{x}_1; \mu_1^I, \mu_2^I) + \sum_{t=2}^{\tau-1} ln\, f(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}; \phi_1, \phi_2, \mu_1^I, \mu_2^I) \\ &\quad + \sum_{t=\tau}^{s} ln\, f(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}; \phi_1, \phi_2, \mu_1^O, \mu_2^I) \\ &= ln\, f(\boldsymbol{x}_1; \mu_1^I, \mu_2^I) + \sum_{t=2}^{s} ln\, f(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}; \phi_1, \phi_2, \mu_1^I, \mu_2^I) \\ &\quad + \sum_{t=\tau}^{s} ln\, \frac{f(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}; \mu_1^O, \mu_2^I)}{f(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}; \mu_1^I, \mu_2^I)} \\ &= h(\mathbf{x}_s) + \sum_{t=\tau}^{s} lr(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}; \mu_1^O - \mu_1^I) \\ &= h(\mathbf{x}_s) + k(\mathbf{lr}_\tau(\mathbf{x}_s; \mu_1^O - \mu_1^I)) \end{aligned} \tag{6.38}$$

where $f(\boldsymbol{x}_1; \mu_1^I, \mu_2^I)$ is the joint pdf of $R_{1,t}$ and $R_{2,t}$ at $t = 1$.

By the factorization theorem, from equation (6.38), a sufficient sequence $\mathbf{lr}_t(\mathbf{x}_s)$ for monitoring a mean shift in independent bivariate Poisson INAR(1) process is

$$\mathbf{lr}_1(\mathbf{x}_s; \mu_1^O - \mu_1^I) = \{T_1(\boldsymbol{x}_1), ..., T_s(\boldsymbol{x}_s)\}$$

and

$$T_t(\boldsymbol{x}_t) = \frac{\sum_{j=0}^{g_1} \begin{pmatrix} x_{1,t-1} \\ j \end{pmatrix} \phi_1^j (1-\phi_1)^{x_{1,t-1}-j} \cdot \frac{(\mu_1^O)^{x_{1,t}-j}}{(x_{1,t}-j)!}}{\sum_{j=0}^{g_1} \begin{pmatrix} x_{1,t-1} \\ j \end{pmatrix} \phi_1^j (1-\phi_1)^{x_{1,t-1}-j} \cdot \frac{(\mu_1^I)^{x_{1,t}-j}}{(x_{1,t}-j)!}} \tag{6.39}$$

where $\mu_1^O = \mu_1^I + c$ and $t = 1, 2, ..., s$. The derivation of likelihood ratio statistics for this case is provided in Section C.3 in Appendix C.

6.4.4 Case 4 (CWS and CBS)

In this case, we allow for both CWS and CBS, therefore, X_1 and X_2 follow a Poisson BINAR(1) process defined in Section 6.3.2. Let $R_{1,t}$ and $R_{2,t}$ be random variables from a BP distribution with parameters μ_1, μ_2, μ_3, where $\mu_3 \neq 0$. We aim to detect mean shifts occurring in both series reflected from a shift of size c in covariance between $R_{1,t}$ and $R_{2,t}$. Since the Poisson BINAR(1) process is the convolution of two binomial distributions and the BP distribution, the distribution of in control and out of control stages are defined below.

$$\left.\begin{array}{c} X_{1,t} \sim Bin(X_{1,t-1}, \phi_1) \\ X_{2,t} \sim Bin(X_{2,t-1}, \phi_2) \\ R_{1,t}, R_{2,t} \sim BP(\mu_1^I, \mu_2^I, \mu_3^I) \end{array}\right\} \quad t < \tau \quad \text{in control stage}$$

$$\left.\begin{array}{c} X_{1,t} \sim Bin(X_{1,t-1}, \phi_1) \\ X_{2,t} \sim Bin(X_{2,t-1}, \phi_2) \\ R_{1,t}, R_{2,t} \sim BP(\mu_1^I, \mu_2^I, \mu_3^O) \end{array}\right\} \quad t \geq \tau \quad \text{out control stage}$$

where $\mu_3^O = \mu_3^I + c$. The mean and variance and the conditional mean and conditional variance of the in control stages are defined in equations (6.14) and (6.15), respectively, while those of the out of control stage are defined in equations (6.27) and (6.28), respectively.

The conditional pdf of the Poisson BINAR(1) process with the defined parameters can be written as in equation (6.17). A sequence of likelihood ratio statistics derived from the log likelihood function of pdf ($t = 1$) and the conditional pdf ($t = 2, ..., s$) of the

BINAR(1) process is

$$\begin{aligned} l(\boldsymbol{x}_t;\phi_1,\phi_2,\mu_1,\mu_2,\mu_3|\tau) &= ln\, f(\boldsymbol{x}_1:\mu_1^I,\mu_2^I,\mu_3^I)+\sum_{t=2}^{\tau-1} ln\, f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1};\phi_1,\phi_2,\mu_1^I,\mu_2^I,\mu_3^I) \\ &\quad +\sum_{t=\tau}^{s} ln\, f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1};\phi_1,\phi_2,\mu_1^I,\mu_2^I,\mu_3^O) \\ &= ln\, f(\boldsymbol{x}_1;\mu_1^I,\mu_2^I,\mu_3^I)+\sum_{t=2}^{s} ln\, f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1};\phi_1,\phi_2,\mu_1^I,\mu_2^I,\mu_3^I) \\ &\quad +\sum_{t=\tau}^{s} ln\, \frac{f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1};\phi_1,\phi_2,\mu_1^I,\mu_2^I,\mu_3^O)}{f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1};\phi_1,\phi_2,\mu_1^I,\mu_2^I,\mu_3^I)} \\ &= h(\mathbf{x}_s)+\sum_{t=\tau}^{s} lr(\boldsymbol{x}_t|\boldsymbol{x}_{t-1};\mu_3^O-\mu_3^I) \\ &= h(\mathbf{x}_s)+k(\mathbf{lr}_\tau(\mathbf{x}_s;\mu_3^O-\mu_3^I)) \end{aligned} \tag{6.40}$$

where $f(\boldsymbol{x}_1;\mu_1^I,\mu_2^I,\mu_3^I)$ is the joint pdf of the BP distribution between $R_{1,t}$ and $R_{2,t}$ at $t=1$. By the factorization theorem, from equation (6.40), a sufficient sequence $\mathbf{lr}_t(\mathbf{x}_s)$ for monitoring a mean shift in Poisson BINAR(1) process is

$$\mathbf{lr}_1(\mathbf{x}_s;\mu_3^O-\mu_3^I)=\{T_1(\boldsymbol{x}_1),...,T_s(\boldsymbol{x}_s)\}$$

and

$$T_t(\boldsymbol{x}_t)=ln\,\frac{v_t}{w_t}, t=1,2,...,s \tag{6.41}$$

where

$$v_t=\sum_{r_1=0}^{g_1}\sum_{r_2=0}^{g_2}\left\{\binom{x_{1,t-1}}{x_{1,t}-r_1}\phi_1^{x_{1,t}-r_1}(1-\phi_1)^{x_{1,t-1}-x_{1,t}-r_1}\binom{x_{2,t-1}}{x_{2,t}-r_2}\cdot \right.$$
$$\left.\phi_2^{x_{2,t}-r_2}(1-\phi_2)^{x_{2,t-1}-x_{2,t}-r_2}\ \frac{(\mu_1^I)^{r_1}}{r_1!}\frac{(\mu_2^I)^{r_2}}{r_2!}\sum_{j=0}^{m}\binom{r_1}{j}\binom{r_2}{j}j!\left(\frac{\mu_3^O}{\mu_1^I\mu_2^I}\right)^j\right\}$$

, $\mu_3^O=\mu_3^I+c$ and

$$w_t=\sum_{r_1=0}^{g_1}\sum_{r_2=0}^{g_2}\left\{\binom{x_{1,t-1}}{x_{1,t}-r_1}\phi_1^{x_{1,t}-r_1}(1-\phi_1)^{x_{1,t-1}-x_{1,t}-r_1}\binom{x_{2,t-1}}{x_{2,t}-r_2}\cdot \right.$$
$$\left.\phi_2^{x_{2,t}-r_2}(1-\phi_2)^{x_{2,t-1}-x_{2,t}-r_2}\ \frac{(\mu_1^I)^{r_1}}{r_1!}\frac{(\mu_2^I)^{r_2}}{r_2!}\sum_{j=0}^{m}\binom{r_1}{j}\binom{r_2}{j}j!\left(\frac{\mu_3^I}{\mu_1^I\mu_2^I}\right)^j\right\}$$

As mentioned in Section 6.4.2, the parameters of the BP distribution (μ_1,μ_2,μ_3) might be difficult to know or define in practice. However, once we have the estimated covariance

$\widehat{Cov}(X_1, X_2)$ and autoregressive coefficients $(\hat{\phi}_i)$, the estimated parameters, $\hat{\mu}_1^I, \hat{\mu}_2^I, \hat{\mu}_3^I$ for the in control stage, can be calculated from

$$\hat{\mu}_3^I = \widehat{Cov}(X_1, X_2)(1 - \hat{\phi}_1^I \hat{\phi}_2^I)$$
$$\hat{\mu}_i^I = \hat{\mu}_{X_i}^I (1 - \hat{\phi}_i^I) - \hat{\mu}_3^I$$

Thus, the SR calculation for case 4 is based on the estimated parameters calculated as above. The derivation of likelihood ratio statistics for this case is provided in Section C.4 in Appendix C.

6.5 Modified one-sided EWMA chart for Poisson data

A simple Shewhart c chart, proposed for detecting nonconforming products in manufacturing industry, is one of the existing charts used for monitoring a shift in a process of Poisson data. However, a c chart is not very effective for detecting a small shift, since a decision as to whether a process is out of control is made on the basis of a current observation only. To detect a small positive mean shift in a Poisson process, we aim to use a one-sided EWMA chart for Poisson data because it is more sensitive (Gan, 1990; Borror *et al.*, 1998; Testik *et al.*, 2006). The development of Poisson EWMA charts for independent and dependent observations exists in the literature, with some limitations. For example, some of them are proposed for detecting both upward and downward shifts (i.e. two-sided charts) (Borror *et al.*, 1998; Testik *et al.*, 2006; Weiß, 2009). In this section, we aim to modify one-sided EWMA charts for monitoring a mean shift in a process of independent Poisson data (CWS = 0) and of Poisson INAR(1) data (CWS $\neq$ 0). Details of the modification for each chart are provided in Sections 6.5.1 and 6.5.2 separately.

6.5.1 Modified one-sided EWMA chart for independent observations

The EWMA chart for Poisson data was introduced by Gan (1990) by rounding EWMA statistics so that they are positive integers. Later Borror *et al.* (1998) proposed a two-sided Poisson EWMA chart which proceeded without losing any information from rounding EWMA statistics and had smaller ARL compared to Shewhart's c chart and Gan's chart. Hu *et al.* (2011) compared four modified one-sided EWMA charts for detecting a linear drift in a Poisson mean. The four methods were a standard EWMA chart with only an upper control limit, resetting EWMA statistics below the target mean by the target mean, resetting observations below the target mean by the target mean and the adaptive EWMA chart. However, none of these perform uniformly better than others for various drift sizes (Hu *et al.*, 2011) and we are not primarily concerned here with a mean drift scenario, but rather a step change.

In this study we extend the idea of Borror *et al.* (1998) to develop a one-sided EWMA chart for independent Poisson data. Let X_t be a random variable from Poisson distribution with parameter μ^I and λ be a constant smoothing parameter required in EWMA charts. The in control mean is denoted by μ^I, while μ^O is the out of control mean of the process and we assume, since it is the practically interesting case, that $\mu^O > \mu^I$.

One-sided EWMA statistics for detecting a positive shift in a process can calculated from

$$z_t = max(\mu^I, \lambda x_t + (1-\lambda) z_{t-1}),\ t > 0 \tag{6.42}$$

where $z_0 = \mu^I$. The mean and the exact variance of Z_t are

$$E(Z_t) = \mu^I$$

$$V(Z_t) = \frac{\lambda}{2-\lambda}(1-(1-\lambda)^{2t})\mu^I \tag{6.43}$$

For large t, $t \to \infty$, from equation (6.43) the asymptotic variance of Z_t is

$$V(Z_t) = \frac{\lambda \mu^I}{2-\lambda} \tag{6.44}$$

In this study we construct one-sided EWMA chart based on this asymptotic variance. Thus, the upper control limit (UCL) can be calculated from

$$UCL = \mu^I + L\sqrt{\frac{\lambda \mu^I}{2-\lambda}} \tag{6.45}$$

where L is the width of the control limits for a specified λ and a desired ARL_0.

The width L for a specific λ and ARL_0 can be obtained by ARL calculation using the Markov chain approach (Brook and Evans, 1972; Lucas and Saccucci, 1990; Borror *et al.*, 1998). Let (LCL, UCL) be an interval divided to K subintervals, where the j^{th} subinterval, (LCL_j, UCL_j), can be defined as

$$LCL_j = LCL + \frac{(j-1)(UCL-LCL)}{K}$$

$$UCL_j = LCL + \frac{j(UCL-LCL)}{K}$$

and m_i is the midpoint of the i^{th} subinterval

$$m_i = LCL + \frac{(2j-1)(UCL-LCL)}{2K}$$

A $(K+1)^{st}$ interval is an absorbing state representing an out of control region above UCL, where a process is stopped when points fall outside the UCL. Thus, the ARL is the expected time to absorption of the Markov chain (Borror *et al.*, 1998).

The run length distribution of an EWMA chart is determined by its initial probability vector and transition probability matrix (Lucas and Saccucci, 1990). Let $\boldsymbol{p}$ be a $K \times 1$ initial probability vector, where p_j represents the probability that z_t starts in state j. For a upper or one-sided EWMA chart, $\boldsymbol{p}$ usually has 1 as the first entry and 0s at every other entry so that the probabilities sum to one (Brook and Evans, 1972; Lucas and Saccucci, 1990; Shu *et al.*, 2007).

Let $\mathbf{Q}$ be a $K \times K$ transition probability matrix among the in control stages where q_{ij} is the probability of moving from state i to state j in one step defined as

$$q_{ij} = \begin{cases} P(Z_t < LCL_j),\ j = 1 \\ P(LCL_j < Z_t < UCL_j | z_{t-1} = m_i),\ j > 1 \end{cases}$$

Thus,

$$q_{ij} = \begin{cases} P(X_t < LCL + \frac{UCL-LCL}{2K\lambda}(-(1-\lambda)(2i-1))) & j = 1 \\ P(X_t < LCL + \frac{UCL-LCL}{2K\lambda}(2j-(1-\lambda)(2i-1))) & \\ \quad -P(X_t < LCL + \frac{UCL-LCL}{2K\lambda}(2(j-1)-(1-\lambda)(2i-1))) & j > 1 \end{cases} \tag{6.46}$$

(Borror *et al.*, 1998).

From Lucas and Saccucci (1990), the ARL_0 for one-sided Poisson EWMA chart can be calculated from

$$\mathrm{ARL}_0 = \boldsymbol{p}'(\mathbf{I} - \mathbf{Q})^{-1}\mathbf{1} \tag{6.47}$$

Having specified $\lambda = 0.3$ (see Section 4.2.5), the in control mean (μ^I) and number of subintervals (K) and, the ARL's calculation using the Markov chain approach defined above can be summarized as follow.

step 1 specify initial value of L, $LCL = \mu^I$ and calculate UCL from equation (6.45)

step 2 calculate the transition probability matrix, $\mathbf{Q}$ from equation (6.46)

step 3 calculate ARL_0 from equation (6.47)

According to the ARL calculation defined above, the limit L is chosen to achieve a desired in control average run length, $\mathrm{ARL}_0 = 370$ or $\mathrm{ARL}_0 = 741$, adjusted for multiplicity in parallel methods. For $\lambda = 0.3$, $L = 3.093$ for $\mathrm{ARL}_0 = 370$ and $L = 3.374$ for $\mathrm{ARL}_0 = 741$, for an in control mean of 5. Limits (L) for various values of λ are given in Table C.1 in Appendix C.

6.5.2 Modified one-sided EWMA chart for dependent observations

While control charts for monitoring for mean shift in autocorrelated process of normal data exist in the literature (Schmid, 1997; Lu and Reynolds, 1999; Shiau and Hsu, 2005; Montgomery, 2009), those for autocorrelated process of Poisson data are rare (Weiß, 2009, 2011). Weiß (2009) proposed a two-sided EWMA chart for Poisson INAR(1) process. This chart is called the combined EWMA chart since it combined the use of a Shewhart c chart and the Poisson EWMA chart proposed by Gan (1990). This chart suffers limitations from the choices of six design parameters (from two charts) which need to be defined with care and sometime appears to be overparameterized (Weiß, 2009). Later (Weiß, 2011) proposed a one-sided EWMA chart for the Poisson INAR(1) process using only three design parameters, however, it suffers from rounding EWMA statistics as in Gan (1990)'s approach, resulting in oversmoothed data. He then proposed

a one-sided s EWMA chart by using different types of rounding operators. It shows an improvement and a good sensitivity for detecting a large shift; however, our interest might well be in small shifts, so we pursue our own method.

To modify a one-sided EWMA chart for Poisson INAR(1) process with positive autocorrelation, we use the same approach as set out in Section 5.2. Let X_t be a random variable from a Poisson INAR(1) process, with marginal Poisson distribution with parameter μ, $R_t \sim P(\mu)$, and ϕ an autoregressive coefficient of degree 1. The in control mean is denoted by μ^I, while μ^O is the out of control mean. One-sided EWMA statistics can calculated from

$$z_t = \max(\mu^I, \lambda x_t + (1-\lambda) z_{t-1}),\ t > 0. \tag{6.48}$$

The in control mean and the variance of Z_t are

$$E(Z_t) = \mu^I = \frac{\mu}{1-\phi}$$

$$V(Z_t) = \frac{\mu}{1-\phi}\left(\frac{\lambda}{2-\lambda}\right)\frac{(1+\phi(1-\lambda))}{(1-\phi(1-\lambda))}$$

where ϕ is an autoregressive coefficient, $0 < \phi < 1$ (Wieringa, 1999; Shiau and Hsu, 2005). Thus, the upper control limit (UCL) can be calculated from

$$UCL = \mu^I + L\sqrt{\frac{\mu}{1-\phi}\left(\frac{\lambda}{2-\lambda}\right)\frac{(1+\phi(1-\lambda))}{(1-\phi(1-\lambda))}} \tag{6.49}$$

where L be the width of the control limits for a specified λ and a desired ARL_0.

The calculation for ARL_0 for a specific L and ϕ is carried by using Monte Carlo simulation defined in Section 5.2. Let N be a number of iterations, μ^I and ϕ be parameters of Poisson INAR(1) process and RL_0 be an in control run length. The ARL's calculation are defined as follows

step 1 specify initial values of L and λ and calculate UCL from equation (6.49)

step 2 set $t = 1$

step 3 generate x_t which follow Poisson INAR(1) process with parameters μ^I and ϕ and calculate EWMA statistics from equation (6.48)

step 4 if $z_t > UCL$, process is out of control and $\mathrm{RL}_0 = t$; otherwise, set $t = t+1$ and go back to step 3

step 5 repeat steps 2-4 N times $(N = 100,000)$

step 6 average RL_0 from N iterations to get the in control average run length (ARL_0)

Having specified $\lambda = 0.3$ (see Section 4.2.5) and defined a desired ARL_0, the limits L for different values of ϕ can be obtained by using the same procedure defined above. The limits L used for one-sided EWMA chart for Poisson INAR(1) process for $\mu = 5$ and $\mathrm{ARL}_0 = 370, 741$ are summarized in Table 6.1.

Table 6.1: The limits (L) for the modified one-sided EWMA charts ($\text{ARL}_0 = 370, 741$) for Poisson INAR(1) process for $\mu = 5$

ARL_0	λ	Autoregressive coefficient (ϕ)								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
370	0.3	3.104	3.062	3.013	2.949	2.875	2.782	2.669	2.516	2.270
741	0.3	3.458	3.357	3.349	3.286	3.161	3.075	2.923	2.836	2.595

6.6 Simulation study

The performance of detecting a mean shift in a Poisson process is evaluated from a simulation study by assuming that

- bivariate Poisson series are generated according to the distribution and parameters defined in each section
- simultaneous change at time τ is considered
- mean shifts of sizes 2 and 3 are investigated by adding them to the BP series at time τ and afterwards
- if two series are independent, a shift occurs in X_1 series only (i.e. model residuals, case 1 and case 3), otherwise shifts occur in both series due to CBS (i.e. case 2 and case 4) (see Section 6.3.3)
- a one-sided EWMA chart is chosen according to the nature of data or derived statistics to be monitored. For example, in the parallel method, since the original Poisson data are monitored separately in parallel, a one-sided EWMA chart modified for independent and dependent Poisson data (Sections 6.5.1 and 6.5.2) is used with $\lambda = 0.3$ and $\text{ARL}_0 = 741$. For Frisén, Wessman and our methods, one-sided EWMA charts for independent or dependent normal data (Sections 4.2.5 and 5.2), with $\lambda = 0.3$ and $\text{ARL}_0 = 370$, are used instead, since the likelihood ratio statistics from these methods look more normal than Poisson (see results in Section 6.6.2).
- the detection performances are evaluated by comparing the conditional expected delay (CED), true alarm rate (TAR), false alarm rate (FAR) and non-detection rate (NDR) between the three methods. Details of each measurement are summarized in Section 3.4).

The detection performance between the methods are investigated by comparing the differences of CED, TAR, FAR and NDR between the methods estimated from the simulations against the critical values defined in Section 3.6. Section 6.6.1 provides the results of using residuals from fitting Poisson regression models with SR methods proposed for normal data, while the results of four cases of our proposed SR methods for Poisson data are provided in Section 6.6.2. In each section, examples of the EWMA charts are illustrated. The vertical dotted line indicates the change point of the process

and the horizontal dashed line is the upper control limit. The comparison of the detection performance between the methods is illustrated in bar charts, where P, F, W and Ck stand for Parallel, Frisén, Wessman and our proposed method Case k, respectively. The full results from the simulation study are given in Tables C.2 - C.7 in Appendix C.

6.6.1 Model residuals

In this section, we aim to show that residuals from Poisson regression model might be used with SR methods proposed for normal data. A simple example is illustrated with two independent Poisson series, X_1 and X_2, defined as follows.

$$\left.\begin{array}{l} X_{1,t} \sim P(\mu_1^I) \\ X_{2,t} \sim P(\mu_2^I) \end{array}\right\} \quad t < \tau \quad \text{in control stage}$$

$$\left.\begin{array}{l} X_{1,t} \sim P(\mu_1^O) \\ X_{2,t} \sim P(\mu_2^I) \end{array}\right\} \quad t \geq \tau \quad \text{out control stage}$$

where $\mu_i^I = 5$ and $\mu_1^O = \mu_1^I + c$.

Two Poisson regression models, M1 and M2, are fitted as defined in Section 6.2. Residuals from each model are checked as whether they are approximately independent and normal distributed. Diagnostic plots for M1 and M2 are shown in Figures C.1 and C.2, respectively, in Appendix C. It can be seen that residuals from M1 model are approximately normally distributed, while X_1 residuals from M2 model are far from normal, resulting in slow detection.

Residuals from the two models are used with SR methods proposed for normal data and monitored with standard one-sided EWMA charts for independent normal data. The example of detection charts of using residuals from M1 and M2 models are shown in Figures 6.1 and 6.2, respectively. The performance of detecting mean shifts of sizes 2 and 3 between three methods, parallel adjusted for multiplicity, Frisén and Wessman methods, is illustrated in Figure 6.3, while the full results from the simulation are given in Table C.2 in Appendix C.

Overall, Frisén and Wessman methods perform significantly better than the parallel method, as expected. Using residuals from M1 models gives significantly shorter $\widehat{\text{ced}}$, though with lower $\widehat{\text{tar}}$ and higher $\widehat{\text{ndr}}$, than for the M2 model. Refitting the model at each time point in M2 model is computationally time consuming and also slows the detection because parameters at each time point are re-estimated over time. Since in control and out of control parameters are not estimated separately as in the M1 model, refitting and re-estimating parameters at each time point smooths a step change between in control and out of control stages which results in the underestimation of the parameter in the out of control stage.

This is an example of monitoring BP series by using model residuals assumed normally distributed with SR methods proposed for independent normal data. Further cases of detecting a mean shift in BP series when CWS $\neq 0$ and/or CBS $\neq 0$ can be investigated

in a similar manner to that set out in Section 6.2 (i.e. with modified EWMA charts if necessary).

6.6.2 Actual Poisson data

In this section, actual bivariate Poisson data are used with our SR methods proposed for detecting a mean shift in Poisson data. Our proposed methods are compared with the Frisén method and the parallel method adjusted for multiplicity. We consider the Frisén method because it was proposed for detecting a mean shift in any exponential family distribution, including the Poisson distribution, while the Wessman method illustrated in this study was only explicitly derived for detecting a change in a normal mean (see more details in Chapter 4).

Several types of one sided EWMA charts are used depending on the distribution of data or the derived likelihood ratio statistics by the SR methods. It can be seen later that even though the actual data are Poisson distributed, the derived sequence of likelihood ratio statistics from Frisén and our proposed methods look more normal than Poisson. Therefore, the statistics from Frisén and our methods are monitored with a standard one-sided EWMA chart for normal data, while the actual data in the parallel method are monitored with one-sided EWMA chart for Poisson data. More details of the charts are given in each section separately.

6.6.2.1 Case 1: no CWS or CBS

Assume that CWS = 0 and CBS = 0, so that two independent Poisson series, X_1 and X_2, as defined in Section 6.6.1 are considered. X_1 and X_2 are generated from Poisson distribution with parameter $\mu_i = 5$. Since CBS = 0, the data considered for the Frisén and our methods in this case are from the X_1 series only. The diagnostic plots of the raw data and likelihood ratio statistics from Frisén and our methods are shown in Figure C.3 in Appendix C. From Figure C.3, parallel data are monitored with one-sided EWMA chart modified for independent Poisson data (Section 6.5.1), while the statistics from Frisén and our methods, which look more normal than Poisson, are monitored with a one-sided EWMA chart for normal data (Section 4.2.5). An example of detection plot for detecting a mean shift of size 2 is illustrated in Figure 6.4. The performance in detecting mean shifts of sizes 2 and 3 of the three methods are illustrated in Figure 6.5, while the original results from the simulation are summarized in Table C.3 in Appendix C.

As expected, Frisén and our methods perform equally well and significantly differently from the parallel method. A large shift size is more likely to be detected with shorter $\widehat{\text{ced}}$, higher $\widehat{\text{tar}}$ and lower $\widehat{\text{ndr}}$. Compared with the results from Section 6.6.1, the three methods show a significant improvement in detection with shorter delay and lower $\widehat{\text{far}}$, but with lower $\widehat{\text{tar}}$ and higher $\widehat{\text{ndr}}$ than using model residuals from M1 and M2 models for all shift sizes. However, in prospective surveillance, early detection with low false alarm rate is considered much more important. Thus this case shows that even though residuals from fitting Poisson regression model might be used as a simple alternative way to detect a mean shift in BP series, it delays the detection in Section 6.6.1.

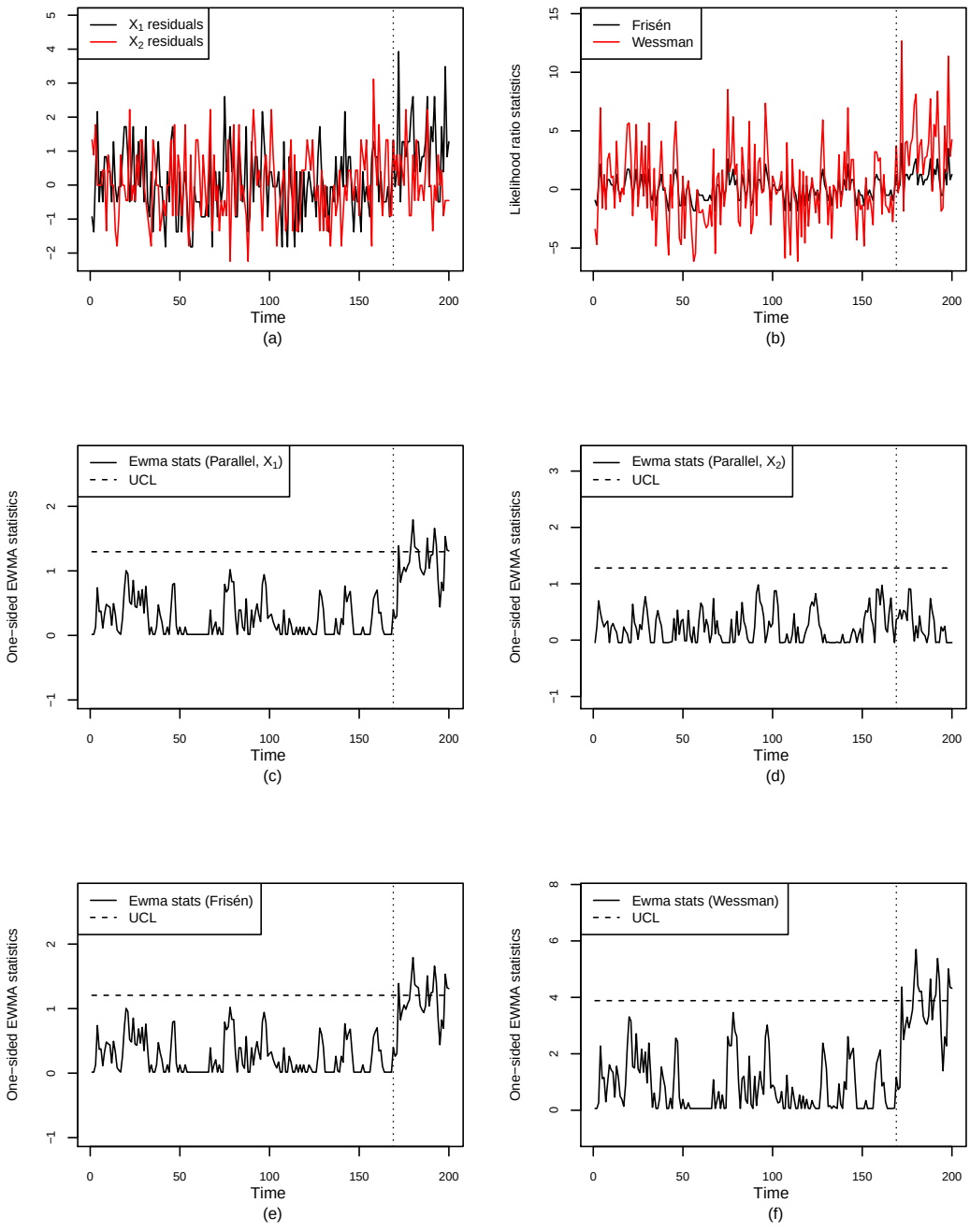

Figure 6.1: An example of detection performance for the M1 model (shift size 2, CWS = 0 and CBS = 0): (a) residual plot, (b) plot of likelihood ratio statistics from Frisén and Wessman methods, (c) - (d) one-sided EWMA charts of parallel method and (e) - (f) one-sided EWMA charts for Frisén and Wessman methods.

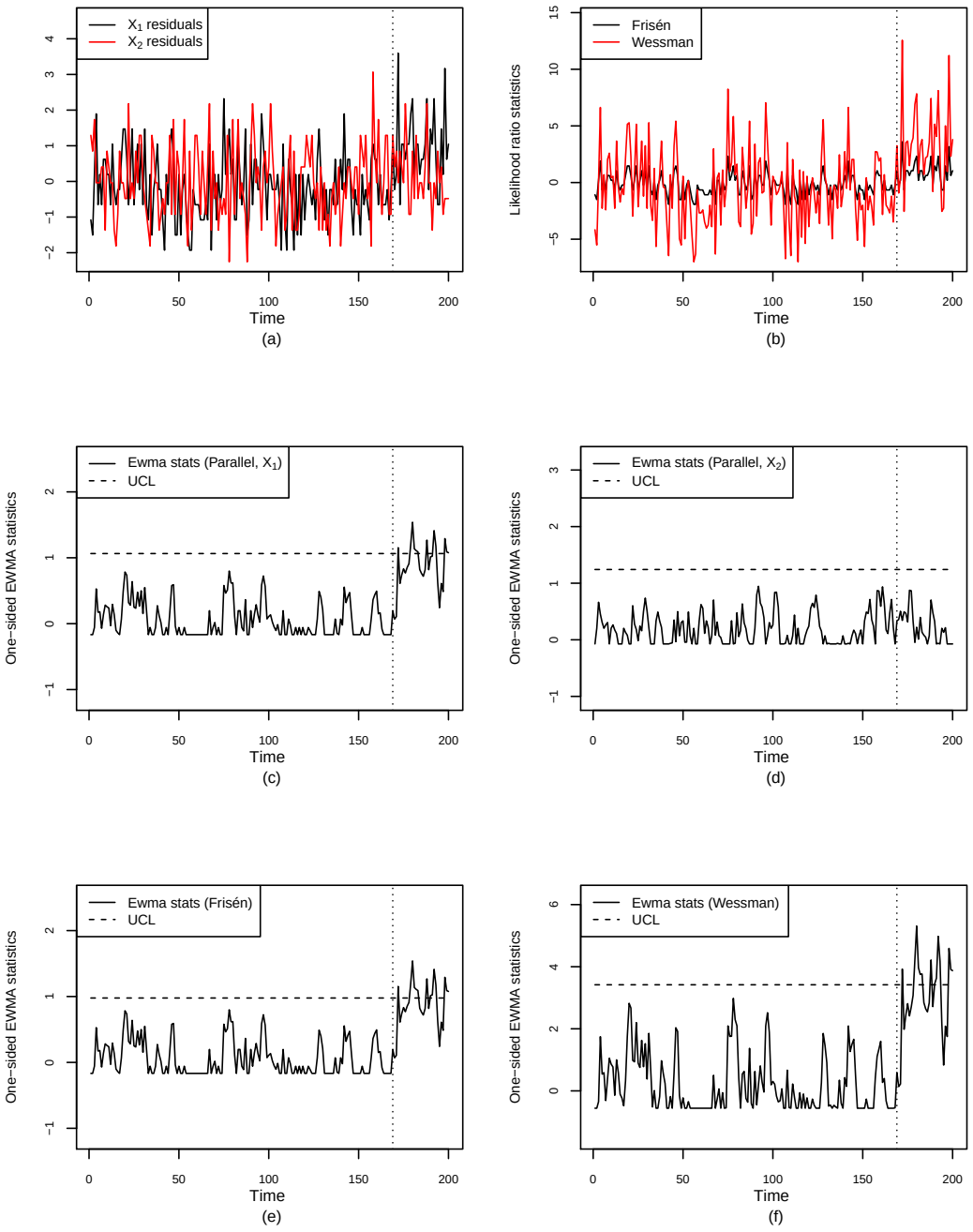

Figure 6.2: An example of detection performance for the M2 model (shift size 2, CWS = 0 and CBS = 0): (a) residual plot, (b) plot of likelihood ratio statistics from Frisén and Wessman methods, (c) - (d) one-sided EWMA charts of parallel method and (e) - (f) one-sided EWMA charts for Frisén and Wessman methods.

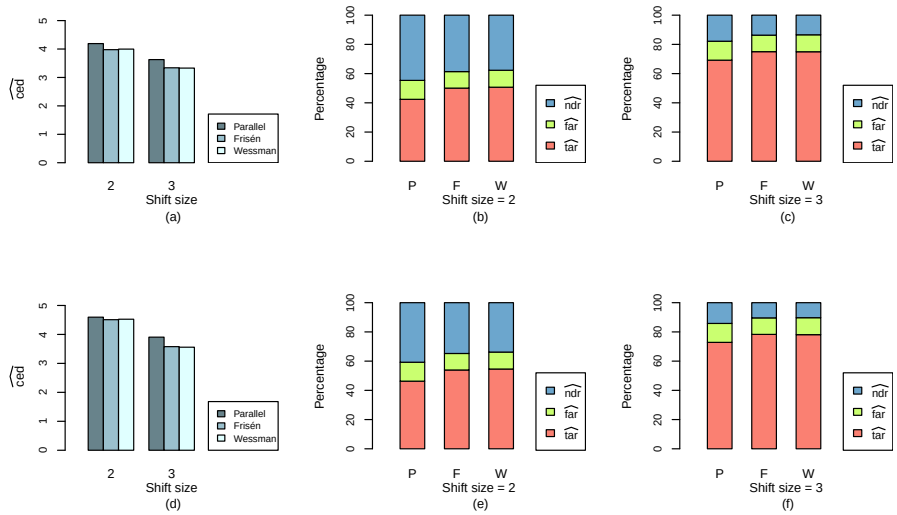

Figure 6.3: A bar chart comparing the performance in detecting the mean shift of sizes 2 and 3 in the process: (a) - (c) M1 model and (d) - (f) M2 model.

6.6.2.2 Case 2: no CWS but CBS

Having assumed that the two series are correlated with each other, X_1 and X_2 are generated from a BP distribution with three parameters, μ_1, μ_2 and μ_3. Three different sets of parameters considered in this case illustrate differing levels of dependency between series. The mean level and CBS in each set for the in control stage are summarized in Table 6.2. Diagnostic plots of actual data and likelihood ratio statistics of Frisén and our methods are shown in Figure C.4 in Appendix C. The one-sided EWMA charts defined for each method in case 1 can also be applied in this case. The example of the detection plots of the three methods is shown in Figure 6.6, while the detection performance in detecting a mean shift of sizes 2 and 3 of the three methods are illustrated in Figure 6.7. The full results from the simulation are summarized in Table C.4 in Appendix C.

Table 6.2: Set of parameters of BP distribution for case 2

Parameters (μ_1, μ_2, μ_3)	Mean level $(E(X_i))$	CWS	CBS
4,4,1	5	0	0.2
3,3,2	5	0	0.4
2,2,3	5	0	0.6

Unsurprisingly, a larger shift size is detected more quickly than smaller shift size. Our method performs significantly differently from the parallel and Frisén methods by show-

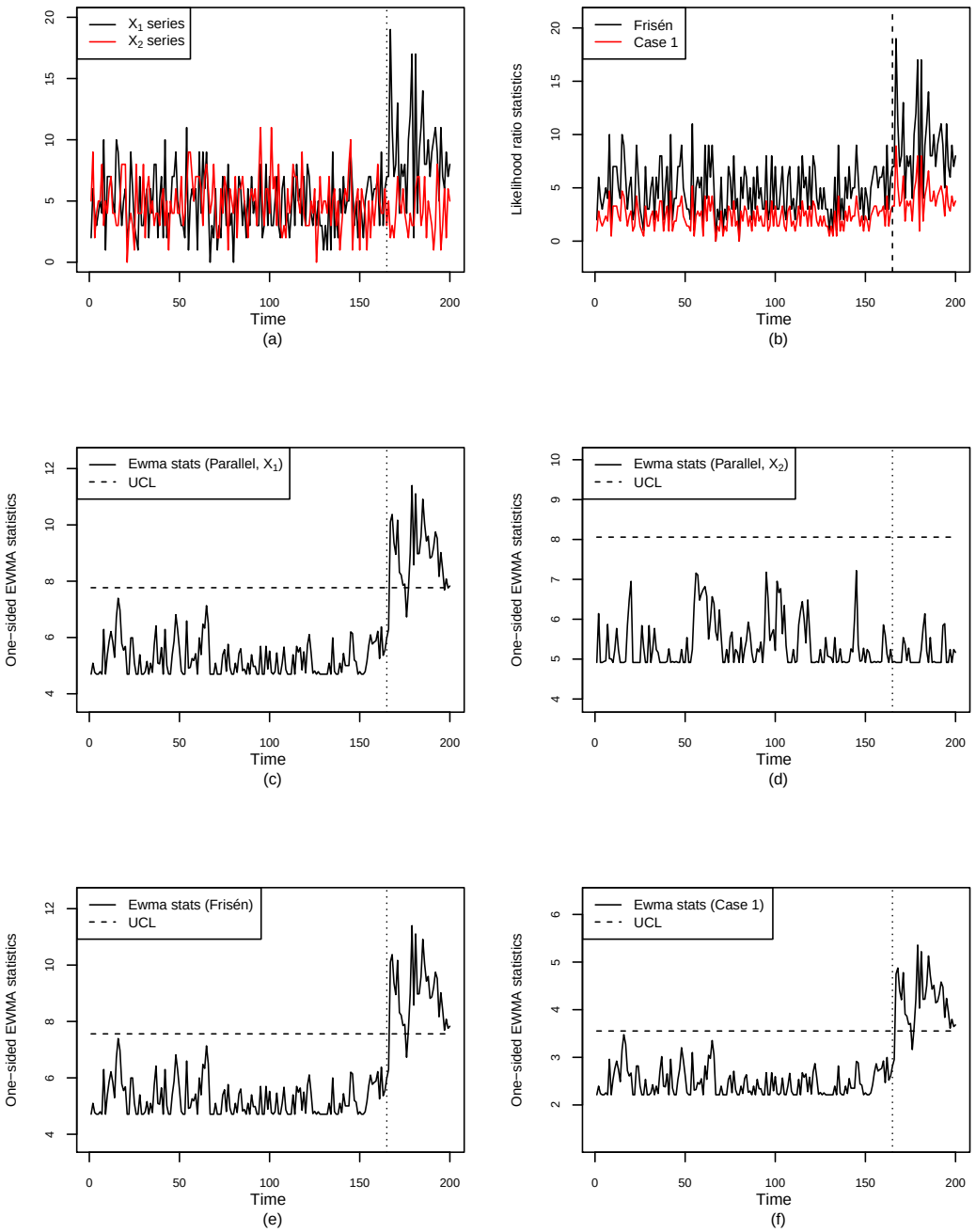

Figure 6.4: An example of detection performance for case 1 (shift size 2, CWS = 0 and CBS = 0): (a) data plot, (b) plot of likelihood ratio statistics from Frisén and our methods, (c) - (d) one-sided EWMA charts for parallel method and (e) - (f) one-sided EWMA chart for Frisén and our methods.

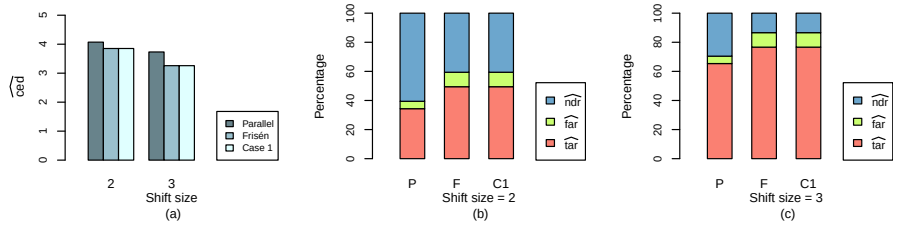

Figure 6.5: A bar chart comparing the performance in detecting the mean shifts of sizes 2 and 3 of case 1 (CWS = 0 and CBS = 0).

ing shorter $\widehat{\text{ced}}$ and lower $\widehat{\text{ndr}}$ for all shift sizes. Since CBS is not incorporated in the Frisén and parallel methods, high $\widehat{\text{ndr}}$s are produced here when CBS increases. Overall, from Figure 6.7, even though our proposed method gives higher $\widehat{\text{far}}$, the CED and TAR of our method is significantly better for detecting a small shift size for all levels of CBS and a large shift size when CBS is high.

Compared to the results of the normal case (scenario 1 in Chapter 4, when CBS $\neq 0$ and CWS $= 0$), all three methods perform worse in detecting a shift in the Poisson process, giving lower $\widehat{\text{tar}}$ and higher $\widehat{\text{ndr}}$. However, this might be due to the nature of Poisson data, where the mean and variance are not separable. A (positive) shift is unlikely to be detected because of the larger variation resulting from the mean shift in the process. The larger variance after the change point can be seen from the data plot and the plot of likelihood ratio statistics (Figures 6.6 (a) and (b)).

6.6.2.3 Case 3: CWS but no CBS

In this case, we assume that observations in each series are dependent over time. Three different levels of CWS are investigated (0.2, 0.4 and 0.6). Two types of EWMA charts, one-sided EWMA charts for independent and dependent data are used as detection tools in this case. This is because, we note the results in Chapter 4, where parallel and Frisén methods, which do not take CWS into account, are found to have a high false alarm rate if data are autocorrelated. Thus, we aim to use both charts to show how detection performance improves when the proper control chart, accounting for autocorrelation, is used.

Here we assume that the X_1 and X_2 series are independent (CBS $= 0$) and are from the Poisson INAR(1) model with parameter μ_i and ϕ_i defined in Section 6.4.3. The in control parameters of the model in each series and its mean level are summarized in Table 6.3.

Since X_1 and X_2 are from the Poisson INAR(1) model, the parallel method is used with modified one-sided EWMA charts for independent and dependent Poisson data (Sections 6.5.1 and 6.5.2). Likelihood ratio statistics derived from Frisén and our method (case

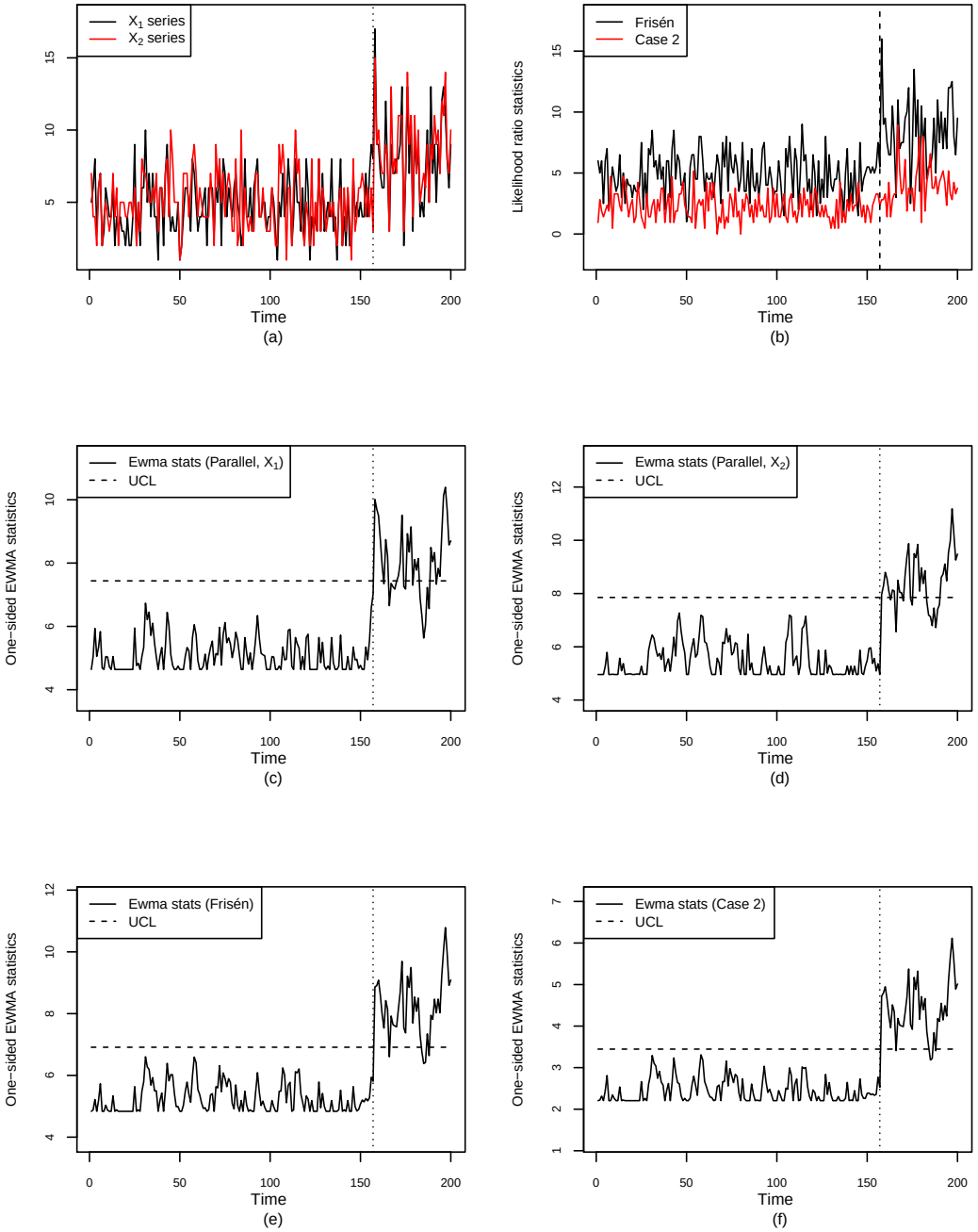

Figure 6.6: An example of detection performance of case 2 (shift size 3, CWS = 0 and CBS = 0.6): (a) data plot, (b) plot of likelihood ratio statistics from Frisén and our methods, (c) - (d) one-sided EWMA charts for parallel method and (e) - (f) one-sided EWMA chart for Frisén and our methods.

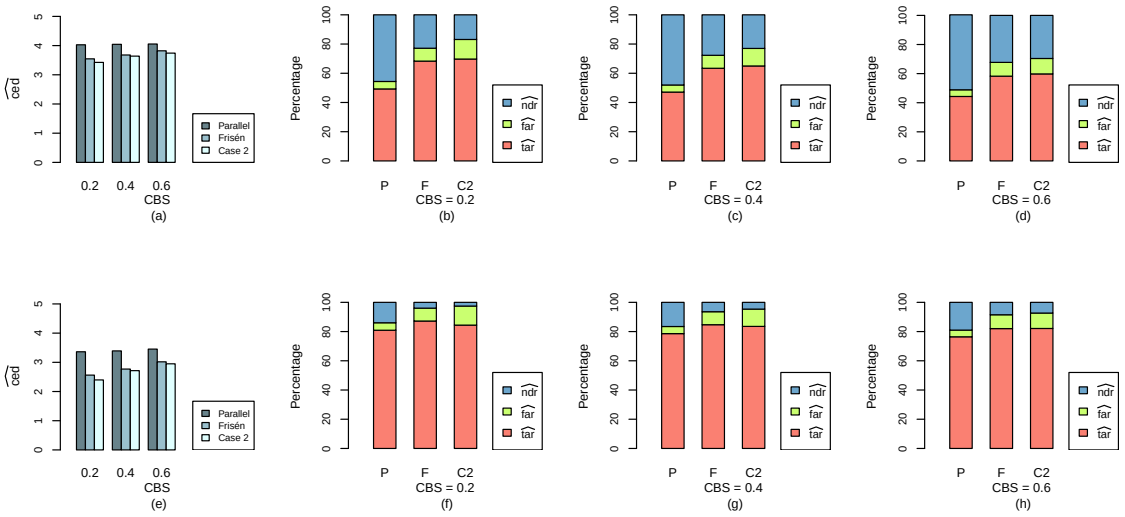

Figure 6.7: A bar chart comparing the detection performance of case 2 (CWS = 0 and CBS $\neq$ 0): (a) - (d) shift size = 2 and (e) - (h) shift size = 3.

Table 6.3: Parameters of Poisson INAR(1) model for case 3

Parameter		CBS	Mean level
μ_i	ϕ_i		$(E(X_i))$
5	0.2	0	6.25
5	0.4	0	8.33
5	0.6	0	12.5

3), which look more normally than Poisson distributed (Figure C.5 in Appendix C), are monitored with a standard one-sided EWMA chart for independent normal data (Section 4.2.5) and a modified one-sided EWMA chart for dependent normal data (Section 5.2). An example of detecting a mean shift of size 3 is shown in Figure 6.8. The upper control limits of the modified charts for the parallel and Frisén methods are higher than those of the standard charts due to the adjustment for the autocorrelation, resulting in slow detection. The detection performances of the three methods for detecting mean shift of sizes 2 and 3 are illustrated in Figures 6.9 and 6.10, respectively. The full results from the simulation are given in Tables C.5 and C.6, respectively, in Appendix C. Note that, in the tables, standard and modified for Frisén and our method represent standard and modified charts for independent and dependent normal data, while those for parallel method represent modified charts for independent and dependent Poisson data, respectively.

Overall, the CED of our method performs is significantly different from the CED of

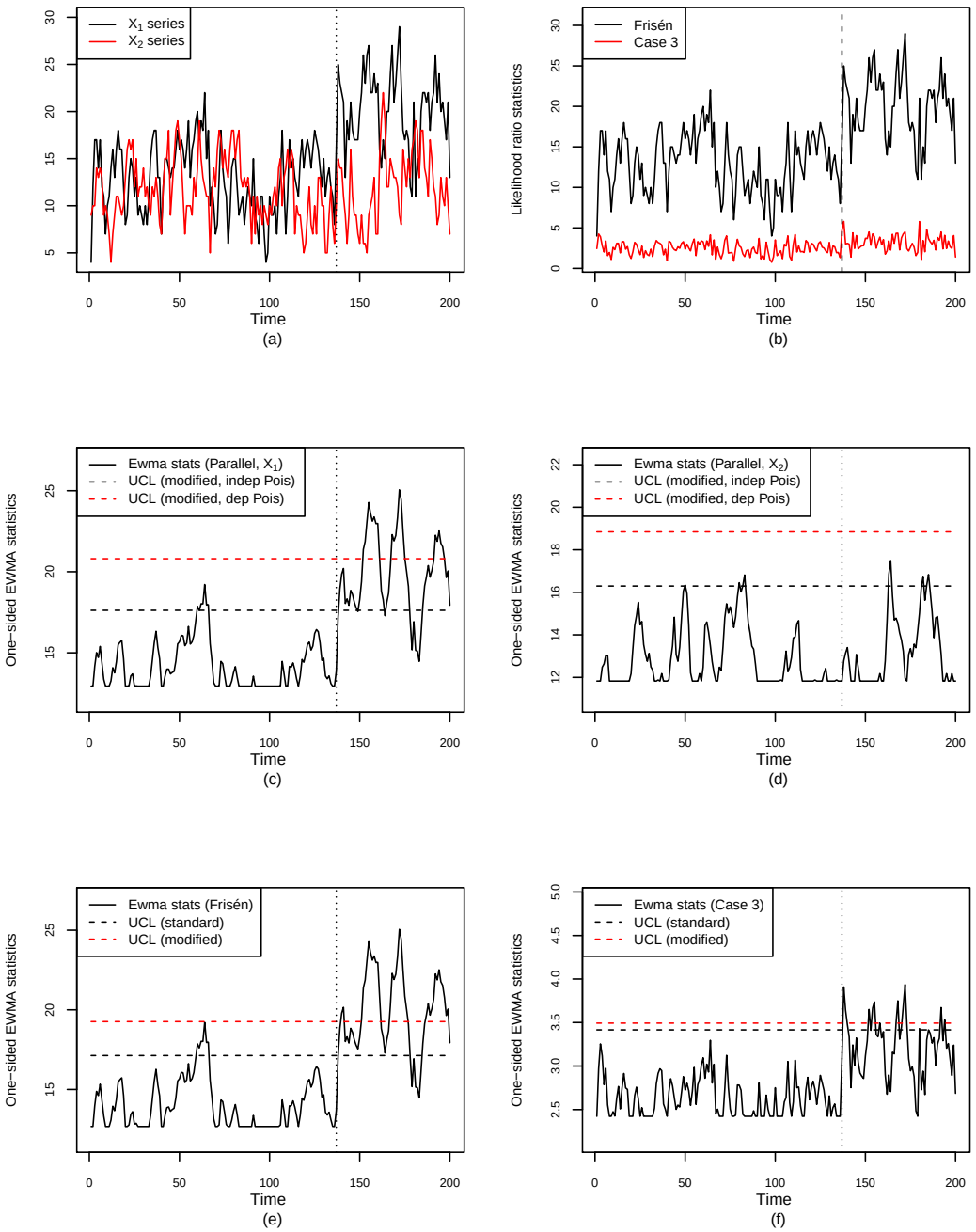

Figure 6.8: An example of detection performance of case 3 (shift size 3, CWS = 0.6 and CBS = 0): (a) data plot, (b) plot of likelihood ratio statistics from Frisén and our methods, (c) - (d) one-sided EWMA charts for parallel method and (e) - (f) one-sided EWMA chart for Frisén and our methods.

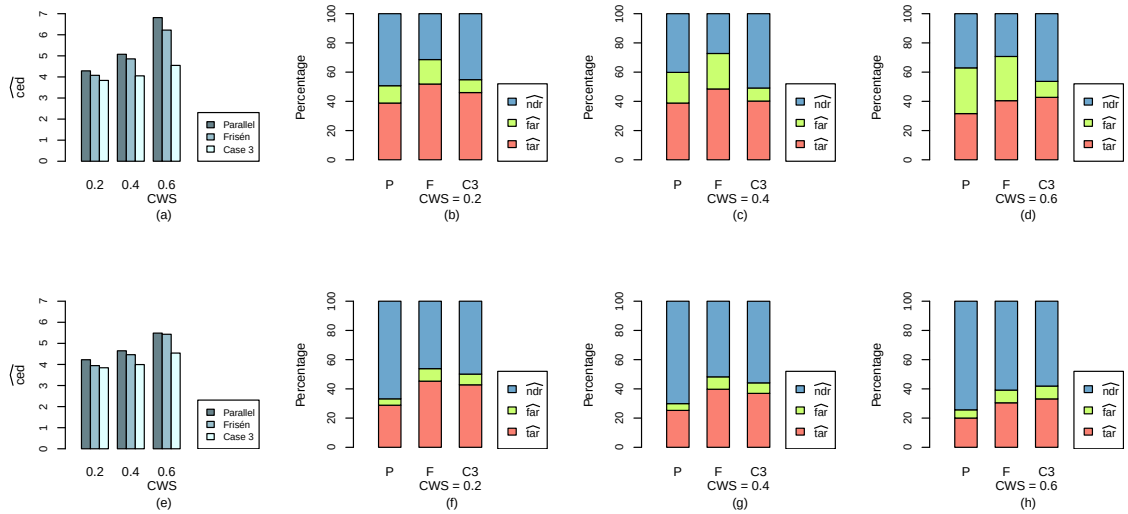

Figure 6.9: A bar chart comparing the detection performance of case 3 (CWS $\neq$ 0, CBS = 0 and shift size = 2): (a) - (c) the results from standard one-sided EWMA charts and (d) - (f) the results from modified one-sided EWMA charts.

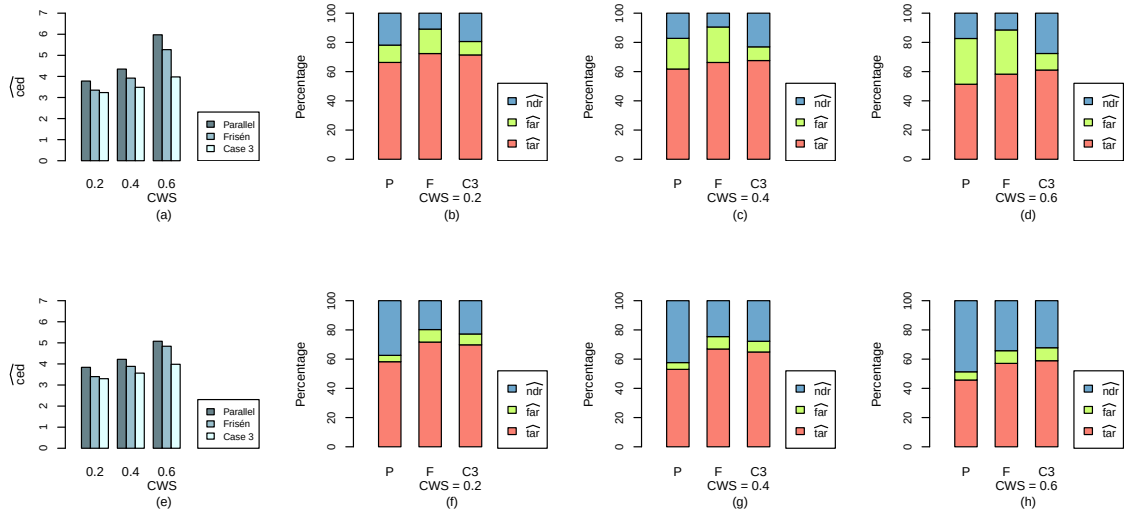

Figure 6.10: A bar chart comparing the detection performance of case 3 (CWS $\neq$ 0, CBS = 0 and shift size = 3): (a) - (c) the results from standard one-sided EWMA charts and (d) - (f) the results from modified one-sided EWMA charts.

other methods. Our method gives shorter delays for all levels of CWS and shift sizes. As expected, parallel and Frisén methods give high false alarm when standard EWMA charts for independent data are used. Irrespective of whether the standard or modified chart is used, the CED of our method is significantly different from the others. Even though our method give a quicker detection (shorter $\widehat{\text{ced}}$), the $\widehat{\text{tar}}$, $\widehat{\text{far}}$ and $\widehat{\text{ndr}}$ are not uniformly better than other methods. For example, our method performs significantly differently from other methods by giving higher $\widehat{\text{tar}}$ and lower $\widehat{\text{ndr}}$ for all shift sizes when CWS is high (Figures 6.9 (d and f) and 6.10 (d and f)). The plots of the autocorrelation function of actual Poisson data (X_1 and X_2) and likelihood ratio statistics from Frisén and our methods (Figure C.7 in Appendix C) show that our method clearly removes CWS, while the effect of CWS are still present in parallel data and the statistics from Frisén methods. The $\widehat{\text{far}}$ of parallel and Frisén methods are substantially improved when modified EWMA charts for dependent data are used, while those of our method are slightly improved since CWS has been removed in the sufficient reduction. The parallel method gives the lowest $\widehat{\text{far}}$, however, overall it performs significantly differently from other methods by showing the longest $\widehat{\text{ced}}$, lowest $\widehat{\text{tar}}$ and highest $\widehat{\text{ndr}}$ due to high upper control limits adjusted for multiplicity.

As mentioned previously in case 2, due to the large variation resulting from the shift in mean, the three methods perform moderate compared to their performances in the normal case (scenario 2, Chapter 4) with lower $\widehat{\text{tar}}$ and higher $\widehat{\text{ndr}}$. Due to the large variance after the mean shifted, a shift is unlikely to be detected resulting in high $\widehat{\text{ndr}}$.

6.6.2.4 Case 4: CWS and CBS

In this case we want to investigate the detection performance when X_1 and X_2 are dependent with both CWS and CBS. X_1 and X_2 follow the Poisson BINAR(1) model defined in Section 6.3.2. The parameters of the in control stage of each series are summarized in Table 6.4. Like case 3, both standard and modified one-sided EWMA charts are used the show how the FAR is improved.

Table 6.4: Set of parameters of Poisson BINAR(1) model for case 4

Parameters (μ_1, μ_2, μ_3)	Mean level $(E(X_i))$	CWS	CBS
3,3,2	12.5	$\phi = 0.6$	$\rho = 0.25$
1,1,4	12.5	$\phi = 0.6$	$\rho = 0.5$

Diagnostic plot for each method (Figure C.6 in Appendix C) shows that the parallel data are still Poisson distributed, while likelihood ratio statistics for Frisén look more normal than Poisson. Even though the statistics of our method are not quite normally distributed, with the presence of heavy tails and some outliers, a one-sided EWMA chart for normal data is still used, though with some concern over its applicability. Therefore, the standard and modified one-sided EWMA charts defined in case 3 for the three

methods are also applied in this case. An example of detecting a mean shift of size 3 in autocorrelated process are shown in Figure 6.11. The detection performances in detecting mean shift of sizes 2 and 3 are illustrated in Figures 6.12 and 6.13, respectively, while the full results from the simulation are provided in Table C.7 in Appendix C.

Since CWS is recognized and allowed for, our proposed method performs significantly differently from other methods by showing shorter $\widehat{\text{ced}}$ for all shift sizes. However, the $\widehat{\text{tar}}$, $\widehat{\text{far}}$ and $\widehat{\text{ndr}}$ are not uniformly better than the others. For example, with the standard chart, the $\widehat{\text{tar}}$ of our method is higher than other methods when shift size is large. On the other hand, with the modified chart, the $\widehat{\text{tar}}$ of our method is lower than other methods, especially when shift size is large.

Overall, even though our method gives a shorter delay for all shift sizes, it still produces high $\widehat{\text{far}}$ compared to the others even if the modified EWMA chart for autocorrelated data is used. This can be investigated from the plots of the autocorrelation functions of parallel data and likelihood ratio statistics from Frisén and our methods (Figure C.8 in Appendix C). Figure C.8 shows that our method clearly removes CWS, while the effect of CWS is still present in parallel and Frisén methods. Thus, the high $\widehat{\text{far}}$ in our methods is not because of CWS, but it might be because of the presence of outliers in normal Q-Q plot of likelihood ratio statistics derived from our method. More generally, using a one-sided EWMA chart for normal data with our methods might not be inappropriate, since the statistics are not quite normal. Thus, a proper control chart for the corresponding distribution of likelihood ratio statistics might be considered in order to improve the detection performance.

6.7 Conclusions

In this chapter we aim to detect a mean shift in a bivariate Poisson process. Two possible approaches considered are using model residuals with SR methods proposed for normal data and deriving four possible cases of SR methods for Poisson data. The detection performance of each method or case is evaluated by simulation and compared with results of parallel, Wessman and Frisén methods. Modified one-sided EWMA chart for independent and dependent Poisson data are proposed for monitoring a shift by the parallel method, while the derived statistics of Wessman, Frisén and our proposed methods are monitored with one-sided EWMA charts for normal data.

The use of methods specific to Poisson data has advantages over the use of (normally distributed) model residuals from a Poisson regression model. Using the residuals slows the detection due to correlated residuals from long step ahead forecasts and oversmoothed estimated parameters from refitting a model for prospective surveillance. Results from using actual Poisson data in the SR methods show that if there is no CBS or CWS in the process, our and the Frisén methods perform equally and better than the parallel method since the Frisén method is also proposed for detecting a shift in a any exponential family distribution, including the Poisson.

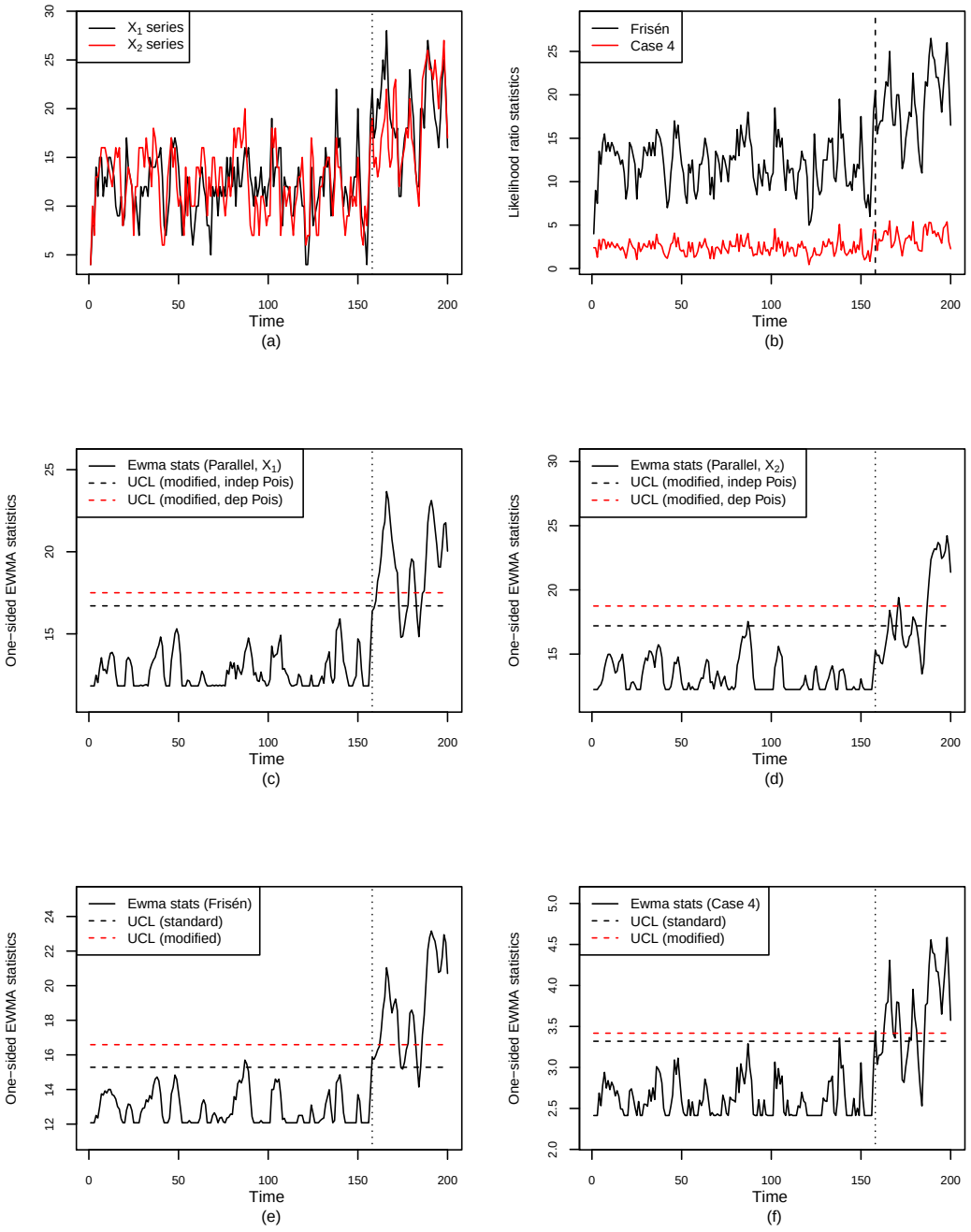

Figure 6.11: An example of detection performance of case 4 (shift size 3, CWS = 0.6 and CBS = 0.5): (a) data plot, (b) plot of likelihood ratio statistics from Frisén and our methods, (c) - (d) one-sided EWMA charts for parallel method and (d) - (f) one-sided EWMA chart for Frisén and our methods.

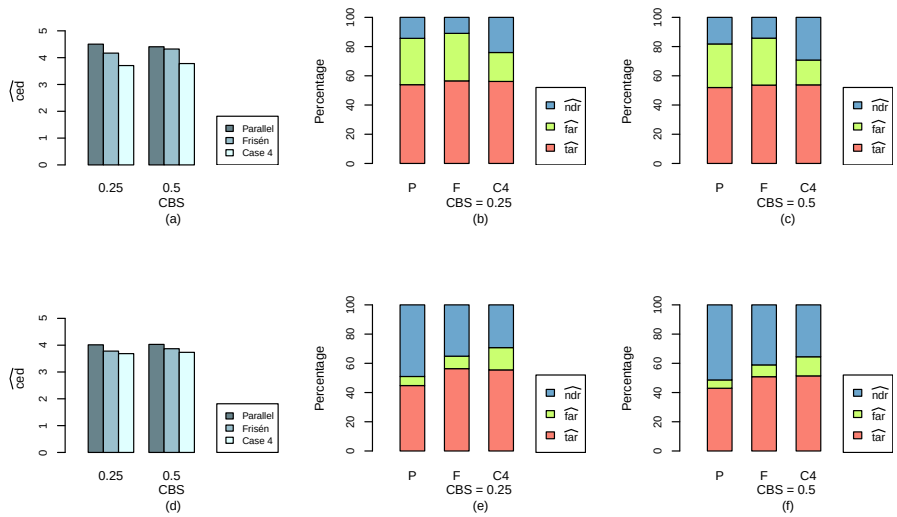

Figure 6.12: A bar chart comparing the detection performance of case 4 (CWS $\neq$ 0, CBS $\neq$ 0 and shift size = 2): (a) - (d) the results from standard one-sided EWMA charts and (e) - (h) the results from modified one-sided EWMA charts.

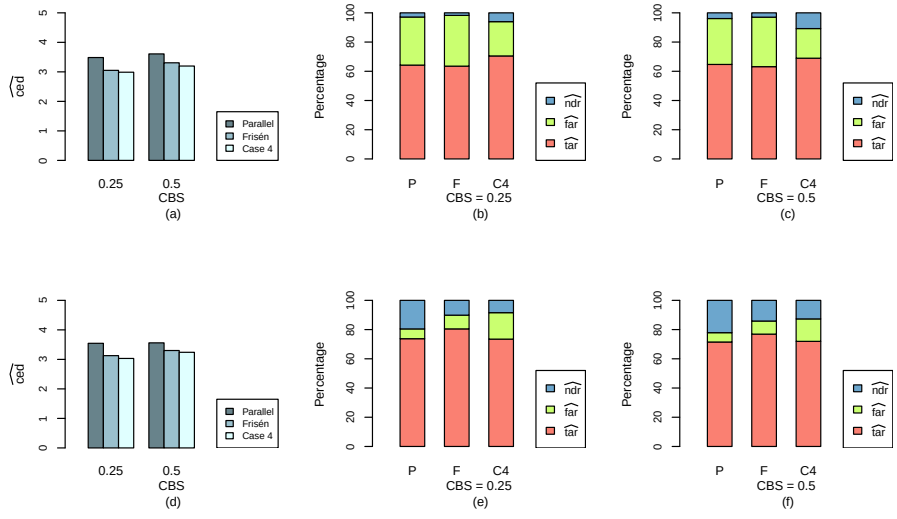

Figure 6.13: A bar chart comparing the detection performance of case 4 (CWS $\neq$ 0, CBS $\neq$ 0 and shift size = 3): (a) - (d) the results from standard one-sided EWMA charts and (e) - (h) the results from modified one-sided EWMA charts.

Overall, either if CWS or CBS or both are present, our proposed method gives shorter delay than other methods, however, the TAR, FAR and NDR of such methods are not uniformly better than those of the others. For example, if CBS only is present, our method gives shorter delay and higher TAR for detecting a small shift size for all levels of CBS and for detecting a large shift size when CBS is high.

If CWS is clearly present, either with or without CBS, our method still gives shorter delay than the others no matter whether a standard or modified chart is used. The FARs of the parallel and Frisén are high due to the effect of CWS being present in the data and derived statistics. However, they can be improved by using a modified chart for autocorrelated data. As a result, NDR is also improved, but the delays of detection are still longer than our method. While FARs of the parallel and Frisén are improved with the modified charts, the FAR of our method is only slightly better since CWS has already been removed in the sufficient reduction. So with the modified chart, the TAR, FAR and NDR of the Frisén method might be better than our method at some level of CWS or CBS, though our method still gives quicker detection (i.e. shorter delay).

Apparently, from the simulation results, the high FAR of our method occurs when CBS is present. There might be two possible reasons to consider. Firstly, our SR methods derive a sequence of likelihood ratio statistics according to a shift in a covariance which results in a mean shift in the process. So if the two series are not independent and the mean shift is due to a shift in the covariance, our methods might be more sensitive to the shift, producing FAR, than the other methods which do not incorporate CBS in the sufficient reduction. It can be observed from the diagnostic plots that the derived statistics from our method are not quite normally distributed, with the presence of some outliers which tend to produce false alarms easily. Thus, alternative modification of the one-sided EWMA chart for monitoring a mean shift in a bivariate Poisson process should be considered, with due regard to the exact distribution of likelihood ratio statistics.

Secondly, due to the equivalence of mean and variance of Poisson data, which cannot be separated, the variance of the process is shifted correspondingly to a shift in the process mean. A shift is unlikely to be detected because of the large variation (or variance) resulting from the mean shift in the process. Monitoring a mean shift in the process with a one-sided EWMA chart either for normal and Poisson data, where the control limit is calculated based on the parameters estimated in the in control stage, might not be relevant for the Poisson case, since not only the mean change is shifted, but also the variance. Thus results should be interpreted with care or other types of control charts, i.e. charts for monitoring both mean and variance, might be considered.

Chapter 7

Case studies: The implementation of SR methods

7.1 Background

Having evaluated the detection performance of SR methods via simulation study, this chapter aims to illustrate the implementation of sufficient reduction methods in practice. Three case studies from four real data sets observed for health and environmental surveillance are used in this chapter.

- Case study 1: Influenza mortality data, USA
- Case study 2: Scarlet fever notifications, UK
- Case study 3: Greek pollution data, Greece
- Case study 4: Swedish radiation and snow depth data, Sweden

The influenza mortality data in the USA are available online from CDC WONDER website (CDC WONDER, 2012), while the others were restricted data, but were made available by request for the purpose of academic research. The scarlet fever notifications were kindly provided by Dr Paul Clearly, the Health Protection Agency North West Regional Epidemiology Unit (HPA, 2012). The Greek pollution data were provided by Dr Sotiris Bersimis (Bersimis, pers. comm.). The Swedish radiation data were kindly provided by Dr Eric Järpe with authorization from the Swedish Radiation Safety Authority (SRSA, 2012), while the snow depth data were made available by the Swedish Meteorological and Hydrological Institute (SMHI, 2012).

The first two case studies provide evidence of a real shift in the data (section 7.2), while there is no evidence of a real or significant shift in the two last data sets (section 7.3). Since there is no significant shift in the latter, simulated outbreak signals are injected to both data in order to be able to evaluate the detection performance of SR methods. Also, due to the non-stationarity in the Greek and Swedish data sets, actual data cannot be directly used with SR methods since they do meet the assumptions of the SR methods. Therefore, in case studies 3 and 4, residuals from fitting time series models are

used instead of observed data. Details of data preparation for this cases are provided in section 7.3.1.

Since we do not have long historical data, for the purpose of system evaluation data in each series are divided into two parts: training and test sets. The training data are used to investigate correlation within and between series, lag between series and to assess the parameters used in the SR methods and EWMA charts, while the test data is used to test the detection performance of the methods. To evaluate this performance, an appropriate one sided EWMA chart (i.e. modified if necessary, as in sections 5.2 and 6.5.2), with $\lambda = 0.3$ and $\text{ARL}_0 = 370$ for Wessman, Frisén and our proposed methods and $\text{ARL}_0 = 741$ for the parallel method adjusted for multiplicity, is chosen according to the nature and characteristics of data in each case study.

The detection performance of the four methods is evaluated by measuring the delay before the system detects a shift once it has started (i.e. measured from the defined change point of the process). Zero delay means that the system triggers an alarm on the day, or week, that a change began (i.e. an alarm is flagged on the same day or week as the change point). A negative delay indicates a false alarm; but see further discussion below.

Details of the data, definition of training and test sets and the appropriate one-sided EWMA chart for each case study are summarized in separate sections. Section 7.2 covers case studies 1 and 2, , while case studies 3 and 4 are in section 7.3.

7.2 Actual data with real shifts

In this section, actual weekly data with real shifts from two data sets, the USA influenza and UK scarlet fever data sets, are used to illustrate SR methods. The USA influenza data provides an example of detecting a mean shift with correlation and time lag between series in a process of normal data, while the UK scarlet fever data set is an example of detecting a mean shift with no lag between series in a process of Poisson data.

In order to evaluate detection performance with actual data, we need to know the exact date of disease outbreak in order to measure how fast the system can detect a mean shift. However, in practice, it is difficult to identify the date of the outbreak (Meynard *et al.*, 2008; Burkom, 2003; Unkel *et al.*, 2012). Therefore, in order to evaluate detection performance, we choose a suitably dramatic shift to be regarded as the change point of the process, τ. The training set is defined as the non-epidemic or non-peak period up to 4 weeks (an arbitrary 'guard region') prior to the defined change point of the process in order to allow for uncertainty in subjectively defining the date of change point. The training set is thus the data observed at times $t = 1, ..., \tau - 4$, while the test set is the data observed at times $\tau - 3, ..., s$. Details of training and test sets for each data set are provided below. Note that since the detection performance of the methods are compared against each other, subjectively choosing a change point of the process is not a critical issue for the system comparison in this study.

7.2.1 Influenza mortality data, USA

The weekly influenza mortality data set collected by CDC WONDER (2012) in the USA is considered. The full data set consists of influenza mortality data reported weekly from 122 cities in the USA for 1996 - 2013. From historical data (1993 - 1998), a peak influenza period is defined as a 4-consecutive-week period during the last few weeks of one year and the first few weeks of the following year. It is reported that influenza notifications in Los Angeles and several cities in late December, 1997 and early January 1998 are higher than those in the last six years (Glaser *et al.*, 2002).

In this example, we identify the data from San Francisco and Las Vegas in 1997 - 1998, cities not far from Los Angeles which also show an increase in the mortality in late December in 1997, as a suitable pair of series to illustrate the implementation of SR methods, since the non-peak periods of the two series can be identified clearly and meet the requirements of SR methods (i.e. both are normal distributed with evidence of correlation and lag between series, see more detail in next section). The data from Los Angeles was considered but is unsuitable because it contains a lot of missing values during the non-peak period. Although long historical data are available, we use only two-years data, 1997 - 1998, since data are non-stationary with a seasonal pattern every year. Defining a non-peak period or training data from the full historical data might not be easy or practical since suitable weeks would not be consecutive. Additionally, in many applications, very limited prior data is available. Thus, it is more realistic if we use the non-peak period in 1997, as a training data, to test whether or not we can detect a shift in the last few weeks of 1997 and the first few weeks of 1998. More details are given in the next section.

7.2.1.1 Training and test sets and detection tools

The weekly influenza mortality data in San Francisco and Las Vegas in 1997 - 1998 are plotted in Figure 7.1. It is clear that there is an increase in the mortality at the beginning of each year during weeks 1 - 11 in both series, with evidence of time lag between series with the San Francisco series running slightly ahead of the Las Vegas series. Due to the dramatic shift in the second week of 1998 in San Francisco, we chose the second week of 1998 as the change point of the process. Since the change point is subjectively chosen, we exclude 4 weeks prior for the training data, which must be chosen from a non-peak period, and define

- the training data: weeks 12 - 51, 1997
- the test data: weeks 52 - 53, 1997 and weeks 1 - 53, 1998
- the change point: week 2, 1998

The plot of the defined training and test sets is shown in Figures 7.2. The vertical dashed line separates the training data and test data, while the vertical dotted line represents the supposed change point of the process.

In order to choose a proper SR method and detection tool for this data set, both series are checked as to whether or not they are independent and normally distributed with

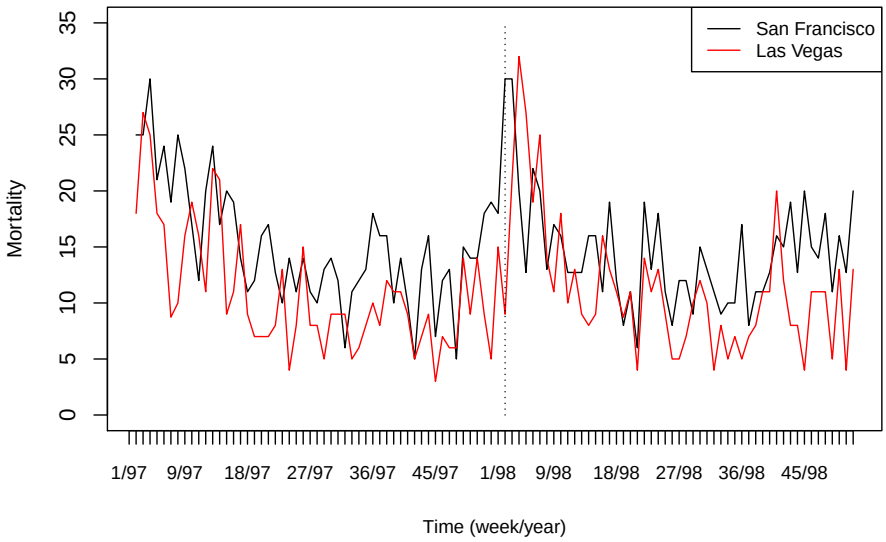

Figure 7.1: Plots of the influenza mortality data in San Francisco and Las Vegas. The vertical dotted line represents the change point of the process.

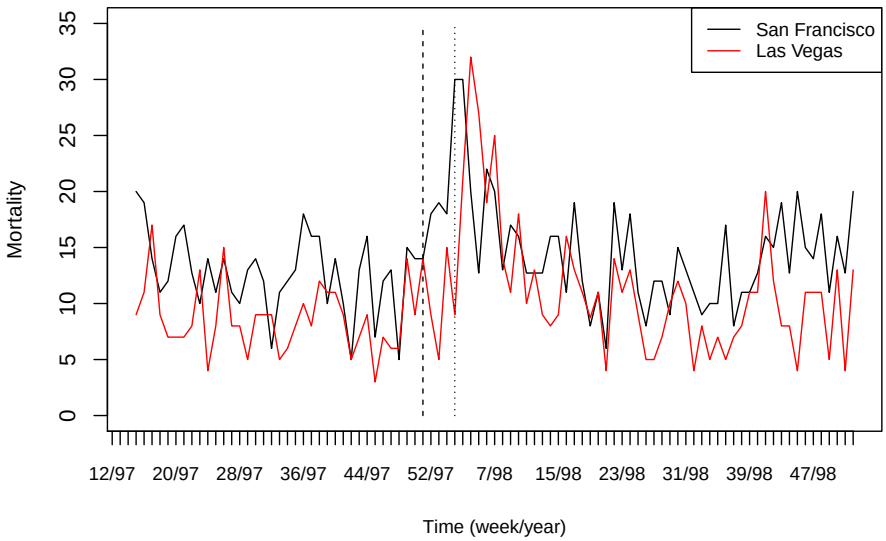

Figure 7.2: Plots of the training and test data for San Francisco and Las Vegas series.

Table 7.1: Parameter estimation for the training sets (case study 1)

Series	Mean	Standard deviation	Cross correlation (Lag = 2)
San Francisco	12.722	3.493	0.431
Las Vegas	8.729	3.105	

or without the evidence of lag between series. The check plots of the training data for San Francisco and Las Vegas series are shown in Figures D.1 and D.2 in appendix D, respectively. The lag between series is investigated from the plot of cross-correlation function in Figure D.3. From the plots, observations in each series are independent and normally distributed. The series are correlated with 2 weeks lag. The mean and standard deviation of each training set are summarized in Table 7.1

In view of our decision regarding the nature of the data, the appropriate version of our SR method is that in case 1 of chapter 4, i.e. that for detecting changes with time lag in a process of independent normal data. It is compared with other three methods: parallel adjusted for multiplicity, Frisén and Wessman methods. Since the data are independent and normally distributed, a standard one-sided EWMA chart is used.

Regarding the relation of shift sizes between series defined in section 4.3.2, if the two series are correlated, the shift size in one series reflects in the shift size in another due to the correlation between series. Let c_i be a shift size we aim to detect in series i and $\hat{\rho}$ be the estimated correlation between series. If two series are not independent ($\rho \neq 0$), the shift in the second series is determined by the shift in the first series and the CBS, i.e. $c_2 = \hat{\rho} c_1$. As shown in section 5.3.2, the SR method is fairly robust to mis-specification of the c_i. Thus, in this example we aim to detect a mean shift of size 2 times the standard deviation of the training data. Due to the time lag between series, the shift size, c_1 is calculated from the standard deviation of the San Francisco series, so $c_1 = 6.986$, while the shift size we will be seeking to identify in the Las Vegas series is $c_2 = 0.431 c_1 = 3.011$.

7.2.1.2 Results

Relative to the defined change point of the process in week 2, 1998, Table 7.2 summarizes the actual delays (in weeks) from the four methods. Since lag and correlation between series are incorporated, our method gives earlier detection than the Frisén and Wessman and parallel methods. Although the negative delays might usually be regarded as false alarms, since the change point (week 2, 1998) is subjectively chosen, this might be considered as evidence that the process has shifted gradually from week 53 in 1997 which might possibly be detected by our method which takes both lag and correlation between series into account. Also it is reported that influenza notification in Los Angeles was substantially high in weeks 52 - 53, 1997 and weeks 1 - 2, 1998 (Glaser *et al.*, 2002). Therefore, the date of the disease outbreak in San Francisco and Los Angeles might

Table 7.2: Actual week delay for detecting the mean shift in the process (case study 1)

Method	Delays in weeks (week detected)
Adj. Parallel	0 (week 2, 1998)
Frisén	-1 (week 1, 1998)
Wessman	-1 (week 1, 1998)
Case 1	-2 (week 53, 1997)

actually be in the last two weeks of 1997.

Even though the change point of the process is difficult to identify, recognizing and incorporating the correlation and lag between series properly in the sufficient reduction gives more chance of detecting a shift in the process. In this case, the Frisén method, which incorporates the lag between series, but not the correlation between series, gives longer delay by failing to detect gradual shift in the two series. The plots of detection performance of the four methods are shown in Figure 7.3.

7.2.2 Scarlet fever notifications, UK

This section provides an example of using SR methods for detecting a mean shift in bivariate Poisson data. The anonymized aggregated counts of scarlet fever notifications in Greater Manchester (GM) and Cheshire & Merseyside (C&M) in 2011 - 2012 provided by the Health Protection Agency North West Regional Epidemiology Unit (HPA, 2012) are used. Since the data are low counts, the SR method proposed for Poisson data (chapter 6) is considered in this example.

7.2.2.1 Training and test sets and detection tools

Figure 7.4 shows the weekly scarlet fever notifications in GM and C&M in weeks 1 - 53, 2011 and weeks 1 - 47, 2012. Since this disease tends to be most common during winter and spring (HPA, 2011), there is a peak a few weeks in January in 2011 and a dramatic shift in the 2^{nd} week of 2012. Due to the high peak in 2012, we chose to regard the 2^{nd} week of 2012 as the change point of the process. The training set is chosen from the non-peak period of 2011. As in case study 1, due to subjectively choosing the change point, we allow 4 weeks prior to the change point as a part of the test data. The training and test periods are defined as below.

- the training data: weeks 15 - 49, 2011
- the test data: weeks 50 - 52, 2011 and weeks 1 - 47, 2012
- the change point: week 2, 2012

The plot of the defined training and test sets is shown in Figure 7.5. The vertical dashed line separates the training data and test data, while the vertical dotted line represents the change point of the process.

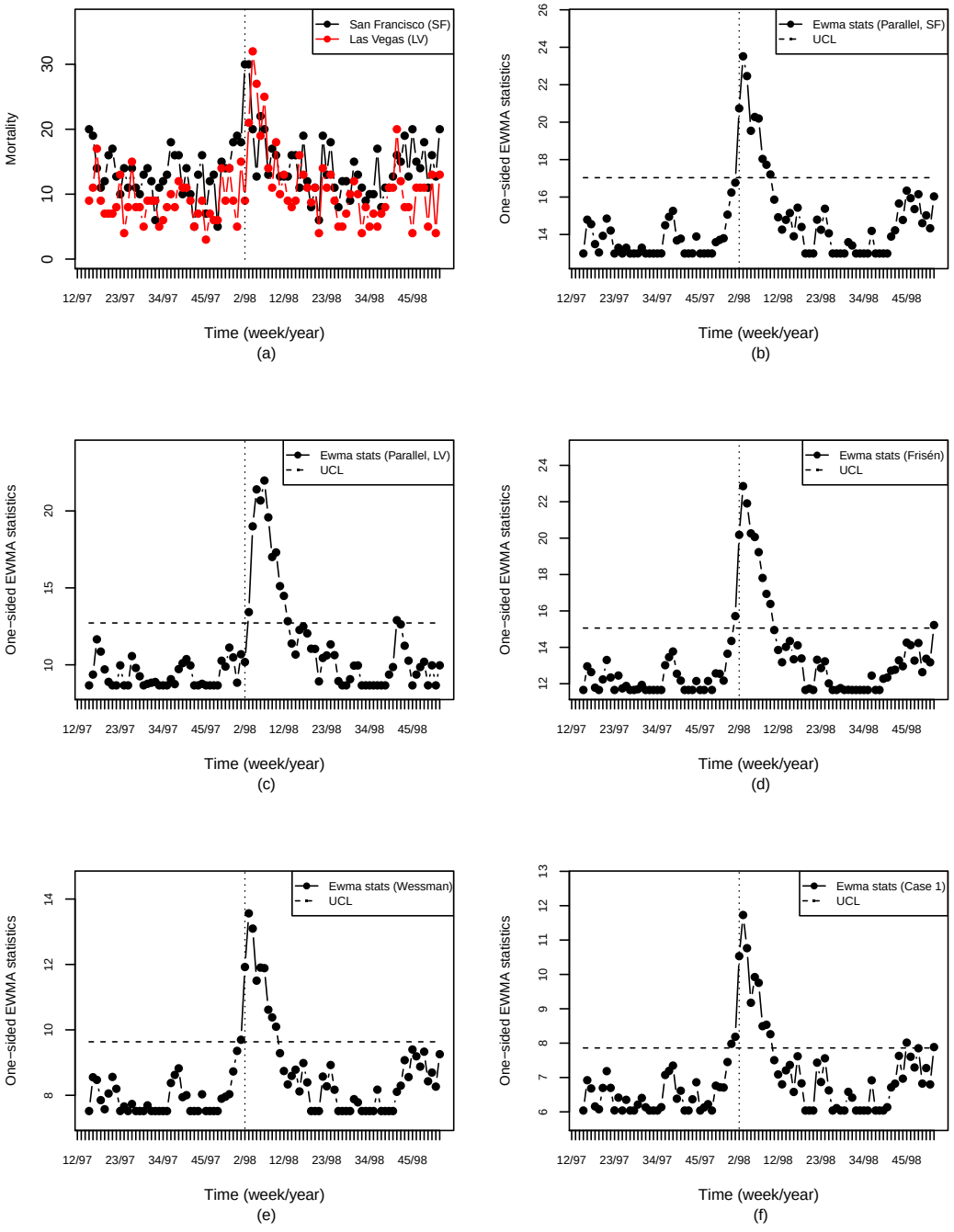

Figure 7.3: Results for detecting a mean shift in the San Francisco (SF) and Las Vegas (LV) series: (a) data plot, (b) - (c) EWMA plots of parallel methods for SF and LV, respectively, (d), (e) and (f) EWMA plots of Frisén, Wessman and our methods, respectively.

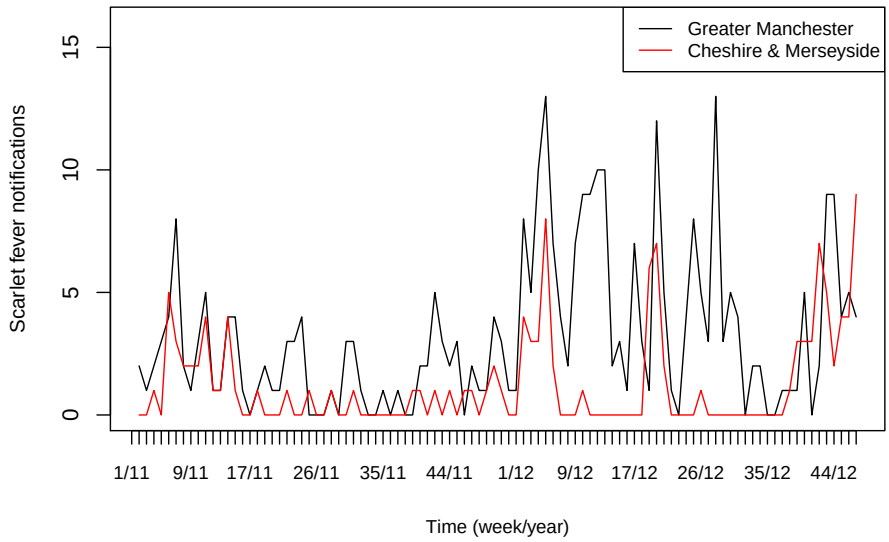

Figure 7.4: Plots of the scarlet fever notifications in Greater Manchester and Cheshire & Merseyside.

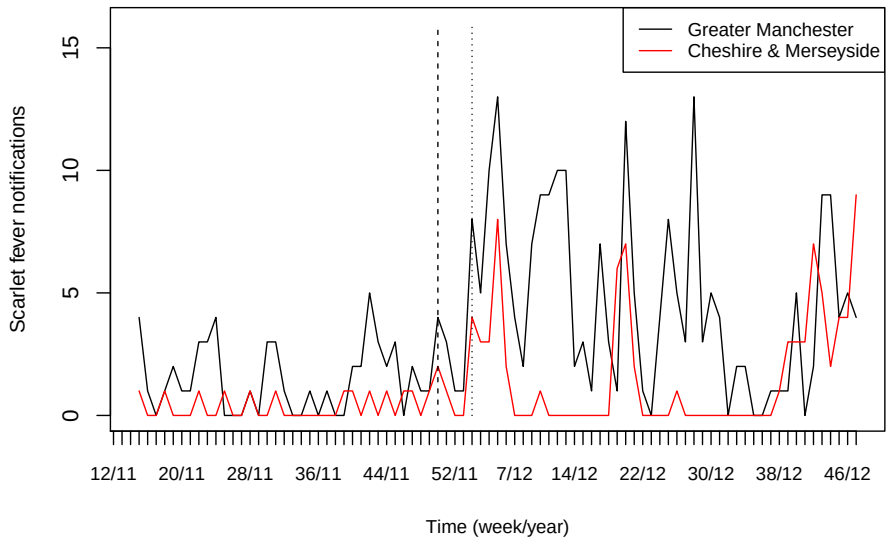

Figure 7.5: Plots of the training and test data for Greater Manchester and Cheshire & Merseyside series.

The histograms and Poisson Q-Q plots of training set of the two series (Figures D.4 and D.5 in appendix D) show that data in each series are Poisson rather than normal, with no evidence of correlation within series. The cross-correlation function plot (Figure D.6) indicates that the series are correlated with zero lag. The mean and standard deviation of each training set are summarized in Table 7.3

Table 7.3: Parameter estimation for the training sets (case study 2)

Series	Mean	Standard deviation	Correlation (Lag = 0)
Greater Manchester	1.528	1.444	0.344
Cheshire & Merseyside	0.417	0.554	

Since the two series are dependent, the relation of mean shift sizes between series is considered as defined in section 4.3.2 (i.e. a shift in one series reflects in a shift in another due to CBS). From Table 7.3, the mean and variance of GM are quite high compared to those of C&M, so the mean shift in the GM series might obviously be detected easily. We then subjectively define GM series as the first series and C&M as the second. Thus, the size of the mean shift of C&M series 'depends' on the shift size of GM series and CBS, i.e. $c_2 = \rho c_1$).

As shown in section 5.3.2, the SR method is fairly robust to mis-specification of the c_i. In this example we aim to detect a mean shift of size twice the standard deviation of training data of GM series. From Table 7.3, the shift size we want to detect in GM series, c_1 is 2.888, while the shift size we aim to detect in the C&M series is $c_2 = 0.344c_1 = 0.993$, respectively. Since the data are reasonably Poisson distributed, our method (case 2, chapter 6) proposed for detecting a mean shift in a bivariate Poisson process, Frisén and parallel (adjusted for multiplicity) methods are used. Due to the distribution of the data and derived statistics from SR methods, the parallel data are monitored with modified one-sided EWMA charts for independent Poisson data (section 6.5.1, chapter 6), while the derived statistics from Frisén and our proposed methods are monitored with a standard one-sided EWMA charts for independent normal data.

7.2.2.2 Results

The detection performance of the three methods is summarized in Table 7.4. Our method gives earlier detection than other methods, although it is technically a false alarm relative to the defined change point. Frisén and parallel methods perform equally detecting a mean shift in the 2^{nd} week of 2012 where the dramatic shift occurs. Once again, if we allow for the uncertainty of subjectively choosing the change point, the false alarm might be considered as a genuine warning of disease outbreak. It can be seen from the data plot (Figure 7.6 (a)) and the EWMA plot for our method (Figure 7.6(e)) that the alarm was given in week 50, 2011, where there are small shifts occurring in both series. By taking correlation between series into account, our methods might be able correctly

Table 7.4: Actual week delay for detecting the mean shift in the process (case study 2)

Method	Delays in weeks (week detected)
Adj. Parallel	0 (week 2, 2012)
Frisén	0 (week 2, 2012)
Case 2	-4 (week 50, 2011)

to recognize these small simultaneous shifts as indicative of an outbreak, while they fail to trigger alarms under the other two methods.

Note that due to the nature of the Poisson distribution where the mean and variance cannot be separated, the shift in mean results in the shift in the variance. This can be seen from the plot of the training and test data in Figure 7.5. The variance of the two series substantially increases after the change point, so a control chart for monitoring a variance shift might be considered as an alternative detection tool. Also, some of the data are zero counts, the zero inflated Poisson EWMA chart accounting for rare events or zero counts in the process might also be considered instead (Fatahi *et al.*, 2012).

7.3 Residual data with simulated shifts

In this section, we use the Greek pollution data and the Swedish radiation data to illustrate the implementation of SR methods. The data set has been collected routinely due to environmental concerns. For the Greek pollution data set, two variables (humidity and ozone levels) measured in Athens, provides an example of detecting a mean shift with no time lag between series, while the Swedish radiation data set, one variable (radiation level) measured from two different areas, exhibits both correlation and time lag between series. However, the actual data for both data sets cannot be directly used with the SR methods proposed for detecting a shift in mean levels (i.e. from μ_0 to μ_1), since they show evidence of non-stationarity. Preliminary analysis is used to investigate whether derived data might be suitable to use with these methods.

Since there is no real or significant mean shift in the data set, artificial signals of different shift sizes are injected into the data for the purpose of system evaluation. The details of conducting preliminary analysis, calculating appropriate mean shift sizes, defining the training and test data sets, investigating correlation and lag between series, specifying detection tools and system evaluation are given in Section 7.3.1.

7.3.1 Data preparation and evaluation tools

7.3.1.1 Preliminary analysis

The preliminary analysis is conducted to investigate whether the data set can be used to demonstrate SR methods. As mentioned above, the Greek data sets are not stationary. Figure 7.7 shows the non-stationarity in humidity and ozone level in Athens in 2006. Due

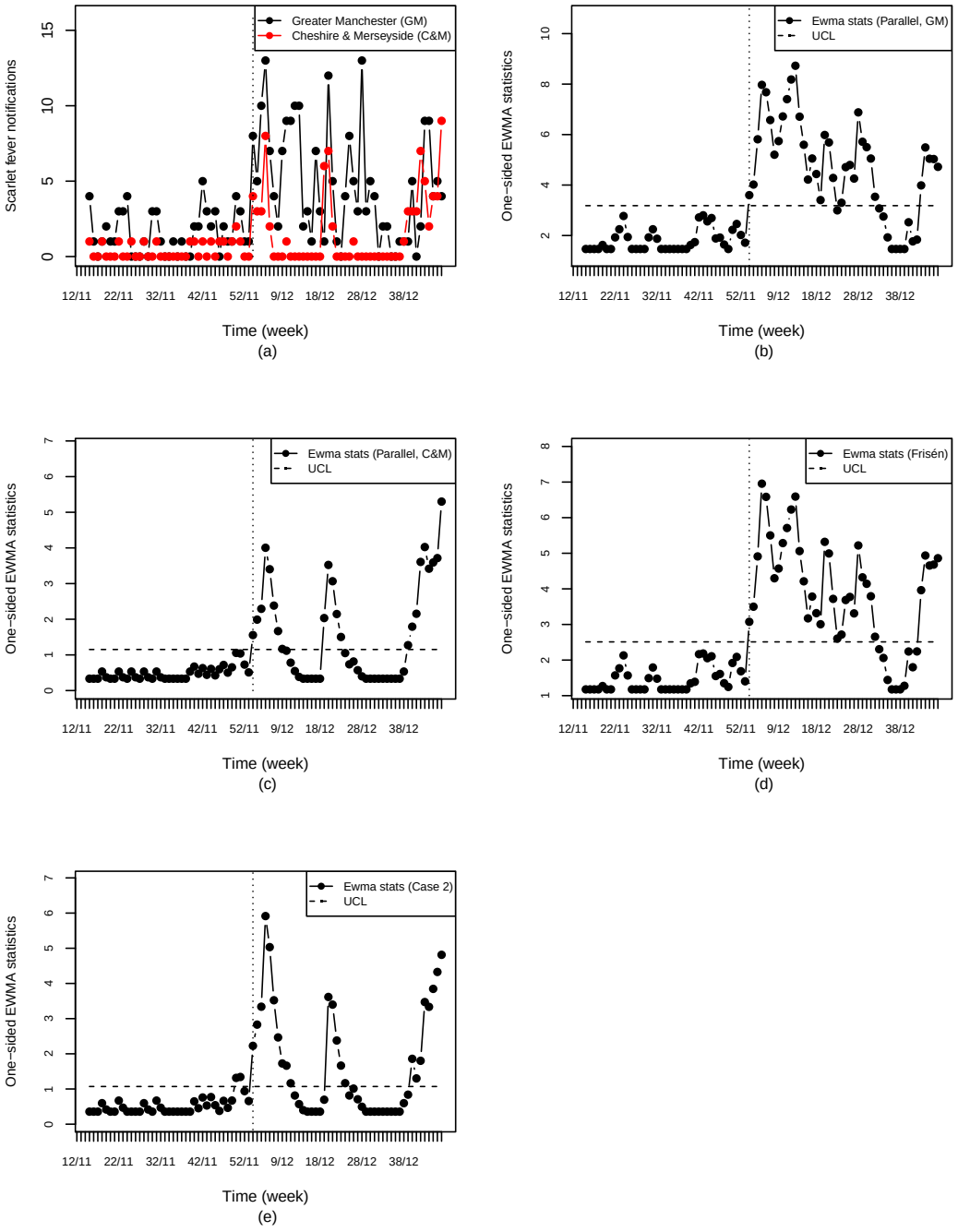

Figure 7.6: Results for detecting a mean shift in Greater Manchester (GM) and Cheshire & Merseyside (C&M) series: (a) data plot, (b) - (c) EWMA charts of parallel methods for GM and C&M, respectively, (d) and (e) EWMA charts of Frisén and our methods, respectively.

to the non-stationarity, the original data might not be suitable to use to demonstrate the SR methods. In order to detect a step change, for which SR method is proposed, we aim to use model residuals, assumed approximately identically and normally distributed, from fitting time series models to illustrate the SR methods. Since SR methods are proposed under different assumptions, the residuals are then assessed in order to choose a suitable SR method for monitoring a shift in the process. More details of the fitted model are provided for each set separately.

7.3.1.2 Mean shift size

As has been noted, there is no real or significant mean shift in either data set, and so artificial signals are calculated and added to the residuals for the purpose of system evaluation. The shift size we want to detect in each series is pre-specified and calculated proportional to the standard deviation of residuals. However, the standard deviation of the residuals in each series is quite different due to the variation in the actual data. So the mean shift in each series will not be equal since it depends on how large the standard deviation is.

To make a sensible comparison in detecting the same 'size' of mean shift for all series, residuals in each series are standardized by their own mean and variance. This makes the standardized residuals for all series have the same variance of 1. Four different shift sizes are investigated in this study, namely 0.5, 1, 2, and 3 times the standard deviation of the series. These shift sizes then are injected to the test data, detailed in the next section, for the purpose of system evaluation.

7.3.1.3 Training and test data sets

Since we have data only one year for each data set, we divide each series into two parts: a training set (before injected mean change), and test set (after change). The training data are used to investigate correlation within and between series, lag between series and to estimate the parameters used in the SR methods and EWMA charts. The test data is where artificial signals have been added. The first point of the test set is defined as the change point of the process. Due to no real or significant shift in either data set, the training set in each data set is chosen from a non-peak period, while the rest is defined as a test set.

The artificial signals will be added to the test data according to the specified change point and lag between series. If there is no lag between the two series, signals will be added on the same day and after, the change point, to both series. On the other hand, if there is an evidence of a lag, l, between series, the signals will be added to the first series according to the change point defined above, then l time points later, signals will be added to the other. Details of training and test sets for pollution and radiation data sets are summarized in each section separately.

7.3.1.4 Detection tools and system evaluation

To compare the performance in detection, four different methods: parallel method adjusted for multiplicity, Frisén method (Frisén *et al.*, 2011), Wessman method (Wessman, 1998) and our proposed methods (case 3 and case 5 in chapter 4) are considered in this section. The derived sequence of statistics will be monitored with two types of EWMA chart: a standard EWMA chart for independent normal data and a modified EWMA for dependent normal data. An appropriate EWMA chart (i.e. standard or modified charts) is chosen according to the nature of the residuals. More details of the EWMA chart used in each example is defined in each section separately.

The four methods are evaluated by comparing the actual delay measured from the day the signal is first added (the change point of a process) until the system gives a signal. For example, if the actual delay is 2, it means that the system can detect a mean shift 2 days after the mean has shifted. Negative actual delay means a false alarm. In public health surveillance, we want to detect changes as soon as possible. In this study we aim to detect a mean shift within 7 days (an arbitrary but sensible period), therefore, if the actual delay is longer than 7 days, it is considered unacceptable and reported as non-applicable (NA) in this study.

7.3.2 Greek pollution data

Due to concerns over air pollution, several substances in the air, such as Ozone, CO, NO and SO, are measured daily in Athens, Greece. In this study, we choose to monitor ozone and humidity levels since ozone is one of important indicators of air pollution and it is also influenced by humid air conditions. Data were made available by Dr Sotiris Bersimis (Bersimis, pers. comm.). Humidity and ozone levels ($\mu g/m^3$) measured daily at Thrakomakedones station in Athens during 2006 are plotted in Figure 7.7. It can be clearly seen that they are negatively correlated. The actual ozone and humidity levels might not be a good example to illustrate the SR methods since the mean levels in both series are not quite stable. Therefore, using model residuals instead of actual data might be more suitable in this case.

Since the use of the model residuals in the surveillance relies on how well the model fits, in this example we fit two different models to the Greek data set in order to show how SR methods can be used under different conditions. Generally, independent and normally distributed residuals are required because they can be simply used with the SR methods proposed for normal data and monitored with a standard chart. However, sometime it might be difficult to find a perfectly fitting model. Some patterns or variations might still be present in the model residuals which might not be independent over time. To handle this dependence, even of the 'residual data', the SR methods proposed for dependent observations might have to be considered.

7.3.2.1 Preliminary analysis

From Figure 7.7, both series are not stationary and quite varied during the year. We consider two different models: time series and regression models called 'Model A' and

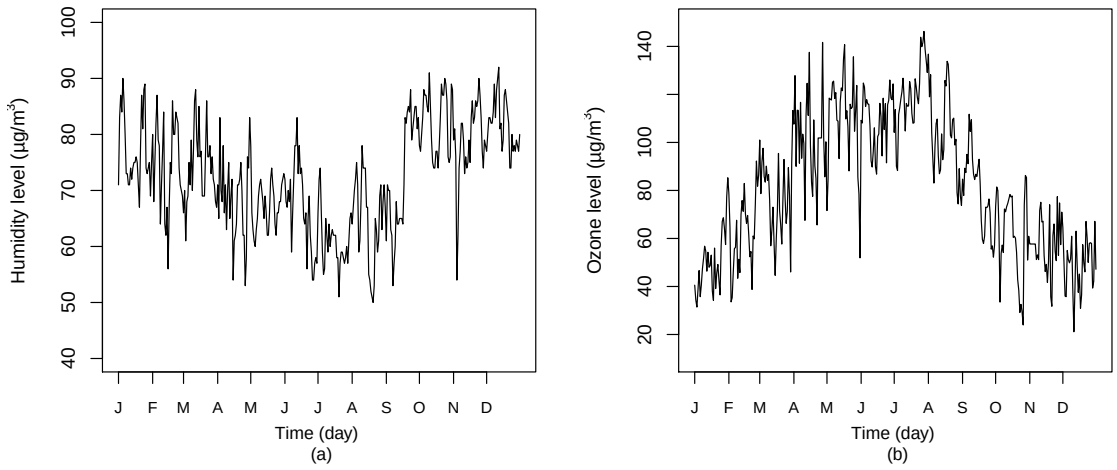

Figure 7.7: Plots of (a) humidity level ($\mu g/m^3$) and (b) ozone level ($\mu g/m^3$).

'Model B', respectively. The details of each model are described as follows.

Model A: Time series models

The lagged scatter plots (Figure D.7) and the plots of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of both series (Figure D.8), in Appendix D, present a trend and non-white noise sequence with a slow decay in the ACF, representing a non-stationarity, thus differencing is needed.

Let X_{H_t} and X_{O_t} be humidity and ozone levels measured on day t. After first order differencing, the differenced data for each series, Y_{H_t} and Y_{O_t}, where

$$Y_{H_t} = X_{H_t} - X_{H_{t-1}},$$

$$Y_{O_t} = X_{O_t} - X_{O_{t-1}}$$

and $t = 1, 2, ..., 265$, show stationarity. The stationarity can be observed from the lagged scatter plots of the differenced data for humidity and ozone series shown in Figure D.9 in Appendix D. The plots of the differenced data and the plots of the ACF and PACF of the differenced data for humidity and ozone series are given in Figures D.10 and D.11, respectively, in Appendix D.

The appropriate orders for a time series model for each series are considered from the plots of the ACF and PACF of the differenced data. For the humidity series, there is a sharp cut-off in the plot of ACF with significant ACF at lags 1 and 4 (Figures D.10 (b)). We consider the Moving Average (MA) model of order 4 as the appropriate time series model for describing the variation in the differenced humidity data since this is a much

simpler method because the residuals of the MA(4) model are look more independent and normally distributed than those of MA(1) model. The MA(4) model of the differenced data can be written as

$$Y_{H_t} = 0.0078 + \epsilon_t - 0.278\epsilon_{t-1} - 0.1916\epsilon_{t-2} - 0.0913\epsilon_{t-3} - 0.2338\epsilon_{t-4},\ \epsilon_t \sim N(0, 27.56),$$

or

$$X_{H_t} = 0.0078 + X_{H_{t-1}} + \epsilon_t - 0.278\epsilon_{t-1} - 0.1916\epsilon_{t-2} - 0.0913\epsilon_{t-3} - 0.2338\epsilon_{t-4}.$$

The diagnostic plots of the model are given in Figure D.12 in in Appendix D.

For the ozone series, there is a sharp cut off in the plot of the ACF with significant ACF at lags 1, 2 and 4 (Figure D.11 (b)). Since the residuals from the MA(4) are more independent and normally distributed than those from the MA(1) and MA(2) models, we choose the MA(4) model as the appropriate time series for describing the variation in the differenced ozone data. The MA(4) model of the differenced data can be written as

$$Y_{O_t} = -0.0335 + \epsilon_t - 0.3640\epsilon_{t-1} - 0.2020\epsilon_{t-2} - 0.0565\epsilon_{t-3} - 0.1376\epsilon_{t-4},\ \epsilon_t \sim N(0, 177.4),$$

or

$$X_{O_t} = -0.0335 + X_{O_{t-1}} + \epsilon_t - 0.3640\epsilon_{t-1} - 0.2020\epsilon_{t-2} - 0.0565\epsilon_{t-3} - 0.1376\epsilon_{t-4}.$$

The diagnostic plots of the model are given in Figure D.13 in in Appendix D.

Model B: Regression models

Since data in both series are quite varied during the year, the data can be fitted by using either sine and cosine functions or a factor month. Since we do not have long historical data to define a proper number of harmonics or frequencies for sine and cosine functions and ozone level in each month is quite different, we then fit regression models using a factor month as a covariate. The regression model for each series can be written as

$$\begin{aligned} X_{H_t} &= \beta_1 + \beta_2(Feb) + \beta_3(Mar) + \beta_4(Apr) + ... + \beta_{11}(Nov) + \beta_{12}(Dec) + \epsilon_t \\ X_{O_t} &= \beta_1 + \beta_2(Feb) + \beta_3(Mar) + \beta_4(Apr) + ... + \beta_{11}(Nov) + \beta_{12}(Dec) + \epsilon_t \end{aligned}$$

where β_i are regression coefficient, $i = 1, 2, ..., 12$ and ϵ_t is a random error assumed independent and normally distributed. The estimates and R^2 for each model are summarized in Table D.1 with the diagnostic plot for each model in Figures D.14 and D.15 in appendix D. Residuals for both series look much more constant in mean once the variation due to the factor month have been taken out. Even though they are approximately normal distributed, they are not independent. The plots of ACF and PACF indicate the AR patterns.

The residuals in each series from both models are standardized in order to permit sensible comparison for detecting the same shift sizes and so the humidity and ozone series have the same variance of 1. The mean and variance of the standardized residuals for Models A and B are summarized in Table 7.5. The standardized residuals for both series are then used as the data in the sufficient reduction. The training and test sets for each series are defined in next section.

Table 7.5: Means and variance of standardized residuals of the two series (case study 3)

Series	Model A		Model B	
	Mean	Variance	Mean	Variance
Humidity	1.407e-17	1.0	-7.089e-18	1.0
Ozone	2.338e-17	1.0	5.449e-18	1.0

7.3.2.2 Training and test sets

Since there is no real shift in either series, the change of the process is chosen arbitrarily on 1 July, 2006 which divides the training and test sets for each series as below.

- the training set: 1 January - 30 June 2006
- the test set: 1 July - 31 December 2006
- the change point: 1 July 2006

The standardized residuals for each series from the previous section are divided into two parts according to the change point defined above. The training data, the standardized residuals, for each series from each model were checked in order to asses whether they follow the SR assumptions and whether or not there is a lag between series.

Consider the plots of training data from Model A. The normal QQ plots and the histograms of each series (Figures D.16 and D.17, in Appendix D) show that the standardized residuals in each series are independent and normally distributed. The two series are negatively correlated with lag zero, which can be observed from the plot of the cross correlation function (CCF) in Figure D.18, in Appendix D. The appropriate versions of the SR method are the Frisén method (Frisén *et al.*, 2011), Wessman method (Wessman, 1998) and our proposed method case 1 of chapter 4 (where the lag between series is zero). Both methods were proposed for detecting a mean shift with no lag in a process of independent normal data. The mean and variance of training data for each series are summarized in Table 7.6.

Table 7.6: Parameter estimation of the training set (Model A)

Series	Mean	Variance	Correlation (Lag = 0)
Humidity	-0.073	1.017	-0.445
Ozone	0.095	1.098	

For the Model B residuals, the plot of the cross correlation function (CCF) in Figure D.19, in appendix D, also shows that two series are negatively correlated with lag zero. The normal QQ plots of each series (Figures D.20 and D.21, in Appendix D) show that

the standardized residuals are not quite normally distributed, with evidence of AR(1) patterns visible in the plots of the ACF and PACF. Due to the correlation within and between series, the appropriate version of our SR method considered in this example is that in case 3 of chapter 4, i.e. that for detecting a mean shift with no lag in a process of dependent normal data. The mean and variance and autoregressive coefficient of AR(1) model calculated from the training data for each series are summarized in Table 7.7.

Table 7.7: Parameter estimation of the training set (Model B)

Series	Mean	Variance	Autoregressive coefficient	Correlation (Lag = 0)
Humidity	2.068e-18	0.972	0.550	-0.522
Ozone	6.725e-18	1.051	0.417	

Regarding the relation of shift sizes between series defined in section 4.3.2, the shift size in one series is reflected in the shift size in another due to the correlation between series, i.e. $c_2 = \hat{\rho} c_1$. Even though there is no lag between series, we define the humidity level as the first series since ozone level depends on the humidity in the air. Four different shift sizes, which are 0.5, 1, 2 and 3 times the standard deviation of the training data of the first series, are investigated. Since the standard deviation of the training data set for each series is roughly 1. Thus, for simplicity, the shift sizes we aim to detect in the humidity series, c_1, are taken as exactly 0.5, 1, 2 and 3. Correspondingly, due to CBS, the shift sizes we aim to detect in ozone series from the Model A and Model B are $c_2 = -0.445 c_1$ and $c_2 = -0.522 c_1$, respectively.

Due to the negative correlation between series, shifts occur in the opposite directions, i.e. an upward shift in the humidity series and a downward shift in ozone. Since there is no lag between series, artificial signals of sizes c_1 and c_2 are added into the test set in both series from 1 July to 31 December 2006. For example, the plot of standardized residuals from Model A with added signals of size 2 (humidity series) and of size -0.844 (ozone series) are shown in Figure 7.8.

Since a shift might occur in either direction, the residuals from Model A are monitored with a standard two-sided EWMA chart for independent normal data (Section 2.3.1.3). Due to the CWS being present in the Model B residuals, a standard two-sided EWMA chart for independent normal data (section 2.3.1.3) and a modified two-sided EWMA chart for dependent normal data (section 5.2) are used as detection tools in order to avoid the false alarms. The limit L for the standard two-sided chart can be obtained from the '`spc`' package in `R`, while the limit L for the modified two-sided chart can be obtained from Shiau and Hsu (2005). However, Shiau and Hsu (2005) provided the limit L of the modified two-sided EWMA chart for a specific $\mathrm{ARL}_0 = 370$, the parallel method adjusted for multiplicity, which might be preferable, but requires the limit L for $\mathrm{ARL}_0 = 741$, is not considered in this example.

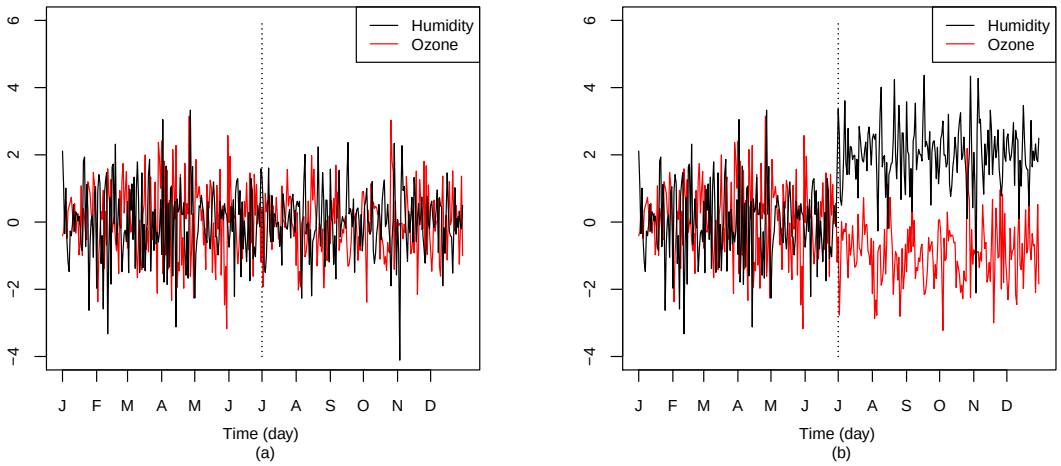

Figure 7.8: Plots of a) standardized residuals and b) standardized residuals with added signals (Model A).

7.3.2.3 Results

For Model A, the actual delays (days) of the four methods for detecting four different shift sizes with standard two-sided EWMA charts are summarized in Table 7.8. The example of detection plots for detecting a mean shift of sizes 2 and -0.89 in humidity and ozone series, respectively, is shown in Figure 7.9. As expected, larger mean shift sizes are detected more quickly than smaller shifts. All methods fail to detect the small shift sizes. The Wessman and our method perform equally and better than the parallel and Frisén methods for detecting the largest shift size since CBS is taken into account in the Wessman and our methods. All methods perform equally to detect a shift of size 2.

Table 7.8: Actual delays for monitoring four shift sizes (Model A)

Method	EWMA chart	Shift size in humidity series (c_1)			
		0.5	1	2	3
Parallel	standard	NA	NA	1	1
Frisén	standard	NA	NA	1	1
Wessman	standard	NA	NA	1	0
Proposed	standard	NA	NA	1	0

For Model B, the actual delays (days) of the three methods for detecting four different shift sizes with both standard and modified two-sided EWMA charts are summarized in Table 7.9. The example of detection plots for detecting a mean shift of sizes 2 and -1.044

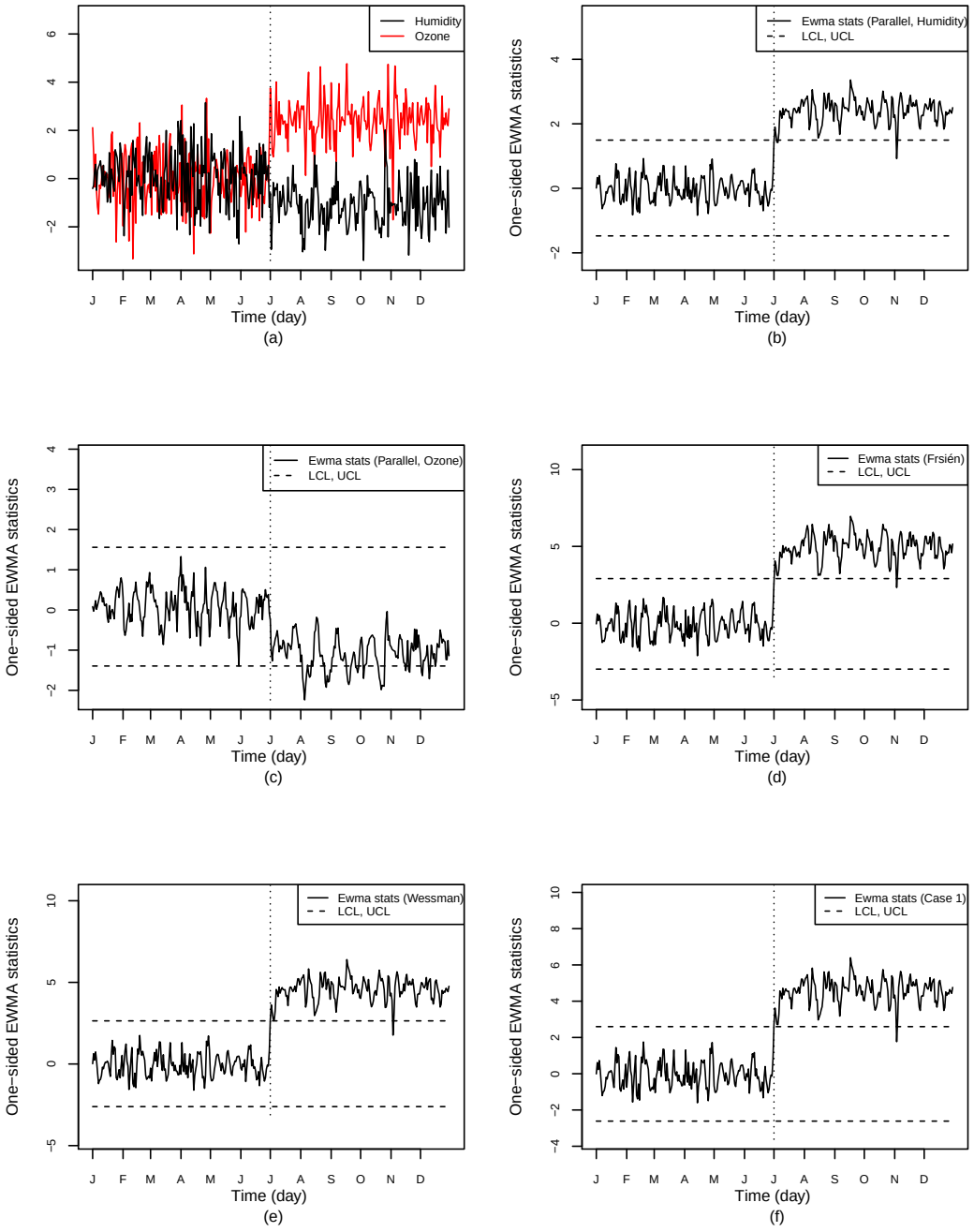

Figure 7.9: Results for detecting a mean shift of sizes 2 and -0.844 in humidity and ozone series (Model A): (a) plot of standardized residuals, (b) and (c) plots of parallel methods with standard chart, respectively, (d), (e) and (f) plots of Frisén, Wessman and our proposed methods, respectively.

in humidity and ozone series, respectively, is shown in Figure 7.10. All methods fail to detect the small shift sizes, 0.5 and 1, however, our proposed method gives shorter delays than other methods for detecting the larger shift sizes. The parallel, Frisén and Wessman methods give false alarms due to the autocorrelation present in the parallel data and derived sequence of statistics from the Frisén and Wessman methods. However, they can be improved by using a modified EWMA chart accounting for the autocorrelation instead of standard EWMA chart for independent data (Figure 7.10 (d) - (e)), but still respond slightly more slowly than our method. While the other methods can be improved by using modified EWMA chart, our proposed method, which incorporates both CBS and CWS, gives the same result for both charts since CWS has already been removed in the sufficient reduction. This makes the use of modified EWMA chart is unnecessary in this case.

Table 7.9: Actual delays for monitoring four shift sizes (Model B)

Method	EWMA chart	Shift size in humidity series (c_1)			
		0.5	1	2	3
Parallel	standard	-136	-136	-136	-136
	modified	-	-	-	-
Frisén	standard	-136	-136	-136	-136
	modified	NA	NA	2	2
Wessman	standard	-136	-136	-136	-136
	modified	NA	NA	2	2
Proposed	standard	NA	NA	1	1
	modified	NA	NA	1	1

7.3.3 Swedish radiation data

The Swedish radiation data set used in this example is the radiation levels measured at two stations (Overtörnea and Pajala) in Sweden during 1998 by the Swedish Radiation Safety Authority for research purposes (SRSA, 2012). The plot of daily radiation levels (nSv/h (nanoSievert per hour)) from the two stations are shown in Figure 7.11 (a). It can be clearly seen that radiation levels for both stations increase from May and remain fairly stable until October before decreasing in November and December. As in the previous example, the actual data cannot be directly monitored, as they are not stationary during the year and are related to other climate covariates. In this study we consider daily snow depth levels (cm) in 1998, measured by the Swedish Meteorological and Hydrological Institute (SMHI, 2012), as a covariate. The plot of snow depth levels in Overtörnea and Pajala are shown in Figure 7.11 (b).

These data were analyzed previously by Järpe (2000) and it was planned to conduct a comparative study. The radiation level was fitted by considering snow depth level and the period of the year as covariates. The period of the year was considered as another covariate because it corresponds to the nature of the snow (i.e. due to accumulated

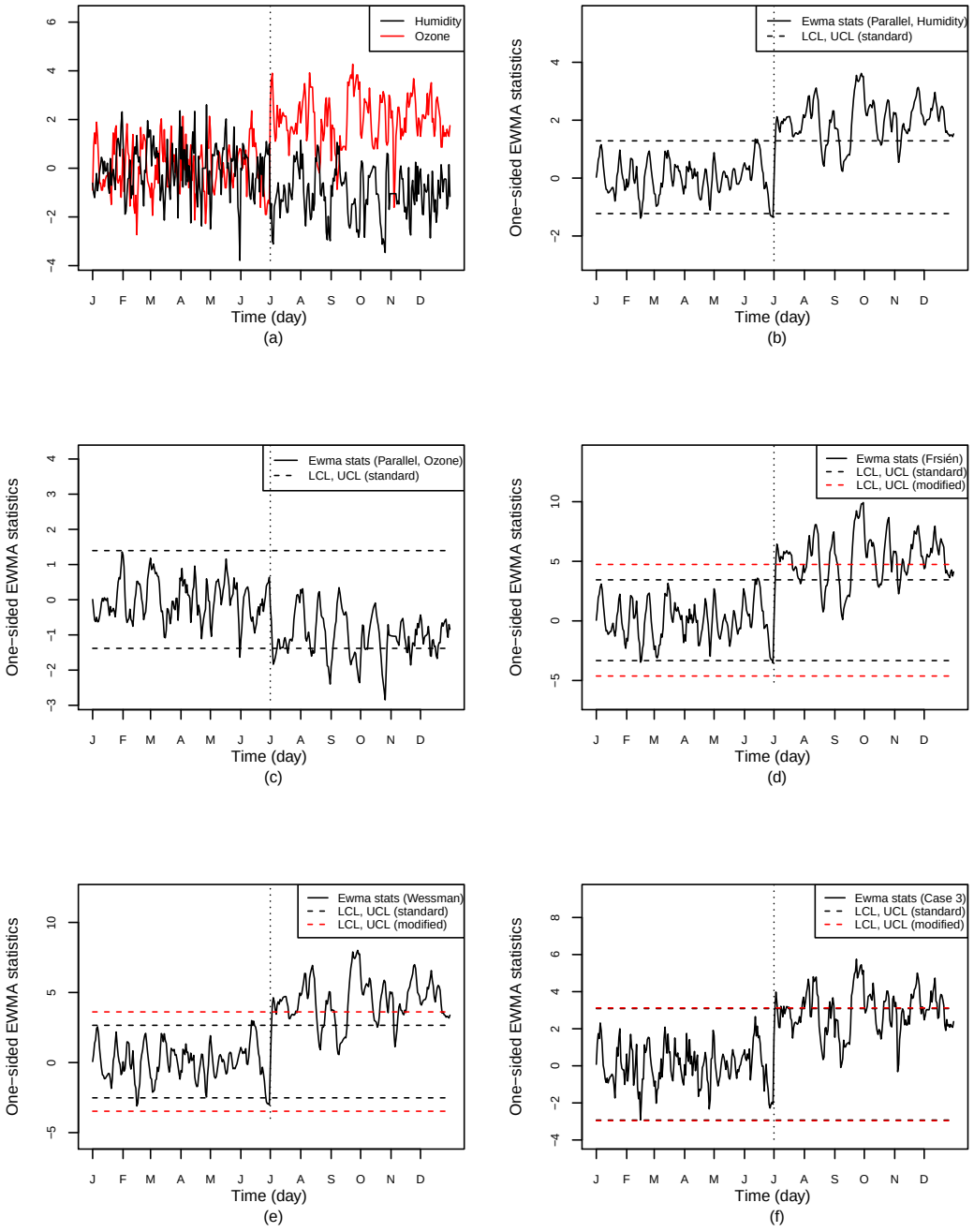

Figure 7.10: Results for detecting a mean shift of sizes 2 and -1.044 in humidity and ozone series (Model B): (a) plot of standardized residuals, (b) and (c) plots of parallel methods with standard chart, respectively, (d), (e) and (f) plots of Frisén, Wessman and our proposed methods, respectively.

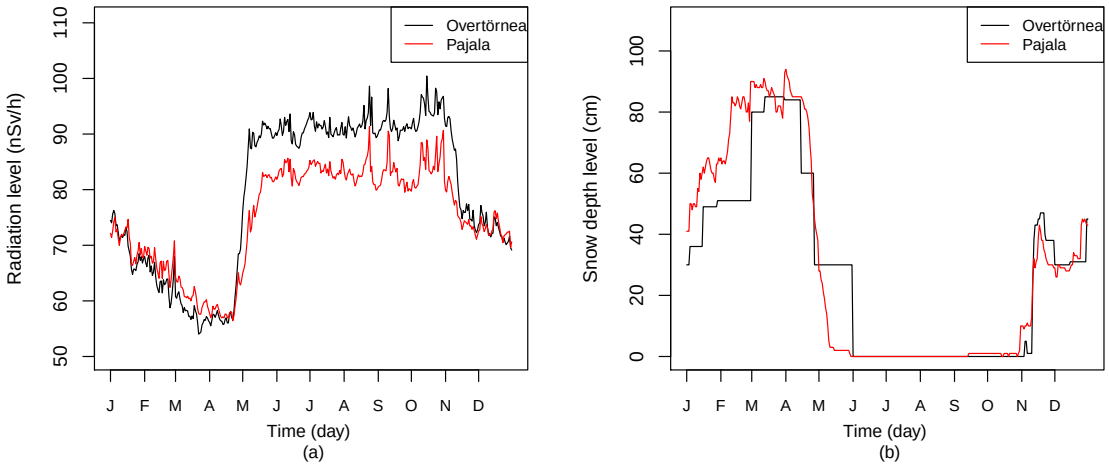

Figure 7.11: Plots of a) radiation level and b) snow depth level measured from Overtörnea and Pajala.

snow during winter, snow is mostly soft and porous in the first half year and more hard and icy during the second) (Järpe, 2000). However, the residuals from the fitted models are not quite stationary during the surveillance period. Thus, the residuals from the Swedish data set might be not a good example to illustrate the use of SR methods which are proposed to detect a shift in mean level during the process. Therefore, we do not conduct any further analysis of this data set.

7.4 Conclusions

Four data sets are used to illustrate the SR methods. Due to the nature and variation in data, two types of data sets are used: actual data with real shifts and model residuals with added artificial signals. The former has the advantage over the latter since the result can be interpreted in practice. However, the system evaluation is quite difficult because a change point of the process is unknown and difficult to identify. Since the change point is chosen arbitrarily for the purpose of system evaluation, the false alarm can be considered either a real false alarm or a genuine warning of disease outbreak which has actually occurred before the time we declared the official start (change point).

Alternatively, in order to avoid a false alarm occurring during an arbitrary 'guard region', which is defined as a part of the test data to allow for uncertainty in subjectively defining the date of change point, one might exclude this region from the test data and consider this region as a gap between the training and test data. A decision as to whether or not a process is out of control will not be made during this gap in order to avoid a false alarm.

On the other hand, evaluating system performance by using model residuals with added signals is more straightforward since the change point of the process and the shift sizes are assumed known. However, the difficulty of using the residuals is that it is difficult to interpret the results in practice. For example, in public health we aim to detect a positive mean shift in a process, the negative correlation between the two residual series might not be easy to interpret in this case. Unfortunately, we were unable to source any real data with a definitively dated outbreak.

No matter what type of data is used to detect a mean shift in the process, the methods and one-sided EWMA chart should be chosen carefully according to the nature of the data and the derived likelihood ratio statistics (e.g. since SR methods proposed for normal or Poisson data differ). For example, ignoring CWS in parallel, Frisén and Wessman methods give false alarms when a standard chart is used. Even though the use of modified chart improves this problem, they give longer delays compared to our methods due to higher control limits.

Chapter 8

Discussion

For the purpose of detecting an abberation or anomaly which might signal the possibility of disease outbreak in public health surveillance, multivariate surveillance is preferable to univariate surveillance, since the correlation between series (CBS) is recognized and incorporated. A small, but persistent, shift is more likely to be detected if CBS is present and all the series are monitored together. The sufficient reduction method is considered. This is one of the dimensionality reduction tools for multivariate surveillance and has proved promising for handling CBS and lag between change points (LCP). Respective the principle of sufficiency, the SR method reduces a p-dimensional multivariate series to a univariate series of statistics by taking all relevant information from the original series into account. By the factorization theorem, the derived univariate series has been shown to be sufficient for monitoring a sudden, but persistent, in the multivariate process. To date this method has been proposed for handing CBS and lag between change point (LCP) in a process of independent data, but they had not previously been used when correlation within series (CWS) is present.

In public health surveillance, health data are often dependent over time and might also be low counts for rare events or sparse diseases. To handle these issues, we firstly proposed SR methods under different distributions: either data are normal or Poisson distributed. Three parameters (CWS, CBS and LCP) are incorporated in the sufficient reduction. In this thesis, we assume that data are autocorrelated (CWS), coming from AR1 model. The relation of the mean shift between series is determined by the correlation between series. Two types of change point (simultaneous changes and changes with time lag) and various shift sizes are investigated. Since an upward shift is typically of interest than a downward shift in health surveillance, a suitable one-sided EWMA chart (e.g. a standard chart or a modified chart to allow for the autocorrelation) is used as a detection tool.

The system evaluation is conducted by the simulation study. Early detection with a low false alarm rate is required. Four measures, the conditional expected delay (CED), true alarm rate (TAR), false alarm rate (FAR) and non-detection rate (NDR), are considered. The detection performance of the proposed SR methods are compared against the existing SR methods (Frisén method (Frisén *et al.*, 2011) and Wessman method (Wessman, 1998)) and the parallel method adjusted for multiplicity. The results, with some limitations and suggestions for further study, are discussed in separate sections below.

8.1 SR methods for normal data

In order to overcome the limitations of the Frisén and Wessman methods, which assume that observations are independent over time and do not incorporate CBS in the former and LCP in the latter, the SR methods proposed for detecting a mean shift in a process of autocorrelated data are proposed by incorporating any or all of CWS and CBS and LCP in the sufficient reduction. As expected, the parallel method performs poorly compared to the others due to the higher control limit adjusted for multiplicity. Also, as expected, the Frisén method performs better for detecting changes with time lag than the Wessman method and worse than the Wessman and our methods when there is CBS.

If CWS is cleary present, our proposed methods perform better with shorter CED, higher TAR and lower FAR than the Frisén, Wessman and parallel methods. Even though, our proposed methods give higher NDR for detecting a small shift size, the NDR improves when shift size is large. From the plot of the autocorrelation function, the high FAR in the parallel, Frisén and Wessman methods is due to the effect of CWS being present in the parallel data and the derived sequence of likelihood ratio statistics, violating the assumption of a standard one-sided EWMA chart, namely that the data need to be independent over time.

The FAR and TAR of such methods are substantially improved by monitoring the parallel data and the derived sequences with a one-sided EWMA chart modified to allow for autocorrelation in a process. However, due to the higher control limits of the modified charts, those three methods still give longer CED than our method. Since the CWS is incorporated in our proposed methods, whether the standard or modified chart is used, there is little difference in their performance. This makes the modified chart unnecessary for our methods. Overall, the four methods give higher NDR for detecting a small shift size. However, NDR is improved when the shift size is large, i.e. a small shift size is more unlikely to be detected than a large shift size. Even though the illustration of SR methods in based on the bivariate case, the multivariate case can be extended in a similar manner.

The SR methods are proposed assuming that data are autocorrelated with the AR1 pattern. Due to the nature of diseases in public health surveillance, different types of time series models might be considered. For example, the nature of endemic disease might be related to the seasons or some other periods of a year, therefore, time series models with a seasonal component might be considered. Larger term dependence models (AR(p), $p > 1$) might also be relevant in some circumstances. Also, we proposed the SR methods assuming that the variance and the estimates of CWS, CBS and LCP estimated from the in control stage are unchanged over time. However, this might not be realistic due to the non-stationarity in the data. Thus time series models, which allow for the uncertainty in variance (heteroscedasticity or volatility), might be considered as alternatives. For example, the heteroscedasticity or volatility might be incorporated by considering an autoregressive conditional heteroscedastic model (ARCH model) instead of an AR1 model.

In principle, the SR method is proposed for detecting a step, but persistent, shift in a mean parameter by deriving the likelihood ratio statistics from the likelihood ratio between the out of control and in control stages, assuming that other parameters are unchanged over time. However, in some cases, the CWS, CBS or variance cannot be controlled due to the nature of diseases or types of shifts (e.g. gradual, linear or exponential shift). Updating parameters or re-defining the stages of the process might be investigated. More discussion is provided in section 8.3.

8.2 SR methods for Poisson data

The idea of SR methods has been extended to detect a simultaneous mean shift in a process of Poisson data since health data for rare events or sparse diseases are low counts. Previous SR methods (Frisén method (Frisén *et al.*, 2011) and Wessman method (Wessman, 1998)) considered this possibility since they were derived for exponential family distribution, but here we have incorporated CWS as well as CBS. The SR methods for detecting a change with time lag can be extended as a special case of simultaneous changes of known time lags, but have not been fully worked through here. Since the mean and variance of Poisson data cannot be separated, a shift in mean also induces a variance shift in the process. Therefore, the assumptions of SR methods defined for normal data (e.g. variance is unchanged over time) do not hold in this case. Two one-sided EWMA charts are modified for monitoring independent and dependent Poisson data via the parallel method.

The use of residuals from Poisson regression model, assumed identically and independently normally distributed, with the SR methods proposed for normal data is considered and conducted as the alternative way for monitoring a shift in a Poisson process. However, due to the limitations of monitoring residuals in prospective surveillance, the use of a method specific to Poisson data has an advantage over the use of the residuals which slows detection and gives high FAR.

The detection performance of our methods proposed for Poisson data are compared against the Frisén method and parallel method adjusted for multiplicity. Again, the parallel method performs worse than the others due to the high control limit after adjustment for multiplicity. Our method performs equally to the Frisén method, when there is no CWS or CBS in the process, and better than the others for detecting a small shift size when CBS is present. As expected, whether a standard or modified chart is used, the proposed methods perform better, by giving shorter CED, than the others when CWS is clearly present. However, the TAR, FAR and NDR of our methods are not uniformly better than the other methods for all levels of CBS and CWS. The FAR and NDR of the Frisén and parallel methods are substantially improved with a modified chart, while there is little difference in our methods since CWS has been removed in the sufficient reduction. Even though the FAR and NDR of the Frisén and parallel methods can be improved with a modified chart, their delays in detection are still longer than our methods.

Several limitations, with some suggestions for SR methods for Poisson data, are considered. Firstly, our methods produce higher FAR than the others when CBS is present. Since the likelihood ratio statistics is derived according to the shift in covariance, which results in the mean shift in the process, our method might be more sensitive to detecting a shift in a process if two series are not independent. Outliers observed from the density plot of the derived statistics tend to produce false alarm easily than the other methods. Also, even though we monitor the derived likelihood ratio statistics with the standard EWMA chart for normal, the statistics are not perfectly identically normally distributed. A suitable modification to the EWMA chart should be considered carefully with regard to the exact distribution of the likelihood ratio statistics.

Secondly, due to the complex probability function of the bivariate Poisson distribution and the Poisson bivariate integer-valued autoregressive model of degree 1, the likelihood ratio statistics cannot be factorized easily according to the factorization theorem. This makes the calculation of our methods more complicated than the Frisén methods. However, for the process of independent Poisson data (no CBS or CWS), our method is similar to the Frisén method.

Thirdly, according to the trivariate reduction method, which is used to construct the bivariate Poisson series from three mutually independent Poisson random variables, negative CBS is not allowed since the parameters for those variables need to be positive. Even though this limitation of using the trivariate reduction method is fairly unimportant in public health surveillance where we mainly focus on a shift in a positive direction, it might be useful to pursue this possibility for other applications. A method used to generate multivariate Poisson data with negative correlation can be found in Yahav and Shmueli (2012b), however, CWS is not allowed.

Finally, assuming that the mean and variance of Poisson data are equal and cannot be separated might be unrealistic in health surveillance. Thus, the negative binomial distribution allowing for the overdispersion in Poisson data might be considered instead of the Poisson distribution. Also frequent zero count for rare events in Poisson data should be taken in the consideration in further study.

8.3 Outbreak characteristics

In this thesis, we proposed the SR methods for detecting a sudden, but persistent, mean shift in a process. However, in practice, there might be several types of outbreak shapes to consider, for example, a gradual, linear or exponential mean shift. These types of shifts have been used in several studies to evaluate the system performance of several methods proposed for public surveillance (Goldenberg *et al.*, 2002; Mandl *et al.*, 2004; Wang *et al.*, 2005; Buckeridge, 2007; Jackson *et al.*, 2007). In order to implement the SR methods for detecting those shifts, it should be noted that the likelihood ratio statistics is derived based on the likelihood ratio between the out of control and in control stages. Unlike the sudden, but persistent, case where the process is divided into two-steps according to a single step shift, other shifts make more than two-step changes

in a process. Therefore, definition of simple in control and out of control stages with the corresponding parameters in each stage is difficult or impossible. In this case the optimality of sufficient reduction methods for detecting a shift in parameter between two distributions might not be hold. Before attempting to develop such methods for detecting these more complex shifts, one might investigate those SR methods; proposed for detecting a simple sudden, but persistent, shift to see how robust they are; though presumably longer delays should be expected.

In addition, the SR methods can be implemented in several research areas such as environmental surveillance and financial surveillance where detecting a change in a process is of interest. Due to the data used in environmental and financial surveillance, continuous surveillance, where decision is made continuously or more frequently than the discrete surveillance common in public health (where a decision is made daily or weekly), might be considered.

8.4 Modified one-sided EWMA chart

Even though the statistical process control is not the main topic of this thesis, three types of one-sided EWMA charts are modified in order to use as detection tools in the system evaluation of SR methods for normal data and of the parallel method for Poisson data. The smoothing parameter and in control average run length are arbitrarily defined in order to make a sensible and reliable comparison in this thesis. Further study of the robustness of such charts should be conducted by investigating the performance of the charts with different values of smoothing parameter and in control average run length.

Also, there are several ways to modify a one-sided EWMA chart, depending on how one-sided EWMA statistics are calculated: for example, standard EWMA statistics, resetting EWMA statistics below the target mean by the target mean, resetting observations below the target mean by the target mean and the adaptive EWMA chart (Harris and Ross, 1991; Schmid, 1997; Hu *et al.*, 2011). Different ways of calculating the EWMA statistics might affect the detection performance, especially for different types of shift in the process. Thus, a robustness study of the modified one-sided EWMA charts and the calculation of EWMA statistics for different types of shifts might be considered.

8.5 Conclusions

Sufficient reduction methods have been proposed for reducing a multivariate process of normal or Poisson data paying due regard to the purpose of health surveillance and the nature of health data. Correlation within and between series and lag between change points are incorporated. A derived sequence of likelihood ratio statistics from the SR methods summarizes all relevant information from the original series and has proved to be sufficient for detecting a mean shift in the multivariate process.

The results from a simulation study show that recognizing and incorporating the parameters appropriately in the SR methods improved the detection performance, with

shorter delay and lower false alarm rate. Ignoring correlation within series gives worse results, by producing high false alarms, than ignoring correlation between series, which causes failure to detect a small shift in the process. If two series are correlated with time lag between series, incorporating the lag and the correlation between series properly also improves the detection ability of such methods.

Even though, sufficient reduction methods have been proposed under specific assumptions (e.g. data have a pattern of AR1 model), the idea can be extended depending on the nature of the data and types of change point. Sufficient reduction methods can therefore be used not only in public health surveillance, but also to detection of a change point in a process in several disciplines.

Appendix A

SR methods for normal data

A.1 Sufficient statistics for case 1 (no CWS but CBS and LCP)

From equation (4.10), due to the known time lag and the availability of data, $T_t(\boldsymbol{x}_t)$ for $t = 1, 2, ..., s-l$ can be derived from

$$\begin{aligned}\prod_{t=\tau_1}^{s-l} \frac{f^O(\boldsymbol{y}_t;\boldsymbol{\mu}^O)}{f^I(\boldsymbol{y}_t;\boldsymbol{\mu}^I)} &= \prod_{t=\tau_1}^{s-l} \frac{exp - \frac{1}{2}\{(\boldsymbol{y}_t-\boldsymbol{\mu}^O)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}^O)\}}{exp - \frac{1}{2}\{(\boldsymbol{y}_t-\boldsymbol{\mu}^I)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}^I)\}} \\ &= exp\{\sum_{t=\tau_1}^{s-l}((\boldsymbol{\mu}^O-\boldsymbol{\mu}^I)'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_t) \\ &\quad - \frac{1}{2}\sum_{t=\tau_1}^{s-l}(\boldsymbol{\mu}^{O'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^O - \boldsymbol{\mu}^{I'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^I)\} \qquad \text{(A.1)}\end{aligned}$$

Hence by the factorization theorem, the likelihood ratio statistics $T_t(\boldsymbol{x}_t)$ for $t = 1, 2, ..., s-l$ are derived from equation (A.1) as

$$\begin{aligned}T_t(\boldsymbol{y}_t) &= (\boldsymbol{\mu}^O-\boldsymbol{\mu}^I)'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_t \\ &= \begin{pmatrix} \mu_1^O-\mu_1^I \\ \mu_2^O-\mu_1^I \end{pmatrix}' \boldsymbol{\Sigma}^{-1} \begin{pmatrix} x_{1,t} \\ x_{2,t+l} \end{pmatrix} \\ &= \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}' \boldsymbol{\Sigma}^{-1} \begin{pmatrix} x_{1,t} \\ x_{2,t+l} \end{pmatrix}\end{aligned}$$

From equation (4.10), the likelihood ratio statistics $T_t(\boldsymbol{x}_t)$ for $t = s-l+1, ..., s$ can be derived from

$$\begin{aligned}\prod_{t=s-l+1}^{s}\frac{f^O(x_{1,t};\mu_1^O)}{f^I(x_{1,t};\mu_1^I)} &= \prod_{t=s-l+1}^{s}\frac{exp\ -\frac{1}{2\sigma_1^2}\{(x_{1,t}-\mu_1^O)^2\}}{exp\ -\frac{1}{2\sigma_1^2}\{(x_{1,t}-\mu_1^I)^2\}}\\ &= \prod_{t=s-l+1}^{s}exp\ -\frac{1}{2\sigma_1^2}\{-2\mu_1^O x_{1,t}+2\mu_1^I x_{1,t}+(\mu_1^O)^2-(\mu_1^I)^2\}\\ &= exp\{\sum_{t=\tau}^{s}((\mu_1^O-\mu_1^I)x_{1,t}/\sigma_1^2)-\frac{1}{2\sigma_1^2}\sum_{t=\tau}^{s}((\mu_1^O)^2-(\mu_1^I)^2)\} \qquad (A.2)\end{aligned}$$

By the factorization, the likelihood ratio statistics for detecting changes with time lag for $t = s-l+1, ..., s$ are

$$T_t(\boldsymbol{x}_t) = c_1 x_{1,t}/\sigma_1^2$$

where $t = s-l+1, ..., s$.

A.2 Sufficient statistics for case 2 (CWS but no CBS or LCP)

From equation (4.12), according to the change point of a process at time τ, a likelihood ratio statistics $T_t(\boldsymbol{x}_t)$ for $t = 1, 2, ..., s$ can be derived from

$$\begin{aligned}\prod_{t=\tau}^{s}\frac{f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1};\boldsymbol{\mu}^O)}{f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1};\boldsymbol{\mu}^I)} &= \prod_{t=\tau}^{s}\frac{exp\ -\frac{1}{2}\{(\boldsymbol{x}_t-(\boldsymbol{\phi}\boldsymbol{x}_{t-1}+\mathbf{c}))'\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{x}_t-(\boldsymbol{\phi}\boldsymbol{x}_{t-1}+\mathbf{c}))\}}{exp\ -\frac{1}{2}\{(\boldsymbol{x}_t-\boldsymbol{\phi}\boldsymbol{x}_{t-1})'\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{x}_t-\boldsymbol{\phi}\boldsymbol{x}_{t-1})\}}\\ &= \prod_{t=\tau}^{s}\frac{exp\ -\frac{1}{2}\{((\boldsymbol{x}_t-\boldsymbol{\phi}\boldsymbol{x}_{t-1})-\mathbf{c})'\boldsymbol{\Sigma}_\epsilon^{-1}((\boldsymbol{x}_t-\boldsymbol{\phi}\boldsymbol{x}_{t-1})-\mathbf{c})\}}{exp\ -\frac{1}{2}\{((\boldsymbol{x}_t-\boldsymbol{\phi}\boldsymbol{x}_{t-1})-0)'\boldsymbol{\Sigma}_\epsilon^{-1}((\boldsymbol{x}_t-\boldsymbol{\phi}\boldsymbol{x}_{t-1})-0)\}}\\ &= \prod_{t=\tau}^{s}exp\ -\frac{1}{2}\{-2\mathbf{c}\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{x}_t-\boldsymbol{\phi}\boldsymbol{x}_{t-1})+\mathbf{c}'\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{c}\}\\ &= exp\{\sum_{t=\tau_1}^{s}(\mathbf{c}'\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{x}_t-\boldsymbol{\phi}\boldsymbol{x}_{t-1}))-\frac{1}{2}\sum_{t=\tau_1}^{s}(\mathbf{c}'\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{c})\} \qquad (A.3)\end{aligned}$$

where $\boldsymbol{x}_t = \begin{pmatrix} x_{1,t} \\ x_{2,t}\end{pmatrix}$ and $\mathbf{c} = \begin{pmatrix} c_1 \\ c_2\end{pmatrix}$.

By the factorization theorem, the likelihood ratio statistics in this case, derived from equation (A.3) can be written as

$$\begin{aligned}T_t(\boldsymbol{x}_t) &= \mathbf{c}'\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{x}_t-\boldsymbol{\phi}\boldsymbol{x}_{t-1})\\ &= \begin{pmatrix} c_1 & c_2\end{pmatrix}\boldsymbol{\Sigma}_\epsilon^{-1}\begin{pmatrix} x_{1,t}-\phi_1 x_{1,t-1} \\ x_{2,t}-\phi_2 x_{2,t-1}\end{pmatrix}\end{aligned}$$

where $t = 1, 2, ..., s$.

A.3 Sufficient statistics for case 4 (CWS, LCP but no CBS)

From equation (4.14), due to the known time lag and the availability of data, the likelihood ratio statistic $T_t(\boldsymbol{x}_t)$ for $t = 1, 2, ..., s-l$ can be derived from

$$\begin{aligned}
\prod_{t=\tau_1}^{s-l} \frac{f(\boldsymbol{y}_t|\boldsymbol{y}_{t-1};\boldsymbol{\mu}^O)}{f(\boldsymbol{y}_t|\boldsymbol{y}_{t-1};\boldsymbol{\mu}^I)} &= \prod_{t=\tau_1}^{s-l} \frac{exp\ -\frac{1}{2}\{(\boldsymbol{y}_t-(\boldsymbol{\phi}\boldsymbol{y}_{t-1}+\mathbf{c}))'\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{y}_t-(\boldsymbol{\phi}\boldsymbol{y}_{t-1}+\mathbf{c}))\}}{exp\ -\frac{1}{2}\{(\boldsymbol{y}_t-\boldsymbol{\phi}\boldsymbol{y}_{t-1})'\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{y}_t-\boldsymbol{\phi}\boldsymbol{y}_{t-1})\}} \\
&= \prod_{t=\tau_1}^{s-l} \frac{exp\ -\frac{1}{2}\{((\boldsymbol{y}_t-\boldsymbol{\phi}\boldsymbol{y}_{t-1})-\mathbf{c})'\boldsymbol{\Sigma}_\epsilon^{-1}((\boldsymbol{y}_t-\boldsymbol{\phi}\boldsymbol{y}_{t-1})-\mathbf{c})\}}{exp\ -\frac{1}{2}\{((\boldsymbol{y}_t-\boldsymbol{\phi}\boldsymbol{y}_{t-1})-0)'\boldsymbol{\Sigma}_\epsilon^{-1}((\boldsymbol{y}_t-\boldsymbol{\phi}\boldsymbol{y}_{t-1})-0)\}} \\
&= \prod_{t=\tau_1}^{s-l} exp\ -\frac{1}{2}\{-2\mathbf{c}'\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{y}_t-\boldsymbol{\phi}\boldsymbol{y}_{t-1})+\mathbf{c}'\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{c}\} \\
&= exp\{\sum_{t=\tau_1}^{s-l}(\mathbf{c}'\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{y}_t-\boldsymbol{\phi}\boldsymbol{y}_{t-1}))-\frac{1}{2}\sum_{t=\tau_1}^{s-l}(\mathbf{c}'\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{c})\} \qquad \text{(A.4)}
\end{aligned}$$

where $\mathbf{y}_t = \begin{pmatrix} x_{1,t} \\ x_{2,t+l} \end{pmatrix}$ and $\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$.

Hence by the factorization theorem, the likelihood ratio statistics $T_t(\boldsymbol{x}_t)$ for $t = 1, 2, ..., s-l$ are derived from equation (A.4) as

$$\begin{aligned}
T_t(\boldsymbol{x}_t) &= \mathbf{c}'\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{y}_t-\boldsymbol{\phi}\boldsymbol{y}_{t-1}) \\
&= \begin{pmatrix} c_1 & c_2 \end{pmatrix}\boldsymbol{\Sigma}_\epsilon^{-1}\begin{pmatrix} x_{1,t}-\phi_1 x_{1,t+l-1} \\ x_{2,t}-\phi_2 x_{2,t+l-1} \end{pmatrix}
\end{aligned}$$

where $t = 1, 2, ..., s-l$.

From equation (4.12), the likelihood ratio statistic $T_t(\boldsymbol{x}_t)$ for $t = s-l+1, ..., s$ can be derived from

$$\begin{aligned}
\prod_{t=s-l+1}^{s} \frac{f(x_{1,t}|x_{1,t-1};\mu_1^O)}{f(x_{1,t}|x_{1,t-1};\mu_1^I)} &= \prod_{t=s-l+1}^{s} \frac{exp\ -\frac{1}{2}\{(x_{1,t}-(\phi_1 x_{1,t-1}+c_1))^2/\sigma_{1,\epsilon}^2\}}{exp\ -\frac{1}{2}\{(x_{1,t}-(\phi_1 x_{1,t-1}))^2/\sigma_{1,\epsilon}^2\}} \\
&= \prod_{t=s-l+1}^{s} \frac{exp\ -\frac{1}{2}\{((x_{1,t}-\phi_1 x_{1,t-1})-c_1)^2/\sigma_{1,\epsilon}^2\}}{exp\ -\frac{1}{2}\{((x_{1,t}-\phi_1 x_{1,t-1})-0)^2/\sigma_{1,\epsilon}^2\}} \\
&= \prod_{t=s-l+1}^{s} exp\ -\frac{1}{2}\{-2c_1(x_{1,t}-\phi_1 x_{1,t-1})/\sigma_{1,\epsilon}^2+c_1^2/\sigma_{1,\epsilon}^2\} \\
&= exp\{\sum_{t=s-l+1}^{s}(c_1(x_{1,t}-\phi_1 x_{1,t-1})/\sigma_{1,\epsilon}^2)-\frac{1}{2}\sum_{t=s-l+1}^{s}(c_1^2/\sigma_{1,\epsilon}^2)\}
\end{aligned}$$

By factorization theorem, the likelihood ratio statistics $T_t(\boldsymbol{x}_t)$ for detecting a mean shift with time lag for $t = s - l + 1, ..., s$ are

$$T_t(\boldsymbol{x}_t) = c_1(x_{1,t} - \phi_1 x_{1,t-1})/\sigma^2_{1,\epsilon}.$$

Table A.1: Results of scenario 1 (CBS but no CWS) for detecting simultaneous changes

CWS = 0 LCP = 0		Estimate	Method		
Shift size	CBS		Parallel	Frisén	Wessman
2	$\rho = 0.2$	$\widehat{\text{ced}}$	2.284	1.999	1.929
		$\widehat{\text{tar}}$	94.0	94.3	94.7
		$\widehat{\text{far}}$	5.6	5.3	5.1
		$\widehat{\text{ndr}}$	0.4	0.4	0.2
	$\rho = 0.4$	$\widehat{\text{ced}}$	2.252	2.075	1.929
		$\widehat{\text{tar}}$	94.7	94.1	94.7
		$\widehat{\text{far}}$	4.9	5.4	5.1
		$\widehat{\text{ndr}}$	0.4	0.5	0.2
	$\rho = 0.6$	$\widehat{\text{ced}}$	2.215	2.066	1.929
		$\widehat{\text{tar}}$	94.8	94.0	94.7
		$\widehat{\text{far}}$	4.9	5.5	5.1
		$\widehat{\text{ndr}}$	0.3	0.5	0.2
3	$\rho = 0.2$	$\widehat{\text{ced}}$	0.978	0.852	0.804
		$\widehat{\text{tar}}$	94.4	94.7	94.9
		$\widehat{\text{far}}$	5.6	5.3	5.1
		$\widehat{\text{ndr}}$	0.0	0.0	0.0
	$\rho = 0.4$	$\widehat{\text{ced}}$	0.966	0.916	0.803
		$\widehat{\text{tar}}$	95.1	94.6	94.9
		$\widehat{\text{far}}$	4.9	5.4	5.1
		$\widehat{\text{ndr}}$	0.0	0.0	0.0
	$\rho = 0.6$	$\widehat{\text{ced}}$	0.942	0.943	0.804
		$\widehat{\text{tar}}$	95.1	94.5	94.9
		$\widehat{\text{far}}$	4.9	5.5	5.1
		$\widehat{\text{ndr}}$	0.0	0.0	0.0

Table A.2: Results of scenario 1 (CBS but no CWS) for detecting changes with time lag)

No CWS ($\phi = 0$)		Estimate	Method			
Shift size	CBS		Parallel	Frisén	Wessman	Case 1
2	$\rho = 0.2$	$\widehat{\text{ced}}$	2.295	1.926	1.931	1.924
		$\widehat{\text{tar}}$	94.5	94.1	93.4	94.0
		$\widehat{\text{far}}$	5.1	5.5	6.2	5.6
		$\widehat{\text{ndr}}$	0.4	0.4	0.3	0.4
	$\rho = 0.4$	$\widehat{\text{ced}}$	2.284	1.923	2.160	1.919
		$\widehat{\text{tar}}$	95.1	93.9	93.6	93.9
		$\widehat{\text{far}}$	4.5	5.7	5.7	5.7
		$\widehat{\text{ndr}}$	0.4	0.4	0.7	0.4
	$\rho = 0.6$	$\widehat{\text{ced}}$	2.287	1.924	2.690	1.917
		$\widehat{\text{tar}}$	94.9	93.8	92.5	93.9
		$\widehat{\text{far}}$	4.7	5.8	6.1	5.7
		$\widehat{\text{ndr}}$	0.4	0.4	0.4	0.4
3	$\rho = 0.2$	$\widehat{\text{ced}}$	0.982	0.806	0.799	0.803
		$\widehat{\text{tar}}$	95.0	94.5	93.7	94.4
		$\widehat{\text{far}}$	5.0	5.5	6.2	5.6
		$\widehat{\text{ndr}}$	0	0	0	0
	$\rho = 0.4$	$\widehat{\text{ced}}$	0.977	0.810	0.934	0.808
		$\widehat{\text{tar}}$	95.5	94.3	94.3	94.3
		$\widehat{\text{far}}$	4.5	5.7	5.7	5.7
		$\widehat{\text{ndr}}$	0	0	0	0
	$\rho = 0.6$	$\widehat{\text{ced}}$	0.979	0.810	1.228	0.807
		$\widehat{\text{tar}}$	95.3	94.2	92.9	94.3
		$\widehat{\text{far}}$	4.7	5.8	6.1	5.7
		$\widehat{\text{ndr}}$	0	0	0	0

Table A.3: Results of scenario 2 (CWS but no CBS) for detecting simultaneous changes

CBS = 0 LCP = 0		Estimate	Method			
Shift size	CWS		Parallel	Frisén	Wessman	Case 2
2	$\phi = 0.2$	$\widehat{\text{ced}}$	2.288	2.237	2.255	2.132
		$\widehat{\text{tar}}$	86.4	88.7	88.0	92.4
		$\widehat{\text{far}}$	12.6	10.6	11.3	6.2
		$\widehat{\text{ndr}}$	1.0	0.7	0.7	1.4
	$\phi = 0.4$	$\widehat{\text{ced}}$	2.611	2.463	2.481	2.139
		$\widehat{\text{tar}}$	75.2	80.1	79.7	85.2
		$\widehat{\text{far}}$	23.4	18.8	19.4	6.5
		$\widehat{\text{ndr}}$	1.4	1.1	0.9	8.3
	$\phi = 0.6$	$\widehat{\text{ced}}$	4.658	4.121	4.149	3.202
		$\widehat{\text{tar}}$	62.9	71.3	70.9	76.8
		$\widehat{\text{far}}$	35.6	27.2	27.7	6.3
		$\widehat{\text{ndr}}$	1.5	1.5	1.4	16.9
3	$\phi = 0.2$	$\widehat{\text{ced}}$	1.144	1.118	1.127	1.066
		$\widehat{\text{tar}}$	87.4	89.4	88.7	93.8
		$\widehat{\text{far}}$	12.6	10.6	11.3	6.2
		$\widehat{\text{ndr}}$	0.0	0.0	0.0	0.0
	$\phi = 0.4$	$\widehat{\text{ced}}$	1.305	1.231	1.241	1.069
		$\widehat{\text{tar}}$	76.6	81.2	80.6	93.4
		$\widehat{\text{far}}$	23.4	18.8	19.4	6.5
		$\widehat{\text{ndr}}$	0.0	0.0	0.0	0.1
	$\phi = 0.6$	$\widehat{\text{ced}}$	3.106	2.747	2.766	2.134
		$\widehat{\text{tar}}$	64.4	72.8	72.3	93.1
		$\widehat{\text{far}}$	35.6	27.2	27.7	6.3
		$\widehat{\text{ndr}}$	0.0	0.0	0.0	0.6

Table A.4: Results of scenario 2 (CWS but no CBS) for detecting changes with time lag

CBS = 0 LCP = 5		Estimate	Method			
Shift size	CWS		Parallel	Frisén	Wessman	Case 4
2	$\phi = 0.2$	$\widehat{\text{ced}}$	2.270	2.249	2.288	2.132
		$\widehat{\text{tar}}$	87.1	88.2	86.8	92.2
		$\widehat{\text{far}}$	11.9	11.1	12.6	6.2
		$\widehat{\text{ndr}}$	1.0	0.7	0.6	1.6
	$\phi = 0.4$	$\widehat{\text{ced}}$	2.581	2.469	2.528	2.143
		$\widehat{\text{tar}}$	76.1	79.9	78.0	85.8
		$\widehat{\text{far}}$	22.5	19.0	20.9	6.7
		$\widehat{\text{ndr}}$	1.4	1.1	1.1	7.5
	$\phi = 0.6$	$\widehat{\text{ced}}$	4.615	4.161	4.281	3.202
		$\widehat{\text{tar}}$	63.2	70.7	68.5	78.1
		$\widehat{\text{far}}$	35.0	27.9	30.1	6.3
		$\widehat{\text{ndr}}$	1.8	1.4	1.4	15.6
3	$\phi = 0.2$	$\widehat{\text{ced}}$	1.132	1.124	1.144	1.066
		$\widehat{\text{tar}}$	88.3	88.9	87.4	93.8
		$\widehat{\text{far}}$	11.7	11.1	12.6	6.2
		$\widehat{\text{ndr}}$	0.0	0.0	0.0	0.0
	$\phi = 0.4$	$\widehat{\text{ced}}$	1.285	1.235	1.264	1.071
		$\widehat{\text{tar}}$	77.8	81.0	79.1	93.3
		$\widehat{\text{far}}$	22.2	19.0	20.9	6.6
		$\widehat{\text{ndr}}$	0.0	0.0	0.0	0.1
	$\phi = 0.6$	$\widehat{\text{ced}}$	2.403	2.330	2.308	2.134
		$\widehat{\text{tar}}$	65.7	72.1	69.9	93.1
		$\widehat{\text{far}}$	34.3	27.9	30.1	6.3
		$\widehat{\text{ndr}}$	0.0	0.0	0.0	0.6

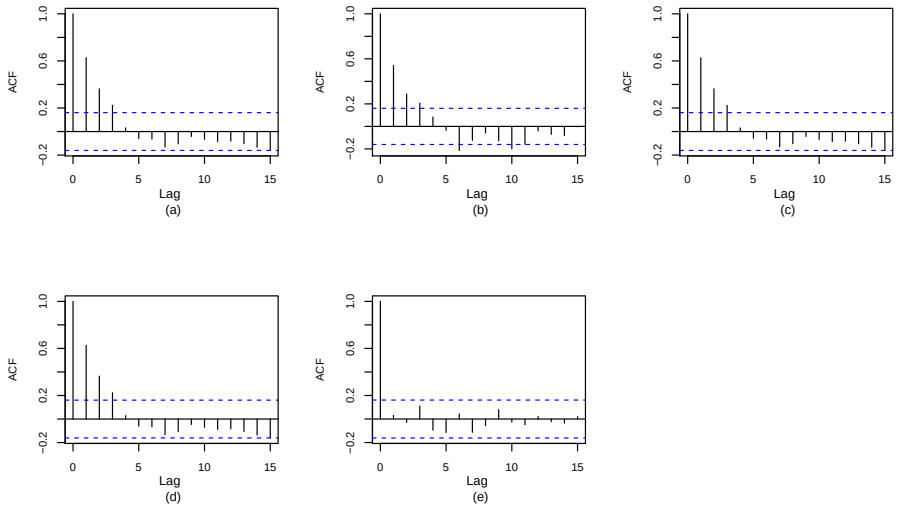

Figure A.1: a) and (b) ACF plots of X_1 and X_2 series and (c), (d) and (e) ACF plots of likelihood ratio statistics from Frisén, Wessman and and proposed method (case 2), respectively (scenario 2: simultaneous changes and shift size = 2).

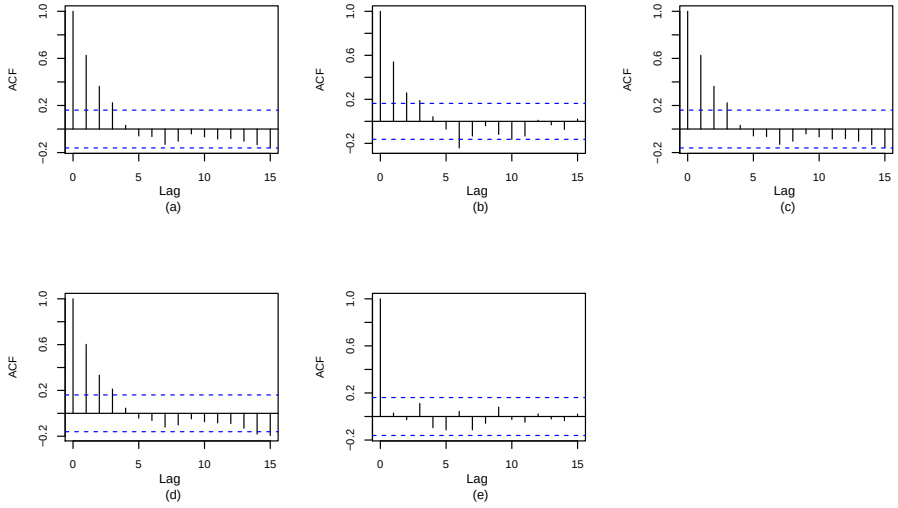

Figure A.2: a) and (b) ACF plots of X_1 and X_2 series and (c), (d) and (e) ACF plots of likelihood ratio statistics from Frisén, Wessman and and proposed method (case 4), respectively (scenario 2: changes with time lag and shift size = 2).

Table A.5: Results of scenario 3 (CWS and CBS) for detecting simultaneous changes

CWS = 0.6 LCP = 0		Estimate	Method			
Shift size	CBS		Parallel	Frisén	Wessman	Case 3
2	$\rho = 0.2$	$\widehat{\text{ced}}$	4.587	4.054	4.149	3.202
		$\widehat{\text{tar}}$	63.9	72.3	70.9	76.9
		$\widehat{\text{far}}$	34.6	26.0	27.7	6.3
		$\widehat{\text{ndr}}$	1.5	1.7	1.4	16.8
	$\rho = 0.4$	$\widehat{\text{ced}}$	4.451	4.048	4.149	3.202
		$\widehat{\text{tar}}$	65.9	71.9	70.9	77.1
		$\widehat{\text{far}}$	32.6	25.9	27.7	6.3
		$\widehat{\text{ndr}}$	1.5	2.2	1.4	16.6
	$\rho = 0.6$	$\widehat{\text{ced}}$	4.267	4.048	4.149	3.205
		$\widehat{\text{tar}}$	68.8	72.0	70.9	77.4
		$\widehat{\text{far}}$	29.7	25.9	27.7	6.4
		$\widehat{\text{ndr}}$	1.5	2.1	1.4	16.2
3	$\rho = 0.2$	$\widehat{\text{ced}}$	3.058	2.702	2.766	2.134
		$\widehat{\text{tar}}$	65.4	74.0	72.3	93.2
		$\widehat{\text{far}}$	34.6	26.0	27.7	6.3
		$\widehat{\text{ndr}}$	0.0	0.0	0.0	0.5
	$\rho = 0.4$	$\widehat{\text{ced}}$	2.967	2.699	2.766	2.134
		$\widehat{\text{tar}}$	67.4	74.1	72.3	93.2
		$\widehat{\text{far}}$	32.6	25.9	27.7	6.3
		$\widehat{\text{ndr}}$	0.0	0.0	0.0	0.5
	$\rho = 0.6$	$\widehat{\text{ced}}$	2.844	2.699	2.766	2.137
		$\widehat{\text{tar}}$	70.3	74.1	72.3	93.0
		$\widehat{\text{far}}$	29.7	25.9	27.7	6.4
		$\widehat{\text{ndr}}$	0.0	0.0	0.0	0.6

Table A.6: Results of scenario 3 (CWS and CBS) for detecting changes with time lag

CWS = 0.6 LCP = 5		Estimate	Method			
Shift size	Cross-CBS		Parallel	Frisén	Wessman	Case 5
2	$\rho = 0.2$	$\widehat{\text{ced}}$	4.538	4.115	4.285	3.202
		$\widehat{\text{tar}}$	64.4	71.5	68.5	78.1
		$\widehat{\text{far}}$	33.9	27.1	30.0	6.3
		$\widehat{\text{ndr}}$	1.7	1.4	1.5	15.6
	$\rho = 0.4$	$\widehat{\text{ced}}$	4.412	4.149	4.213	3.212
		$\widehat{\text{tar}}$	66.2	70.9	69.1	77.8
		$\widehat{\text{far}}$	32.0	27.7	28.8	6.6
		$\widehat{\text{ndr}}$	1.8	1.4	2.1	15.6
	$\rho = 0.6$	$\widehat{\text{ced}}$	4.129	4.138	4.189	3.219
		$\widehat{\text{tar}}$	69.3	71.1	69.2	77.6
		$\widehat{\text{far}}$	28.9	27.5	28.4	6.8
		$\widehat{\text{ndr}}$	1.8	1.4	2.4	15.6
3	$\rho = 0.2$	$\widehat{\text{ced}}$	3.003	2.743	2.857	2.134
		$\widehat{\text{tar}}$	66.6	72.9	69.9	93.1
		$\widehat{\text{far}}$	33.4	27.1	30.0	6.3
		$\widehat{\text{ndr}}$	0.0	0.0	0.1	0.6
	$\rho = 0.4$	$\widehat{\text{ced}}$	2.924	2.766	2.809	2.141
		$\widehat{\text{tar}}$	68.4	72.3	71.1	92.8
		$\widehat{\text{far}}$	31.6	27.7	28.8	6.6
		$\widehat{\text{ndr}}$	0.0	0.0	0.1	0.6
	$\rho = 0.6$	$\widehat{\text{ced}}$	2.801	2.758	2.793	2.146
		$\widehat{\text{tar}}$	71.4	72.5	71.4	92.6
		$\widehat{\text{far}}$	28.6	27.5	28.4	6.8
		$\widehat{\text{ndr}}$	0.0	0.0	0.2	0.6

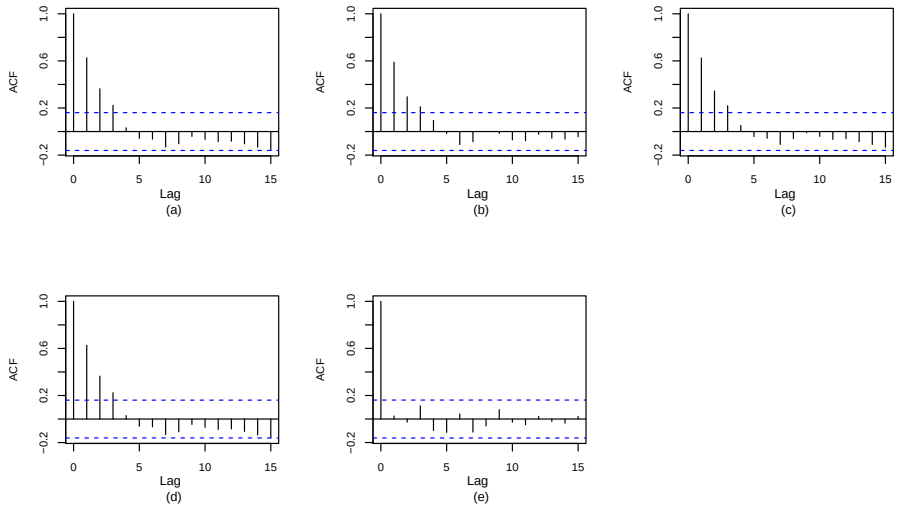

Figure A.3: a) and (b) ACF plots of X_1 and X_2 series and (c), (d) and (e) ACF plots of likelihood ratio statistics from Frisén, Wessman and and proposed method (case 3), respectively (scenario 3: simultaneous changes and shift size = 2).

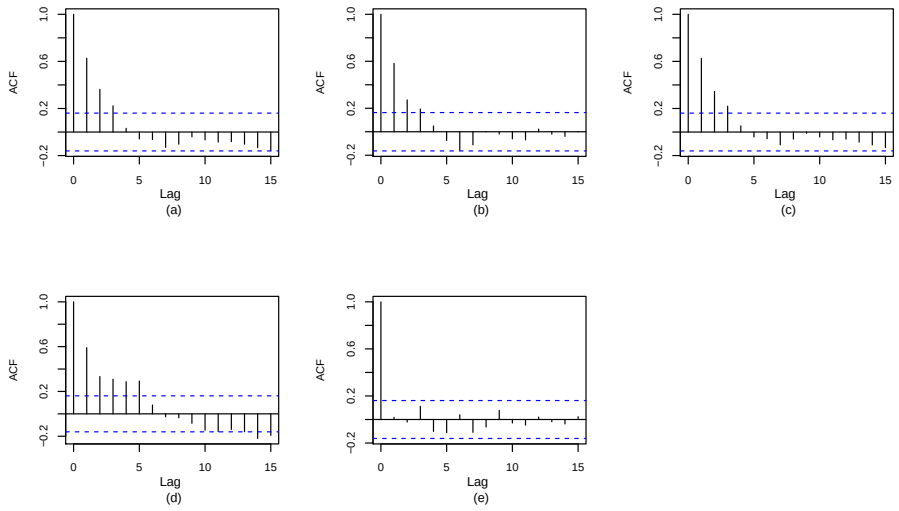

Figure A.4: a) and (b) ACF plots of X_1 and X_2 series and (c), (d) and (e) ACF plots of likelihood ratio statistics from Frisén, Wessman and and proposed method (case 5), respectively (scenario 3: changes with time lag and shift size = 2).

Table A.7: Critical values for scenario 1 (simultaneous changes)

CBS	Comparison	Critical value (method i - method j)		
		P - F	P - W	F - W
0.2	$CED_i - CED_j$	0.031	0.031	0.030
	$TAR_i - TAR_j$	0.60	0.58	0.56
	$FAR_i - FAR_j$	0.53	0.52	0.51
	$NDR_i - NDR_j$	0.30	0.27	0.24
0.4	$CED_i - CED_j$	0.031	0.031	0.031
	$TAR_i - TAR_j$	0.59	0.57	0.57
	$FAR_i - FAR_j$	0.52	0.50	0.52
	$NDR_i - NDR_j$	0.30	0.27	0.24
0.6	$CED_i - CED_j$	0.031	0.031	0.031
	$TAR_i - TAR_j$	0.59	0.57	0.57
	$FAR_i - FAR_j$	0.52	0.51	0.52
	$NDR_i - NDR_j$	0.29	0.26	0.24

Table A.8: Critical values for scenario 1 (changes with time lag)

CBS	Comparison	Critical value (method i - method j)					
		P - F	P - W	P - C1	F - W	F - C1	W - C1
0.2	$CED_i - CED_j$	0.031	0.031	0.031	0.031	0.030	0.031
	$TAR_i - TAR_j$	0.58	0.60	0.58	0.59	0.57	0.59
	$FAR_i - FAR_j$	0.52	0.54	0.52	0.55	0.53	0.55
	$NDR_i - NDR_j$	0.27	0.29	0.27	0.23	0.21	0.22
0.4	$CED_i - CED_j$	0.031	0.033	0.031	0.033	0.031	0.033
	$TAR_i - TAR_j$	0.57	0.60	0.57	0.60	0.58	0.60
	$FAR_i - FAR_j$	0.51	0.51	0.51	0.54	0.54	0.54
	$NDR_i - NDR_j$	0.28	0.31	0.27	0.27	0.21	0.26
0.6	$CED_i - CED_j$	0.031	0.036	0.031	0.035	0.031	0.035
	$TAR_i - TAR_j$	0.58	0.63	0.58	0.63	0.58	0.63
	$FAR_i - FAR_j$	0.52	0.55	0.52	0.57	0.54	0.57
	$NDR_i - NDR_j$	0.28	0.33	0.27	0.28	0.21	0.28

Appendix B

The extension of SR methods for normal data

Table B.1: Results of scenario 2 (CWS but no CBS) for detecting simultaneous changes, monitored with modified one-sided EWMA chart for autocorrelated data

Shift size	CWS	Estimate	Method			
			Parallel	Frisén	Wessman	Case 2
2	0.2	$\widehat{\text{ced}}$	3.158	2.117	2.119	2.119
		$\widehat{\text{tar}}$	90.1	93.3	93.0	92.3
		$\widehat{\text{far}}$	5.0	5.1	5.6	5.6
		$\widehat{\text{ndr}}$	4.9	1.6	1.4	2.1
	0.4	$\widehat{\text{ced}}$	3.198	3.175	3.195	2.218
		$\widehat{\text{tar}}$	83.0	88.2	87.9	85.3
		$\widehat{\text{far}}$	6.2	5.5	6.1	5.3
		$\widehat{\text{ndr}}$	10.8	6.3	6.0	9.4
	0.6	$\widehat{\text{ced}}$	4.329	4.264	4.301	3.181
		$\widehat{\text{tar}}$	75.6	81.1	80.5	76.0
		$\widehat{\text{far}}$	7.6	6.2	7.0	5.7
		$\widehat{\text{ndr}}$	16.8	12.7	12.5	18.3
3	0.2	$\widehat{\text{ced}}$	1.053	1.054	1.059	1.059
		$\widehat{\text{tar}}$	95.0	94.9	94.4	94.4
		$\widehat{\text{far}}$	5.0	5.1	5.6	5.6
		$\widehat{\text{ndr}}$	0.0	0.0	0.0	0.0
	0.4	$\widehat{\text{ced}}$	2.186	2.116	2.130	1.056
		$\widehat{\text{tar}}$	91.4	94.4	93.9	94.6
		$\widehat{\text{far}}$	8.5	5.5	6.1	5.3
		$\widehat{\text{ndr}}$	0.1	0.1	0.0	0.1
	0.6	$\widehat{\text{ced}}$	3.247	2.132	2.150	2.121
		$\widehat{\text{tar}}$	91.7	93.5	92.9	93.5
		$\widehat{\text{far}}$	7.6	6.3	7.0	5.7
		$\widehat{\text{ndr}}$	0.7	0.2	0.1	0.8

Table B.2: Results of scenario 2 (CWS but no CBS) for detecting changes with time lag, monitored with modified one-sided EWMA chart for autocorrelated data

Shift size	CWS	Estimate	Method			
			Parallel	Frisén	Wessman	Case 4
2	0.2	$\widehat{\text{ced}}$	3.148	2.105	2.148	2.116
		$\widehat{\text{tar}}$	90.3	93.3	91.6	92.3
		$\widehat{\text{far}}$	4.7	5.0	6.9	5.5
		$\widehat{\text{ndr}}$	5.0	1.7	1.5	2.2
	0.4	$\widehat{\text{ced}}$	3.198	3.175	3.226	2.304
		$\widehat{\text{tar}}$	82.9	88.1	87.5	85.8
		$\widehat{\text{far}}$	6.2	5.5	7.0	5.3
		$\widehat{\text{ndr}}$	10.9	6.4	5.5	8.9
	0.6	$\widehat{\text{ced}}$	4.324	4.267	4.315	3.185
		$\widehat{\text{tar}}$	75.6	80.9	81.2	76.5
		$\widehat{\text{far}}$	7.5	6.3	7.3	5.8
		$\widehat{\text{ndr}}$	16.9	12.8	11.5	17.7
3	0.2	$\widehat{\text{ced}}$	1.048	1.053	1.074	1.058
		$\widehat{\text{tar}}$	95.4	95.0	93.1	94.5
		$\widehat{\text{far}}$	4.6	5.0	6.9	5.5
		$\widehat{\text{ndr}}$	0.0	0.0	0.0	0.0
	0.4	$\widehat{\text{ced}}$	2.121	2.116	2.1518	1.056
		$\widehat{\text{tar}}$	94.2	94.4	92.9	94.6
		$\widehat{\text{far}}$	5.7	5.5	7.0	5.3
		$\widehat{\text{ndr}}$	0.1	0.1	0.1	0.1
	0.6	$\widehat{\text{ced}}$	3.243	2.134	2.157	2.121
		$\widehat{\text{tar}}$	91.8	93.4	92.4	93.7
		$\widehat{\text{far}}$	7.5	6.3	7.3	5.7
		$\widehat{\text{ndr}}$	0.7	0.3	0.3	0.6

Table B.3: Results of scenario 3 (CWS and CBS) for detecting simultaneous changes, monitored with modified one-sided EWMA chart for autocorrelated data

Shift size	CWS	Estimate	Method			
			Parallel	Frisén	Wessman	Case 3
2	0.2	$\widehat{\text{ced}}$	4.324	4.293	4.231	3.181
		$\widehat{\text{tar}}$	76.0	80.0	80.5	76.0
		$\widehat{\text{far}}$	7.5	6.4	7.0	5.7
		$\widehat{\text{ndr}}$	16.5	13.6	12.5	18.3
	0.4	$\widehat{\text{ced}}$	4.329	4.299	4.231	3.168
		$\widehat{\text{tar}}$	76.6	79.1	80.5	76.6
		$\widehat{\text{far}}$	7.6	6.3	7.0	5.3
		$\widehat{\text{ndr}}$	15.8	14.6	12.5	18.1
	0.6	$\widehat{\text{ced}}$	4.320	4.299	4.231	3.168
		$\widehat{\text{tar}}$	77.0	79.0	80.5	76.5
		$\widehat{\text{far}}$	7.4	6.3	7.0	5.3
		$\widehat{\text{ndr}}$	15.6	14.7	12.5	18.2
3	0.2	$\widehat{\text{ced}}$	3.243	3.205	2.151	2.121
		$\widehat{\text{tar}}$	91.8	93.2	92.9	93.5
		$\widehat{\text{far}}$	7.5	6.4	7.0	5.7
		$\widehat{\text{ndr}}$	0.7	0.4	0.1	0.8
	0.4	$\widehat{\text{ced}}$	3.248	3.202	2.151	2.112
		$\widehat{\text{tar}}$	91.7	93.0	92.9	94.1
		$\widehat{\text{far}}$	7.6	6.3	7.0	5.3
		$\widehat{\text{ndr}}$	0.7	0.7	0.1	0.6
	0.6	$\widehat{\text{ced}}$	3.240	3.204	2.151	2.112
		$\widehat{\text{tar}}$	91.9	93.0	92.9	94.1
		$\widehat{\text{far}}$	7.4	6.3	7.0	5.3
		$\widehat{\text{ndr}}$	0.7	0.7	0.1	0.6

Table B.4: Results of scenario 3 (CWS and CBS) for detecting changes with time lag, monitored with modified one-sided EWMA chart for autocorrelated data

Shift size	CWS	Estimate	Method			
			Parallel	Frisén	Wessman	Case 5
2	0.2	$\widehat{\text{ced}}$	4.320	4.292	4.357	3.178
		$\widehat{\text{tar}}$	75.7	80.5	79.8	76.7
		$\widehat{\text{far}}$	7.4	6.8	8.2	5.6
		$\widehat{\text{ndr}}$	16.9	12.7	12.0	17.7
	0.4	$\widehat{\text{ced}}$	4.320	4.302	4.367	3.171
		$\widehat{\text{tar}}$	75.9	79.9	80.1	76.8
		$\widehat{\text{far}}$	7.4	7.0	8.4	5.4
		$\widehat{\text{ndr}}$	16.7	13.1	11.5	17.8
	0.6	$\widehat{\text{ced}}$	4.306	4.301	5.507	3.171
		$\widehat{\text{tar}}$	76.1	79.9	78.2	76.7
		$\widehat{\text{far}}$	7.1	7.0	9.2	5.4
		$\widehat{\text{ndr}}$	16.8	13.1	12.6	17.9
3	0.2	$\widehat{\text{ced}}$	3.240	2.146	2.179	2.119
		$\widehat{\text{tar}}$	91.9	92.9	91.3	93.7
		$\widehat{\text{far}}$	7.4	6.8	8.2	5.6
		$\widehat{\text{ndr}}$	0.7	0.3	0.5	0.7
	0.4	$\widehat{\text{ced}}$	3.236	2.150	3.275	2.114
		$\widehat{\text{tar}}$	92.0	92.7	90.9	94.0
		$\widehat{\text{far}}$	7.3	7.0	8.4	5.4
		$\widehat{\text{ndr}}$	0.7	0.3	0.7	0.6
	0.6	$\widehat{\text{ced}}$	3.226	2.151	3.304	2.114
		$\widehat{\text{tar}}$	92.3	92.7	89.9	94.0
		$\widehat{\text{far}}$	7.0	7.0	9.2	5.4
		$\widehat{\text{ndr}}$	0.7	0.3	0.9	0.6

Table B.5: Scenario R1 (mis-specification of CWS: CWS = 0.6, CBS = 0 and LCP = 0)

Method	Shift size	Estimate	Specified CWS = 0.6	Estimated CWS	Mis-specified CWS			
					0.0	0.2	0.4	0.8
Case 2	2	$\widehat{\text{ced}}$	2.482	2.542	3.140	2.903	2.656	2.352
		$\widehat{\text{tar}}$	63.3	64.0	60.3	61.8	63.2	23.8
		$\widehat{\text{far}}$	5.3	6.3	27.7	23.3	14.7	1.6
		$\widehat{\text{ndr}}$	31.4	29.7	12.0	14.9	22.1	74.6
	3	$\widehat{\text{ced}}$	1.930	1.908	2.292	2.127	1.985	2.042
		$\widehat{\text{tar}}$	87.8	86.5	71.4	75.7	83.4	61.2
		$\widehat{\text{far}}$	5.3	6.3	27.7	23.3	14.7	1.6
		$\widehat{\text{ndr}}$	6.9	7.2	0.9	1.0	1.9	37.2

Table B.6: Scenario R2 (mis-specification of CWS: CWS = 0.6, CBS = 0 and LCP = 5)

Method	Shift size	Estimate	Specified CWS = 0.6	Estimated CWS	Mis-specified CWS			
					0.0	0.2	0.4	0.8
Case 4	2	$\widehat{\text{ced}}$	2.502	2.538	3.169	2.928	2.681	2.338
		$\widehat{\text{tar}}$	63.5	64.5	60.2	62.1	63.1	23.1
		$\widehat{\text{far}}$	5.5	6.3	28.1	23.5	15.3	1.7
		$\widehat{\text{ndr}}$	31.0	29.2	11.7	14.4	21.6	75.2
	3	$\widehat{\text{ced}}$	1.928	1.904	2.314	2.119	1.963	2.034
		$\widehat{\text{tar}}$	87.6	87.0	70.8	75.5	82.7	62.2
		$\widehat{\text{far}}$	5.5	6.3	28.1	23.4	15.8	1.6
		$\widehat{\text{ndr}}$	6.9	6.7	1.1	1.1	2.0	36.2

Table B.7: Scenario R3 (mis-specification of CWS: CWS = 0.6, CBS = 0.6 and LCP = 0)

Method	Shift size	Estimate	Specified CWS = 0.6	Estimated CWS	Mis-specified CWS			
					0.0	0.2	0.4	0.8
Case 3	2	$\widehat{\text{ced}}$	2.482	2.539	3.140	2.903	2.656	2.352
		$\widehat{\text{tar}}$	63.3	64.5	60.3	61.8	63.2	23.8
		$\widehat{\text{far}}$	5.3	6.4	27.7	23.3	14.7	1.6
		$\widehat{\text{ndr}}$	31.4	29.1	12.0	14.9	22.1	74.6
	3	$\widehat{\text{ced}}$	1.930	1.897	2.292	2.127	1.945	2.042
		$\widehat{\text{tar}}$	87.8	86.4	71.4	75.7	83.4	61.2
		$\widehat{\text{far}}$	5.3	6.4	27.7	23.3	14.7	1.6
		$\widehat{\text{ndr}}$	6.9	7.2	0.9	1.0	1.9	37.2

Table B.8: Scenario R4 (mis-specification of CWS: CWS = 0.6, CBS = 0.6 and LCP = 5)

Method	Shift size	Estimate	Specified CWS = 0.6	Estimated CWS	Mis-specified CWS			
					0.0	0.2	0.4	0.8
Case 5	2	$\widehat{\text{ced}}$	2.508	2.564	3.187	2.936	2.670	2.438
		$\widehat{\text{tar}}$	63.5	64.2	59.9	62.2	63.1	23.0
		$\widehat{\text{far}}$	5.8	6.8	28.5	23.7	15.1	1.8
		$\widehat{\text{ndr}}$	30.7	29.0	11.6	14.3	21.7	75.2
	3	$\widehat{\text{ced}}$	1.953	1.925	2.324	2.124	1.942	2.036
		$\widehat{\text{tar}}$	87.3	86.5	70.4	75.3	82.9	62.2
		$\widehat{\text{far}}$	5.8	6.8	28.5	23.6	15.2	1.7
		$\widehat{\text{ndr}}$	6.9	6.7	1.1	1.1	1.9	36.1

Table B.9: Scenario R5 (mis-specification of CBS: CBS = 0.6, CWS = 0 and LCP = 0)

Method	Shift size	Estimate	Specified CBS = 0.6	Estimated CBS	Mis-specified CBS			
					0.0	0.2	0.4	0.8
Wessman	2	$\widehat{\text{ced}}$	1.771	1.772	1.908	1.871	1.844	1.876
		$\widehat{\text{tar}}$	90.5	91.1	88.7	89.3	90.1	89.8
		$\widehat{\text{far}}$	5.3	5.1	5.5	5.3	5.2	5.7
		$\widehat{\text{ndr}}$	4.5	3.8	5.8	5.4	4.7	4.5
	3	$\widehat{\text{ced}}$	0.863	0.804	0.941	0.913	0.816	0.861
		$\widehat{\text{tar}}$	94.7	94.9	94.5	94.7	94.8	94.3
		$\widehat{\text{far}}$	5.3	5.1	5.5	5.3	5.2	5.7
		$\widehat{\text{ndr}}$	0.0	0.0	0.0	0.0	0.0	0.0

Table B.10: Scenario R6 (mis-specification of CBS: CBS = 0.6, CWS = 0 and LCP = 5)

Method	Shift size	Estimate	Specified CBS = 0.6	Estimated CBS	Mis-specified CBS			
					0.0	0.2	0.4	0.8
Case 1	2	$\widehat{\text{ced}}$	1.771	1.763	1.835	1.803	1.789	1.796
		$\widehat{\text{tar}}$	90.7	90.7	90.3	90.5	90.4	90.0
		$\widehat{\text{far}}$	5.3	5.7	5.7	5.5	5.6	6.1
		$\widehat{\text{ndr}}$	4.0	3.6	4.0	4.0	4.0	3.9
	3	$\widehat{\text{ced}}$	0.802	0.799	0.822	0.812	0.807	0.816
		$\widehat{\text{tar}}$	94.7	94.3	94.3	94.5	94.4	94.0
		$\widehat{\text{far}}$	5.3	5.7	5.7	5.5	5.6	6.0
		$\widehat{\text{ndr}}$	0.0	0.0	0.0	0.0	0.0	0.0

Table B.11: Scenario R7 (mis-specification of CBS: CBS = 0.6, CWS = 0.6 and LCP = 0)

Method	Shift size	Estimate	Specified CBS = 0.6	Estimated CBS	Mis-specified CBS			
					0.0	0.2	0.4	0.8
Case 3	2	$\widehat{\text{ced}}$	2.511	2.500	2.586	2.559	2.549	2.538
		$\widehat{\text{tar}}$	55.2	55.5	50.8	51.6	53.4	52.8
		$\widehat{\text{far}}$	6.8	7.1	7.3	7.3	6.8	7.4
		$\widehat{\text{ndr}}$	38.0	37.4	41.9	41.1	39.8	39.8
	3	$\widehat{\text{ced}}$	1.928	1.921	2.043	2.015	1.960	1.955
		$\widehat{\text{tar}}$	86.1	85.6	82.9	83.8	85.6	83.1
		$\widehat{\text{far}}$	6.8	7.1	7.3	7.3	6.8	7.4
		$\widehat{\text{ndr}}$	7.1	7.3	9.8	8.9	7.6	9.5

Table B.12: Scenario R8 (mis-specification of CBS: CBS = 0.6, CWS = 0.6 and LCP = 5)

Method	Shift size	Estimate	Specified CBS = 0.6	Estimated CBS	Mis-specified CBS			
					0.0	0.2	0.4	0.8
Case 5	2	$\widehat{\text{ced}}$	2.516	2.514	2.519	2.519	2.513	2.521
		$\widehat{\text{tar}}$	55.1	55.5	55.6	55.4	55.9	55.6
		$\widehat{\text{far}}$	7.1	7.1	8.4	8.2	7.6	7.7
		$\widehat{\text{ndr}}$	37.8	37.3	36.0	36.4	36.5	36.7
	3	$\widehat{\text{ced}}$	1.919	1.926	1.931	1.929	1.923	1.925
		$\widehat{\text{tar}}$	85.3	85.3	84.2	84.3	84.9	85.0
		$\widehat{\text{far}}$	6.9	7.0	8.2	8.0	7.5	7.4
		$\widehat{\text{ndr}}$	7.8	7.7	7.6	7.7	7.6	7.6

Table B.13: Scenario R9 (mis-specification of LCP: LCP = 5, CBS = 0.6 and CWS = 0)

Method	Shift size	Estimate	Specified LCP = 5	Estimated LCP	Mis-specified LCP				
					0	3	4	6	7
Case 1	2	$\widehat{\text{ced}}$	1.779	1.779	2.207	1.965	1.872	1.911	1.936
		$\widehat{\text{tar}}$	90.2	90.2	77.4	85.7	82.4	83.8	82.7
		$\widehat{\text{far}}$	5.7	5.7	6.7	10.5	13.8	12.4	13.8
		$\widehat{\text{ndr}}$	4.1	4.1	15.9	3.8	3.8	3.8	3.5
	3	$\widehat{\text{ced}}$	0.807	0.807	1.238	0.903	0.860	0.857	0.877
		$\widehat{\text{tar}}$	94.3	94.3	92.5	89.5	86.3	86.4	89.1
		$\widehat{\text{far}}$	5.7	5.7	6.7	10.5	13.7	13.6	10.9
		$\widehat{\text{ndr}}$	0.0	0.0	0.8	0.0	0.0	0.0	0.0

Table B.14: Scenario R10 (mis-specification of LCP: LCP = 5, CBS = 0.6 and CWS = 0.6)

Method	Shift size	Estimate	Specified LCP = 5	Estimated LCP	Mis-specified LCP				
					0	3	4	6	7
Case 5	2	$\widehat{\text{ced}}$	2.564	2.564	2.676	2.673	2.626	2.716	2.695
		$\widehat{\text{tar}}$	54.2	54.2	39.4	52.9	51.7	50.5	51.1
		$\widehat{\text{far}}$	6.8	6.8	7.4	11.9	13.6	13.6	12.0
		$\widehat{\text{ndr}}$	39.0	39.0	53.2	35.2	34.7	35.9	36.9
	3	$\widehat{\text{ced}}$	1.925	1.925	2.275	2.027	1.967	2.053	2.044
		$\widehat{\text{tar}}$	86.5	86.5	72.9	82.2	80.5	80.9	82.2
		$\widehat{\text{far}}$	6.8	6.8	7.4	11.8	13.2	12.9	11.6
		$\widehat{\text{ndr}}$	6.7	6.7	19.7	6.0	6.3	6.2	6.2

Table B.15: Scenario R11 (specification of shift size: CBS = 0, CWS = 0 and LCP = 0)

Method	Estimate	Shift size in the process				
		0.5	1	2	3	4
Parallel	$\widehat{\text{ced}}$	4.289	4.143	2.270	0.986	0.454
	$\widehat{\text{tar}}$	12.8	46.4	92.9	94.0	94.0
	$\widehat{\text{far}}$	6.0	6.0	6.0	6.0	6.0
	$\widehat{\text{ndr}}$	81.2	47.6	1.1	0	0
Frisén	$\widehat{\text{ced}}$	4.203	3.979	1.921	0.815	0.370
	$\widehat{\text{tar}}$	16.9	55.2	93.8	94.6	94.6
	$\widehat{\text{far}}$	5.4	5.4	5.4	5.4	5.4
	$\widehat{\text{ndr}}$	77.7	39.4	0.8	0	0

Table B.16: Scenario R12 (specification of shift size: CBS = 0.6, CWS = 0 and LCP = 0)

Method	Estimate	Shift size in the process				
		0.5	1	2	3	4
Parallel	$\widehat{\text{ced}}$	4.292	4.083	2.182	0.942	0.426
	$\widehat{\text{tar}}$	14.7	50.4	94.3	95.1	95.1
	$\widehat{\text{far}}$	4.9	4.9	4.9	4.9	4.9
	$\widehat{\text{ndr}}$	80.4	44.7	0.8	0	0
Wessman	$\widehat{\text{ced}}$	4.209	3.913	1.895	0.804	0.322
	$\widehat{\text{tar}}$	17.9	56.2	94.2	94.9	94.9
	$\widehat{\text{far}}$	5.1	5.1	5.1	5.1	5.1
	$\widehat{\text{ndr}}$	77.0	38.7	0.7	0	0

Table B.17: Scenario R13 (specification of shift size: CBS = 0, CWS = 0.6 and LCP = 0)

Method	Estimate	Shift size in the process				
		0.5	1	2	3	4
Parallel	$\widehat{\text{ced}}$	4.355	4.817	4.423	3.092	2.152
	$\widehat{\text{tar}}$	8.3	21.5	71.2	90.8	92.4
	$\widehat{\text{far}}$	7.6	7.6	7.6	7.6	7.6
	$\widehat{\text{ndr}}$	84.1	70.9	21.2	1.6	0
Case 2	$\widehat{\text{ced}}$	4.057	4.241	3.547	2.177	1.223
	$\widehat{\text{tar}}$	10.6	27.0	74.1	92.8	93.7
	$\widehat{\text{far}}$	6.3	6.3	6.3	6.3	6.3
	$\widehat{\text{ndr}}$	83.1	66.7	19.6	0.9	0

Table B.18: Scenario R14 (specification of shift size: CBS = 0.6, CWS = 0.6 and LCP = 0)

Method	Estimate	Shift size in the process				
		0.5	1	2	3	4
Parallel	$\widehat{\text{ced}}$	4.164	4.728	4.302	2.999	2.086
	$\widehat{\text{tar}}$	9.7	23.8	73.2	91.3	92.6
	$\widehat{\text{far}}$	7.4	7.4	7.4	7.4	7.4
	$\widehat{\text{ndr}}$	82.9	68.8	19.4	1.3	0
Case 4	$\widehat{\text{ced}}$	3.980	4.234	3.543	2.160	1.218
	$\widehat{\text{tar}}$	10.9	27.3	74.6	92.5	93.6
	$\widehat{\text{far}}$	6.4	6.4	6.4	6.4	6.4
	$\widehat{\text{ndr}}$	82.7	66.3	19.0	1.1	0

Appendix C

SR methods for Poisson data

C.1 Sufficient statistics for case 1 (no CWS or CBS)

From equation (6.34), $T_t(\boldsymbol{x}_t)$ can be derived from

$$\begin{aligned}
\sum_{t=\tau}^{s} lr(\boldsymbol{x}_t; \mu_1^O - \mu_1^I) &= \sum_{t=\tau}^{s} ln\left(\frac{f(\boldsymbol{x}_t; \mu_1^O, \mu_2^I)}{f(\boldsymbol{x}_t; \mu_1^I, \mu_2^I)}\right) \\
&= \sum_{t=\tau}^{s} ln\left(\frac{exp(-\mu_1^I - \mu_2^I - c)(\mu_1^I + c)^{x_{1,t}}(\mu_2^I)^{x_{2,t}}/x_{1,t}!x_{2,t}!}{exp(-\mu_1^I - \mu_2^I)(\mu_1^I)^{x_{1,t}}(\mu_2^I)^{x_{2,t}}/x_{1,t}!x_{2,t}!}\right) \\
&= \sum_{t=\tau}^{s} ln\ exp(-c)\left(\frac{\mu_1^I + c}{\mu_1^I}\right)^{x_{1,t}} \\
&= \sum_{t=\tau}^{s} -c + x_{1,t}\ ln\left(\frac{\mu_1^I + c}{\mu_1^I}\right) \qquad \text{(C.1)}
\end{aligned}$$

By the factorization theorem, the likelihood ratio statistics $T_t(\boldsymbol{x}_t)$ derived from equation (C.1) are

$$T_t(\boldsymbol{x}_t) = x_{1,t}\ ln\left(\frac{\mu_1^I + c}{\mu_1^I}\right),\ t = 1, 2, ..., s$$

C.2 Sufficient statistics for case 2 (no CWS but CBS)

From equation (6.36), $T_t(\boldsymbol{x}_t)$ can be derived from

$$\begin{aligned}
\sum_{t=\tau}^{s} lr(\boldsymbol{x}_t; \mu_3^O - \mu_3^I) &= \sum_{t=\tau}^{s} ln\left(\frac{f(\boldsymbol{x}_t; \mu_1^I, \mu_2^I, \mu_3^O)}{f(\boldsymbol{x}_t; \mu_1^I, \mu_2^I, \mu_3^I)}\right) \\
&= \sum_{t=\tau}^{s} ln\left(\frac{e^{-(\mu_1^I+\mu_2^I+\mu_3^I+c)}\frac{(\mu_1^I)^{x_{1,t}}}{x_{1,t}!}\frac{(\mu_2^I)^{x_{2,t}}}{x_{2,t}!}\sum_{j=0}^{m}\binom{x_{1,t}}{j}\binom{x_{2,t}}{j} j!\left(\frac{\mu_3^I+c}{\mu_1^I\mu_2^I}\right)^j}{e^{-(\mu_1^I+\mu_2^I+\mu_3^I)}\frac{(\mu_1^I)^{x_{1,t}}}{x_{1,t}!}\frac{(\mu_2^I)^{x_{2,t}}}{x_{2,t}!}\sum_{j=0}^{m}\binom{x_{1,t}}{j}\binom{x_{2,t}}{j} j!\left(\frac{\mu_3^I}{\mu_1^I\mu_2^I}\right)^j}\right) \\
&= \sum_{t=\tau}^{s} ln\left(exp(-c)\frac{\sum_{j=0}^{m}\binom{x_{1,t}}{j}\binom{x_{2,t}}{j} j!\left(\frac{\mu_3^I+c}{\mu_1^I\mu_2^I}\right)^j}{\sum_{j=0}^{m}\binom{x_{1,t}}{j}\binom{x_{2,t}}{j} j!\left(\frac{\mu_3^I}{\mu_1^I\mu_2^I}\right)^j}\right) \\
&= \sum_{t=\tau}^{s} -c + ln\left(\frac{\sum_{j=0}^{m}\binom{x_{1,t}}{j}\binom{x_{2,t}}{j} j!\left(\frac{\mu_3^I+c}{\mu_1^I\mu_2^I}\right)^j}{\sum_{j=0}^{m}\binom{x_{1,t}}{j}\binom{x_{2,t}}{j} j!\left(\frac{\mu_3^I}{\mu_1^I\mu_2^I}\right)^j}\right) \qquad \text{(C.2)}
\end{aligned}$$

where $m = min(x_{1,t}, x_{2,t})$. By the factorization theorem, the likelihood ratio statistics $T_t(\boldsymbol{x}_t)$ derived according to the change in covariance $(\mu_3^O - \mu_3^I)$ are

$$T_t(\boldsymbol{x}_t) = ln\left(\frac{\sum_{j=0}^{m}\binom{x_{1,t}}{j}\binom{x_{2,t}}{j} j!\left(\frac{\mu_3^I+c}{\mu_1^I\mu_2^I}\right)^j}{\sum_{j=0}^{m}\binom{x_{1,t}}{j}\binom{x_{2,t}}{j} j!\left(\frac{\mu_3^I}{\mu_1^I\mu_2^I}\right)^j}\right)$$

.

C.3 Sufficient statistics for case 3 (CWS but no CBS)

From equation (6.38), $T_t(\boldsymbol{x}_t)$ can be derived from

$$\begin{aligned}
\sum_{t=\tau}^{s} lr(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \mu_1^O - \mu_1^I) &= \sum_{t=\tau}^{s} ln\left(\frac{f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \mu_1^O, \mu_2^I)}{f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \mu_1^I, \mu_2^I)}\right) \\
&= \sum_{t=\tau}^{s} ln\left(\frac{\sum_{j=0}^{g_1}\binom{x_{1,t-1}}{j}\phi_1^j(1-\phi_1)^{x_{1,t-1}-j}\cdot\frac{e^{-(\mu_1^I+c)}(\mu_1^I+c)^{x_{1,t}-j}}{(x_{1,t}-j)!}}{\sum_{j=0}^{g_1}\binom{x_{1,t-1}}{j}\phi_1^j(1-\phi_1)^{x_{1,t-1}-j}\cdot\frac{e^{-\mu_1^I}(\mu_1^I)^{x_{1,t}-j}}{(x_{1,t}-j)!}}\right) \\
&= \sum_{t=\tau}^{s} -c + ln\left(\frac{\sum_{j=0}^{g_1}\binom{x_{1,t-1}}{j}\phi_1^j(1-\phi_1)^{x_{1,t-1}-j}\cdot\frac{(\mu_1^I+c)^{x_{1,t}-j}}{(x_{1,t}-j)!}}{\sum_{j=0}^{g_1}\binom{x_{1,t-1}}{j}\phi_1^j(1-\phi_1)^{x_{1,t-1}-j}\cdot\frac{(\mu_1^I)^{x_{1,t}-j}}{(x_{1,t}-j)!}}\right) \qquad \text{(C.3)}
\end{aligned}$$

By the factorization theorem, the likelihood ratio statistics $T_t(\boldsymbol{x}_t)$ derived from equation (C.3) are

$$T_t(\boldsymbol{x}_t) = \frac{\sum_{j=0}^{g_1} \left(\begin{array}{c} x_{1,t-1} \\ j \end{array} \right) \phi_1^j (1-\phi_1)^{x_{1,t-1}-j} \cdot \frac{(\mu_1^I + c)^{x_{1,t}-j}}{(x_{1,t}-j)!}}{\sum_{j=0}^{g_1} \left(\begin{array}{c} x_{1,t-1} \\ j \end{array} \right) \phi_1^j (1-\phi_1)^{x_{1,t-1}-j} \cdot \frac{(\mu_1^I)^{x_{1,t}-j}}{(x_{1,t}-j)!}}, \; t = 1, 2, ..., s$$

C.4 Sufficient statistics for case 4 (CWS and CBS)

From equation (6.40), $T_t(\boldsymbol{x}_t)$ can be derived from

$$\begin{aligned} \sum_{t=\tau}^{s} lr(f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \mu_3^O - \mu_3^I) &= \sum_{t=\tau}^{s} ln \, \frac{f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \phi_1, \phi_2, \mu_1^I, \mu_2^I, \mu_3^O)}{f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \phi_1, \phi_2, \mu_1^I, \mu_2^I, \mu_3^I)} \\ &= \sum_{t=\tau}^{s} -c + ln \, \frac{v_t}{w_t} \end{aligned}$$

where

$$\begin{aligned} f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \phi_1, \phi_2 \mu_1^I, \mu_2^I, \mu_3^I) = \sum_{r_1=0}^{g_1} \sum_{r_2=0}^{g_2} \Bigg\{ & \left(\begin{array}{c} x_{1,t-1} \\ x_{1,t} - r_1 \end{array} \right) \phi_1^{x_{1,t}-r_1} (1-\phi_1)^{x_{1,t-1}-x_{1,t}-r_1} \\ & \cdot \left(\begin{array}{c} x_{2,t-1} \\ x_{2,t} - r_2 \end{array} \right) \phi_2^{x_{2,t}-r_2} (1-\phi_2)^{x_{2,t-1}-x_{2,t}-r_2} \\ & \cdot \; e^{-(\mu_1^I + \mu_2^I + \mu_3^I)} \frac{(\mu_1^I)^{r_1}}{r_1!} \frac{(\mu_2^I)^{r_2}}{r_2!} \\ & \cdot \sum_{j=0}^{m} \left(\begin{array}{c} r_1 \\ j \end{array} \right) \left(\begin{array}{c} r_2 \\ j \end{array} \right) j! \left(\frac{\mu_3^I}{\mu_1^I \mu_2^I} \right)^j \Bigg\} \end{aligned}$$

and

$$\begin{aligned} f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}; \phi_1, \phi_2 \mu_1^I, \mu_2^I, \mu_3^O) = \sum_{r_1=0}^{g_1} \sum_{r_2=0}^{g_2} \Bigg\{ & \left(\begin{array}{c} x_{1,t-1} \\ x_{1,t} - r_1 \end{array} \right) \phi_1^{x_{1,t}-r_1} (1-\phi_1)^{x_{1,t-1}-x_{1,t}-r_1} \\ & \cdot \left(\begin{array}{c} x_{2,t-1} \\ x_{2,t} - r_2 \end{array} \right) \phi_2^{x_{2,t}-r_2} (1-\phi_2)^{x_{2,t-1}-x_{2,t}-r_2} \\ & \cdot \; e^{-(\mu_1^I + \mu_2^I + \mu_3^I + c)} \frac{(\mu_1^I)^{r_1}}{r_1!} \frac{(\mu_2^I)^{r_2}}{r_2!} \\ & \cdot \sum_{j=0}^{m} \left(\begin{array}{c} r_1 \\ j \end{array} \right) \left(\begin{array}{c} r_2 \\ j \end{array} \right) j! \left(\frac{\mu_3^I + c}{\mu_1^I \mu_2^I} \right)^j \Bigg\} \end{aligned}$$

By the factorization theorem, the likelihood ratio statistics $T_t(\boldsymbol{x}_t)$ derived from equation (C.3) are

$$T_t(\boldsymbol{x}_t) = ln \, \frac{v_t}{w_t}, \; t = 1, 2, ..., s$$

Table C.1: The limits (L) for the modified one-sided EWMA charts for independent Poisson data for an in control mean of 5

ARL_0	smoothing parameter (λ)								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
370	2.758	2.976	3.093	3.178	3.195	3.278	3.301	3.244	3.203
741	3.051	3.264	3.374	3.448	3.485	3.556	3.542	3.523	3.560

Table C.2: Detection performance of using model residuals with SR methods

Model	Shift size	Estimate	Method		
			Parallel	Frisén	Wessman
M1	2	$\widehat{\text{ced}}$	4.191	3.977	3.997
		$\widehat{\text{tar}}$	42.4	50.1	50.7
		$\widehat{\text{far}}$	13.0	11.3	11.6
		$\widehat{\text{ndr}}$	44.6	38.6	37.7
	3	$\widehat{\text{ced}}$	3.628	3.339	3.329
		$\widehat{\text{tar}}$	69.2	75.0	75.0
		$\widehat{\text{far}}$	13.0	11.3	11.6
		$\widehat{\text{ndr}}$	17.8	13.7	13.4
M2	2	$\widehat{\text{ced}}$	4.596	4.509	4.525
		$\widehat{\text{tar}}$	46.3	53.9	54.6
		$\widehat{\text{far}}$	13.0	11.3	11.6
		$\widehat{\text{ndr}}$	40.7	34.8	33.8
	3	$\widehat{\text{ced}}$	3.906	3.578	3.559
		$\widehat{\text{tar}}$	72.8	78.3	78.1
		$\widehat{\text{far}}$	13.0	11.3	11.6
		$\widehat{\text{ndr}}$	14.2	10.4	10.3

Table C.3: Detection performance for case 1

Shift size	Estimate	Method		
		Parallel	Frisén	Case 1
2	$\widehat{\text{ced}}$	4.071	3.852	3.852
	$\widehat{\text{tar}}$	34.4	49.5	49.5
	$\widehat{\text{far}}$	5.1	9.9	9.9
	$\widehat{\widehat{\text{ndr}}}$	60.5	40.6	40.6
3	$\widehat{\text{ced}}$	3.732	3.259	3.259
	$\widehat{\text{tar}}$	65.4	76.7	76.7
	$\widehat{\text{far}}$	5.1	9.9	9.9
	$\widehat{\widehat{\text{ndr}}}$	29.5	13.4	13.4

Table C.4: Detection performance for case 2

Shift size	CBS	Estimate	Method		
			Parallel	Frisén	Case 2
2	$\rho = 0.2$	$\widehat{\text{ced}}$	4.027	3.546	3.424
		$\widehat{\text{tar}}$	49.3	68.3	69.7
		$\widehat{\text{far}}$	5.1	8.8	13.4
		$\widehat{\text{ndr}}$	45.6	22.9	16.9
	$\rho = 0.4$	$\widehat{\text{ced}}$	4.043	3.678	3.639
		$\widehat{\text{tar}}$	47.1	63.4	65.0
		$\widehat{\text{far}}$	4.9	8.9	12.0
		$\widehat{\text{ndr}}$	48.0	27.7	23.0
	$\rho = 0.6$	$\widehat{\text{ced}}$	4.054	3.821	3.742
		$\widehat{\text{tar}}$	44.3	58.3	59.8
		$\widehat{\text{far}}$	4.6	9.5	10.6
		$\widehat{\text{ndr}}$	51.1	32.2	29.6
3	$\rho = 0.2$	$\widehat{\text{ced}}$	3.361	2.562	2.397
		$\widehat{\text{tar}}$	80.9	87.2	84.4
		$\widehat{\text{far}}$	5.1	8.8	13.1
		$\widehat{\text{ndr}}$	14.0	4.0	2.5
	$\rho = 0.4$	$\widehat{\text{ced}}$	3.391	2.769	2.715
		$\widehat{\text{tar}}$	78.6	84.7	83.5
		$\widehat{\text{far}}$	4.9	8.9	11.8
		$\widehat{\text{ndr}}$	16.5	6.4	4.7
	$\rho = 0.6$	$\widehat{\text{ced}}$	3.453	3.017	2.949
		$\widehat{\text{tar}}$	76.4	82.0	82.1
		$\widehat{\text{far}}$	4.6	9.5	10.7
		$\widehat{\text{ndr}}$	19.0	8.5	7.3

Table C.5: Detection performance for case 3 (shift size = 2)

Shift size	CWS	Estimate	Chart	Method		
				Parallel	Frisén	Case 3
2	$\phi = 0.2$	$\widehat{\text{ced}}$	standard	4.287	4.075	3.834
			modified	4.224	3.945	3.843
		$\widehat{\text{tar}}$	standard	38.8	51.9	46.0
			modified	28.8	45.3	42.8
		$\widehat{\text{far}}$	standard	11.9	16.7	8.8
			modified	4.4	8.6	7.3
		$\widehat{\text{ndr}}$	standard	49.3	31.4	45.2
			modified	66.8	46.1	49.9
	$\phi = 0.4$	$\widehat{\text{ced}}$	standard	5.074	4.856	4.051
			modified	4.648	4.463	3.995
		$\widehat{\text{tar}}$	standard	38.8	48.5	40.2
			modified	25.3	39.8	36.9
		$\widehat{\text{far}}$	standard	21.0	24.3	8.9
			modified	4.6	8.5	7.2
		$\widehat{\text{ndr}}$	standard	40.2	27.2	50.9
			modified	70.1	51.7	55.9
	$\phi = 0.6$	$\widehat{\text{ced}}$	standard	6.814	6.221	4.546
			modified	5.487	5.432	4.540
		$\widehat{\text{tar}}$	standard	31.6	40.5	42.7
			modified	20.0	30.5	33.1
		$\widehat{\text{far}}$	standard	31.3	30.2	10.9
			modified	5.6	8.7	8.8
		$\widehat{\text{ndr}}$	standard	37.1	29.3	46.4
			modified	74.4	60.8	58.1

Table C.6: Detection performance for case 3 (shift size = 3)

Shift size	CWS	Estimate	Chart	Method		
				Parallel	Frisén	Case 3
3	$\phi = 0.2$	$\widehat{\text{ced}}$	standard	3.779	3.348	3.233
			modified	3.839	3.399	3.300
		$\widehat{\text{tar}}$	standard	66.3	72.4	71.4
			modified	58.2	71.7	69.8
		$\widehat{\text{far}}$	standard	11.9	16.7	9.2
			modified	4.4	8.6	7.4
		$\widehat{\text{ndr}}$	standard	21.8	10.9	19.4
			modified	37.4	19.7	22.8
	$\phi = 0.4$	$\widehat{\text{ced}}$	standard	4.347	3.919	3.482
			modified	4.218	3.884	3.564
		$\widehat{\text{tar}}$	standard	61.8	66.3	67.6
			modified	53.0	66.9	64.0
		$\widehat{\text{far}}$	standard	21.0	24.3	9.4
			modified	4.6	8.5	7.3
		$\widehat{\text{ndr}}$	standard	17.2	9.4	23.0
			modified	42.4	24.6	27.8
	$\phi = 0.6$	$\widehat{\text{ced}}$	standard	5.973	5.270	3.976
			modified	5.076	4.839	3.983
		$\widehat{\text{tar}}$	standard	51.3	58.3	61.1
			modified	45.7	57.1	58.9
		$\widehat{\text{far}}$	standard	31.3	30.3	11.3
			modified	5.6	8.7	8.9
		$\widehat{\text{ndr}}$	standard	17.3	11.4	27.6
			modified	48.7	34.2	32.2

Table C.7: Detection performance for case 4

Shift size	CWS CBS	Estimate	Chart	Method		
				Parallel	Frisén	Case 4
2	$\phi = 0.6$ $\rho = 0.25$	$\widehat{\text{ced}}$	standard	4.503	4.169	3.705
			modified	4.013	3.779	3.685
		$\widehat{\text{tar}}$	standard	53.9	56.5	56.1
			modified	44.8	56.4	55.5
		$\widehat{\text{far}}$	standard	31.7	32.5	19.8
			modified	6.2	8.5	15.2
		$\widehat{\text{ndr}}$	standard	14.4	11.0	24.1
			modified	49.0	35.1	29.3
	$\phi = 0.6$ $\rho = 0.5$	$\widehat{\text{ced}}$	standard	4.406	4.322	3.780
			modified	4.029	3.868	3.733
		$\widehat{\text{tar}}$	standard	52.0	53.6	53.8
			modified	42.9	50.8	51.4
		$\widehat{\text{far}}$	standard	29.8	32.1	17.0
			modified	5.7	8.1	13.1
		$\widehat{\text{ndr}}$	standard	18.2	14.3	29.2
			modified	51.4	41.1	35.5
3	$\phi = 0.6$ $\rho = 0.25$	$\widehat{\text{ced}}$	standard	3.485	3.051	2.988
			modified	3.545	3.127	3.032
		$\widehat{\text{tar}}$	standard	64.2	63.5	70.5
			modified	73.8	80.5	73.5
		$\widehat{\text{far}}$	standard	32.9	34.7	23.5
			modified	6.7	9.4	18.1
		$\widehat{\text{ndr}}$	standard	2.9	1.8	6.0
			modified	19.5	10.1	8.4
	$\phi = 0.6$ $\rho = 0.5$	$\widehat{\text{ced}}$	standard	3.608	3.307	3.197
			modified	3.560	3.301	3.240
		$\widehat{\text{tar}}$	standard	64.7	63.2	69.0
			modified	71.5	76.9	71.9
		$\widehat{\text{far}}$	standard	31.4	33.9	20.3
			modified	6.4	9.0	15.4
		$\widehat{\text{ndr}}$	standard	3.9	2.9	10.7
			modified	22.1	14.1	12.7

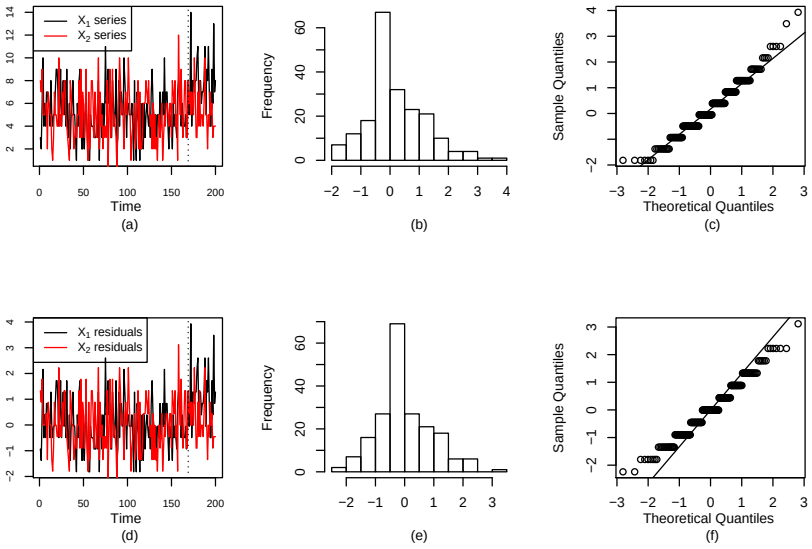

Figure C.1: Diagnostic plots for model M1: (a) data plot, (b) residual plot, (c) - (d) histograms of X_1 and X_2 residuals, respectively, (e) - (f) normal Q-Q plots of X_1 and X_2 residuals, respectively.

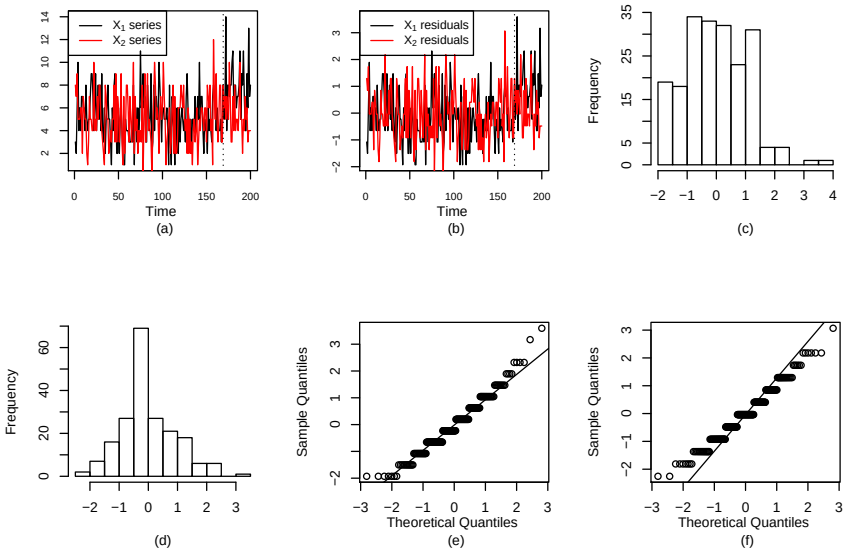

Figure C.2: Diagnostic plots for model M2: (a) data plot, (b) residual plot, (c) - (d) histograms of X_1 and X_2 residuals, respectively, (e) - (f) normal Q-Q plots of X_1 and X_2 residuals, respectively.

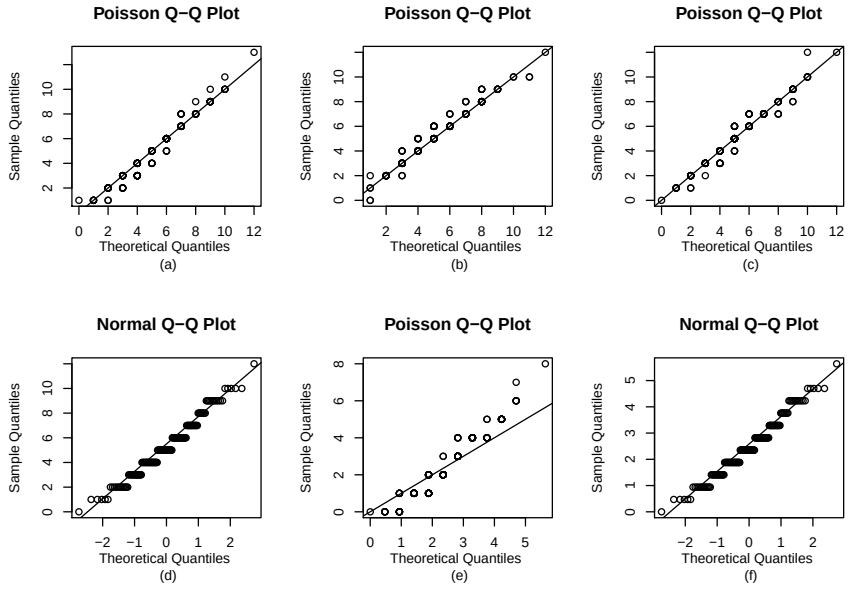

Figure C.3: Diagnostic plots for case 1: (a) - (b) Poisson Q-Q plots of X_1 and X_2 series, respectively, (c) - (d) and (e) - (f) Poisson and Normal Q-Q plots of likelihood ratio statistics from Frisén and our methods, respectively.

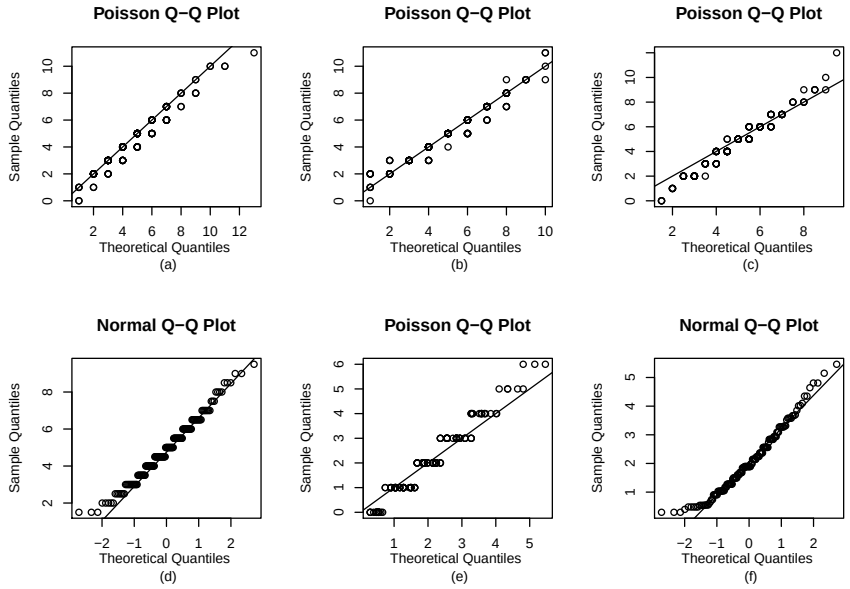

Figure C.4: Diagnostic plots for case 2: (a) - (b) Poisson Q-Q plots of X_1 and X_2 series, respectively, (c) - (d) and (e) - (f) Poisson and Normal Q-Q plots of likelihood ratio statistics from Frisén and our methods, respectively.

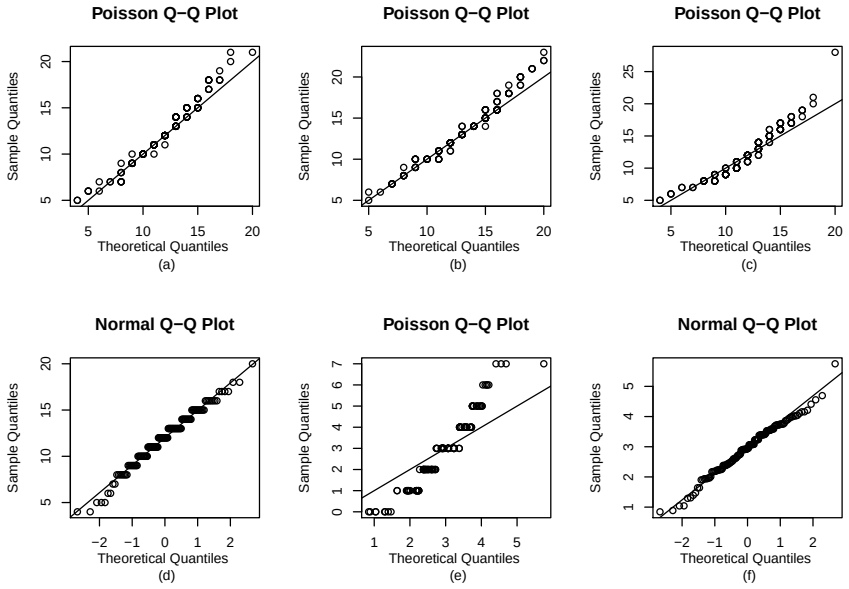

Figure C.5: Diagnostic plots for case 3: (a) - (b) Poisson Q-Q plots of X_1 and X_2 series, respectively, (c) - (d) and (e) - (f) Poisson and Normal Q-Q plots of likelihood ratio statistics from Frisén and our methods, respectively.

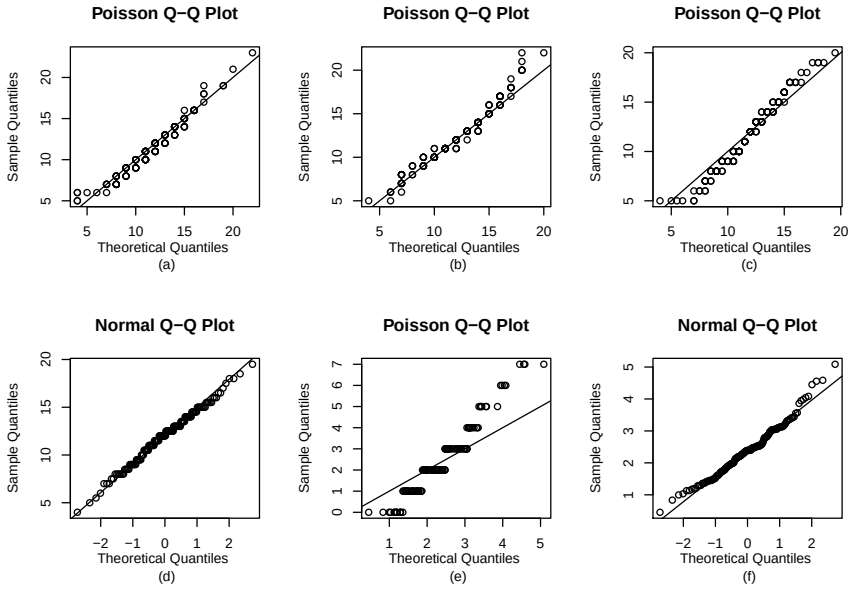

Figure C.6: Diagnostic plots for case 4: (a) - (b) Poisson Q-Q plots of X_1 and X_2 series, respectively, (c) - (d) and (e) - (f) Poisson and Normal Q-Q plots of likelihood ratio statistics from Frisén and our methods, respectively.

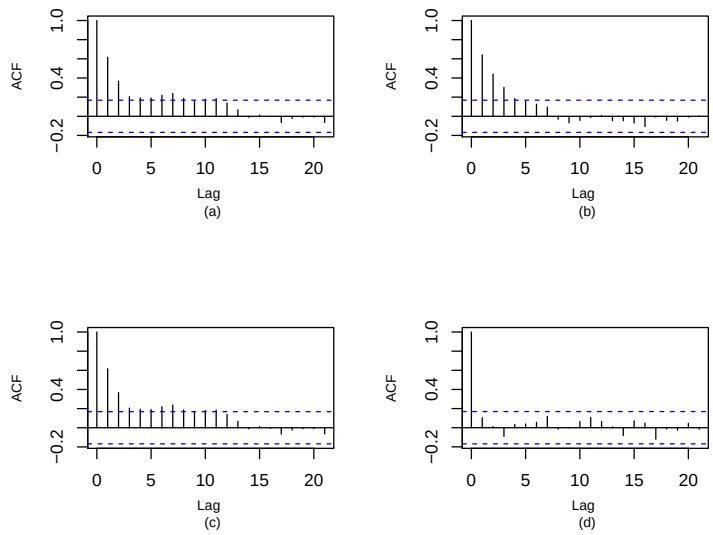

Figure C.7: The autocorrelation function plots of (a) X_1 series, (b) X_2 series, (c) sufficient statistics from Frisén method and (d) likelihood ratio statistics from case 3.

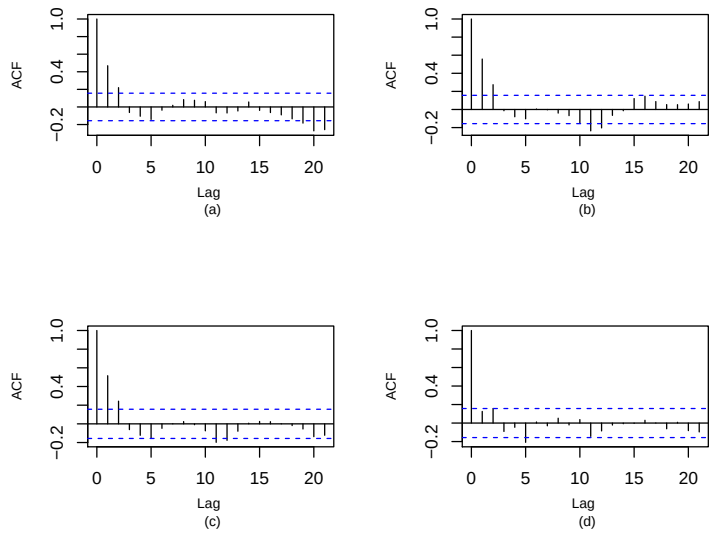

Figure C.8: The autocorrelation function plots of (a) X_1 series, (b) X_2 series, (c) sufficient statistics from Frisén method and (d) likelihood ratio statistics from case 4.

Appendix D

The implementation of SR methods

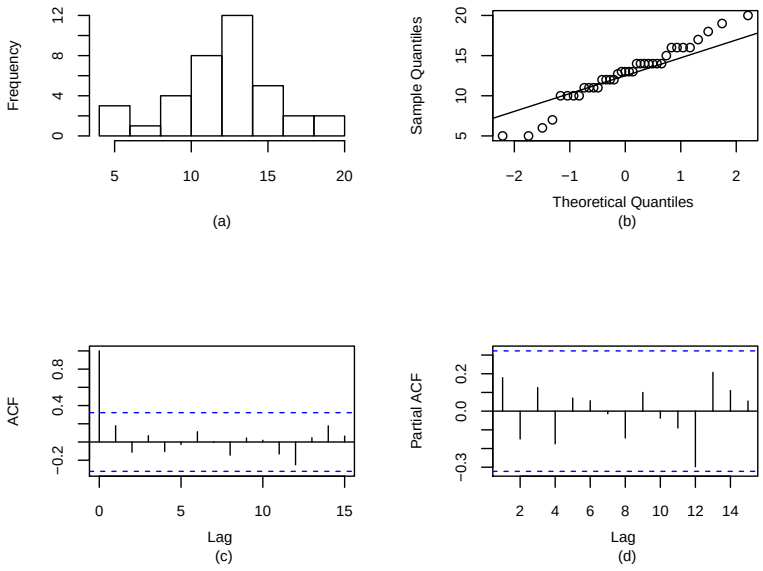

Figure D.1: Plots for the training data set for San Francisco series: (a) histogram, (b) Normal Q-Q plot, (c) and (d) plots of the autocorrelation and partial correlation functions, respectively.

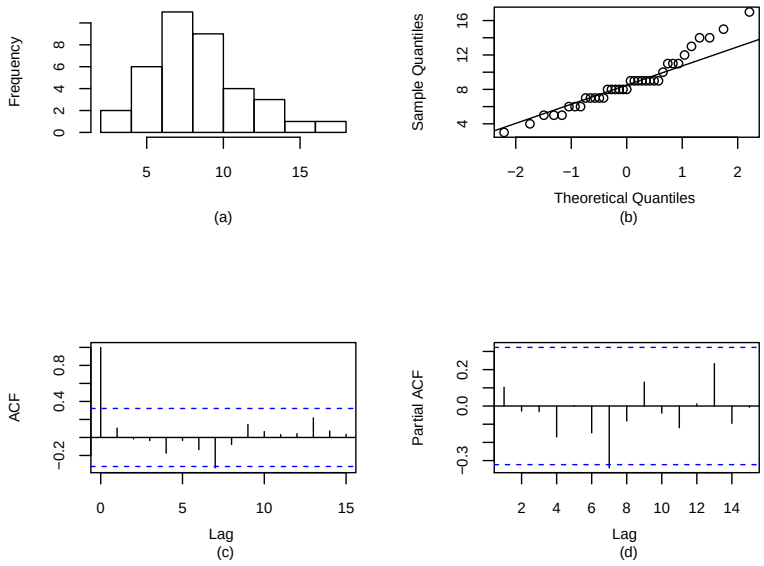

Figure D.2: Plots for the training data set for Las Vegas series: (a) histogram, (b) Normal Q-Q plot, (c) and (d) plots of the autocorrelation and partial correlation functions, respectively.

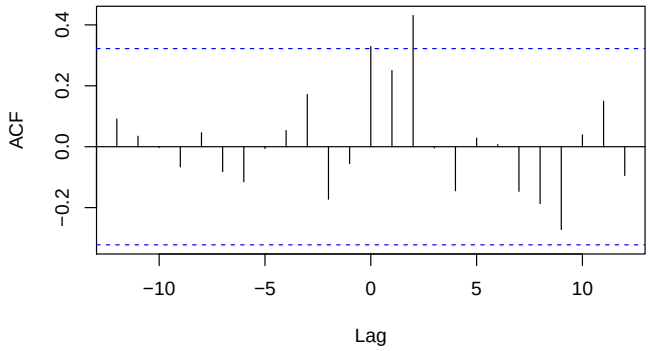

Figure D.3: Plot of the cross correlation function of the training sets for San Francisco and Las Vegas series.

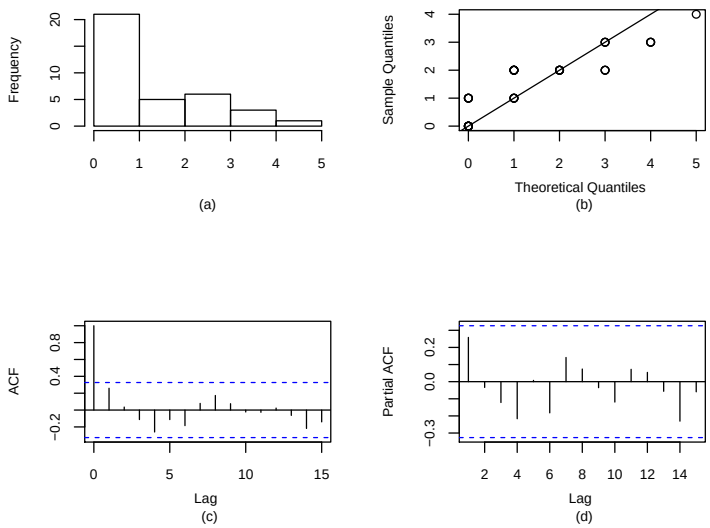

Figure D.4: Plots of training data set for Greater Manchester series: (a) histogram, (b) Poisson Q-Q plot, (c) and (d) plots of the autocorrelation and partial correlation functions, respectively.

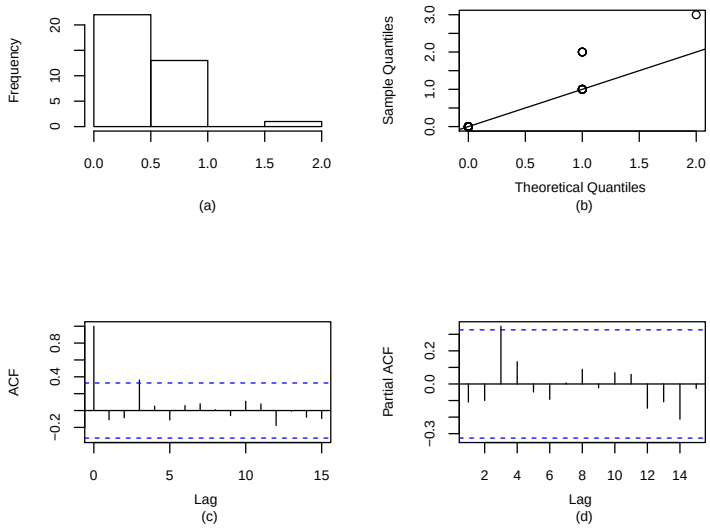

Figure D.5: Plots of training data set for Cheshire & Merseyside series: (a) histogram, (b) Poisson Q-Q plot, (c) and (d) plots of the autocorrelation and partial correlation functions, respectively.

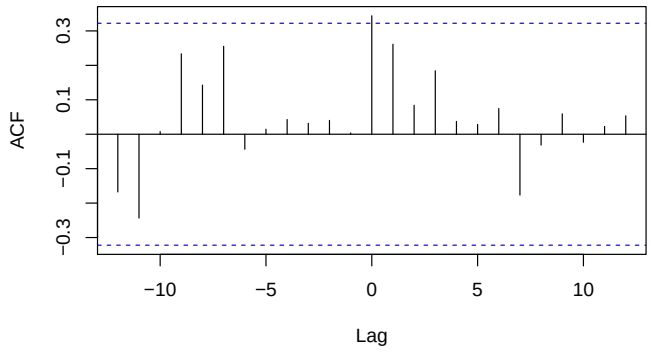

Figure D.6: Plot of the cross correlation function of the training sets for Greater Manchester and Cheshire & Merseyside series.

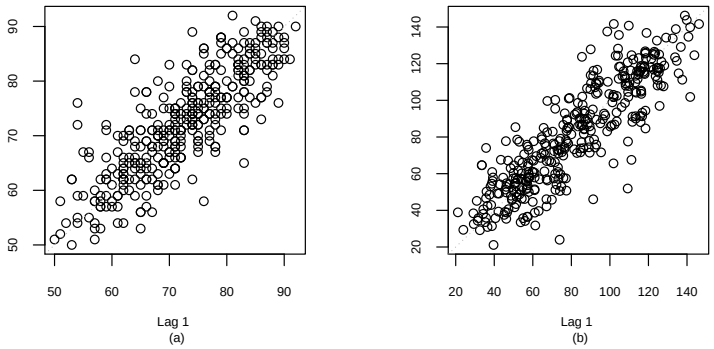

Figure D.7: Scatter plots of the data for (a) humidity and (b) ozone series.

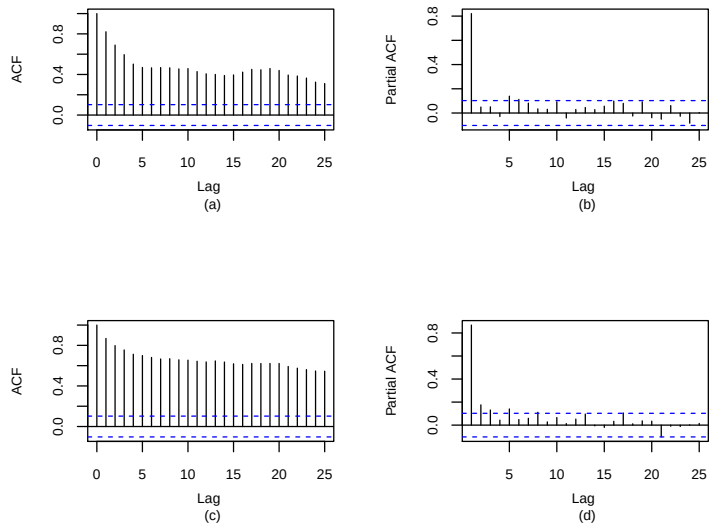

Figure D.8: Plots of ACF and PACF of (a) - (b) humidity series and (c) - (d) ozone series.

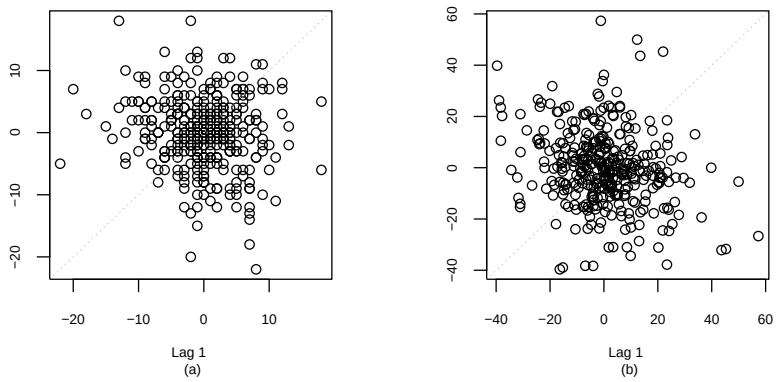

Figure D.9: Scatter plots of the differenced data for (a) humidity and (b) ozone series.

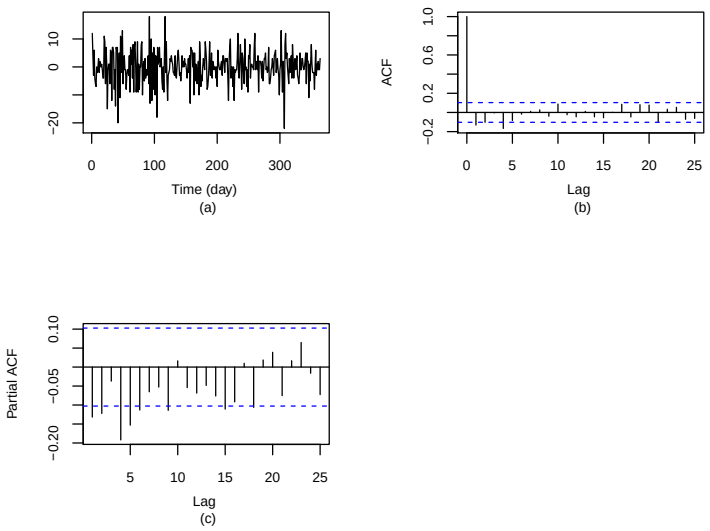

Figure D.10: Plots of (a) differenced data, (b) ACF and (c) PACF for humidity series.

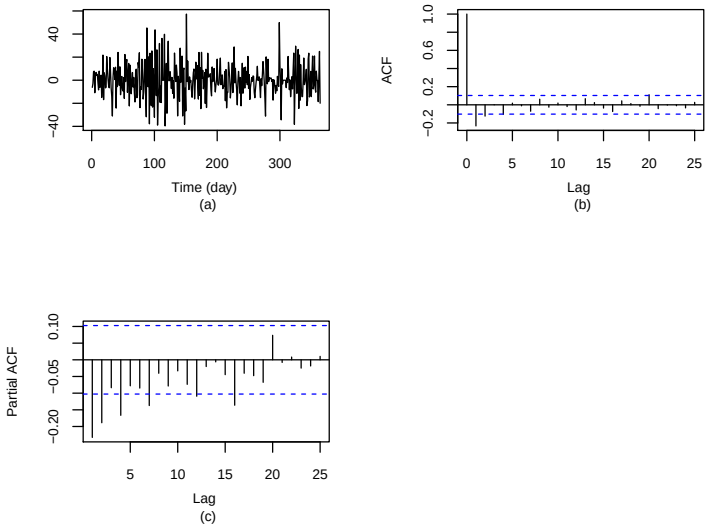

Figure D.11: Plots of (a) differenced data, (b) ACF and (c) PACF for ozone series.

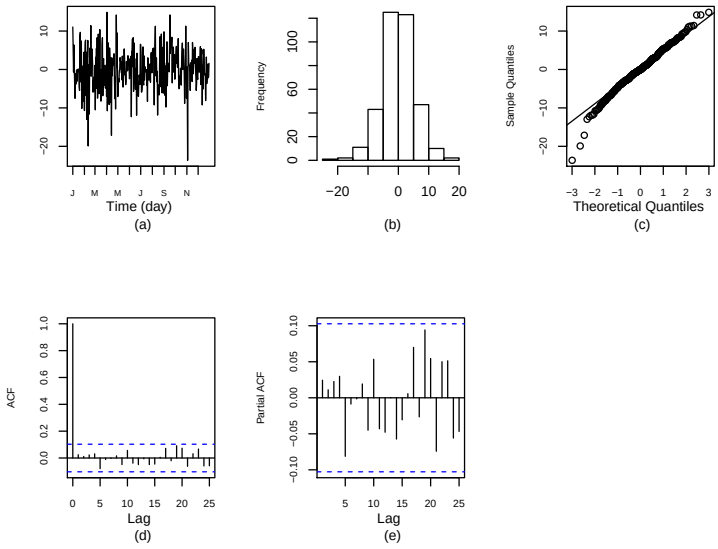

Figure D.12: Diagnostic plots for the humidity model (Model A): (a) residual plot, (b) histogram, (c) normal Q-Q plot, (d) and (e) plots of autocorrelation and partial correlation functions, respectively.

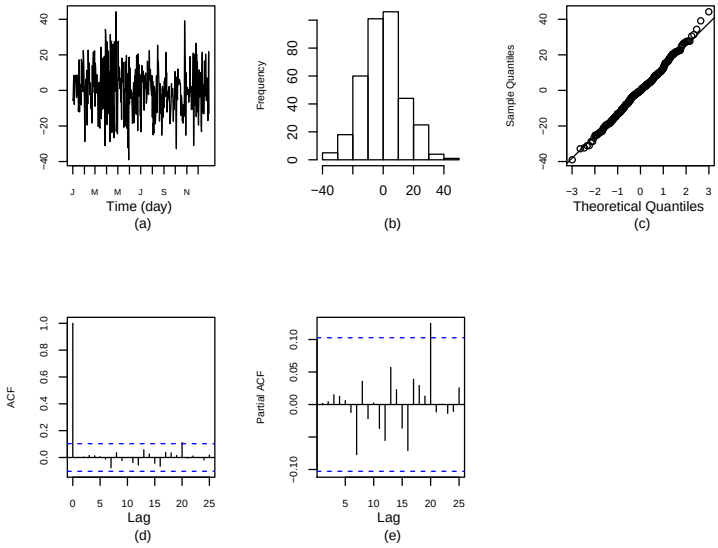

Figure D.13: Diagnostic plots for the ozone model (Model A): (a) residual plot, (b) histogram, (c) normal Q-Q plot, (d) and (e) plots of the autocorrelation and partial correlation functions, respectively.

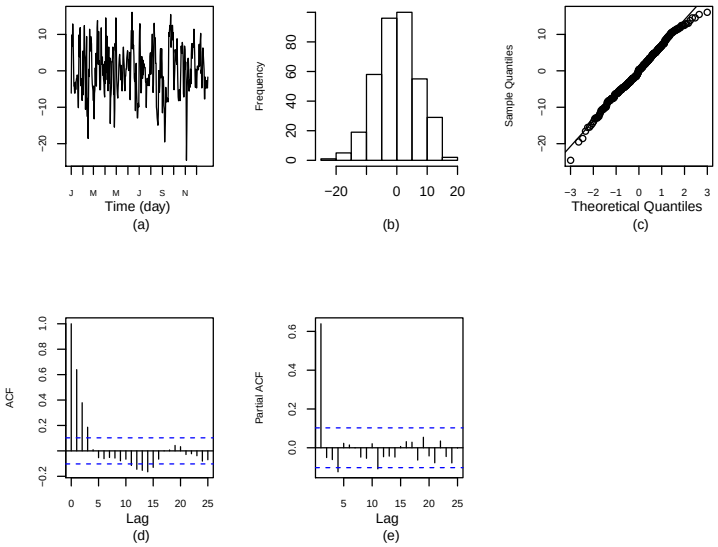

Figure D.14: Diagnostic plots for the humidity model (Model B): (a) residual plot, (b) histogram, (c) normal Q-Q plot, (d) and (e) plots of autocorrelation and partial correlation functions, respectively.

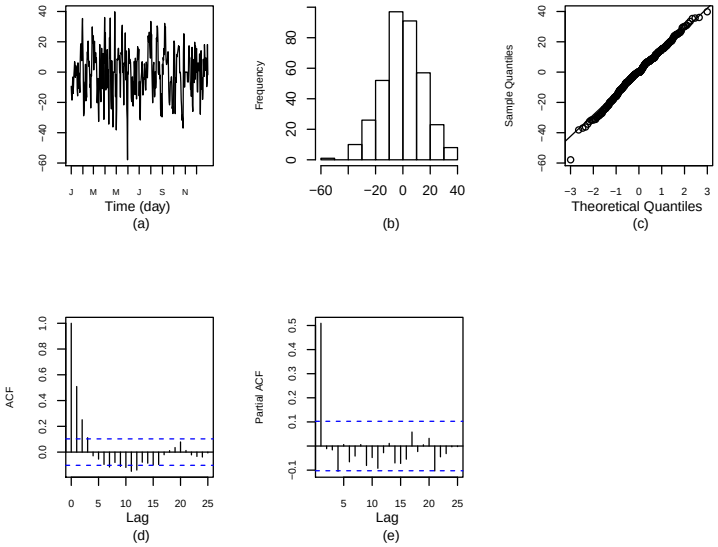

Figure D.15: Diagnostic plots for the ozone model (Model B): (a) residual plot, (b) histogram, (c) normal Q-Q plot, (d) and (e) plots of autocorrelation and partial correlation functions, respectively.

Table D.1: Parameter estimation for the regression models

Coefficient	Humidity model	Ozone model
β_1	77.13	49.90
β_2	-2.56	12.52
β_3	-2.93	27.32
β_4	-8.66	51.96
β_5	-9.64	59.89
β_6	-10.19	59.32
β_7	-16.06	69.99
β_8	-12.19	53.42
β_9	-4.59	29.67
β_{10}	5.06	11.01
β_{11}	1.44	7.79
β_{12}	4.58	-0.97
R^2	0.4633	0.7178

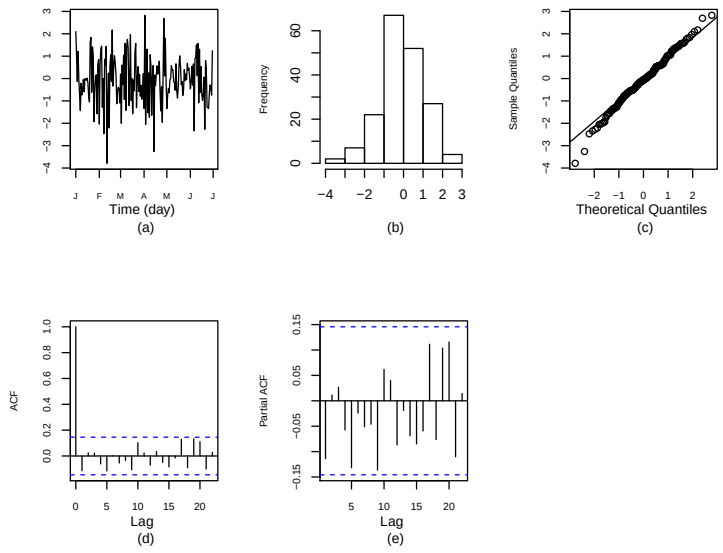

Figure D.16: Plots for the training data set for humidity series (Model A): (a) residual plot, (b) normal Q-Q plot, (c) and (d) plots of the autocorrelation and partial correlation functions of the training data set.

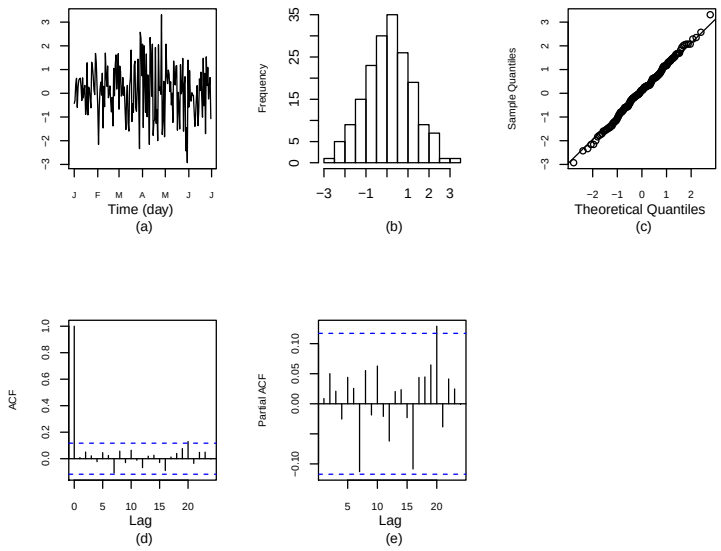

Figure D.17: Plots for the training data set for ozone series (Model A): (a) residual plot, (b) normal Q-Q plot, (c) and (d) plots of the autocorrelation and partial correlation functions of the training data set.

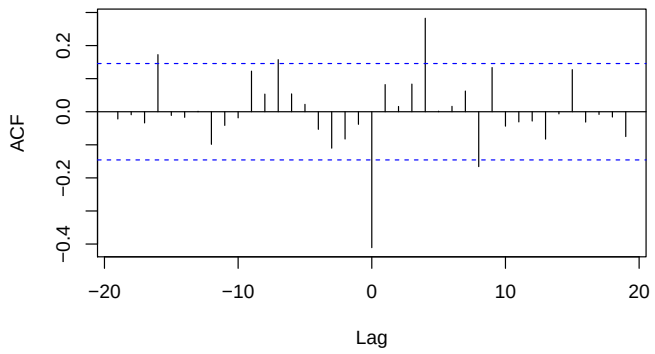

Figure D.18: Plot of the cross correlation function of the training data set for humidity and ozone series (Model A).

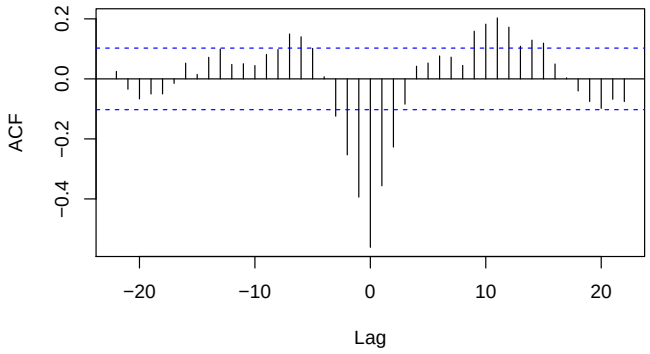

Figure D.19: Plot of the cross correlation function of the training data set for humidity and ozone series (Model B).

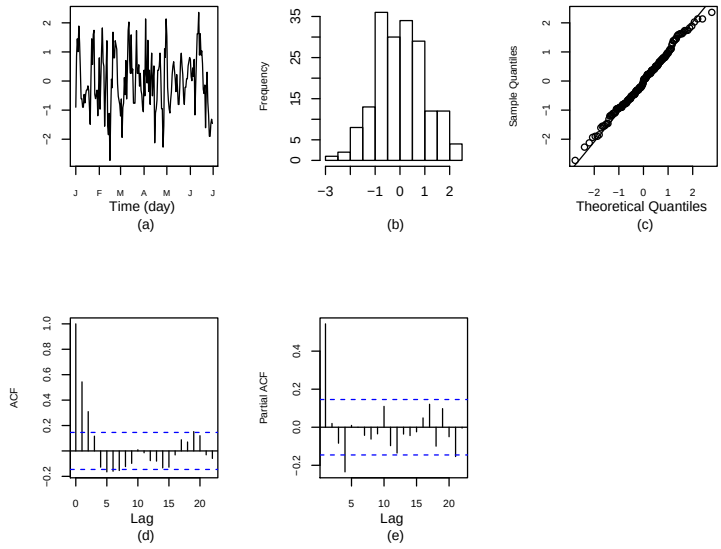

Figure D.20: Plots for the training data set for humidity series (Model B): (a) residual plot, (b) histogram, (c) normal Q-Q plot, (d) and (e) plots of the autocorrelation and partial correlation functions of the training data set.

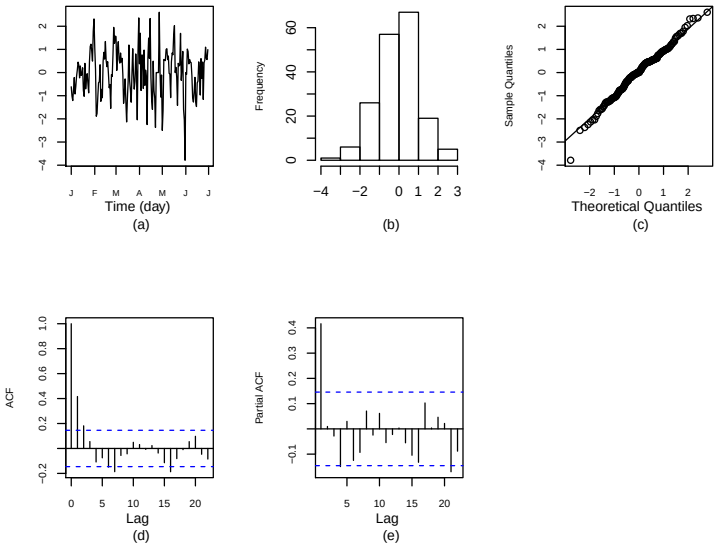

Figure D.21: Plots for the training data set for ozone series (Model B): (a) residual plot, (b) histogram, (c) normal Q-Q plot, (d) and (e) plots of the autocorrelation and partial correlation functions of the training data set.

Bibliography

Adams, B. M. and I. T. Tseng (1998). Robustness of forcast-based monitoring schemes. *Journal of Quality Technology 30*, 328–339.

Agresti, P. (2007). Generalized linear models for count data. In *An Introduction to Categorical Data Analysis* (2nd ed.). John Wiley & Sons, Inc.

Al-Osh, M. and A. A. Alzaid (1987). First-order integer-valued autoregressive process. *Journal of Time Series Analysis 8*(3), 261–275.

Alwan, L. C. and H. V. Roberts (1988). Time series modelling for statistical process control. *Journal of Business and Economics Statistics 6*, 87–95.

Alzaid, A. A. and M. Al-Osh (1990). An integer-valued pth-order autoregressive structure (inar(p)) process. *Journal of Applied Probability 27*(2), 314–324.

Andersson, E. (2007). Effect of dependency in systems for multivariate surveillance. Technical report, Statistical Research Unit, the University of Gothenburg.

Andersson, E. (2008). Hotelling's T^2 method in multivaraite on-line surveillance. on the delay of an alarm. Technical report, Statistical Research Unit, the University of Gothenburg.

Andrews, N. (2010). Detecting outbreaks using a national laboratory reporting network: statistical methods and practical issues. In *Statistical Methods for Outbreak Detection*, The Open University, Milton Keynes, UK.

Arnold, S. F. (1988). *Encyclopedia of Statistical Science*. Wiley, New York.

Bersimis, S., S. Psarakis, and J. Panaretos (2007). Multivariate statistical process control chart: An overview. *Quality and Reliability Engineering International 23*, 517–543.

Bodnar, O. and W. Schmid (2007). Surveillance of the mean behavior of multivariate time series. *Statistica Neerlandica 61*(4), 583–406.

Borror, C. M., C. P. Champ, and R. S. E. (1998). Poisson EWMA control chart. *Journal of Quality Technology 30*(4), 352–361.

Box, G. P., G. M. Jenkins, and G. C. Reinsel (1994). *Time Series Analysis: Forecasting and Control* (3 ed.). Prentice-Hall: Englewood Chiffs, New Jersey.

Brockwell, P. J. and R. A. Davis (2002). *Introduction to Time Series and Forecasting* (2nd ed.). Springer, NewYork.

Brook, D. and D. Evans (1972). An approach to the probability distribution of CUSUM run length. *Biometrika 59*(3), 539–549.

Buckeridge, D. (2007). Outbreak detection through automated surveillance: A review of the determinants of detection. *Journal of Biomedical Informatics 40*, 370–379.

Buehler, J., J. Hopkins, J. Overhage, D. Sosin, and V. Tong (2004). Framework for evaluating public health surveillance systems for early detection of outbreaks. *Morbidity and Mortality Weekly Report 53*, 1–11.

Burkom, H. (2003). Development, adaptation, and assessment of alerting algorithms for biosurveillance. *Johns Hopkins APL Technical Digest 24*(4), 335–342.

Burkom, H., Y. Elbert, A. Feldman, and J. Lin (2004). Role of data aggregation in biosurveillance detection strategies with applications for essence. *Morbidity and Mortality Weekly Report 53*, 67–73.

Burkom, H., Y. Elbert, S. F. Magruder, A. H. Najmi, W. Peter, and M. W. Thompson (2008). Developments in the roles, features, and evaluation of alerting algorithms for disease outbreak monitoring. *Johns Hopkins APL Technical Digest 27*(4), 313–331.

Burkom, H., S. Murphy, and G. Shmueli (2007). Automated time series forecasting for biosurveillance. *Statistics in Medicine 26*(2), 4202–4218.

Cardinal, M., R. Roy, and J. Lambert (1999). On application of INAR for the analysis of disease incidence. *Statistics in Medicine 18*, 2025–2039.

CDC (2009, July). The centers for disease control and prevention: Public health surveillance: Overview. `http://www.cdc.gov/ncphi/disss/nndss/phs/overview.htm`.

CDC WONDER (2012, November). Cdc wonder. `http://wonder.cdc.gov/mmwr/mmwrmort.asp`.

Cooper, D., N. Verlander, G. Smith, A. Charlett, E. Gerard, and L. Willocks (2006). Can syndromic surveillance data detect local outbreaks of communicable disease? A model using a historical cryptosporidiosis outbreak. *Epidemiology and Infection 134*, 13–20.

Cowling, B., I. Wong, L. Ho, S. Riley, and G. Leung (2006). Methods for monitoring influenza surveillance data. *International Journal of Epidemiology 35*, 1314–1321.

Cowpertwait, P. S. P. and A. V. Metcalfe (2009). *Introductory Time Series with R*. Springer, NewYork.

Cox, D. R. and D. V. Hinkley (1974). *Theoretical Statistics*. Chapman & Hall, London.

Cox, K. (2004). Tools and techniques for enhanced health surveillance in deployed settings. Technical report, NATO Medical Surveillance and Response, Research and Technology Opportunities and Options.

Crosier, R. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics 30*(3), 291–303.

Crowder, S. V. (1989). Design of exponentially weighted moving average schemes. *Journal of Quality Technology 21*, 155–162.

Czado, C., T. Gneiting, and L. Held (2009). Predictive model assessment for count data. *Biometrics 65*(4), 1254–1261.

Dafni, U., S. Tsiodras, D. Panagiotakos, K. Gkolfinopoulou, G. Kouvatseas, and Z. Tsourti (2004). Algorithm for statistical detection of peaks — syndromic surveillance system for the Athens 2004 Olympic Games. *Morbidity and Mortality Weekly Report 53*, 86–94.

Diggle, P. and M. Stanton (2008). Meningitis incidence in ethiopia: association with environment/meteorology. Technical report, Lancaster University.

Diggle, P. J. (1990). *Time Series: A Biostatistical Introduction*. Oxford University Press, New York.

Diggle, P. J. (2007). Spatio-temporal point processes: Methods and applications. In *Statistical Methods for Spatio-Temporal Systems*. Chapman & Hall/CRC, London.

Dobson, A. J. and A. G. Barnett (2008). *An introduction to generalized linear models* (3 ed.). CRC Press, Boca Raton.

EARS (2010, June). Early aberration reporting system. `http://www.bt.cdc.gov/surveillance/ears/`.

Elliot, A. (2010). Syndromic surveillance in the uk: real-time monitoring of morbidity in the community. In *Statistical Methods for Outbreak Detection*, The Open University, Milton Keynes, UK.

English, J. R., S. Lee, T. W. Martin, and C. Tilmon (2000). Detecting changes in autoregressive processes with $\bar{X}$ and EWMA charts. *IIE Transactions 32*, 1103–1113.

Farrington, C., N. Andrews, A. Beale, and M. Catchpole (1996). A statistical algorithm for the early detection of outbreak for infectious disease. *Journal of the Royal Statistical Society 159*(3), 547–563.

Farrington, P. and N. Andrews (2004). Outbreak detection: Applocation to infectious disease surveillance. In *Monitoring the Health of Population Statistical Principles and Methods for Public Health Surveillance*. Oxford University Press, Oxford.

Fatahi, A. A., R. Noorossana, P. Dokouhaki, and B. F. Moghaddam (2012). Zero inflated Poisson EWMA control chart for monitoring rare health-related events. *Journal of Mechanics in Medicine and Biology 12*(4).

Fienberg, S. and G. Shmueli (2005). Statistical issues and challenges associated with rapid detection of bio-terrorist attacks. *Statistics in Medicine 24*, 513–529.

Fox, D. (2007, August). Statistical methods for biosecurity monitoring & surveillance: Control chart. http://www.acera.unimelb.edu.au/materials/endorsed/0605.pdf.

Fricker, R. (2007). Directionally sensitive multivariate statistical process procedures with applications to syndromic surveillance. *Advances in Disease Surveillance 3*(1), 1–17.

Fricker, R., M. Knitt, and C. Hu (2007). Comparing directionally sensitive MCUSUM and MEWMA procedures with application to biosurveillance. *Quality Engineering 20*(4), 478–494.

Fricker, R. D. and H. Rolka (2006). Protecing against biological terrorism: Statistical issues in electronic biosurveillance. *Chance 19*(4).

Frisén, M. (1992). Evaluation of method for statistical surveillance. *Statistics in Medicine 11*, 1489–1502.

Frisén, M. (2003). Statistical surveillance. Optimality and methods. *International Statistical Review 71*(2), 403–434.

Frisén, M. (2005). Optimal surveillance. In *Spatial & Syndromic Surveillance for Public Health*. Wiley, Chichester.

Frisén, M. (2007). Properties and use of the Shewhart method and followers. *Sequential Analysis 26*, 171–193.

Frisén, M. (2009). Evaluation of multivariate surveillance. Technical report, Statistical Research Unit, the University of Gothenburg.

Frisén, M., E. Andersson, and L. Schioler (2007). Robust outbreak surveillance of epidemics in sweden. *Statistics in Medicine 28*, 476–493.

Frisén, M., E. Andersson, and L. Schioler (2011). Sufficient reduction in multivariate surveillance. *Communications in Statistics - Theory and Methods 40*(10), 1821–1838.

Frisén, M. and J. de Maré (1991). Optimal surveillance. *Biometrika 78*, 271–280.

Gan, F. F. (1990). Monitoring Poisson observations using modified exponentially weighted moving average control charts. *Communications in Statistics - Simulation and Computation 19*(1), 103–124.

Glaser, C. A., S. Gilliam, W. W. Thompson, D. E. Dassey, S. H. Waterman, M. Saruwatari, S. Shapiro, and K. Fukuda (2002). Medical care capacity for influenza outbreaks, Los Angeles. *Emerging Infectious Diseases 8*(6), 569–574.

Gneiting, T. and A. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Goldenberg, A., G. Shmueli, R. Caruana, and S. Fienberg (2002). Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. *Proceedings of the National Academy of Sciences 99*(8), 5237–5240.

Google (2010). Explore flu trends around the world. http://www.google.org/flutrends/.

Hanslik, T., P. Boelle, and A. Flahault (2001). The control chart: an epidemiological tool for public health monitoring. *Public Health 115*, 277–281.

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

Held, L., M. Hofmann, and M. Höhle (2006). A two-component model for counts of infectious disease. *Biostatistics 7*(3), 422–437.

Held, L., M. Höhle, and M. Hofmann (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling 5*, 187–199.

Held, L., K. Rufibach, and F. Balabdaoui (2010). A score regression approach to assess calibration of continuous probabilistic predictions. *Biometrics*.

Höhle, M. (2007). Surveillance: An R package for the surveillance of infectious diseases. *Computational Statistics 22*(4), 571–582.

Höhle, M. (2008, May). Short course on statistical surveillance of infectious disease. http://www.stat.uni-muenchen.de/~hohle/surv-short2008_4.pdf.

Höhle, M. (2010). Online change-point detection in categorical time series. In *Statistical Modelling and Regression Structures*. Springer.

Höhle, M. and M. Paul (2007). Count data regression charts for the monitoring of surveillance time series. *Computational Statistics and Data Analysis 52*(9), 4357–4368.

Höhle, M., M. Paul, and L. Held (2009). Statistical approaches to the monitoring and surveillance of infectious disease for veterinary public health. *Preventive Veterinary Medicine 91*, 2–10.

Holgate, P. (1964). Estimation for the bivariate Poisson distribution. *Biometrika 51*(1/2), 241–245.

Hotelling, H. (1974). Multivariate quality control, illustrated by the air testing of sample bombsights. In *Techniques of Statistical Analysis*, pp. 113–184. McGraw-Hill, New York.

HPA (2011, December). Health protection agency, gerenal information - scarlet fever. http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/ScarletFever/GeneralInformationScarletFever/.

HPA (2012, November). Health protection agency, regional epidemiology unit - north west. http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/HPAweb_C/1195733821245.

Hu, J., L. Shu, P. W. Tse, and K. Tsui (2011). A comparison of EWMA-type charts under linear drifts in Poisson means. In *Quality and Reliability (ICQR), 2011 IEEE International Conference*, pp. 297–301.

Hulth, A. (2010). Practical usage of computer-supported outbreak detection in europe. In *Statistical Methods for Outbreak Detection*, The Open University, Milton Keynes, UK.

Hunter, J. S. (1986). The Exponentially Weighted Moving Average. *Journal of Quality Technology 18*(4), 203–210.

Hutwagner, L., T. Browne, G. Seeman, and A. Fleischauer (2005). Comparing aberration detection methods with simulated data. *Emerging Infectious Diseases 11*(2), 314–316.

Hutwagner, L., W. Thompson, G. Seeman, and T. Treadwell (2005). A simulation model for assessing aberration detection methods used in public health surveillance for systems with limited baselines. *Statistics in Medicine 24*, 543–550.

Jackson, M., A. Baer, I. Painter, and J. Duchin (2007). A simulation study comparing aberration detection algorithms for syndromic surveillance. *BMC Medical Informatics and Decision Making 7*(6).

Järpe, E. (2000). *On univariate and spatial surveillance*. Ph. D. thesis, University of Göteborg, Swenden.

Johnson, N., S. Kotz, and N. Balakrishnan (1997). Discrete multivariate distributions. John Wiley & Sons, Inc., New York.

Joner, M., W. Woodall, M. Reynolds, and R. Fricker (2008). A one-sided MEWMA chart for health surveillance. *Quality and Reliability Engineering International 24*, 503–518.

Karlis, D. and I. Ntzoufras (2005). Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R. *Journal of Statistical Software 14*(10).

Kavanagh, K., C. Roberson, and J. McMenamin (2010). Exception reporting of influenza-like syndromed in scotland using nhs24 data, experience during the influenza a h1n1v pandemic. In *Statistical Methods for Outbreak Detection*, The Open University, Milton Keynes, UK.

Kavanagh, K., C. Roberson, H. Murdoch, G. Crooks, and J. McMenamin (2012). Syndromic surveillance of influenza-like illness in Scotland during the influenza a H1N1v pandemic. *Journal of the Royal Statistical Society: Series A (Statistics and Society) 175*(4), 939–958.

Kleiman, K., R. Lazarus, and R. Platt (2004). A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. *American Journal of Epidemiology 159*(3), 217–224.

Kleinman, K. and A. Abrams (2006). Assessing surveillance using sensitivity, specificity and timeliness. *Statistical Methods in Medical Research 15*, 445–464.

Knoth, S. (2012, September). Package 'spc'. `http://cran.r-project.org/web/packages/spc/spc.pdf`.

Kramer, H. and W. Schmid (1997). Control charts for time series. *Nonlinear Analysis, Theory, Methods & Applications 30*(7), 4007–4016.

Kuang, J., W. Z. Yang, D. L. Zhou, Z. J. Li, and Y. J. Lan (2012). Epidemic features affecting the performance of outbreak detection algorithms. *BMC Public Health 12*(418).

Kullaa, J. (2003). Damage detection of the z24 bridge using control charts. *Mechanical Systems and Signal Processing 17*(1), 163–170.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics, Theory and Methods 26*, 1481–1496.

Kulldorff, M. (2001). Prospective time periodic geographic disease surveillance using scan statistics. *Journal of Royal Statistics Society. Series A 164*, 61–72.

Kulldorff, M., F. Mostashari, L. Duczmal, K. Yih, K. Kleinman, and R. Platt (2007). Multivariate scan statistics for disease surveillance. *Statistics in Medicine 26*(8), 1824–1833.

Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society, Series B 57*(4), 613–658.

Lawson, A. and K. Kleinman (2005). Introduction: Spatial and syndromic surveillance for public health. In *Spatial & Syndromic Surveillance for Public Health*. Wiley, Chichester.

Le Strat, Y. and F. Carrat (1999). Monitoring epidemiologic surveillance data using hidden Markov models. *Statistics in Medicine 18*, 3463–3478.

Lotze, T., G. Shmueli, and I. Yahav (2010). Simulating and evaluating biosurveillance datasets. In *Biosurveillance: Methods and Case studies*. Taylor & Francis, CRC Press.

Lowry, C., W. Woodall, C. Champ, and H. Rigdon (1992). A multivariate exponentially weighted moving average control chart. *Technometrics 34*(1), 46–53.

Lu, C. W. and J. Reynolds, M. R. (1999). Control charts for monitoring the mean and variance of autocorrelated processes. *Journal of Quality Technology 31*, 259–274.

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

Lütkepohl, H. (1993). *Introduction to multiple time series analysis* (2nd ed.). Springer-Verlag, NewYork.

MacCarthy, B. L. and T. Wasusri (2001). Statistical process control for monitoring scheduling performance — addressing the problem of correlated data. *The Journal of the Operational Research Society 52*(7), 810–820.

MacGregor, J. (1995). Statistical process control of multivariate process. *Control Engineering Practice 3*(3), 403–414.

Maciejewski, R., R. Hafen, S. Rudolph, G. Tebbetts, W. Cleveland, and D. Ebert (2009). Generating synthetic syndromic surveillance data for evaluating visual-analysis techniques. *IEEE Computer Graphics and Applications 29*(3), 18–28.

Mandl, K., B. Reis, and C. Cassa (2004). Measuring outbreak-detection performance by using controlled feature set simulations. *Morbidity and Mortality Weekly Report 53*, 130–136.

Marshall, C., N. Best, A. Bottle, and P. Aylin (2004). Statistical issues in the prospective monitoring of health outcomes across multiple units. *Journal of the Royal Statistical Society 167*(3), 541–559.

Martínez-Beneito, M., D. Conesa, A. López-Qu´lez, and A. López-Maside (2008). Bayesian Markov switching models for the early detection of influenza epidemics. *Statistics in Medicine 27*, 4455–4468.

Mason, R. and J. Young (2008, March). It depends. the effect of dependent observations on process control. `http://www/asq.org.quality-progress/2008/04/statistics-roundtable-it-depends.html`.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin 21*(4), 645–650.

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability 20*(4), 822–835.

Meyer, N., J. McMenamin, C. Robertson, M. Donaghy, G. Allardice, and D. Cooper (2008). A multi-data source surveillance system to detect a bioterrorism attack during the G8 summit in Scotland. *Infection and Epidemiology 136*(7), 876–885.

Meynard, J., H. Chaudet, G. Texier, V. Ardillon, F. Ravachol, and X. Deparis (2008). Value of syndromic surveillance within the armed forces for early warning during a Dengue fever outbreak in French Guiana in 2006. *BMC Medical Informatics and Decision Making*, 8–29.

Mohtashemi, M., K. Kleinman, and W. Yih (2007). Multi-syndrome analysis of time series using PCA: A new concept for outbreak investigation. *Statistics in Medicine 26*, 5203–5224.

Montgomery, D. (2009). *Statistical quality control: A modern introduction*. Wiley, New York.

Montgomery, D. C. and C. M. Mastrangelo (1991). Some statistical process control methods for autocorrelated data. *Journal of Quality Technology 23*, 179–193.

Morais, M. C. and A. Pacheco (2001). Some stochastic proporties of upper one-sided and EWMA charts for μ in the presence of shift in σ. *Sequential Analysis: Design Methods and Applications 20*, 1–2.

Neill, D., A. Moore, and G. Cooper (2005). A Bayesian spatial scan statistic. In *Advances in Neural Information Processing Systems*, Volume 18, pp. 1003–1010.

Neill, D., A. Moore, and G. Cooper (2007). A multivariate Bayesian scan statistic. *Advances in Disease Surveillance 2*.

NHS (2012, May). Hand, foot and mouth disease symptoms. `http://www.nhs.uk/Conditions/Hand-foot-and-mouth-disease/Pages/new_Symptoms.aspx`.

NHS24 (2012, April). Nhs24 — health information and self care advice for scotland. `http://www.nhs24.com/`.

Oakland, J. S. (1990). *Statistical process control: A really practical guide* (3rd ed.). Butterworth-Heinemann., Oxford.

Paul, M. (2010). A Poisson autoregressive model for prospective outbreak detection. statistical methods for outbreak detection. In *Statistical Methods for Outbreak Detection*, The Open University, Milton Keynes, UK.

Paul, M., L. Held, and A. Toschke (2008). Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine 27*, 6250–6267.

Pavlin, J. (2003). Investigation of disease outbreaks detected by "syndromic" surveillance systems. *Journal of the Urban Health: Bulletin of the New York Academy of Medicine 80*(2), 107–114.

Pedeli, X. and D. Karlis (2005). A bivariate INAR(1) process with application. *Statistical Modelling 11*(4), 325–349.

Pelat, C., P. Boelle, B. Cowling, F. Carrat, A. Flahault, and S. Ansart (2007). Online detection and quantification of epidemics. *BMC Medical Informatics and Decision Making 7*(29).

Pérez, C., J. Tejada, and M. Carrasco (1998). Multivariate time series analysis in nosocomical infection surveillance: a case study. *Statistics in Medicine 26*, 1834–1856.

Perry, B. and T. Allen (2005). Causal reasoning engine: An explanation-based approach to syndromic surveillance. In *Proceedings of the 38th Annual Hawaii International Conference*.

Pignatiello, J. and G. Runger (1990). Comparisons of multivariate CUSUM charts. *Journal of Quality Technology 22*, 173–186.

Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*. Springer, New York.

Reis, B. and K. Mandl (2003). Time series modelling for syndromic surveillance. *BMC Medical Informatics and Decision Making 3*(2).

Roberts, S. W. (1966). A comparison of some control chart procedures. *Technometrics 8*, 411–430.

Robertson, C. (2010). Statistical aspects of syndromic surveillance in Scotland. In *Statistical Methods for Outbreak Detection*, The Open University, Milton Keynes, UK.

Robinson, P. B. and T. Y. Ho (1978). Average run lengths of geometric moving average charts by numerical methods. *Technometrics 20*(1), 85–93.

RODS (2010, July). Real-time outbreak and disease surveillance. `https://www.rods.pitt.edu/site/component/option,com_frontpage/Itemid,34/`.

Rogerson, P. (1997). Surveillance systems for monitoring the development of spatial patterns. *Statistics in Medicine 66*, 2081–2093.

Rogerson, P. (2001). Monitoring point patterns for the development of space-time clusters. *Journal of Royal Statistics Society. Series A 1*, 87–96.

Rogerson, P. and I. Yamada (2004). Approaches to syndromic surveillance when data consist of small regional counts. *Morbidity and Mortality Weekly Report 53*, 79–85.

Rolka, H., H. Burkom, G. Cooper, M. Kulldorff, D. Madigan, and W. Wong (2007). Issues in applied statistics of public health bioterrorism surveillance using multiple data streams: research needs. *Statistics in Medicine 26*, 1834–1856.

Rossi, G., L. Lampugnani, and M. Marchi (1999). An approximate CUSUM procedure for surveillance of health events. *Statistics in Medicine 18*, 2111–2122.

SatScan (2010, June). Satscan. `http://www.satscan.org`.

Schmid, W. (1997). On EWMA chart for time series. *Frontiers in Statistical Quality Control 5*, 114–137.

Schmid, W. and A. Schöne (1997). Some properties of the EWMA control chart in the presence of autocorrelation. *The Annals of Statistics 3*, 1277–1283.

Scranton, R., G. Runger, J. Keats, and D. Montgomery (1996). Efficient shift detection using multivariate exponentially-weighted moving average control charts and principal components. *Quality and Reliability Engineering International 12*, 165–171.

Serfling, R. (1963). Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Reports 78*(6), 494–506.

Shephard, R. (2006, June). The development of a syndromic surveillance system for the extensive beef cattle producing regions of Australia. `http://ses.library.usyd.edu.au/bitstream/2123/2210/2/02whole.pdf`.

Shiau, J. H. and Y. Hsu (2005). Robustness of the EWMA control chart to non-normality for autocorrelated processes. *Quality Technology and Quantative Management 2*(2), 125–146.

Shiryaev, A. N. (1963). On optimum methods in quickest detection problems. *Theory of Probability and its Applications 8*, 22–46.

Shmueli, G. and H. Burkom (2010). Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics 52*(1), 39–51.

Shmueli, G. and S. Fienberg (2006). Current and potential statistical methods for monitoring multiple data streams for biosurveillance. In *Statistical Methods in Counterterrorism*. Springer.

Shu, L., W. Jiang, and S. Wu (2007). A one-sided EWMA control chart for monitoring process means. *Communications in Statistics - Simulation and Computation 36*, 901–920.

SMHI (2012). Swedish meteorological and hydrological institute. `http://www.smhi.se/en`.

Sonesson, C. and D. Bock (2003). A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistics Society 166*, 5–21.

Sonesson, C. and M. Frisén (2005). Multivariate surveillance. In *Spatial & Syndromic Surveillance for Public Health*. Wiley, Chichester.

Spencer, C. E., J. Marshall, R. Pirie, D. Campbell, and N. P. French (2011). The detection of sparially localised outbreaks in Campylobateriosis notification data. *Spatial and Spatio-temporal Epidemiology 2*, 173–183.

SRSA (2012). Swedish radiation safety authority. `http://www.stralsakerhetsmyndigheten.se/In-English/`.

Steutel, H. J. and K. Van Harn (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability 7*, 893–899.

Stoto, M., S. Matthias, and T. Louis (2004). Syndromic surveillance. an effective tool for detecting bioterrorism? *Chance 17*(1), 19–24.

Strat, Y. (2005). Overview of temporal surveillance. In *Spatial & Syndromic Surveillance for Public Health*. Wiley, Chichester.

Stroup, D., R. Brookmeyer, and T. William (2004). Public health surveillance in action: A framework. In *Monitoring the Health of Populations, Statistical Principles and Methods for Public Health Surveillance*. Oxford University Press, Oxford.

Stroup, D., M. Wharton, K. Kafadar, and A. Dean (1993). Evaluation of a method for detecting aberrations in public health surveillance data. *American Journal of Epidemiology 137*(3).

Stroup, D., G. Williamson, J. Herndon, and J. Karon (1989). Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Statistics in Medicine 8*(3), 323–329.

Ternhag, A., A. Tegnell, B. Lesko, K. Skaerlund, and P. Penttinen (2004). Basic surveillance network, a European database for surveillance data on infectious diseases. *Eurosurveillance 9*(7).

Testik, M. and G. Runger (2006). Robustness properties of multivariate EWMA control charts. *IEEE Transactions 38*, 635–645.

Testik, M. C., B. D. McCullough, and C. M. Borror (2006). The effect of estimated parameters on Poisson EWMA control charts. *Quality Technology & Qualitative Management 3*(4), 513–527.

Triantafyllopoulos, K. (2009). Time series. Technical report, School of Mathematics and Statistics, the University of Sheffield, UK.

Unkel, S., C. Farrington, P. Garthwaite, and N. Andrews (2012). Statistical methods for the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal Statistical Society: Series A 31*(1), 49–82.

Wang, L., M. Ramoni, K. Mandl, and P. Sebastiani (2005). Factors affecting automated syndromic surveillance. *Artificial Intelligence in Medicine 34*, 269–278.

Watkins, R. E., S. Eagleson, R. G. Hall, L. Dailey, and A. J. Plant (2006). Approaches to evaluation of outbreak detection methods. *BMC Public Health 6*(263).

Watkins, R. E., S. Eagleson, B. Veenendaal, G. Wright, and A. J. Plant (2009). Disease surveillance using a hidden markov model. *BMC Public Health 9*(39).

Weiß, C. H. (2009). EWMA monitoring of correlated process of Poisson counts. *Quality Technology & Qualitative Management 6*(2), 137–153.

Weiß, C. H. (2011). Detecting mean increases in Poisson INAR(1) process with EWMA control charts. *Journal of Applied Statistics 38*(2), 383–398.

Wessman, P. (1998). Some principles for surveillance adopted for multivariate processes with a common change point. *Communications in Statistics - Theory and Methods 27*, 1143–1161.

Wieringa, J. E. (1999). *Statistical Process Control for Serially Correlated Data*. Labyprint Publication.

Woodall, W. H. (2006). Use of control charts in health-care and public-health surveillance. *Journal of Quality Technology 38*(2), 89–104.

Yahav, I. and G. Shmueli (2012a). Directionally-sensitive multivariate control charts in practice: Application to biosurveillance. *To appear in: Quality and Reliability Engineering International*.

Yahav, I. and G. Shmueli (2012b). On generating multivariate Poisson data in management science applications. *Applied Stochastic Models in Business and Industry 28*, 91–102.

Zhou, Q., Y. Luo, and Z. Wang (2010). A control chart based on likelihood ratio test for detecting patterned mean and variance shifts. *Computational Statistics and Data Analysis 54*, 1634–1645.