Optimisation of Patient Monitor Alarm Settings Using Annotated Hospital Data

Matias Fontana

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 8.1.2017

Thesis supervisor:

Prof. Jari Saramäki

Thesis advisor:

D.Sc. (Tech.) Kimmo Uutela

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
MASTER'S THESIS

Author: Matias Fontana

Title: Optimisation of Patient Monitor Alarm Settings Using Annotated Hospital Data

Date: 8.1.2017 Language: English Number of pages: 7+61

Department of Computer Science

Professorship: Computational Science

Supervisor: Prof. Jari Saramäki

Advisor: D.Sc. (Tech.) Kimmo Uutela

Alarms are a key functionality of any clinical patient monitoring system. When developing a new system, alarm behaviour must be tuned towards maximising patient safety from all points of view. False alarms are a major problem in patient wards, and produce an effect called alarm stress, which may desensitise care staff and may endanger patients. On the other hand, systems must be sensitive enough to detect any clinically relevant alarm situation.

The purpose of this thesis is to study and optimise the alarm behaviour of a patient monitoring system. Annotated monitoring data from hospital tests is used as a reference when tuning system parameters. The platform can be used to rerun hospital test cases in the development environment and produce results.

The goal is to have a system that minimises false alarms and does not give false negatives. In this way patient safety is guaranteed from both ends; clinically relevant situations are detected, but care staff is not desensitised by too many false alarms.

Main results include optimised alarm configurations for the monitoring system, and information about the subjectivity of alarm relevance classification in a clinical monitoring context.

Keywords: clinical alarms, patient monitoring, false alarms, alarm stress, alarm fatigue, nuisance alarms

Preface

I would like to thank my advisor Kimmo Uutela for his guidance in this endeavour. Many thanks to Jari Saramäki, who supervised the work and gave valuable insight and comments. Also thanks to my family who supported me throughout this entire process.

Herttoniemi, 8.1.2018

Matias Fontana

Contents

Symbols and abbreviations

Abbreviations

GE	General Electric
SpO2	Peripheral capillary oxygen saturation
PR	Pulse Rate
RR	Respiratory rate
ECG	Electrocardiogram
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative

1 Introduction

Alarms are essential to any patient monitoring system. Their primary purpose is to ensure patient safety during possibly fatal situations by alerting care staff of imminent danger. Modern monitors are often configured to be extremely sensitive to changes in patient vital signs, and false alarms are common. Studies suggest that 80-99% of alarms in hospital wards are clinically irrelevant. Alarm may be technically true, yet clinically irrelevant. These alarms are called nuisance alarms. [26]

Frequent false alarms can desensitise care staff and cause an effect called alarm fatigue, which may affect the quality of treatment and even patient safety. Nurse response times are delayed, and overall faith in the alarm system is decreased. A noisy hospital environment can also affect patients negatively. Solutions for decreasing the number of false alarms are urgently needed. Concurrently, monitor sensitivity to actual dangerous situations must not be hindered. [20]

During development of new patient monitoring technologies and devices, a continuous dialogue between project stakeholders is being conducted. Engineers and clinical specialists discuss the configuration of the patient monitor with regards to alarm behaviour and other life-critical systems. This process is difficult, as there is no single truth about how monitors should be configured. Eventually, hospitals make the decision on configurations to be used. Apart from the clinical aspect, usability and marketing realities must be considered. In addition, current global regulation on medical technology is strict and may slow down or prevent the development of much needed improvements in monitoring alarm systems.

Continuous monitoring is traditionally used in high-acuity care environments such as Intensive care units and Operating rooms. Low-acuity continuous monitoring is less common, and usually intermittent spot checks are the prevalent monitoring technique. Different monitoring requirements associated with these environments are considered; it may not be feasible to apply similar alarming rules to low-acuity monitoring.

The purpose of this thesis is to use annotated patient data for optimising alarm system configurations of a patient monitor. The monitoring is continuous and done in low-acuity wards. Data is collected in an undergoing clinical trial and analysed. The goal is to study and improve the alarm behaviour of the system being tested, and improve patient safety.

The main results of this study are optimised alarm settings for low-acuity patient monitoring which minimise clinically irrelevant alarms. Valuable information about the nature of clinical alarm relevance and how to determine it was also found. The research methods applied proved viable for studying patient monitoring alarms, and can be reapplied in the future.

2 Background

Both scientific interest in the clinical monitoring field and the medical engineering industry contribute to this work. Being part of ongoing product development which may affect the lives of clinical patients makes the work feel especially meaningful and urgent.

The chapter will cover backgrounds of clinical patient monitoring and alarms as well as vital signs measurements briefly. A relevant physiological background is given with regards to principal issues related with the development. Proposed methods for the improvement of clinical alarms are listed as well.

2.1 Physiological background

There are several physiological realities which must be taken into consideration when developing patient monitoring systems and their alarms. Alarming conditions have to be determined so that patient safety is maximised. The measurements used in this thesis are mostly associated with tissue oxygenation, respiration and pulse rate. This section covers the basics of these topics, and identifies patient safety issues related to them. These factors affect the configuration and optimisation of alarm settings, and provide a medical point of view.

2.1.1 Tissue oxygenation

The human body needs oxygen for various metabolic reactions. Cellular respiration is the reaction where oxygen is used together with glucose to create the basic energy molecules used by cells. This process expels carbon dioxide, which is later exhaled from the body via respiratory pathways. Without oxygen, human cells cannot survive.

Tissue oxygenation is a circular process which starts and ends in the lungs. Its purpose is to supply oxygen (O2) to cells and expel carbon dioxide (CO2) from cells and the body. The process combines both respiratory and cardiovascular systems into a loop. The respiratory system handles gas exchange between air and blood, and the cardiovascular system transports oxygen and carbon dioxide between alveoli and other tissue cells. The complete process is shown in Figure 2.1.1.

The process begins with pulmonary ventilation (breathing), which brings oxygen-rich air into the respiratory system. Respiratory muscle activation creates a pressure difference between the lungs and ambient air, which moves air in and out of it. The respiratory system is a complex network of tracts and cavities. It starts with larger tubes and gradually divides into smaller airways, eventually connecting to small

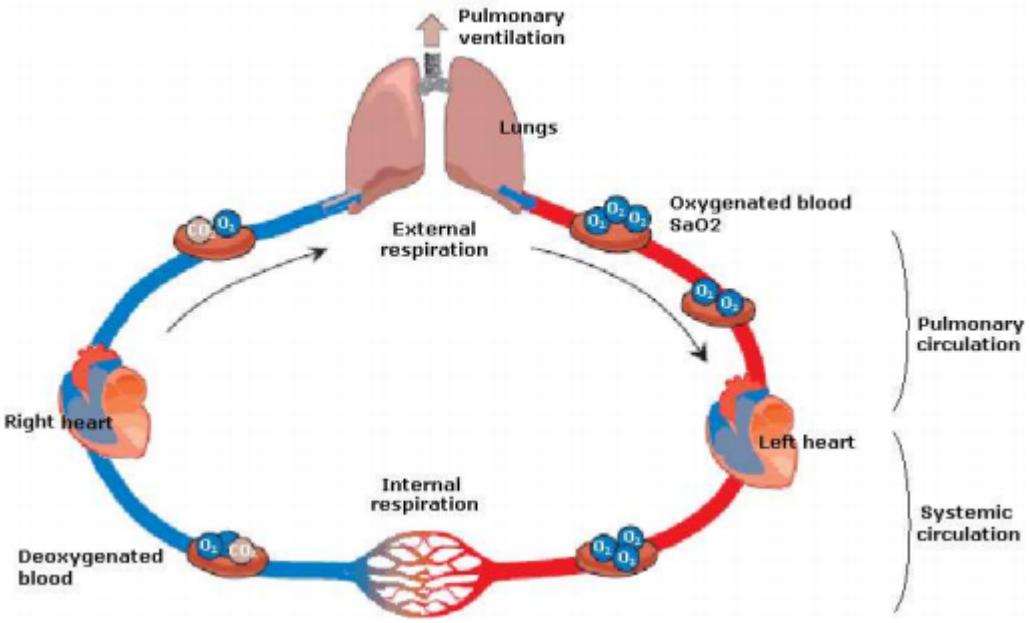

Figure 1: Tissue oxygenation process [35]

structures called alveoli, where the gas exchange between air and blood takes place. This part of the process is called external respiration.

In the alveoli, oxygen diffuses through thin alveolar capillary walls and binds to haemoglobin molecules in the blood. The haemoglobin molecule has a specialised iron (Fe) structure to which oxygen binds. A small fraction of oxygen simply dissolves into blood plasma. The main transporter of oxygen is the haemoglobin molecule. Oxygenated blood then exits the pulmonary circulation system into the heart and gets pumped into the systemic circulation system. There it is distributed throughout the body. Gas exchange between the systemic circulation system and tissues is called internal respiration, and happens only through thin capillary walls. Interstitial fluid mediates gas exchange between capillaries and tissue cells.

Blood loses oxygen and gains carbon dioxide during internal respiration. Carbon dioxide is also carried by haemoglobin, but does not bind to the same place as oxygen. Carbon dioxide is bound to protein structures in the haemoglobin molecule. Deoxygenated blood continues its flow from small capillaries into larger veins. Eventually blood flows back to the heart and is pumped back into the pulmonary circulation system and alveolar capillaries, where carbon dioxide diffuses back into air and gets expelled by pulmonary ventilation. After this the process starts all over again.

2.1.2 Hypoxaemia and hypoxia

Hypoxaemia is a state of abnormally low levels of oxygen in arterial blood. Hypoxaemia is typically caused by some kind of respiratory disorder, and may lead to hypoxia, which is a more general state of oxygen deficiency in the body. Common causes of hypoxaemia are various respiratory related disorders, ventilation/perfusion imbalance (air to blood ratio in alveoli), drug induced low respiration rate, shunting, heart problems, anemia, and apnoea. Symptoms of hypoxaemia are fatigue, shortness of breath, confusion, headache, and skin colour changes. Hypoxaemia may lead to hypoxia, but this is not always the case. Hypoxia is more severe than hypoxaemia. Both hypoxaemia and hypoxia are treated by giving the patient supplemental oxygen as soon as possible. [22]

Blood oxygenation monitoring aims at detecting hypoxaemia. Studies distinguish several levels of hypoxaemia and thresholds for said levels. Hypoxaemia can be split into three tiers of severity: mild, moderate and severe. Studies suggest thresholds for these are 85–90%, 80–85%, and 70–80%, respectively. There are no universally accepted limits for hypoxaemia. Most studies, however, seem to suggest that and SpO2 value of 90% is a good indicator of hypoxaemia. Monitoring guidebooks typically define two suggested thresholds, one for normal patient and one for patients with respiratory disorders. The latter group have lower thresholds suggested. This kind of adjustment can be made for any patient group that for some reason has generally lower blood oxygenation levels. For example, people who live in high terrain are accustomed to lower amounts of oxygen in the air they breathe, and have therefore lower SpO2 values as well. Supplemental oxygen is typically administered when saturation drops under the threshold consistently. [23] [16]

Hypoxia is a state of oxygen deficiency in cells. It happens when the amount of oxygen supplied to cells is less than the amount consumed by them. The main symptoms of hypoxia are fatigue, headache, nausea, hyperventilation, cyanosis (skin and membranes turn blue), seizures, coma, and death. Hypoxia can be difficult to diagnose, especially if its onset is slow. Extreme hypoxia is called anoxia, which means total oxygen deprivation of tissues. Anoxia can quickly cause irreversable damage to tissues. Those with fast metabolic rates are more quickly affected, for example brain tissue.[22]

2.1.3 Pulse rate and respiratory rate

Normal ranges for pulse rate and respiratory rate are defined as about 60–100 and 12–20, respectively. Alarming thresholds are wider, and are commonly set at about 50–150 and 8–35. There isn't as much debate on how to set these limits, as they are already well outside the normal range. For sick patients in hospitals however, both pulse rate and respiratory rate can be abnormal due to health status, medication, or even the hospital environment and care setting. Being in a hospital is stressful.[18]

[22]

Studies suggest that respiratory rate can be an indicator of serious health problems. Individuals with respiratory rates higher than 24 breaths per minute are probably critically sick and in life-threatening danger. Even going higher than 20 breaths per minute is a sign of being unwell. Alarm limits for respiratory rate are highly varied, but typically rates above 35 are alarmed. On the low end, alarming rates are normally under 10, with limits set to about 8 breaths per minute. [12]

Apnoea means a state without breathing, which is generally alarmed after about 20 seconds of no breaths detected. Apnoea typically happens during sleep or heavy sedation. Apnoea can be categorised into two distinct subcategories: central apnoea and obstructive apnoea.

Central apnoea is caused by the central nervous system not sending proper signals to trigger breathing. This may be caused by sedative medication or other drugs, but also various diseases. Central apnoea is a principal concern in post-operative monitoring, as patients are often sedated, and may experience central apnoea. Central apnoea is easier to detect with unobstructive respiratory measurements, as no breathing reflex happens.

Obstructive sleep apnoea happens when throat muscles relax too much and obstruct breathing airways. Symptoms range from simple snoring to longer periods of apnoea. Detecting obstructive apnoea may be tricky, as the breathing reflex happens normally, but airflow is obstructed, this requires a direct measurement of airflow. Apnoea may lead to serious complications and even death.

Typical patterns death in hospitals feature three kinds of respiratory behaviour:

1. Slowly accelerating RR value ending in rapid decrease and death (e.g. sepsis, congestive hearth failure, pulmonary embolism).
2. Slowly decreasing RR value ending in death (CO2 narcosis) .
3. Intermittent apnea periods with normal RR in between, ending in terminal apnoea (Sleep apnea with failed arousal).

These situations emphasise the importance of respiratory monitoring in preventing unexpected hospital deaths. SpO2 alone may not be able to detect these deadly patterns, as the measurement value does not decrease early enough. [27]

Tachycardia (high heart rate) means a heart rate of over 100 beats per minute. It can be caused by drugs or various hypermetabolic states. For example exercise, fever or sepsis. Tachycardia is usually a sign of systemic physiological perturbations, and underlying causes should always be identified.

Bradycardia (low heart rate) is when heart rate is under 60 beats per minute. Young adults and trained athletes can have bradycardia. Bradycardia may be caused by drugs and some physiologic states. Hypoxia and hypothermia can lead to bradycardia. Heart diseases may cause bradycardia as well.

2.2 Clinical monitoring

Clinical monitoring of patients can be done during hospital care or at home. This thesis focuses on hospital patient monitoring, but the underlying principles are transferable and can be applied to continuous home monitoring as well. The main difference factor is the status of the patient, which affects the requirements of the monitoring system. Patient monitors are used throughout hospitals from Intensive Care Units (ICU) to regular wards. Criticality requirements may differ, but the technical equipment is similar. Used vital sign measurements may also vary depending on the clinical needs.

The purpose of clinical monitoring is to communicate information on patients to care staff. Multiple vital measurements can be followed, and medical personnel can determine patient health status or make decisions on treatment based on said measurements. The four main vital measurements monitored are heart rate (HR), arterial blood oxygenation (SpO2), respiratory rate (RR) and blood pressure (BP).

Critical situations like surgery and intensive care units typically have continuous patient monitoring. Lower-acuity environments, however, often lack continuous monitoring and rely perhaps only on intermittent spot checks of the patient. Studies show that applying continuous monitoring may lower the number of adverse events in all wards.[3] [2]

Monitoring is done in real time, but collected data can also be analysed afterwards for various purposes. In this thesis, data collected from hospital tests is used to research patient alarm situations and further improve alarm configurations of the monitoring system. Numerous data-driven methodologies can be applied to further utilise monitoring data, ranging from simple patient population studies to predicting patient status by combining multiple monitoring parameters.

Doctors can also use collected data to analyse patient health better. An example of this is ECG monitoring, where the monitor makes heart condition classification decisions in real time. As mistakes are possible, doctors can re-examine the data manually afterwards.

Monitoring vital parameters can be used as a control tool for medical therapy equipment such as anaesthesia machines. In this case, the machine has monitoring sensors, which help control various functions of the machine itself. This combined system is critical to the well-being of anaesthesia patients during surgery or alike,

and offers invaluable help to anaesthesiologists.

Many other therapeutic devices have physiologic monitoring functionality as well, and create a closed loop system making therapy decisions. Separate monitor machines could also be used as controllers for other devices. Coupling monitoring functionality with therapeutically active devices establishes very high accuracy requirements for the monitoring system.

Monitoring provides various data outputs. Technical information about the monitoring system and process is displayed as needed. Clinical information about the patient is also shown, and can be interpreted by clinical staff during treatment. These information outputs are by nature passive, and require the user to actively watch the information stream. This is time-consuming, and often not continuously feasible in busy hospital environments. This is why monitors also have an active medium for transmitting patient information: alarms. Alarm are the most prominent and perhaps best known feature of clinical monitoring systems. They are the last link in the chain, letting care staff know something has changed for the worse.

2.2.1 Alarms

The word 'alarm' is originally from the Latin word 'ad arma', which means 'to arms' or 'to your weapons'. It is, therefore, a call to immediate action. Alarms have been around throughout history, and can be observed even in animals. At the most basic level, alarms are simply signals delivering information in one form or another.

The reality of alarms in the context of clinical monitoring is not as black and white as the previous etymological and biological definition suggests, but often quite fuzzy. The underlying principle, however, is the same.

Alarms are an essential part of any patient monitoring system. Their purpose is to inform care staff of some abnormality related to the monitored patient or monitoring system. They monitor vital organ functions and clinical device operation. The goal is to increase patient safety by early detection of possibly dangerous abnormalities. Finally, according to the definitions above, alarms always prompt some action as well. This can be immediate or delayed, but some kind of care staff action is expected following an alarm. Depending on the severity levels of the alarm, this action can be as subtle as slightly raised awareness of the patient's status, or as visible as resuscitation.

Monitoring systems communicate alarms to care staff primarily in two ways:

- Auditory: the alarm is indicated by a warning sound. Different sounds may be used for different alarms, or even spoken messages. Alarms type and severity

can be determined by sound.

- Visual: the monitor screen displays a text and/or flashes the alarmed parameter name or some other area. Monitors with many parameters may show multiple alarm conditions at the same time.

The required user action following an alarm depends on it's severity level. The hierarchy by which alarm signals are ordered is based on standards, which are defined by clinical risk analyses. For simplicity, they can be reduced low, medium and high levels. Additionally, some information messages may be transmitted by a visible or audible signal. These messages are not alarms, even though their information might still be important to the treatment of patients.

The idea behind alarm severity hierarchy is to determine the kind of action it prompts. Current alarm standards define the way an alarm affects care staff work.

1. Low: This severity level is meant to raise the awareness of care staff. Nurse workflow may be interrupted in the future, but no immediate action is required.
2. Medium: Requires prompt response from care staff. Current workflow likely has to be planned differently in order to tend to the event.
3. High: Disrupts current task and required immediate response from care staff.

Alarm severity differentiation is a useful tool for placing more emphasis on certain situations. The severity system of alarm may also be dynamic. This enables alarm states to escalate their severity level depending on the situation. Contributing factors may be further changes in parameter values, other parameters (multivariate systems) or time of threshold violation.

Alarm goals can be categorised hierarchically into five different classes. In this context, a goal means the situation or condition that the alarm is supposed to reveal. This categorisation contains information messages as well, and can be viewed as a holistic characterisation of the type of information that monitoring devices actively transmit to users.

- Life-threatening physiological situation: detecting critical changes in vital signs parameters which may lead to patient harm or death. Examples of this are asystole, apnoea and severe hypoxia.
- Life-threatening device malfunction: detecting critical problems with medical devices. For example disconnection from the patient or loss of power.
- Imminent physiological danger: detecting gradual changes in vital signs parameters which signal patient status deterioration over time.

- Imminent device malfunction: detecting problems with medical devices which may lead to malfunctions eventually.
- Diagnostic alarms: various information messages about non-critical clinical or technical conditions which may occur during care and affect treatment. Examples may be pathophysiological states of the patient, or information messages from medical devices.

Based on the list above, alarms can be categorised to clinical and technical alarms. The former is a diagnostic of the patients physiological status, and the latter focusing on the functionality of medical equipment.

The relationship between alarm goals and alarm severity levels is fuzzy at best. While the above list is intuitive in nature, it is not applicable to all types of alarms and situations. Some patient parameters behave differently than others, which affects the way alarm conditions can be defined and evaluated. Nevertheless, alarm severity is the main tool for distinguishing the criticality of an event.

This thesis focuses more on clinical alarms[26], as they are more complicated to evaluate. Humans are complex, and clinical alarm conditions are harder to define and categorise. Technical alarms are considered as well, but the main experimental work is focused on clinical alarms.

As invaluable as alarms are in preventing patient harm, they are not without problems. False alarms are commonplace throughout all monitoring systems, and cause significant problems in hospital environments.

Currently, the prevalent logic behind most monitoring systems is to make alarm behaviour as sensitive to potentially dangerous situations as possible. This compromises the specificity of alarm condition evaluation and leads to a high false alarm rate. The approach can be defined as 'better safe than sorry'. False alarms and their effects are covered in the next section.

2.2.2 False alarms

Frequent false alarms have been described as a "top ten" health technology safety concern by the ECRI Institute [25]. Why are false alarms such a hazard? First the underlying reasons behind the situation must be understood better.

As medical devices used during treatment grow in numbers, the amount of alarm generating systems increases as well. More vital parameters are monitored and different therapeutic machines used. Medical technologies are often technically uncoordinated, which may lead to alarm redundancies. This leads to a hospital environment with lots of alarms, which cause noise, distraction and overall disturbance.[26]

The majority of alarms in hospital are clinically irrelevant. Studies suggest the percentage of false alarms is around 80–99%. The main reason behind false alarms are movement artefacts and incorrect readings. Different studies report slightly different numbers, but the trend is clear. This means that only about 10% of alarms are clinically relevant. [26] [32]

Typical alarm systems have high sensitivity but low specificity. This trend can be observed in both positive and negative predictive values, which are low and high respectively. A common range for sensitivity and negative predictive value is about 90–99%. In contrast, positive predictive value and specificity are usually 5–27% and about 50% respectively. [26]

Metric	Formula
Sensitivity	TP / (TP + FN)
Specificity	TN / (TN + FP)
Positive predictive value	TP / (TP + FP)
Negative predictive value	TN / (TN + FN)

Table 1: Accuracy metric formulas. TP = True positive, TN = True negative, FP = False positive, FN = False negative

High sensitivity means that alarm systems are extremely sensitive to any abnormalities, and therefore almost never miss a real alarm. Low specificity values mean the system is inaccurate in distinguishing between real and false alarm conditions, which causes high false alarm rates. Positive and negative predictive values represent the percentage of true positives and true negatives out of all positives and negatives, respectively. [26]

The rationale of never wanting to miss a real alarm is clear behind these figures. However, technical challenges associated with signal acquisition and disturbances in the measurements are also a contributing factor. [26] [14]

False alarms are typically caused by either technical issues with the system, or artefacts in the analysed signal. Alarm limits, the alarm itself, and alarm algorithms can also contribute to false alarms being generated. As the underlying reasons can be so diverse, so must the classification when evaluating what exactly is false about the alarm. From a medical point of view, false alarms have two categories:

- False positive: alarms caused by a malfunction or defect in medical equipment.

Can also be caused by patient activity. For example, SpO2 alarms due to motion artefacts or a broken sensor.

- Irrelevant: Alarms caused by a correct alarm condition, but which do not have any clinical relevance and thus require no intervention. They are also called non-actionable or nuisance alarms. For example, SpO2 alarm with patients who have constant low blood oxygenation levels. Staff intervention may also cause these alarms. These alarms are technically true, but clinically false.

The need for clinical action is a good measure for the validity of an alarm. Alarm are sometimes categorised as *actionable* or *non-actionable*, depending on the need for clinical intervention.[14]

Nuisance alarms, alarms that are technically correct but clinically irrelevant, are the main problem with current alarm management. Technically false alarms and intervention alarms are usually resolved by fixing equipment and training staff in its correct use.

Nuisance alarms, on the other hand, are not so simple to tackle. They arise from the aforementioned rationale of extreme sensitivity in alarm systems, and therefore the only way to avoid them would be to change the prevalent alarm logic. This is more difficult a challenge than it may seem, as medical realities and patient safety issues quickly arise when contemplating these topics.[14]

Several studies have been conducted on the validity of alarms in both ICU and low acuity settings. Results of these studies all suggest high false alarm rates.

Studies show that care staff may be subjected to about 150–700 alarms per patient per day in the ICU and 5–10 in lower acuity settings.[14] [26] [36] When factoring in the lower nurse to patient ratio of low-acuity wards, it is unfeasible to use similar alarm criteria as in the ICU environment. [2] The load on clinical staff may become overwhelming and cause numerous potentially hazardous effects.

2.2.3 Alarm fatigue

Frequent alarms cause a phenomenon called alarm fatigue. A combination of excessive noise and constant alarms can stress nurses and other care staff to the point of affecting quality of care negatively. If the alarms generated are also clinically irrelevant, or false, the effect is amplified.

Nurses often describe their care environments as chaotic and noisy. A cacophonic and stressful mix of alarms, patients and other care staff. This kind of environment is difficult for people to be in and makes it difficult for nurses to do their job well.[20]

Alarm fatigue causes an effect called "Crying Wolf", named after a story of a

shepherd boy who cried "Wolf!" too many times as a joke. One day, when the wolf actually appeared and the boy again called for help, nobody came to his rescue. The story teaches that people become easily desensitised because of false alarms, and stop reacting to alarms correctly, or try to avoid them altogether. [14]

What are the principal effects of alarm fatigue? How does it affect the treatment of patients? How can it be managed? All of these questions are at the heart of current clinical monitoring alarm research. Answers are urgently needed.

The main cause for alarm fatigue is simply the huge alarm burden that medical personnel have to endure. Coupled with low nurse to patient ratios, care staff struggle to cope with the overwhelming situation. There are simply too many calls to action for people to respond to. Furthermore, as the majority of alarms are false, nurses lose faith in the alarm system and become understandably frustrated with it constantly alarming.[26][14]

This leads to overall desensitisation towards all alarms, which lengthens response times considerably and often causes personnel to set wider alarm limits, silencing alarms or even disabling them altogether. While changing alarm limits and disabling useless alarms is recommended, it must be done without decreasing patient safety. Overall, it would be better if the system would not require the user to do this too often.[26]

Longer reaction times and silenced alarms threaten patient safety. The basic function of the alarm system changes too as care staff bend rules and make up their own system for dealing with the alarm load. This is necessary from the nurses' point of view, but not intended when designing the alarm system. Potentially dangerous situations may be missed because of these interferences.[14]

Studies done on response times suggest that typical response times are long. Of course this depends on the hospital environment, more critical care environments have a much faster reaction time than less critical ones. Alarm severity levels also affect response times significantly.

The probability that nurses will respond to alarms in under 60 seconds may be as low as 10%. Common response reaction time in wards is between 1–10 minutes.[4][26]

An example of the dangers of alarm fatigue is a patient who died of cardiac arrest at a major Boston hospital in January 2010. The patient's heart rate had begun decreasing prior to the incident, and an alarm had gone off in the nurses' central station. Despite this, care staff didn't notice anything before a routine check-up later. A separate crisis alarm at the patient's bed had also been turned off for unknown reasons. A report released on the incident suggested "alarm fatigue" contributed to the death of the patient. [24]

In total it has been reported that around 200–500 patient deaths have been related to alarm fatigue between 2005 and 2008. Due to the complex way alarm fatigue affects care, it is possible that the actual number is higher.[24]

Hospital noise levels are on average around 72 dB during the day and 60 dB at night. This is much higher than the 40 dB guideline given by the World Health Organization. While it is unclear how significant a factor alarms are in hospital sound levels, they are nevertheless strongly associated with causing excessive noise.[31]

Alarm audibility is a serious issue. Too many overlapping alarm sounds and overall noisiness makes it difficult to distinguish alarm sounds from one another, or even hear them at all. This is an obvious risk, as sound is a primary way alarms are transmitted. Studies done in hospital environments suggest that, on average, clinical staff can correctly identify 50% of alarm sounds. This situation means a more coherent and coordinated alarm sound system is urgently needed. [4]

Alarm fatigue can be observed in patients as well. The heightened sound levels caused by constant alarming disturb patient sleep and overall wellbeing. A hospital is already a stressful place for patients, and additional cacophony and chaos certainly doesn't help. A noisy environment and lack of sleep can increase patient stress levels and worsen their conditions.[17]

Stress has been known to cause multiple adverse effects in humans. Although the mechanism of action is not perfectly known, stress effects can inhibit immune system functionality and cause various disorders. People under stress are at greater risk of getting sick or even dying. [22]

In summary, alarm hazards consist of a causal chain originating from having too many false alarms, and ending in various adverse effects. These effects can be observed not only in care staff, but patients as well. [5] The following table summarises problems, causes, effects and possible solutions to this multi-faceted issue.

Problems	Causes	Effects	Solutions
• Irrelevant alarms • Noise • Alarm interpreting	• Sensitive systems • Simple alarm logic	• Staff and patient annoyance • Low faith in alarm system • Increased response time • Decreased safety and performance	• Better audio • Smart alarms • Alarm protocols • Integration between alarms • Alarm configurations

2.2.4 Alarm improvements

It is clear that something needs to be done to improve current alarm characteristics of monitoring systems. Modern monitor alarms are lagging behind the development level of other medical technologies by many years. There are many reasons for this, ranging from technical challenges to legal and regulatory restrictions. [26] [7] [19] Several solutions have been suggested to combat high false alarm rates. These methods can be roughly divided into three levels of operation:

1. Technical level (signal acquisition)
2. Alarm logic level (alarm generation)
3. Alarm validation level (alarm validation)

The first level deals with sensor technology and how the actual measurement is done. After sensor contact, the parameter is typically filtered somehow as well. Improving the system at this level means developing better sensor technologies and filtering strategies.

The second level is where the alarm state is evaluated. Common monitoring systems use a simple threshold, which triggers the alarm. Pattern recognition is used for arrhythmia monitoring. Asystole of apnoea states are determined by the pause time of events, beats and breaths, respectively.

The third level is the validation of the alarm against some other criteria. The simplest case is no validation at all, which means alarms are activated instantly. Typical systems apply some time delay for actually sounding the alarm. Multivariate alarm methods may compare parameter values together to validate an alarm state. For example, heart rate and pulse rate may be compared to determine heartbeat frequency.[26] Proposed methods for improving alarm behaviour can be further divided into two principal categories:

1. Technological
    - Technical sensor improvements: improves measurement accuracy
    - Statistical methods: Mean- or median-filtering, linear trend approximation, regression analysis, artifact filtering.
    - Autoregressive models and self adjusting thresholds: Low order AR models for physiologic variable modelling.
    - Phase-space embedding: Parameter time series transformation to phase-space models.
    - Trend detection and curve fitting: Trend detection, dimension reduction for multivariate data.
    - Artificial intelligence: Knowledge-approach, machine learning, fuzzy logic, bayesian networks, neural networks.
2. Hospital protocols
    - Alarm response protocols: Alarm management, care staff knowledge of alarms.
    - User configuration of alarms: Patient customisation, criticality settings, optimal settings.
    - Proper use of medical equipment: Medical device positioning, alarm management, patient context.
    - Standardisation and coordination of alarms: Eliminating redundancies, alarm standards.

Methods from both categories are applied in this work, focusing mostly on configurations and statistical methods for false alarm reduction. The alarms evaluated in this work are univariate threshold alarms (except apnoea and technical alarms), so experimenting is limited to configuring thresholds, delays and averaging.

2.3 Alarm configurations

When determining optimal configurations for a clinical monitoring alarm system, many physiological and medical issues must be taken into consideration. This thesis

aims at finding optimal solutions for minimising false alarm rate without missing clinically relevant alarms. As a result, patient safety is improved.

The main factors considered are alarm *thresholds*, alarm generation *delays* and *averaging filter lengths*. While these can be viewed as three separate factors, they also contribute to one singular question: *If something changes, how long until an alarm is triggered, if at all?*. All three factors basically move the alarm generation trigger further away from normal values. We can therefore construct a relative measure for the total delay of the alarm. First, it takes time for the vital parameter to reach the alarm threshold. Second, the absolute amount of delay for the alarm to trigger is added to that time. Third, parameter value averaging effectively lengthens threshold violation time as well. Of course we must also view threshold, delay, and averaging as separate factors when determining optimal configurations. However, the *total delay* compound measure is a handy tool for modelling the effect of different alarm configurations. It is relative, which means that a base level must be first set before calculating it.

Some vital sign values change differently than others. Respiration rate and heart rate changes can be so quick that modelling them impossible and unnecessary, and therefore the alarm limit's effect on total delay is negligible. In a crisis, they may change instantly, unlike SpO2 which is slower and monotonic. Blood oxygenation level has a relatively low change rate measured in seconds per unit, which makes it possible to derive models of blood desaturation during apnoea or insufficient oxygen supply. A model was developed by A.D. Farmery and P.G. Roe on blood oxygen desaturation rate during apnoea. From it we can derive that it takes about four seconds for a 1 unit drop in SpO2, 2 seconds for post-op or obese individuals. Children have a faster rate as well. Figure 2.3 shows a graph of the desaturation model.The model is used in the practical section of this work. [1]

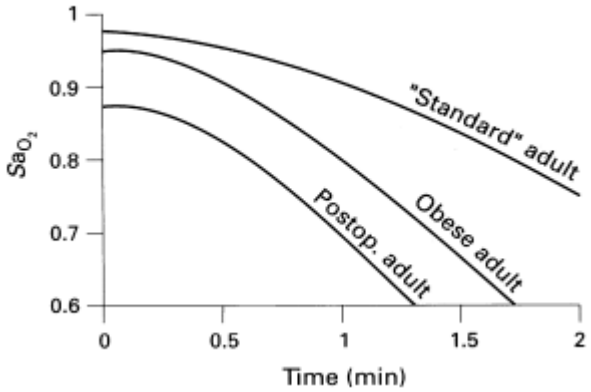

Figure 2: Blood oxygen desaturation rate [1]

There have been numerous studies on how changing these factors affects alarm

behaviour, and they suggest that even with relatively small relaxations of alarm settings, significant alarm burden reductions can be achieved. However, patient safety issues often arise if settings are relaxed too much, and different results may conflict.

This thesis only covers pulse oximetry and respiratory monitoring, so all electrocardiograph related alarms are left out. Without ECG-events, the bulk of the alarm burden seems to be pulse oximeter related, and other studies confirm this.[9] [15] Most studies focus on the SpO2 alarm, which is more common than the pulse rate or respiratory rate alarm. However, the underlying principles of configuring the alarm are the same, and can be applied to all alarms. Usually its a question of either *safe thresholds* or *accepted time of violation*. Common thresholds for SpO2, pulse rate and respiratory rate are:

Table 2: Common alarm thresholds

	SpO2 (%)	PR (bpm)	RR (bpm)
Low	90	50	8
High	N/A	150	35

Common thresholds and alarm rates for SpO2, pulse rate and respiratory rate are shown in Table 2. Some hospital protocols allow alarm thresholds to be varied by the user, others do not. Changing thresholds enables clinical staff to adapt settings for each patient. For example, patients with continuously low blood oxygen saturation may need to have alarm thresholds changed in order to avoid false alarms. Alarm settings have been defined using medical knowledge of normal ranges and patient population observations. Figure 2.3 shows typical distributions of parameter values. It it hard to define universally valid limits for these vital signs, especially with different patients and criticality levels. [11]

A study by James Welch et al yields plenty of results on the effects of threshold and delay changes. A large multi-parameter database was collected and used for simulating alarm behaviour depending on various settings. The database contained 94 575 hours of patient data. Results suggested that false alarm rates could be reduced significantly by changing configurations. The study concluded that low alarm rates could be achieved by selecting alarm limits as the 0.5–1.0% points of the parameter value distribution ends, depending on desired limit. Delays should be chosen so that quickly resolving violations are not evaluated as alarms. Delays ranged from 0 to 120 seconds. The methodology used in this study proved to be viable for data-driven optimisation and development of patient monitoring alarm systems.[8] [28]

Other studies seem to report similar results on the effects of threshold and delay configuring. Alarm were often reduced by over 50% and up to 90% with some configurations. Delays and averaging seem to have a stronger effect than thresholds only. This may be due to the *small and brief* nature of threshold violations. The

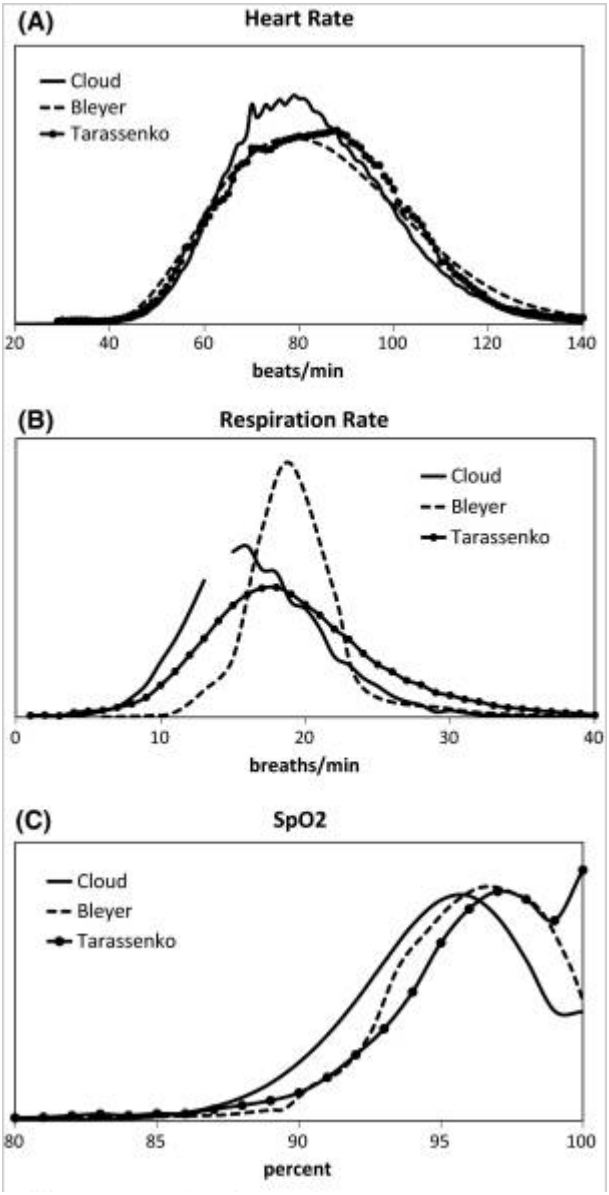

Figure 3: Parameter value distributions from different studies [8]

parameter value drops or jumps suddenly out of the limits for a short period of time. This can be effectively countered by delaying triggering of the alarm. [6] Usually these brief threshold violations can be physiologically normal states that correct themselves quickly. Artefacts may also cause brief violations.

In summary, most pursuits for reducing false alarm burden suggest lowering limits, settings delays, and averaging the parameter value. How these configurations should be made is another matter. One study suggested having default alarm settings, of which nurses could tweak limits by 10%. Further adjustments would have to be approved by physicians.[3] Delay and averaging changes done by users were not suggested in any studies. However, clinical staff can often silence alarms for a short

period or completely. This is normal and helps reduce irrelevant alarms that are caused by treatment.[14]

2.4 Monitoring parameters

The most common vital measurements monitored are pulse oximetry, electrocardiography, blood pressure and respiration. The monitoring system used while making this thesis is limited to pulse oximetry and respiratory monitoring.

2.4.1 Pulse oximetry

Pulse oximetry is a noninvasive method for measuring blood oxygen saturation and pulse rate. It provides two important vital sign parameters: SpO2 and heart pulse rate. SpO2 is arterial oxygen saturation (SaO2) measured by pulse oximeter. Actual SaO2 value may vary slightly, and is only measured accurately by sampling the blood itself. The SpO2 value is generally correlated well enough to SaO2 that it can be used in clinical monitoring as a vital parameter of patient status. Pulse rate is a measurement of arterial blood pulses, which result from heart beats. Pulse rate is usually the same as heart rate, which is derived from the electrical activity of the heart. Some arrhythmias may cause the two parameters to behave differently, which is why monitoring both is useful.[29]

Blood saturation measurements are based on the principle of different absorption spectra of oxyhaemoglobin and reduced haemoglobin. Typically two wavelengths of light are used for the measurement, red and infrared are the most common. The absorption of these two wavelengths after transmission through the skin is measured. This is called plethysmographic signal, from which SpO2 and pulse rate are extracted.[29]

Pulse oximetry sensors are highly sensitive to motion artefacts, incident ambient light and other disturbances. The extracted plethysmographic signal can be very noisy, and parameter values may change rapidly. When the signal is normal, however, pulse oximetry is a very reliable and good measurement. Because it's non-invasive, pulse oximetry is an easy parameter to monitor and does not obstruct the patient too much either. It is therefore suitable for continuous monitoring as well.[29]

2.4.2 Respiratory monitoring

Respiratory rate and gas exchange monitoring can be performed with multiple different techniques. These can be divided into three groups:

- Direct methods: these methods measure respiratory gas flow directly. For example, capnography, where the CO2 concentration of respiratory gas is measured with a nasal cannula.
- Indirect methods: measurement is done by looking at other physiological changes in the body associated with respiration. For example, impedance pneumography, where impedance changes over the chest are measured to get respiration rate.
- Derived methods: these methods use physiological signals primarily meant for other measurements, and extracts information that can be further processed to derive respiratory rate. An example is ECG-derived respiration.

The monitoring system tested in this thesis is based on the second group. Impedance pneumography is the base technology applied. The measurement is based on the principle that the impedance over the chest area is modulated by respiration. Electrodes are used to measure impedance, and the resulting signal is analysed and respiration rate is determined from it. Impedance respiration is extremely susceptible to movement artefacts, but provides a practical and unobstructive method for measuring respiratory rate. Therefore it is also suited for continuous monitoring.[30] [13]

2.5 Mathematical optimisation

This section briefly covers basic concepts of mathematical optimisation. First, different optimisation problems are characterised. Then, common optimisation techniques are discussed. This section supports the decision about which optimisation technique is applied in the practical part of this work. More detailed information on algorithms that were considered and eventually chosen for this thesis in section 3.7.

2.5.1 Optimisation problems

Optimisation problems are ubiquitous in any engineering field. The goal is typically either minimising or maximising a function value. The general definition for an optimisation problem is:

$$\begin{aligned} \underset{x}{\text{minimise}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq b_i,\ i = 1, \ldots, m. \end{aligned}$$

There are many different characteristics that define an optimisation problem. These qualities may also limit the use of some optimisation methods. They range from

real-world limitations to purely mathematical aspects. The most basic definitions of an optimisation problem are:

- Constraints
- Linearity
- Discrete or continuous
- Static or dynamic
- Dimension of input
- Dimension of output

Constraints limit the search space, and must therefore be taken into account when solving the problem. Solutions outside the constrained space are not determined. The linearity of a problem must also be considered, as non-linear functions tend to have local optimums, which makes it more difficult for optimisers to converge on a global optimum. Whether the solution space is discrete or continuous matters too, and limits the ways in which new candidate solutions can be generated. The difference between static and dynamic problems is that the latter changes over time, which means the optimiser must provide an optimal sequence to the problem. For static problems this is not the case. Input dimension adds to search space size. Output dimension refers to the number of targets to be optimised, which adds complexity to the problem.[33]

2.5.2 Optimisation techniques

Below are listed some common optimisation techniques:

- Convex optimisation: Objective function and constraints are convex.
- Integer programming: Some or all input variables are integers.
- Nonlinear programming: Part of the problem is nonlinear.
- Stochastic programming: Function or inputs include random elements.
- Combinatorial optimisation: Set of feasible solutions is discrete.
- Heuristics and metaheuristics: Few assumptions are made on the problem, optimal is sought by educated guess , optimum is not guaranteed to be found.

The ability to use a technique depends on the problem at hand. If the problem can be represented in a convex way, convex optimisation may be used. Linearity and other problem characteristics also shape the technique used for solving it. Most optimisation techniques rely on formulating the objective function and constraints mathematically, and applying various operations on them, like differentiation. Also, if the problem is continuous, it is possible to use mathematical methods to determine which way to look the the optimum. An example of this is the gradient descent method. However, if a problem cannot be represented in this way, or is not continuous, we may have to rely simply on making educated guesses for the optimums. This is where heuristics are useful, and many hard problems have to be solved by heuristic optimisation techniques.[21]

2.5.3 Multi-objective optimisation

Problems which have multiple targets to be optimised are not only more complex than single target problems, but also don't have a singular optimum value. Instead, the optimum is a combination of the output values. This solution set can be further processed in a set of *dominate* solutions, which is known as a Pareto optimal solution set. These solutions are optimal in the sense, that there is no way of improving one target without worsening the other. In the simplest two-dimensional example, we have a line which represents the optimal combinations of the two variables. The final decision is always a trade off, and additional rules are needed to determine which combination is chosen. These rules may include hierarchical ordering of the targets, or using a weighted sum for choosing the optimum. [34] [10]

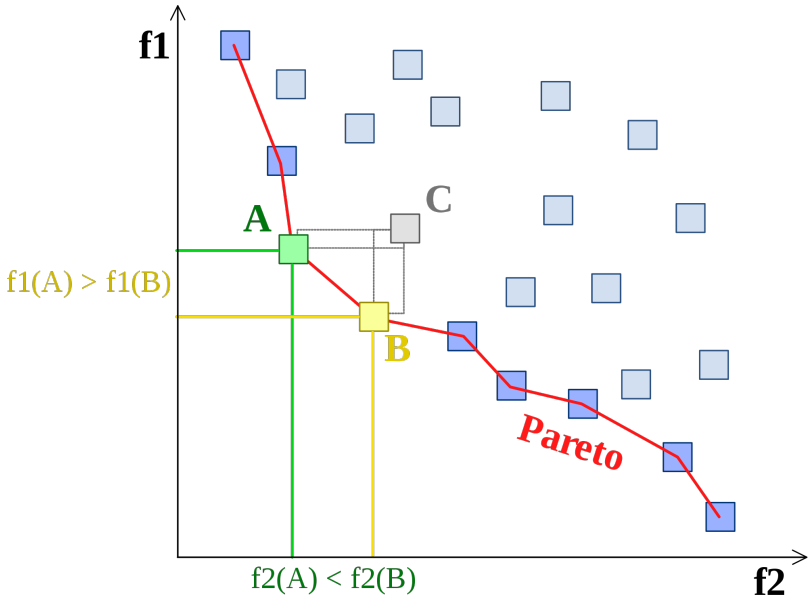

Figure 4: Pareto front of optimal solutions [37]

3 Materials and Methods

This chapter is about the materials and research methods used in this thesis. Hospital test data is analysed and descriptive statistics calculated from it. Alarm behaviour baselines are also shown in this chapter. All material used in this thesis comes from hospital tests, apart from the algorithms themselves, which are developed on site. Hospital material can be divided into raw patient data, vital sign data, alarm data and nurse annotations. The first three are relatively straightforward and are used mainly as is. Nurse annotation data is subject to human-factor effects, which contribute to possible variation, inconsistency and incompleteness of the data. Annotations were eventually reviewed a second time, as inconsistencies with testing protocols were detected.

The main goal of the work was to create a development platform for analysing and optimising the alarm behaviour of the clinical patient monitoring system. The developed platform offers capability to annotate hospital test data, rerun said data with different algorithm versions and system settings, and producing various results and analysing those results further. Main outputs are overall alarm statistics and optimised configurations results. The platform can also be used for investigating problems within the system. It is a tool for both low level development work and higher level system evaluation.

3.1 Hospital data

Hospital data consisted of multiple signals, trends and logs. During hospital tests, the monitoring system collects all patient data and derived parameter values with logs. This data is first saved on the monitoring device itself. After the test case, the data is transferred from the monitor to a computer. The data is subsequently transferred to a local engineering database, where it is preprocessed before being analysed by the testing framework. The data used by the testing framework is listed below.

1. Raw signal data
    - Composite plethysmograph (100 Hz, *absorption*)
    - Electrocardiograph (500 Hz, mV)
    - Impedance pneumograph (125 Hz, $mOhm$)
2. Parameter trend data
    - SpO2 (1 Hz, %)
    - Pulse rate (1 Hz, *beats per minute*)

- Respiratory rate (1 Hz, *breaths per minute*)

3. Log data
    - Alarm log (1 hz, *active alarms*)

3.2 Patient population

Clinical tests were conducted at Helsinki University Hospital in Meilahti, Helsinki. Patiets were being treated in three different wards: Emergency Ward, Vascular Surgical Ward, and Cardiac Surgical Ward.

There were 70 patients enrolled in the study. Of those, 63 were adults and 58 of those were included in the study. Three enrolled patients refused the study afterwards. Other reasons for data exclusion were corrupted data and experimental testing protocols which were not studied further. Child patient data (n = 7) is treated separately, and may be used in future studies.

Data used for evaluation was in total 118.2 hours. Patient gender distribution was 79.3 % male, 20.7 % female. Descriptive statistics of patient age, height and weight are shown in table 3. Patient distribution between the three wards was: emergency ward (n = 27), vascular surgical ward (n = 11), and cardiac surgical ward (n = 20). No adverse effects were recorded during the study.

Hospital tests were conducted with one or two research nurses per case for reducing possible variation between testing staff. Some variation may still exist due to cases with only one testing nurse. A thorough review of test results was conducted afterwards, which aimed at reducing variance of results between different testing staff.

Table 3: Descriptive statistics of patient population

Patient info (n = 58)	Mean	Median	SD
Age	61.48	64.0	13.83
Height	174.40	175.5	8.17
Weight	80.68	79.0	79.0

3.3 Development platform overview

The platform and processes used for research and development in this thesis have three main phases of operation. First, patient data is collected during hospital tests. Alarm behaviour is annotated by clinical specialists on-site, and all the data is then transferred to the engineering database.

Second, system monitor algorithms are compiled as MATLAB MEX functions, which enables the algorithms to be run within the development framework in MATLAB as mostly black-box functions. The annotated alarm log is further post processed and reviewed into a reference for correct alarm behaviour.

Third, the original patient data from the hospital is rerun on the platform with different algorithm versions or system settings. The parameter values and alarms from this run are then compared against the reference data and final results are produced. The system offers high flexibility in analysing results further. An optimisation framework built around the base functionality of the system can be used to find optimal configurations for alarm configuration settings.

Figure 3.3 displays the high level structure of the entire system. The practical work done in this thesis is mainly focused on Phase II and Phase III sections in the diagram, namely data rerunning and optimisation against references, finally producing statistics and optimised configurations.

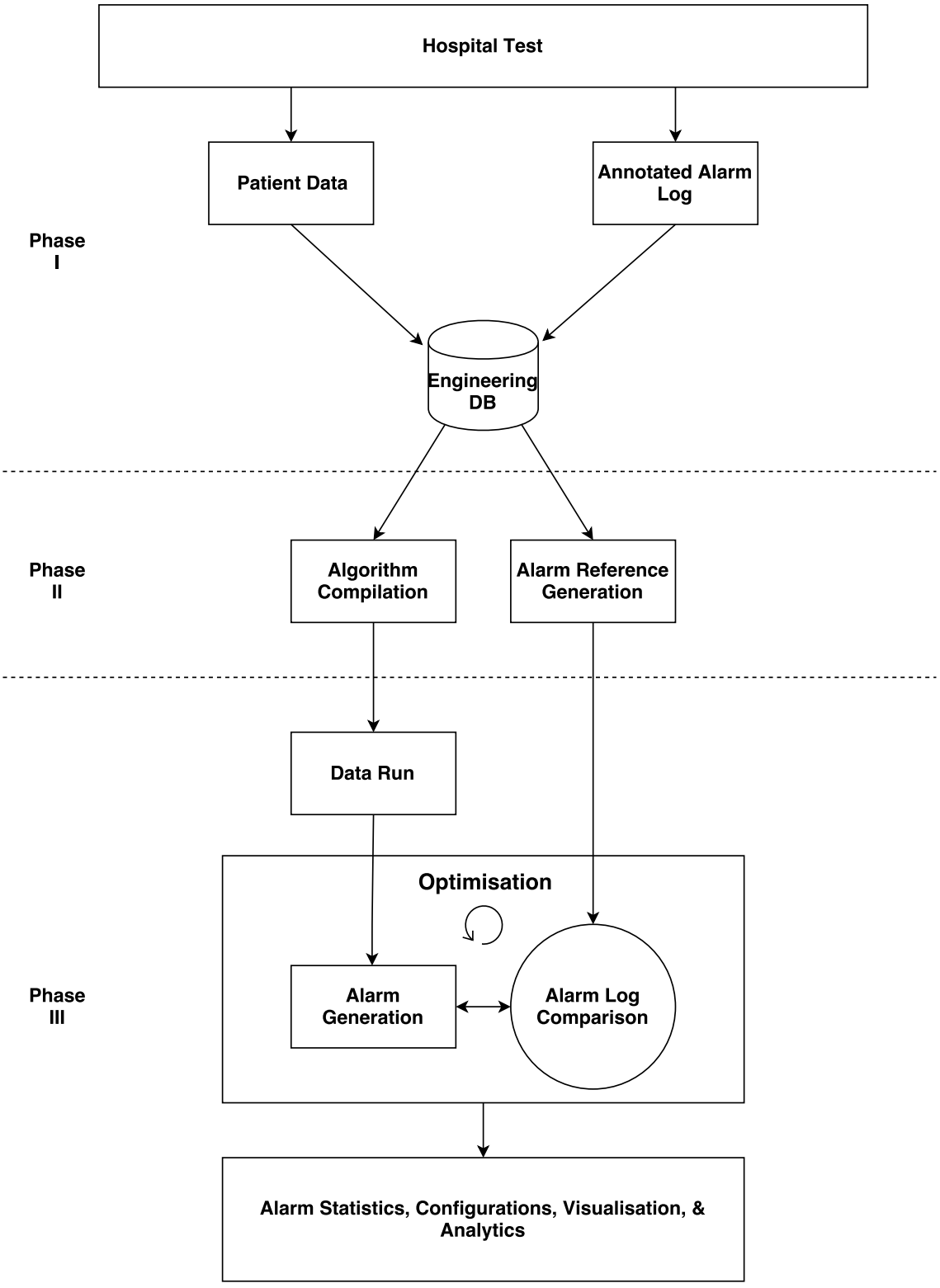

Figure 5: High level platform structure

3.4 Hospital data annotation

The monitoring system is being tested in Helsinki University Hospital in Meilahti, Helsinki. The tests produce a lot of different data ranging from usability results to raw patient vital signals. For the purposes of this work, the most relevant data are raw patient data, calculated parameter values and alarm logs. Overall system logs are also collected for debugging purposes.

3.4.1 Hospital tests

Hospital tests were typically 2.5 - 3 hours long, on average 2.7 hours. Due to various experimental procedures and test scenarios, some test cases were not fully analysed in terms of alarm behaviour, and are therefore truncated by a specific time interval. For example, the pulse oximetry probe was tested at multiple different locations, some of which weren't as good as others. Other anomalies during testing were also discarded. These included hardware-related problems and various test protocol abnormalities especially in the first phases of system testing. Taking intervals into consideration test cases time mean was 2.0 hours.

Tests were conducted in three different wards and with different patients. See 3.1 for more information. Patients can be both bedridden or ambulatory, and the system is tested during patient transport as well. For algorithm and alarms development it is important to collect data from many different monitoring scenarios, which gives a more holistic picture its actual use environment.

3.4.2 Alarm event annotations

The collected patient data is sent back from the hospital as is. The system alarm log, however, is more scrutinised. Clinical specialists on site use a tablet computer with an annotation program to annotate any alarm situation that the system under test may report. This first phase annotation program was developed by René Coffeng at GE Healthcare. It produces timestamped events which contain various information. The test nurses are able to add anything of interest to the annotation log file. This file uses a specific format, which enables the MATLAB framework to automatically read and parse it, extracting the required data. Typically, the annotated log contained information on the patient's physical context, activity or overall state. Reference measurements were also conducted to compare with the monitors own measurements. Lastly, the log contained annotations for alarm situations that happened during the testing period.

Alarm events are classified into six categories with the following criteria:

Table 4: Classification of alarm events and their explanations

Class	Definition
True	Clinically relevant and actionable alarm.
False	Technically incorrect alarm, caused by malfunction or wrong measurement.
Possible	Alarm condition correct, clinical relevance depends on preference.
Irrelevant	Alarm condition correct, but clinically irrelevant and non-actionable.
N/A	Caused by testing protocols or interventions. Left out of analysis.
Unknown	Alarm cause unknown, should be investigated. Left out of analysis.

For optimisation purposes, only True, False, Possible, and Irrelevant are important. "Not applicable" and "Unknown" classes are diagnostic classes used to discard events completely or prompt further investigation, respectively. If the cause of an Unknown alarm remains unclear, it was left out from the optimisation process, but added to the final alarm statistics as such.

Classes True, False, Possible and Irrelevant are used for both analysis and optimisation by the system. The main factors differentiating the classes is the type of clinical action needed and technical correctness of the alarm. The required action may range from heightened awareness of the patient's status to a physical intervention. In the end, the classification depends on the nurse's interpretation of the event.

The difference between classes True and Possible is that the latter may be viewed as useful, even though it doesn't prompt further clinical intervention. True cases always prompt intervention. For this study, alarms were annotated True only if testing nurses felt they had to intervene somehow with the treatment of the patient, or if clinical intervention was required following an alarm. Alarms classified as Possible were cases where the alarm condition was correct, and deemed clinically relevant. These situations did not have to prompt an immediate response or change in therapy. However, they were considered clinically relevant in the sense that they reflected the patient's health status and provided valuable information. This process is better

explained in section 3.4.3

In this thesis, results are presented as two system profiles, *sensitive* and *insensitive*, which regard possible alarm as useful alarms or nuisance alarms, respectively. In practice, the sensitive profile considers Possible alarms strictly required, while the Insensitive profile only requires True alarms. Table 5 illustrates how the classes relate to clinical and technical relevance.

Table 5: Matrix of clinical relevance and alarm classes

	Relevant	Irrelevant
Clinically	True/Possible	False/Possible/Irrelevant
Technically	True/Possible/Irrelevant	False

These categories are used to create statistics of the system's alarm behaviour. The difference between the classes is quite fuzzy, and it is understandably difficult to reach a consensus. In fact, annotations had to be reviewed again during the study. Different interpretations of defined classes led to changing annotation procedures, which had to be reviewed again to achieve a more consistent testing protocol. As testing matured, annotations were more in line with definitions. Unclear annotation cases were reviewed again and a final decision was reached. Guidelines for annotations were given to clinical specialists prior to testing. These included the definitions in table 4, excluding the *Possible* class, which was applied to annotations using the questionnaire mentioned in the next section. However, even with the guidelines, alarm classification is often subjective.

During development, different stakeholders and parties may even disagree on the right answer, which complicates things further. It is, however, the nature of life sciences to be inexact at times. The common goal is still to maximise patient safety. There are numerous factors in this equation, which makes any solution a compromise of sorts.

3.4.3 Annotation review process

The first round of annotations can be challenging, especially if the system reports multiple alarms with high frequency. The nurse might also need to tend to the patient or system hardware, which might cause incomplete annotations for some alarms. This indeed happened frequently, especially in early stages of testing. To combat this, a second two-stage annotation round is done offline after the test has terminated. The system does some automatic preprocessing and heuristic analysis on the annotations in order to correctly complete the log, but user input is often required nonetheless.

For these purposes, the platform offers a simple Microsoft Excel and PDF workflow

for reviewing online annotations. Users can look at each alarm event separately with annotations and parameter waveforms around that specific event displayed. Once the annotations are completely filled, the system post-processes the completed log into a reference file ready for comparison. Figure 3.4.3 shows an example of the annotation process. The second round of annotations improved the situation by filling gaps in the annotations and yielding a complete reference for the system to analyse.

10-11-2016 12:57:31	10-11-2016 12:58:25	medium	spo2_low	N/A
10-11-2016 12:58:36	10-11-2016 12:58:37	low	spo2_probe_off	N/A
10-11-2016 12:58:45	10-11-2016 13:02:45	low	spo2_check_probe	TRUE
10-11-2016 13:02:45	10-11-2016 13:03:16	medium	spo2_low	FALSE
10-11-2016 13:02:58	10-11-2016 13:03:18	medium	spo2_pr_low	
10-11-2016 13:04:04	10-11-2016 13:04:25	medium	spo2_low	FALSE
10-11-2016 13:05:21	10-11-2016 13:05:22	low	spo2_check_probe	Unknown
10-11-2016 13:05:31	10-11-2016 13:05:48	low	spo2_check_probe	Unknown
10-11-2016 13:05:48	10-11-2016 13:07:00	low	spo2_low	
10-11-2016 13:05:59	10-11-2016 13:07:00	medium	spo2_pr_low	
10-11-2016 13:09:10	10-11-2016 13:11:09	medium	impresp_apnea	Irrelevant
10-11-2016 13:11:09	10-11-2016 13:11:29	low	impresp_rr_low	Irrelevant
10-11-2016 13:11:23	10-11-2016 13:11:29	low	spo2_low	FALSE
10-11-2016 13:11:29	10-11-2016 13:12:20	medium	impresp_apnea	
10-11-2016 13:12:20	10-11-2016 13:12:37	low	impresp_rr_low	Irrelevant
10-11-2016 13:12:30	10-11-2016 13:12:36	low	spo2_low	TRUE

Figure 6: Example snippet from annotation the tool

A final step was taken in order to finalise annotations. Nurses were asked about situations where they had to intervene somehow. During the test case, the research nurse is not supposed to treat the patient in any clinical way, but only focus on the equipment under test and test protocols. However, if a clinically relevant situation arises which may indicate a potential adverse situation, the testing nurse will intervene if others haven't done so already. This questionnaire resulted in one single event where the research nurse informed care staff about a physiological alarm, which led clinical action. This single event was the only clinically relevant and actionable alarm recorded during the study.

Alarm events which were originally classified as clinically relevant were renamed *possible*, and used as constraints for the aforementioned *sensitivity profiles* of the system.

3.5 Algorithm compilation

The algorithms used in the patient monitoring systems are written in either C or C++ programming languages. They are compiled for the target hardware architecture, either the monitor itself or a sensor. In order to execute the algorithms on the development platform, they are compiled as MATLAB MEX functions, which can be executed as normal functions from MATLAB. The platform automatically compiles both the original algorithm libraries and MATLAB MEX functions. The algorithms are not touched after this, and apart from some input parameters upon function calling, they are a black-box from the system's point of view.

3.6 Platform structure

The development platform itself is a collection of MATLAB functions. System execution is semi-automatic or automatic, depending on the required action. This ensures flexibility and enables easy debugging of possible problems. Most information is saved during system execution, within space and convenience limits. It is therefore possible to monitor system behaviour across the entire data path. This functionality is invaluable for debugging purposes and also for more in-depth analysis of test cases and algorithm malfunction.

3.6.1 Typical workflow

Typical system workflow begins with defining execution settings and data in the main configuration file. The platform offers extensive configuration options and also a sweep capability across multiple settings values, which is useful for comparing different settings. After configuration, and provided that hospital reference data and raw patient data is available, the run is executed automatically and results are returned to the user.

3.6.2 System output

There are two principal outputs of the system. The first is overall alarm behaviour, which is reported as a table and various graphical visualisations. The second is information about optimal settings as calculated by the optimisation framework, see the next section. Optimisation results are partly reported visually as well.

The system also records intermediate results and data for each test case. This can be accessed in MATLAB for more in-depth analysis of possible anomalies in the algorithms or patient data. Additionally, the system features some specialised

reporting tools for visualising test case specific alarm behaviour and results for sweep runs. Below is a list of the main data outputs of the system:

1. Statistical results
    - Descriptive statistics of data
    - Alarm rates categorised as required
    - Accuracy values
2. Graphics
    - Histograms
    - Bar charts
    - Tradeoff pareto plots
    - Alarm rate surfaces
    - Heatmaps of alarm occurrences

3.7 Optimisation

The main goal of the optimisation is to find optimal settings for the system. The main settings to be optimised are alarm limits, delays, and filtering configurations. Alarm limits and delays are the two most influential factors to alarm behaviour. Moving averaging is applied to each parameter, and can affect the end results as well. On the other hand, the total delay applied to generating the alarm must also be kept small enough in order to ensure patient safety. This makes the optimisation problem multi-objective, and means that any solution is a compromise between alarm count and applied delay. The algorithm result is a trade off plot with a pareto-line, showing best possible combinations of both targets. These results should be interpreted and applied, taking clinical requirements into consideration.

There are some algorithm specific inputs that could be optimised as well, but this system is not perfectly suitable for that, as it's only working against alarm count, which is not accurate enough for optimising them. The effect of varying said parameters would probably be too small and inaccurate to actually find optimal solutions. Algorithm developers are better suited at tuning those, as they can go down to signal and parameter value level and compare those with similar references. The framework is capable of analysing signal level differences, but for the scope of this study, it was left out of the optimisation process. Alarm related behaviour of signals is evaluated, however, and threshold violation statistics are reported.

3.7.1 Target functions

The optimisation framework uses two target functions, one for alarm count and one for optimising values related to the input configurations, i.e. alarm limit, delay, and averaging. These input parameters are denoted in the following functions as l, d, and a, respectively. Other variables used are parameter *defaults* and constraints such as *maximum delay*, and *limitRange*, the size of the threshold search space. The targets are therefore alarm count and applied total delay to alarm generation. The multi-objective optimisation problem is then scalarised to a single function as a weighted sum of the two target functions. The final optimisation problem is defined as follows:

$$\begin{aligned} \text{minimize} \quad & w_1 f_1(l,d,a) + w_2 f_2(l,d,a) \\ \text{subject to} \quad & l,d,a \in \mathbb{N}, \\ & f_2 \leq maxDelay, \\ & l \leq l_{max}, \\ & l \geq l_{min}, \\ & d \leq d_{max}, \\ & d \geq d_{min}, \\ & a \leq a_{max}, \\ & a \geq a_{min}, \\ & \text{No Missed Alarms} \end{aligned} \tag{1}$$

The problem is a constrained weighted sum that needs to be minimised. The primary target function is essentially the process of taking patient parameter signals from the algorithms and applying filtering and alarm evaluation to them. The output is alarm count, which should be minimised. The first function also evaluates inputs against constraints, mainly *search space* and system sensitivity profile. There is also a constraint for *maximum delay value*, which is calculated in different ways for different vital parameters. The formulas are shown below. The constraint for system sensitivity profile is evaluated against the number of Possible alarms. In practice, the sensitive profile treats Possible alarms as True alarms, and does not allow any to be missed.

Constraints on input parameters and their combinations are used as safeguards when the amount of annotated alarms is not large enough. If there were plenty of *True* and *Possible* alarms, they would not be needed to limit optimiser. The weights can also be used to control the movements of the optimiser.

The first target function is defined as follows:

$$f1(\text{limit}, \text{delay}, \text{averaging}) = n_{alarms} \tag{2}$$

The second target function calculates the total delay applied to alarm generation, which should also be minimised. This calculation is different for the SpO2 alarm, because the limit value can be embedded in the total delay (the SpO2 value change over time model). With other alarms, the total delay is first only calculated from delay and averaging values (to be checked against maximum delay constraints). The effect of the limit is added afterwards, to be used by the optimiser for choosing potentially narrower limits, if possible. The total delay variable is multiplied with a weight in order to reduce its size compared to alarm count. This is done to emphasise the effect of alarm count in the weighted sum to be optimised. In the function definition, the weights are the *maxDelay* and *limitRange* variables, these can be altered if needed. More on the weights used later in this section. The total delay function is defined as follows:

$$f2_{spo2}(l, d, a) = (4 * (l - l_{default}) + d + a/2)/\text{maxDelay} \tag{3}$$

$$\begin{aligned} f2_{others}(l, d, a) = {} & (d + a/2)/\text{maxDelay} + \\ & (l - l_{default})/\text{limitRange} \end{aligned} \tag{4}$$

These functions are given to the annealing algorithm. A separate wrapper was created on top of this core functionality, which serves as helper for applying search-space constraints and other heuristic controllers. The annealing function itself runs inside the wrapper, and results are output from it as well. See the pseudocode below for the annealing algorithm. The weights are applied inside the functions. Different weights were tested for the delay measure, to see how much they affected the end result of the optimisation. Increasing the weight on the second function increased the number of alarms but reduced the applied delay. There is no right answer as to which value is correct, the solution is always a trade off.

It seems that weights around 1/4 and 1/2 caused the optimiser to move towards minimising the delay value more, while smaller weights clearly caused alarm count to be minimised first. In the end, the weights were chosen so that alarm count is given more emphasis. With f1 the weight is 1, and the f2 weight is either the inverse of SpO2 maximum delay or inverse of total delay and inverse of the limit range. Different weights for the target functions could be used to create different profiles for the end system, and could make the *system sensitivity* profiles more nuanced.

The outputs of the target functions are *alarm count*, *total relative delay*. The two outputs are the main multi-objective optimisation outputs, which are used to create optimal solutions. Alarm count is simply the number of alarms produced with

the settings. Total relative delay was calculated relative to default settings. It's formula depended on the parameter. Specifically, with SpO2 the threshold effect is added to the total delay during calculation, based on the oxygenation drop model mentioned in section 2. For other parameters, only delay and averaging are used at first. To bring the effect of changing the limit also the the second relative delay value, it is simply added afterwards. This is because the delay value is first checked against maximum delay constraint, and only after is it evaluated by the algorithm. Constraint checks are done within the functions, and it violated, the function returns a very large value, which makes it impossible for the optimiser to choose it as a feasible solution.

To differentiate between sensitive and insensitive system profiles, an additional constraint is added that all also all *possible* alarms need to be detected. This could have been done with a third target function for possible alarms, but because of the absolute requirement of not wanting to miss any *possible* alarms, this was not necessary.

3.7.2 Methods

The optimisation problem had many characteristics that had to be taken into account when choosing a suitable optimisation method. The most restrictive were the non-differentiable target function, integer solution space, and non-linearity. Traditional optimisation methods such as gradient descent and simplex algorithm could not be applied in this situation. Heuristic methods like genetic algorithms and simulated annealing were considered, and the latter chosen because of simple implementation even for multi-objective problems.

The calculation for optimal solutions was not time-sensitive in this project. This means that brute force sweeping is also a viable option for solving the optimisation problem. If there were more parameters to be optimised, the situation would be radically different. A larger data set would also affect running time, but only linearly. Brute force sweep was also applied to the problem, yielding a perfect solution. This result could be used as a reference for the simulated annealing algorithm.

Simulated annealing is a probabilistic optimisation algorithm that applies heuristics for accelerating and improving the optimisation process. It is defined as a metaheuristic algorithm. The main principle is based on metallurgical processes, where metals are treated in order to alter their physical properties. Annealing is a heat-treatment process which eventually cool down metals. The algorithm can be configured with various stopping criteria and other settings, which can affect the end results. Being a probabilistic search-algorithm, increasing the number of iterations usually yields better solutions. Below is a pseudocode example of the simulated annealing process used in this study. The main inputs are the number of maximum iterations, which can be broken down into different sub-iterations, the first guessed

solution, which may affect the speed of convergence, and the cooling schedule, which may also affect the speed of convergence.

$T \leftarrow T_0$
$S \leftarrow S_0$
$oldcost \leftarrow loss(S)$
while $n_{iter} \geq n_{max}$ **do**
    $T \leftarrow temperature(coolingschedule)$
    $S_{new} \leftarrow neighbourhood(S, T)$
    $newcost \leftarrow loss(S_{new})$
    $newcost2 \leftarrow loss2(S_{new})$
    $delta \leftarrow costDiff(oldcost, newcost, oldcost2, newcost2)$
    if $delta > 0$ **then**
        $S \leftarrow S_{new}$
        $oldcost \leftarrow newcost$
        $oldcost2 \leftarrow newcost2$
    else
        if $P(delta, T) \geq rand(0, 1)$ **then**
            $S \leftarrow S_{new}$
            $oldcost \leftarrow newcost$
            $oldcost2 \leftarrow newcost2$
        end if
    end if
end while

The annealing heuristic is mainly used for enabling the algorithm to jump out of local minima. It is based on the idea of possibly accepting a worse solution in order to search in a different location of the solution space. The probability with which a worse solution is chosen is calculated as:

$$P = e^{-\Delta Cost/T} \tag{5}$$

The probability of accepting a worse solution decreases as the temperature cools down. Other heuristics have been tailored for the specific needs of this problem. First, step size is determined by temperature and size of the search space in each dimension. The size of the step in each dimension is proportional to the total dimension size and scaled by the current annealing temperature. The cooler it is, the shorter steps are taken. The outputs of the two target functions, alarm count and total delay, are handled differently, giving precedence to alarm count. This focuses the search to areas where the alarm count is small enough. When comparing the new cost to the old one, the total delay difference is weighted differently from alarm count, which limits its effect to cases where the alarm count does not change or changes very little.

3.7.3 Challenges

The main challenges with the optimisation were related to step sizes and overall computation time. The target function was computationally costly, taking about 1.5 – 2.0 seconds per iteration. This meant that the number of iterations had to be balanced with the optimality of the result. Any stopping criteria applied to the algorithm is a compromise. By favouring the alarm count result of the algorithm, it was possible to direct the search towards lower alarm counts more quickly, which saved time. Adaptive step-size also enabled the algorithm to find optimal regions quicker, which ensured a better result. The running time of the algorithm could be anything from two minutes to two hours, depending on selected search criteria. As the main benefit of this method was speed, it was often used as a quick tool for determining optimal configurations for a particular alarm.

Choosing the weights for the weighted sum factors was also challenging. The decision is largely based on clinical preference and trial-and-error testing of different weights. As mentioned above, using different weights allows us to guide the optimiser towards different solutions. In this case, it was chosen to prioritise alarm count, and give delay minimisation less emphasis. Other constraints were used for controlling the total delay value from growing too much.

Cross validation was applied in order to see how the optimiser results depended on the data. The leave-one-out (LOO) method was used, which runs the optimisation over the combinations of data where one item is left out. Results gave some estimate to the range of the optimal solution space given by the optimisation process. Section 4.2.1 shows the results of this cross-validation. Only the ranges of limits, delays and averaging lengths are shown, as they had the largest impact on the alarm counts. Principal component analysis and correlation analysis were also applied to the input variables and results, in order to determine which inputs correlate most with the end results.

4 Results

This section presents the principal results of this study and thesis. First, the alarm statistics obtained with default settings are presented. Other interesting statistics on the alarm behaviour of the system are shown as well. Finally, effects of configuring and optimising the alarm configurations are explained.

4.1 Alarm statistics

The test system provided a plethora of different results. The most useful of which were simple alarm counts and rates per patient per day. All results were further subcategorised by either alarm type, parameter, test case, or any required category.

Alarm statistics were first calculated with factory default settings for the system, and can be viewed as a baseline for the system's alarm behaviour. The default settings for clinical alarms are shown in table 6. Total alarm rates and their descriptive statistics are displayed in table 7.

Table 6: Factory default alarm configuration

Parameter (unit)	Limits	Delays (s)	Averaging (s)
SpO2 (%)	90 –	30	10
SpO2 critical (%)	85 –	10 (0*)	10
Pulse Rate (bpm)	50 – 150	30	10
Respiratory rate (bpm)	6 – 35	30	10
Apnoea (s)	20	0	0
Check Probe	N/A	25	0
Probe Off	N/A	30	0

* SpO2 critical alarm delay is zero if SpO2 alarm is active

Table 7: Alarm rate with defaults. Total alarm rate is for all patients combined. Mean, median and standard deviation is between patients.

Statistic	Value (n/day)
Alarm rate	47.6
Mean	41.3
Median	6.4
Standard deviation	90.9

Table 8: Rates of alarms by class

Alarm rate (n / day)	Total*	Relevant	Possible	Irrelevant	False
SpO2 low	13.4	0.2	3.5	0.8	6.5
SpO2 critical	10.6	0.2	1.8	0.2	4.7
Pr low	9.1	0.0	5.3	0.2	2.8
Pr high	0.4	0.0	0.0	0.0	0.4
Rr low	0.6	0.0	0.0	0.0	0.6
Rr high	2.8	0.0	2.4	0.0	0.4
Apnoea	0.0	0.0	0.0	0.0	0.0
Check probe	9.3	0.0	0.4	0.8	8.3
Probe off	1.4	0.0	0.4	0.0	1.0
Clinical	36.9	0.4	13.0	1.8	15.4
Technical	10.7	0.0	0.6	0.9	9.3
Total	47.6	0.4	13.6	3.7	24.7

* Alarms classified as unknown are added to total rates.

Figure 7 shows the distribution of alarms and classes. It is important to note, that the *possible* class is ambiguous in nature, meaning that the alarm might have been useful in some circumstances, depending on the system sensitivity profile mentioned in section 3.4.2.

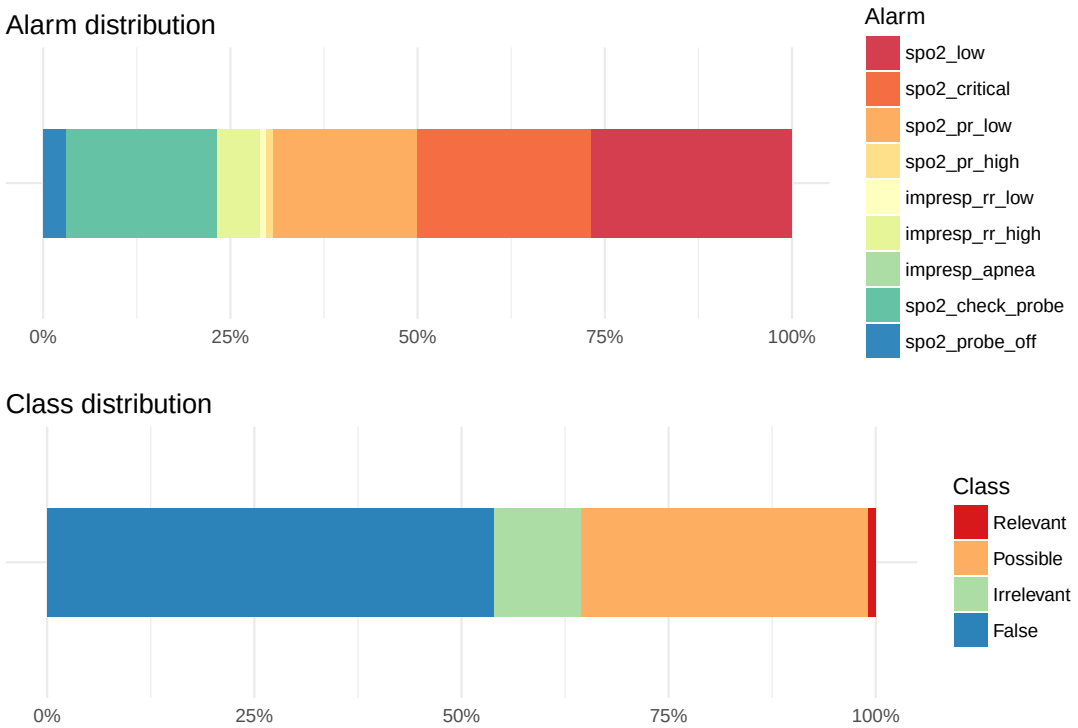

Figure 7: Alarms and classes distribution

With default settings, the system generated lots of clinically irrelevant alarms. This was expected, as clinical monitor systems tend to be configured to be very sensitive to alarm conditions. Alarm statistics results suggested most false alarms were caused by the pulse oximetry measurement. Impedance respiration was much less prone to cause false alarms. Technical false alarms were only generated by the pulse oximetry probe. Figure 8 summarises alarm behaviour of all test cases. It seems that some patients did not produce any alarms, while other produced a lot. The overall alarm burden is largely due to some patients producing numerous alarms. This was expected, as patients tend to either cause little to no alarm burden, or the opposite. The underlying distribution is fat-tailed, almost exponential, which suggests the bulk of the population is similar to each other, while few individuals are very different.

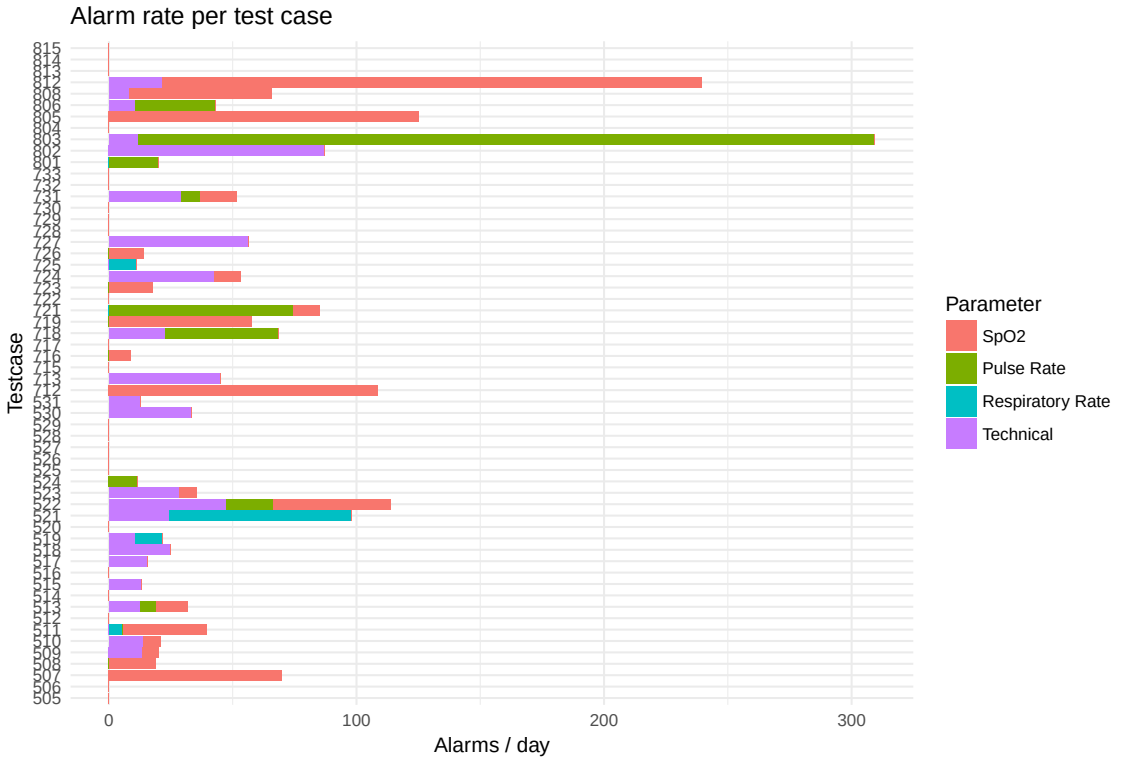

Figure 8: Alarm rates for each test case subcategorised by parameter

4.1.1 Alarm configurations

Different alarm configurations were tested with the hospital data. Settings were swept through their determined range to create plots which show the effect of different configurations. As expected, widening thresholds and lengthening delays and the averaging window reduced the number of nuisance alarms. The plots can easily be used to determine an acceptable configuration for a specific alarm.

Figure 9 shows the effect of limits and delays on the SpO2 Low alarm rate. The effect of changing alarm configurations is in line with previous studies covered in Section 2.3. Similar results were observerd with other parameter alarms as well. Pulse rate low alarms were drastically reduced when lowering the low limit from the default 50.

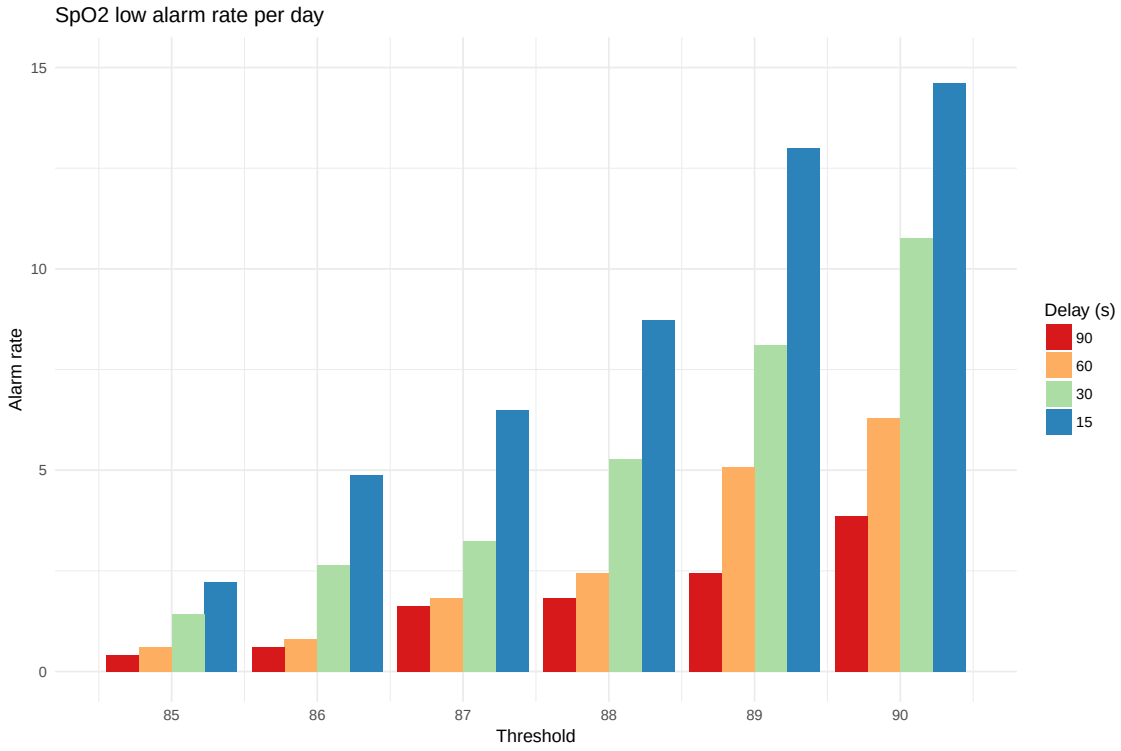

Figure 9: Spo2 low alarm rates depending on threshold and delay

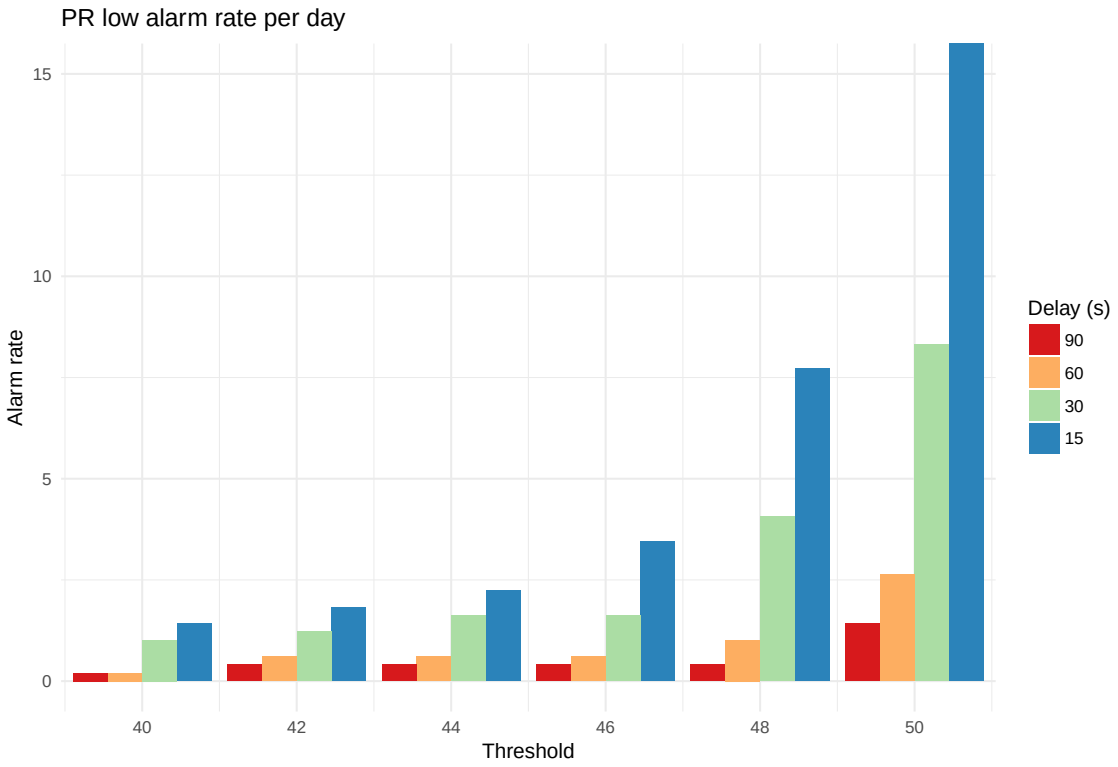

Figure 10: Pulse rate low alarm rates depending on threshold and delay

For clarity, the effect of averaging was separately plotted, as results suggested it's effect on alarm rates was smaller. Figure 11 shows the effect of the moving average on alarm rates. We can observe a large drop in alarm count when applying the filtering. Total alarm rate was reduced by 8 alarm per patient per day. Lengthening the averaging window further reduced the alarms, but reductions were not so significant anymore. The effect of filtering length is further studied in the next section.

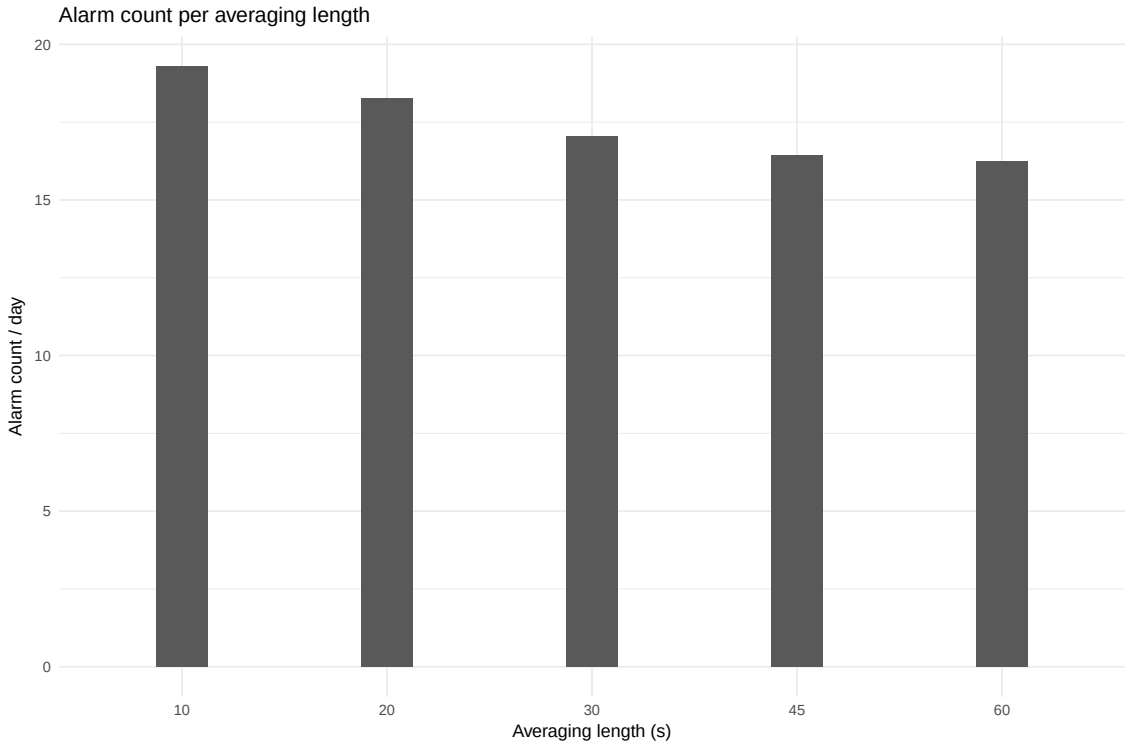

Figure 11: Total alarm rates depending on averaging filter length

We can also see how it affects parameter values as a whole, when plotting distributions them before and after filtering. Figure 12 shows the effect of applying averaging to the parameter values. These histograms also show the number of physiologically outlying values found in each parameter. Filtering reduces the number of these threshold violating values, which reduces total violation time and subsequently the number of alarms.

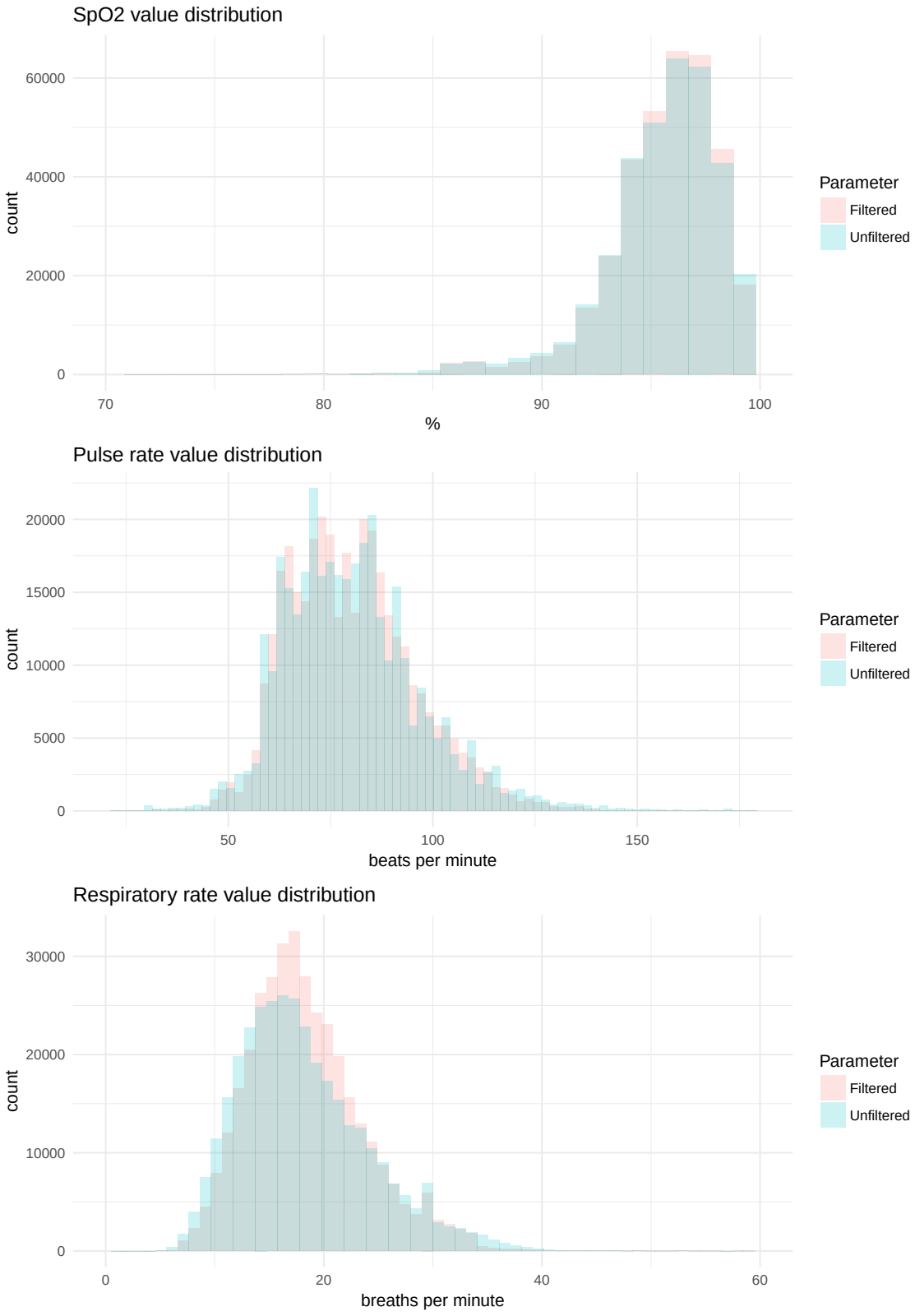

Figure 12: Parameter value distributions with or without filtering

The nature of threshold violation was analysed as well, and it seems that most violations are short, which suggests they are caused by artefacts or temporary anomalies in the physiological parameters. The amplitude of the violation periods was analysed as well. Violation amplitude tends to be small as well. A clinically relevant change in parameter value would typically be long lasting and quite significant in amplitude too. This information, coupled with alarm threshold violation lengths shown in figure 13, suggest that widening thresholds, applying delays and lengthening the averaging window might be effective in reducing clinically irrelevant threshold violations, while not affecting those which result from a relevant physiological change.

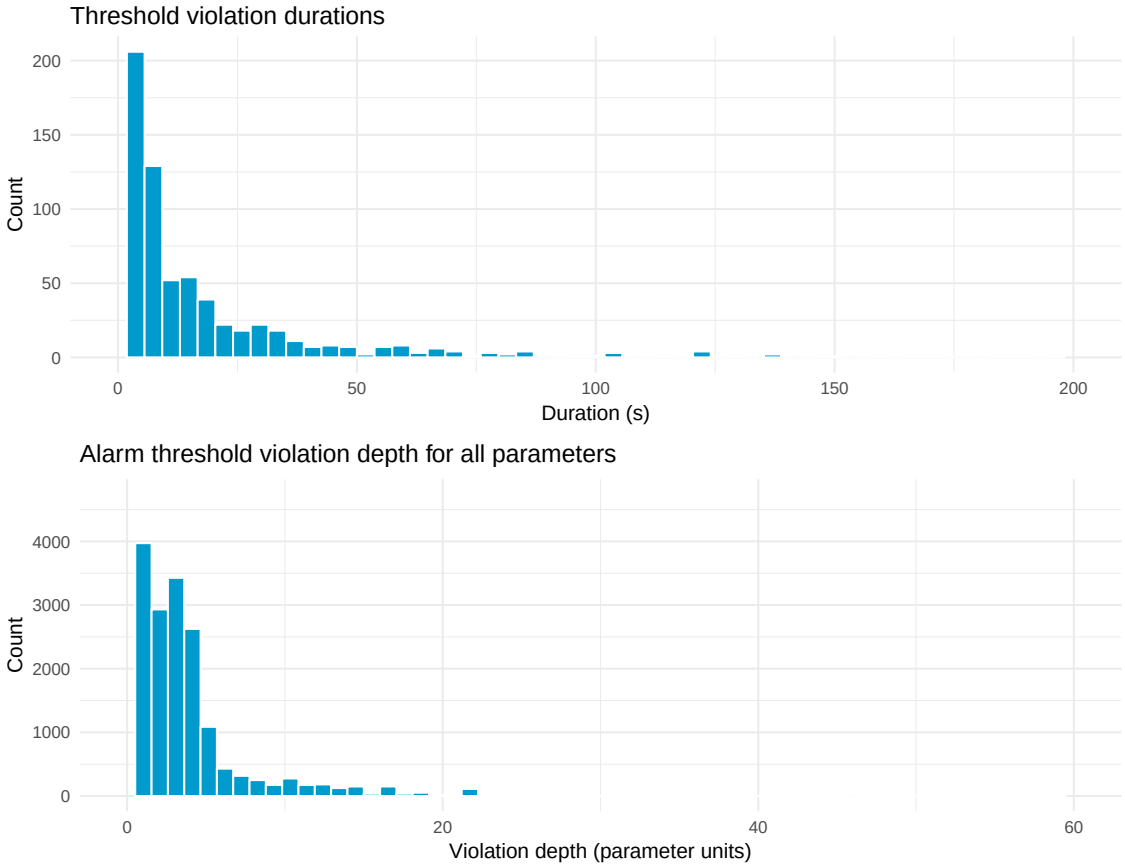

Figure 13: Alarm threshold violations. Parameter units have been normalised.

Alarm burstiness is a useful metric to describe system stability. It is the number of separate alarm events happening within a certain period of time. High alarm burstiness results from sensitive systems which easily trigger alarms often, while low burstiness means fewer alarms are present in a certain time period. In the clinical monitoring context, low burstiness is preferable, as the monitored physiological system is relatively stable, and possibly relevant alarm-triggering conditions don't change that quickly. The first picture 14 displays system alarm burstiness with default settings. We can observer as high as 10 separate alarms in a 5 minute interval, which is extremely high. The goal is to cut the higher values from the histogram.

The duration of clinically relevant alarms is typically relatively long. Short alarms contribute to burstiness, and are usually caused by artefacts and temporary anomalies. The second figure in 14 shows that most alarms are short in duration. Short alarms can be effectively managed with delays.

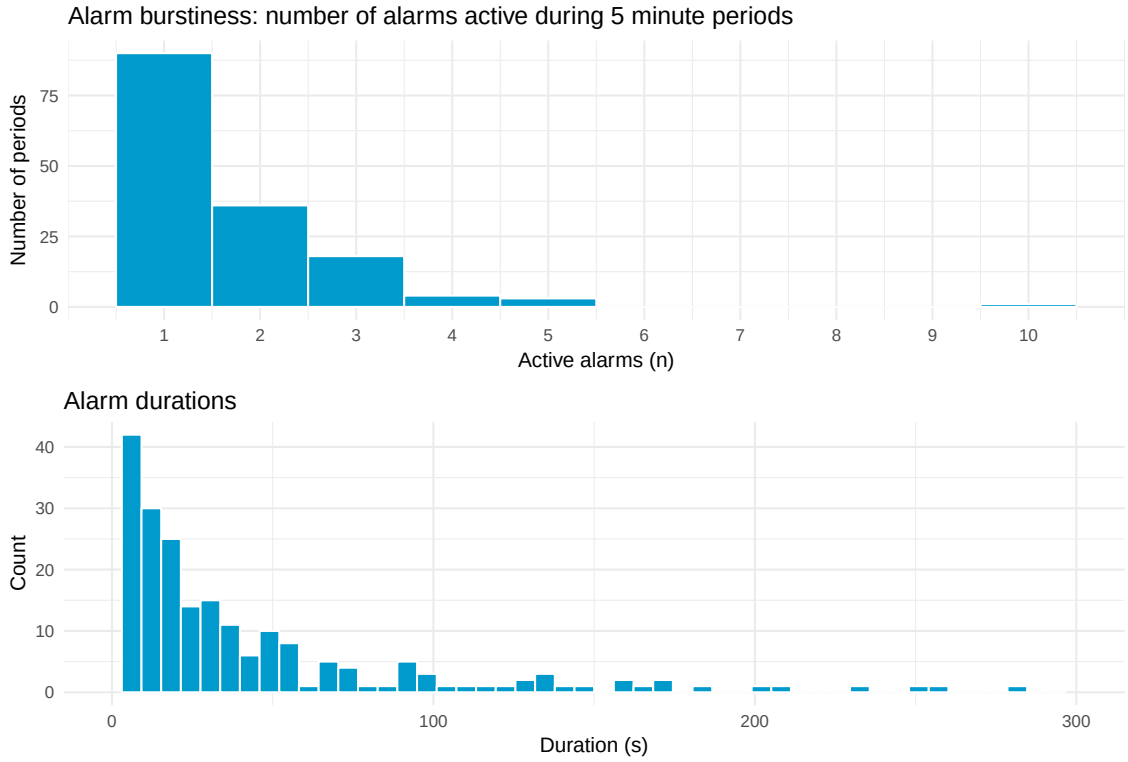

Figure 14: Alarm burstiness and durations

The next section presents results of the optimisation process, which sought to find optimal configurations for alarm generation, with settings that minimise the number of nuisance alarms while not sacrificing patient safety.

4.2 Optimisation

Optimisation was performed with both brute force and simulated annealing methods. The former obviously yielded perfect solutions, but was very time consuming. Simulated annealing optimisation produced sub-optimal solutions compared to brute force, but was able to provide solutions to the problem much more quickly than a complete search.

4.2.1 Optimisation results

Complete searches over the possible configurations yielded tradeoff plots. Figure 15 shows these results for SpO2 alarms. In the plot, the total relative delay is plotted against alarm count. Details of this metric are explained in the Methods section. The tradeoff plot can be used to choose best combinations of projected alarm count and applied delay to alarm generation. Similar plots were created for all optimised alarm configurations. We can clearly see the pareto-front in the plot below. As discussed in the optimisation methods section, changing the weights of the optimised sum would shift our choice of optimum value on this line.

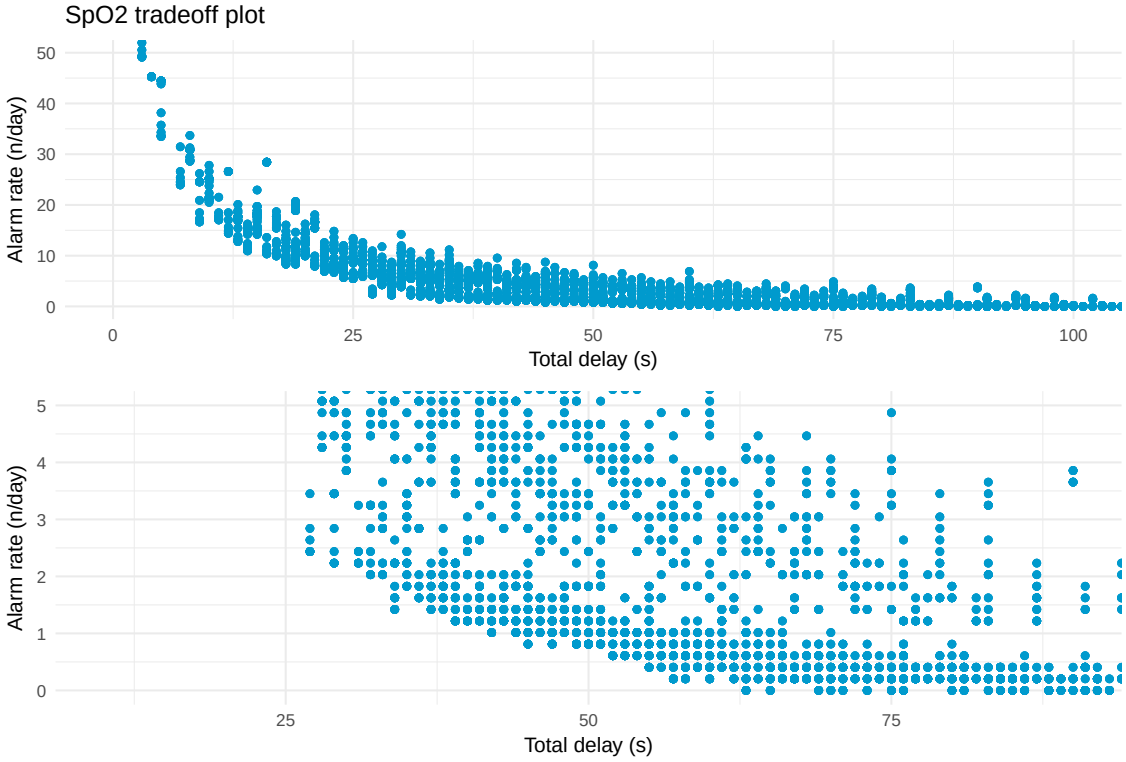

Figure 15: Tradeoff plot of spo2 alarm count and total relative delay

The results of simulated annealing are not as complete as those of brute force, but producing them was much less time-consuming. Simulated annealing produced satisfactory results in a fraction of the time taken by brute force. Brute force runs took about 4 – 5 hours, while the average time of convergence for the optimiser was around 10 minutes. Sensitivity to the data set was estimated by cross-validating the results. Graphic results for SpO2 optimiser cross validation are displayed in figure 16. Correlation analysis and Principal component analysis were applied to brute-force results, in order to study the significance of alarm configuration input factors.

Cross-validation results gave insight into the optimisers dependence on the data

set. Combined with statistical analysis of brute force results, information about the importance of configuration inputs was revealed. In general, the optimiser seemed to converge pretty well. The relationship between alarm delay and parameter averaging seems to be a linear tradeoff. For example, the SpO2 optimiser always converged on the same limits, but delay and averaging optimums had some variance. Limits where very consistent with all parameters. Correlation and principal component analysis suggested the strongest correlation between alarm count and alarm setting was with limit, followed by delay and averaging.

From the following plot, we can determine confidence intervals for the input parameters, specifically those that have higher variance and dependencies between each other. Confidence intervals for these parameters are shown in the final configuration table 10. The points in the plots below are optimal solutions given by the optimiser using different data sets. The plot clearly shows a linear dependence between the two input variables plotted.

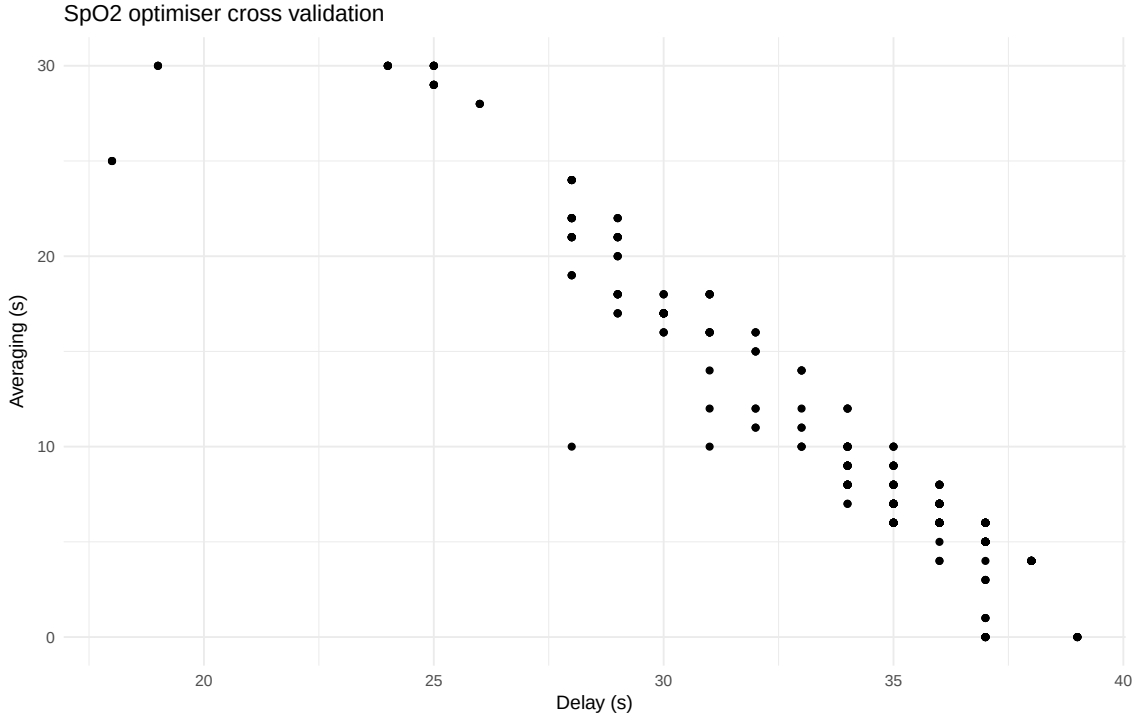

Figure 16: Cross validation results for SpO2 alarm optimiser. Individual dots are optimal solutions given by the optimiser.

A similar relationship between delay and averaging was also true with other parameters as well. Below the results for Pulse rate optimisation.

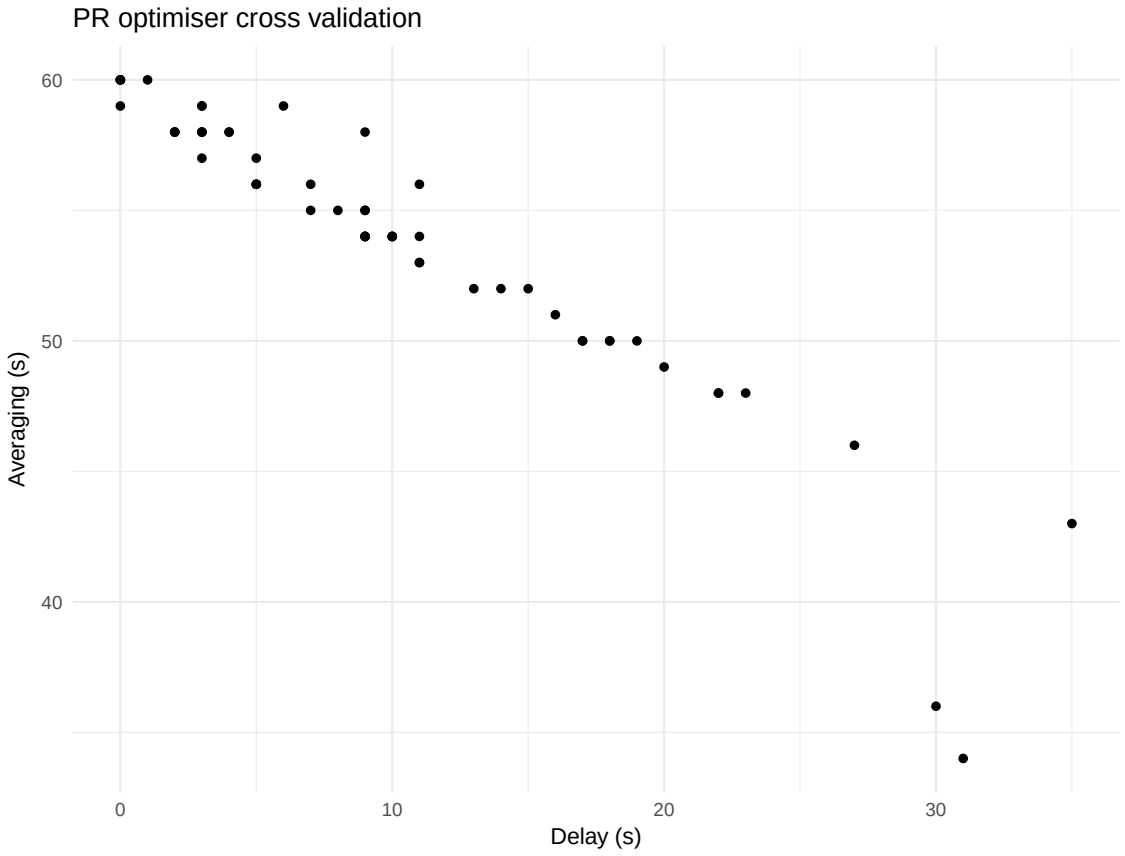

Figure 17: Cross validation results for Pulse Rate alarm optimiser. Individual dots are optimal solutions given by the optimiser.

With the optimised configuration parameters obtained from the optimisation process, we were able to reduce alarm rates considerably. Table 10 shows alarm rates subcategorised by parameters and alarm types. The total reduction in nuisance alarms was significant, and no clinically relevant alarms were missed. The distribution of patient alarm rates was deemed unreliable due to small sample size, so a bootstrapping technique was applied for estimating the underlying distribution of mean patient alarm rates. Figure 18 shows the result of this process. The underlying patient alarm rate distribution is fat-tailed, which is typical of populations where few individuals contribute a large portion of the total number of events. This property can still be seen in the distribution of *mean* alarm rates, which is slightly skewed.

This model estimates the mean alarm rate that end users may expect to see with the system. It is the mean calculated from a bootstrapping sample of patient alarm rates. The largest value of each bootstrap sample was removed, because we expect individuals with very high alarm rate will be handled differently by changing their alarm settings in order to reduce alarm burden. This graphic is useful for estimating the expected mean alarm rate. The 95 % confidence interval was also calculated to be at 2.1 – 15.0.

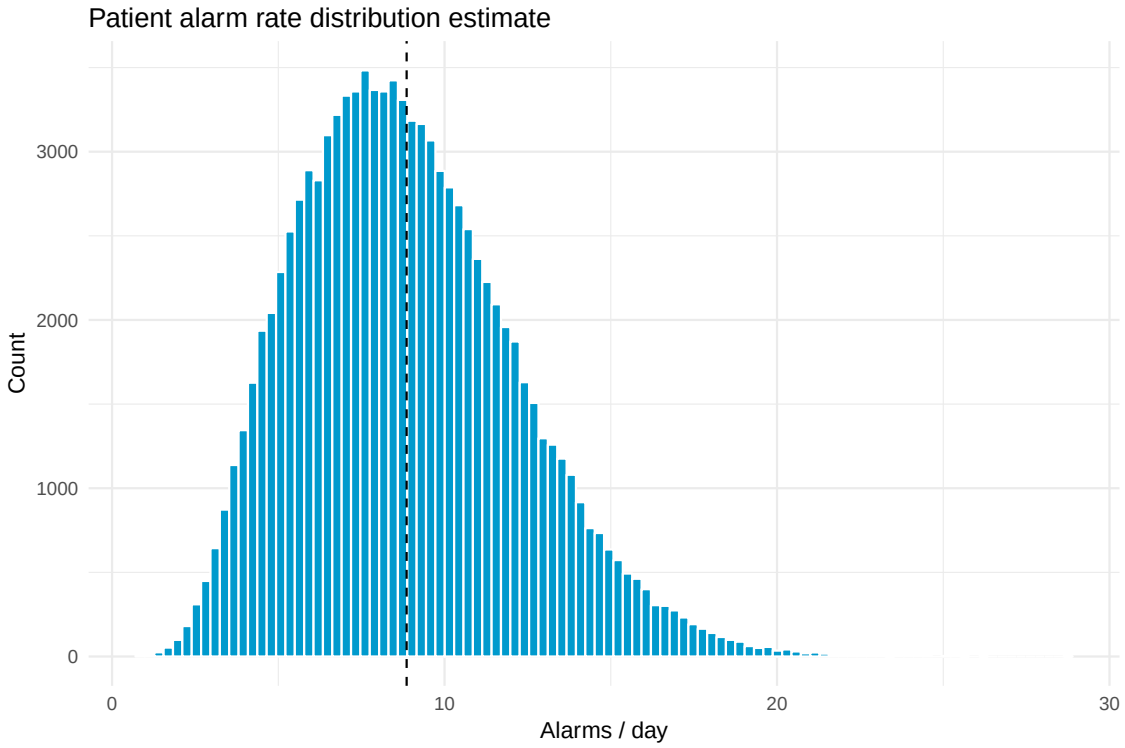

Figure 18: Distribution of mean alarm rates for system. 95 % confidence interval: 2.1 – 15.0. The dashed line is the mean.

4.2.2 Optimised parameters

There can be many interpretation as to what the optimal parameters for this system are. However, to facilitate the decision-making process, we developed two separate sensitivity profiles for the system, sensitive and insenstive. The former aims to make the system more sensitive to anomalies, therefore accepting more alarms to be sounded. Alarms that have been classified as "Possible" by clinical specialists are considered important.

The insensitive profile minimises all alarms except clinically relevant ones, which have been classified as True by nurses. In this study, only a single SpO2 low alarm was classified as True and clinically relevant, so the insensitive profile is not ideal for the optimiser. Futhermore, some alarms don't have relevant events, so we must rely on other constraints when minimising.

The other constraints depend very much on the clinical perspective taken on the matter. Alarm thresholds and delays must be kept within a reasonable range. Discussions with clinical specialists and examining literature suggested that delays and thresholds cannot be lowered too much, even if the alarm classification would allow the optimiser to do so. A total relative delay of 60 seconds was chosen as a

good constraint for relaxing the alarm rules. With this constraint, two sets of alarm configuration profiles were calculated, *insensitive* and *sensitive*. The former is shown under, and is the official recommendation of this thesis work. The sensitive profile includes *possible* alarms, which might be useful in some situations.

Table 9: Table of suggested configurations and their effect on alarm rate for *insensitive* system profile. 95 % confidence intervals in parentheses for optimised alarms.

Alarm	Limit	Delay (s)	Averaging (s)	Rate (n/day)	Reduction (%)
SpO2 low (%)	85	30 (20–32)	30 (28–40)	3.0	78
SpO2 critical (%)	80	10	30	2.6	75
Pr low (bpm)	44	45 (40–50)	30 (25–35)	1.6	82
Pr high (bpm)	150	10 (7–11)	30 (29–36)	0.4	0
Rr low (bpm)	6	30	30	0.6	0
Rr high (bpm)	37	45 (40–48)	5 (3–10)	0.0	100
Apnoea (s)	20	0	0	0.0	0
Check probe	N/A	25	0	1.0	89
Probe off	N/A	30	0	1.4	0

Table 10: Table of suggested configurations and their effect on alarm rate for *sensitive* system profile. 95 % confidence intervals in parentheses for optimised alarms.

Alarm	Limit	Delay (s)	Averaging (s)	Rate (n/day)	Reduction (%)
SpO2 low (%)	85	30 (20–32)	30 (28–40)	3.0	78
SpO2 critical (%)	80	10	30	2.6	75
Pr low (bpm)	49	30 (28–37)	45 (40–50)	5.9	35
Pr high (bpm)	150	10 (7–11)	30 (29–36)	0.4	0
Rr low (bpm)	6	30	30	0.6	0
Rr high (bpm)	36	20 (17–24)	22 (10–24)	2.8	0
Apnoea (s)	20	0	0	0.0	0
Check probe	N/A	25	0	1.0	89
Probe off	N/A	30	0	1.4	0

It is clear that by relaxing alarm thresholds, we can greatly reduce the generated load. Reduction percentages are also in line with previous studies. No clinically relevant event would be missed with these settings, based on data collected in this study. The number of *possible* alarms with the above system profiles is 9 and 34, respectively.

Figure 19: Alarm rates with sensitive settings subcategorised by class

When plotting results for the sensitive system settings, we can see a clear increase in both false and possible alarms, specifically in Pulse rate low and Respiratory rate high alarms, which had events of this kind. SpO2 also had a couple of them, which explains the increase of those too. Essentially keeping the Possible alarms in causes a in increase in collateral Irrelevant and False alarms as well, especially with Pulse rate and Respiratory rate. On the other hand, these increases were due to patients with specific conditions which caused the alarms. These could be treated with individual patient limits as well.

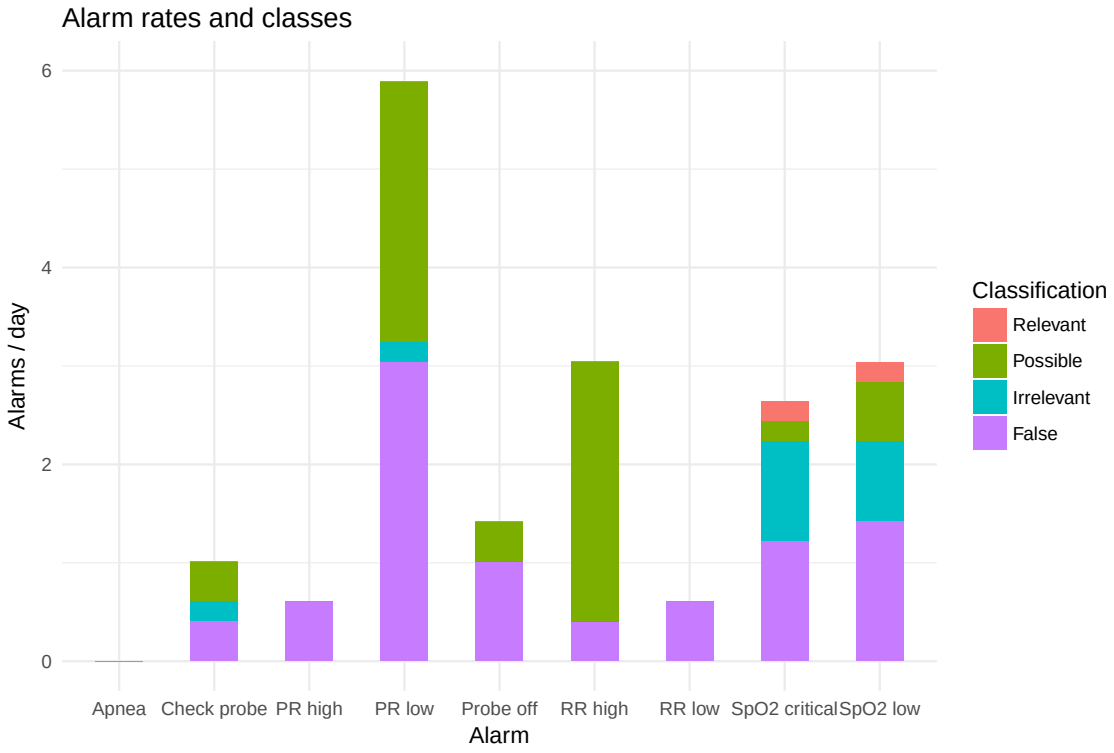

Figure 20: Alarm rates with insensitive settings subcategorised by class

The two sensitivity profiles both succeed in lowering the overall alarm load. The final question remains a clinical one, whether or not we want to include lower criticality alarm situations or not. Another question are the individual settings which may be applied ad hoc to specific patients. The next image shows the total alarm load reduction from the original default settings against the sensitive, sensitive individual, and insensitive settings. We can see a clear reduction even with sensitive settings, which can be further improved by using individual settings for patients with special circumstances. However, the insenstive settings got the lowest alarm rates.

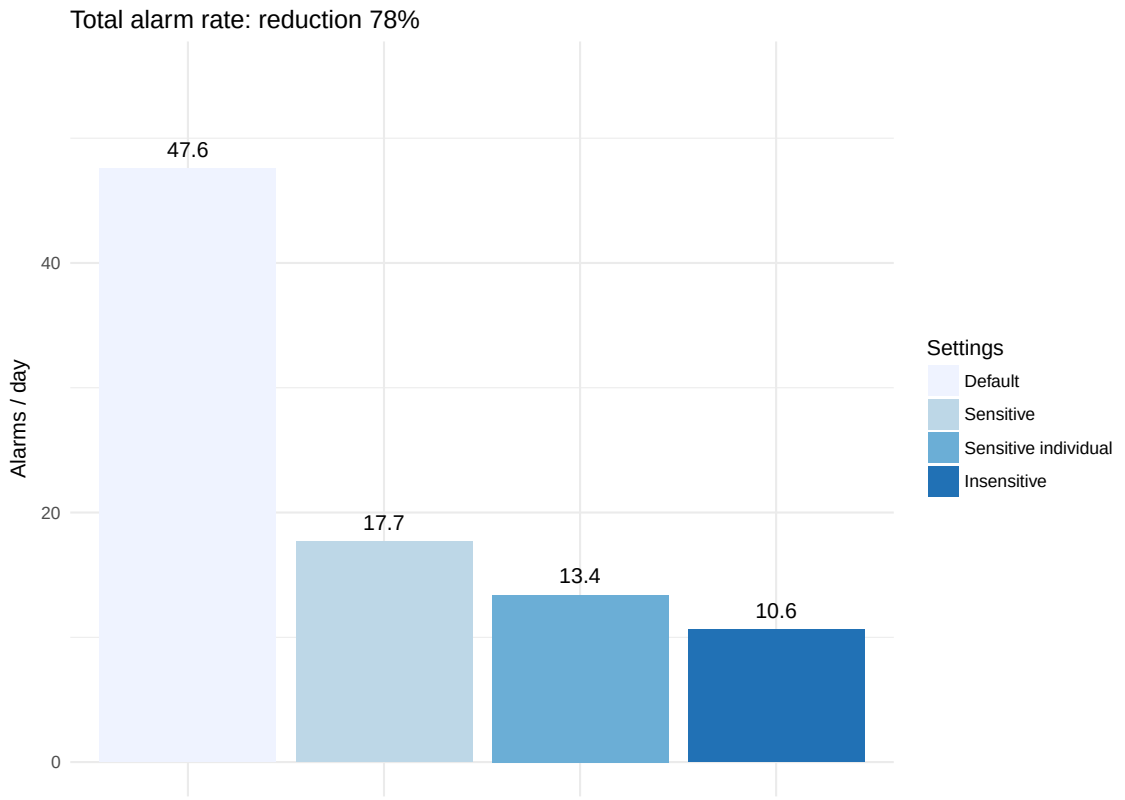

Figure 21: Comparison of total alarm rates with different settings: Factory default, Sensitive, Sensitive with individual limits, and Insensitive

5 Discussion

This section discusses the research project, problems, and results. The principal questions are related to the validity of the study protocol and produced results. The dilemma of system sensitivity is also a matter that merits further discussion. Lastly, the combination of clinical and technical perspectives is interesting as well.

The aim of this thesis was to investigate ways to reduce clinically irrelevant alarms of a patient monitoring system. Results of this thesis suggest patient monitoring alarm configurations can be optimised in order to reduce irrelevant alarms. Based on collected data, no relevant alarms were missed, and irrelevant alarms were reduced on average by about 80 %.

The alarm configurations proposed in this study were based on a model constructed from observations of a patient population. The main factors in this model were the annotation process applied on the original patient population, and the clinical knowledge used to further guide the research project. The practical work itself was straightforward, as was the optimisation process.

The study conducted during the making of this thesis was valid in principle, but had some limitations. The amount of data is not large, which leads to a small number of observed events in general. Especially the number of potentially adverse events and clinically relevant alarms is low. In fact, only a single event was recorded. This means all other parameters did not have a reference for a potentially dangerous event. The credibility of the optimised parameters suffers from this, which is why the *possible* alarm class was also used to constrain the optimisation process. With it, we were able to produce reasonable optimal configurations, which minimised alarms while keeping clinical safety interests in mind. However, the optimisation process would definitely benefit from more data. When estimating the expected alarm load, methods such as bootstrapping were used in order to address this issue, but the fact remains that 118 hours of patient data is not enough.

Annotation of patient data proved to challenging as well. Ambiguous guidelines and their varying interpretations led to high variance of annotations, which eventually forced us to re-evaluate all annotations. Fortunately this did solve the issue. For future research, it is important to define annotation protocols strictly and clearly. Of course, annotation is also a very subjective process which is affected by the individual, environment and patient; this makes the technique an inherent uncertain factor in the research method. Increasing the number of observers might help with this issue. More data would also help smoothing out intraobserver differences.

Another point of interest regarding this study was the question of system sensitivity. There is always a tradeoff between alarm load and sensitivity to measurement change, and determining the optimal position between them is tricky. The flexible way to define system sensitivity protocols proved an elegant solutions for bringing different

requirements and needs to system configurations. Also, the possibility to define individual settings for specific patients is strongly recommended.

From a technical and engineering point of view, it is clear that we can simply use collected data to optimise the behaviour of the system, and trust that the data contains the truth about different patients that might be monitored by the system. The clinical point of view is, understandably, quite different. Patients vary greatly, and nurses are not keen to risk their safety. *Better safe than sorry* is the prevalent approach to clinical alarms. Maybe it should be so, but research such as this does bring about an additional, practical point of view to the question.

It is important to determine clear definitions and requirements for continuous monitoring in low acuity environments such as surgical wards. Because this practice is relatively new and rare, it is not yet well defined. Traditional high acuity protocols, stemming from Intensive Care Units or Operating rooms, can be applied, but clearly result in too many alarms. Low acuity wards with less critical patients and lower nurse to patient ratios cannot sustain high acuity alarm rates, which is why monitoring alarm behaviour should be adjusted accordingly. This study methodology is a viable tool for supporting decisions on low-acuity continuous monitoring practices.

Further research is still needed to better shed light on these issues. Additional annotated patient data will reinforce the model for monitoring requirements in low acuity hospital environments.

6 Summary

This thesis is about a research study done during the clinical testing phase of a new patient monitor. It aims at optimising the alarm behaviour of the system, thus adding to overall patient safety. The optimisation is done by utilising an annotated reference of alarms recorded during the hospital test cases. Hospital data is rerun afterwards with different alarm configurations, and these are optimised using the reference. The goal is to minimise clinically irrelevant alarms while keeping relevant ones.

Principal results of the study are various information about the alarm rate and validity of the system. Optimisation results comprise of optimal configurations for alarms. For these, two separate profiles are provided, as some situations require more sensitive monitoring. Results suggest the number of clinically irrelevant alarms is high, and there is room for relaxing the alarm thresholds.

The study methodology proved viable for this kind of research, and data collection and analysis will continue towards improving the model further.

References

[1] A.D. Farmery et al. "A model to describe the rate of oxyhaemoglobin desaturation during apnoea". In: *British Journal of Anaesthesia* 76 (1996), pp. 284–291.

[2] Andreas H. Taenzer et al. "Impact of Pulse Oximetry Surveillance on Rescue Events and Intensive Unit Transfers". In: *Anesthesiology* 112 (2010), 282—287. DOI: 10.1097/ALN.0b013e3181ca7a9b.

[3] Andreas H. Taenzer et al. "Postoperative Monitoring – The Darthmouth Experience". In: *Apsf Newsletter* 27.1 (2012), 1—28. DOI: https://www.apsf.org/newsletters/html/2012/spring/01_postop.htm.

[4] Avinash Konkani et al. "Reducing Hospital Noise: A Review of Medical Device Alarm Management". In: *Biomedical Instrumentation and Technology* November/December (2012), pp. 478–487. DOI: 10.2345/0899-8205-46.6.478.

[5] Barbara J. Drew at al. "Insights into the Problem of Alarm Fatigue with Physiologic Monitor Devices: A Comprehensive Observational Study of Consecutive Intensive Care Unit Patients". In: *PLOS ONE* 9 (2014). DOI: 10.1371/journal.pone.0110274.

[6] Felix Schmid et al. "Reduction of clinically irrelevant alarms in patient monitoring by adaptive time delays". In: *Journal of Clinical Monitoring and Computing* 31 (2015), 213—219. DOI: 10.1007/s10877-015-9808-2.

[7] Imhoff et al. "Smart alarms from medical devices in the OR and ICU". In: *Best Practice and Research Clinical Anaesthesiology* 23 (2009), pp. 39–50. DOI: 10.1016/j.bpa.2008.07.008.

[8] James Welch et al. "Multi-parameter vital sign database to assist in alarm optimization for general care units". In: *Journal of Clinical monitoring and computing* 30 (6 2016), 895—900. DOI: http://dx.doi.org/10.1007/s10877-015-9790-8.

[9] Kelly Creighton Graham et al. "Monitor alarm fatigue: Standardizing use of physiological monitors and decreasing nuisance alarms". In: *American Journal of Critical care* January (2010), pp. 28–37. DOI: 10.4037/ajcc2010651.

[10] Mahmoud H. Alrefaei et al. "A simulated annealing technique for multi-objective simulation optimization". In: *Applied mathematics and computation* 215 (2009), 3029—3035. DOI: 10.1016/j.amc.2009.09.051.

[11] Matthias Görges et al. "Imroving Alarm Performance in the Medical Intensive Care Unit Using Delays and Clinical Context". In: *Technology, Computing, and Simulation* 108 (2009), 1546—1552. DOI: 10.1213/ane.0b013e31819bdfbb.

[12] Michelle A. Cretikos et al. "Respiratory rate: the neglected vital sign". In: *Medical Journal of Australia* 188.11 (2008), 657—659. URL: https://www.ncbi.nlm.nih.gov/pubmed/18513176.

[13] Robert M. Kacmarek et al. *Monitoring in respiratory care*. St. Louis, MO, USA: Mosby–Year Book, 1993.

[14] Schmid F et al. "Patient monitoring alarms in the ICU and in the operating room". In: *Critical care* 216.17 (2013). DOI: 10.1186/cc12525.

[15] Sylvia Siebig et al. "Collection of annotated data in a clinical validation study for alarm algorithms in intensive care–a methodologic framework". In: *Journal of Critical care* 25 (2010), pp. 128–135. DOI: 10.1016/j.jcrc.2008.09.001.

[16] C.J. Kalkman A.T. Rheineck-Leyssius. "Influence of pulse oximeter lower alarm limit on the incidence of hypoxaemia in the recovery room". In: *British Journal of Anaesthesia* 79 (1997), 460—464.

[17] Barbara Oakley Avinash Konkani. "Noise in hospital intensive care units–a critical review of a critical topic". In: *Journal of Critical Care* 27 (2012), 522.e1–522.e2. DOI: http://dx.doi.org/10.1016/j.jcrc.2011.09.003.

[18] Casey D. Blitt and Roberta L. Hines. *Monitoring in Anesthesia and Critical Care Medicine*. New York, New York: Churchill Livingstone, 1995.

[19] Frank E. Block. "Why we do not have – and will not have – the integrated and "smart" alarm systems that technology would allow us to have today". In: *Journal of Electrocardiology* 45 (2012), pp. 592–595. DOI: http://dx.doi.org/10.1016/j.electrocard.2012.08.002.

[20] Maria Cvach. "Monitor alarm fatigue An Integrative Review". In: *Biomedical Instrumentation and Technology* July/August (2012), pp. 268–275. DOI: 10.2345/0899-8205-46.4.268.

[21] Graham Kendall Edmund K. Burke. *Search Methodologies*. Springer, pp. 403 –451.

[22] Gerard J. Tortora Sandra Reynolds Grabowski. *Principles of Anatomy and Physiology*. Hoboken, New Jersey: Wiley Sons, Inc.

[23] British Thoracic Society Emergency Oxygen Guideline Development Group. *BTS Guide for oxygen use in adults in healthcare and emergency settings*. Vol. 72. 2017. URL: https://www.brit-thoracic.org.uk/document-library/clinical-information/oxygen/2017-emergency-oxygen-guideline/bts-guideline-for-oxygen-use-in-adults-in-healthcare-and-emergency-settings/.

[24] P.R. Barach J.M. Solet. "Managing alarm fatigue in cardiac care". In: *Progress in Pediatric Cardiology* 33 (2012), pp. 85–90. DOI: 10.1016/j.ppedcard.2011.12.014.

[25] James P. Keller. "Clinical alarm hazards: a 'top ten' health technology safety concern". In: *Journal of Electrocardiology* 45 (2012), pp. 588–591. DOI: http://dx.doi.org/10.1016/j.jelectrocard.2012.08.050.

[26] Michael Imhoff Silvia Kuhls. "Alarm Algorithms in Critical Care Monitoring". In: *Anesthesia Analgesia* 102 (2006), pp. 1525–1537. DOI: 10.1213/01.ane.0000204385.01983.61.

[27] J. Paul Curry Lawrence A. Lynn. "Patterns of unexpected in-hospital deaths: a root cause analysis". In: *Patient Safety in Surgery* 5 (2011). URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3045877/.

[28] Masimo. *Optimizing Masimo SET SpO2 Alarm Settings on Select GE Monitors*. Tech. rep. URL: http://www.masimo.com/siteassets/uk2/pdf/clinical-evidence/whitepaper/lab6145a.pdf.

[29] John T.B. Moyle. *Pulse Oximetry*. London, United Kingdom: BMJ Books, 1998.

[30] Mika Mäntykangas. "Improving respiratory rate esimation by combining information from impedance pneumography and electrocardiography". MA thesis. Aalto University, 2016.

[31] Susan F. Paparella. "Signals or Noise: Maximizing the use of clinical alarm systems for patient safety". In: *Journal of Emergency Nursing* 40.2 (2014), pp. 169–171. DOI: http://dx.doi.org/10.1016/j.jen.2013.11.003.

[32] Joanne Phillips. "Clinical Alarms: Complexity and Common sense". In: *Critical Care Nursing Clinics of North America* 18 (2006), 145—156. DOI: 10.1016/j.ccell.2006.01.002.

[33] R. Clark Robinson. *Introduction to Mathematical Optimization*, pp. 1–10. URL: http://www.math.northwestern.edu/~clark/publications/opti.pdf.

[34] Paolo Serafini. "Simulated Annealing for multi objective optimization problems". In: (). URL: https://users.dimi.uniud.it/~paolo.serafini/MCDMsimann.pdf.

[35] Katja Urpalainen. "Fraktionaalisen moniaallonpituuspulssioksimetrian algoritmikehitys". MA thesis. Aalto University, 2011.

[36] James Welch. "An Evidence-Based Approach to Reduce Nuisance Alarms and Alarm Fatigue". In: *Biomedical Instrumentation and Technology: Alarm Systems* 45.s1 (2011), pp. 46–52. DOI: https://doi.org/10.2345/0899-8205-45.s1.46.

[37] Wikipedia, the free encyclopedia. *Pareto front*. [Online; accessed December 29, 2017]. 2013. URL: https://upload.wikimedia.org/wikipedia/commons/b/b7/Front_pareto.svg.