Investigating Real-Time Monitoring of Fatigue Indicators of New Zealand Forestry Workers

Judy Bowen, Annika Hinze, Christopher Griffiths

Department of Computer Science, University of Waikato, New Zealand
{jbowen,hinze}@waikato.ac.nz, cjgg1@students.waikato.ac.nz

Abstract

The New Zealand forestry industry has one of the highest fatality and injury rates of any industrial sector in the country. Worker fatigue has been identified as one of the main contributing factors. Currently no independent and objective large data source is available that might support an analysis of this, or provide the basis for ongoing monitoring to further investigate. In order to successfully manage fatigue in the forestry workplace, we must identify suitable ways of detecting it. Industry partners are increasingly looking at monitoring solutions (particularly lightweight, wearable technology) that aim to measure worker activities and physiological metrics in order to determine if they are fatigued.

In this article we present the results of studies which investigate whether or not such technology can capture meaningful data in a reliable way that is both practical and usable within the forestry domain. Two series of studies were undertaken with in-situ forestry workers using reaction and decision-making times as a measure of potential impairment, while considering activity levels (via step count and heart rate) and job-roles. We present the results of these studies and further provide a comparison of results across different ambient temperatures (winter vs summer periods). Our studies show that while it is possible to see correlations between workloads (based on both physical and cognitive stresses) and fatigue measures using in-situ measurements, results are highly personalised to individual workers and can be misleading if the wider context is not also taken into consideration.

Keywords: forestry, fatigue, reaction-time, safety, workplace monitoring

1. Introduction

Forestry has one of the highest fatality and injury rates of any industrial sector in New Zealand.[1] Since 2008 there have been 32 fatalities, with claims to the NZ accident compensation scheme (ACC) in excess of two million NZ dollars each year. There were 12,921 active ACC claims between 2008 and 2013.[2] As a result of these poor safety statistics an independent review was conducted by Adams et al. (2014) for the Forestry Industry Contractors Association

[1] based on official data available up to 2014

[2] At the point of writing, there are no newer final ACC statistics available.

(FICAday). Data was gathered through interviews and self-reporting of all involved in the sector, from forestry owners and managers to the workers themselves, although the actual number of respondents was relatively small in comparison to the size of the industry. The initial report identified a number of factors contributing to the high accident rate (such as fatigue, lack of training, poor health and safety cultures) and included recommendations primarily aimed at increased participation in training and certification for workers and contractors. However, the report did not consider how the potential underlying causes of accidents might be identified or monitored, nor did it address potentially unsafe work practices or question why these continue to exist.

Aside from the FICA report there has been no large-scale data collection on fatigue in the forestry industry in NZ; information to date has been based on questionnaires and self-reporting of selected groups of workers. This method of data collection has many pitfalls, being subjective by nature and therefore susceptible to response bias, for example it is known that the structure of a questionnaire can have an impact on results as individuals tend to agree with questions (Morrel-Samuels, 2002).

The poor safety statistics and lack of large-scale data collection or analysis of accident causes beyond worker compliance provided the initial motivation for our work. In order to collect a meaningful amount of data from workers going about their everyday tasks, we needed to identify appropriate types of data that could be collected unobtrusively as well as suitable ways to collect such data from forestry workers. In conjunction with this, the industry health and safety organisations began to take an interest in such an approach, and began looking at technological solutions used in industries such as the military, mining, haulage driving etc. Before committing to large-scale purchases of either bespoke or off-the-shelf solutions they also wanted more information about the feasibility of monitoring workers in this way.

We initially considered the use of lightweight mechanisms, such as off-the-shelf activity trackers (e.g. Fitbit, Jawbone etc.), to gather data on levels of activity (via steps and hill-climbing measurements) and quality and quantity of sleep of forestry workers (Bowen et al., 2015). The goal of this work was two-fold: firstly we wanted to develop an actual data set (rather then self-reported data) from which to understand the working environment and identify worker fatigue (a known cause of accidents and contributor to risk). Secondly we aimed at identifying data that may be used by real-time technological interventions to identify potentially hazardous situations. We began with a series of experiments conducted within the research team and with a small group of forestry workers to validate equipment, methods and data types before moving on to the larger studies described in this article.

From these initial experiments we found that the steps and distances walked by workers did not seem to be significant factors leading to fatigue; relevant aspects seemed to lie in the stress caused by the necessity to pay close and ongoing attention to tasks performed in a potentially hazardous environment. Furthermore, although sleep data was relevant, there were issues of privacy and ethics that needed to be considered when collecting such data, particularly if we were to retain worker buy-in to our studies. We also found that the off-the-shelf solutions for sleep tracking were in some cases unreliable or impractical (levels of discomfort made it unlikely study participants would engage with longer studies involving their use). In these initial investigations we also encountered some resistance to this out-of-work tracking (forestry workers were concerned about privacy and how the data may be interpreted) and as such we decided that we would not include sleep data in our larger studies. We discuss this in more detail in Section 7.

Some observations from the initial investigations (Bowen et al., 2015), relating to cumulative fatigue were confirmed by analysis of accident statistics made available by one of the independent forestry management companies. The data, collected over 8.75 years, indicates an increasing incident rate throughout the working day, mitigated by breaks taken for lunch (see Figure 1). As well as the date and time of day, the accident data contains information about which activity and task was involved (we discuss roles and activities of workers later in Section 4) as well as a self-reported reason for the accident. We do not discuss the details of these reports further in this article other than to consider the time-of-day for accident occurrence in Section 7.

Contributions. In this article, we present the results of two studies designed to investigate methods for identifying and measuring contributors to fatigue in the workplace. Each study consisted of a series of visits undertaken with groups of in-situ forestry workers. We use reaction and decision-making times as a measure of potential impairment and compare this with activity levels (measured by step count and heart rate) and job-roles, to try and identify contributors to fatigue. We also introduce a comparison of results across different ambient temperatures (taken during Summer and Winter periods) to consider any amplification effect this may have on the same measures. We present the results of our studies and discuss the limiting factors of capturing and using such data in a meaningful way.

Structure of the article. The article is structured as follows: Section 2 provides background information on NZ forestry and the concepts of fatigue and reaction time. In Section 3 we discuss related work in the context of both forestry and workplace monitoring more generally. Section 4 describes the methodology used for our studies and the results are given in Section 5 (Study 1) and Section 6 (Study 2). We also provide a comparison across different working temperatures (seen in Winter and Summer). A discussion of the results is given in Section 7 with a final summary, conclusions and future work presented in Section 8

2. Background

This section briefly discusses aspects of managerial structures and work organisation in NZ forestry, and introduces concepts and measures of fatigue that are relevant to this article.

2.1. NZ Forestry

According to the New Zealand Treasury (2016), the forestry industry contributes about 1% of New Zealand's GDP and provides about 10% of New Zealand's total merchandise exports. There are roughly 1.8 million hectares of plantation forests split between state and private ownership. The industry is governed by four primary associations. The New Zealand Institute of Forestry is the main professional body and there are three separate

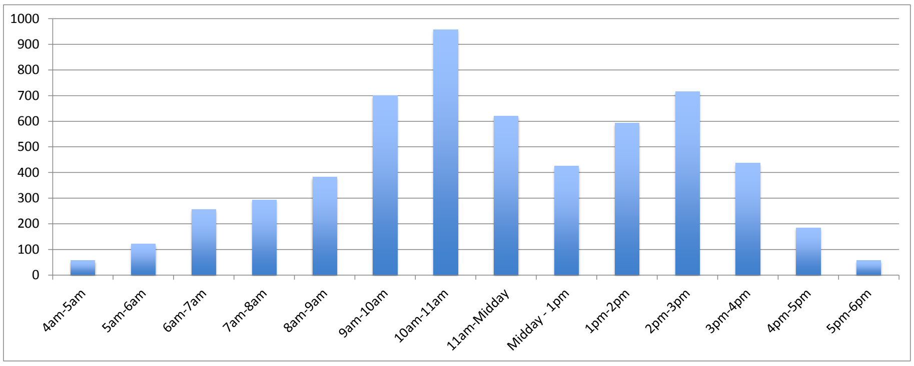

Figure 1: Reported Accidents in Forestry by Time of Day Over an 8 Year Period

associations beneath it: Forest Owners Association (FOA); Forest Industry Contractors Association (FICA); the New Zealand Farm Forest Association (NZFFA). Management and consultancy companies exist to source and manage contractors for operations such as harvesting, planting and forest maintenance. In addition these bodies monitor contractors, ensuring on-site operations comply with both operational standards and relevant legislation such as the Health and Safety at Work Act (The Parliament of New Zealand, 2015).

Work is distributed by way of tender for which smaller companies compete for contracts covering the operational aspects of harvesting, transport and siviculture. Each of these smaller companies generally have a workforce of less than 20 employees, with roles split between manual and mechanical operators.

An industry census conducted in 2012 by FITEC (the independent trade organisation now merged with Competenz) found the workforce is biased towards both younger and Māori[3] employees: 21.5% are aged 15 to 24 (5.6% higher than in the total New Zealand workforce) with 38.5% identifying as Māori (27.2% higher than the total New Zealand workforce) (Competenz, 2012). Low levels of educational achievement are prevalent within the industry with over 60% of workers having no formal post-school qualifications. However, current competency requirements are seeing more and more individuals completing industry-specific training.

The work is both physically and mentally demanding, with operations being performed irrespective of the weather. This combined with a poor pay rate of $17.50 per hour median (Payscale, 2014), which is only slightly higher than the New Zealand minimum wage of $15.25 per hour (Legislation, New Zealand) has contributed to a high rate of worker attrition, 45% of forestry workers change jobs within the first 12 months leading to a start rate of around 3,000 individuals annually (Ministry for Business and Innovation, 2014). The structure and workforce demographics are relevant to the wider consideration of safety as they provide the context of the workforce and working conditions. For example, one of the contractors involved in our early studies stated that he would like his workers to do fewer hours and to pay them more, but if he did "the contracting company down the road would get

[3]the indigenous population of New Zealand

all my business as his guys would do more for less and so they would win all the tenders for work".

2.2. Fatigue

Fatigue is a subjective physiological state experienced by individuals as a result of either physical or mental exertion (Hockey, 2013). *Mental fatigue* affects an individual's cognitive processes and *physical fatigue* affects an individual's ability to maintain physical actions. Both types are cumulative by nature, i.e., the level of fatigue generated by a task increases the more a task is performed. Both types of fatigue lessen after a period of rest.

Fatigue is quantifiable by indicators such as reduction of muscle strength (physical fatigue) and slower response times (mental fatigue). Studies into physical fatigue typically seek to replicate high levels of exertion by taxing muscle groups and measure how much force can be exerted over time. For example, Kumar et al. (2002) found that as the duration of muscle contractions increased the level of fatigue also increased.

Mental fatigue can impact the execution of complex thoughts and behaviour and reduce their efficiency (Alvarez and Emory, 2006). It can be induced by long periods of cognitive processing and the effects are particularly pertinent in high-risk work environments. Norman and Shallice (1986) identified five situations that are potentially impacted when an individual is suffering from mental fatigue: (1) planning and decision making, (2) error correction and troubleshooting, (3) unrehearsed or novel sequences of actions, (4) dangerous or technically difficult situations, and (5) situations that require the overcoming of a strong habitual response or resisting temptation.

The levels of fatigue experienced are determined by an individual's perception of these physiological changes. The energy required for a task (physical or cognitive) are personal to the individual concerned; they are also dependent upon factors such as age, physical fitness and mental ability.

2.3. Measuring fatigue

Physiological changes occurring as one enters a fatigued state can be used as indicators of reduced performance. For example, increased heart rate and core temperature may indicate physical activity, while a reduced reaction time may indicate mental fatigue.

In physically demanding roles, physiological monitoring may provide a better understanding of the impact of the roles and tasks upon an individual. Cumulative measurements taken throughout the workday can provide an understanding of both role- and task-based activities allowing for causes of fatigue to be identified and the associated risk to be managed.

We consider reaction time measurements as a potential indicator of fatigue within an individual over the course of a work period, where a reduction in reaction time indicates the effects of fatigue. The two 'gold-standard' tests for reaction time are simple reaction time and choice reaction time. We describe these next.

The *Simple Reaction Time (SRT)* test by Galton (1889) is a measure of the (averaged) length of time it takes an individual to respond to either a visual or auditory stimulus. The reaction is typically the pressing of a button. This test is appropriate for gaging general alertness and motor speed of an individual, with an uncertainty factor being introduced by the variable time period between the individual trials.

Brisswalter et al. (1997) examined the influence of physical exercise on SRT with participants exercising at 20%, 40%, 60% and 80% of their maximal aerobic power. During the exercise period they identified a decrease in cognitive performance measured by SRT, although on completion of the exercise period no significant difference in simple reaction time was apparent. This suggests that while we might identify workers as being impaired during the task, post-task measurement may not be able to identify this.

Choice reaction time (CRT) is similar to SRT, but measures the (averaged) response time of an individual by requiring them to make a choice when presented with a visual stimulus. For example, a visual representation of an arrow pointing either left or right is presented and the participant is required to select the corresponding button on a keypad. Having to choose the correct response for CRT requires more cognitive processing than SRT.

Tests are typically run as a number of trials, with uncertainty being added by the choice and a variable time between the stimuli.

3. Related Work

This section discusses related work on safety in forestry, as well as fatigue and worker monitoring in general.

3.1. NZ Forestry Safety

NZ forestry work-sites are generally remote, requiring the working day be lengthened to incorporate travel time (occasionally up to several hours). These long working days combined with a physically demanding workload can contribute to fatigue in workers. This may in turn adversely impact safe working practices and be a contributor to the high numbers of incidents within the industry (Spurgeon et al., 1997).

Lilley et al. (2002) undertook an investigation into the role that rest and recovery play in accidents and injury of workers. This study relied on self-reporting and involved 367 workers responding to a self-administered questionnaire. 78% of participants reported experiencing fatigue at work at least some of the time and the study concluded that the combination of slim margin for error and impairment due to fatigue constituted a significant risk factor within the industry.

In an attempt to gain more detailed data, Parker (2010) conducted a study using wearable video cameras to capture forestry worker behaviours. This work was limited by the small number of participants (due to equipment costs) and the time and expertise required to analyse the footage to understand what was being observed.

As a result of the forestry sector's poor safety statistics, FICA conducted the aforementioned independent review and interim results were published in (Adams et al., 2014).The focus of the review was on identifying and analysing the factors that impact on health and safety within the forestry industry and to produce guidelines designed to minimise the amount of incidents. The interim report identified a number of factors which may be contributing to the high accident rate, for example, worker fatigue, lack of training, poor health and safety cultures in the workplace. The report also included a number of recommendations which were primarily based around the creation of new processes, action groups and codes of practice to support and increase participation in training and certification for workers and contractors. However the report did not consider how the potential underlying causes might be identified or monitored.

3.2. Fatigue and Reaction Time

Levels of activity, both physical and mental, have been shown to increase the level of fatigue experienced by an individual during the course of their day (Pichot et al., 2002; Murata et al., 2005b). In a workplace context, especially where employees are working in remote locations such as those found within the forestry industry, detrimental effects from mental fatigue can invariably lead to situations where individuals can come to harm. From the situations identified by Norman and Shallice (1986) (described in Section 2.2) we consider Points 1, 3 and 4 as being particularly pertinent for forestry workers.

Increasing levels of mental fatigue have also been shown to adversely affect task motivation and mood. A study by van der Linden et al. (2003) using the Wisconsin Card Sorting Test, and the Tower of London Test designed to induce mental fatigue, found that the willingness to apply oneself to a task and do one's best was "significantly lower" for fatigued participants. As such, a higher level of mental fatigue can result in lower productivity levels. This conclusion is reinforced by Murata et al. (2005a), who observe that "fatigue is usually related to a loss of efficiency and disinclination to work."

Williamson et al. (2011) conducted a survey of research into different categories of fatigue and their effect on safety. They concluded that there was strong evidence to link task-related fatigue and performance to safety outcomes. They also noted that there was only limited data available and concluded that more research was needed to understand which roles and activities in the workplace may be more affected by fatigue. We seek to address this by considering different roles within the forestry environment to see if there are differences in the fatigue measures and effects we measure.

Sabzi (2012) investigated the effect of exercise-induced fatigue on choice reaction time. Using a mix of exercise types (aerobic, anaerobic, mixed, prolonged-intermittent and super-maximal-intermittent), they measured the choice reaction time of 15 participants both before and after exercise periods. An increased reaction time was identified across all exercise types with anaerobic, mixed and super maximal intermittent producing the largest differences. They concluded that "exercise-induced fatigue could reduce choice reaction time".

The related work discussed here, as well as many similar studies, (Saito (1999), Williamson et al. (2001), Lin et al. (2008)) all suggest a relationship between work-induced fatigue (either physical or mental) and reaction times. This relationship informs our understanding of the forestry accident rate and its causes.

3.3. Observing challenging workplaces

There are many challenges inherent in collecting observational data in workplace environments. The Human Work Interaction Workshop (2015) specifically focussed on design for challenging work environments. Discussions around data gathering in such environments identified common themes from a variety of different work domains studied, such as safe access to industrial sites, ethical considerations of monitoring employees (including use of, and access to, data) and finding suitable unobtrusive study methods.

Table 1: Days per crew per Visit for Studies 1 (winter) and 2 (summer)

Study 1	**Crew 1**	Jul-20	Jul-21	Jul-22	Jul-23		4 days
		Aug-24	Aug-25	Aug-25	Aug-27	Aug-28	5 days
							9 days
	Crew 2	Jul-27	Jul-28	Jul-29			**3 days**
	Crew 3	Aug-10	Aug-11	Aug-12			**3 days**
Study 2	**Crew 3**	Feb-22	Feb-23	Feb-24	Feb-25	Feb-26	**5 days**

As an example, although there is a reasonable amount of evidence to suggest that activities outside of work (particularly sleep quality and quantity) have an effect on workers, collecting such data raises many ethical issues (as it involves tracking workers during their personal time) and may also lead to resistance on the part of workers to take part in such studies. We discuss this further in Section 7.

4. Methodology

We conducted two studies with forestry workers in their working environment, one in the winter and one in the summer (for details see Table 1). We describe next the methodology of both the winter study, which we refer to as study 1 (see Section 5) and the summer study (see Section 6), which is study 2. The aim of both studies was to investigate:

(A) if we could identify a measurable correlation between levels of physical activity and cognitive response times, using lightweight and unobtrusive measurement equipment, and

(B) if there were measurable differences between roles with a high physical load (the manual workers) and those with a high cognitive load (machine and plant operators).

This would then enable us to further consider the most suitable measurements and corresponding equipment for an in-situ solution designed to detect fatigue in forestry workers.

4.1. Participants, Location and Timing

Participants were sourced from three separate forestry industry sub-contractors, all performing harvesting operations (for details on individual roles see Section 4.2). All participants were male. The studies were performed with forestry workers at their place of work in the forest during their normal working hours. Study 1 was conducted with three crews (see upper part of Table 1), one at each of one of three operational sites which were all located in the North Island of New Zealand. The participants ranged in ages from 17 to 62 years; participant demographics are presented in Table 2. Study 1 took place in July and August 2015 (during the NZ winter) and was conducted over 3, 4 and 5 day visits (as operational conditions allowed). Table 1 provides a summary of days for each visit per crew in each study. Monitoring commenced at the start of the working day, generally 06.45am, and concluded at the end of the working day, generally around 3.45pm.

Study 2 was conducted during the NZ summer months (see lower part of Table 1). It largely replicated the methodology of the winter study. This study was intended to re-investigate (A) and (B) above to see if results were consistent and reproducible, and also provide a comparison of reaction

Table 2: Participant demographic for Studies 1 (winter) and 2 (summer)

Crew	Participant	Role	Age Group	Winter	Summer
1	1A	Loader Driver	40 – 50	×	
1	1B	Loader Driver	50 – 60	×	
1	1C	Quality Control	< 20	×	
1	1D	Manual Feller	> 60	×	
1	1E	Process Operator	20 – 30	×	
2	2A	Loader Driver	40 – 50	×	
2	2B	Quality Control	20 – 30	×	
2	2C	Quality Control	20 – 30	×	
2	2D	Manual Feller	30 – 40	×	
2	2E	Process Operator	20 – 30	×	
3	3A	Quality Control	50 – 60	×	×
3	3B	Quality Control	20 – 30	×	×
3	3C	Manual Feller	30 – 40	×	×
3	3D	Loader Driver	30 – 40	×	×
3	3E	Loader Driver	40 – 50	×	×

times between different ambient temperatures. Anecdotally we had heard from the workers during the first study that they felt more impaired when working in the heat of the summer months than in the cold temperatures over winter. The participants of this second study were sourced from one of the crews participating in the winter study (crew 3), see Table 2 last column. The study took place in February 2016.

The number of days per visit, as well as access times to workers to conduct reaction time measurements were out of the control of the research team. One of the challenges of collecting data from hazardous workplaces is site access, and restrictions in place on different sites. In addition, the needs of the workers with respect to work performance targets and mandatory break times meant that we could not structure the studies around our preferred requirements, but rather to fit in with the crews. We discuss this further in Section 7.

4.2. Participant Roles

Participants performed a variety of tasks on site ranging from manual roles through to mechanised operations. Forestry roles are generally dictated by the operational conditions with small areas limiting the amount of plant and dictating the production methodology. We have included an overview of the roles monitored during this study and include images to further demonstrate harvesting methods.

Loader driver. Using mechanised plant these operators are responsible for skid site maintenance; stacking of logs for transport, laying out logs for quality control and loading processed logs for transport. Generally, loader driver commence site operations earlier than other crew members to facilitate early transport of processed logs and prepare the operational area for the day's production. Figure 2A provides an example of product loading for removal from the worksite.

A) Example of loader driver role

B) Example of quality control role - manual

C) Example of quality control role - mechanised

D) Example of typical terrain type

Figure 2: Examples of role and terrain types

Quality control. Quality control operations consist of log grading (sizing by diameter) and removal of any remaining branch stems prior to shipping. Quality controllers generally work in close proximity to mechanised plant requiring high levels of spatial awareness. In operations where log making is performed manually the quality control operations include length cutting. Figure 2B illustrates manual quality control; Figure 2C illustrates mechanised quality control methods. Quality control roles can be the most varied in terms of the physical activity levels. The amount of walking required is highly dependent on the site, and the role of quality control is often combined with other roles (such as safety observer) as required which also effects the physical nature of the role.

Manual felling. Manual felling generally occurs in remote locations where environmental conditions prevent the use of mechanised felling (steep slopes, inaccessible areas). These locations are generally remote from other site operations and as such the feller typically works alone. On completion of felling operations trees are transported to the skid site for processing using mechanised methods. Figure 2D illustrates the terrain type requiring manual felling operations be undertaken.

Process operator. The process operators are machinery operators who de-branch and trim the harvested trees. In areas where machinery cannot operate the trees are hauled to a central location for this processing.

4.3. Measurements

We captured three types of data: (1) physiological data as a measure for the level of *activity*, (2) reaction times as a measure for *fatigue*, and (3) ambient temperature at the workplace as a measure of *environmental factors.*

Activity: Physiological Data Measurement. The level of activity of the participants was measured using two types of physiological data: number of steps taken and heart rate. The measuring was executed using FitBit Charge HR activity trackers (Fitbit Inc, San Francisco, CA, USA). This is a wrist worn device utilising a triaxial accelerometer, vibration monitor, and altimeter to determine activity. Heart rate is collected at the wrist using Pure Pulse, a proprietary technology based on photoplethysmography. LED lights are used to illuminate the skin, and an electro-optical cell monitors the change in intensity of reflected light which, in turn, is interpreted as pulse.

These devices are capable of monitoring steps, heart rate, distance travelled, calories expended and sleep. Furthermore, the manufacturers allow third party access to stored data via an API, allowing developers to access stored data for incorporation into other applications.

Each participant was assigned an alphanumeric identifier (also shown in column 'participant' in Table 2) in order to protect their identities and an account was created for them on the `Fitbit.com` web application. Accounts were created using the same alpha-numeric identifier which provided the link between the reaction time component of the study and the physiological component.

A Fitbit Charge HR was given to participants at the start of each work day and they were instructed to wear it on the wrist of their non-dominant hand. Throughout the day, the Fitbit Charge HR collected physiological data from the participant. At the end of each working day, the Fitbit Charge HR devices were collected for synchronizing with the `Fitbit.com` web application and for re-charging.

Fatigue: Reaction Time Measurement. Reaction times were measured using the Deary-Liewald Reaction time test application[4] that was developed by the Centre for Cognitive Ageing and Cognitive Epidemiology based at the University of Edinburgh (Deary et al., 2011). For the SRT, the software shows a white square against a blue backdrop; the stimulus is the appearance of a diagonal cross in the square to which the participants respond with pressing any key quickly. For the CRT, four white squares are shown next to each other, each corresponding to one of four keys. The stimulus is the appearance of a cross in one of the squares, to which the participant has to respond by pressing the appropriate key.

The software records response times, the inter-stimulus interval for each trial, the keys that were pressed and if the response was correct.

We used an HP Laptop with Windows 10 for the testing using the Deary-Liewald Reaction time software. Simple and Choice Reaction Time measurements for each participant were undertaken three times a day: prior to the commencement of the employees' work period, on completion of their break, and at the end of their work day. Donders (1969), proposed a simple subtraction method for determining decision making times. Splitting the decision-making process into four stages, detection, discrimination, response and motor, Donders

[4]The software is freely available at `http://www.ccace.ed.ac.uk/research/software-resources/software`

used the difference in timings between tests where no choice is required (simple reaction time) and tests where a choice is required (choice reaction time) to infer the speed of mental processing, or decision making time. Using this method in conjunction with our testing data we may gain an insight into the speed of mental processing or, how long it takes an individual to make a decision. We include this calculation as part of our aggregated reaction time results.

Environment: Ambient temperature measurements. Temperature readings were undertaken using a McGregors digital thermometer at the same time as the reaction time measurements (described above). We aimed to capture the range of temperatures that the employees were exposed to throughout the course of a work period.

5. Study 1 (Winter): Results

This study was performed with three crews over varying periods of time (see Table 1). We first present the results for each of the crews, and then discuss the aggregated data.

5.1. Results for Crew 1

Crew 1 is primarily a mechanised crew, i.e, the majority of tasks were performed by plant (participants 1A, 1B, 1E) with the exception of manual felling (1D) and quality control (1C), see Table 2.

Monitoring took place over two separate sessions; 20th July to 23rd July 2015 and 24th to 28th August 2015. Both studies commenced at 06.45 and finished at 15.45 each day. Where participants were absent from work no data was recorded.

Two of the participants were loader drivers (1A and 1B) who started work on

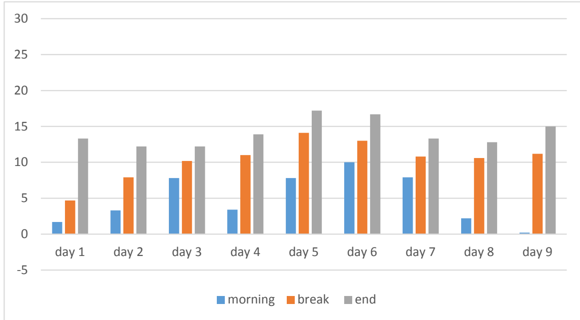

Figure 3: Temperature during study days (Crew 1)

site at 4.00am to load the trucks for early departure. The other participants from this crew commenced work at 6.45am. For practical reasons, the first data collection for all participants of this crew was performed at 6:45am. As such the data collected and reported as *start of the day* for these two participants does not represent the actual starting time, but rather are the start of daily data collection.

Travel time to get to their place of work for a 6:45am start was not considered for any participants.

Ambient Temperature. The temperatures across the study period varied between 0°C and 17.1°C with the weather being mainly fine with no rain during the monitoring periods. Figure 3 shows the temperatures in the morning, during break time and at the end of the shift, measured at the time the participants took the reaction time tests reported below.

Physiological Data. Figure 4 shows both the cumulative steps (in blue) and average steps per day (in red) for each participant across the study duration. Note that not all participants were present each day; for example, participant 1D shows a similar average step number to 1C although their cumulative steps for five days are lower than those

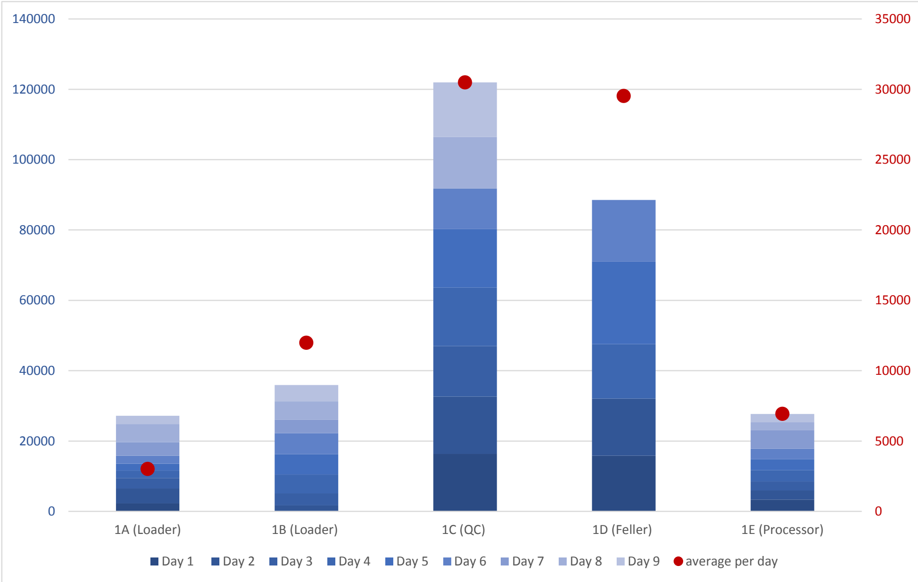

Figure 4: Step count by participant and role (Crew 1); step aggregation in blue (left axis) and average steps in red (right axis)

for 1C's eight days. Overall, we observed a large variation in the number of steps taken due to the respective tasks performed by the participants. For Crew 1, the member with the Quality Control role (crew member 1C) generated the largest number of steps across both study periods with a weekly mean of 77,072 (equivalent to approximately 58 kilometers). The crew member with a Tree Feller role (1D) also presents a high step rate, 52,020 (equivalent to approximately 40 kilometers). These roles are predominantly ground-based, with operational layout and terrain type dictating the required levels of manual work.

Figure 5 shows the mean heart rates recorded during the study. The intensity of activity can be modelled using heart rate (Robergs and Landwehr, 2002), with periods of high and low activity throughout the workday represented by higher and lower heart rates. A maximal heart rate of 220 beats per minute minus the participant's age was used to determine intensity of activity. These were colour-coded in the figure in the following way: a pale green background indicates areas of low intensity, pale yellow indicates medium and red is high intensity. We observe that the heart rates of participants 1A, 1C and 1E indicate low intensity throughout the day, while 1B had peaks of medium and high activity, and participant 1D showed predominantly medium and high activity. While both participants 1C and 1D (QC and feller) walked on average 3 to 4 times as many steps as their three colleagues, the work intensity of 1D, the feller, is much higher than that of 1C. This is an example of where context is important, but missing, from individual measures. Although both the quality controller and feller walk similar distances, the other activity levels within their roles have very different intensities. The QC role is mainly walking and visually checking, whereas the feller is operating a chainsaw,

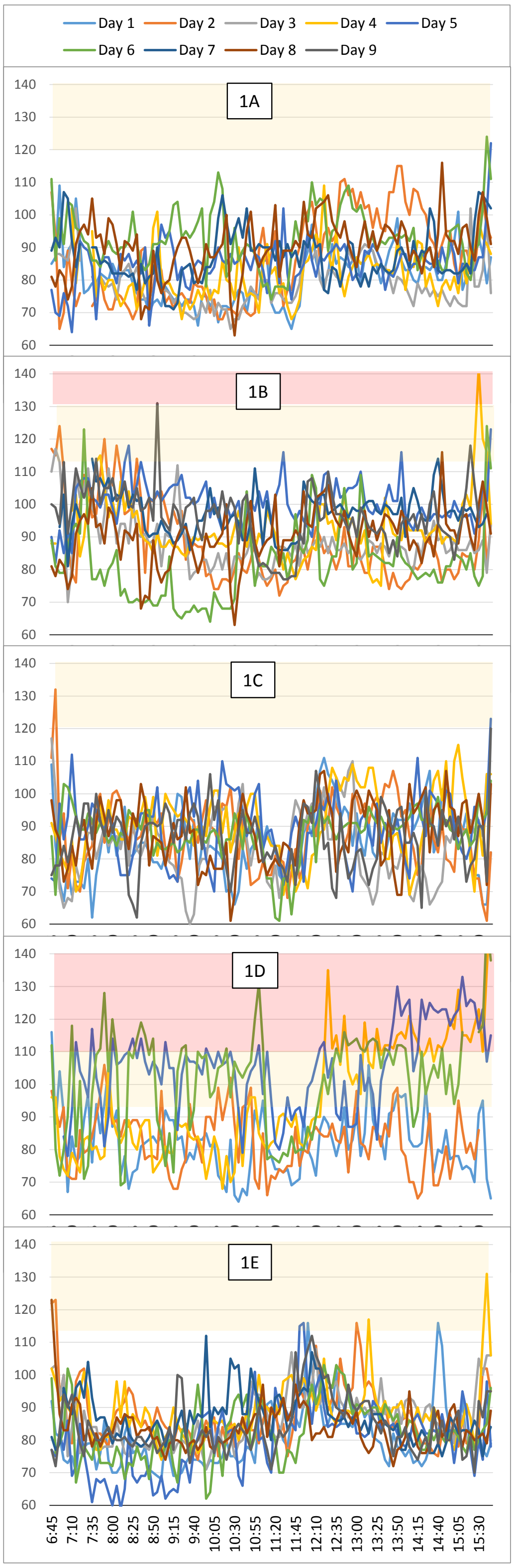

Figure 5: Heart rate (Crew 1)

cutting branches, moving logs etc. This more physically demanding role can be indicated by the heart rate profile but not the step count.

Participants 1A and 1B perform the same role, however, 1B has a higher average step count and his heart rate profile indicates higher levels of exertion. It may be that the location of his loading tasks requires more frequent movements in and out of his vehicle, or that he is in an area where higher levels of concentration are required (more people and plant around), but there is no way to determine this from the data alone. All of the participants show some increase in heart rate during the afternoon, although 1E (processor) tails off towards the end of day, perhaps suggesting a reduction in workload as the day comes to an end, or self-pacing towards the end of day.

It should also be noted that there are some gaps in the heart rate data. Participant 1B has no data for day 1, and also has some smaller gaps across other days which appear to be where the monitoring device was removed for short periods of time (although we have no confirmation of this). We are missing data for 4 days for 1D, and again there are gaps in some of the days where we do have data. In previous studies we had already noticed that there was a propensity for unexplained gaps in data as well as equipment being lost temporarily which affected continuity of data, we discuss this further in Section 7.

Reaction Time. Both Simple Reaction Time (SRT) and Choice Reaction Time (CRT) was measured at three periods throughout the working day (at the start of the shift, at the end of break time and end of the shift). SRT was measured at each instance 15 times; CRT was measured 20 times. For CRT, only those instances with the correct choice selection were considered. For both SRT and CRT, outliers were removed. We considered as outliers those data points below the first quartile minus 1.5 times the inner quartile range and those data points above the third quartile plus 1.5 times the inner quartile range. For the final CRT or SRT measurement, the mean across the remaining test instances was taken. Figures 6 and 7 show those values for SRT and CRT, respectively, for Crew 1. Note that the values for participant 1B for SRT on day 1 (break and evening) are omitted here as they are much higher (763 and 1048) and out of the range shown. Higher values represent longer response times, i.e. slower reactions.

Considering the SRT and CRT performance of each participant across the duration of the study we can better identify variation across work periods. All our performance measurements indicated variation across the workday, although some participants had flatter graph profiles indicating less variation than others.

Individual results for SRT show that there are no defined patterns across each working day for any of the workers. For example, although participant 1A has the slowest reaction times across all days compared with the other participants, on some days he gets faster over the course of the day (day 1), whilst on others he gets slower (day 5). Similarly, we cannot see any direct correlation between heart rate and reaction times. On day 5, participant 1D has a high heart rate consistently in the high intensity zone all afternoon, but his SRT at the end of the day is one of his fastest sets of results. Two of the participants who are the least physically active (by way of step count) are 1A and 1B and these participants also have the slowest SRT results in general, 1B also has the slowest CRT results. Their heart rate intensities reflect their low activity levels (although 1B does spike almost into intense activity at day 4), but they are performing tasks that may be more mentally tiring as it requires them to be alert and aware of their surroundings at all times. However, 1E who has a similar role to the loader 1B (being mostly a machine operator) has the second lowest step count does not exhibit similarly slow SRT and has one of the fastest and most consistent set of results for CRT.

5.2. Results for Crew 2

Monitoring took place over a three day period; 27th July to 29th July 2015 commencing at 06.45 and finishing at 15.45 each day. Crew 2 has a higher level of manual operation than Crew 1 (see Table 2). Participant 2E was absent on the 29th July (day 3).

Participant 2A was a loader driver starting at 4.00am, as such data collected and reported as start of the day does not represent the actual starting time for the same reasons as for Crew 1. Remaining participants commence their duties at 6.45am and therefore data collected encompasses the full working day. Participant 2E has a multi-functional role, mostly operating plant, but occasionally providing assistance to the feller.

Ambient Temperature. Temperature across the course of this study varied between 6°C and 18°C with the weather being mainly fine apart from occasional showers on the 28th July (see Figure 8). Ambient temperatures were warmer during monitoring than those encountered with crew 1, as such comparisons between the two crews at lower ambient temperatures could not be performed.

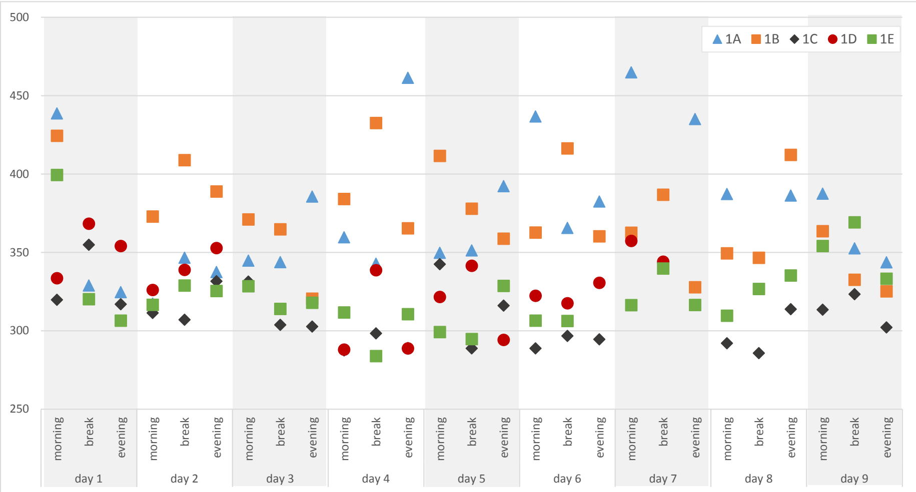

Figure 6: Simple Reaction Times (Crew 1)

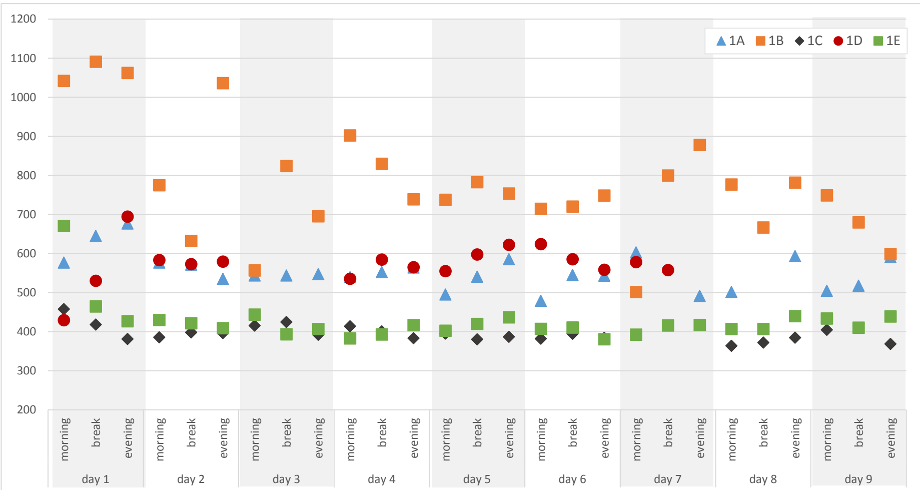

Figure 7: Choice Reaction Times (Crew 1)

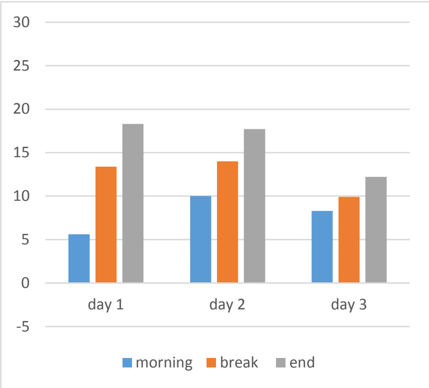

Figure 8: Temperature during study days (Crew 2)

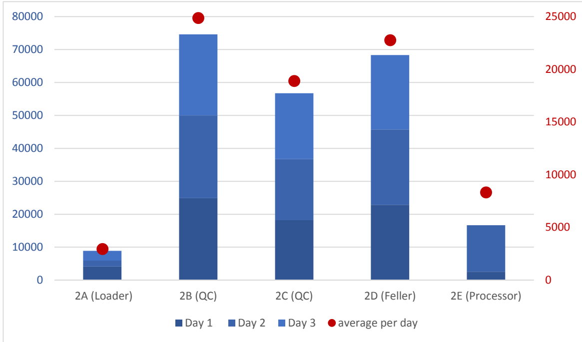

Figure 9: Step count by participant (Crew 2): aggregated (blue) and average (red)

Physiological Data. Figure 9 shows the cumulative recorded steps for participants in Crew 2 over the study period. Similar patterns to the roles of participants in Crew 1 (Figure 4) can be observed with higher step rates in roles such as felling and quality control.

Heart rate monitoring results are presented in Figure 10; as in Crew 1 activity intensity is indicated using colour bands. Heart rate intensity for all participants in Crew 2 was generally lower than those of Crew 1. In particular the participant with the most physically demanding role, the feller (2D), did not exceed a low level of heart rate intensity apart from a couple of occasions. It may be that he was somehow 'pacing' himself so as to not exert too much physical effort, or it could be that he has a high-level of physical fitness which can affect heart rate intensity. Once again the lack of contextual information means that we cannot determine which, if either, is the case.

Reaction Time. Figure 11 presents the results of the mean reaction time by period for participants in Crew 2 across the course of the work period. Similar to the results for Crew 1 there is no discernible pattern in the SRT results for any participant. In the CRT results, however, participant 2C (feller) shows improvement in speed during the day for each of the three days monitored. Also, the slowest participants were again those with the least physically demanding roles (2A and 2E, loader and processor). The feller (2D) has the second fastest set of SRT and CRT results, both quality controllers (2B and 2C) have fast SRT and CRT results, with 2C showing the fastest CRT response times (end of day 3) across all crews.

5.3. *Results for Crew 3*

Monitoring took place over a three day period, 10th August to 12th August 2015, commencing at 6.45am and finishing at 3.45pm each day, secondary visits were made on the 29th August and 4th September to expand on the data collected during the initial visit. Participant 3E was the loader driver whose duties commenced at 2.00am, data collection for this participant did not begin until 6.45am when the remaining participants commenced their work duties.

Ambient Temperature. Temperature across the course of the study with crew 3 varied

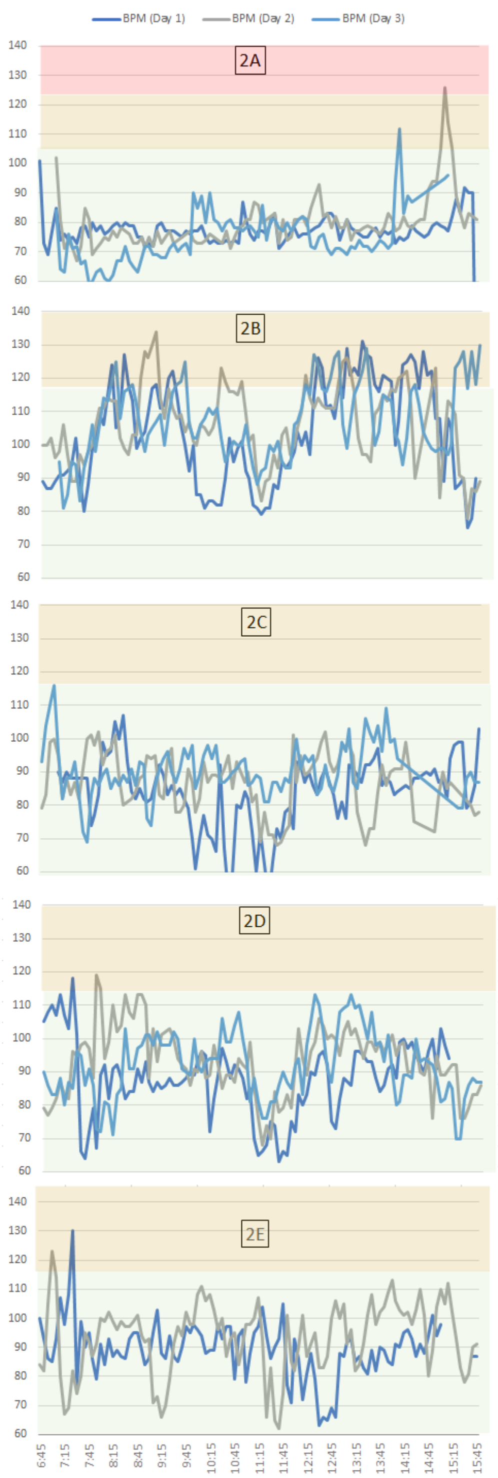

Figure 10: Crew 2 – Heart Rate

between -4°C and 11°C with the weather being fine throughout the monitoring period. Days 1, 2 and 3 showed the lowest morning temperatures seen in the study, 0°C, -4°C and -1°C respectively. There is no evidence that this had any effect on reaction time measurements for any of the participants.

Physiological data. Activity was measured in the same way as for Crews 1 and 2 via step counts and heart rate monitoring. This crew performs harvesting operations using a cable hauling system to deliver felled trees to an elevated processing platform; as such the majority of operations are performed using mechanised techniques.

Figure 13 show the levels of activity (measured as steps) for the differing roles.

The feller role has the highest step rates (some 109,000 equivalent to 83km) due to the remote locations of stock for harvesting.

Heart rate data is presented in Figure 14. As with the other crews activity intensity is represented by colour. This crew showed similar heart rate levels across roles to Crew 1. Participant 3E exhibits some large data spikes around the middle of day 1, and also has missing data for the afternoons of days 1–3 which suggests he may have removed the monitor for some reason. This is more likely than an intermittent failure of the monitoring device at the same time on 3 occasions, although this cannot be entirely ruled out. The feller, 3C, has greater variance between days than the fellers from the first two crews who exhibit a more regular pattern across days. However, unlike the feller from Crew 1, he barely hits the maximum exertion levels on most days.

Reaction Time. There are no discernible patterns in the simple reaction times for any of the crew members (Figure 15). 3A and

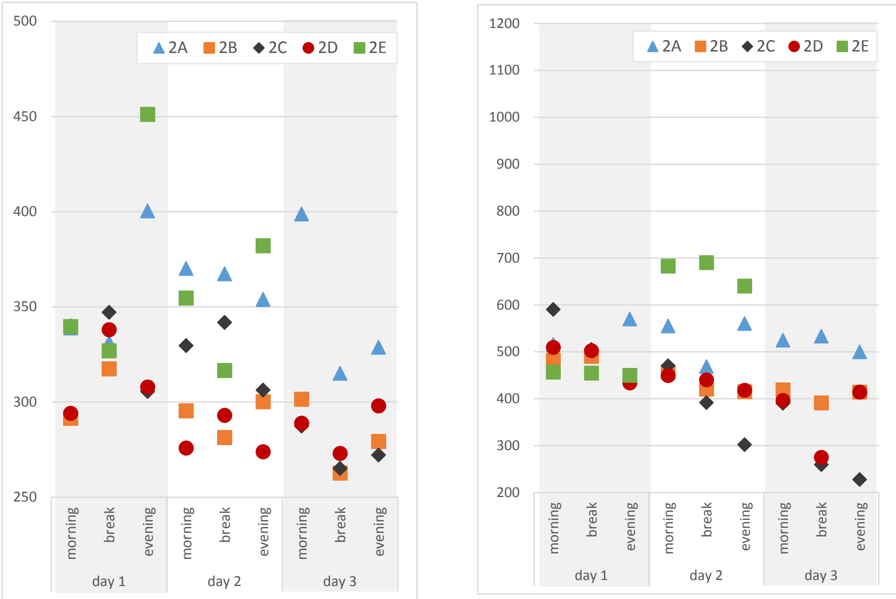

Figure 11: Reaction Times: SRT – left and CRT – right, (Crew 2)

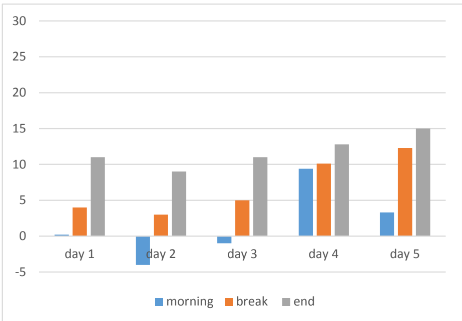

Figure 12: Temperature during study days (Crew 3)

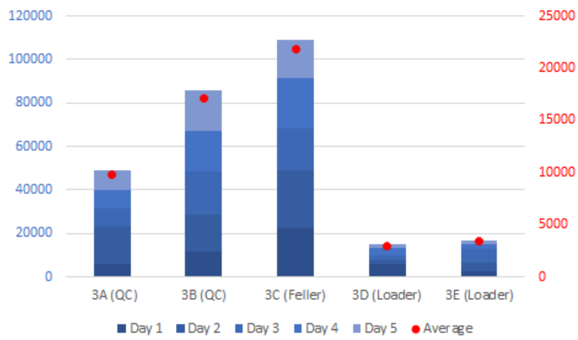

Figure 13: Step count by participant (Crew 3)

3E were always slowest first thing in the morning but sometimes got slower again at the end of the day and sometimes not, they also had slower simple reaction times generally than the other participants. 3D had the fastest, and most consistent simple reaction times, but again there was no fixed pattern. The CRT graph, Figure 16 is flatter, showing more consistent scores for participants across each day. The biggest change is seen in 3A who starts slow but improves gradually as the week progresses. Of interest is the fact that both 3A and 3E were at their slowest on the morning of day 2, which was the coldest morning seen across all studies. However, there is no direct evidence that these are related as we do not see the same effect with the other participants.

5.4. Aggregated Analysis across Study 1 (Winter)

Our analysis of results is comparative by nature. For example, do high levels of activity result in a decrease in performance? (in-

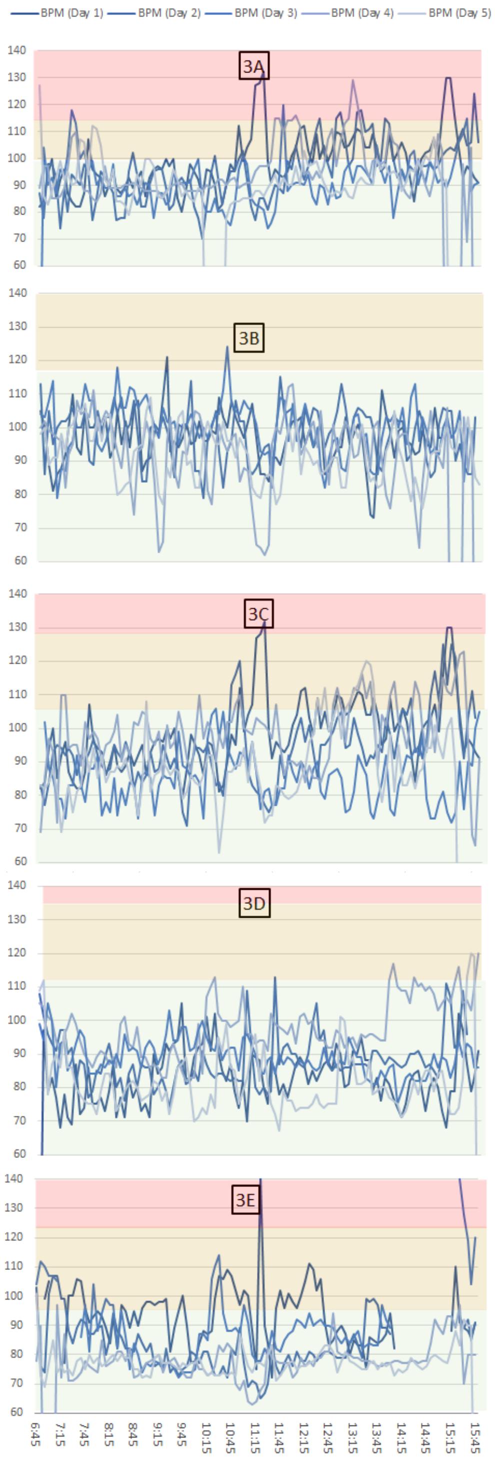

Figure 14: Crew 3 – Heart Rate

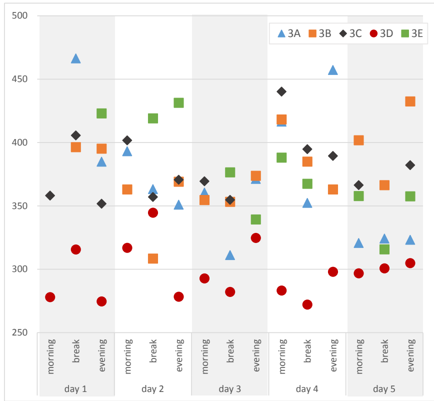

Figure 15: Simple Reaction Times (Crew 3)

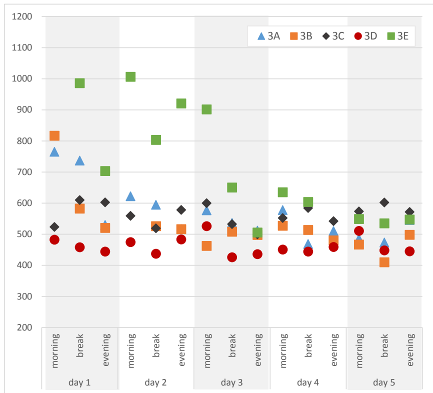

Figure 16: Choice Reaction Times (Crew 3)

dicated by increasing reaction times). Or, does temperature impact an individual's ability to perform? We further examine the differences in performance by role.

Steps counted. When considering activity, our investigation shows large differences for physical activity levels (measured as step counts), see Figure 17. Mechanised operations generate the lowest step rates (i.e., average step counts per day for all loaders is under the average step count of 14,459 across the 15 participants) as operators perform tasks from inside plant in seated positions preventing high levels of activity. The average per day of all fellers is above the average step count across all three crews.

Participant	Average steps	Role
1C	30493.0	quality control
1D	29517.7	feller
2B	24874.7	quality control
2D	22777.0	feller
3C	21821.6	feller
2C	18906.7	quality control
3B	17118.6	quality control
1B	11977.3	loader
3A	11766.8	quality control
2E	8333.0	processor
1E	6923.3	processor
3E	3395.6	loader
1A	3019.1	loader
3D	3002.6	loader
2A	2964.7	loader

Figure 17: Comparison mean steps per day

Reaction times. Figures 18 and 19 show the mean reaction times of each participant.

Note that the aggregation is across 9 days for crew 1, across 3 days for crew 2 and across 5 days for crew 3. The colour schemes indicate crews (blue – crew 1, orange – crew 2 and green – crew 3). The marker shapes indicate roles (square – loader, circle – quality control, triangle – feller, and diamond – process operator)

For selected participants, overall patterns can be detected (e.g., participant 1B shows a slowing down of both simple reaction time and choice reaction time in the evening while participant 3E drastically improves over the course of the day). 1C and 2C have similar trends for both SRT and CRT, but for 3B and 3D they are reversed. As the participants are potentially affected by a number of factors (such as physical exhaustion, cognitive load or even temperature), it is difficult to observe specific patterns by role or across crews.

Figure 20 shows the decision making times calculated as difference in reaction time. While this averaged data is interesting in identifying if there are trends that can be more obviously detected, it does not nec-

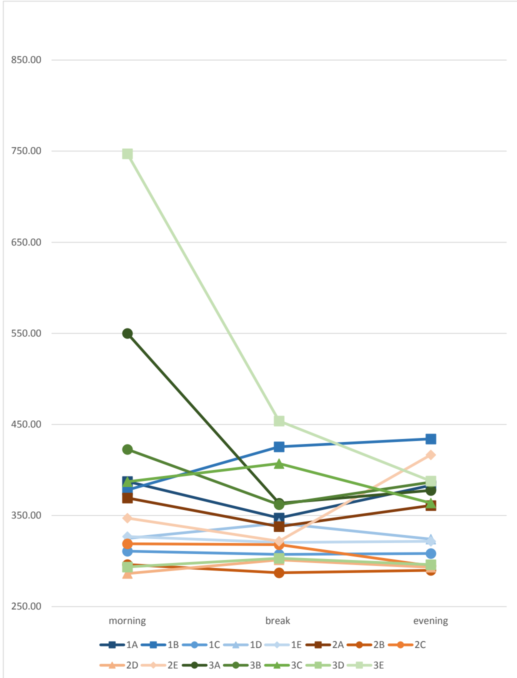

Figure 18: Average SRT per person

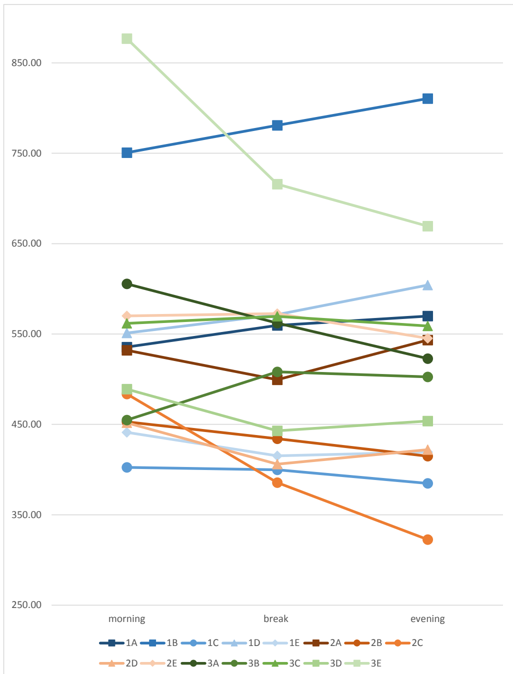

Figure 19: Average CRT per person

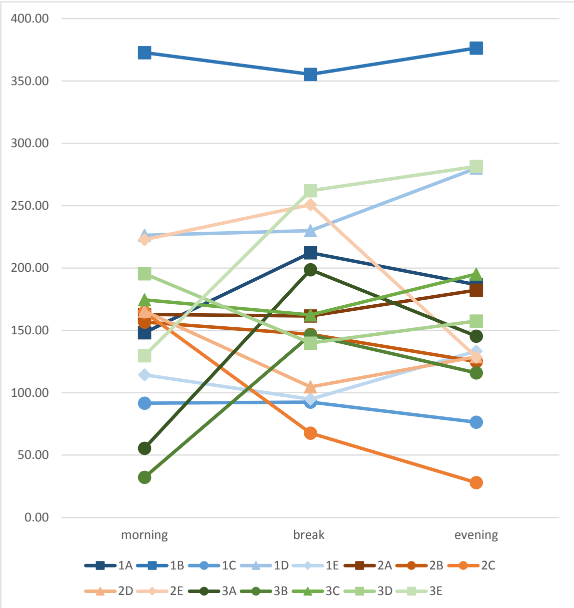

Figure 20: Decision making time per person

essarily help if the goal is ascribing meaning to data during real-time collection. We discuss this further in Section 7.

Table 3 shows the difference between decision making time in the evening and in the morning. We can see which participants have (on average) faster decision making times at the end of the day (indicated in green in column 1), as opposed to those who get slower (yellow, red). This is not consistent across roles, but rather appears as a personalised measurement. However, Table 4 which orders the decision-making time just on speed in the evening, does show some correlation between role types and decision-making time. We can see that the role-types are clustered based on this ordering, apart from the fellers. Given that actual speed of mental processing is an individual measure (irrespective of alterations over the course of a day due to fatigue etc.) it is interesting to note this clustering. It could be an artefact of the types of roles people choose to do based on their perceived abilities. There has been some interest from the forestry industry in whether or not specific traits related to both decision making and visual acuity may be useful in assigning workers to different roles, but there is no evidence that this is possible or useful at this stage.

6. Study 2 (Summer): Results

The purpose of the Summer study was comparative along two axes. Firstly to see if we could identify similar trends and patterns in the data to those seen in Study 1 (as a means of identifying if results were generally reproducible). Secondly to see if we could discern differences in the data that could be attributed to the differences in temperature, i.e. to find out if extremes of heat or cold had any amplification effect. The study was conducted with crew 3 from the winter study so that participants (and roles) were the same and could be directly compared. However, there were some differences in the working environment that were outside of our control and which directly impact comparison of results. The crew were working in a different location, as such the terrain and distances between parts of the work site were different from the first study. In addition, one of the quality controllers (3B) performed other duties during the summer and acted in a more multifunctional role (assisting the feller and log makers primarily on day 3) which affected his physiological data. The location itself also had a direct impact on our reaction time monitoring as we were only able to gain access to the participants at the start and end of day and not at the lunch break. We first present the results for Crew 3 during the Summer, then we compare the results

Participant	Diff Morning / Evening	Role	Morning	Break	Evening
2C	-136.77	quality	164.70	67.56	27.93
2E	-94.22	processer	222.81	250.81	128.60
3D	-37.89	loader	195.44	139.88	157.55
2D	-36.62	feller	165.47	104.71	128.85
2B	-31.65	quality	156.69	146.90	125.03
1C	-15.20	quality	91.57	92.49	76.37
1B	3.76	loader	372.65	355.33	376.40
1E	19.04	processer	114.32	94.81	133.37
2A	19.57	loader	162.73	161.55	182.30
3C	20.59	feller	174.57	162.42	195.16
1A	38.39	loader	148.21	212.22	186.60
1D	53.77	feller	226.28	229.99	280.05
3B	83.72	quality	32.18	146.24	115.89
3A	89.75	quality	55.55	198.67	145.30
3E	151.73	loader	129.75	262.06	281.48

Table 3: Difference Decision Making Time (DMT): morning and evening

Participant	Role	Morning	Break	Evening
2C	quality	164.70	67.56	27.93
1C	quality	91.57	92.49	76.37
3B	quality	32.18	146.24	115.89
2B	quality	156.69	146.90	125.03
2E	processer	222.81	250.81	128.60
2D	feller	165.47	104.71	128.85
1E	processer	114.32	94.81	133.37
3A	quality	55.55	198.67	145.30
3D	loader	195.44	139.88	157.55
2A	loader	162.73	161.55	182.30
1A	loader	148.21	212.22	186.60
3C	feller	174.57	162.42	195.16
1D	feller	226.28	229.99	280.05
3E	loader	129.75	262.06	281.48
1B	loader	372.65	355.33	376.40

Table 4: Decision Making Time (DMT): ordered by evening speed

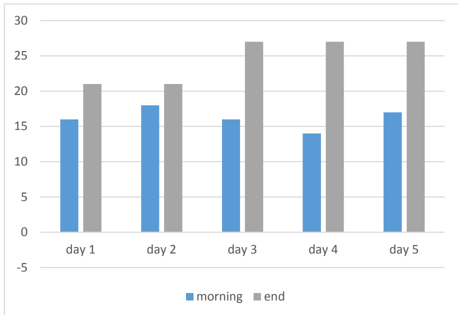

Figure 21: Temperature during study days, only morning and evening available (Crew 3 Summer)

of the two studies in terms of patterns that can be observed and the influence of different temperatures.

6.1. Results for Crew 3

Ambient Temperature. Temperature across the course of this study varied between 12°C and 27°C with the weather being fine throughout the monitoring period. Figure 21 shows the temperature at the start and end of each day for the 5 day study period.

Physiological data. Activity levels were determined using the same techniques employed during the winter study with data being collected on step and heart rates. The longer duration of this summer monitoring period provides data for a full working week (40hrs) providing a better indication of weekly activity by role. Figure 22 details the cumulative step rates encountered by role. It should be noted that the unusually high step rates performed by one of the quality control operators (3B) (>100,000) is likely to be a result of a more multi-functional role during the study period.

As in our winter studies, we monitored

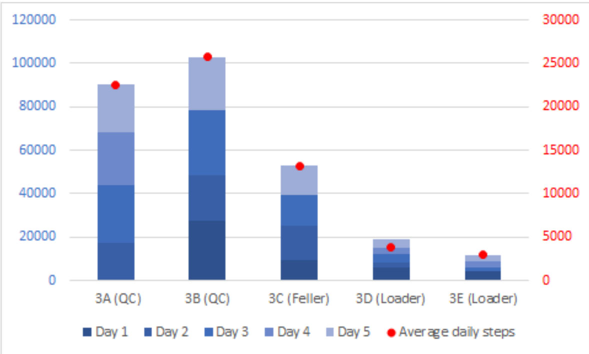

Figure 22: Step count by participant (Summer)

heart rates throughout the workday to assess if any significant difference was present between summer and winter months. Figure 23 presents the results of the heart rate measurements across the duration of the study. Participant 3A (Quality Control), who has the second highest step count, shows the highest levels of heart rate exertion. Participant 3B (Quality Control) however, with the highest step rate, has a more consistently low level of heart rate exertion, although also high variability (which is also seen with 3C). The feller, 3C is also at higher exertion levels which is to be expected given the physical nature of his role. The two loader drivers, 3D and 3E have heart rates that generally remain in the low exertion range, although both are typically higher in the afternoon, and 3E in particular has increased levels across two afternoons and several spikes during Days 4 and 5.

Reaction Time. Due to site access limitations reaction time measurements were only taken at the start and end of the working day. Figures 24 and 25 present the results of the reaction time testing for the summer study. Of interest are the results for Loader 3D. He has the fastest SRT times for days 1, 4 and 5, but his corresponding CRT times

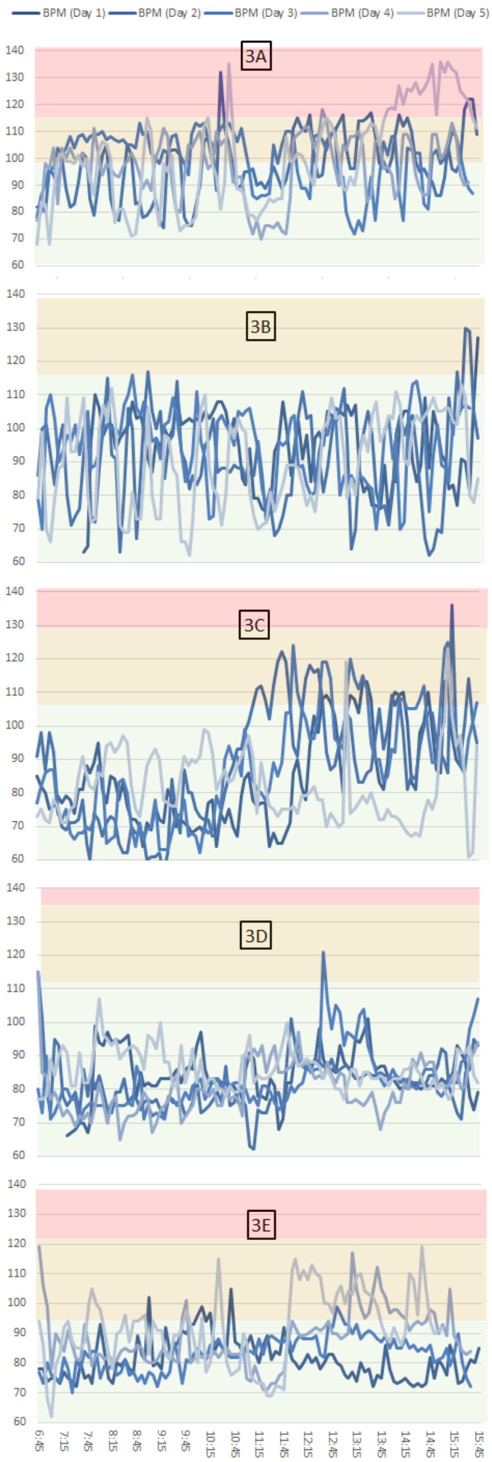

Figure 23: Crew 3 Summer study

are the slowest. The other Loader, 3E has very erratic (and slow) SRT times (on the morning of day 1 his speed is 804ms which is too slow to show on the graph). However his CRT results are consistent (and among the fastest). No data was recorded for the evening on day 5 for both 3D and 3E. Given the early start the loader drivers have they are keen to leave site at the end of the working day and it is not always possible to persuade them to stay and undertake the reaction time tests.

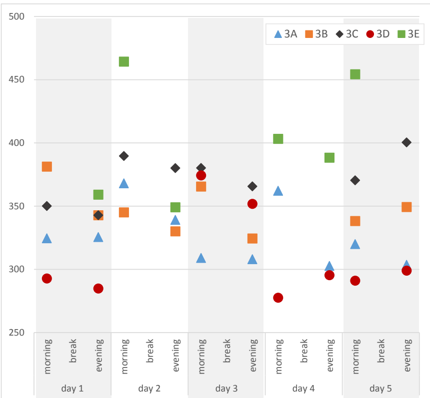

Figure 24: Simple Reaction Times (Summer)

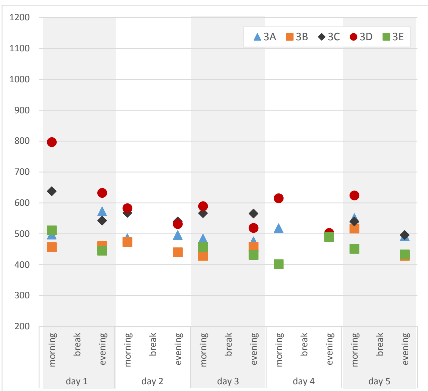

Figure 25: Choice Reaction Times (Summer)

6.2. *Analysis of Summer vs Winter*

We now present a comparison of the Winter and Summer data to see if there are similar patterns that can be identified, or obvious indications of an amplification effect due to temperature.

Steps. Figure 26 shows a comparison of average steps taken by the participants of Crew 3 in summer and winter. We found

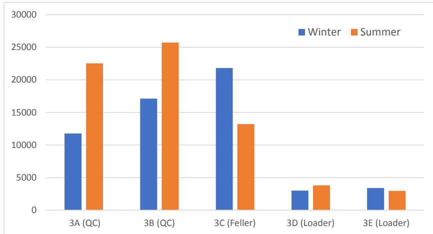

Figure 26: Average Steps Winter (blue) vs Summer (orange)

comparison of step rates across studies can be problematic, especially for Quality Control roles, the distance to the operational area can be significantly different, requiring higher step rates between the safe zone and skid where operations are performed. This is similar for the fellers who are walking between operational areas which may differ greatly across different sites. There was a decrease in mean daily steps for the feller of 7,500 during the summer session. While this may indicate the feller was self pacing (due to higher temperature and humidity levels encountered in their operational area), it could equally be an artefact of the different working location where perhaps the site area being covered was smaller.

There is some evidence that heart rate patterns are consistent across summer and winter, suggesting that the data capture methods are providing reproducible results. Figure 27 presents a comparison for each participant's heart rate between the two studies. Participants 3B, 3C, 3D and 3E show similar trends across the daily average, while 3A has a larger variation between summer and winter averages. Winter heart rates are on average slightly higher, and for four of the participants are higher at the start of day in the coldest temperatures, which may suggest that the colder temperatures do have an amplification effect on heart rate exertion measures.

When we compare the results of simple and choice reaction times between the two studies (see Figure 28), we can see similar results for both CRT and SRT, apart from 3E (Loader) who is noticeably slower in Winter, particularly at the start of day. 3A (Quality control) is also slower (SRT) at the start of day in Winter.

The heart rate and reaction time comparisons across Winter and Summer suggest that the data is reproducible (similar patterns seen) which gives some confidence in the measurements and collection methods. However, there are no discernible differences between different ambient temperatures (or the extremes at either end) that show this is having any effect generally. Table 5 provides a comparison of decision making times at start and end of day in winter and summer. The lighter colours indicate faster speeds so we can see individual patterns for each worker in the two different seasons. 3A, 3D and 3E exhibit the same patterns (fast-slow, slow-fast, fast-slow respectively) in both winter and summer, but 3B and 3C reverse their winter fast-slow pattern to slow-fast in summer. We discuss this further in the next section.

Participant	Role	Morning (Winter)	Evening (Winter)	Morning (Summer)	Evening (Summer)
3A	quality	55.55	145.30	170.44	191.01
3B	quality	32.18	115.89	89.41	88.39
3C	feller	174.57	195.16	164.39	131.09
3D	loader	195.44	157.55	301.61	121.25
3E	loader	129.75	281.48	39.95	263.45

Table 5: Decision Making Time (DMT): winter vs summer

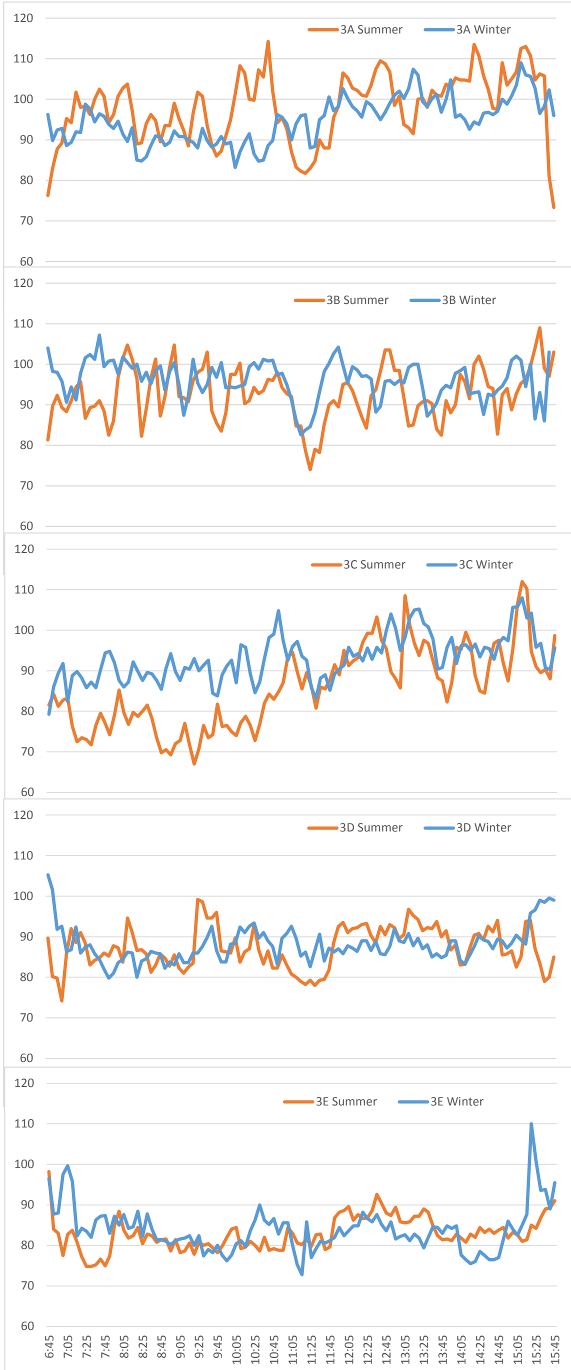

Figure 27: Heart rate winter (blue) v summer (orange), from top: 3A, 3B, 3C, 3D, 3E

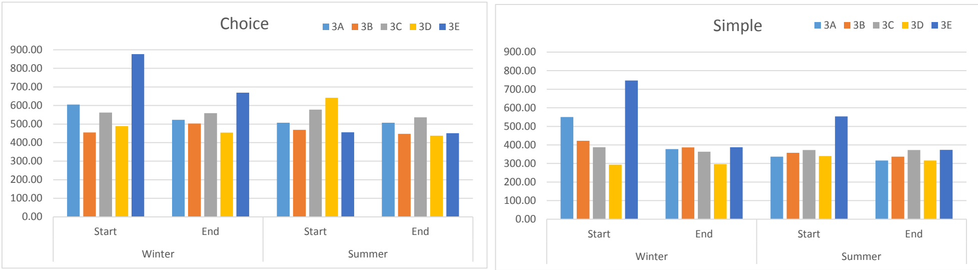

Figure 28: Mean difference in reaction time winter v summer

7. Discussion

We here discuss the results of our two studies in the light of related work and the context of workplace monitoring ethics.

Steps taken. When examining the amount of steps taken across the working day we find that manual roles result in high step rates that exceed those found in other employment types. Porcari and Ekhwan (2007) conducted a study for the American Council on Exercise into the amount of steps taken by employees in 10 common occupations. They found a wide variation in the step rates of these professions ranging from 4,300 steps for secretaries through to 15,251 for mail carriers. As we can see, harvesting roles generate levels of activity (measured as steps) that far outweigh those encountered in other professions. However, there is no clear indication that these high step counts equate to higher levels of fatigue.

Heart Rate. We consider whether the heart rate measurements enable us to differentiate between roles encountered in our study. We find mechanised roles do not produce overly high heart rates. We hypothesise that high points seen around the start and end of the lunch break only occur as a result of individuals climbing in and out of machinery cabs. We would need to confirm this with visual analysis at the same times to be sure of this.

For the manual roles, the variation in heart rate across the workday is more pronounced (e.g., 1D, 3C). In addition, the higher heart rates of these workers enhance the step count information by indicating that fellers (who are performing high levels of manual labour as well as walking) typically work at a higher rate of physical exertion than quality control roles with similar step counts. However, it is not the case that the most physically active workers have significantly slower reaction times at the middle or end of day so we cannot infer that they are more fatigued or impaired than workers in other, less manual, roles.

Bates and Schneider (2008) used the observed differences in heart rates to identify break periods throughout the course of a workday. Used in this way it may help compliance with legislation by ensuring workers take adequate breaks. Although some similar observations of lower exertion can be found in our results for the lunch break period the data is not sufficiently contextualised or significant to warrant its potential use in questions of compliance. Also, heart rate in particular and physically induced fatigue in general are highly personalised to an individual. Age, gender, fitness-level and overall health are all contributors to these, and as such the data must be considered within this context. Commercial solutions, such as the Readiband sleep tracker from Fatigue Science[5] typically develop an initial personalised baseline for each individual and consider subsequent data in relation to this. This, however, requires additional study time (to set the baseline), requiring more time commitments from participants which is often problematic. We discuss this further shortly.

Reaction times. When we examine the available workplace accident data (see Figure 1 in Section 1) by time of day, we see two definite spikes in time periods of incident occurrence: the first occurred between 10am and 11am, the second between 2pm and 3pm.

Brisswalter et al. (1997) found that the effects of physical exercise led to a decrease in cognitive performance (i.e., choice reaction time), however simple reaction time alone showed no significant difference. Our results show differences between SRT and CRT, with greater variation in SRT results. However, neither SRT or CRT show a clear correlation to physical activity. As such, while workplace fatigue, manifested as slower reaction times, may be a contributor to the periods of increased incident rates, we cannot currently identify a direct correlation between reaction time trends and workplace incidences. Measures for CRT and SRT are known to be highly variable across different timescales for individuals. Even with outliers removed (as discussed in Section 5) the variability of the data means that it cannot be used individually as an indicator of either fatigue or work-induced stress.

[5]https://www.fatiguescience.com

Influence of Temperature. Our investigations of the effect of temperature on performance were driven by comments from work crews during our winter studies. All of the crews indicated that they felt that summer work had the greatest impact on their fatigue levels due to higher temperatures. (Pilcher et al., 2002) concluded from a meta-study that both hot (>32°C) and cold temperatures (<10°C) have negative effects on performance. Similar temperature extremes are reached in our studies during several mornings (e.g., in day 1,2,3, 8 and 9 for Crew 1, day 1 for Crew 2, and all five days for Crew 3). However, we found that although they feel the work is harder in the summer months, their performance is not significantly impacted detrimentally by higher temperatures. If anything, our comparisons in Section 6.2 rather shows a small decline at the lower temperatures during winter months but this is not significant.

Sleep and fatigue. As we discussed in the introduction to this work, we are aware of the importance of sleep and the role it can play in workplace fatigue. The importance of sleep in relation to waking performance has been investigated as a contributor to fatigue many times (Williamson and Feyer (2000), Åkerstedt et al. (2002), Belenky et al. (2003)) with conclusions that low duration or poor sleep quality adversely affecting ones mood, physical performance and cognitive processing abilities. Several studies have identified that increases in reaction time occur with sleep deprivation (Van Dongen et al. (2003), Lim and Dinges (2008), Kim et al. (2011)) suggesting that levels of sleep achieved by an individual can influence performance. More importantly it is further suggested that moderate sleep deprivation can impair performance similar to those levels found in alcohol intoxication Williamson and Feyer (2000).

Anecdotally, forestry management have expressed concerns about sleep quantity and quality of their workers. At one of our initial meetings, one individual stated that younger workers might “party all weekend” and “turn up for work on Monday morning exhausted”. However, there is no evidence for this. Our early attempts to include sleep tracking as part of our studies led us to understand that there is a high level of resistance for this from workers who feel it is an invasion of their privacy. This is further supported by the forestry industry’s own experience of trying to conduct a study using the Fatigue Science Readiband [6] sleep tracking and fatigue monitoring solution, which is seen as the ‘gold-standard’ method for collecting sleep data outside of a dedicated sleep laboratory. At the time of writing, the health and safety organisation trying to run the sleep study had been unable to recruit enough volunteers among the forestry workers to conduct the study, despite delaying it several times. This reluctance from the workers to be monitored in this way is reflected in our own ethical concerns about the collection and use of such data, leading to questions such as how will a management team deal with a worker who is identified as being fatigued due to a period of poor sleep? Further discussions of data privacy and ethical concerns in the context of monitoring forestry workers are presented in (Bowen et al. (2017)).

Contextualising monitoring data. There are still many ‘myths’ in the consideration of fatigue and accident rates and how these can be reduced by worker monitoring. We have found an increase in the desire from management teams to find ways of monitoring workers (both from those involved in our studies, and those seeking to utilise other options such as the Readiband or other proprietary solutions) but this is not necessarily coupled with evidence to support such monitoring. Our studies so far suggest that data must be both contextualised and individualised before it can be meaningfully used to indicate fatigue, and even then it should be understood that indicators of fatigue do not directly correlate to accidents or near misses.

When analysing the type of data we have presented here we must take into account the lack of context. For example although we can see the details of the physiological data throughout the day, if a particular data point is directly affected by an event in the field (e.g. a worker’s heart rate spikes because of a near miss accident) we cannot determine this. Even the SRT/CRT testing is prone to the effect of distraction, on at least one occasion we were aware of a worker being distracted by someone else entering the break room as the were doing the response-time testing, and it is likely that more minor distractions were also present that we were unaware of. In addition although we may be more successful at identifying relationships across aggregated data, this is not useful if we wish to capture real-time data and analyse it in the field to provide instant feedback to (or about) individual workers.

Study limitations. The final consideration is the lack of control we have over when and how to perform monitoring. Both workers and management are mindful of work performance targets, so we cannot interrupt worker time to any great extent. Conversely, as workers get limited break time

[6] https://www.fatiguescience.com/

they are resistant to undergoing any testing that eats into that time. We would have liked to have performed reaction time testing prior to their lunch break as well as at the end, but this was not possible due to the workers' need to take the break, have lunch etc. before they would engage in our activities. The accident data presented in Figure 1 shows two peaks, one immediately prior to the lunch break and the other mid-afternoon. Ideally we would like to have gathered reaction time data at exactly these points, but it proved impossible for us to collect more detailed data at these times other than the automatically collected physiological data.

8. Summary and Conclusions

Our studies aimed to investigate the use of in-situ data collection in the New Zealand forestry industry. Specifically, we wanted to identify suitable measurements and measuring techniques from two perspectives: (1) could we reliably capture data from forestry workers over the course of their working day; and (2) could we find meaningful correlations in the data to suggest that we could identify fatigue in workers based on their reaction times and perceived workloads.

Reliable data capture. Results from our two studies show that we can collect data using lightweight off-the-shelf equipment, although there are some restrictions to this. For example, our studies have used wrist-worn commercial activity trackers to collect heart rate data, whereas heart rate variability is likely to provide more reliable data for considering fatigue. However measurement of heart rate variability in an accurate manner requires the use of chest-straps (wrist-worn light-based heart monitors are not accurate enough in this domain) and these can be uncomfortable to wear for long periods of time by workers engaged in more physical roles. For our next series of studies (future work) we are looking at incorporating such chest-strap monitors into compression shirts so that they are more comfortable to wear.

Meaningful correlations. The data that we collect automatically (including step counts and heart rate) can be compared to reaction time tests which use simple and choice reaction time as an indicator of impairments. However, we did not find significant correlations in our data to show that we can determine fatigue-based impairments from our measurements. Not only do personal factors have a large influence on the physiological data, but there are contextual elements for both types of data (e.g. distractions when a worker undertakes the reaction-time tests, or the desire to finish quickly at the end of the day affecting mindset) which also have an effect. While this may seem to be a disappointing outcome it does provide valuable information for forestry health and safety bodies who are keen to adopt such monitoring approaches. Although some commercial solutions do promise to be able to accurately identify fatigue based on similar measures to our own, our results suggest that they should be cautious in adopting them without investing significant time to study their use in the forestry domain.

It is also important to note that our monitoring periods were short and our participant numbers small. As such our studies can only provide a snap shot of physiological and reaction time data across a limited time period. Extended data collection over longer time frames with larger groups of workers may enable us to better identify any trends that maybe present as well as evaluate the reliability of the data collection by way of repeated results. Our future work includes conducting longer studies with larger groups of workers and extending the measurements taken. In addition they will include comparisons with some of the commercial solutions being considered by the industry to see if these produce results that are different or can somehow be validated as more accurate.

In the longer term we are also investigating how the data may be used as part of a larger solution based around an Internet-of-Things solution which captures a wider variety of data (mixing automatic measures with self-reporting and ambient sensors).

References

G. Adams, H. Armstrong, and M. Cosman. Independent forestry safety review – an agenda for change in the forestry industry, October 2014.

T. Åkerstedt, P. Fredlund, M. Gillberg, and B. Jansson. Work load and work hours in relation to disturbed sleep and fatigue in a large representative sample. *Journal of psychosomatic research*, 53(1):585–588, 2002.

J. A. Alvarez and E. Emory. Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review*, 16(1):17–42, 2006.

G. P. Bates and J. Schneider. Hydration status and physiological workload of uae construction workers: A prospective longitudinal observational study. *Journal of occupational medicine and toxicology*, 3(1):1, 2008.

G. Belenky, N. J. Wesensten, D. R. Thorne, M. L. Thomas, H. C. Sing, D. P. Redmond, M. B. Russo, and T. J. Balkin. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: A sleep dose-response study. *Journal of sleep research*, 12(1): 1–12, 2003.

J. Bowen, A. Hinze, S. J. Cunningham, and R. Parker. Evaluating low-cost activity trackers for use in large-scale data gathering of forestry workers. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, OzCHI '15, pages 474–482, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3673-4.

J. Bowen, A. Hinze, C. Griffiths, V. Kumar, and D. Bainbridge. Personal data collection in the workplace: ethical and technical challenges. In *British Human Computer Interaction Conference, (BHCI)*, page 10 pages, 2017.

J. Brisswalter, R. Arcelin, M. Audiffren, and D. Delignières. Influence of physical exercise on simple reaction time: effect of physical fitness. *Perceptual and Motor Skills Journal*, 85:1019–1027, 1997.

Competenz. Forestry cluster needs analysis, 2012. online available at `http://troq.competenz.org.nz/assets/TROQ/Documents/Forestry/Needs-Analysis-Report-Forestry-22112012-final.pdf`.

I. J. Deary, D. Liewald, and J. Nissan. A free, easy-to-use, computer-based simple and four-choice reaction time programme: the deary-liewald reaction time task. *Behavior Research Methods*, 43 (1):258–268, 2011.

F. C. Donders. On the speed of mental processes. *Acta psychologica*, 30:412–431, 1969.

F. Galton. An instrument for measuring reaction time. *Report of the British Association for the Advancement of Science*, 59:784–785, 1889.

R. Hockey. *The Psychology of Fatigue, Work, Effort and Control.* University of Sheffield, 2013. ISBN 9780521762656.

Human Work Interaction Workshop. position papers, 2015. available online at `https://projects.hci.sbg.ac.at/hwid2015/position-papers`.

H. J. Kim, J. H. Kim, K.-D. Park, K.-G. Choi, and H. W. Lee. A survey of sleep deprivation patterns and their effects on cognitive functions of residents and interns in korea. *Sleep medicine*, 12(4):390–396, 2011.

S. Kumar, Y. Amell, T.and Narayan, and N. Prasad. Indices of muscle fatigue. In *Proceeding of Human Factors and Ergonomics Society Annual Meeting*, volume 13, 2002.

Legislation, New Zealand. Acts, 2008. *Minimum Wage Act, 1983. New Zealand.*

R. Lilley, A. Feyer, P. Kirk, and P. Gander. A survey of forest workers in new zealand. do hours of work, rest, and recovery play a role in accidents and injury? *Journal of Safety Research*, 33(1): 53–71, 2002.

J. Lim and D. F. Dinges. Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences*, 1129(1):305–322, 2008.

Y.-H. Lin, C.-Y. Chen, S.-Y. Lu, and Y.-C. Lin. Visual fatigue during vdt work: Effects of time-based and environment-based conditions. *Displays*, 29(5):487–492, 2008.

Ministry for Business and Innovation. Independent forestry safety review an agenda for change in the forestry sector, 2014.

P. Morrel-Samuels. Getting the truth into workplace surveys. *Harvard Business Review*, 80(2): 111–118, 2002.

A. Murata, A. Uetake, and Y. Takasawa. Evaluation of mental fatigue using feature parameter extracted from event-related potential. *International journal of industrial ergonomics*, 35 (8):761–770, 2005a.

A. Murata, A. Uetake, and Y. Takasawa. Evaluation of mental fatigue using feature parameter extracted from event-related potential. *International Journal of Industrial Ergonomics*, 35(8): 761–770, 2005b.

New Zealand Treasury. New zealand economic and financial overview 2016, 2016.

D. A. Norman and T. Shallice. *Attention to Action*, pages 1–18. Springer US, Boston, MA, 1986. ISBN 978-1-4757-0629-1.

R. Parker. *Technological advances in the analysis of work in dangerous environments: Tree felling and rural fire fighting*. PhD thesis, 2010.

Payscale. Hourly rate for industry: logging, 2014. online available at `http://www.payscale.com/research/NZ/Industry=Logging/Hourly_Rate`.

V. Pichot, E. Bourin, F. Roche, M. Garet, J.-M. Gaspoz, D. Duverney, A. Antoniadis, J.-R. Lacour, and J.-C. Barthélémy. Quantification of cumulated physical fatigue at the workplace. *Pflügers Archiv*, 445(2):267–272, 2002.

J. J. Pilcher, E. Nadler, and C. Busch. Effects of hot and cold temperature exposure on performance: a meta-analytic review. *Ergonomics*, 45(10):682–698, 2002.

J. Porcari and R. Ekhwan. Do you do 10k a day? *American Council on Exercise, Fitness Matters*, 12(4):9–12, 2007.

R. A. Robergs and R. Landwehr. The surprising history of the" hrmax= 220-age" equation. *Journal of Exercise Physiology Online*, 5(2):1–10, 2002.

A. H. Sabzi. The effect of different fatigue protocols on choice reaction time. *Middle-East Journal of Scientific Research*, 12, 2012.

K. Saito. Measurement of fatigue in industries. *Industrial health*, 37(2):134–142, 1999.

A. Spurgeon, J. M. Harrington, and C. L. Cooper. Health and safety problems associated with long working hours: a review of the current position. *Occupational and Environmental Medicine*, 54 (6):367–375, 1997.

The Parliament of New Zealand. Health and safety at work act 2015, 2015. online available at `http://www.legislation.govt.nz/act/public/2015/0070/latest/DLM5976660.html`.

D. van der Linden, M. Frese, and T. F. Meijman. Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica*, 113(1):45 – 65, 2003.

H. P. Van Dongen, G. Maislin, J. M. Mullington, and D. F. Dinges. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *SLEEP-NEW YORK THEN WESTCHESTER-*, 26(2):117–129, 2003.

A. Williamson, D. A. Lombardi, S. Folkard, J. Stutts, T. K. Courtney, and J. L. Connor. The link between fatigue and safety. *Accident Analysis & Prevention*, 43(2):498 – 515, 2011. ISSN 0001-4575. doi: http://dx.doi.org/10.1016/j.aap.2009.11.011. Advancing Fatigue and Safety Research.

A. M. Williamson and A.-M. Feyer. Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. *Occupational and environmental medicine*, 57(10):649–655, 2000.

A. M. Williamson, A.-M. Feyer, R. P. Mattick, R. Friswell, and S. Finlay-Brown. Developing measures of fatigue using an alcohol comparison to validate the effects of fatigue on performance. *Accident Analysis & Prevention*, 33(3):313–326, 2001.