

Fatigue consideration in operational task assignment

A Bi-Objective Job-Shop Scheduling Problem Considering Worker Fatigue and Productivity in Cobotic Order Picking Systems

MSc Robotics, MSc Multi-Machine Engineering,
Delft University of Technology

B.L.Vermin



Fatigue consideration in operational task assignment

A Bi-Objective Job-Shop Scheduling Problem Considering Worker Fatigue and Productivity in Cobotic Order Picking Systems

by

B.L.Vermin

Student:	Berry Lance Vermin
Student Number:	4443012
Chair Committee:	Prof. dr. ir. David Abbink
Daily Supervisor:	Dr. Frederik Schulte
Committee member:	Dr. Oscar Oviedo Trespalacios
Committee member:	Dipl. ir. Christoph Schmidt
Institution:	Delft University of Technology
Faculty:	Faculty of Mechanical, Maritime and Materials Engineering, Delft
Company mentor:	Duco van Holthe
Company:	Picnic Technologies
Project Duration:	June, 2022 - March, 2023

Cover Image: Courtesy of Picnic Technologies



"But I would learn from my previous mistakes: I would not make them again; I would make completely new mistakes. That's really all life is, making newer, better mistakes."

SuperfastMatt

Abstract

The increase in online retail demand has stimulated automation in order picking systems, leading to new challenges and opportunities in task assignment and scheduling. In partially automated order picking systems, such challenges and opportunities exist regarding human factors implementation in the job-shop scheduling problem, an optimisation problem essential in operations. Workplace fatigue is a human factor often overlooked in scheduling research and application, despite hurting employees' well-being and costing U.S. employers up to €127 billion annually. With the opportunities that automation offers, cobotic order picking systems could actively consider human fatigue development, mitigating its negative effects in operation.

This thesis investigates the possibility and potential benefits of fatigue consideration in the job-shop scheduling problem for a partially automated order picking system. We present a new bi-objective mixed integer nonlinear programming problem formulation to represent system constraints and a predictive fatigue model while considering worker fatigue and productivity during schedule optimisation. To put the results of simulated optimisation in perspective, we experimentally validate the fatigue model predictions and fatigue mitigation capabilities of the scheduling approach using heart rate measurements and qualitative fatigue ratings. These experiments occur with employees in a real-life partially automated order picking system.

Our mathematical model can find solutions that the conventional single-objective optimisation approach cannot, allowing fractional energy expenditure distribution improvements more than 4x larger than the decrease in productivity they require in 53% of the considered virtual cases. This is a promising result for fatigue mitigation in operations only by altering operational decision-making. However, the validation experiments show that our predictive fatigue model has an average RMSE of 2.20 kcal/min in estimating energy expenditure rates compared to heart rate measurements while also showing a low correlation. When assessing 10 minute intervals, a time span that fits a scheduling scope, the estimations improve slightly (avg. deviation of -1.85 kcal/min, avg. correlation of 0.17) but still underestimate the measured values. The experiments also show no significant differences in experienced fatigue between existing schedules and those with fatigue mitigation measures applied.

We conclude that the current scheduling formulation is not yet fit for application with a predictive fatigue model. However, real-life operations can benefit from energy expenditure estimation via heart rate measurements and a different approach for implementation is proposed. Research opportunities lie in further fatigue model development and validation, extension to indirect fatigue effects and other human factors, and further development of the mathematical formulation.

Preface

In late 2021, I started my thesis internship at Picnic. After two years of studying by myself, under the shadow of Covid-19 restrictions, my colleagues at Picnic were very welcoming and made me feel included in the automation team from day one. I feel lucky to have been a part of this amazing team and the friends that I made along the way.

Maarten, Luuk, Pita, Daniel, Annemarijn, Ariëla, Cris, Emi and Bram, thanks for making every day at the office enjoyable, even when my thesis was not. Furthermore, I want to thank Wessel, Julia, Ruben, Kevin and Corneel for their critical feedback, input and support when I needed it, even during busy (Christmas) periods. Also, I want to thank my Picnic supervisor Duco for his support and guidance during this lengthy thesis process - and for trying to make sense of my overextensive stories.

Second, I thank my TU Delft supervisors David and Frederik for their guidance and the interesting conversations we shared, keeping me on my toes at all times. Your contributions helped me bring my research to a higher level. A big thanks too to Eline and Lara for sharing their specialised knowledge.

Thirdly, I want to thank my dear friends for their support, feedback and much-needed distraction. To Hugo, Suus, Wouter, Timo, Bram, Kwikka, Michiel, Olger and Manon: I am proud to call you my friends and have you around.

Lastly, I thank my parents and sister for their never-ending love, time and mostly for giving me the freedom and opportunity to follow my passion and become the person I am today, both professionally and personally. No words could express the gratitude I feel for having you in my life.

*B.L. Vermin
Delft, March 2023*

Contents

Abstract	i
Preface	iii
Nomenclature	vi
1 Introduction	1
1.1 Picnic's order picking system	1
1.1.1 Picnic's zone picking process	1
1.1.2 Operational challenges at zone picking stations	2
1.2 Fatigue in the workplace	3
1.3 Human factors in the workplace	4
1.4 Problem statement	4
1.5 Structural outline	5
2 Literature review	7
2.1 Fatigue and recovery models	7
2.1.1 Fatigue quantification	7
2.1.2 Practical application	12
2.2 Human factors in job-shop scheduling	15
2.2.1 Job-shop scheduling problems	15
2.2.2 Implementation of fatigue in the job-shop scheduling problem	17
2.2.3 Solving JSP with human factors	19
2.3 Existing research gaps	20
3 Modelling approach	23
3.1 Fatigue model	23
3.2 Mathematical job-shop scheduling formulation	23
3.2.1 Setting up the bi-objective	24
3.2.2 Constraints	26
3.3 Model optimisation	28
3.3.1 Selecting a solution approach	29
3.3.2 Computational complexity	29
3.3.3 Verification of constraints and model output	30
3.3.4 Estimating the Pareto front	33
4 Empirical study	37
4.1 Environment	38
4.2 Test subjects	39
4.2.1 Qualitative ratings and personal information	39
4.3 Fatigue model validation	39
4.3.1 Heart rate data measurement and preprocessing	39
4.3.2 From sensor data to energy expenditure rate	40
4.3.3 Order data	42
4.3.4 Pick events	42
4.4 Schedule validation	44
5 Results	45
5.1 Modelling study	45
5.1.1 Pareto front estimate with global optima	45
5.1.2 Pareto front estimate without global optima	47

5.2	Empirical study	48
5.2.1	Heart rate and %HRR	48
5.2.2	Energy expenditure rate	51
5.2.3	Rest allowance	58
5.2.4	Qualitative fatigue ratings	58
6	Discussion	61
6.1	Modelling study	61
6.1.1	Implications of the results	61
6.1.2	Limitations	62
6.2	Empirical study	62
6.2.1	Implications of the results	62
6.2.2	Limitations	63
6.3	Relevance to human factors and operations research	63
6.4	Relevance to operational management practice	64
6.5	Opportunities for further research	64
6.6	Managerial recommendations	65
7	Conclusion	67
7.1	Summary of the main contributions	67
7.2	The answer to the research questions	67
7.3	Final remarks	68
	Bibliography	69
	Appendices	75
A	Paper format	76
B	Workload questionnaire report	99
B.1	Methods	99
B.2	Key takeaways	99
B.3	Results and discussion	100
C	Generation of model inputs	103
C.1	Shopper properties	103
C.2	Inventory	103
C.3	Item properties	104
C.4	Sets and parameters	104
C.4.1	Task decomposition	104
C.4.2	From task decomposition to pick info	109
D	Additional results from the mathematical optimisation	111
E	Graphical user interface for experiments	115
F	Additional experimental results	121

Nomenclature

Acronyms

Abbreviation	Definition
AGV	Automatic Guided Vehicle
BMR	Basal Metabolic Rate
BPM	Beats Per Minute
DRC	Dual-Resource Constrained
DV	Decision Variable(s)
ECG	Electrocardiogram
EE	Energy Expenditure
EMG	Electromyography
FC	Fulfilment Centre
FSP	Flow-Shop (Scheduling) Problem
GA	Genetic Algorithm
GTP	Goods-To-Person
GUI	Graphical User Interface
HF	Human Factors
HRR	Heart Rate Reserve
HRV	Heart Rate Variability
JSP	Job-Shop (Scheduling) Problem
KPI	Key Performance Indicator
LFFR	Learning-Forgetting-Fatigue-Recovery (Model)
MET	Maximum Endurance Time
MIP	Mixed Integer Programming (problem)
MILP	Mixed Integer Linear Programming (problem)
MINLP	Mixed Integer Nonlinear Programming (problem)
MOO	Multi-Objective Optimisation
MVC	Maximum Voluntary Contraction
MWR	Mean Work Rate
OP	Order Picking
RA	Rest/Relaxation Allowance
RHR	Resting Heart Rate
RMSE	Root-Mean-Squared Error
RPE	Rating of Perceived Exertion
SA	Simulated Annealing
WMS	Warehouse Management System

Introduction

With customers doing more online shopping than ever before, retail is shifting its business away from traditional brick-and-mortar stores and into an online setting (Coppola, 2021). This online setting necessitates retailers to prepare orders per customer, therefore needing to pick all selected items before delivery. With labour accounting for up to 55% of operational costs in order picking (OP), businesses are looking for new, more efficient ways to prepare customers' orders (de Koster et al., 2007).

By introducing automation in the OP process, retailers can increase their productivity per labour hour, allowing them to fulfil a greater demand at reduced operational costs. Robotic picking systems are fast-evolving but still limited in application to a wide range of items (Vijayakumar and Sgarbossa, 2021; Bormann et al., 2019). In the online grocery sector, which has seen increasing demand in recent years (Verschueren et al., 2021), automation is often only partially applied, and reliance on humans continues to exist. In a partial automation environment, we find resources related to humans and automation, making this a dual-resource constrained (DRC) system.

One operational challenge in DRC OP systems is a combinatorial optimisation problem called the job-shop scheduling problem (JSP). The JSP describes assigning specific OP tasks to employees, a picking station and a time to optimise some objective while preparing a product in a unique sequence of steps (Thürer et al., 2020). In OP systems, these steps are individual item picks for an order. Generally, the objective is to minimise the total production time or 'makespan', but many other objectives exist (Xiong et al., 2022). Optimisation of the JSP in a DRC system comes with additional challenges, but also ways in which the automated resources can support humans in ways previously impossible.

1.1. Picnic's order picking system

This thesis is written in conjunction with Picnic, an online grocery company. In their OP process, individual items are picked by 'shoppers' into special crates called 'totes' in a fulfilment centre (FC). From the FC, the totes are batched and transported by trucks to distribution hubs, from which smaller batches are shipped to customers in small electric vehicles. FCs can service several hubs and therefore are a valuable object for optimisation and benefit from economies of scale.

Picnic has two types of FCs: manual FCs and (partially) automated FCs (see Figure 1.1). Currently, there is one automated FC in the Netherlands, located in Utrecht. We will refer to this OP system as 'FCA'. FCA is comprised of a number of different components that are interconnected by a network of conveyors. Order picking in FCA, or the zone pick area in particular, resembles a DRC system where the operational task assignment characterises as a JSP. Since FCA is still in ramp-up after opening nearly one year ago, the level of flexibility that the conveyor network offers in task assignment is new for Picnic and it offers many challenges for optimisation that are different from their more mature manual fulfilment centres. We will therefore focus this thesis on operations in FCA's zone pick system.

1.1.1. Picnic's zone picking process

In zone picking, shoppers are provided with customer totes (totes that contain items for a customer order) by conveyors to a picking station. The picking station contains multiple item storage locations, or 'slots', on pallets or rolling containers (see Figure 1.2a). A touch screen instructs the shopper in which slot the item



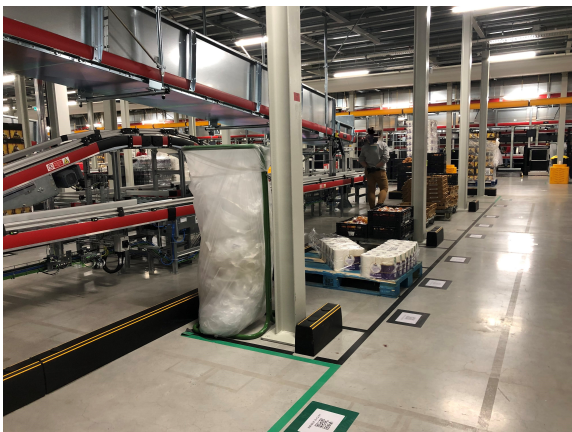
(a) Manual fulfilment centre.



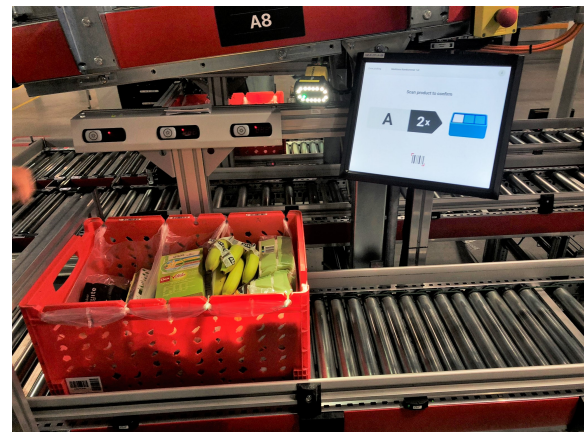
(b) Automated fulfilment centre.

Figure 1.1: Different types of order fulfilment systems by Picnic.

is stored, how many items to pick and in which of the three bags to place the item (see Figure 1.2b). After walking to this slot and back, shoppers scan the item and confirm the pick on the touch screen once it is placed in the correct bag. If all picks for the tote for the current station have been completed, it returns to the conveyor network and a new tote arrives automatically.



(a) Zone picking station with items stored in slots.



(b) Customer tote and instruction screen at zone picking station.

Figure 1.2: Overview of the zone picking station.

1.1.2. Operational challenges at zone picking stations

From a scheduling point of view, the operational task assignment in zone picking is interesting, as it requires discarding some assumptions that are common in JSP formulations. To better understand the operational challenges that Picnic faces at its zone pick stations, we must consider the separate stations as an element of the bigger picture. Each customer tote needs to pass multiple sections of the OP system before completion. As tote travel between different areas in FCA costs time and takes up space on the limited conveyor network, FCA's operational strategy is to minimise tote movements by having them only leave a picking area once all picks have been executed. For each area, the goal is to perform all required actions as quickly as possible to maintain a low order lead time and free up space for new order totes in the areas.

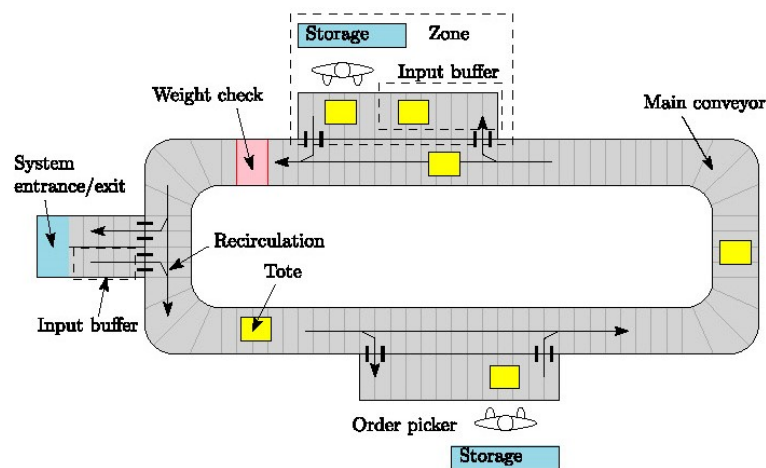


Figure 1.3: Schematic overview of the zone pick loop (Van Der Gaast et al., 2020).

More specifically, all zone pick stations are connected to one conveyor sub-loop (see Figure 1.3), for which we want to minimise the time spent by order totes. Each of the zones contains a subset of items, requiring most totes to visit multiple stations before completion. When we also consider the shoppers that execute the picks, the resulting optimisation problem can be considered a job-shop scheduling problem, as we need to assign the picking tasks to stations and shoppers. Because items can be picked on any station where the item is 'slotted', by any available shopper, this can be categorised as a double flexible job-shop scheduling problem (DFJSP) (Gong et al., 2018). FCA's zone pick being a DRC system, we can also expect constraints that relate to both humans and machines. This means that the resulting JSP should deal with job sequence, station assignment and shopper assignment altogether (Zheng and Wang, 2016).

At the same time, there can be a larger imbalance in physical workload between employees than in manual systems, as only a subset of items is slotted at each station. This storage location assignment can lead to large, heavy or large numbers of items being picked in one station, while other stations can have much less physically demanding picking tasks. A pilot test with a workload and fatigue questionnaire has pointed out that shoppers experience fatigue during their shifts and experience their work as physically demanding (see appendix B). Fatigue has consequences for employee well-being, productivity and Picnic's chances of retaining shoppers.

1.2. Fatigue in the workplace

Workplace fatigue can have costly effects on employers, with some studies estimating that the overall cost of fatigue-related productivity loss in the workplace is as high as 127 billion euros per year in the U.S. alone (National Safety Council, 2022). Fatigue in the workplace is a multidimensional construct in which we can identify external and occupational factors (Sedighi Maman et al., 2017). External factors include outside work activities, sleeping disorders, climate and other personal factors, while occupational factors include shift work, long hours and overtime, time on task, workload and break schedules.

A survey study among 4188 employees at four U.S. corporations investigated the negative effects of sleep deprivation in a working environment. The researchers estimated the fatigue-related productivity losses to be as high as €1850,- per employee annually while also seeing considerable safety risks (Rosekind et al., 2010). Although employers can take some steps to address workplace sleepiness due to sleep disturbances, a large portion of the external fatigue factors lies outside their influence.

This is not the case for occupational fatigue, which, as the name suggests, results from the work environment or shift schedule. Like sleep deprivation, occupational fatigue is negatively correlated with performance, as concluded in a study of 745 registered nurses (Barker et al., 2010). According to Fruggiero et al. (2016), occupational fatigue at a rate of 80% - corresponding to about 20 hours of wakefulness - results in effects similar to blood alcohol concentrations that exceed the legal limit. Next to obvious safety hazards, this leads to considerable economical impact. Not all occupational fatigue is completely avoidable in an environment that requires physical effort during work, but it is possible to minimise the negative implications on the design, organisational and operational level (Thürer et al., 2020).

1.3. Human factors in the workplace

Considering human factors (HF), such as fatigue, in work environments can simultaneously improve productivity and employee well-being (Kadir et al., 2019). The human factors research field concerns itself with:

"The understanding of interactions among human and other elements of a system and the profession that applies theory, principles, data, and methods to design in order to optimise human well-being and overall system performance." (IEA, 2000)

HF has an engineering focus, which is why the term human factors engineering is often used to describe the use of HF knowledge to "fit the task to the human" (Shorrock, 2019). While the disciplines that focus on the functioning of the individual in multiple contexts (e.g. anthropology, sociology and psychology) also focus on behavioural aspects, they would have a much more holistic view of one individual's behaviour than human factors research. The scope of the research field strictly considers the elements of behaviour that influence the individual's performance *within the system*. The goal of HF research is threefold: describing, predicting and controlling the human performance in the system structure. HF engineering has two primary purposes: increasing the effectiveness and efficiency of the tasks performed by employees and enhancing their well-being (Meister, 1989). According to Neumann and Dul (2010), HF consideration in real operations can simultaneously improve both of these aspects. However, as argued by Kadir et al. (2019), the conceptual tools, methods and designs for HF research are rarely applied or tested outside a controlled laboratory environment.

HF are essential in a partial automation environment but generally not or insufficiently taken into account in existing operations (Winkelhaus et al., 2021). If HF are considered, this is often done using mathematical HF models on the design level, such as workstation design (Neumann and Dul, 2010; Vijayakumar and Sgarbossa, 2021). However, in existing facilities like Picnic's OP system, this would require retrofitting the design level, which is not always possible and often expensive (Grosse et al., 2017). We also see that the current operational decision-making methods do not allow for fatigue consideration (Vijayakumar and Sgarbossa, 2021). Furthermore, operational decisions like line balancing and sequencing are now often done based on time, while the balance might be insufficient when physical effort or general fatigue of workers is taken into account (Finco et al., 2021). HF engineering can offer a solution to consider shopper fatigue while optimising the JSP, through application of mathematical models describing the causes and effects of HF. With the opportunities that automation offers, 'cobotic' OP systems - ones with active human-robot collaboration - could then actively consider human fatigue development, mitigating its negative effects in operation (de Koster, 2023).

1.4. Problem statement

In this thesis, we address this knowledge gap on occupational fatigue and the possibility of mitigating its negative effects through the implementation of fatigue models in the JSP. For Picnic's OP system, this entails taking into account worker, task and station heterogeneity and constraints in a partial automation environment, while optimising productivity and fatigue simultaneously. The problem statement is formulated as:

How can we simultaneously optimise worker fatigue and productivity in the job-shop scheduling problem for cobotic order picking systems?

We solve this problem using three research questions:

1. **How can we quantify, predict and measure fatigue for scheduling purposes in a heterogeneous workforce?**

This provides us with the personalised fatigue input that we need for the scheduling model.

2. **How can we apply fatigue models to a bi-objective JSP formulation that represents a partially automated order picking system?**

This provides a mathematical formulation of real-life constraints and can be used to generate fatigue-conscious schedules for the OP system.

3. **How do the fatigue model and resulting schedules perform in a real-life operation?**

This validation step provides context to the used fatigue model and pointers towards practical application of the proposed scheduling model.

We hypothesise that fatigue consideration in the JSP can be achieved by implementing a quantitative fatigue model in the JSP formulation and setting up a bi-objective optimisation with fatigue and productivity indicators. We expect predictive fatigue models to show some inaccuracies but still allow for appropriate fatigue mitigation in scheduling. Also, we expect fatigue-conscious schedules to lead to lower measured and perceived fatigue in real operations.

1.5. Structural outline

The remainder of this thesis is structured as follows. Chapter 2 reviews existing literature and provides the background knowledge related to our problem statement: fatigue quantification and measurement, human factors in job-shop scheduling problems, and solution techniques for scheduling problems. Chapter 3 describes the modelling approach used in the simulation study. First, we use the theoretical background to select an appropriate predictive fatigue model, after which we develop our hybrid mathematical model. We describe the model inputs and constraints, solution techniques, model verification and bi-objective development in detail. Then, in Chapter 4, we describe the empirical validation study, also with direct links to Chapter 2. In Chapter 5, we present the results of the scheduling model and fatigue model validation experiments. Here we also compare the results for different experimental circumstances and review our attempt to schedule with appropriate rest allowance. These results, our limitations and recommendations are discussed in Chapter 6. Finally, Chapter 7 concludes this thesis and provides an answer to the research questions.

2

Literature review

This thesis combines two branches of literature: human fatigue quantification and modelling and human factors in job-shop scheduling, which we will review separately in this section. Section 2.1 introduces and discusses ways to quantify physical fatigue, as well as the practical application of fatigue models. Section 2.2 gives an overview of the main methods of human factors consideration in JSP literature, as well as summarising common solution approaches. Finally, Section 2.3 highlights the existing research gaps in these two fields.

2.1. Fatigue and recovery models

In this section, we overview the primary methods for fatigue quantification in modelling and how these models can be used for fatigue integration in the JSP. Then, we will list the steps required for the practical application of such models.

2.1.1. Fatigue quantification

Occupational fatigue is the multidimensional loss of performance and can be classified into physiological and psychological elements. Physiological fatigue describes the reduction in muscle force generation or muscle response resulting from preceding physical exertion (performance fatigability). Psychological fatigue is the accumulated mental load that an operator experiences during a task, and it can lead to reduced cognitive performance, increased reaction times and increased error rates. The two types of fatigue are closely related and have been shown to influence each other (Lorist et al., 2000, 2002; Jongbloed et al., 2022).

Psychological fatigue is only indirectly measurable by performance in psychomotor vigilance tests or through subjective evaluations (Aryal et al., 2017; Daria et al., 2017). For physiological fatigue, the rating of perceived exertion (RPE) scale by Borg (1982) is often used to quantify subjective symptoms of fatigue. However, it is a semi-quantitative index of combined factors (physiological, psychological and situational) that are sensitive to subjective noise (Battini et al., 2016; De Souza E Silva et al., 2016; Aryal et al., 2017; Sedighi Man et al., 2017; Hajifar et al., 2021). Still, when aiming to improve employee well-being, perceived fatigue is a key performance indicator (KPI) to consider, even if it is a mix of mental, physical and environmental factors.

Given the similarities in cognitive and environmental task characteristics at the zone pick stations, quantification of physical workload and fatigue is more relevant. By doing so, we can gain valuable insights into the ways in which physical workload and fatigue can impact worker performance and identify potential solutions to improve their working conditions. There is no clear, unambiguous way to quantify a fatigue 'level', but ways exist in which we could get an indirect measure instead. In doing so, we would be looking at three different aspects of physical effort:

- **Capacity:** The energy that the neuromusculoskeletal system can exert
- **Ability:** The work that an individual can do in theory
- **Output:** The actual work output that is delivered

In this distinction of physical effort aspects, the performance fatigability would influence the transfer from ability to output. Assuming that an individual's ability is constant over time, an output performance benchmark could

be compared to a performance test in a fatigued state. This could be quantified as a task completion time or a measure of mechanical output (Jongbloed et al., 2022). A fatigue measure that has to do with performance fatigability in terms of mechanical output is the maximum endurance time.

Maximum endurance time-based fatigue models

Physical fatigue can be categorised into static and dynamic fatigue. As the name suggests, static fatigue results from holding a particular body orientation for a prolonged period, while dynamic fatigue results from movements. Although muscles do not technically do work while they are stationary, the internal (antagonistic) muscle tension required to keep a position does wear them out over time. The maximum time a muscle can exert a static force level is called the maximum endurance time (MET). Once this moment is reached, the respective muscles are at maximum fatigue, and the orientation cannot be held any longer. The MET is often calculated as a function of the percentage of maximum voluntary contraction (%MVC or fMVC). This is the maximum force the muscles can exert (El Ahrache et al., 2006; Jaber and Neumann, 2010; Daria et al., 2017).

Many models exist to calculate the MET, with predominantly linear and power forms (Jaber and Neumann, 2010). An well-known example is $MET = \beta_0 \times e^{-\beta f_{MVC}}$ from Rose et al. (1992). β and β_0 are MET calculation parameters equal to 7.96 and 4.16, respectively. According to El Ahrache et al. (2006), there is no clear consensus or link to detailed experimental data in literature describing such models. This hinders the use of MET models in practical applications like the JSP. The authors identify four causal factors:

- Some studies use subjective scales, unrepresentative groups of test subjects or secondary data sets, leading to different results. Most models do acknowledge the negative exponential tendency of the MET-%MVC relationship.
- There is no consensus on whether there is fatigue development below a certain %MVC (the endurance limit). While some studies identify an endurance limit of 15%, others indicate a lower value or no such effect at all.
- There are studies indicating that the general MET model is not valid for all muscle groups, while many studies assume this. This can lead to the model being invalid for tasks that require different muscle groups than those it has been calibrated for.
- Not all studies take into account that the MET-%MVC relation can vary between individuals. The β and β_0 parameters from Rose et al. (1992) could be different for each person.

The authors develop a table of MET percentile values from a large number of models to allow for MET estimation among a group of people. Although this gives better MET estimation for a population, this does not solve one of the fundamental problems of the MET:

1. The MET alone has no real value in a scheduling context other than setting a limit on the time-on-task. Even if full MET is used, there is no notion of the required resting time.
2. As stated by Jaber and Neumann (2010), the MET gives no information about the shape of the fatigue development function between $t = 0$ and $t = MET$. This knowledge would be a good first step towards using it in a scheduling context.
3. Because the MVC is different for each muscle (pair), the MET determination will automatically be muscle-specific rather than full-body. In some settings, this does not pose an immediate problem because the task concerns few muscles (e.g., sawing motion). However, in jobs comprising multiple sub-tasks with different muscles, MET estimation is laborious. This is amplified by individual differences, especially when dealing with a large employee pool.
4. MVC is measured in a static load scenario and, therefore, only valid for static exertion. This is not a valid assumption for most physical tasks, and we need to look at dynamic fatigue.

Addressing the flaws of maximum endurance time fatigue models

In the work by Abdous et al. (2018), a mathematical model is formulated for an assembly line scheduling problem with fatigue incorporated. They then try to solve the problem with a dual objective function minimising the number of workstations and the fatigue level, which is defined as:

$$F_{cem_{V'}} = MVC e^{-C \sum_{j \in V'} I_j} \quad (2.1)$$

, with $F_{cem_{V'}}$ the current muscle capacity after set of tasks V' , C a parameter that represents the worker's factors and I_j the load integral of task j . The exponential nature of the model was "theoretically validated

and compared with existing static endurance time models” and deemed precise for fatigue evaluation in static and dynamic tasks. The fatigue 'level' is then defined as the ratio between MVC and $F_{cem_{V'}}$, making its unit Newtons although it could also be expressed as a percentage¹.

Another formulation focused on finding a fatigue level is given by Konz (1998), who state that the fatigue and recovery processes take the following form (formulation from Jaber et al. (2013)):

$$\begin{aligned} F(t) &= 1 - e^{-\lambda t} \\ R(\tau_i) &= F(t)e^{-\mu\tau_i} \end{aligned} \quad (2.2)$$

, with $F(t)$ the (fractional) fatigue level at time t and $R(\tau_i)$ the fatigue level after a break of length τ_i in cycle i . λ and μ are the fatigue and recovery parameters, respectively. We can see a very similar formula to (2.1), but now the fatigue level is presented as a fraction, and λ replaces the MVC value², parameter C and summation of the load integrals I_j simultaneously. This means λ varies with the worker and task, whereas parameter μ only varies with the worker.

If a break is insufficient to allow $R = 0$ (full recovery), the residual fatigue level in the next cycle ($i + 1$) can be calculated using:

$$F_{i+1}(t) = R(\tau_i) + (1 - R(\tau_i)) \left(1 - e^{-\lambda(t_n - t_i)}\right) \quad (2.3)$$

, with t_n the length or production time of cycle i and t_i to be determined using:

$$t_i = -\ln(1 - R(\tau_i)) / \lambda \quad (2.4)$$

The fatigue-recovery effect on the accumulated fatigue level is shown in Figure 2.1.

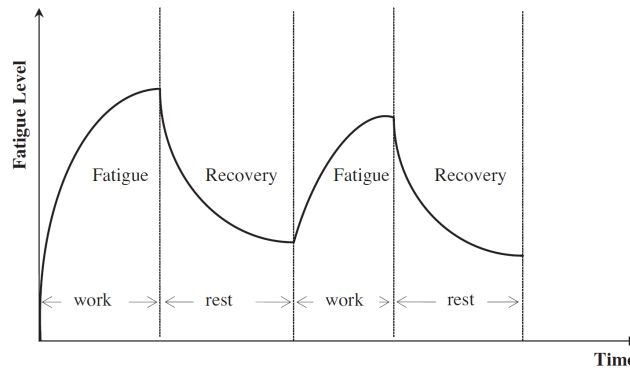


Figure 2.1: The fatigue level under the fatigue-recovery effect (Jaber et al., 2013).

Although these fatigue models look promising, they do not specify the required resting time per task and still require the MVC and other personal parameters to be known for each employee. This impracticality can be challenged by introducing the term rest or relaxation allowance (RA). The RA is the fraction of the time a static muscle position is held that is needed on break to recover from the load. Like the MET, there exists a multitude of methods for RA calculation. One critical factor that El Ahrache et al. (2006) has already mentioned is the debate around low %MVC values and their influence on the fatigue level and, consequently, the required recovery time. This same debate can be seen in RA calculation literature. An overview of existing models by Daria et al. (2017) states some measure of the MET and %MVC is generally present, but the available models still vary considerably in shape and form (El ahrache and Imbeau, 2009). On top of that, the models generally only hold for very simple, static scenarios.

An example of such a simplification in application is given by Jaber and Neumann (2010), who use a mathematical formulation to describe the scheduling problem. The RA estimations here are based on an assumption of the MET and the equation $RA = 3 \cdot MET^{-0.41479}$. This model is then used to solve the scheduling problem for multiple policies for break scheduling. Although the fatigue level can be known from this model at any moment, there is no distinction in RA or MET for the different tasks in the scheduling problem. Another example

¹In that case, it could be calculated by $\%F = \frac{1 - F_{cem_{V'}}}{MVC} = \frac{1}{MVC} - e^{-C \sum_{j \in V'} I_j}$

²Note that these are still related through $fMVC = i\lambda$

is the work by [Fruggiero et al. \(2016\)](#), who use an agent-based model to simulate the effects of different job switching (scheduling) policies using the MET-based RA principle. They identify differences per job but do not consider individual differences and only model static effort for one muscle group at a time.

As we can see, the first two flaws of MET-based fatigue and recovery models can be addressed. However, the inherent problems regarding fatigue evaluation for full-body and dynamic tasks still stand.

Energy expenditure-based fatigue models

This shortcoming is mitigated in the work by [Price \(1990\)](#), where the rest allowance is a function of the mean work rate (MWR). The MWR is the average (full-body) energy expenditure (EE) rate over the length of a task or set of tasks, measured in Watts. However, the term energy expenditure is much more widely used in the relevant literature, where EE_{task} denotes the energy expenditure for a specific task. The acronym 'EE' is used for both the energy expenditure [kcal] and energy expenditure rate [kcal/min] (= MWR) in literature, but for clarity, we will only use the first definition. The energy expenditure rate will be written as $\dot{E}E$.

For any shift made up of several different tasks, the RA can be calculated by summing the RAs for the task elements. If we calculate the RA through the $\dot{E}E$ per task, this means that our input is a piece-wise function of the $\dot{E}E$ over time. The RA percentage can then be calculated using the following formula:

$$RA = \frac{MWR - AWL}{AWL - RR} \quad (2.5)$$

, where all elements on the right side are given in Watts, AWL is the acceptable work level (300 W), and RR is the relaxation rate (130W when standing, 105W when seated). [Finco et al. \(2021\)](#) use an adaptation (in kcal/min) on the formula from [Price \(1990\)](#), written as:

$$RA = \max \left\{ 0; \frac{\dot{E}E - MAEE}{MAEE - \dot{E}E_R} \right\} \quad (2.6)$$

, with $\dot{E}E$ the mean work rate during a period, $\dot{E}E_R$ the energy expenditure rate during rest and $MAEE$ the maximum allowable energy expenditure rate (= AWL), which is the maximum level an individual can work without fatigue effects. Figure 2.2 shows what this relationship would look like in practice.

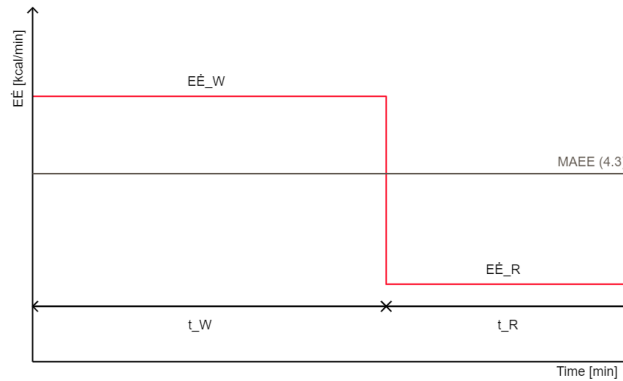


Figure 2.2: Graph relating $\dot{E}E$ and RA, with $\dot{E}E_W$ denoting the energy expenditure rate during the work period, t_W the working time and t_R the resting time.

[Battini et al. \(2016\)](#) have yet another, but functionally equivalent form:

$$RA = \frac{41 \cdot \dot{E}E_o - 176}{100} \quad \text{if } \dot{E}E_o > 4.3 \frac{\text{kcal}}{\text{min}}; \quad 0 \quad \text{otherwise} \quad (2.7)$$

, where $\dot{E}E_o$ denotes the average energy expenditure rate for operation o . Only if this is higher than the maximum allowable energy expenditure rate ($MAEE = 4.3$ kcal/min), RA is needed. The inequality condition on the right-hand side is no different from the $\max\{0; \dots\}$ formulation from (2.6). By filling in the values for $\dot{E}E_R$ and $MAEE$, [Battini et al. \(2016\)](#) have simplified the formula but do not account for interpersonal differences in the values for $\dot{E}E_R$ and $MAEE$.

MAEE

However, the $MAEE$ is dependent on sex, age, body weight and height, so it would show inherent differences between different people. [Berti et al. \(2021\)](#) use different values for the $MAEE$ to account for the characteristics of employees of different age. Their input values come from a simplification of the formulae by [De Souza E Silva et al. \(2016\)](#), who give the $MAEE$ as:

$$\begin{aligned} MAEE &= 0.0016 ((60 - 0.55 \cdot AGE) \cdot BW) && \text{Men} \\ &= 0.0016 ((48 - 0.37 \cdot AGE) \cdot BW) && \text{Women} \end{aligned} \quad (2.8)$$

, with BW the body weight of the individual. When considering a realistic, mixed workforce, it makes sense to evaluate the $MAEE$ as accurately as possible. This requires including all of these factors in the estimation.

$\dot{E}E_R$

The energy expenditure rate during rest is also not a constant value. [Price \(1990\)](#) differentiates standing and seated rest rates (130W and 105W or 1.86kcal/min and 1.50kcal/min, respectively), but the work by [Garg et al. \(1978\)](#) goes one step further to also include the body weight BW :

$$\begin{aligned} \text{Sitting: } \dot{E}E_R &= 0.023BW \\ \text{Standing: } \dot{E}E_R &= 0.024BW \\ \text{Standing, bent position: } \dot{E}E_R &= 0.028BW \end{aligned} \quad (2.9)$$

This, again, can make considerable differences when dealing with a mixed workforce.

Non-linear shape of $\dot{E}E$

The RA model from Price treats the $\dot{E}E$ function as a piece-wise step function (see [Figure 2.2](#)), meaning the momentary energy expenditure of an individual immediately jumps to the new situation. This notion is challenged by [Calzavara et al. \(2019\)](#), who state that the energy expenditure rates do not change instantly between periods of work and rest but follow a nonlinear shape.

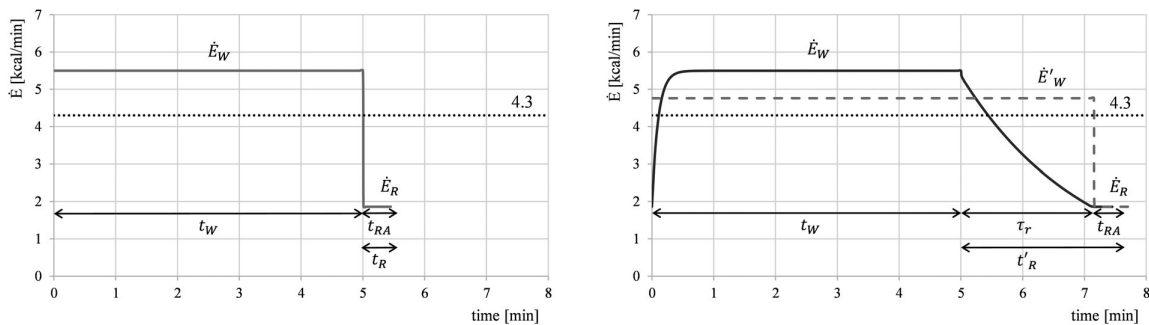


Figure 2.3: Difference in RA calculation between the two energy expenditure rate models ([Calzavara et al., 2019](#)).

A comparison between RA calculation using the normal $\dot{E}E$ (written as \dot{E} in the figure) curve and the nonlinear curve can be seen in [Figure 2.3](#). New in the right figure is variable τ_r , which is the time it takes to go from $\dot{E}E_W$ to $\dot{E}E_R$ and can be calculated using:

$$\tau_r = \frac{\ln F(t_W) - \ln \dot{E}E_R}{\mu} \quad (2.10)$$

, with μ the recovery rate of the individual and $F()$ the (exponential) fatigue function. As can be seen from the figure, the real full recovery time t'_R in the exponential energy expenditure rate function is much longer than the recovery time t_R in the constant shape function. Also, there is a slight decrease in the fatigue buildup rate at the start of the working period. For a practical application, we would need to know the μ values for each employee before being able to calculate the fatigue and recovery functions. In addition, we need to calculate the integral of the fatigue and recovery functions to get to the actual rest allowance. Although this method is more accurate than the one from [Price \(1990\)](#), it requires data and personalised information that are not easily gathered.

2.1.2. Practical application

Now that we have seen different methods to model fatigue in a scheduling context, we will look into their suitability for practical application. For MET-based fatigue models, a practical application would require MET value measurement and %MVC determination per required muscle group for all employees and movements/positions needed to execute a task. While their model solves some of the problems mentioned by [El Ahrache et al. \(2006\)](#), [Jaber et al. \(2013\)](#) still require many measurements to determine λ and μ . Even when the measurement of these inputs would be possible, they may differ due to a combination of factors and need to be repeated multiple times to get a good average estimation. This makes practical application on a large scale considerably more difficult.

Although EE-based fatigue models rarely output a fatigue level - which would not translate to the definition of performance fatigability from Section 2.1.1 - they allow for integration in a scheduling algorithm without the flaws common for MET-based models. Another significant advantage is that they require no data gathering on fatigue or recovery rates for each individual but can work with personal characteristics ($MAEE$ and $\dot{E}E_R$) as the model input. One crucial factor missing in the RA calculation so far, however, is the determination of $\dot{E}E$. Two main methods exist for estimating a task's energy expenditure: bodily measurements or mathematical task modelling.

Finding $\dot{E}E$ through bodily measurements

Localised muscle fatigue can most directly be evaluated using electromyography (EMG), where electrodes placed on the skin measure the action potential of muscles. This action potential gives information about the muscle contraction and, consecutively, the MVC fraction of that muscle. This method directly relates to the difference between ability and output from Section 2.1.1, but the measurement is muscle-specific and does not hold information about full-body fatigue, as was the case for MET-based fatigue models. In addition, EMG measurements are subjective to (electrical) sensor noise and muscle crosstalk because the sensors are placed on the skin above a muscle rather than on the muscle itself. Furthermore, the sensor placement on the body could influence the work in a real-life setting and would be costly when applied in a working environment ([Steinebach et al., 2021](#); [Battini et al., 2020](#)).

Another fatigue evaluation approach is through machine learning methods, analysing raw bodily measurements to predict RPE scores. The work by [Hajifar et al. \(2021\)](#) focuses on predicting the perceived fatigue of workers in a material handling environment. Their work uses wearable gait sensors and time series methods to predict the RPE scores of workers in different settings. Although their results look promising in detecting fatigue, the model outputs are limited in their ability to predict fatigue in the longer term and cannot easily be linked to the fatigue models we have seen in Section 2.1.1.

Another work aiming to detect perceived fatigue directly is [Aryal et al. \(2017\)](#), which monitors fatigue indicators in construction work using a variety of sensors. The physiological measurements in their study are used in a boosted tree classifier that differentiates between low, medium, high and very high RPE scores for the four classes. Using a boosted tree classifier entails the requirement for a training phase with RPE questionnaires, which is out of the scope for a model validation step. However, some of their physiological measurements can be found in works that describe fatigue model validation methods.

For EE-based fatigue models, the most accurate way of measuring the $\dot{E}E$ in real-time is through measurement of the oxygen uptake. A method called indirect calorimetry requires measuring respiratory gas exchange rates to calculate oxygen uptake (VO_2). According to [Hackney \(2016\)](#), the relation between the oxygen uptake and energy expenditure rate is a linear one (see Figure 2.4). Although VO_2 sensors have become much more easily wearable over the years, the process of VO_2 measurement is still an inconvenience outside of a research lab setting. For practical applications, therefore, one might better use the relation between VO_2 and heart rate as these have demonstrated a linear relationship with a correlation coefficient between 0.7 and 0.8 ([Calzavara et al., 2018b](#); [Daria et al., 2017](#)).

Using the HR measurements, [Li et al. \(1993\)](#) states that the energy expenditure rate can be estimated using:

$$EE = (m \cdot HR + b) \cdot 20.48 \quad (2.11)$$

, where m and b are the slope and initial value of the HR- VO_2 relationship curve, respectively. These parameters would be different for each individual and can only be determined through experiments. When implementing this method for EE measurement, this experiment needs to be performed separately for each individual

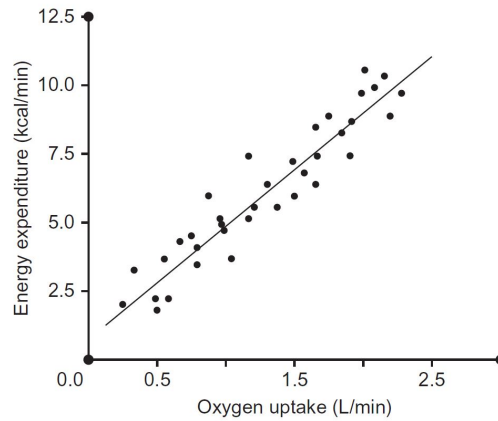


Figure 2.4: Relation between an individual's energy expenditure rate and the oxygen uptake (Hackney, 2016).

and task. For the same task but different task intensity, the existing relationship holds as well (Calzavara et al., 2019). This still burdens the individuals for whom EE estimation is proposed, as VO₂ tests are cumbersome, demanding of the participant, and time-consuming.

Recent work by Battini et al. (2020) suggests that the VO₂ calibration step might not be necessary when using HR as a measure of energy expenditure. The wide availability and ease of use of HR sensors could make up for the decrease in estimation accuracy compared to VO₂ measurements. Previously, de la Riva et al. (2015) have shown to be able to match the RA predictions from existing works when using RA lookup tables and HR measurements for a dynamic task. More recent methods have focused on the percentage of heart rate reserve (%HRR) that relates the momentary HR to the maximum HR (HR_{max}) and resting HR (HR_{rest}) as follows:

$$\%HRR = \frac{HR - HR_{rest}}{HR_{max} - HR_{rest}} \quad (2.12)$$

, where HR_{max} can be calculated using $HR_{max} = 208 - 0.7 \cdot AGE$ (Tanaka et al., 2001). According to Calzavara et al. (2018b), the %HRR indicates the relation between aerobic and anaerobic conditions for individuals, where 80 %HRR indicates full anaerobic condition. This condition can only be held for a limited period, as it leads to lactic acid buildup. Using both HR_{max} and HR_{rest} in the calculation of %HRR normalises the value, thereby making it a personalised measure; a heart rate of 150 beats per minute could indicate different %HRR between individuals. In a fatigue classification study on a supply pick and insertion task by Sedighi Maman et al. (2020), features related to %HRR (coefficient of variation and mean) have the highest appearance in the models. However, due to the nature of their classifiers, no specific function is formulated to relate the feature values to the resulting fatigue state. Also, the classifier distinguishes two fatigue states: fatigued and not fatigued; this is not directly usable for $\dot{E}E$ estimation. Still, this is a promising result for the validity of using %HRR as a potential input for $\dot{E}E$ estimation.

A study by Hiilloskorpi et al. (2003) presents and evaluates an $\dot{E}E$ estimation method using the %HRR, sex and body weight (BW) as follows:

$$\begin{aligned} \dot{E}E &= 0.744 + 0.0265 \cdot HRR + 0.00699 \cdot BW + 0.00102 \cdot HRR \cdot BW && \text{Women, light activity} \\ \dot{E}E &= 0.165 + 0.0688 \cdot HRR + 0.02666 \cdot BW + 0.00050 \cdot HRR \cdot BW && \text{Women, heavy activity} \\ \dot{E}E &= 0.449 + 0.0627 \cdot HRR + 0.00743 \cdot BW + 0.00100 \cdot HRR \cdot BW && \text{Men, light activity} \\ \dot{E}E &= 1.044 + 0.0250 \cdot HRR + 0.01088 \cdot BW + 0.00177 \cdot HRR \cdot BW && \text{Men, heavy activity} \end{aligned} \quad (2.13)$$

This would allow for $\dot{E}E$ estimation for all participants given the HR data and the body weight. One must note that the standard error of the estimate was 1.01 kcal/min for this model, which is relatively high considering that the EE in the experimental results in Figure 2.4 go up to about 11 kcal/min. However, this method is much easier to apply than the more accurate estimation using VO₂ measurements.

Finding \bar{EE} through mathematical task modelling

A mathematical approach for estimating the energy expenditure of a picking task is the formula presented by Calzavara et al. (2018a):

$$EE_i = \frac{EE_{p_i} \cdot Q_i + EE_{fix} \cdot Z_i}{Z_i} \quad (2.14)$$

, with EE_i the total energy expenditure required for picking item i , EE_{p_i} the energy expenditure required for the picking task per picked carton, Q_i the number of cartons to be picked for item i , EE_{fix} the fixed energy expenditure for the picking round and Z_i the number of times item i shows up in orders.

To calculate the energy expenditure of a task, we need to find EE_{p_i} and EE_{fix} . Calculating EE_i would also require the picking time. Another approach is the use of the metabolic equivalent of a task. This unit, defined by Álvarez-García et al. (2020), is expressed in $[\frac{kcal}{kg \cdot h}]$ and represents the energy consumed by a person per kilogram of body weight. Through a series of experiments, they find the values for multiple task scenarios, e.g. office work, shovelling snow or running.

A less detailed but more holistic approach is to relate the activity intensity to the basal metabolic rate (BMR) of a person (Hackney, 2016). The BMR is expressed in kcal energy intake per day. For example, the BMR of a man with body weight BW , height H and age AGE can be calculated by the Harris-Benedict equation:

$$BMR = 88.362 + 13.397BW + 4.799H - 5.677AGE \quad (2.15)$$

The required kcal intake for a moderate exercise type of physical activity is $1.55 \cdot BMR$. The inclusion of height and age in the BMR is rarely seen in the estimation of \bar{EE}_R , although the two concepts overlap considerably. Still, this method does not allow for detailed estimation of \bar{EE} and has no real value for RA estimation.

A widely recognised EE estimation method is proposed by Garg et al. (1978), who describe a method for task decomposition into positional elements. Their basic formula is:

$$\bar{EE}_{job} = \frac{\sum_{pos \in n_{pos}} \dot{EE}_{pos} x t_{pos} + \sum_{i \in n} \Delta EE_i}{T} \quad (2.16)$$

, with \bar{EE}_{job} the average EE rate [kcal/min] of the job, n_{pos} the total number of body postures during the job, \dot{EE}_{pos} the EE rate [kcal/min] of the posture, t_{pos} the duration [min] of the posture, n the number of tasks i in the job, ΔEE_i the net EE [kcal] of task i and T the total time duration [min] of the job. In their appendix, the writers then set up formulae describing the \bar{EE} of different task elements that are used by a multitude of authors for the RA estimation (see Figure 2.5) (Battini et al., 2016; Calzavara et al., 2018a, 2019; Berti et al., 2021).

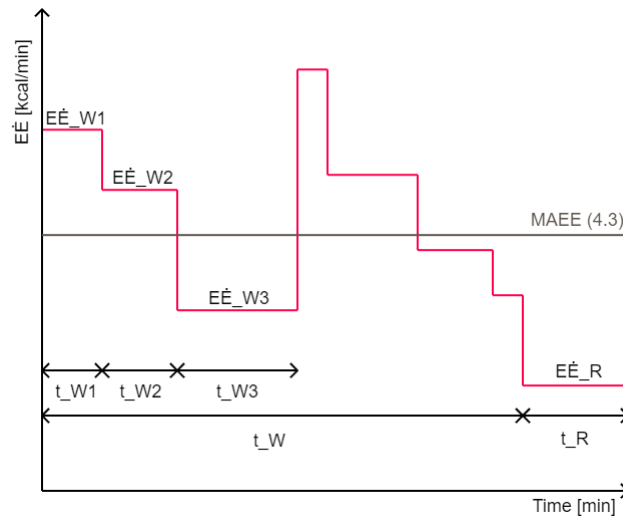


Figure 2.5: Graph relating \bar{EE} and RA with task decomposition, with $\dot{EE}_{W(x)}$ denoting the energy expenditure rate during work period x , $t_{W(x)}$ the working time during that period, \dot{EE}_R the resting energy expenditure rate and t_R the resting time.

Although task decomposition is the most comprehensive method for energy expenditure rate estimation, this level of detail is a major flaw too. When applying the method in real life, one must evaluate every individual movement of an employee during their work. For varying, complex tasks, this can be very time-consuming, and this method cannot monitor the real-time \dot{E} of an employee (Battini et al., 2020).

2.2. Human factors in job-shop scheduling

Since we have already established the DRC nature of Picnic's OP system in Section 1, we will dive deeper into DRC-related JSP literature. The work by Xu et al. (2011) gives a summary of the different challenges that are addressed in the existing literature. They come to five categories: job release, job dispatching, worker flexibility, worker assignment and transfer costs. Following the research motivation and problem statement in chapter 1.4, our focus will be on the worker assignment category.

In their work, Xu et al. identify two aspects of worker assignment: *when* and *where* rules. *When* rules concern the start and end time of a worker's shift at a (work)station and thus also dictate the timing of transfer between stations. *Where* rules then determine where the transfer will have the worker assigned. The authors state that the *when* rules "have a more significant effect on overall system performance" than the *where* rules, which is why the latter is often considered after the former. However, this assessment overlooks the *who* rules, which play a crucial role in systems where worker heterogeneity is considered (Thürer et al., 2020). When looking at real-life DRC systems, the practical machine scheduling problems that arise will have at least some worker heterogeneity, so solving these will require decision-making on all three aspects instead of only those mentioned by Xu et al. (2011).

2.2.1. Job-shop scheduling problems

These three aspects come together in the job-shop scheduling problem of DRC systems. In the JSP, there is a set M of m machines and a set J of n jobs consisting of a sequence of N operations O to be processed. For the operations within one job, precedence relations dictate the order in which these need to be processed. Also, each operation i can be processed only on a subset of machines (M_i) and has a specific processing time p_i . In the JSP, the decision variable is the starting time of each operation t_i , to be optimised for minimal makespan C_{max} (Šeda, 2007):

$$\min C_{max} = \max_{i \in O} (t_i + p_i) \quad (2.17a)$$

$$\text{s.t. } t_i \geq 0, \quad \forall i \in O, \quad (2.17b)$$

$$t_j \geq t_i + p_i, \quad \forall i, j \in O, i \rightarrow j, \quad (2.17c)$$

$$(t_j \geq t_i + p_i) \vee (t_j \geq t_i + p_j), \quad \forall i, j \in O, i \neq j, M_i = M_j. \quad (2.17d)$$

(2.17a) shows the objective to minimise the makespan, defined as the maximum of the starting + processing time for all operations i . This starting time is strictly positive, as ensured by (2.17b). (2.17c) shows the constraint from the precedence relation between operations i and j , which is indicated by $i \rightarrow j$. Finally, (2.17d) ensures that no unique operations i and j can be processed on the same machine at the same time.

The JSP is inherently different from the flow-shop scheduling problem (FSP), another common scheduling problem in operations research (Šeda, 2007). In this case, there are m machines in set M and an equally large set of operations O ($1, \dots, m$). All jobs J are made up of the same operations with a set order and thus do not impose precedence constraints on the FSP. Flow-shops produce identical or very similar products, such as electronics or car manufacturing lines. Even though many different variants of the FSP exist, some even introducing worker heterogeneity Benavides et al. (2014), the FSP is strictly a sequencing problem with low product variety and a fixed production structure. An OP system, therefore, cannot be described as a flow-shop because orders can be made up of various items.

The JSP applies to a more versatile production system where production cycles differ between products. Shortcomings of the current formulation, however, are numerous. For example, this is not a closed formulation and therefore cannot be used to generate a schedule. This would require the reformulation of (2.17c) and (2.17d). Also, this formulation does not include many constraints relevant to a real-life operation, such as worker scheduling, machine switching and non-predetermined precedence relations. If we would include these constraints, the problem formulation would require a combination of continuous and discrete (binary) decision variables, making it a mixed integer programming problem (MIP). In fact, most JSP formulations in

literature use only linear constraints and objectives and are therefore mixed integer linear programming problems (MILP).

Because systems the JSP applies to are so widespread, we find a variety of JSP in literature. A survey by [Xiong et al. \(2022\)](#) lists the common assumptions in JSP models and then gives an overview of the common sub-types from this basic model. Examples of assumptions are:

- **Jobs**
 1. All jobs are released at time zero
 2. The set of jobs is predetermined, static and known
 3. Jobs do not have due dates or priority among each other
 4. The sequence of operations in a job is predetermined
 5. Jobs are independent of one another
- **Operations**
 1. Operations have a unique and predetermined processing time
 2. Operations are pre-assigned to only one machine
 3. Operations can only be processed by one machine at a time
 4. Operations must be completed in one stint
- **Machines**
 1. All machines are available at time zero
 2. Machines can only process one operation at a time
 3. Machines are independent of one another
 4. Machines are always available
- **Workers**
 1. All workers are available at time zero
 2. No worker constraints are considered
- Buffer, warehouse or supply capacity is unlimited

Evident from these assumptions is that not all hold in a real-life JSP, depending on the system. The common basic sub-types of JSP are variants of the basic formulation with one or some assumptions altered. For example, there is the dynamic JSP, which does not assume all attributes of the system or inputs are unchanged during the scheduling process, thereby altering job assumptions 2 and 4, among others. Another sub-type is the flexible JSP, which would add flexibility to jobs assumption 4 and operations assumption 2. Furthermore, regarding HF, one would consider JSP with nonconstant or nondeterministic processing times (operations assumption 1) or with dual-resource constraints implemented (workers assumption 2). Still, the general formulation of these JSP sub-types would be similar to the one in (2.17). Although the formulation may look simple because it contains few constraints, the JSP is "one of the hardest combinatorial optimisation problems" ([Šeda, 2007](#)). The JSP is known to be NP-hard, as no algorithms exist that could solve it optimally in polynomial time. In practice, this means that exact solving methods could solve only small problem instances. Therefore, a significant challenge addressed in JSP literature is to find an appropriate solving method for JSPs (see Section 2.2.3).

However, as identified in the literature survey by [Xiong et al. \(2022\)](#), there are more "hot topics" in JSP research. This includes some topics that have become relevant because of changing markets and technology (e.g. pro-environment and energy-saving optimisation, distributed manufacturing, AGV transport between machines), but also includes efforts to make the JSP represent reality as closely as possible. Each assumption in the list presented above is an inaccuracy when comparing the model to the real world. The more of these assumptions are relaxed in the JSP formulation, the more complex the problem becomes and the closer it comes to a real-world scenario. The same goes for the optimisation objective. In a real-life production system, there is rarely only the makespan to optimise, but the objective is a combination of makespan, costs, lateness or others. However, most JSPs in literature concern themselves with only a single objective. Multi-objective optimisation (MOO) is currently finding its way into JSP literature to address this mismatch between literature and reality.

When implementing human factors in the JSP, the relaxation of multiple assumptions and MOO are both relevant. Although this topic has not been addressed explicitly by [Xiong et al. \(2022\)](#), there has been an increasing volume of literature published in recent years.

2.2.2. Implementation of fatigue in the job-shop scheduling problem

Generally, there exist two main approaches to implement human fatigue models in scheduling problems. We will introduce the two methods to implement human factors into scheduling problems in general, followed by an overview of the most prominent works in Table 2.1.

Human factors focus

First, there is the HF modelling-focused approach as taken by [Jaber and Bonney \(1996\)](#); [Jaber et al. \(2013\)](#); [Givi et al. \(2015b\)](#); [Fruggiero et al. \(2016\)](#). In these works, the starting point of the problem formulation is a detailed representation of human learning or fatiguing. This approach's result is that the performance indicators in the mathematical model concern human characteristics such as fatigue level or experience-dependent production time.

As learning-forgetting and fatigue-recovery models have been represented far better than other HF mechanisms ([Grosse et al., 2017](#)), a combined model of the two is proposed by [Jaber et al. \(2013\)](#). This human factors modelling-focused paper combines the work by [Jaber and Bonney \(1996\)](#) and [Konz \(1998\)](#) into the learning-forgetting-fatigue-recovery (LFFR) model. Due to the complexity of these calculations, the authors simplify the manufacturing process to be a single-machine, single-worker system, and the scheduling problem concerns break and production batch size. Although their calculation of the processing time has great detail, incorporating prior experience and personal characteristics like the learning exponent and history-dependent forgetting exponent, it requires the recursive calculation of the remembered experience and adds nonlinearity to the mathematical formulation. The same is true for the constraints on the production batch size and break time, where recursive calculation of the fatigue level is required. In a nutshell, this method imposes a different calculation of the processing time of items and human factors constraints on the scheduling problem.

The constraint-based human factors inclusion by [Jaber et al. \(2013\)](#) allows for a highly customised calculation of learning and fatiguing for different individuals. However, their mathematical model is no example of a JSP formulation. The work by [Givi et al. \(2015b\)](#) takes a similar approach while introducing fatigue-dependent productivity, human error, and more system constraints to the scheduling problem in a VBA program but still considers a DRC flow-shop. The same is true for [Fruggiero et al. \(2016\)](#), who set up a multi-agent model architecture to generate scheduling for a DRC flow shop.

No work on JSP solving using this level of detail in the HF models has been published to our knowledge. There are, however, works that include HF in a JSP formulation in a different way.

Scheduling focus

In this type of mathematical model, the starting point is a scheduling formulation as we have seen in (2.17) and constraints related to human factors are introduced. [Mosheiov and Sidney \(2003\)](#) have implemented job-dependent learning curves in, among others, an unrelated parallel machine scheduling problem. They do not perform a real-life application of this approach but instead focus on estimating the solving time for a small example problem. [Cheng et al. \(2008\)](#) add a 'deteriorating' effect to this scheduling problem and apply it to a flow-shop scheduling problem. The paper then focuses on the mathematical proof for the solution when the processing time is station-independent.

A more inclusive work by [Othman et al. \(2012\)](#) combines learning, fatigue, recovery and motivation in a virtual job-shop scheduling problem. They impose MET-based maximum fatigue load constraints on a JSP formulation with a multi-objective optimisation (MOO) goal, including hiring, training, skill level, overtime, and salary/cost. The overall fatigue accumulation level is one of the optimisation sub-goals. The authors then make many simplifications in the (virtual) application of the HF models, resulting in an unrealistic version of the scheduling problem. For example, they only consider jobs without operations, thereby making it seem more like an FSP. However, to our knowledge, they are the first to suggest multiple objectives for JSP optimisation.

[Mouayni et al. \(2019\)](#) introduce a fatigue level using a combined fatigue and recovery function:

$$F_{t+1} = w_t \cdot (1 - (1 - F_t) \cdot e^{-\lambda_i}) + (1 - w_t) \cdot F_t \cdot e^{-\mu_{i'}} \quad (2.18)$$

, where F_t is the fatigue fraction ($0 \leq F < 1$) at the end of period t , w_t a binary variable indicating the worker state (1 is busy, 0 is idle), λ_i the fatigue rate when processing job i and $\mu_{i'}$ the recovery rate after processing job i' . This fatigue fraction estimates the human error probability in a simple JSP formulation with job earliness and tardiness. This is one of the very few works where a fatigue fraction or level is calculated in a paper with

a scheduling focus. However, doing so also requires recursive calculation of this fatigue level.

The work by [Yazdani et al. \(2015\)](#) introduces a worker-dependent processing time in a DRC flexible JSP. This changes p_i in (2.17c) to $p_{i,j,k,l}$, for job i , operation j , machine k and worker l . [Battini et al. \(2016\)](#) dive into fatigue quantification using the principle of energy expenditure and review its usability in a storage location assignment problem where both makespan and fatigue alleviation are considered. Their approach looks promising for an extension towards scheduling but requires real-life verification before application.

[Finco et al. \(2021\)](#) consider two different decision levels: tactical, concerning line balancing, and operational, concerning line sequencing. In this article, the RA calculation by [Calzavara et al. \(2019\)](#) is linearised to ensure the mixed-model assembly line balancing problem can be solved linearly. The assumption here still is built around the acceptable work level of energy expenditure (or $MAEE$ in the article) being 4.3 kcal/min, even though the authors debated exactly this earlier in their work. However, they deem this average sufficient for the line balancing problem and opt for the individual differences of the $MAEE$ to be taken into account in the line sequencing problem. The exponential shape of the energy expenditure function from [Calzavara et al. \(2019\)](#) is not considered in this model, so the calculation of \dot{E}'_W is based on the constant shape of [Price \(1990\)](#).

What sets this work apart is the split in their approach, first laying down the granular groundwork for the line sequencing problem to optimise the short-term planning with individual characteristics of workers taken into account. The second problem is approached with heuristics. Using this method, workers with the highest $MAEE$ values - those who can endure the highest physical effort - are assigned to the most intensive workstations/tasks. All in all, this leads to the least total fatigue and, consequently, the lowest overall RA. Unfortunately, the model lacks some realism because, like [Calzavara et al. \(2019\)](#), there is still the assumption that the whole RA period is scheduled after each task, which is not always feasible in actual work planning. Also, as far as optimisation is concerned, the two-step approach does not allow for global optimisation but will only find piece-wise optimality.

Finally, [Berti et al. \(2021\)](#) incorporate fatiguing, heterogeneous workforce attributes (different $MAEE$) and three different RA scheduling methods into a heuristic for DRC JSPs. The rest allowance is included in the recovery time parameter $r_{i,j,k,l}$ that the authors calculate using the RA model by [Price \(1990\)](#) which we have seen in Section 2.1.1.

This paper's approach allows for differentiation between operations and jobs regarding the fatiguing process. Since the individual, heterogeneous fatiguing process is captured in the recovery time parameter, and there is no notion of workload or fatigue level, the paper does not allow companies to apply the heuristics to their JSP with specific task and workforce characteristics. If a workload or fatigue measure were present, further development could also include fatigue-dependent productivity, as this has been shown to have a negative correlation ([Sgarbossa et al., 2020](#); [Calzavara et al., 2019](#); [Dode et al., 2016](#); [Lorist et al., 2002](#); [Fruggiero et al., 2016](#)).

In Table 2.1, we present an overview of prominent works regarding the implementation of HF in scheduling problems. These works are characterised by the HF modelling approaches presented in this section, as well as key properties relating to our system characteristics and research goals proposed in 1.4. As can be seen, none of the existing works focuses on all key properties that we value for this study.

Article	SF	OPF	JSP	WTHE	AV
Othman et al. (2012)	✓	✓	✓	✗	✗
Jaber et al. (2013)	✗	✗	✗	✓	✗
Battini et al. (2016)	✓	✓	✗	✗	✗
Calzavara et al. (2019)	✓	✗	✗	✓	✗
Finco et al. (2021)	✓	✓	✗	✗	✓
Berti et al. (2021)	✓	✗	✓	✓	✓
Our Study	✓	✓	✓	✓	✓

Table 2.1: Literature table of applicative HF in scheduling studies. SF=Scheduling Focus, OPF=Optimise Productivity and Fatigue, JSP=Job-Shop Scheduling Problem, WTHE=Worker and Task Heterogeneity, AV=Applicative Validation.

2.2.3. Solving JSP with human factors

We have seen multiple types of JSP and two ways to approach HF integration, but one thing all these different JSP formulations have in common is the difficulty of finding an optimal solution. As mentioned in Section 2.2.1, the JSP is NP-hard in the basic form, and the computational complexity grows exponentially with increasing problem size (Hoitomt et al., 1993). Adding more parameters, factors and details will only add to this computational complexity, and analytical solution methods may no longer be feasible even for small problem sizes (Xu et al., 2011). In search for a solution method that suits DRC JSP with larger problem sizes, we give an overview of the approach categories that exist in literature (Sellers, 1996; Xu et al., 2011). We highlight some notable works in these categories. A more comprehensive overview can be found in the works by Xiong et al. (2022) and Xu et al. (2011).

Classical optimisation methods

As categorised by Sellers (1996), classical optimisation methods can find optimal solutions to a variety of optimisation problems. These analytical approaches are exact and use various techniques from differential calculus (Rao, 2019). Two main forms are applied in a scheduling context: branch and bound and dynamic programming.

Branch and bound can be used to find optimal solutions to combinatorial optimisation problems by building a rooted decision tree of the solution space. This algorithm is guaranteed to give the globally optimal solution for mixed integer problems, but is exponential in terms of time complexity; in the worst case, each possible solution needs to be explored (Datta, 2022).

Dynamic programming can solve problems by solving a set of simpler sub-problems. This does require the problem to be expressed as a function of solutions to similar smaller problems (Otasevic, 2018). This could be the case if all machines and workers were completely independent, but for more complex JSPs this is often not the case.

In the work by Othman et al. (2012), the Lingo 13.0 commercial solver is used, which uses neither of these two methods (Lindo Systems Inc., 2022), but it must be said that this scheduling problem has single jobs (without operations) that are independent. Academic solver SCIP (Maher et al., 2016) is suitable for JSPs and uses the branch and bound method.

Simulation

Xu et al. (2011) finds that simulation-based solving techniques are among the most used methods, although their focus is not solely on the JSP. A JSP example where simulation is used is the work by Thürer et al. (2020); Thürer and Stevenson (2021), with 135 scenarios to be simulated. Another example is given by Wang et al. (2020), who use infinitely repeated game-based simulation to solve a multi-objective for a flexible JSP. Simulation is an excellent tool to solve multiple problems in DRC systems at once, the JSP being only a part of this, but requires complete simulation of all jobs of the problem to get results. This can be very time-consuming for larger problem sizes.

(Meta-)Heuristics

In many traditional scheduling approaches, a set of general rules can be applied to generate a schedule. Examples of such simple heuristics are the *when*, *where* and *who* rules mentioned in Section 2.2, but a more extensive overview of scheduling heuristics can be found in Demirkol et al. (1998). Heuristic optimisation can navigate large solution spaces with low computational cost but guarantees neither feasibility nor optimality; a global optimum is often not found due to their locally greedy strategies (Xu et al., 2011; Wang and Chen, 2013). Also, heuristics require knowledge of the optimisation problem before development and are often derivatives of best practices or practical knowledge about a problem. For example, Hoitomt et al. (1993) uses Lagrangian relaxation to solve a JSP divided into sub-problems that are then solved by separate heuristics that are specific per sub-problem and Zhang et al. (2017) apply game theory using their developed heuristic in a multi-layer scheduling approach.

Metaheuristics, on the other hand, are methods that search for a heuristic that fits the problem. These generally do not need prior knowledge of the optimisation problem to solve it but can perform better than man-made heuristics. In DRC scheduling, the most commonly used meta-heuristics are simulated annealing (SA) and the genetic algorithm (GA) (Xu et al., 2011).

SA is an optimisation technique where an analogy is drawn between the search for mathematical optimality and the crystalline restructuring of metals during the annealing phase of the production process. The

parallels are drawn between the metal reaching global minimum energy states through the reduction of atomic movements and an algorithm finding a global optimum of some objective by changing the decision variables from a previous state. SA can be used for both discrete and continuous optimisation and nonlinear models (Aleksendrić and Carlone, 2015).

The genetic algorithm is an optimisation technique that mimics the evolution of a species by survival of the fittest. Like natural evolution, GA combines different parameter settings in the form of chromosomes to form the next solution generation. Using the objective function as a selective tool, the algorithm continues evolution through reproduction, crossover and mutation. In the reproduction step, selecting the best chromosomes will have only the $x\%$ best remaining for the virtual 'fertilisation' in the crossover step. Then, on a random basis, some of the genes on the chromosomes are mutated. This ensures we move on from the limited set of genes present in the initial set of chromosomes. Genetic algorithms are suitable for nonlinear model optimisation (Murray-Smith, 2012).

Multiple forms of these two algorithms can be found in DRC JSP literature, as well as some custom meta-heuristics. Yazdani et al. (2015) propose two meta-heuristic algorithms to solve a DRC flexible JSP, namely simulated annealing, and vibration damping optimisation. The latter is an algorithm borrowed from the damping optimisation for a randomly vibrating system, based on a physical process like SA (Mehdizadeh et al., 2015). What is interesting about this work is that the authors review their MILP formulation using an exact solver (CPLEX) for small instances, after which they successfully apply meta-heuristics for larger problem sizes. The authors note that MOO would be an interesting topic for further research, which is attempted with the use of a multi-objective function in a JSP with worker fatigue integrated by Mouayni et al. (2019) (which we have already seen in Section 2.2.2). Their method, the greedy randomised adaptive search procedure (GRASP), is a two-phase multi-start meta-heuristic where a local search is performed after constructing a solution with a greedy function. Since the elemental weights in their objective function are constant, this is not strictly an example of MOO. The work by Gong et al. (2018) applies MOO through a new hybrid genetic algorithm for a double-flexible JSP. Although their algorithm development is described in detail, their implementation of human fatigue through a 'human fatigue indicator' does not show the same attention it received in their literature review or in the work by Mouayni et al. (2019). The authors also denote this as a valuable topic for further research.

Machine learning approaches

A less-mentioned branch of research is the field of machine learning approaches to JSP optimisation. According to Willems and Rooda (1994), one of the primary functions that, for example, neural networks can perform is constraint satisfaction and optimisation tasks, thereby considering them suitable for solving JSPs. Although research on the application of machine learning on JSPs is not novel (e.g., Niu et al. (2021)), multiple authors have stated, even in newer works, that there is a lack of validation or applicability to real-world problems Aytug et al. (1994); Lee et al. (2010); Cunha et al. (2020).

2.3. Existing research gaps

The literature review shows two main research gaps in the area:

- The existing validated methods that can estimate fatigue development through energy expenditure via bodily measurements require constant monitoring for every individual and are cumbersome when applied to a large workforce. For scheduling purposes, predictive methods using mathematical task modelling are a better option, with Garg et al. (1978) being applicable to activities that have not been previously evaluated. The method is applied in some existing works, but these lack applicative detail or do not consider both worker and task heterogeneity (Battini et al., 2016; Calzavara et al., 2018a, 2019; Berti et al., 2021). For implementation in a scheduling model, this method would require validation in a real-life scenario.
- In Table 2.1 we found no works that combine all requirements from Section 1.4, meaning no existing works combine bi-objective optimisation of productivity and fatigue in a double flexible JSP with worker and task heterogeneity.

This thesis fills these research gaps by applying a personalised, predictive fatigue model to a mathematical JSP formulation where real-life DRC OP system constraints are considered. Because this fatigue model has not yet been validated in a similar setting, nor have they been applied to similar scheduling problems, there

is a need for real-life applicative validation of the fatigue model and resulting schedules. In the next chapter, we will elaborate on the methodology for this problem.

3

Modelling approach

In this chapter, we describe our methodology to answer the first two research questions from Section 1.4. In Section 3.1, we address the first question, selecting an appropriate predictive fatigue model for scheduling purposes. Section 3.2 describes the development of the mathematical problem formulation, followed by Section 3.3, which discusses the optimisation approach. Together, these last two sections aim to answer the second research question.

3.1. Fatigue model

In Section 2.1.1 we have seen that EE-based fatigue models are most useful in a scheduling context because they can serve as a direct input to the RA model in (2.6) and do not share the flaws present in MET-based models. From Section 2.1.2, we can conclude that the task decomposition by Garg et al. (1978) is an $\dot{E}E$ estimation method that does not require live or prior measurements to be applied. For the highly similar tasks that shoppers do, this detailed task decomposition method is suitable for estimating the average $\dot{E}E$ during picking, though this is still a simplification of the real world. Each body movement is coupled to a specific equation dictating the energy expenditure of that task. For example, a two-arm lift with load L is represented by:

$$EE = 10^{-2} [0.062BW (h_2 - 0.81) + (3.19L - 0.52S \cdot L) (h_2 - h_1)] \text{ for } 0.81 < h_1 < h_2 \quad (3.1)$$

, with BW the body weight of the shopper, h_2 the endpoint height of the lift, S the sex of the shopper and h_1 the starting point height of the lift. The formulae from Garg et al. (1978) have been adjusted to fit our specific situation, with assumptions representing Picnic's OP system. In appendix C.4.1 we show the application of the task decomposition method in greater detail.

Each elemental movement is also estimated in terms of time, therefore allowing us to take the average $\dot{E}E$ during the pick action. The personal characteristics, a pick speed parameter, different item weights and walking distances ensure this calculation embodies both worker and task heterogeneity. The worker heterogeneity is also considered in the calculation of $\dot{E}E_R$, also from Garg et al. (1978), and $MAEE$ calculation from (2.8). The resulting values for $\dot{E}E$, $\dot{E}E_R$ and $MAEE$ are then used to calculate the RA value for each pick using (2.6).

Another advantage of this approach is that EE-based models can be linked to a measure of heart rate for validation of the model predictions, as opposed to MET-based models that require EMG measurements (see Section 2.1.2). The latter is cumbersome to practice, especially when measuring multiple muscles in a working environment. The fatigue model validation will be discussed in Chapter 4.

3.2. Mathematical job-shop scheduling formulation

In this section, we develop a mathematical formulation for the scheduling problem. We use different terminology than we have seen in Section 2.2.1 to correspond to the specific situation at Picnic. This entails the following comparison:

Literature term	Picnic term
Job	(Order) tote
Operation	Order line
Machine	(Picking) station
Worker	Shopper

Our mathematical model is based on the work by [Berti et al. \(2021\)](#) and [Yazdani et al. \(2015\)](#), but adds decision variables and constraints to improve the completeness of their formulations and the fit to Picnic's OP system. This means taking the JSP assumptions by [Xiong et al. \(2022\)](#) from Section 2.2.1 and adjusting these to fit the circumstances specific to Picnic as explained in Section 1.1. The resulting attributes and assumptions of this model are:

1. Set of *shoppers* process a unique set of *order lines* for a set of *totes* on a set of *stations*.
2. All information on shoppers, stations, totes and order lines is pre-known and fixed.
3. There are no due dates or different priorities between totes or order lines.
4. Order lines are specific items to be picked per tote, which can be any of the items in the assortment.
5. Any shopper can pick any order line, but only on a subset of stations.
6. Totes are independent of each other, order lines are not; they are subject to precedence constraints based on pre-known item fragility categories (1-5, with 5 being the most fragile category).
7. An order line can be picked once, on only one station at a time; there is no interruption or preemption for picking.
8. Waiting is permitted between order lines.
9. Each order line has a specific processing time and rest allowance, both item- and shopper-dependent.
10. Order lines of the same tote cannot be picked simultaneously.
11. Rest allowance is to be scheduled directly after picking an order line.
12. Shoppers can switch stations, which costs (travel) time.
13. Stations and shoppers only process one order line at a time and are independent of each other.
14. No buffer, setup time, warehouse or machine availability constraints are considered.

In 3.1, we describe the notation that is used in our mathematical model. First, we introduce the indices, sets and some auxiliary variables. Appendix C describes the generation of the sets in the virtual problem instances. The parameters are pure inputs and are also described in the appendix. Finally, the decision variables describe the degrees of freedom that our mathematical model offers during optimisation. We use a mix of continuous and binary decision variables.

3.2.1. Setting up the bi-objective

The mathematical model starts with the objective function, which includes two objectives relating to productivity and shopper fatigue as discussed in Section 1.4. Including multiple objectives in an optimisation problem is not new but is applied in many different fields. Multi-objective optimisation often does not lead to one optimal solution for all elements of the combined objective, as these can have opposing interests. For example, operational cost and product quality in a production process can negatively influence each other.

One of the works from Table 2.1 that applies MOO for storage assignment with human energy expenditure taken into account is the work by [Battini et al. \(2016\)](#). In their bi-objective of time and energy, the authors estimate the optimal solutions in the trade-off between these two objectives using an objective function that is a weighted sum with fractional single-objective deviation, following [Marler and Arora \(2004\)](#) (see Section 3.3.4). Translated into objectives and decision variables specific to our problem, this takes the following form:

Objective:

$$\min x \frac{C_{max} - C_0}{C_0} + (1 - x) \frac{EE_{max} - EE_0}{EE_0} \quad (3.2)$$

, where x is the weight of the individual objectives C_{max} for makespan and EE_{max} for energy expenditure, which are made non-dimensional by their singular optimal values, C_0 and EE_0 . This would be their optimal value if they made up the objective function alone or when $x = 1$ and $x = 0$, respectively. The two objectives C_{max} and EE_{max} are defined in (3.3) and (3.4).

Notation	Description
i, i', i''	Tote index
j, j', j''	Order line index
k, k', k''	Station index
l, l'	Shopper index
I	Set of totes
J	Set of order lines per tote, J_i is a set of all order lines for tote i
$J_{i,j}$	j th order line of tote i
K	Set of stations
$K_{i,j}$	Set of stations where order line $J_{i,j}$ can be picked
L	Set of shoppers
$C_{i,j}$	Completion time of order line $J_{i,j}$
C_{max}	Maximum completion time, makespan
EE_{max}	Maximum sum of required energy expenditure for all individuals
Parameters	
C_0	Optimal value for C_{max} if only optimising for C_{max} (or $x = 1$)
EE_0	Optimal value for EE_{max} if only optimising for EE_{max} (or $x = 0$)
$EE_{i,j,l}$	Energy expenditure for order line $J_{i,j}$ when shopper l is executing it
$f_{i,j}$	Item fragility category of order line $J_{i,j}$
$p_{i,j,l}$	Processing (picking) time of order line $J_{i,j}$ when shopper l is executing it
$r_{i,j,l}$	Recovery time for shopper l after picking order line $J_{i,j}$
$s_{k,k'}$	Switching time between station k and k'
x	Bi-objective weight parameter
Decision variables	
$S_{i,j}$	Starting time of order line $J_{i,j}$ (sec.) measured from start of operation ($t = 0$)
$\alpha_{i,j,k,l} \in \{0, 1\}$	1 if station k and shopper l are selected to process order line $J_{i,j}$, 0 otherwise
$\beta_{TO:i,j,j'} \in \{0, 1\}$	1 if tote i order line $J_{i,j}$ is performed right before $J_{i,j'}$, 0 otherwise
$\beta_{ST:i,j,i',j',k} \in \{0, 1\}$	1 if order line $J_{i,j}$ is performed right before $J_{i',j'}$ on station k , 0 otherwise
$\beta_{SH:i,j,i',j',l} \in \{0, 1\}$	1 if order line $J_{i,j}$ is performed before $J_{i',j'}$ by shopper l , 0 otherwise
$\gamma_{i,j,i',j',k,k',l} \in \{0, 1\}$	1 if shopper l moves from station k to k' to pick order line $J_{i',j'}$ after picking $J_{i,j}$, 0 otherwise

Table 3.1: Notations, parameters and decision variables for the mathematical JSP formulation with fatigue consideration.

With:

First, we define the terms used in the objective function. C_{max} is the makespan, defined as the maximum of all individual picks' $J_{i,j}$ completion times $C_{i,j}$.

$$C_{max} = \max_{i \in I; j \in J_i} (C_{i,j}) \quad (3.3)$$

The definition of EE_{max} is the maximum sum of the required energy expenditure for all individuals $EE_{i,j,l}$, depending on personal characteristics and the item to pick. The calculation of $EE_{i,j,l}$ and $r_{i,j,l}$ as its result is explained in detail in appendix C.

$$EE_{max} = \max_{l \in L} \left(\sum_{k \in K} \sum_{j \in J_i} \sum_{i \in I} EE_{i,j,l} \cdot \alpha_{i,j,k,l} \right) \quad (3.4)$$

Minimising EE_{max} results in an equally distributed sum of energy expenditure over shoppers which will also likely decrease the total among all shoppers. This is different from [Battini et al. \(2016\)](#), who only minimise the total required energy expenditure because their goal is to develop a storage assignment policy. In our case, we also want to choose the right shopper for a task and thus need to take the distribution into account too. Calculating the standard deviation of $EE_{i,j,l}$ is not possible with our chosen solver (see Section 3.3.1) because it would require illegal operations with decision variables, so we resort to using EE_{max} in our optimisation.

The completion time of an order line is the starting time plus processing time and recovery time (adjusted from [Berti et al. \(2021\)](#), representing assumptions 8, 9, and 11).

$$C_{i,j} = S_{i,j} + \sum_{l \in \mathbf{L}} \sum_{k \in \mathbf{K}} (p_{i,j,l} + r_{i,j,l}) \alpha_{i,j,k,l} \quad (3.5)$$

3.2.2. Constraints

No double work constraint

Each order line is executed once, by one shopper and on one station (representing assumption 7).

$$\sum_{l \in \mathbf{L}} \sum_{k \in \mathbf{K}} \alpha_{i,j,k,l} = 1 \quad \forall i \in \mathbf{I}; j \in \mathbf{J}_i \quad (3.6)$$

Item slotting constraint

Order lines can only be picked on the stations where the item is slotted (representing assumption 5).

$$\sum_{l \in \mathbf{L}} \sum_{k \in \mathbf{K}_{i,j}} \alpha_{i,j,k,l} = 1 \quad \forall i \in \mathbf{I}; j \in \mathbf{J}_i \quad (3.7)$$

Sequencing constraints

Tote and order line sequencing are constrained on three distinct levels: tote, station, and shopper. The resulting constraints look different but have the same basic structure.

Tote level sequencing

No two order lines j and j' belonging to the same tote i can be picked simultaneously - no matter which station and shopper are involved. Therefore, the starting time of the order line that is picked next is at least the completion time of the previous one (representing assumptions 8 and 10). To check which order line is picked first, we make use of decision variable β_{TO} (see also (3.9) and (3.10)): only if this equals 1 the constraint needs to hold. This equation needs to hold for all totes and all order lines in those totes.

$$S_{i,j'} \geq C_{i,j} \cdot \beta_{TO:i,j,j'} \quad \forall i \in \mathbf{I}; j \in \mathbf{J}_i; j' \in \mathbf{J}_i \setminus \{j\} \quad (3.8)$$

With β_{TO} defined as a binary decision variable dependent on the order line sequencing. If order line $J_{i,j'}$ is the next order line of tote i to be picked after $J_{i,j}$, $\beta_{TO:i,j,j'}$ equals 1, 0 if otherwise. Of two opposing β_{TO} , only one can be equal to 1; therefore, the sum of the two is 0 or 1 at most (see (3.9)). If j and j' denote the same order line, β_{TO} equals 0.

$$\begin{aligned} \beta_{TO:i,j,j'} + \beta_{TO:i,j',j} &\leq 1 \quad \forall i \in \mathbf{I}; j \in \mathbf{J}_i; j' \in \mathbf{J}_i \setminus \{j\} \\ \beta_{TO:i,j,j'} &= 0 \quad \forall i \in \mathbf{I}; j = j' \in \mathbf{J}_i \end{aligned} \quad (3.9)$$

As we know that each order line of tote i needs to be picked, we will have a sequence of J_i order lines to consider for $\beta_{TO:i,j,j'}$. The number of order lines for tote i will therefore exceed the number of binary sequence indicators β_{TO} that equal 1 by one, as formulated in (3.10). Note that the number of order lines in tote i can be related to α by summation over j, k and l .

$$\sum_{j' \in \mathbf{J}_i} \sum_{j \in \mathbf{J}_i} [\beta_{TO:i,j,j'}] + 1 = \sum_{l \in \mathbf{L}} \sum_{k \in \mathbf{K}} \sum_{j \in \mathbf{J}_i} \alpha_{i,j,k,l} \quad \forall i \in \mathbf{I} \quad (3.10)$$

Finally, there can be at most one order line that precedes or follows $J_{i,j}$.

$$\begin{aligned} \sum_{j' \in \mathbf{J}_i} \beta_{TO:i,j,j'} &\leq 1 \quad \forall i \in \mathbf{I}; j \in \mathbf{J}_i \\ \sum_{j' \in \mathbf{J}_i} \beta_{TO:i,j',j} &\leq 1 \quad \forall i \in \mathbf{I}; j \in \mathbf{J}_i \end{aligned} \quad (3.11)$$

Station level sequencing

Since two order lines can never be picked at the same time on one station, this puts additional restrictions on the schedule (assumption 13). Similar to the tote level sequencing constraint, the starting time of the following order line ($J_{i',j'}$) in the sequence of station k is at least equal to the completion time of the previous ($J_{i,j}$). This needs to hold given that:

- $J_{i',j'}$ is picked after $J_{i,j}$
- $J_{i,j}$ is indeed picked at station k , irrespective of l
- $J_{i',j'}$ is also picked at station k , irrespective of l

By definition of β_{ST} (see (3.13)), we know that all three conditions hold when $\beta_{ST:i,j,i',j',k} = 1$. Which shopper picks either order line is irrelevant in this constraint. This constraint is adjusted from Berti et al. (2021) and Yazdani et al. (2015).

$$S_{i',j'} \geq C_{i,j} \cdot \beta_{ST:i,j,i',j',k} \quad \forall i, i' \in \mathbf{I}; j \in \mathbf{J}_i; j' \in \mathbf{J}_{i'}; k \in \mathbf{K} \quad (3.12)$$

With β_{ST} defined as a binary decision variable dependent on the order line sequencing. If order line $J_{i',j'}$ is the next order line at station k to be picked after $J_{i,j}$, $\beta_{ST:i,j,i',j',k}$ equals 1, 0 if otherwise. Of two opposing β_{ST} , only one can be equal to 1, so the sum of the two is 0 or 1 at most. However, we have an additional condition for the station as there is no guarantee that order lines $J_{i,j}$ and $J_{i',j'}$ are both picked at station k . Therefore, the right-hand side from (3.9) is replaced by the product of two summed α corresponding to $J_{i,j}$ and $J_{i',j'}$ in (3.13). If $J_{i,j}$ and $J_{i',j'}$ denote the same order line, β_{ST} equals 0.

$$\begin{aligned} \beta_{ST:i,j,i',j',k} + \beta_{ST:i',j',i,j,k} &\leq \sum_{l \in \mathbf{L}} \alpha_{i,j,k,l} \cdot \sum_{l \in \mathbf{L}} \alpha_{i',j',k,l} \quad \forall i \in \mathbf{I}; j \in \mathbf{J}_i; [i', j'] \in \mathbf{J} \setminus \{[i, j]\}; k \in \mathbf{K} \\ \beta_{ST:i,j,i',j',k} &= 0 \quad \forall [i, j] = [i', j'] \in \mathbf{J} \end{aligned} \quad (3.13)$$

As we know that each order picked at station k needs to be picked, we can sum this number using decision variable α like in (3.10). The number of binary sequence indicators β_{ST} that equal 1 will be exceeded by one by the sum of α over station k , as formulated in (3.14).

$$\sum_{j' \in \mathbf{J}_{i'}} \sum_{i' \in \mathbf{I}} \sum_{j \in \mathbf{J}_i} \sum_{i \in \mathbf{I}} [\beta_{ST:i,j,i',j',k}] + 1 = \sum_{l \in \mathbf{L}} \sum_{j \in \mathbf{J}_i} \sum_{i \in \mathbf{I}} \alpha_{i,j,k,l} \quad \forall k \in \mathbf{K} \quad (3.14)$$

Finally, we also need to constrain the number of sends/receives that an order line can have: there can be at most 1 order line that precedes or follows $J_{i,j}$ at station k .

$$\begin{aligned} \sum_{k \in \mathbf{K}} \sum_{j' \in \mathbf{J}_{i'}} \sum_{i' \in \mathbf{I}} \beta_{ST:i,j,i',j',k} &\leq 1 \quad \forall i \in \mathbf{I}; j \in \mathbf{J}_i \\ \sum_{k \in \mathbf{K}} \sum_{j' \in \mathbf{J}_{i'}} \sum_{i' \in \mathbf{I}} \beta_{ST:i',j',i,j,k} &\leq 1 \quad \forall i \in \mathbf{I}; j \in \mathbf{J}_i \end{aligned} \quad (3.15)$$

Shopper level sequencing

One shopper can not handle two different order lines at the same time (assumption 13), so we get a constraint similar to (3.12). There are two major differences:

- The constraint needs to hold for all shoppers but is irrespective of the stations where $J_{i,j}$ and $J_{i',j'}$ are picked. k and l thus swap places in the formula.
- If a shopper switches stations, there is a station switching time added to the minimum starting time of $J_{i',j'}$.

The station switching time is only added if station switching actually takes place. This is indicated by decision variable γ (see (3.17)).

$$S_{i',j'} \geq \left(C_{i,j} + \sum_{k \in \mathbf{K}} \sum_{k' \in \mathbf{K}} s_{k,k'} \cdot \gamma_{i,j,i',j',k,k',l} \right) \cdot \beta_{SH:i,j,i',j',l} \quad \forall i, i' \in \mathbf{I}; j \in \mathbf{J}_i; j' \in \mathbf{J}_{i'}; l \in \mathbf{L} \quad (3.16)$$

Station switching happens if the same shopper l subsequently picks two order lines $J_{i,j}$ and $J_{i',j'}$ at different stations k and k' . If so, we state that $\gamma_{i,j,i',j',k,k',l} = 1$ and the following needs to hold:

- Order line $J_{i',j'}$ is picked at station k' and by shopper l

- There is an order line $J_{i,j}$ for which holds:
 - Order line $J_{i,j}$ is picked at station k and by shopper l
 - Order line $J_{i',j'}$ is picked right after $J_{i,j}$
- Order lines $J_{i,j}$ and $J_{i',j'}$ are not picked on the same station

These are all satisfied if $\beta_{SH:i,j,i',j',l} = 1$ (see (3.18)). Therefore γ can be defined as:

$$\begin{aligned} \gamma_{i,j,i',j',k,k',l} &= \sum_{j \in \mathbf{J}_i} \sum_{i \in \mathbf{I}} [\beta_{SH:i,j,i',j',l} \cdot \alpha_{i,j,k,l} \cdot \alpha_{i',j',k',l}] \quad \forall i' \in \mathbf{I}; j' \in \mathbf{J}_{i'}; k \in \mathbf{K}; k' \in \mathbf{K} \setminus \{k\}; l \in \mathbf{L} \\ \gamma_{i,j,i',j',k,k',l} &= 0 \quad \forall k = k' \in \mathbf{K} \end{aligned} \quad (3.17)$$

With β_{SH} defined as a binary decision variable dependent on the order line sequencing, similar to (3.13) and (3.14) but with l and k swapping places (see (3.18) and (3.19)).

$$\begin{aligned} \beta_{SH:i,j,i',j',l} + \beta_{SH:i',j',i,j,l} &\leq \sum_{k \in \mathbf{K}} \alpha_{i,j,k,l} \cdot \sum_{k \in \mathbf{K}} \alpha_{i',j',k,l} \quad \forall i \in \mathbf{I}; j \in \mathbf{J}_i; [i',j'] \in \mathbf{J} \setminus \{[i,j]\}; l \in \mathbf{L} \\ \beta_{SH:i,j,i',j',l} &= 0 \quad \forall [i,j] = [i',j'] \in \mathbf{J} \end{aligned} \quad (3.18)$$

$$\sum_{j' \in \mathbf{J}_{i'}} \sum_{i' \in \mathbf{I}} \sum_{j \in \mathbf{J}_i} \sum_{i \in \mathbf{I}} [\beta_{SH:i,j,i',j',l}] + 1 = \sum_{k \in \mathbf{K}} \sum_{j \in \mathbf{J}_i} \sum_{i \in \mathbf{I}} \alpha_{i,j,k,l} \quad \forall l \in \mathbf{L} \quad (3.19)$$

Finally, there can be at most one order line that precedes or follows $J_{i,j}$ at station l .

$$\begin{aligned} \sum_{k \in \mathbf{K}} \sum_{j' \in \mathbf{J}_{i'}} \sum_{i' \in \mathbf{I}} \beta_{SH:i,j,i',j',l} &\leq 1 \quad \forall i \in \mathbf{I}; j \in \mathbf{J}_i \\ \sum_{k \in \mathbf{K}} \sum_{j' \in \mathbf{J}_{i'}} \sum_{i' \in \mathbf{I}} \beta_{SH:i',j',i,j,l} &\leq 1 \quad \forall i \in \mathbf{I}; j \in \mathbf{J}_i \end{aligned} \quad (3.20)$$

Precedence constraints

We need to adjust the precedence constraint as used by Berti et al. (2021) to fit our scenario. Order lines within one tote need to be picked in the order of the fragility category (assumption 6), but unlike Berti et al. (2021) we cannot assume that this means $J_{i,j}$ needs to be executed before $J_{i,j+1}$. This means that we need a set of constraints that relates to tote pick sequence decision variable β_{TO} . For order lines within one tote, we should have:

- If order line $J_{i,j'}$ has a lower fragility category than $J_{i,j}$, it should not be picked right after $J_{i,j}$
- If order line $J_{i,j'}$ has a higher fragility than $J_{i,j}$, it can either be picked right after or later
- If the fragility category of two order lines is equal, there is no precedence constraint on β_{TO}

Together, these rules are written as the following constraint:

$$\beta_{TO:i,j,j'} \cdot f_{i,j} \leq f_{i,j'} \quad \forall i \in \mathbf{I}; j, j' \in \mathbf{J}_i \quad (3.21)$$

Variable constraints

The start and thus completion time of any order line is at least 0 seconds (taken from Yazdani et al. (2015)).

$$S_{i,j} \geq 0 \quad \forall i \in \mathbf{I}; j \in \mathbf{J}_i \quad (3.22)$$

Decision variables α , β_{TO} , β_{ST} , β_{SH} and γ are binary.

$$\alpha_{i,j,k,l}, \beta_{TO:i,j,j'}, \beta_{ST:i,j,i',j',k}, \beta_{SH:i,j,i',j',l}, \gamma_{i,j,i',j',k,k',l} \in \{0, 1\} \quad \forall i, i' \in \mathbf{I}; j \in \mathbf{J}_i; j' \in \mathbf{J}_{i'}; k, k' \in \mathbf{K}; l \in \mathbf{L} \quad (3.23)$$

3.3. Model optimisation

Unlike the works by Berti et al. (2021) and Yazdani et al. (2015) that it was based on, our formulation is not an example of a MILP problem. These formulations are common in JSP literature and can be solved by most exact solvers. Due to constraints like (3.16) and (3.17), where we multiply decision variables, not all constraints are linear and our formulation categorises as a mixed-integer nonlinear programming (MINLP) problem.

3.3.1. Selecting a solution approach

In Section 2.2.3 we have seen that classical optimisation methods are only suitable for very small problem instances, especially when more constraints are added. For medium-sized instances, simulation could be an option where many KPIs can be easily tracked, although there are not many examples with realistic JSPs where this is applied. Meta-heuristics are the most reasonable option for larger problem instances, weighing solving speed for large problem instances and optimality. Machine learning approaches have yet to be proven fit for application, although their potential may be even more significant than meta-heuristics.

However, our mathematical formulation represents a specific JSP that has not been previously addressed to our best knowledge, so it requires model verification. To be able to verify the model constraints and outcomes, we need an optimisation approach that can guarantee global optimality of the solutions (see Section 3.3.3). Thus, even with the knowledge that this can only be applied to very small problem instances, we apply an exact solver to the mathematical model. Because of the presence of binary variables, the MINLP is non-convex by definition, which leaves us with few options when looking for a solver [Kronqvist et al. \(2019\)](#). The two best-performing options are SCIP and BARON, but SCIP is the only one that is freely available without restrictions (BARON allows for up to 10 constraints and 50 nonlinear operations in demo mode). Other advantages are SCIP's integrated implementation in Python through PySCIPOpt ([Maher et al., 2016](#)), easy iterative constraint generation, insights in the solving process and a wide variety of solver parameter settings. Therefore, we use the PySCIPOpt interface to solve the scheduling problem in SCIP.

SCIP, or Solving Complex Integer Problems, uses a spatial branch-and-bound algorithm with linear relaxations. This allows the solver to find global optima for non-convex MINLPs [Bestuzheva et al. \(2021\)](#). SCIP was installed on Windows Subsystem for Linux and executed on an Intel 10700F octa-core 2.9/4.7GHz processor with 16GB DDR4 RAM. Our workflow to set up the scheduling problem in PySCIPOpt is as follows:

1. Input parameter generation with the model run settings: number of totes, order lines, stations and shoppers, random seed (see appendix C for details). The random seed allows us to run multiple different instances of the same problem size, with repeatable results.
2. Initialisation of the DV by iterating over indices i, j, k, l and using SCIP expression `addVar`.
3. Setting the model objective using SCIP expression `setObjective`.
4. Initialisation of the constraints by iterating over indices i, j, k, l and using SCIP expression `addCons` or `addConsAnd`.

From here, there will be a SCIP model instance that can be solved in Python to give a solution to the problem. Even for small instances, this takes a long time to find an optimal solution. We will therefore discuss the computational complexity of our model in the next section.

3.3.2. Computational complexity

As one can imagine from looking at the decision variables, their indices and the number of constraints, the current formulation requires a large number of individual decision variables (DV) and constraints to be generated by a solver. We can calculate these numbers using (3.24) and (3.25), respectively, given the following notation:

Model input name	Abbreviation
n_{totes}	I
$n_{orderlines}$	J
$n_{stations}$	K
$n_{shopper}$	L

$$\text{No. DV} = IJ(IJ(K^2L + K + L) + J + KL + 1) \quad (3.24)$$

$$\text{No. Cons.} = J(IJ(I((K^2 + 2)L + 2K) + 3) + 9I) + I + K + L \quad (3.25)$$

For large problem instances, the number of DV and constraints is roughly equal to $I^2J^2K^2L$ as it is the highest order term in both (3.24) and (3.25). In Table 3.2, we show the increase for several problem instances. For more realistic problem instance sizes for Picnic's zone pick OP system, we would be in the range of 10E+13 DV and constraints.

I	J	K	L	Decision variables	Constraints
2	2	2	2	220	322
3	2	3	2	882	1,106
3	3	3	2	1,953	2,438
3	4	3	2	3,444	4,292
4	4	3	2	6,064	7,513
4	4	4	3	14,352	16,219
5	5	5	5	85,150	91,240
10	10	5	5	1,353,600	1,453,920
60	40	5	5	777,758,400	835,509,670

Table 3.2: Number of decision variables and constraints for different problem sizes.

Even for relatively small instances (e.g. row 4 in Table 3.2), SCIP takes a long time to find an optimal solution, not having found the optimum after 6000 seconds. To maximise the problem instance sizes that we can run, we impose three solver speed improvement measures.

- First, we run the optimisation process in the SCIP interactive shell to be able to use its concurrent solve option. This means that SCIP uses multiple solvers concurrently, each with a different random seed for the branch-and-bound process. Because the separate solvers can communicate with each other, this can speed up the process because they can traverse the solution space more quickly. Also, with random solver instance creation, the dependency on a favourable set of initial values is decreased by using multiple instances. On the contrary, because this requires the concurrent creation of nodes in the branch-and-bound process, memory is a restricting factor when using many processor cores. We found decent results for low memory usage with four concurrent solvers.
- Second, we impose a stopping criterion based on the runtime of the process (6000 seconds per optimisation run). As this early stopping does not guarantee a feasible solution, we set feasibility emphasis arguments and check the validity of the solution before saving. If no feasible solution is found within the time limit, the program reruns the optimisation with double the time limit and different solver instances. This is useful when selecting an appropriate runtime setting but often not necessary when a sufficiently long runtime was set initially. Generally speaking, SCIP can find near-optimal solutions quickly but cannot guarantee a global optimum when problem instances become larger.
- Third, we impose (3.26) as an additional constraint to limit the search space. This assumes that all order lines are picked by one shopper, for which the longest possible processing time and RA are assumed, plus the maximum travel time after each pick. As we optimise for the makespan during optimisation, this worst-case scenario constraint is not close to a violation in the optimal solution.

$$S_{i,j} \leq (\max(p_{i,j,l}) + \max(r_{i,j,l}) + \max(s_{k,k'})) \cdot \sum_{i \in I} |J_i| \quad \forall i \in I; j \in J_i \quad (3.26)$$

3.3.3. Verification of constraints and model output

Constraint verification can be done by manually checking the generated schedules with the constraints in our model. For example, using the generated tote schedule in Figure 3.1, we can see that the constraint in (3.21) is violated: order lines of fragility category 4 or 5 are scheduled to be picked into tote 1 before an order line of fragility category 1 is picked.

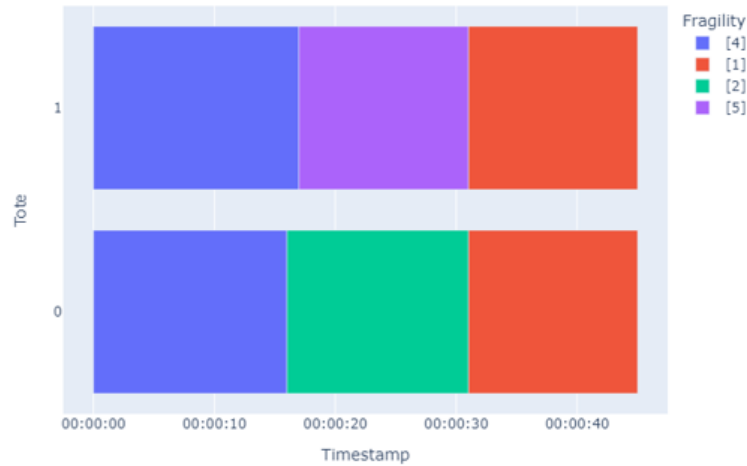


Figure 3.1: 'Tote schedule' with violated precedence constraint. The coloured blocks represent the processing time and RA of single order lines, grouped on a single row per tote that they are picked for. The colours indicate item fragility categories.

Another verification example can be explained by Figures 3.2a and 3.2b. This is a possible result without the constraint in (3.7), meaning that order lines can be picked irrespective of their actual storage location. Figures like these give us information to verify the tote, station, and shopper level sequencing constraints and the no double work and precedence constraints. Coupling these with the DV values in the SCIP solution output allows for elementwise verification of the constraints. We have done this for several different problem instances, consistently passing the tests.

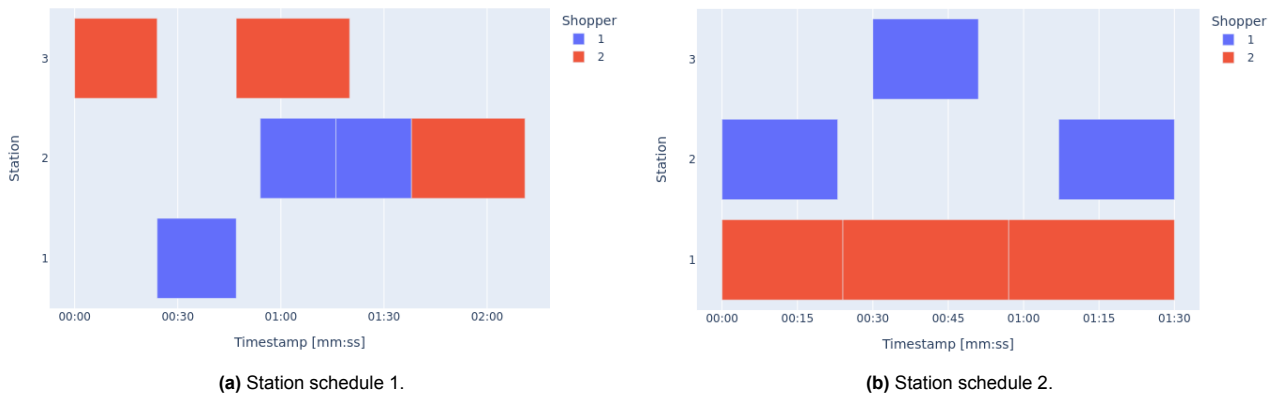


Figure 3.2: Two different 'station schedules' for the same problem instance. The coloured blocks represent the processing time and RA of singular order lines, grouped per station where they are picked. The colours indicate the shopper that picks the items. As can be seen, the number of picks per station is not equal for the two solutions, which should always be the case for feasible solutions to the same problem instance.

To verify the scheduling problem formulation as a whole, we compare 'scheduling with time and RA' and 'scheduling with time' for a set of 63 different problem instances. 'Scheduling with time and RA' implies that we use the mathematical formulation as presented. 'Scheduling with time' differs in how RA is applied to the schedule: this version applies the personalised fatigue - and resulting RA requirement - *after* the optimal schedule is made, following the method by Calzavara et al. (2019). Mathematically, this means the term $r_{i,j,l}$ in (3.5) is considered equal to 0 during optimisation. When a schedule has been generated, the actual value of $r_{i,j,l}$ is added to the resulting schedule without violating the other constraints; the model will be solved another time with fewer DV and constraints. As the first approach has complete information during optimisation, we use the outcomes as a verification step of the model output; more information during scheduling should lead to equal or better schedules, as long as we find guaranteed global optima in both cases. This is also why an exact solver is helpful in constraint and model output verification.

To keep optimisation simple, we only optimise for the makespan by setting weight $x = 1.00$. As mentioned in Section 3.3.2, larger problem instances do not find a guaranteed optimal solution before they outrun their

set time limit. We have set a runtime limit to 6,000 seconds for this comparison, meaning that one comparison takes around 3.5 hours maximum to return a result. Generally, we have found problem instances with up to 2,500 decision variables to find a global optimum within the time limit and problems above 50,000 decision variables to run out of memory before the time limit. Problems between these two values will often find feasible solutions, just no guaranteed global optima. In our comparison between scheduling with and without RA, we distinguish between runs that did or did not find a global optimum within 6,000 seconds. In Figure 3.3, we compare the scheduling outcomes of the two methods for the 63 problem instances.

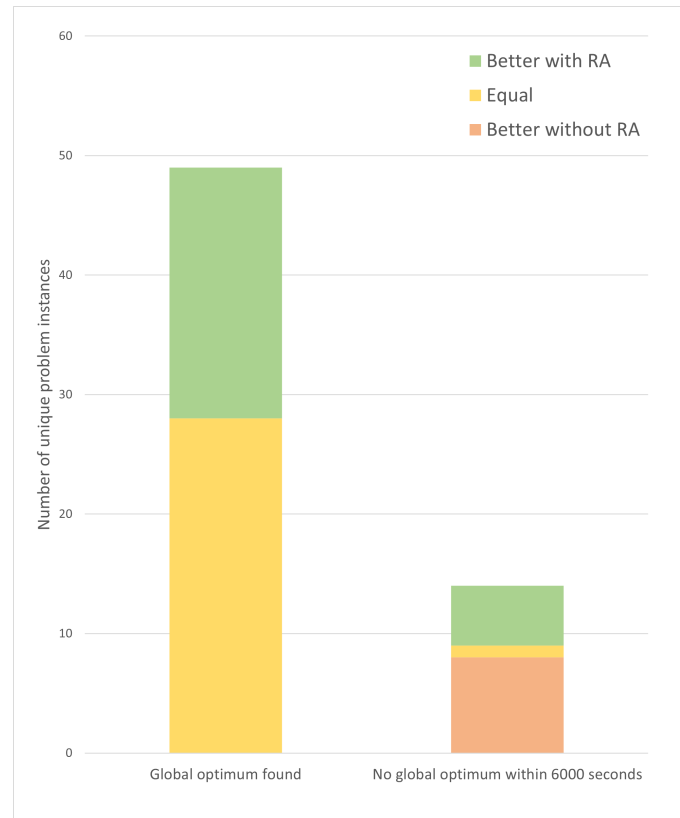


Figure 3.3: Overview of the results from our model verification tests with multiple unique problem instances of different sizes. We categorise optimisation runs that found a global optimum within the 6000 second time limit and those that did not (the columns), and use colours to indicate the comparative results between the 'scheduling with time and RA' and 'scheduling with time' runs for the same problem.

The first column in the figure shows the problem instances for which SCIP returns a global optimum within 6,000 seconds. This column is made up of two categories, showing that all problem instances performed equally or better with 'scheduling with time and RA' than with 'scheduling with time'. This is the expected behaviour of the mathematical model. The second column shows the problem instances for which no global optimum was found within 6,000 seconds. This category is much smaller than the first because these take much longer to run and give no further information for verifying the mathematical formulation. Because the solver may return local optima, it is not surprising that 'scheduling with time' results in a better solution in many cases. This does not disprove our mathematical model's validity because the solution's suboptimality could cause this behaviour.

In Figure 3.4 we visualise the performance difference between the two scheduling approaches. The grey diagonal indicates equal performance for the two methods, the area above the diagonal indicates a shorter makespan when RA is considered during scheduling, and the area below the line indicates the opposite. Each dot represents one problem instance, categorised by colour to indicate the time in which a globally optimal solution was found. Larger problem instances often take longer and have longer makespan because more tasks need to be scheduled. We can see this in the figure as the problems with the shortest solving times are the ones with the shortest makespan. We also see that the absolute difference between the methods' schedule lengths (i.e. the distance to the diagonal) is often larger for longer schedules. This is also to be

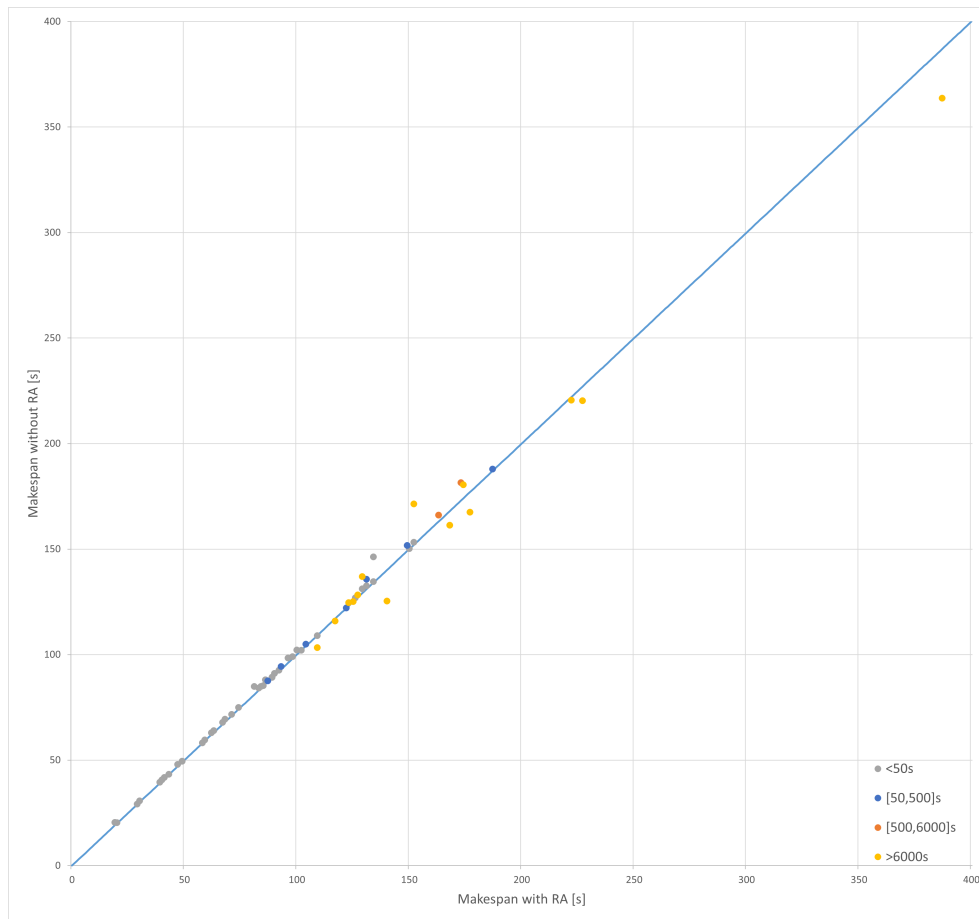


Figure 3.4: Individual results of the two scheduling approaches for numerous unique problem instances. The diagonal line indicates equal performance between 'scheduling with time and RA' and 'scheduling with time'.

expected, as there are more degrees of freedom in the scheduling problem where the scheduling approach can make a difference. The only problem instances below the diagonal belong to the '>6000s' category, as shown in Figure 3.3. Together with the verification of individual constraints, this is convincing evidence that the model behaves as expected.

3.3.4. Estimating the Pareto front

After successful constraint and model output verification, we optimise a series of virtual problem instances. To judge the effectiveness of our bi-objective optimisation approach over the traditional single-objective function with makespan, we estimate the Pareto front for each of these instances. The Pareto front is a popular method to visualise multi-objective optimisation in a plot (see Figure 3.5). One objective is on the x-axis and the other is on the y-axis, and each dot represents a single solution. In a minimisation problem like ours, an ideal solution would be as far as possible in the bottom left corner. However, we are constrained by the feasible solution space, limiting the feasible solutions to the dots in the figure.

If no other solutions exist that, relative to some solution, improve one objective without hurting the other, we call this a non-dominated solution. A line through all non-dominated solutions, or the Pareto set, is called the Pareto front, indicating Pareto-optimal solutions. The Pareto front allows management to weigh the importance of the two objectives and select a solution accordingly. Depending on the priority of the two objectives, the overall preferred solution could be any of the points on the Pareto front. Finding the Pareto front can be time-consuming, as the optimisation must be executed numerous times. Still, often a rough estimation can be found quite quickly.

The weighted sum objective is not the perfect way to estimate the Pareto front, as we can only guarantee that a solution is Pareto optimal if the solver has found the global optimum. However, taking sufficiently small steps

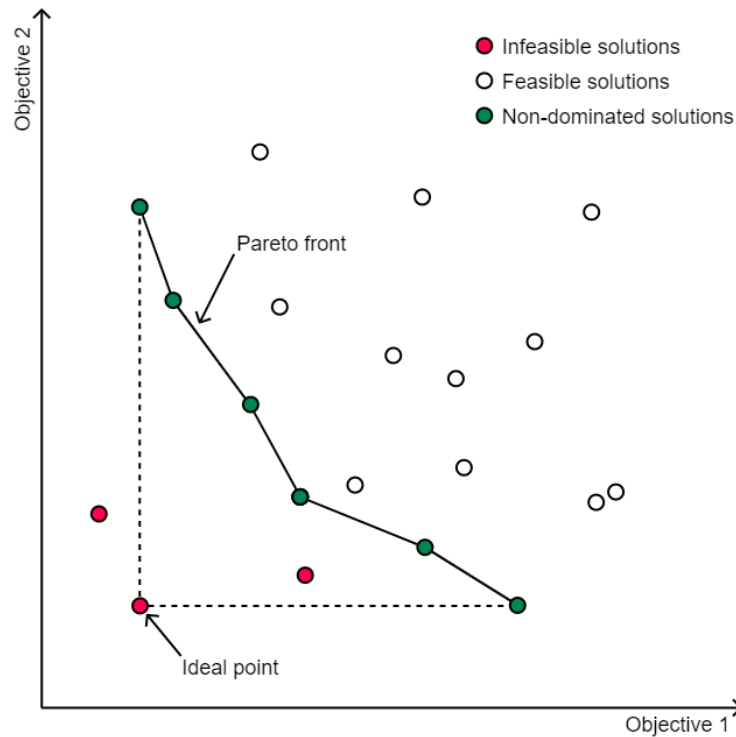


Figure 3.5: Visualisation of a Pareto front with objectives f_1 and f_2 . Each dot represents a solution to the bi-objective optimisation problem.

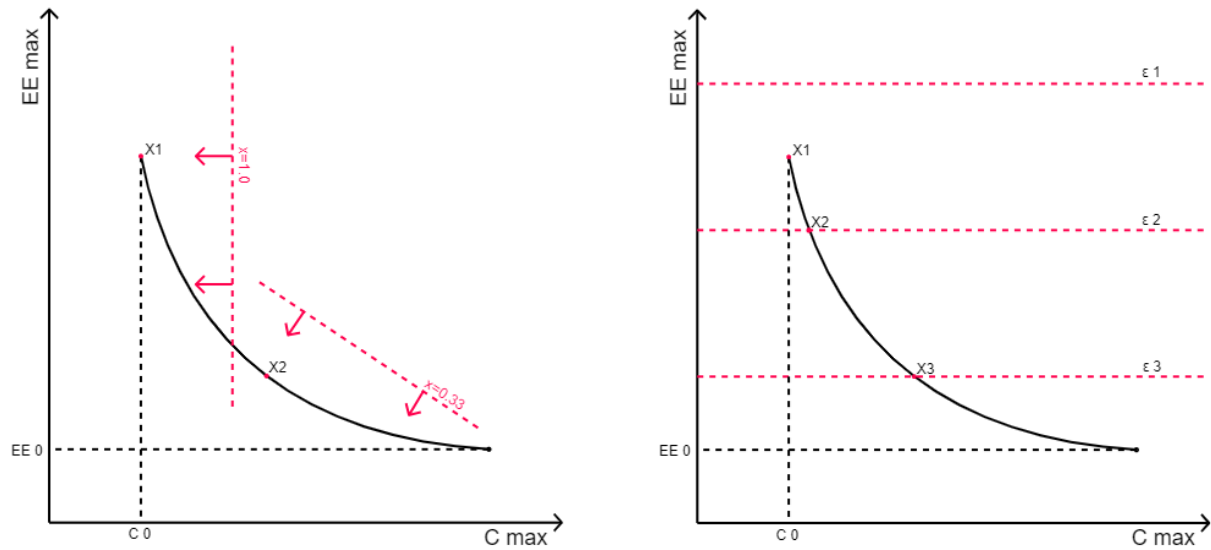
in varying x may get us closer, as shown by [Snyder and Daskin \(2005\)](#). The principle of the weighted sum method is to find a tangent line with slope x of the Pareto front (see [Figure 3.6a](#)). Another problem, therefore, is the method's inability to traverse non-convex objective functions and solution spaces, as this would lead to multiple options to place the tangent.

A way in which this can be avoided is the ϵ -constraint method, which exists in multiple different forms. Generally, the method entails that one of the objectives is kept while the others are transformed into a constraint bounded by ϵ . What remains is a single-objective problem that can generate a Pareto optimal solution if the solution is unique or if the optimisation has been done for all other objectives too. A subset of the Pareto optimal set can be attained by varying the value of all ϵ , as depicted in [Figure 3.6b \(Becerra and Coello Coello, 2006\)](#).

For our specific problem, however, this method does not have a real advantage in estimating the Pareto front over the weighted sum with fractional deviation in [\(3.2\)](#), since we cannot guarantee solution uniqueness or solver optimality for all problem instances, and it would require many optimisation iterations.

For each problem instance we run, we first run the two single-objective versions of objective function [\(3.2\)](#) to get C_0 and EE_0 . Then, we use these values and vary the bi-objective weight x in 25 equally-spaced steps between 0 and 1 and optimise the resulting MINLP instances. Since we receive global optima from the solver, any solution where x is neither 1 nor 0 will return a Pareto-optimal schedule, allowing us to estimate the Pareto front.

With the solving done, we export the solution to a `.sol` file, objective value and solving time to a `.txt` document, and all input and DV values to an Excel worksheet. Also, we create two figures with a Gantt chart of the schedule that was generated: one for an overview per station (like [Figure 3.1](#)), the other for an overview per tote (like [Figure 3.2](#)). Besides the solver instances created by SCIP, all input parameters are the same for multiple runs of the same seed, meaning that the optimal solution should be repeatable.



(a) Optimising weighted sum of C_{max} and EE_{max} given weights x , visualised by moving the red line as close to the origin while still in the feasible solution space. $x = 1.0$ finds X_1 , $x = 0.33$ finds X_2 .

(b) Optimising C_{max} for different $\epsilon_{1,2,3}$ values that constrain EE_{max} , while still in the feasible solution space. ϵ_1 finds X_1 , ϵ_2 finds X_2 and ϵ_3 finds X_3 .

Figure 3.6: Possible Pareto optimal solutions $X_{1,2,3}$ drawn on a hypothetical Pareto front between two 'utopia points', using the weighted sum and ϵ -constraint method. C_0 and EE_0 indicate the optimal value for single-objective optimisation where we find the utopia points, the endpoints of the Pareto front curve.

4

Empirical study

In an attempt to bridge the gap between simulated reality and practical application, we put the fatigue model and fatigue-conscious schedules to the test. The experiments are done during normal picking operations in Picnic's OP system, a real-life facility still in ramp-up. It must be noted that this system is not a lab environment, thus restricting the control over the experimental setting. This is no different from the difficulties that arise when implementing HF-centred scheduling in a working environment, making the lack of control over the experimental environment both a challenge and an opportunity for learning about the requirements for implementation.

Given the current options for detailed schedule control in FCA's zone pick system, we cannot generate, apply and review different picking schedules. On the other hand, recreating all environmental factors from the zone pick system in a mathematical model would increase the model complexity considerably, thereby no longer allowing it to be solved with the current methods. Also, the current scheduling algorithm would not be able to solve realistic order sizes and numbers within a reasonable run time and, therefore, could not output a realistic schedule. Rather than testing the output of our scheduling model, we review our fatigue quantification approach instead; the validity of predictive EE estimations, EE -based RA calculations and their relation to perceived fatigue.



Figure 4.1: Shopper picking during the experiment. The illustrated chest strap is worn underneath the clothes.

To get a first look at the experimental factors, data gathering and cleaning process and measurement reliability, we performed a pilot with only one test subject. This was used to fine-tune the experimental setup, instructions

for participants, data gathering and cleaning process, and analysis steps. During the experiments, we assess shoppers while they work in normal circumstances. Data is gathered using Picnic's warehouse management system (WMS), a laptop and a heart rate chest strap sensor worn underneath their clothes (see Figure 4.1). This chapter describes each aspect of the experiment in more detail.

4.1. Environment

As explained in Section 1.1, the zone pick area in FCA consists of multiple picking stations. Each station has a handful of slots where pallets, rolling containers or other large load carriers can be placed (see Figure 4.2). The slotting of those load carriers over the available stations is based on workload distribution in terms of item order volume so that the pick demand per station is as balanced as possible. If not all slots of a station are used, items are stored in slots that are as close to the instruction screen as possible to minimise walking time. Still, this walking time and the picks' physical demand are not considered in the current slotting algorithm. These experiments aim to test our proposed method for physical workload prediction and measurement, a first step towards considering the physical workload for both scheduling and slotting.

Given that the pilot tests showed there are many factors at play in the zone pick system that can influence the outcome of our experiments, we must monitor the subjects closely to detect any anomalies. Using four shoppers and stations simultaneously during testing allows for close monitoring because these can be within viewing distance of each other. However, testing is done during the busy Christmas period, when there is no room for delay. Also, shift captains often rotate shoppers to different sections of the OP process. Having four shoppers unavailable for reassignment has proven to be too many during most shifts. Therefore, we chose to test two shoppers during the busiest time of day, such that there is sufficient pressure on the stations to have them work continuously. Also, as the OP system at FCA is still in ramp-up, shoppers often serve two stations simultaneously because the pick demand per station is too low to assign a single station to a shopper. We chose adjacent station pairs with the highest daily workload; low-workload stations are often grouped by more than two per shopper.



Figure 4.2: A zone pick station with four different items slotted.

Even for our two designated station pairs for this experiment, we cannot create dummy orders or special operational procedures because the WMS is not easily adjustable; we run a typical operation at the four stations. This means that there could be system-induced idle time. We mitigate this as much as possible by choosing periods right after a batch of order totes has been released into the zone pick conveyor loop to keep constant pressure on the picking stations. Any anomalies during the experiments are noted.

4.2. Test subjects

The selection of test subjects should reflect a real operation as much as possible for multiple reasons. The first is that human-centred scheduling for an OP system only makes sense if it is designed and evaluated for the people working in that system. A big part of the human factors effects in the scheduling algorithm results from worker heterogeneity, and the pool of test subjects should reflect this. Another reason is that by selecting actual employees, many of the initial learning effects no longer influence the experiments. Existing employees are familiar with the processes, best practices and safety regulations and can participate in the experiments without influencing the system performance too much. Therefore, we randomly select a set of 20 healthy employees from the employee pool to participate in our experiments.

Per testing day, the two participants are split up into groups A and B. Group A first performs experimental scenario 1, followed by a break and scenario 2 (see Section 4.4). The order in which these scenarios are performed may influence the outcome of the experiments, which is why group B does the opposite. Since we expect employees to behave and experience the experiment differently, we want them to participate in both experimental scenarios to be able to compare their relative experiences instead of their personal differences. With a total of 20 employees, there is likely a large variance in the results, but we think this number is sufficient to represent the variance in the entire employee pool.

Participants are selected the day before the experiment, so we can only select shoppers that are also working the next day. We do not want shoppers to feel pressured to participate and therefore do not make them decide on this at the start of the experiment. Also, selecting them the day before allows us to select any non-novice shopper (> 1 month of experience) that has not participated before and not only shoppers that are scheduled to work in zone picking. If shoppers agree to participate, we add them to next day's zone picking schedule with the operational staff. The first two days of testing made clear that the operational staff's awareness of the experiment, the selected participants, and the required boundary conditions is crucial to the accuracy of the measurements, so this is discussed with them the day before and before the shift starts.

4.2.1. Qualitative ratings and personal information

Shoppers are asked to fill in a questionnaire before, during and after the experiments, as has also been done during a pilot study (see appendix B). This provides qualitative feedback about their general and experiment-specific experience of the physical workload in FCA and zone pick through the NASA TLX form (Hart and Staveland, 1988), but also contains RPE scores recorded every 10 minutes during the experiments. These questions are embedded in a small graphical user interface (GUI) on a laptop. The GUI is also used to inform the participants about the experiment and scenario, record personal information, and provide a work and break timer and is available in both English and Dutch. This makes filling in the personal (experience) data more manageable during work and less privacy-sensitive, as all data is encrypted upon saving, and there is no unnecessary paper trail. The GUI was written in Python but converted to an executable to work on every laptop in the experiments. We give an overview of the GUI in appendix E.

4.3. Fatigue model validation

As mentioned in Section 2.3, our method for predictive fatigue estimation requires validation if it is applied to a setting where worker and task heterogeneity are concerned. The task decomposition method from Section 3.1 is no exception; we describe its experimental validation in this section. In Section 2.1.2, we explained that HR-based $\dot{E}E$ estimation is most suitable for application in real life, despite inaccuracies in the estimated values. We will therefore use this as our ground truth in the validation of our predictive fatigue model.

4.3.1. Heart rate data measurement and preprocessing

Previous research has shown that the Polar H10 chest strap sensor is the best-performing commercially available device to measure heart rate in dynamic activities (e.g. Pasadyn et al. (2019); Stone et al. (2021)). This device has electrodes embedded in the chest strap, which can make an electrocardiogram (ECG) and has been supplied by TU Delft's TEL research group. Noncommercial ECG measurements are often performed with a Holter monitor, which can have five individual electrodes to be stuck on different chest areas. In contrast, the Polar H10 strap has a single strap that should be worn just below the chest, as shown in Figure 4.3a).

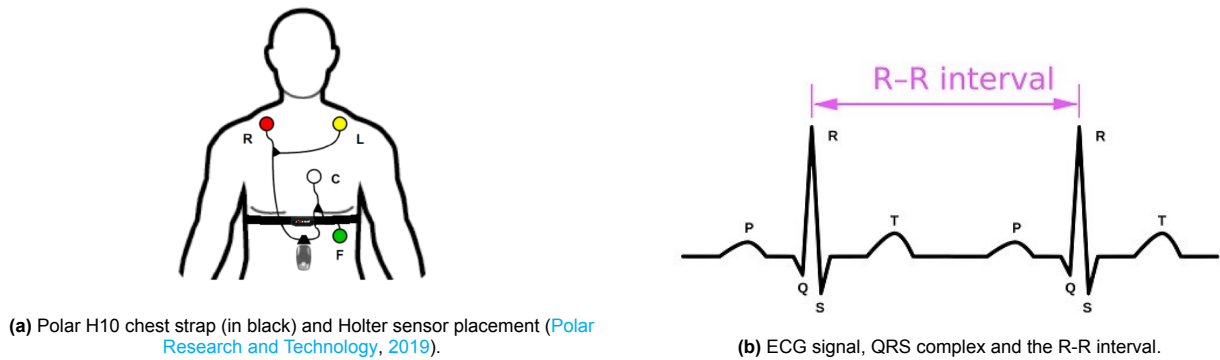


Figure 4.3: HR measurement with Polar H10.

The H10 finds the QRS complex of the person's heartbeat using a detection algorithm on the filtered ECG signal. From consecutive QRS complexes, one can determine the R-R interval, which indicates the time between two heartbeats. This output is filtered once again according to "physiological criteria" before being converted to a heart rate in beats per minute (BPM) (Polar Research and Technology, 2019).

An open-source program called Cardia¹ logs the received BPM and the timestamp with millisecond-accuracy on a laptop. Although heart rate variability (HRV) is an indicator of fatigue (Segerstrom and Nes, 2016), we cannot reliably get an HRV value from the heart rate measurements. We will therefore focus on the BPM and not use the relative timing of the measurements for our analysis. All lines of the logfile where a heart rate measurement is registered are appended to a Pandas data frame in Python. This will save the timestamp in `yyyy-mm-dd HH:MM:SS.xxx` format. Some seconds will see multiple registrations, especially at higher BPM. There are also some seconds without a heart rate registration. When combining data from different sources in our analysis, we want to use the timestamp on second-level accuracy because not all sources measure on sub-second intervals. Therefore, we preprocess the raw heart rate logs to return only one value per second and stitch the gaps of single seconds without a log line. As these measurements are uncommon in production or manufacturing systems, we resort to sports sciences research to find an appropriate method. A common approach is to sample heart rate data every 5 seconds (e.g., Barbero-Alvarez et al. (2008); Ingebrigtsen et al. (2012)), which we put into practice using a centred moving average of 5 seconds.

4.3.2. From sensor data to energy expenditure rate

Looking at a snippet of heart rate data for one of the participants, we see high variability in BPM. Some HRV is typical for healthy hearts, but we are only interested in sustained heart rate response for the purposes of this study. According to Jennings et al. (1981), an appropriate period when looking at sustained heart rate response would be 30 seconds minimum. With individual picks taking between 7.9 and 12.7 seconds in our estimation, this is also the case for our experiments; HRV within the time span of one pick is not of our concern. Another work by Phua et al. (2012) uses 2-minute intervals to calculate average HR for measurement method comparison.

¹Cardia version V1.2.0.1 <https://github.com/uwburn/cardia>

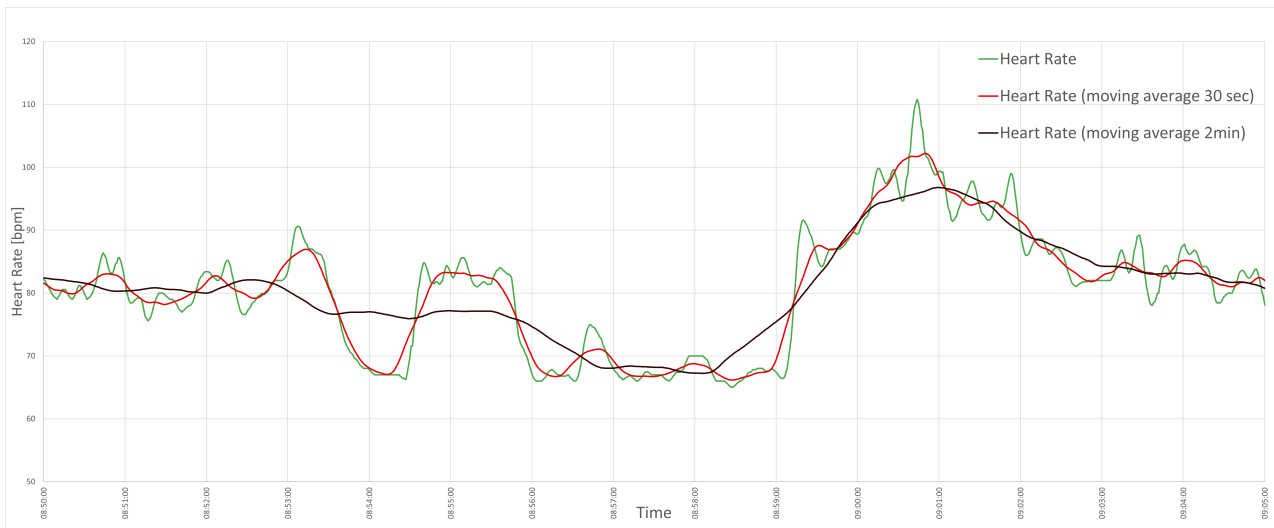


Figure 4.4: Comparing three different representations of heart rate data over time. The green line shows the unedited heart rate, the red line the 30 second centred moving average and the black line the 2 minute centred moving average.

Looking at these three different intervals in Figure 4.4, we see that the 30-second (centred) moving average retains most of the sub-minute detail without the high variability. In contrast, the 2-min (centred) moving average loses much of this detail, especially around fast increases in heart rate. To compare the pick- and HR-based $\dot{E}E$ estimation, we want these to have similar responses to activities, which we find possible with the 30-second interval. Going forward, the 30-second centred moving average is applied to the heart rate, %HRR, $\dot{E}E$ estimations and RA, always after calculation. If we use a different interval, this is specified. Since a moving average can only be calculated if data from the whole period is available, we need non-numeric filtering before our calculations to deal with the beginning and end of the data record. This is done in Excel using the *FILTER* and *ISNUMBER* functions.

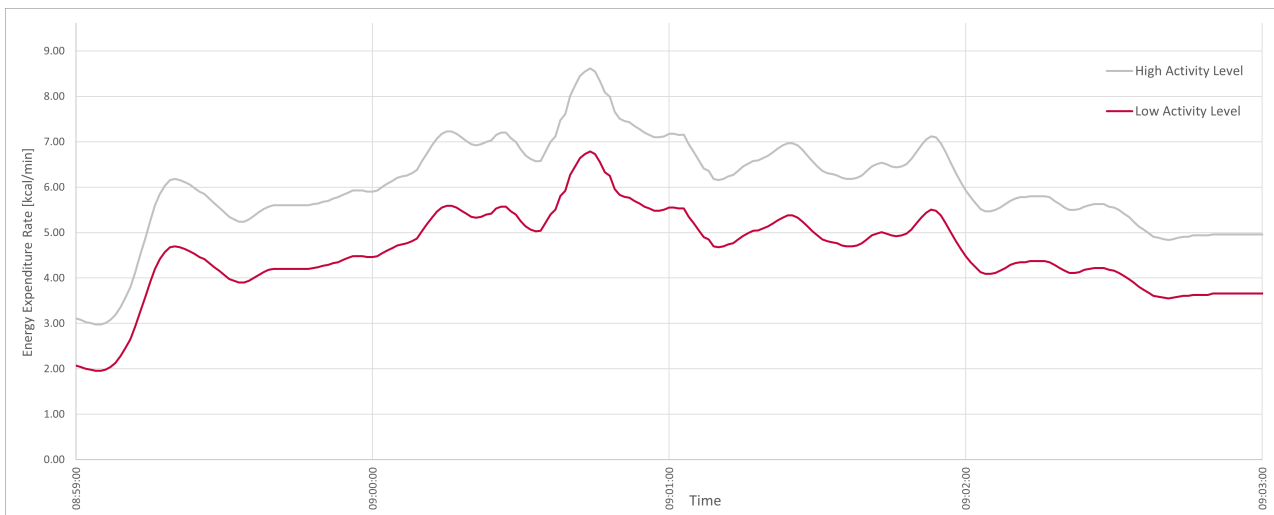


Figure 4.5: Estimation of energy expenditure rate for different activity intensity. Estimate values are generated using %HRR and [Hiilloskorpi et al. \(2003\)](#).

We go through a few steps from the heart rate data to get the estimation of $\dot{E}E$ at any moment. Using personal data from participants, (2.12) (this requires the measured minimum or resting heart rate (RHR) and estimated maximum heart rate HR_{max}), and mathematical models from [Tanaka et al. \(2001\)](#); [Hiilloskorpi et al. \(2003\)](#), we can calculate $\dot{E}E_{HR}$ from the momentary heart rate.

The previously introduced method by [Hiilloskorpi et al. \(2003\)](#) defines two different activity levels: high and low activity intensity. A snippet of the predictions of these two intensity levels is shown in Figure 4.5. For our

analysis, we want to evaluate only one HR-based $\dot{E}E$ estimation, which means we select the most appropriate level. The author describes the boundary region or 'flex point' between the two intensity levels to be in the range of 80-100 BPM; below 80 BPM the low-intensity estimation is most accurate, and above 100 BPM the opposite is true. We look at two methods to traverse the boundary region: a step and linear transition. Figure 4.6, illustrates the different approaches to assess activity intensity around the flex point.

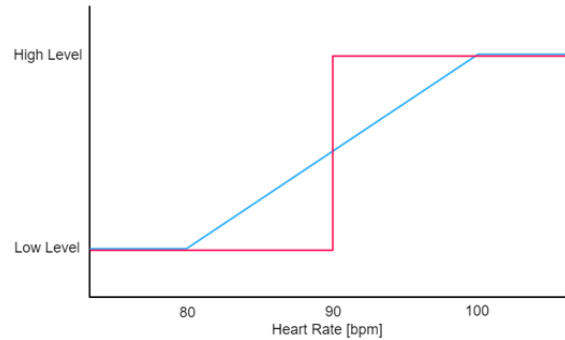


Figure 4.6: Visualisation of a step or linear transition between the high and low activity intensity level.

The step function ensures a direct transition between the low and high activity intensity at the 90 BPM mark. With the linear approach, there is no sharp flex point but a gradual flex 'region' between 80 and 100 BPM. This leads to less drastic jumps in the $\dot{E}E$ values, as shown in Figure 4.7. We deem this transition to be the more realistic variant of the two and use this going forward in the analysis.

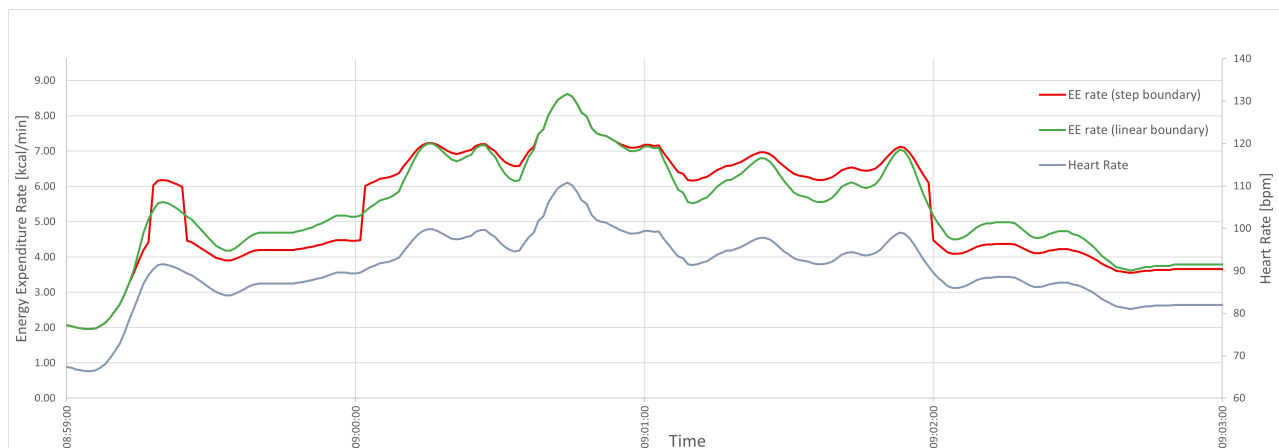


Figure 4.7: The behaviour of the two $\dot{E}E$ calculation approaches when the heart rate varies around the flex point identified by [Hilloskorpi et al. \(2003\)](#).

4.3.3. Order data

To estimate $\dot{E}E$ before the start of the experiment, we need an estimation of the items that will be picked. Using Picnic's WMS, we can extract the order data for that day for all stations. This data gives us a series of items to pick, which we feed into the task decomposition method in appendix C.4.1. Note that this method always assumes shoppers man a single station, which they fill in using the GUI before the start of the experiment. From there, we can calculate the average energy expenditure during the day, $\dot{E}E_{orders}$, which is a rough estimate.

4.3.4. Pick events

To get the $\dot{E}E_{picks}$ estimations, we track all picking data during the experiments using Picnic's WMS. This system registers pick events that are recorded when shoppers confirm a pick on the touchscreen. Pick events

contain a timestamp, tote ID, item ID, item quantity, item slot and station number. Using the separate article info database, we can get the article distribution weight with which we augment the pick event data (distribution weight is item content plus packaging). Although shopper ID numbers are also logged in the pick event data, shoppers only sometimes log in or out when switching stations between shifts, so this ID number can be inaccurate. Instead, we use participant numbers linked to the experiment day number and manually note when participants worked on which station.

With the pick event data, we used another Python script to extract personal information from the GUI logfiles and perform task decomposition on all individuals for the specific testing day. In this process, it is essential to note that station switching and double station working were common during the experiments. This means that, following our notes during the experiments, we need to group stations that were serviced in parallel by a single shopper into one pick event data file and then manually switch the pick data input on the timestamps that our notes describe shopper reallocation. However, before we can do this with the energy expenditure rate outputs of the pick event data, we need to calculate these for specific shoppers.

When given recorded pick events from WMS and shoppers' personal information, our Python script calculates the energy expenditure and pick time per pick event according to appendix C.4.1. When two consecutive picks happen by one shopper on different stations, we add a 'switch pick' EE and time on the timestamp between the two pick timestamps. This results in a list data frame containing timestamps, pick EE values and a predicted picking time. Since there is often system-induced idle time between picks, we cannot assume picking time to be equal to the time between two pick events. Instead, we will use the predictions from the task decomposition in appendix C. If we plot the timestamps and EE requirements, we get Figure 4.8a. This is still not useful to use as input for the analysis because it does not represent the $\dot{E}E_{picks}$ over time, nor is the placement of the arrows representative of the effort spent.

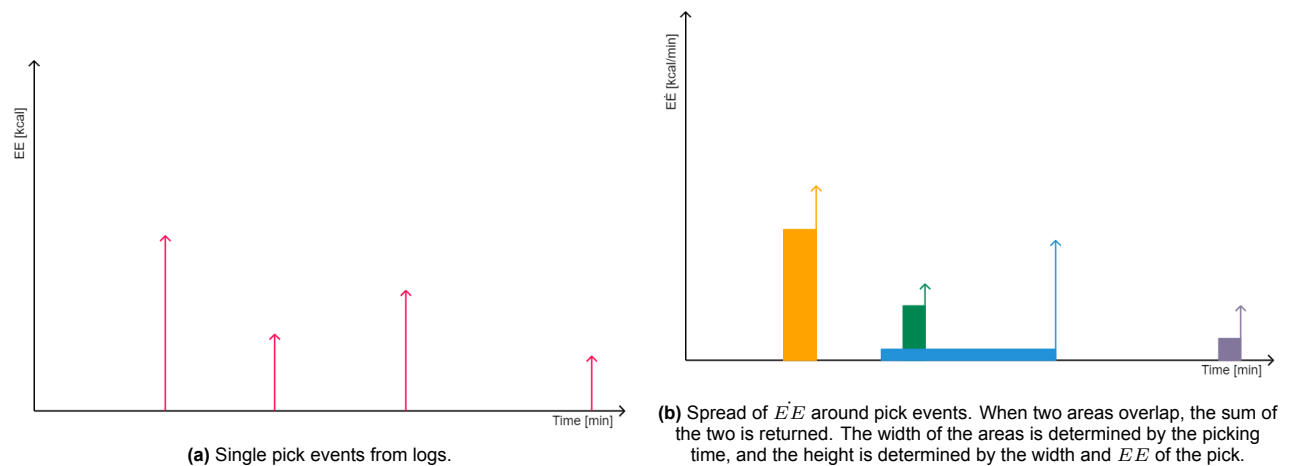


Figure 4.8: Pick event analysis.

Alternatively, we use the predicted picking time to calculate the average $\dot{E}E$ in the time before the pick event was registered. Then, we sum all of these for each second timestamp that we recorded to get the momentary effort during the whole experiment. A conceptual example of this is shown in Figure 4.8b. This can then be coupled to the timestamps in the heart rate data to compare the different methods for $\dot{E}E$ estimation. Picks registered during active break times in scenario 2 are not added to $\dot{E}E_{picks}$. Because resting while standing or sitting has non-zero $\dot{E}E_R$ (see (2.9)), $0.024BW$ is set to be the lowest possible value of $\dot{E}E_{picks}$.

Finally, all different data streams ($\dot{E}E_{HR}$, $\dot{E}E_{orders}$ and $\dot{E}E_{picks}$) are aggregated into one Excel sheet from which the analysis is done. Next to analysing the relative relation between the values through (average) deviations, we investigate if a general bias is present that influences the relationship between the values. We look at the correlation between $\dot{E}E_{HR}$ and $\dot{E}E_{picks}$ and use Excel's **CORREL** function to calculate the correlation coefficient between different values. This function calculates a number between 1 and -1, indicating how one function's behaviour relates to another. If two functions show highly similar behaviour, the correlation coefficient will be close to 1, while opposite behaviour results in values close to -1. If there is no relation between two functions and they behave irrespective of each other, the correlation coefficient is close to 0.

The most granular prediction of $\dot{E}E$ is based on the order data and is constant throughout the day. Sup-

pose we calculate a correlation coefficient between the value of $\dot{E}E_{orders}$ and $\dot{E}E_{HR}$ or $\dot{E}E_{picks}$. In that case, we compare a variable function to a constant one and would see a correlation coefficient of zero. Therefore, we only evaluate the correlation of $\dot{E}E_{HR}$ and $\dot{E}E_{pick}$. If these are both good estimations of $\dot{E}E$, their functions should match closely and have a high, positive correlation coefficient. We expect the estimations to show inaccuracies, but have a positive correlation. A situation in which we expect a near-zero correlation is during inactivity in the pick data and during breaks.

Another step to consider is the relational delay between effort and heart rate response. According to [Hammond and Froelicher \(1985\)](#); [Achten and Jeukendrup \(2003\)](#), this response is not immediate but can take up to 3 to 5 minutes following an increased workload or work rate. Therefore, we also compare $\dot{E}E_{picks}$ and $\dot{E}E_{HR}$ with the heart rate 3, 4 and 5 minutes ahead of time.

4.4. Schedule validation

With $\dot{E}E_{orders}$ known before the start of the experiments, we can calculate the appropriate RA using (2.6). To see the effects of this application, we divide the experiments into two blocks of approximately 1.5 hours each. Both scenarios are under everyday circumstances, with the exception of the heart rate monitoring and qualitative fatigue ratings during the experiments. Scenario 2 differs from scenario 1 because it applies the RA principle after every 20 minutes of picking. This break is held standing still at the station, as shown in Figure 4.9.



Figure 4.9: A shopper resting during the calculated rest allowance in scenario 2.

Using the qualitative fatigue ratings gathered by the GUI, the predicted and the measured $\dot{E}E$ estimates, we can compare the two scenarios for each individual.

5

Results

In this chapter, we present the results of this thesis. Section 5.1 presents the results from the modelling study. The results from the fatigue model and schedule validation in the empirical study are shown in chronological order in Section 5.2.

5.1. Modelling study

Because of the limited speed of our exact solver, we only run small virtual problem instances through our scheduling model. We find globally optimal solutions within 6000 seconds of run time until the number of decision variables exceeds roughly 2,500. This is already the case for problem instances with 3 totes, 3 order lines per tote, 3 stations and 2 shoppers, a problem size tiny compared to real-life problem sizes. However, even for these small problems, we can identify some key outcomes of our developed approach.

Our bi-objective optimisation is able to find non-dominated solutions, although only a few for each problem instance. This can provide granular estimations of the Pareto front in the discrete solution space. The exact shape remains unknown, but the handful of points in the Pareto front plots can already assist managerial decision-making. Note that one point in the Pareto front can be the solution space representation of several unique schedules with the same objective 'value pairs', as shown in Figure 5.1.

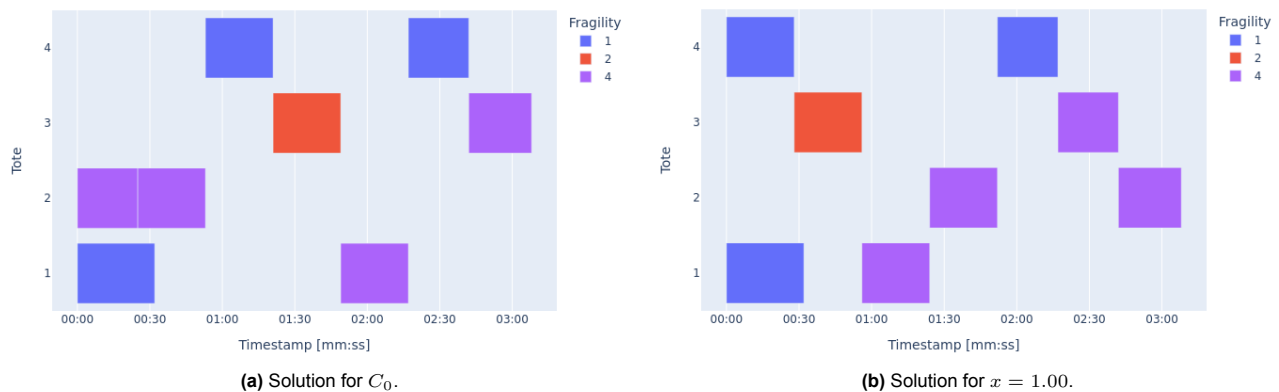


Figure 5.1: Two 'tote schedules' with the same objective value pair ($187.95s, 117.83kcal$). Each row in the Gantt chart represents the pick planning for the two order lines in a tote, while the colours indicate item fragility. The sequencing of picks is different in the two schedules, but both are feasible solutions.

5.1.1. Pareto front estimate with global optima

In Table 5.1 and Figure 5.2, we show the results for a problem instance with 3 totes, 3 order lines per tote, 3 stations and 2 shoppers, leading to 1995 decision variables and 2449 constraints. We get 7 unique objective value pairs and a maximum run time of 168 seconds. The highest values for C_{max} are found in the runs ' EE_0 ' and ' $x = 0.00$ '. The highest value for EE_{max} is found in the runs ' $x = 1.00$ ' and ' C_0 ' and is equal in both runs. Note that optimisation runs ' C_0 ' and ' $x = 1.00$ ' or ' EE_0 ' and ' $x = 0.00$ ' have essentially scaled versions of the

same single objective function. The lowest values for C_{max} and EE_{max} can be found in, among others, their respective single-objective optimisation runs.

Run	C_{max} [s]	EE_{max} [kcal]	Pareto-optimal
C_0	131.57	74.22	✓
EE_0	229.20	58.18	✗
0.00	261.68	58.18	✗
0.04	154.94	58.59	✓
0.42	149.63	60.11	✓
0.63	133.09	70.94	✓

Table 5.1: Unique solution value pairs.

In Figure 5.2, we draw the estimated Pareto front from our 27 optimisation runs, with the relative deviation of C_0 on the x-axis and that of EE_0 on the y-axis. Solutions as close to 0% deviation are preferred. In comparison to single-objective optimisation for the makespan, run C_0 , we can see that we are now able to find some alternative solutions. This allows us to make improvements in terms of EE distribution that single-objective optimisation would not have allowed for, such as a 10% better EE distribution against a 1.5% makespan increase when moving from the solution in run C_0 to the one in run $x = 0.63$. Using weight parameter x , management could decide upon the relative priority of these two objectives. We see comparable results for other small problem instances, with different shapes of the estimated Pareto front. Some examples are shown in appendix D.

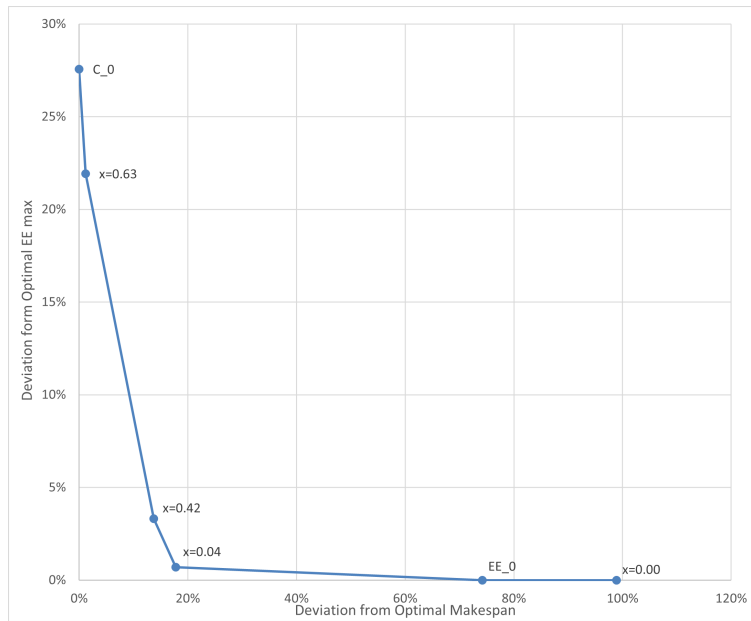


Figure 5.2: Estimated Pareto front with only globally optimal solutions. Each dot may represent multiple solutions, but only one objective value pair.

For all single-objective runs (C_0 , $x = 1.00$, EE_0 and $x = 0.00$), there is no guarantee that the solution is non-dominated, just that they indicate the absolute minimum for C_{max} or EE_{max} , respectively. Even if no dominating points are seen in the figure (this is the case for run C_0), we cannot guarantee Pareto optimality for any of these four runs, as they only optimise for a single objective.

This is not the case for all other solutions with $0.00 < x < 1.00$, as we know all solutions in the figure to be global optima for their respective problem instance. If either of the two objectives could improve without hurting the other - meaning the solution is dominated - we would not be in the global optimum for that optimisation problem, as this would lower the value of the combined objective. This would mean there are three non-dominated points in Figure 5.2. However, this figure does not show the underlying solutions that share the objective value pair with C_0 : $x = 1.00$, $x = 0.96$ and $x = 0.92$. Therefore, we can see four dots representing guaranteed Pareto-optimal solutions: C_0 , $x = 0.04$, $x = 0.42$ and $x = 0.63$. This also means

' C_0 ' is one of the utopia points.

If we compare the first Pareto-optimal solution in these instances to the original, single-objective optimum (c_0), we can calculate the steepness of the line piece between these two solutions (e.g., $c_0-x = 0.63$ in Figure 5.2). Table D.4 shows the resulting trade-off characteristics for 30 unique problem instances. For 53% of all problem instances with purely global optima we considered, this line piece has a steepness of 4 or higher, meaning the fractional decrease of EE_{max} is more than four times the fractional increase of C_{max} that accompanies it. The number is even as high as 699 in one case, where a 49% decrease in EE_{max} required a 0.07% increase of the makespan.

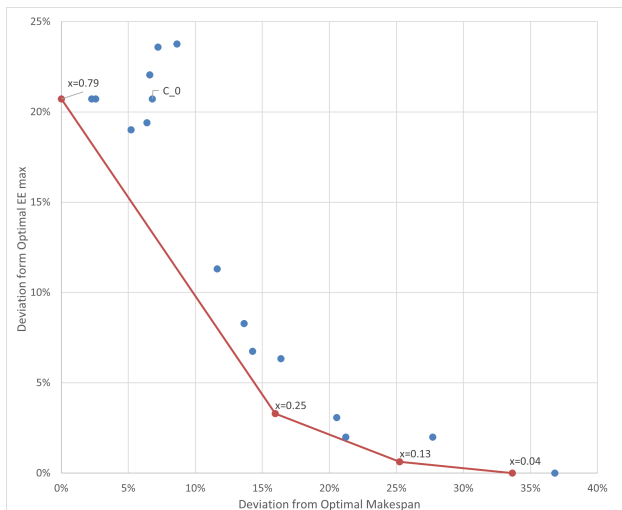
5.1.2. Pareto front estimate without global optima

For a problem instance with four totes, four order lines per tote, three stations and two shoppers, we get 6066 variables and 7531 constraints. The 27 optimisation runs result in 21 unique objective value pairs (C_{max}, EE_{max}) (see Table 5.2). None of the individual optimisation runs returned their global optimum within 3,000 seconds and had to be stopped early, except for ' EE_0 ' and ' $x = 0.00$ '. This means we have no guaranteed Pareto-optimal solutions. We can, therefore, only do a rough estimate of the Pareto front, using points that are non-dominated by any other row in Table 5.2 but may be dominated by a solution found after a longer run time.

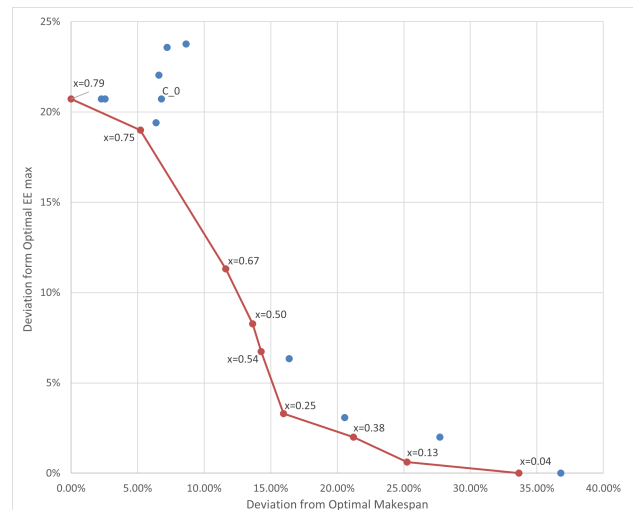
Run	C_{max} [s]	EE_{max} [kcal]	Pareto-optimal
C_0	230.47	121.83	\times
EE_0	802.63	100.92	\times
0.04	288.43	100.92	\times
0.08	295.27	100.92	\times
0.13	270.32	101.55	\times
0.21	260.18	104.03	\times
0.25	250.27	104.25	\times
0.29	275.63	102.93	\times
0.38	261.63	102.93	\times
0.42	251.18	107.32	\times
0.50	245.28	109.27	\times
0.54	246.64	107.72	\times
0.58	229.60	120.50	\times
0.63	220.75	121.83	\times
0.67	240.92	112.34	\times
0.71	221.35	121.83	\times
0.75	227.06	120.10	\times
0.79	215.83	121.83	\times
0.83	234.47	124.90	\times
0.92	230.07	123.17	\times
0.96	231.39	124.72	\times

Table 5.2: Unique solution value pairs.

The resulting Pareto front estimate is strictly a guess and can therefore have different shapes, as shown in Figure 5.3. What can also be seen in both the figure and the table is that run ' C_0 ' is dominated by other runs (e.g. ' $x = 0.71$ ') and does not provide us with the lowest possible value for C_{max} . As run ' EE_0 ' returns a global optimum, we know that 100.92 is the lowest possible value for EE_{max} , although it is dominated by $x = 0.04$.



(a) Estimated convex Pareto front without any guaranteed non-dominated solutions.



(b) Estimated concave Pareto front without any guaranteed non-dominated solutions.

Figure 5.3: Two different shapes for the Pareto front estimate. Red dots indicate Pareto-optimal points, with the red line representing the corresponding Pareto front estimate. Blue dots are other solutions that are sometimes not proven to be dominated and could therefore be part of the Pareto front.

5.2. Empirical study

We present the experimental results in order of calculation from the preprocessed data. This makes it easier to follow the relations and possible explanations in Section 6.

5.2.1. Heart rate and %HRR

In Figures 5.4a and 5.4b, we can see the distribution of age and resting heart rate among participants. We have tested a population sample (11 women, 9 men) of varying ages, skewed towards the 20-35 years range. Apart from some outliers, most measured RHR lie between 60 and 80 BPM, a normal range for healthy adults.

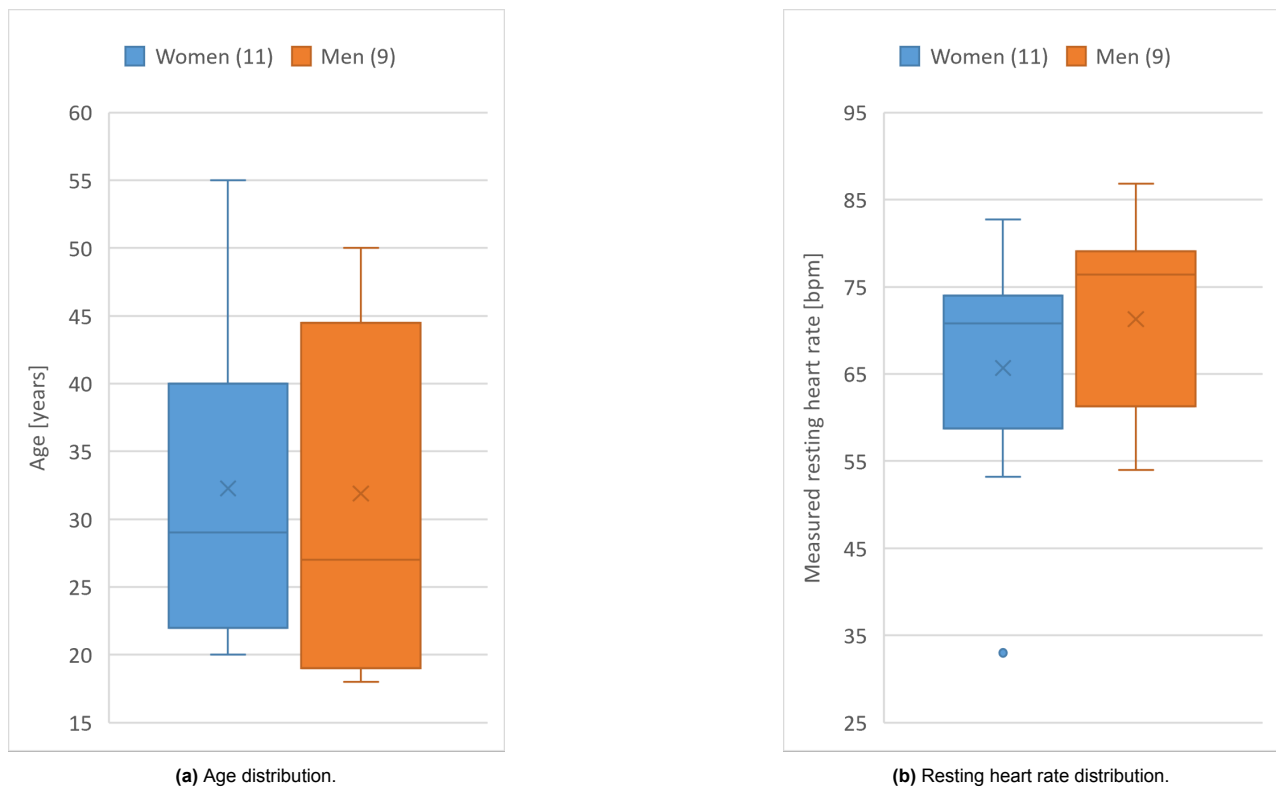


Figure 5.4: Distribution of participant characteristics relevant to the resting heart rate.

As the normal range differs per gender and age category, we also compare these values to population data from previous research. The population-based estimate of the RHR is generated by using shoppers' age and gender and linking them to population categories in the tables from [Ostchega et al. \(2011\)](#). In these tables, the authors present the mean and standard error¹ of the RHR for each population category. In Figure 5.5, we show how our sample relates to this population data regarding resting heart rate estimates. Individuals are grouped into bins representing the difference of their RHR compared to the population mean, expressed in standard deviations from that mean. Looking at the relative deviation from the population means in the figure, we see that the participant data centres around a value slightly below 0. Of all participants, 90% have an RHR within one standard deviation of the population mean.

¹The standard error is calculated by $SE = \frac{s}{\sqrt{n}}$ with s the sample standard deviation and n the sample size.

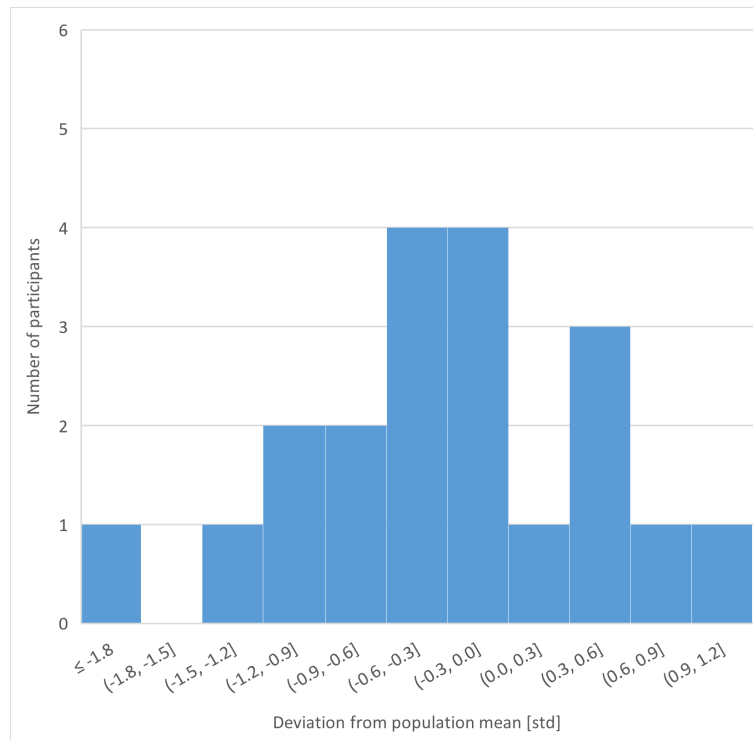


Figure 5.5: The distribution of the resting heart rate in our population sample compared to large population data. The deviation is in the unit of standard deviations from the population mean.

A first indication of the workload participants are exposed to can be found by looking at the distribution of %HRR values during the experiments. We divide this data into five intensity categories, each representing a 20% slice of the individual HRR. An overview of the average occurrence of each category among all participants is given in Figure 5.6.

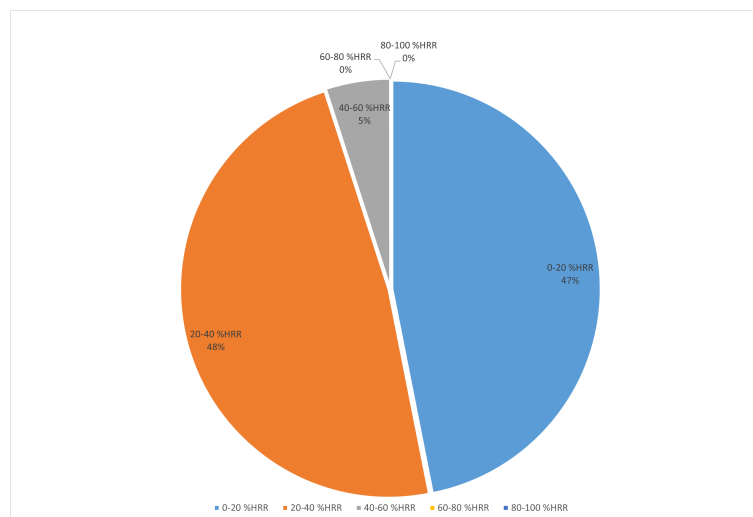


Figure 5.6: The %HRR level occurrence during experiments, grouped into five intensity categories.

%HRR values in the two lowest categories are most frequent during the tests. Only one participant reached a %HRR of over 60%, which sets the highest measurement across all participants and scenarios to 62%. However, these low %HRR levels do not mean that there is no exceedance of the maximum allowable energy expenditure.

5.2.2. Energy expenditure rate

In Table 5.3, we present the individual $\dot{E}E$ estimations for the entire experiment length. The energy expenditure rate is estimated in two ways: before and after the experiments. The values for $\dot{E}E_R$, $MAEE$ and $\dot{E}E_{orders}$ belong to the first category, $\dot{E}E_{HR}$ and $\dot{E}E_{picks}$ belong to the second.

Participant	$\dot{E}E_R$	$MAEE$	$\dot{E}E_{orders}$	Mean $\dot{E}E_{HR}$	Mean $\dot{E}E_{picks}$
1	1.61	2.96	3.95	3.68	2.39
2	1.42	3.13	3.53	5.16	2.41
3	1.51	3.33	3.85	4.11	2.49
4	1.63	3.53	3.82	2.87	2.39
5	1.39	3.60	3.58	2.38	2.22
6	1.56	3.61	3.97	4.07	2.45
7	1.61	3.68	3.92	4.90	2.07
8	1.49	3.70	3.78	3.86	2.14
9	1.42	3.76	3.77	3.34	2.53
10	1.54	4.16	3.77	2.59	2.14
11	1.70	4.57	4.70	4.88	2.52
12	2.40	5.20	5.30	7.55	3.72
13	2.06	5.23	5.57	8.28	3.15
14	2.09	5.37	4.99	5.78	3.11
15	1.80	5.48	4.32	3.20	2.27
16	1.92	5.78	4.85	3.75	2.80
17	2.38	5.96	5.27	4.14	3.37
18	2.04	6.74	5.17	4.05	3.74
19	2.04	6.81	5.26	5.28	2.58
20	2.52	8.32	5.96	6.03	3.46

Table 5.3: Resting, allowable, order-based, heart rate-based, and pick-based energy expenditure rate estimations per participant, presented in ascending order of $MAEE$. The estimations for $\dot{E}E_{HR}$ and $\dot{E}E_{picks}$ are given in the mean of all measurements during active experiments. Bold values indicate $MAEE$ exceedance.

In the table, we can see that the resting energy expenditure rate increases slightly with the $MAEE$, but that the difference between the two gets bigger for higher $MAEE$ values and that the relation differs per individual. Looking at the values in column $\dot{E}E_{orders}$, we see values ranging from about 3.5 to 6.0 kcal/min. This categorises as a low activity intensity in general. For reference, running at >16 km/h or cycling at >32 km/h are among the most vigorous activities in the overview by [American College of Sports Medicine \(2006\)](#) and would equate to an $\dot{E}E$ of 19 kcal/min for a person weighing 75 kg.

However, low activity levels can still exceed the $MAEE$, as shown by the bold values in the column. For 11 out of 20 participants, the predicted requirements exceeded the maximum allowable limits up to 0.98 kcal/min. Other estimates are below $MAEE$, meaning no RA requirements are predicted. The values for $MAEE$ and $\dot{E}E_{orders}$ are close to each other, except for participants with high $MAEE$ values (>5.00 kcal/min). Participants with low values for the $MAEE$ (<4.00 kcal/min) are most likely to be dominated by $\dot{E}E_{orders}$. We see a similar number of exceedances in column $\dot{E}E_{HR}$, although the values can be more extreme. For example, participant 13 has an $MAEE$ exceedance of 3.05 kcal/min, while participant 18 is 2.69 kcal/min below their $MAEE$ level. Finally, column $\dot{E}E_{picks}$ shows the least variance and lowest values overall.

Figure 5.7 shows visually the distribution and relative relation between the different estimates. Generally, we see that the individuals that are below the median in one of the box plots are also below the median in the others and vice versa, especially when comparing $\dot{E}E_{picks}$ and $\dot{E}E_{orders}$.

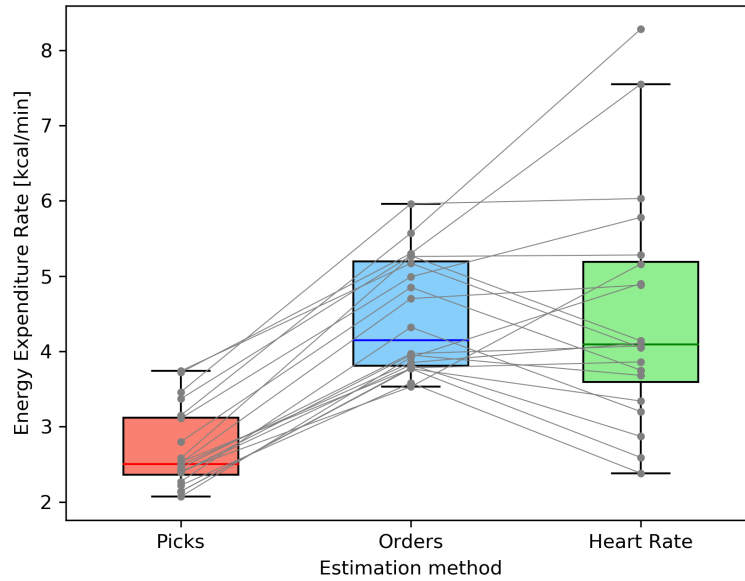


Figure 5.7: Three interconnected box plots representing the average values from the different estimation methods. Each line represents one individual during the tests.

As we treat $\dot{E}E_{HR}$ as the most accurate estimate of the energy expenditure rate in these circumstances, we evaluate the accuracy of $\dot{E}E_{orders}$ and $\dot{E}E_{picks}$ by comparing their estimates to this value. For $\dot{E}E_{orders}$, this results in a root root-mean-square error (RMSE) of 1.14 kcal/min, while $\dot{E}E_{picks}$ has an RMSE of 2.20 kcal/min. These values are in agreement with the visual representation in Figure 5.7.

In Table F.1, we present the correlation data of all individuals throughout the experiment. As can be seen, not every individual shows a high correlation between the two variables, with some even slightly negative. The lowest values also show low significance (high p-values). The highest correlation coefficients are seen mainly in the second column, as indicated by the coloured cells. The time shift approach for the correlation assessment, following the example from Achten and Jeukendrup (2003), does not seem to increase the correlation coefficients significantly. We see the same pattern in the average deviation figures.

To investigate the cause of these differences, we zoom in on the individual participant level. This also allows us to compare values in smaller periods to see where these differences may come from. We show a few examples that show reoccurring patterns in the data.

Individual EE rate comparison

A typical snippet of the three $\dot{E}E$ estimates is shown in Figure 5.8. First, we see that $\dot{E}E_{HR}$, shown as the dark red line, is not too far off from $\dot{E}E_{orders}$ in light red. This corresponds to the participant in Table 5.3, which showed minimal average differences between the two. The grey line represents $\dot{E}E_{picks}$, showing low, mostly singular peaks. These peaks result from individual picks, and the flat regions in between peaks correspond to $\dot{E}E_R$. The frequency of flat areas explains the low average value we saw in the table. This figure shows that the demand fluctuates a lot and does not require constant picking, as we would see few flat areas in the curve if there were a constant pick demand.

Looking at the shape of both graphs, there are some similarities in the graphs for $\dot{E}E_{HR}$ and $\dot{E}E_{picks}$, for example around 15:38:00. Contrastingly, the snippet was taken from scenario 2, and there was RA assigned between 15:36:00 and 15:38:00, where we see few different behaviour of the two plots. There was no break between 15:50:00 and 15:55:00, where we see large differences between the two graphs. Although no picks are registered during this time, the $\dot{E}E_{HR}$ shows a significant increase up to 11 kcal/min. During the entire snippet, $\dot{E}E_{HR}$ and $\dot{E}E_{picks}$ have a correlation coefficient of -0.02 with a P-value of 0.45, indicating low significance.

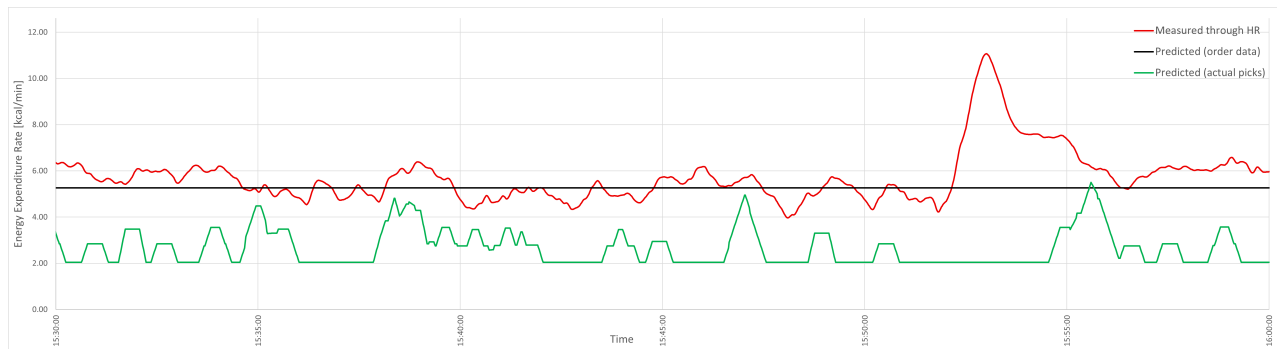


Figure 5.8: Snippet of $\dot{E}E$ estimates over time during scenario 2 for participant 19.

We can see similar patterns for another participant, as shown in Figure 5.9. Even though the averages of the three estimates vary, we can find more similarities in the $\dot{E}E_{HR}$ and $\dot{E}E_{picks}$ estimates than in the previous figure. We also see some brief moments where $\dot{E}E_{picks}$ exceeds $\dot{E}E_{HR}$, which is quite rare in the recorded data. For the time span in this figure, the correlation between these two is 0.34 and is significant at $p < 0.01$.

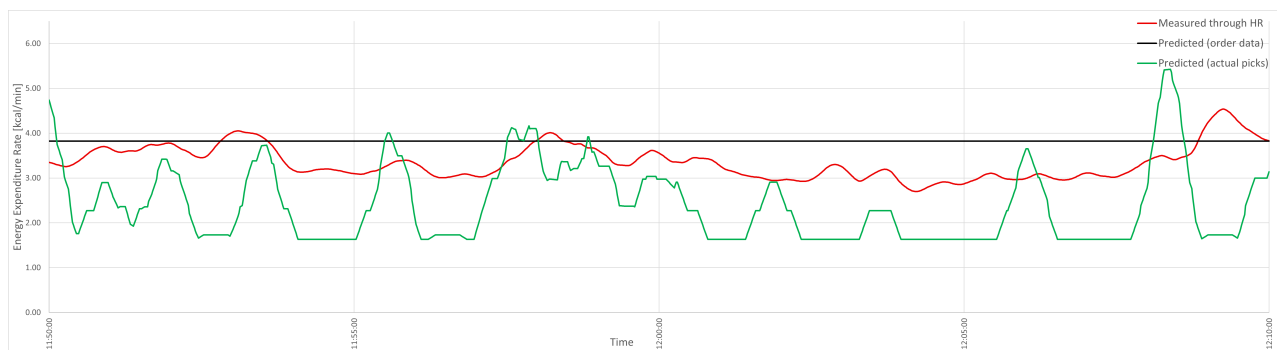


Figure 5.9: Snippet of $\dot{E}E$ estimates over time for participant 4.

In contrast with Figure 5.8, there are also cases where the result of RA periods is more clearly visible. In Figure 5.10, we see a third participant's response to RA periods between 9:56:00 and 9:59:00 and between 10:19:00 and 10:22:00. During these periods, we see $\dot{E}E_{picks}$ drop to resting levels and a less noticeable drop of $\dot{E}E_{HR}$ relative to the time before and after the breaks. Other than that, this figure shows higher pick demand than the previous two pictures, as well as a higher correlation between $\dot{E}E_{HR}$ and $\dot{E}E_{picks}$ (0.51, significant at $p < 0.01$).

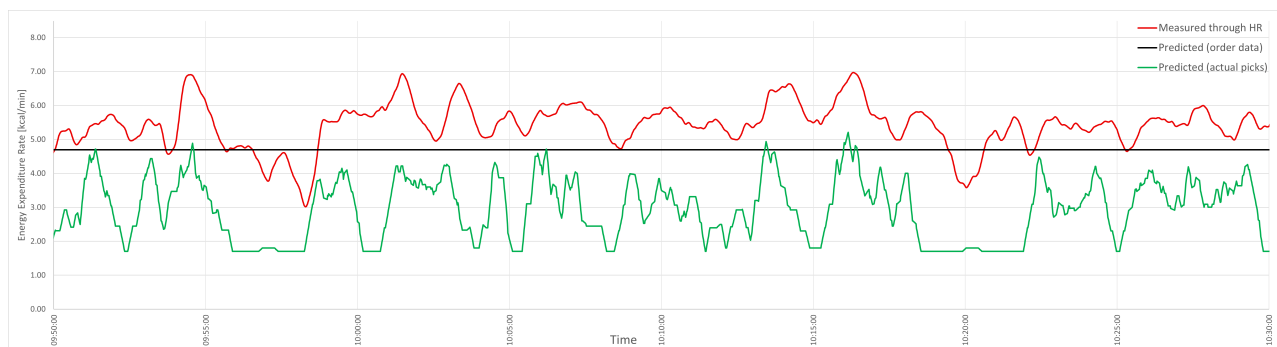


Figure 5.10: Snippet of $\dot{E}E$ estimates over time for participant 11.

Unlike the snippets we have seen so far, Figure 5.11 shows that the $\dot{E}E_{HR}$ value can drop considerably during tests. We see that the $\dot{E}E$ value that was measured through the heart rate, thus using the model by [Hiilloskorpi et al. \(2003\)](#), is lower than the theoretical $\dot{E}E_R$ value from [Garg et al. \(1978\)](#). This is one snippet, but this occurs in the data from other participants too.

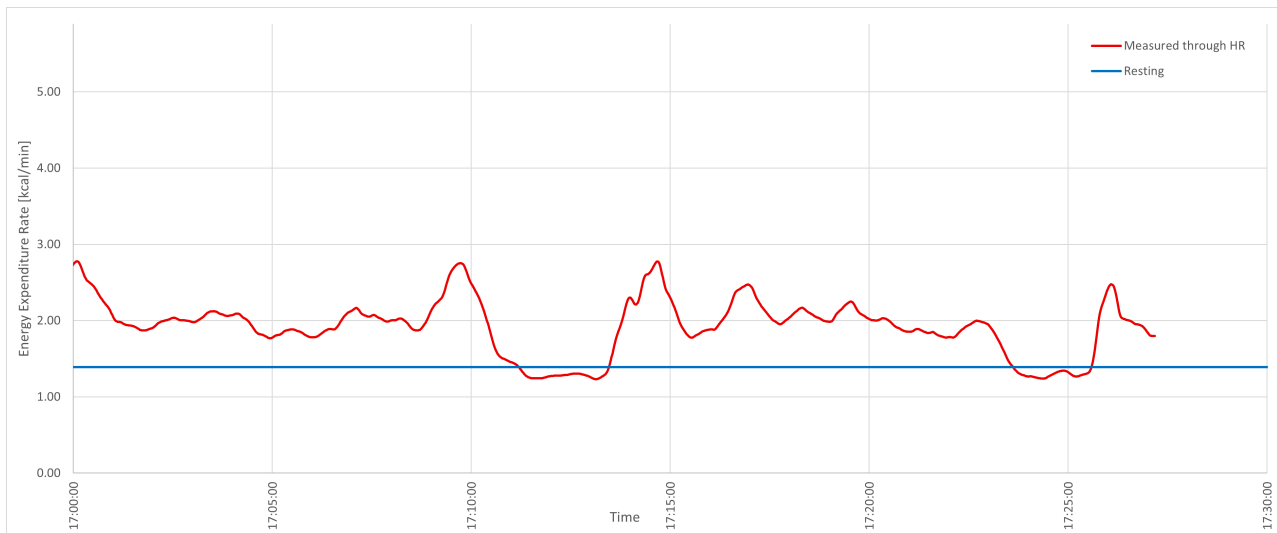


Figure 5.11: Snippet of $\dot{E}E_{HR}$ and $\dot{E}E_R$ over time for participant 5. From 19:12:00 to 19:15:00, there was a period of RA, and the experiment ended at 17:27:00.

Separating the data into 10 minute intervals

The pick demand in individual figures shows high fluctuation in the snippets for individual participants. This fluctuation exists between participants, within the experimental data of a single participant, and between the two scenarios for a given participant. In Figure 5.12, we present the occurrence of pick demand per 10 minute interval across all recorded data, clarifying that these differences also occur on the average level for 10 minute periods. We see varying numbers, but most values are below 60 picks per 10 minutes. According to our model's estimated picking time of 7.9 to 12.7 seconds, this would be the average number of picks per 10 minutes that a shopper could process.

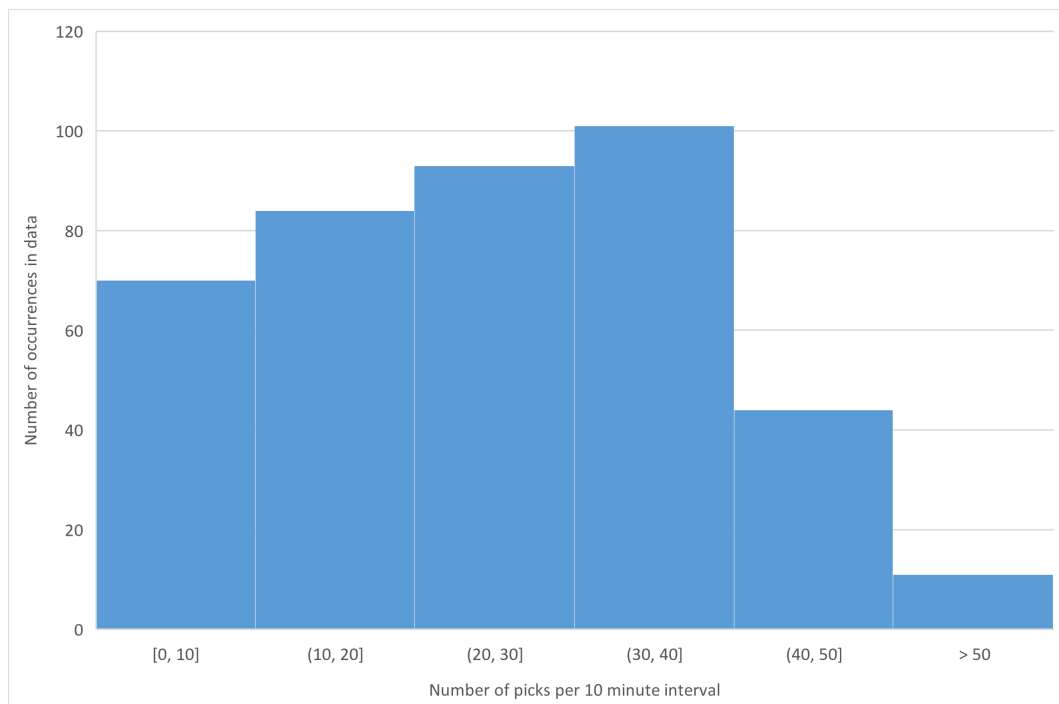


Figure 5.12: Overview of all pick demands per 10 minute interval for all participants. The data has been filtered for periods with constant heart rate registration for 10 minutes, then categorised into bins of picks per 10 minute intervals.

These results highlight the importance of considering distinct intervals in evaluating the $\dot{E}E$ estimates. To

this end, we have analysed 10 minute intervals for individual participants, from which we present two examples.

In Figure 5.13, we show the result of this analysis for participant 20. This overview shows the high variability in pick demand per 10 minute interval. The majority of intervals contained a low workload with up to 20 picks per 10 minutes. Since the predicted average pick time is about 10 seconds, the shopper was busy picking an item up to only a third of the time. Correlation coefficients range from -0.11 to 0.53, although most are positive. This seems to suggest that the pick-based prediction is not too far off, but the bias is an important factor to take into account here too.

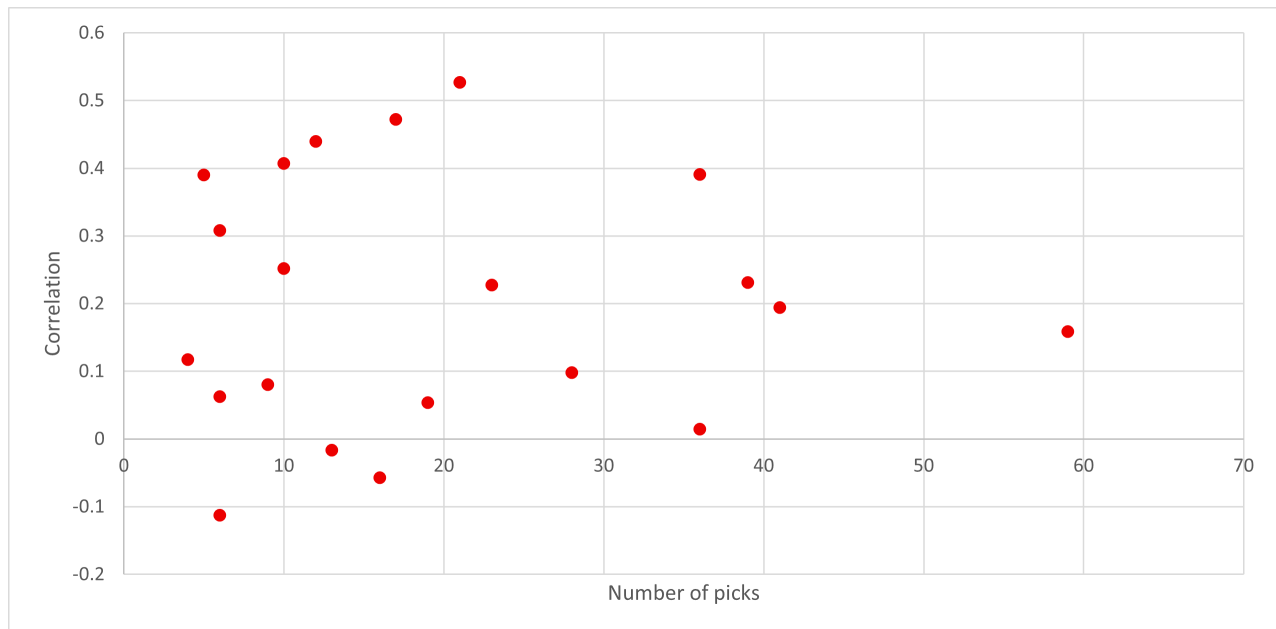


Figure 5.13: Overview of all 10 minute intervals for participant 20. Each dot represents one measurement interval, with the number of pick events on the x-axis and the $\dot{E}E_{HR} - \dot{E}E_{picks}$ correlation coefficient on the y-axis.

On average, for the entire measurement, as shown in Table 5.3, the bias between $\dot{E}E_{HR}$ and $\dot{E}E_{picks}$ is -2.29 kcal/min. This means that, on average, the pick-based estimation is too low. However, when we look at the 10 minute intervals again, we see this bias is much higher in some periods. We plot the correlation and deviation in Figure 5.14. Note that this deviation is expressed as the deviation between the 10 minute average values of $\dot{E}E_{HR}$ and $\dot{E}E_{picks}$, which is fundamentally different from the average deviation between the two variables during that period. For scheduling purposes, the average values of $\dot{E}E_{HR}$ and $\dot{E}E_{picks}$ are valuable KPIs as they relate to the starting point for RA calculation. Therefore, deviation from the interval mean is influential to scheduling.

Points in the top part of the figure and close to the y-axis indicate excellent predictions, as these have low bias and high correlation. These points would lead to accurate RA predictions. The figure shows a few of these points, but many have a deviation of more than -2.00 kcal/min, with some even near 6.00 kcal/min. This means that the predicted $\dot{E}E$ value is much lower than the actual value, and the appropriate RA requirements would not be met with this estimation.

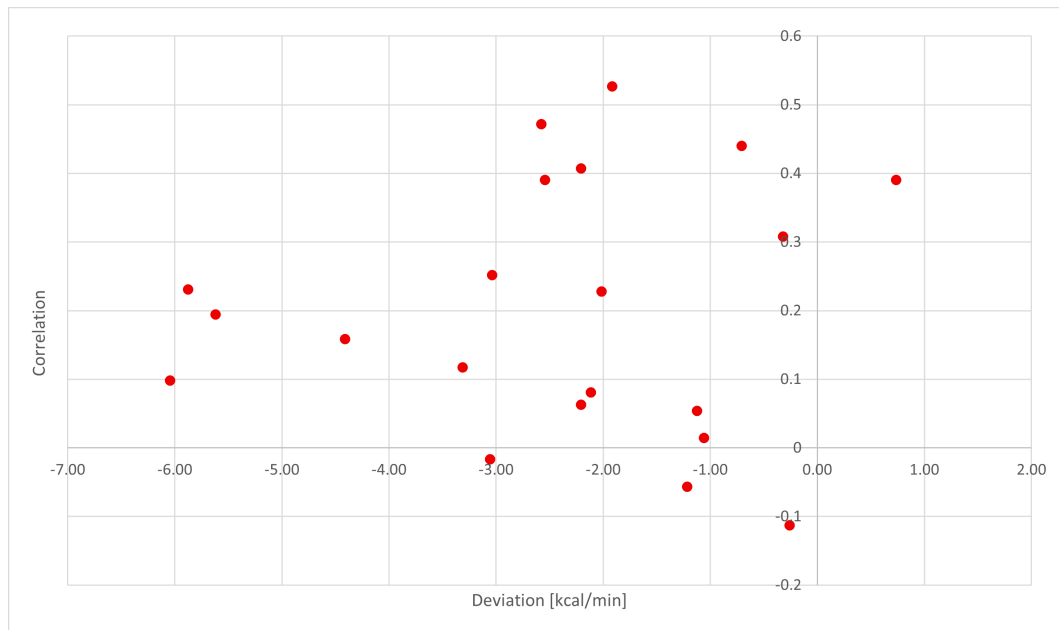


Figure 5.14: Overview of all 10 minute intervals for participant 20. Each dot represents one measurement interval, with the average $\dot{E}E_{HR}-\dot{E}E_{picks}$ deviation on the x-axis and the $\dot{E}E_{HR}-\dot{E}E_{picks}$ correlation coefficient on the y-axis.

The data from participant 16 in Figure 5.15 shows a different outcome, as the 10 minute intervals have a higher average picking workload (24 picks per 10 minutes) with less variation than participant 20. The correlation coefficients are generally also higher, although some low and negative correlations are still found.

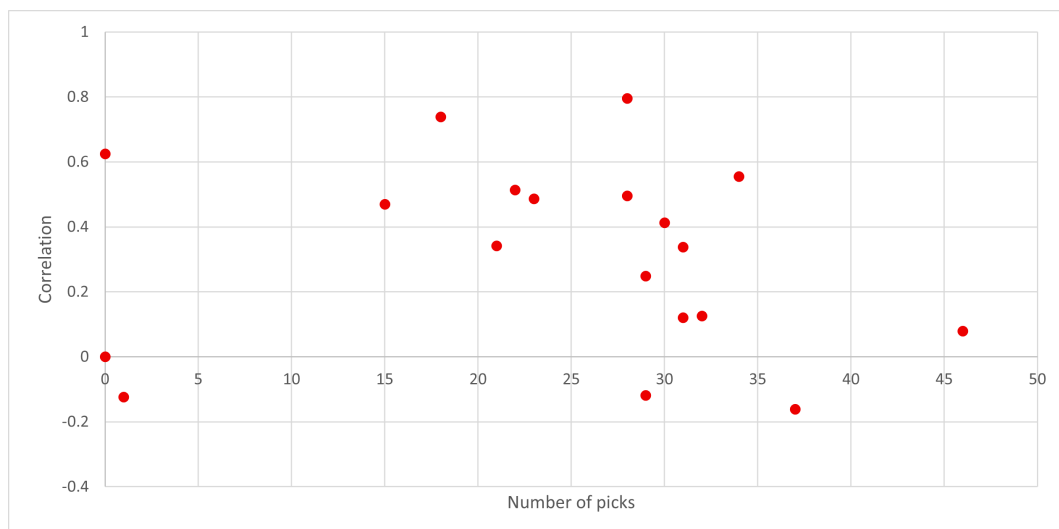


Figure 5.15: Overview of all 10 minute intervals for participant 16. Each dot represents one measurement interval, with the number of pick events on the x-axis and the $\dot{E}E_{HR}-\dot{E}E_{picks}$ correlation coefficient on the y-axis.

This participant's average bias is -2.03 kcal/min for the whole experiment, but the individual interval deviation is much smaller. Figure 5.16 presents the overview of correlations and deviations for the individual 10 minute intervals, which shows several positive correlations that also have low deviations. These estimates are, therefore, more promising for application.

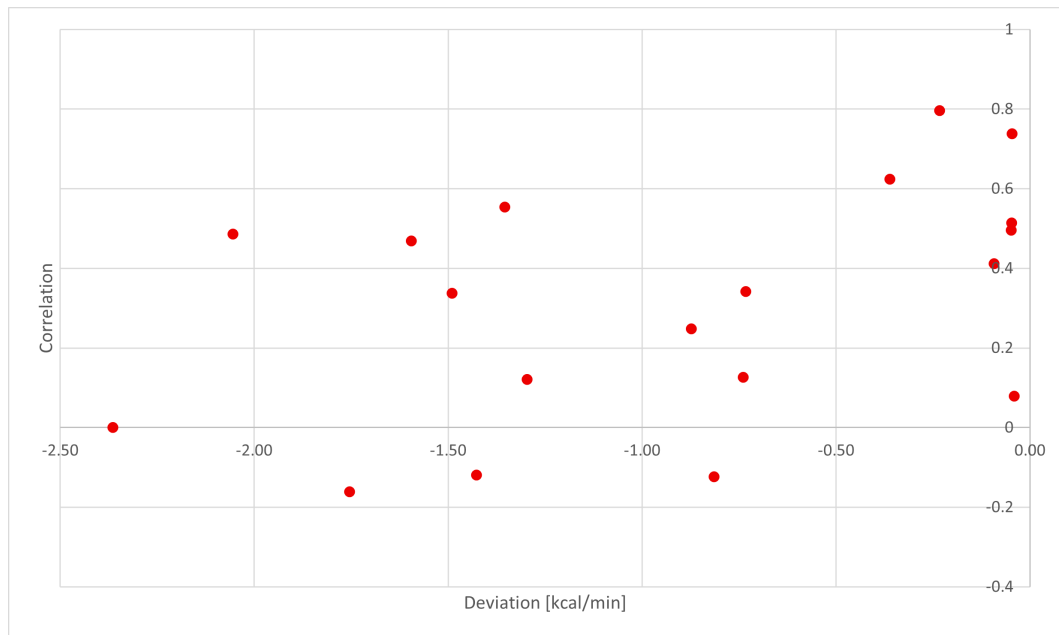


Figure 5.16: Overview of all 10 minute intervals for participant 16. Each dot represents one measurement interval, with the average $\dot{E}E_{HR}-\dot{E}E_{picks}$ deviation on the x-axis and the $\dot{E}E_{HR}-\dot{E}E_{picks}$ correlation coefficient on the y-axis.

Although these two examples show some of the variability in the data, they can be misleading at a population sample level. We point out this argument in Figure 5.17, where we combine all individual data in one big plot. Although some good estimates exist, the vast majority has a negative deviation of more than 1 kcal/min and we also see a high number of intervals with low or negative correlation coefficients. Figures F.1, F.2 and F.3 show the same figure with a distinction between pick demand bins, separating the data at respectively 10, 20 and 30 picks per 10 minutes. These figures show no clear link between pick demand characteristics and estimation quality, which is what the two individual examples could have led us to believe.

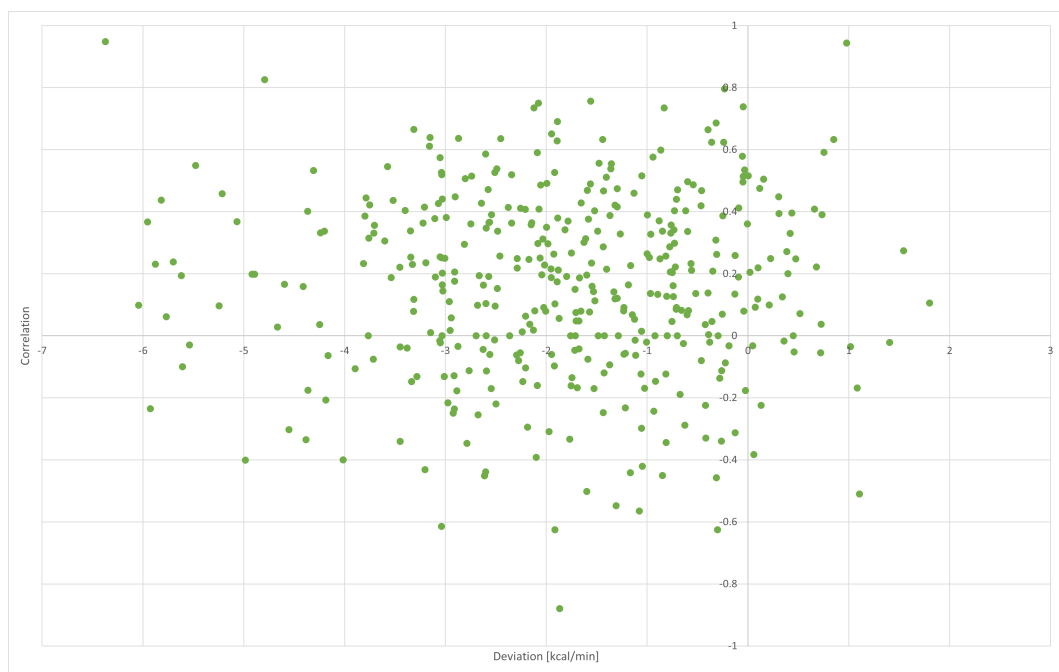


Figure 5.17: Overview of all 10 minute intervals for all participants. Each dot represents one measurement interval, with the average $\dot{E}E_{HR}-\dot{E}E_{picks}$ deviation on the x-axis and the $\dot{E}E_{HR}-\dot{E}E_{picks}$ correlation coefficient on the y-axis.

5.2.3. Rest allowance

The rest allowance is the most useful measure to relate fatigue effects to the scheduling problem, as it can be used as a direct input for scheduling breaks. Also, if the average rest allowance can be lowered by reassigning tasks to shoppers, this can be an influential factor for shoppers' productivity and fatigue buildup. We overview the different rest allowance values calculated from the $\dot{E}E$ estimations in Table F.2.

The second column shows the values based on $\dot{E}E_{orders}$, which are only positive if that value exceeds the $MAEE$. This shows the same participants to require RA as we have seen in Table 5.3 to exceed the $MAEE$. The same similarity is seen for the shift-average RA requirements from $\dot{E}E_{picks}$ and $\dot{E}E_{HR}$, columns 3 and 4, respectively. The table shows an uneven spread of the rest allowance among participants. This is especially true for the heart rate-based estimates, which go up to 124%. This means that, in practice, this participant would require more time on break than working. When comparing these values to Table 5.3, we see that the highest exceedance of the $MAEE$, which was for participants 12 and 13, does not correspond to the highest value for the RA. This is the result of the personal characteristics considered when calculating the RA and means that not all high exceedances lead to high RA values.

5.2.4. Qualitative fatigue ratings

In conjunction with the quantitative measurements, we also gathered qualitative feedback during the tests. For example, Figures 5.18 and 5.19 show the relation between $\dot{E}E_{HR}$ and the RPE scores, where we make the distinction between scenario 1 and 2. The energy expenditure rate is shown in a 20-minute centred moving average to better relate to the low frequency of the RPE scores. Although similarities in the trend of the two KPIs can be seen in Figure 5.18, this is an exception for the entire participant pool. Most figures, like 5.19, show much more variance in $\dot{E}E_{HR}$ than in the RPE scores. No significant similarities can be seen in general.

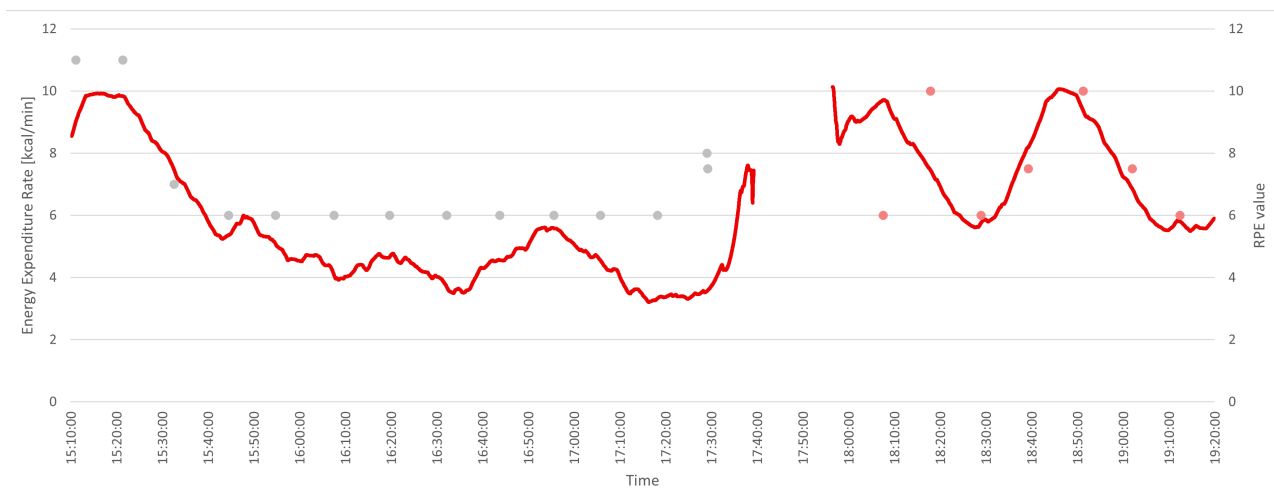


Figure 5.18: Participant 20's $\dot{E}E_{HR}$ values (red line) versus RPE scores for scenarios 1 and 2 (pink and grey dots, respectively).

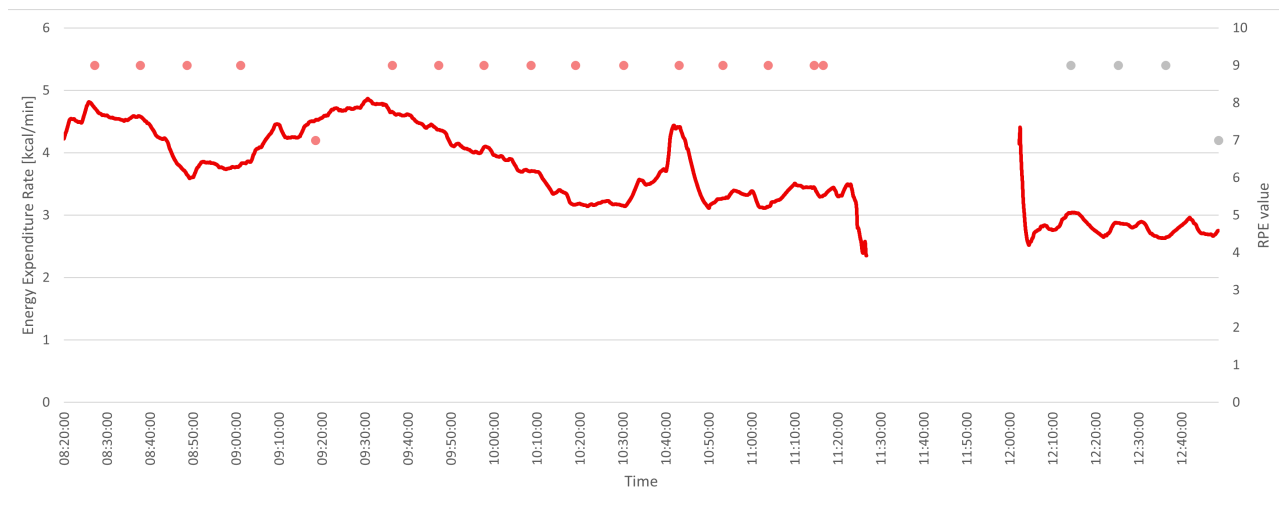


Figure 5.19: Participant 16's $\dot{E}E_{HR}$ values (red line) versus RPE scores for scenarios 1 and 2 (pink and grey dots, respectively).

If no relationship can be found within the ratings of one person, it could be interesting to see how participants compare to each other. In Table F.3, we compare the personal average RPE value, a score derived from the NASA TLX forms and the RA value from the heart rate measurements. We choose the latter over the energy expenditure rate because it shows the relative overload compared to the $MAEE$ and $\dot{E}E_R$.

The table shows no conclusive relationship between the three KPIs. The median value for RPE ratings would be 13, while it is 30 for the NASA TLX score. Keeping that in mind, we see few similarities between these two columns. If we also consider the RA column, the relationship between the three becomes even less visible. For example, participant 15 has one of the lowest RPE scores, one of the highest NASA TLX scores and 0%RA. Overall, no significant relationship between the two scenarios and the resulting quantitative or qualitative KPIs is to be seen, nor do we see differences in the average pick demand, although the provided RA was between 8-24% during the experiments. This value does not correspond with the second column of Table 5.3, because a bug in the GUI code provided erroneous RA percentages to the GUI timers during the experiments.

6

Discussion

This thesis explored how worker fatigue and productivity can simultaneously be considered in job-shop scheduling for a partially automated order picking system. We discuss the results of the modelling study in Section 6.1 and those of the empirical study in Section 6.2. Relevance to the research field and operational management practice are discussed in Sections 6.3 and 6.4, respectively. Finally, we suggest opportunities for future work in Section 6.5 and managerial recommendations in Section 6.6.

6.1. Modelling study

6.1.1. Implications of the results

The results in Section 5.1 show that we can successfully generate schedules for the considered OP system. We are able to draw a rough estimate of the Pareto front for our bi-objective optimisation problem as long as the solver returns a guaranteed global optimum. This is an expected result given our weighted sum approach (Becerra and Coello Coello, 2006). Even though our methods only allow for a few unique non-dominated solutions to be found, we can improve the energy expenditure distribution compared to single-objective optimisation solution ' C_0 '. This choice would depend on managerial insights, but the steep Pareto front line pieces that we found indicate considerable potential for implementation. This is in accordance with our hypothesis.

In every problem instance we present in Section 5.1 and appendix D, we find high values for C_{max} for runs ' EE_0 ' and ' $x = 0.00$ '. These seem out of order in comparison to the rest of the objective pairs that were found (see Figure D.2) because these runs do not care about the value of $C_{i,j}$ except for the broad estimate in (3.26).

The actual returned value for C_{max} from the solver is even higher, a technical artefact from the solver. Because the optional calculations with decision variables in SCIP are limited - we have already discussed this in Section 3.2.1 about the standard deviation of $EE_{i,j,l}$ - we cannot directly calculate the value of C_{max} from all individual $C_{i,j}$. Instead, we create an additional decision variable representing C_{max} and constrain it to be equal to or larger than all $C_{i,j}$. We do the same for EE_{max} and $\sum_l \sum_j \sum_i EE_{i,j,l}$, which is also why the number of DV varies between Section 5.1, Appendix D and Table 3.2. SCIP's solving process initiates these two DV with very high values, which never decrease if C_{max} is not part of the objective function, meaning we need to calculate these by hand.

Although we use the same method for EE_{max} , we do not see the same thing in runs ' C_0 ' and ' $x = 1.00$ '. This can be explained by the fact that this value is tightly connected to the minimisation of C_{max} ; short schedules divide the work equally over all shoppers and unintentionally divide the workload. We may not reach optimal values for EE_{max} in those two runs, but they are in the same order of magnitude.

For the problem instance presented in Section 5.1.2, we see that only runs ' EE_0 ' and ' $x = 0.00$ ' return a global optimum within our time limit. This can be explained by the fact that the principal restricting decision variable to their objective is $\alpha_{i,j,k,l}$, a binary decision variable. All other runs rely on decision variable $S_{i,j}$, a continuous decision variable. This means there are many more options for SCIP to review in the solution space, leading to longer solving times that exceed the time limit, therefore not returning a global optimum.

6.1.2. Limitations

First, our bi-objective optimisation approach is unsuitable for finding a good Pareto front estimate that is fit for providing managerial insights in application. Given the limitations of our exact solver, this is not a problem that another bi-objective optimisation approach would have solved. Now that the principle behind our mathematical formulation has shown to be valid, future work could improve practical applicability by considering a different solving approach that can exhibit the model's full potential in realistic problem instances.

Also, we choose to model energy expenditure, the precursor of fatigue, and therefore not fatigue itself. The model's inputs are based on population data from 1981 which shows inherent flaws, and we look at its output as an absolute value. Looking at energy expenditure or fatigue overload may be more respective to an individual's well-being than the total sum we included in the bi-objective. This relates closely to an ethical question regarding equity and equality: do we really want to schedule heavier tasks to people that can handle them better? And, if so, do we then look at absolute or relative fatigue overload? Both questions are relevant for implementation and offer an interesting direction for future work.

Furthermore, some modelling simplifications were made: every station is the same, there is no tote travel time between stations, a constant picking speed, and no system downtime. These elements could be added, but they represent complex mechanisms and are difficult to apply in the JSP formulation.

Finally, our model does not return a schedule that can directly be applied in real life but shows that this is possible with the existing modelling techniques. One specific example is the approach for break scheduling, as applying the rest allowance after each pick is very impractical in real life. Partial fatigue alleviation is not possible with our approach, although it has already been researched in a different context (Jaber et al., 2013). Still, for a conceptual study like ours, this is a big improvement over the existing works and opens up many opportunities for future work.

6.2. Empirical study

6.2.1. Implications of the results

In Section 5.2.1, we find that our population sample is spread evenly between men and women of different ages. Most participants have resting heart rates that fall within one standard deviation of their predicted value, meaning that no inherent bias is expected in the data due to the population sample. We see that the participants generally do not exceed %HRR levels of 60%. This means none of the participants reached %HRR values that indicate anaerobic dissimulation, which is at 80% (Calzavara et al., 2018b).

Table 5.3 shows that predicted values for $\dot{E}E_R$ and $MAEE$ show to increase together, albeit not with the same incline. This can be explained by the fact that their respective equations, (2.9) and (2.8), contain an element of an individual's body weight. The latter, however, also uses gender, age and other constant values, which results in a different outcome. The interpersonal difference between these two values makes that similar $MAEE$ overload values can lead to different RA requirements. This difference poses a task assignment problem where total RA requirements are minimised, which is what our mathematical model in Section 3.2 is capable of.

The interpersonal difference in values for $MAEE$ in our population sample shows significant differences with the approach by Battini et al. (2016), who use the standard value of 4.3 kcal/min. This shows the value of assessing this value individually when dealing with a heterogeneous workforce.

In Table 5.3, we also see the widely different estimations of $\dot{E}E$ during the experiments. Although $\dot{E}E_{orders}$ is a much rougher estimation of the work to be done during the experiments, it is often a much better estimate on a shift level. The $\dot{E}E_{picks}$ estimation has a consistent negative bias with respect to $\dot{E}E_{HR}$. Even the correlation between $\dot{E}E_{picks}$ and $\dot{E}E_{HR}$, which ignores any inherent bias in the estimation, does not show a clear validity in the $\dot{E}E$ estimations. This is the case on a shift level but also for 10 minute intervals. A first valid suspicion could relate these inaccuracies to the highly volatile pick demand, but our results show no clear relation between the pick demand and estimation accuracy. This contradicts our expectations of the estimation method's performance.

Shoppers do much more than just order picking, such as pallet loading, cleaning and opening boxes. We estimate that about 25% of the shift is not spent according to our task decomposition description. However, the task intensity of the other tasks, we believe, is similar to that of picking. This would explain the better estimations of $\dot{E}E_{orders}$, as this assumes a continuous picking workload. Furthermore, many operational challenges, such as system downtime and shopper reassignment, occurred during the experiments, troubling the experimental circumstances.

These factors could also have influenced the schedule validation results, where we see no clear differences between scenarios or the two groups of participants in terms of $\dot{E}E_{HR}$ or the qualitative fatigue ratings. Therefore, we cannot conclude that the positive results from our modelling study translate to real-life applications.

6.2.2. Limitations

First, we find the pick volume to be very low and have high variability between participants, between scenarios and within scenarios. This is in line with our operational observations during the experiments. FCA is by no means a lab environment, and it is difficult to predict or even measure the actual work done by shoppers; station switching is common, pick demand fluctuates a lot, and there are long periods of inactivity due to operational issues. This also influences the capabilities of $\dot{E}E_{orders}$, as this calculation takes single station demand data, while it was rare for shoppers to only service one station during their work.

The significant demand fluctuation and uncertainty do not allow us to make conclusive statements on the comparison between scenarios or participant groups, as was one of the goals of this experiment. The influence of the demand fluctuation cannot be separated from the possible influences of the breaks or different scenario order, especially if there is so little pick demand that there is often involuntary resting outside the scheduled RA breaks. Also, given the unpredictable nature of the demand, there is a low likelihood that the RA estimations could be accurate enough to create the working circumstances that scenario 2 requires.

The picking task, as it has been described in the task decomposition in appendix C, requires some assumptions of the working environment that can influence the model outcome, such as picking height, walking distances, characterisation of movements and the picking time. In Figures F.4 and F.5, we show some examples of the invalidity of these assumptions. However, we argue that this would most influence the relation between the height of the energy peaks in the $\dot{E}E_{picks}$ and $\dot{E}E_{HR}$, while this does not explain the absence of a peak in the pick-based estimation. For example, Figure 5.8 shows two break periods where no picks are done. We see values drop slightly in $\dot{E}E_{HR}$, but also some peaks that do not correspond to picks. Although a gradual decrease and increase of the $\dot{E}E$ are to be expected (Calzavara et al., 2019), these peaks are not.

While we know that the estimations of $\dot{E}E_{picks}$ and $\dot{E}E_{orders}$ are not perfect, the large differences in RA values in Table F.2 cannot only be attributed to that. Especially if we consider that the most extreme RA-HR values do not correspond to higher RPE or NASA TLX scores, these high values seem out of order. This can have a variety of causes: external influences on the heart rate, too high $\dot{E}E_R$ calculation (Garg et al., 1978), too low $MAEE$ values (De Souza E Silva et al., 2016), too high HR_{max} (Tanaka et al., 2001), or too high calculation of $\dot{E}E_{HR}$ (Hiilloskorpi et al., 2003). These four models rely on population averages that may not apply to a small sample of individuals (Achten and Jeukendrup, 2003) or may have become outdated. With only 20 participants partaking in the experiments, this could have led to inappropriate assumptions and calculations, leading to inaccuracies in not only the predicted $\dot{E}E$ but also the measured values.

6.3. Relevance to human factors and operations research

Our thesis attempts to bridge the gap between theoretical scheduling problem approaches and operational reality in the context of human fatigue. Existing literature often simplifies the human factors model (Price, 1990; Othman et al., 2012; Yazdani et al., 2015; Berti et al., 2021), simplifies the scheduling problem (Jaber et al., 2013; Givi et al., 2015b; Fruggiero et al., 2016), or applies multi-stage scheduling methods that do not allow finding global optima (Calzavara et al., 2019; Finco et al., 2021). Our work establishes the first steps towards human fatigue consideration in scheduling real-life operations where these simplifications are absent. This results from our mathematical formulation of the JSP, which combines and augments the existing works by Berti et al. (2021) and Yazdani et al. (2015) with fatigue models that had not previously been applied in this context.

We also show the importance of fatigue model validation before application. While the proposed methods have been seen before in literature, the existing works do not go into the same grade of detail or verify their models in real life (Battini et al., 2016; Calzavara et al., 2018a, 2019; Berti et al., 2021). By validating our $\dot{E}E$ predictions in Section 5.2, we see that the approach is not fit for purpose in the current real-life setting.

Finally, the lack of similarities between RA calculation through heart rate and subjective fatigue ratings suggests there may be discrepancies in the RA approach. This has not been questioned before in a similar environment and is the basis for many existing scheduling problems that use fatigue models Konz (1998); Zhao et al. (2019); Finco et al. (2021). Our work could set a precedent for fatigue alleviation approaches

in scheduling with a heterogeneous workforce because the current results show that this does not lead to realistic RA estimation for individuals.

6.4. Relevance to operational management practice

We have seen that our mathematical formulation is able to provide solutions that single-objective makespan minimisation, which is the common approach in scheduling problems, would not have found. By making the connection between scheduling models and a real-life order picking environment, we show that applying theoretical scheduling strategies in practice is possible. This allows management to decide on the relative importance of the two objectives and generate a schedule accordingly. Although our work is currently only able to solve small problem instances, and the percentage-wise improvements could be lower in realistic problems, we believe this approach can offer great benefits for future applications. With the knowledge that occupational fatigue is a major cost factor in the working environment, as we have seen in Chapter 2, the value of fatigue consideration in operational management decision-making could also be very large if the improvements only reach a few percent.

While our work only considers fatigue effects, this work could open up opportunities for future research on other human factor integration in a working environment. Looking at the increasing automation in the sector, we argue this to be an increasingly relevant topic in operations management. We hope our work can incite further research, but also managerial considerations of human factors in the working environment.

Apart from the benefits to employers, human factors consideration can also improve the working circumstances for employees. With occupational fatigue being a large influence on the development of musculoskeletal disorders, we believe fatigue consideration in the working environment can improve worker well-being. Whereas this is common in the design level, it is still rare on the operational level and this work is one of the first to consider it in scheduling problems relevant to order picking.

6.5. Opportunities for further research

Our selected fatigue quantification method combines existing models, each with its own inaccuracies and assumptions. Most of these models are known to be valid only for large populations and may not be representative when sampling individuals. Between these models, we also had to make educated guesses of model inputs and calculation steps, as was the case for the task decomposition. Given our goal of assessing the practical applicability of the hybrid energy expenditure quantification method, this was unavoidable within the scope of a thesis. However, future research on methods for applicable task decomposition, such as posture analysis during work, could be useful to increase the accuracy of this approach without requiring constant measurement.

Another opportunity lies in the long-term verification of heart rate-based energy expenditure measurements for a larger set of test subjects to improve the accuracy of $\dot{E}E_{HR}$ and, consequently, RA calculation. If calibrated for external influences, this method could be an attainable opportunity to estimate fatigue effects in operation. Subjective fatigue ratings, especially when recorded over a long period of time, could help couple the heart rate measurements with perceived fatigue. Important to note, however, is that methods need to be developed to deal with the difficult experimental circumstances we encountered during our verification experiments. This, along with measurement noise, low and fluctuating demand, is one of the leading limitations in our attempt to validate the fatigue quantification method.

The fatigue model outcome is still an energy expenditure rate and RA value for both of these measurement methods. To assess the added value of fatigue alleviation during work, a notion of a fatigue level could offer a solution, especially if this can be coupled with fatigue-dependent productivity or errors. While already an active research topic (Givi et al., 2015a,b), the current research will require long-term studies to be applicable in a real-life setting. This could also incorporate assessment of other human factors, such as learning, motivation and stress, all relevant factors for application.

As far as the scheduling model is concerned, the current bi-objective approach offers opportunities to consider multiple relevant objectives during scheduling. However, there are many more objectives to consider in systems like Picnic's OP system. For example, order lateness, cost, employee training, production quality, volume flexibility and employee satisfaction are all objectives that could be relevant to management, although they may be difficult to quantify. Furthermore, many human and system constraints, such as ARBO regulations, system downtime, process balancing, stock levels and even conveyor traffic and routing, could be formulated

to add even more realism to the mathematical model. However, these add complexity to a problem that is already difficult to solve mathematically, so it would also likely require alternative solution strategies. Meta-heuristics and machine learning approaches, as discussed in Section 2.2.3, seem promising alternatives to our current exact solver. This would also allow for solving bigger instances of the current scheduling problem and create more complete Pareto fronts for better insights into the trade-offs between multiple objectives. This would further reveal the potential benefits of fatigue consideration in the scheduling problem.

6.6. Managerial recommendations

Fatigue in the workplace is a multidimensional construct and partially results from the work environment. Occupational fatigue has proven links to productivity and decrease in employee well-being, in both the short and the long term. Our work shows that occupational fatigue can be considered in operational decision making and that there are real benefits in doing so. The implementation of our methods allows for improvements at virtually no cost.

Although we identify multiple opportunities for further research in human fatigue modelling, measurement and incorporation in scheduling problems, we do not have to wait for this research to be finished. The substantial potential benefits of human factors consideration in the working environment motivate to start applying the learnings from the current state of research sooner rather than later.

We have seen that pick- and order-based $\dot{E}E$ estimations do not allow for accurate predictions or fast application to other activity types, especially for the type of environment that is a DRC OP system in ramp-up. Instead, despite the possible inaccuracies on an individual level, heart rate measurements allow operations management to quantify the relative physical workload per type of task in the fulfilment centre. Although these types of measurements are common in sports engineering literature, they have, to our best knowledge, yet to find their way into order picking environments. This can provide quantitative insights to consider during workforce planning. A good start may be to do this on a granular level instead of the meticulous level that our JSP formulation describes. Given the current operational challenges, pick-level task assignment is out of reach for application, while activity-level job rotation is a realistic option. Job rotation is not actively applied in present-day operations, but it could have benefits that exceed fatigue alleviation: worker learning, ordered break schedules, and more diverse work. We think these benefits can improve job satisfaction and employee retention rates.

One applicative limitation of our JSP formulation is that it applies rest allowance after each pick - resulting in RA periods of only a few seconds - and does not allow for partial fatigue alleviation. This is impractical in operation and has no desired effect because the energy expenditure rate does not drop to the resting level immediately during a break. More desirable is to add RA periods after longer periods of time, such as every 30 minutes or longer. Further research on fatigue-dependent productivity and error rates would be valuable to this RA scheduling approach.

The finer scope of our developed JSP formulation is not fit-for-purpose in the current circumstances but may become a relevant topic in future operational decision-making. In conjunction with the knowledge that academic research has gathered in the meantime, a more mature OP system could benefit from a detailed approach towards scheduling with human factors. In that case, ethical aspects of the implementation should also be taken into account. Our fatigue quantification method through pick and order data, or heart rate, in particular, requires gathering and storing sensitive personal information from employees. Although this information can be normalised and anonymised, we believe that the potential risks for individuals may not be taken into consideration by every employer. Finally, there are also conceptual issues to be considered, such as the equity and equality question posed in Section 6.1.

7

Conclusion

In this conclusive chapter, we first summarise the main contributions of this thesis, followed by answers to the research questions. Finally, we round off the thesis with some final remarks.

7.1. Summary of the main contributions

In this thesis, we propose a job-shop scheduling problem formulation for a dual-resource constrained order picking system where total makespan and physical workload distribution can be optimised simultaneously. Based on the results in sections 5.1 and 5.2, this study makes four main contributions:

- First, we combine existing energy expenditure and rest allowance models into a hybrid energy expenditure quantification model that can predict resting requirements. This means that this model can be used in a DRC job-shop scheduling problem without live measurements. With a detailed task element description, this model can be applied to a multitude of scheduling problems. The level of detail and worker heterogeneity consideration that our approach possesses have not been shown prior in the scheduling literature.
- Second, we propose a detailed mathematical formulation for the JSP in a real-life order picking system, incorporating realistic system constraints and a bi-objective that allows for simultaneous consideration of makespan and physical workload. This approach allows us to find schedules that would not have been found using single-objective optimisation which is common in existing JSP literature. These alternative solutions to the scheduling problem can offer better distribution of physical workload without hurting the total production time too much. For scheduling problems in DRC system contexts, existing literature has not shown our level of detail and bi-objective approach towards the scheduling problem.
- Third, we evaluate the practical applicability of our energy expenditure quantification approach in a real-life OP operation. Our results show that neither our order- and pick-based EE estimations provide accurate predictions when compared to the heart rate-based estimation. This may in part be caused by the large variation in pick demand and personal characteristics for the 20 participants, but could also suggest that our estimation method is unsuitable for accurate prediction. Given the experimental difficulties and missing elements in our task decomposition, we are inconclusive about the potential of our energy expenditure quantification approach in a more structured environment.
- Fourth, we investigate the effects that rest allowance application during work has on the energy expenditure and subjective fatigue ratings. We are unable to find significant differences due to external influences during the experiments. This remains an open topic for research, as it could quantify the potential benefits of appropriate break schedules.

7.2. The answer to the research questions

We investigated how worker fatigue and productivity can simultaneously be considered in job-shop scheduling for a partially automated order picking system. We can answer the research questions from Section 1.4:

- **How can we quantify, predict and measure fatigue for scheduling purposes in a heterogeneous workforce?**

We find that energy expenditure-based fatigue models can be applied in scheduling problem formulations. In combination with the rest allowance principle, this allows for a personalised approach to fatigue in a scheduling context. The applied predictive fatigue model allows for detailed task decomposition without prior measurements, able to evaluate energy expenditure requirements for a variety of tasks.

- **How can we apply fatigue models to a bi-objective JSP formulation that represents a partially automated order picking system?**

Our modelling study shows that the proposed bi-objective JSP formulation can successfully generate schedules for partially automated OP systems while considering fatigue and productivity. The results are promising for application, although our results are limited to small problem instances. We argue that the mathematical formulation could benefit from a different solution approach, without the need for reformulation.

- **How do the fatigue model and resulting schedules perform in a real-life operation?**

Our empirical study shows that the applied predictive energy expenditure model was inaccurate in estimating energy expenditure rates in a real operation and that heart rate sensors are a more viable approach at this time. Also, we find no significant changes in measured or qualitative fatigue after applying the generated fatigue-conscious schedules, in contrast with the results from the modelling study.

7.3. Final remarks

In conclusion, this thesis presents a detailed formulation for the job-shop scheduling problem in a dual-resource constrained order picking system, taking into account physical workload distribution and total makespan simultaneously. Our approach provides an answer to the question of how human fatigue models can be implemented in the scheduling problem and highlights the importance of verifying these models in real-life situations. Our results also highlight the potential discrepancies in rest allowance calculation through pick data and heart rate. These findings have practical implications for scheduling in real-life operations, providing management with alternative schedules that consider both productivity and employee physical strain. This thesis sets the first steps towards human fatigue consideration in scheduling for real-life operations and opens up avenues for further research in this field.

Bibliography

- Abdous, M. A., Delorme, X., Battini, D., Sgarbossa, F., and Berger-Douce, S. (2018). Multi-objective optimization of assembly lines with workers' fatigue consideration. *IFAC-PapersOnLine*, 51(11):698–703.
- Achten, J. and Jeukendrup, A. E. (2003). Heart Rate Monitoring Applications and Limitations. *Sports Med*, 33(7):517–538.
- Aleksendrić, D. and Carlone, P. (2015). Composite materials – modelling, prediction and optimization. *Soft Computing in the Design and Manufacturing of Composite Materials*, pages 61–289.
- Álvarez-García, J. A., Cvetkovic, B., and Lustrek, M. (2020). A survey on energy expenditure estimation using wearable devices. *ACM Computing Surveys*, 53(5).
- American College of Sports Medicine (2006). The Compendium of Physical Activities. In *ACSM Resource Manual for Guidelines for Exercise Testing and Prescription*. Lippincott Williams & Wilkins, Baltimore, MD, 5 edition.
- Aryal, A., Ghahramani, A., and Becerik-Gerber, B. (2017). Monitoring fatigue in construction workers using physiological measurements. *Automation in Construction*, 82:154–165.
- Aytug, S., Bhattacharyya, S., Koehler, G. J., and Snowdon, J. (1994). A Review of Machine Learning in Scheduling. *IEEE Transactions on Engineering Management*, 41(2):165–171.
- Barbero-Alvarez, J. C., Soto, V. M., Barbero-Alvarez, V., and Granda-Vera, J. (2008). Match analysis and heart rate of futsal players during competition. <https://doi.org/10.1080/02640410701287289>, 26(1):63–73.
- Barker, L. M., Nussbaum, M. A., and Cpe, M. A. N. (2010). Fatigue, performance and the work environment: a survey of registered nurses. *Journal of Advanced Nursing*, 67(6):1370–1382.
- Battini, D., Finco, S., and Sgarbossa, F. (2020). Human-oriented assembly line balancing and sequencing model in the industry 4.0 era. *International Series in Operations Research and Management Science*, 289:141–165.
- Battini, D., Glock, C. H., Grosse, E. H., Persona, A., and Sgarbossa, F. (2016). Human energy expenditure in order picking storage assignment: A bi-objective method. *Computers & Industrial Engineering*, 94:147–157.
- Becerra, R. L. and Coello Coello, C. A. (2006). Solving hard multiobjective optimization problems using ϵ -constraint with cultured differential evolution. In *Parallel Problem Solving from Nature-PPSN IX*, volume 4193 LNCS, pages 543–552, Berlin, Heidelberg. Springer.
- Benavides, A. J., Ritt, M., and Miralles, C. (2014). Flow shop scheduling with heterogeneous workers. *European Journal of Operational Research*, 237(2):713–720.
- Berti, N., Finco, S., Battaia, O., and Delorme, X. (2021). Ageing workforce effects in Dual-Resource Constrained job-shop scheduling. *International Journal of Production Economics*, 237:108151.
- Bestuzheva, K., Besançon, M., Chen, W.-K., Chmiela, A., Donkiewicz, T., Van Doornmalen, J., Eifler, L., Gaul, O., Gamrath, G., Gleixner, A., Gottwald, L., Graczyk, C., Halbig, K., Hoen, A., Hojny, C., Van Der Hulst, R., Koch, T., Lübbecke, M., Maher, S. J., Matter, F., Mühmer, E., Müller, B., Pfetsch, M. E., Reffeldt, D., Schlein, S., Schlösser, F., Serrano, F., Shinano, Y., Sofranac, B., Turner, M., Vigerske, S., Wegscheider, F., Wellner, P., Weninger, D., and Witzig, J. (2021). The SCIP Optimization Suite 8.0.
- Borg, G. A. (1982). Psychophysical Bases of Perceived Exertion. *Medicine & Science in Sports & Exercise*, 14(5):377–381.

- Bormann, R., de Brito, B. F., Lindermayr, J., Omainka, M., and Patel, M. (2019). Towards Automated Order Picking Robots for Warehouses and Retail. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11754 LNCS:185–198.
- Calzavara, M., Glock, C. H., Grosse, E. H., and Sgarbossa, F. (2018a). An integrated storage assignment method for manual order picking warehouses considering cost, workload and posture. <https://doi.org/10.1080/00207543.2018.1518609>, 57(8):2392–2408.
- Calzavara, M., Persona, A., Sgarbossa, F., and Visentin, V. (2018b). A device to monitor fatigue level in order-picking. *Industrial Management and Data Systems*, 118(4):714–727.
- Calzavara, M., Persona, A., Sgarbossa, F., and Visentin, V. (2019). A model for rest allowance estimation to improve tasks assignment to operators. *International Journal of Production Research*, 57(3):948–962.
- Cheng, T. C., Wu, C. C., and Lee, W. C. (2008). Some scheduling problems with deteriorating jobs and learning effects. *Computers and Industrial Engineering*, 54(4):972–982.
- Coppola, D. (2021). E-commerce worldwide - statistics & facts. *Statista*.
- Cunha, B., Madureira, A. M., Fonseca, B., and Coelho, D. (2020). Deep Reinforcement Learning as a Job Shop Scheduling Solver: A Literature Review. *Advances in Intelligent Systems and Computing*, 923:350–359.
- Daria, B., Martina, C., Alessandro, P., Fabio, S., and Valentina, V. (2017). Fatigue and recovery: research opportunities in order picking systems. *IFAC-PapersOnLine*, 50(1):6882–6887.
- Datta, S. (2022). Branch and Bound Algorithm.
- de Koster, R. (2023). Warehousing 2030. In Merkert, R. and Hoberg, K., editors, *Global Logistics and Supply Chain Strategies for the 2020s: Vital Skills for the Next Generation*, pages 243–260. Springer International Publishing, Cham.
- de Koster, R., Le-Duc, T., and Roodbergen, K. J. (2007). Design and control of warehouse order picking: A literature review. *European Journal of Operational Research*, 182(2):481–501.
- de la Riva, J., Garcia, A. I., Reyes, R. M., and Woocay, A. (2015). Methodology to Determine Time Allowance by Work Sampling Using Heart Rate. *Procedia Manufacturing*, 3:6490–6497.
- De Souza E Silva, C. G., Franklin, B. A., Forman, D. E., Gil, C., and Araújo, S. (2016). Influence of age in estimating maximal oxygen uptake. *Journal of Geriatric Cardiology*, 13:126–131.
- Demirkol, E., Mehta, S., and Uzsoy, R. (1998). Benchmarks for shop scheduling problems. *European Journal of Operational Research*, 109(1):137–141.
- Dode, P., Greig, M., Zolfaghari, S., and Neumann, W. P. (2016). Integrating human factors into discrete event simulation: A proactive approach to simultaneously design for system performance and employees well being. *International Journal of Production Research*, 54(10):3105–3117.
- El ahrache, K. and Imbeau, D. (2009). Comparison of rest allowance models for static muscular work. *International Journal of Industrial Ergonomics*, 39(1):73–80.
- El Ahrache, K., Imbeau, D., and Farbos, B. (2006). Percentile values for determining maximum endurance times for static muscular work. *International Journal of Industrial Ergonomics*, 36(2):99–108.
- Finco, S., Calzavara, M., Sgarbossa, F., and Zennaro, I. (2021). Including rest allowance in mixed-model assembly lines. *International Journal of Production Research*, 59(24):7468–7490.
- Fruggiero, F., Riemma, S., Ouazene, Y., Macchiaroli, R., and Guglielmi, V. (2016). Incorporating the Human Factor within Manufacturing Dynamics. *IFAC-PapersOnLine*, 49(12):1691–1696.
- Garg, A., Chaffin, D. B., and Herrin, G. D. (1978). Prediction of metabolic rates for manual materials handling jobs. *American Industrial Hygiene Association Journal*, 39(8):661–674.

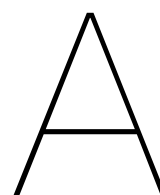
- Givi, Z. S., Jaber, M. Y., and Neumann, W. P. (2015a). Modelling worker reliability with learning and fatigue. *Applied Mathematical Modelling*, 39(17):5186–5199.
- Givi, Z. S., Jaber, M. Y., and Neumann, W. P. (2015b). Production planning in DRC systems considering worker performance. *Computers and Industrial Engineering*, 87:317–327.
- Gong, G., Deng, Q., Gong, X., Liu, W., and Ren, Q. (2018). A new double flexible job-shop scheduling problem integrating processing time, green production, and human factor indicators. *Journal of Cleaner Production*, 174:560–576.
- Grosse, E. H., Glock, C. H., and Neumann, W. P. (2017). Human factors in order picking: a content analysis of the literature. *International Journal of Production Research*, 55(5):1260–1276.
- Hackney, A. C. (2016). Measurement Techniques for Energy Expenditure. *Exercise, Sport, and Bioanalytical Chemistry*, pages 33–42.
- Hajifar, S., Sun, H., Megahed, F. M., Jones-Farmer, L. A., Rashedi, E., and Cavuoto, L. A. (2021). A forecasting framework for predicting perceived fatigue: Using time series methods to forecast ratings of perceived exertion with features from wearable sensors. *Applied Ergonomics*, 90.
- Hammond, H. K. and Froelicher, V. F. (1985). Normal and abnormal heart rate responses to exercise. *Progress in Cardiovascular Diseases*, 27(4):271–296.
- Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, 52(C):139–183.
- Hiilloskorpi, H. K., Pasanen, M. E., Fogelholm, M. G., Laukkanen, R. M., and Mänttari, A. T. (2003). Use of heart rate to predict energy expenditure from low to high activity levels. *International Journal of Sports Medicine*, 24(5):332–336.
- Hoitomt, D. J., Luh, P. B., and Pattipati, K. R. (1993). A Practical Approach to Job-Shop Scheduling Problems. *IEEE Transactions on Robotics and Automation*, 9(1):1–13.
- IEA (2000). What Is Ergonomics?
- Ingebrigtsen, J., Bendiksen, M., Randers, M. B., Castagna, C., Krstrup, P., and Holtermann, A. (2012). Yo-Yo IR2 testing of elite and sub-elite soccer players: Performance, heart rate response and correlations to other interval tests. <https://doi.org/10.1080/02640414.2012.711484>, 30(13):1337–1345.
- Jaber, M. Y. and Bonney, M. (1996). Production breaks and the learning curve: The forgetting phenomenon. *Applied Mathematical Modelling*, 20(2):162–169.
- Jaber, M. Y., Givi, Z. S., and Neumann, W. P. (2013). Incorporating human fatigue and recovery into the learning-forgetting process. *Applied Mathematical Modelling*, 37(12-13):7287–7299.
- Jaber, M. Y. and Neumann, W. P. (2010). Modelling worker fatigue and recovery in dual-resource constrained systems. *Computers and Industrial Engineering*, 59(1):75–84.
- Jennings, J. R., Bberg, W. K., Hutcheson, J. S., Obrist, P., Porges, S., and Turpin, G. (1981). Publication Guidelines for Heart Rate Studies in Man. *Psychophysiology*, 18(3):226–231.
- Jongbloed, K., Van der Kruk, E., De Vos, R., and Harlaar, J. (2022). *Defining and measuring fatigue in endurance sports*. PhD thesis, Delft University of Technology, Delft.
- Kadir, B. A., Broberg, O., and Conceição, C. S. d. (2019). Current research and future perspectives on human factors and ergonomics in Industry 4.0. *Computers and Industrial Engineering*, 137:106004.
- Konz, S. (1998). Work/rest: Part II - The scientific basis (knowledge base) for the guide. *International Journal of Industrial Ergonomics*, 22(1-2):73–99.
- Kronqvist, J., Bernal, D. E., Lundell, A., and Grossmann, I. E. (2019). A review and comparison of solvers for convex MINLP. *Optimization and Engineering*, 20:397–455.

- Lee, C. Y., Piramuthu, S., and Tsai, Y. K. (2010). Job shop scheduling with a genetic algorithm and machine learning. <https://doi.org/10.1080/002075497195605>, 35(4):1171–1191.
- Li, R., Deurenberg, P., and Hautvast, J. G. (1993). A critical evaluation of heart rate monitoring to assess energy expenditure in individuals. *The American Journal of Clinical Nutrition*, 58(5):602–607.
- Lindo Systems Inc. (2022). Powerful LINGO Solvers.
- Llorens-Serrano, C., Pérez-Franco, J., Oudyck, J., Berthelsen, H., Dupret, E., Nübling, M., Burr, H., and Moncada, S. (2019). COPSOQ III Guidelines and questionnaire.
- Lorist, M. M., Kernell, D., Meijman, T. F., and Zijdwind, I. (2002). Motor fatigue and cognitive task performance in humans. *Journal of Physiology*, 545(1):313–319.
- Lorist, M. M., Klein, M., Nieuwenhuis, S., De Jong, R., Mulder, G., and Meijman, T. F. (2000). Mental fatigue and task control: Planning and preparation. *Psychophysiology*, 37(5):614–625.
- Maher, S. J., Miltenberger, M., Pedroso, J. P., Rehfeldt, D., Schwarz, R., and Serrano, F. (2016). PySCIPOpt: Mathematical Programming in Python with the SCIP Optimization Suite. In *Mathematical Software - ICMS 2016*, volume 9725, page 301 – 307.
- Marler, R. T. and Arora, J. S. (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6):369–395.
- Mehdizadeh, E., Tavakkoli-Moghaddam, R., and Yazdani, M. (2015). A vibration damping optimization algorithm for a parallel machines scheduling problem with sequence-independent family setup times. *Applied Mathematical Modelling*, 39(22):6845–6859.
- Meister, D. (1989). *Conceptual Aspects of Human Factors*. John Hopkins University Press.
- Millar, W. J. (1986). Distribution of body weight and height: Comparison of estimates based on self-reported and observed measures. *Journal of Epidemiology and Community Health*, 40:319–323.
- Mosheiov, G. and Sidney, J. B. (2003). Scheduling with general job-dependent learning curves. *European Journal of Operational Research*, 147(3):665–670.
- Mouayni, I. E., Demasure, G., Haouzi, H. B. E., Charpentier, P., and Siadat, A. (2019). Jobs scheduling within Industry 4.0 with consideration of worker's fatigue and reliability using Greedy Randomized Adaptive Search Procedure. *IFAC-PapersOnLine*, 52(19):85–90.
- Murray-Smith, D. J. (2012). Experimental modelling: system identification, parameter estimation and model optimisation techniques. *Modelling and Simulation of Integrated Systems in Engineering*, pages 165–214.
- National Safety Council (2022). What is Fatigue Costing Your Company? - National Safety Council.
- Neumann, W. P. and Dul, J. (2010). Human factors: Spanning the gap between OM and HRM. *International Journal of Operations and Production Management*, 30(9):923–950.
- Niu, Y., Schulte, F., and Negenborn, R. R. (2021). Human Aspects in Collaborative Order Picking - Letting Robotic Agents Learn about Human Discomfort. *Procedia Computer Science*, 180:877–886.
- Ostchega, Y., Porter, K. S., Hughes, J., Dillon, C. F., and Nwankwo, T. (2011). Resting pulse rate reference data for children, adolescents, and adults; United States, 1999-2008.
- Otasevic, N. (2018). Follow these steps to solve any Dynamic Programming interview problem.
- Othman, M., Gouw, G. J., and Bhuiyan, N. (2012). Workforce scheduling: A new model incorporating human factors. *Journal of Industrial Engineering and Management*, 5(2):259–284.
- Pasadyan, S. R., Soudan, M., Gillinov, M., Houghtaling, P., Phelan, D., Gillinov, N., Bittel, B., and Desai, M. Y. (2019). Accuracy of commercially available heart rate monitors in athletes: a prospective study. *Cardiovascular Diagnosis and Therapy*, 9(4):379.

- Phua, C. T., Lissorgues, G., Gooi, B. C., and Mercier, B. (2012). Statistical Validation of Heart Rate Measurement Using Modulated Magnetic Signature of Blood with Respect to Electrocardiogram. *International Journal of Bioscience, Biochemistry and Bioinformatics*, pages 110–116.
- Polar Research and Technology (2019). Polar H10 Heart Rate Sensor System [White Paper]. Technical report, Polar Research and Technology, Kempele.
- Price, A. D. (1990). Calculating relaxation allowances for construction operatives - Part 1: Metabolic cost. *Applied Ergonomics*, 21(4):311–317.
- Rao, S. S. (2019). Classical Optimization Techniques. *Engineering Optimization Theory and Practice*, pages 57–108.
- Rose, L., Ericson, M., Glimskär, B., and Nordgren, B. (1992). Ergo-index-development of a model to determine pause needs after fatigue and pain reactions during work. *Computer Applications in Ergonomics, Occupational Safety and Health*, pages 461–468.
- Rosekind, M. R., Gregory, K. B., Mallis, M. M., Brandt, S. L., Seal, B., and Lerner, D. (2010). The Cost of Poor Sleep: Workplace Productivity Loss and Associated Costs. *Journal of Occupational and Environmental Medicine*, 52(1):91–98.
- Šeda, M. (2007). Mathematical Models of Flow Shop and Job Shop Scheduling Problems. *International Journal of Physical and Mathematical Sciences*, 1(7):307–312.
- Sedighi Maman, Z., Alamdar Yazdi, M. A., Cavuoto, L. A., and Megahed, F. M. (2017). A data-driven approach to modeling physical fatigue in the workplace using wearable sensors. *Applied Ergonomics*, 65:515–529.
- Sedighi Maman, Z., Chen, Y. J., Baghdadi, A., Lombardo, S., Cavuoto, L. A., and Megahed, F. M. (2020). A data analytic framework for physical fatigue management using wearable sensors. *Expert Systems with Applications*, 155:113405.
- Seegerstrom, S. C. and Nes, L. S. (2016). Heart Rate Variability Reflects Self-Regulatory Strength, Effort, and Fatigue. <https://doi.org/10.1111/j.1467-9280.2007.01888.x>, 18(3):275–281.
- Sellers, D. W. (1996). Survey of approaches to the job shop scheduling problem. *Proceedings of the Annual Southeastern Symposium on System Theory*, pages 396–400.
- Sgarbossa, F., Grosse, E. H., Neumann, W. P., Battini, D., and Glock, C. H. (2020). Human factors in production and logistics systems of the future. *Annual Reviews in Control*, 49:295–305.
- Shorrock, S. (2019). ‘Human Factors’ and ‘Human Performance’: What’s the difference?
- Snyder, L. V. and Daskin, M. S. (2005). Reliability Models for Facility Location: The Expected Failure Cost Case. <https://doi.org/10.1287/trsc.1040.0107>, 39(3):400–416.
- Steinebach, T., Wakula, J., and Mehmedovic, A. (2021). The Influence of an Ergonomic Storage Location Assignment on Human Strain in Manual Order Picking. In Black, N. L., Neumann, W. P., and Noy, I., editors, *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021)*, pages 511–521, Cham. Springer International Publishing.
- Stone, J. D., Ulman, H. K., Tran, K., Thompson, A. G., Halter, M. D., Ramadan, J. H., Stephenson, M., Finomore, V. S., Galster, S. M., Rezai, A. R., and Hagen, J. A. (2021). Assessing the Accuracy of Popular Commercial Technologies That Measure Resting Heart Rate and Heart Rate Variability. *Frontiers in Sports and Active Living* | www.frontiersin.org, 3(585870).
- Tanaka, H., Monahan, K. D., and Seals, D. R. (2001). Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*, 37(1):153–156.
- Thürer, M. and Stevenson, M. (2021). Order release, dispatching and resource assignment in multiple resource-constrained job shops: an assessment by simulation. <https://doi.org/10.1080/00207543.2021.1930240>, 60(12):3669–3681.

- Thürer, M., Zhang, H., Stevenson, M., Costa, F., and Ma, L. (2020). Worker assignment in dual resource constrained assembly job shops with worker heterogeneity: an assessment by simulation. *International Journal of Production Research*, 58(20):6336–6349.
- Van Der Gaast, J. P., De Koster, R. B., Adan, I. J., and Resing, J. A. (2020). Capacity analysis of sequential zone picking systems. *Operations Research*, 68(1):161–179.
- Verschueren, C., Läubli, D., Laizet, F., D’Auria, G., Wachinger, T., Simmons, V., Nyssens, J., and Vallöf, R. (2021). Disruption and Uncertainty: The State of Grocery Retail 2021: Europe. Technical Report November, McKinsey & Company.
- Vijayakumar, V. and Sgarbossa, F. (2021). A literature review on the level of automation in picker-to-parts order picking system: Research opportunities. *IFAC-PapersOnLine*, 54(1):438–443.
- Wang, F.-S. and Chen, L.-H. (2013). Heuristic Optimization. *Encyclopedia of Systems Biology*, pages 885–885.
- Wang, J., Yang, J., Zhang, Y., Ren, S., and Liu, Y. (2020). Infinitely repeated game based real-time scheduling for low-carbon flexible job shop considering multi-time periods. *Journal of Cleaner Production*, 247:119093.
- Willems, T. M. and Rooda, J. E. (1994). Neural networks for job-shop scheduling. *Control Engineering Practice*, 2(1):31–39.
- Winkelhaus, S., Grosse, E. H., and Morana, S. (2021). Towards a conceptualisation of Order Picking 4.0. *Computers and Industrial Engineering*, 159:107511.
- Xiong, H., Shi, S., Ren, D., and Hu, J. (2022). A survey of job shop scheduling problem: The types and models. *Computers & Operations Research*, 142:105731.
- Xu, J., Xu, X., and Xie, S. Q. (2011). Recent developments in Dual Resource Constrained (DRC) system research. *European Journal of Operational Research*, 215(2):309–318.
- Yazdani, M., Zandieh, M., Tavakkoli-Moghaddam, R., and Jolai, F. (2015). Two meta-heuristic algorithms for the dual-resource constrained flexible job-shop scheduling problem. *Scientia Iranica E*, 22(3):1242–1257.
- Zhang, Y., Wang, J., and Liu, Y. (2017). Game theory based real-time multi-objective flexible job shop scheduling considering environmental impact. *Journal of Cleaner Production*, 167:665–679.
- Zhao, X., Liu, N., Zhao, S., Wu, J., Zhang, K., and Zhang, R. (2019). Research on the Work-rest Scheduling in the Manual Order Picking Systems to Consider Human Factors. *Journal of Systems Science and Systems Engineering*, 28(3):344–355.
- Zheng, X.-L. and Wang, L. (2016). A knowledge-guided fruit fly optimization algorithm for dual resource constrained flexible job-shop scheduling problem. *International Journal of Production Research*, 54(18):5554–5566.

Appendices



Paper format

A Bi-Objective Job-Shop Scheduling Problem Considering Worker Fatigue and Productivity in Cobotic Order Picking Systems

Berry Lance Vermin^{a,b}, David Abbink^a, Frederik Schulte^b

^aDepartment of Cognitive Robotics, Delft University of Technology, 2628 CD Delft, The Netherlands

^bDepartment of Maritime & Transport Technology, Delft University of Technology, 2628 CD Delft, The Netherlands

Abstract

Increasing online retail has resulted in increased automation in order picking systems, leading to new challenges and opportunities in task scheduling. The job-shop scheduling problem is an optimisation problem essential in such systems, but existing JSP literature often overlooks workplace fatigue, which harms employees' well-being and costs U.S. employers up to €127 billion annually.

This thesis proposes fatigue consideration in the job-shop scheduling problem in a cobotic order picking system to mitigate its negative effects. We present a new bi-objective mixed integer nonlinear programming problem formulation that considers worker fatigue and productivity during schedule optimisation. To put the results of simulated optimisation in perspective, we experimentally validate the fatigue model and scheduling results in a real operation.

The mathematical model finds solutions that conventional single-objective optimisation cannot, allowing fractional fatigue distribution improvements more than 4x larger than the decrease in productivity they require in 53% of the considered virtual cases. The experiments show that our predictive fatigue model has an average RMSE of 2.20 kcal/min in estimating energy expenditure rates compared to heart rate measurements. It also shows a low correlation, meaning it is unfit for application. Also, fatigue-conscious schedules show no clear benefit regarding measured and perceived fatigue. However, the scheduling model could also use heart rate measurements that do not show these inaccuracies.

Our study highlights the need to further develop and validate the mathematical formulation and fatigue model and extend to other human factors and indirect fatigue effects. The findings show promise for fatigue mitigation in operations by altering operational decision-making.

Keywords: Order Picking (OP), Job-Shop Scheduling Problem (JSP), Human Factors (HF), Mixed-Integer Nonlinear Programming (MINLP), Occupational Fatigue, Cobotic Order Picking

1. Introduction

Retail is shifting its business away from traditional brick-and-mortar stores and into an online setting [1]. With labour accounting for up to 55% of operational costs in order picking (OP), businesses are looking for new, more efficient ways to prepare customers' orders [2]. By introducing automation in the OP process, retailers can increase their productivity per labour hour, allowing them to fulfil a greater demand at reduced operational costs. In the online grocery sector, which has seen increasing demand in recent years [3], automation is often only partially applied, and reliance on humans continues to exist.

One operational challenge in partially automated OP systems is a combinatorial optimisation called the job-shop scheduling problem (JSP). In OP systems, the JSP describes assigning individual item picks to employees, a picking station and a time while preparing an order in a unique sequence of steps [4]. Generally, the objective is the total production time or makespan, but many other objectives exist [5]. Optimisation of the JSP in a partially automated OP system comes with additional challenges, but also ways

in which the automated resources can support humans in ways previously impossible.

This thesis is written in conjunction with Picnic, an online grocery company. The conveyor network in Picnic's OP system allows for flexible task assignment to order pickers ('shoppers'), who can also switch easily between (work)stations. At the same time, there can be a larger imbalance in physical workload between employees than in manual OP systems, as only a subset of items is located at each station. This storage location assignment (slotting) can lead to large, heavy or large numbers of items being picked at one station, while other stations can have much less physically fatiguing picking tasks. This fatigue affects employee well-being, productivity and Picnic's chances of retaining shoppers.

Workplace fatigue can have costly effects on employers. Some studies estimate that the cost of fatigue-related productivity loss in the workplace is as high as 127 billion euros per year in the U.S. alone [6]. Fatigue in the workplace is a multidimensional construct in which we can identify external and occupational factors [7]. External factors include outside work activities, sleeping disorders, climate and other personal factors, while occupational factors are within the employer's control and include shift work, long hours and overtime, time on task, workload and break schedules. Not all occupational fatigue is completely avoidable, but it is possible to minimise the negative implications on the design, organisational and operational level [4].

Considering human factors (HF), such as fatigue, in work environments can simultaneously improve productivity and employee well-being [8]. HF are essential in a partial automation environment but generally not or insufficiently considered in existing operations [9]. If HF are considered, this is often done using mathematical HF models on the design level, such as workstation design [10, 11]. However, in existing facilities like Picnic's OP system, this would require retrofitting the design level, which is not always possible and often expensive [12]. We also see that the current operational decision-making methods do not allow fatigue consideration [11]. With the opportunities that automation offers, 'cobotic' OP systems - ones with active human-robot collaboration - could then actively consider human fatigue development, mitigating its negative effects in operation [13].

In this thesis, we address this knowledge gap on occupational fatigue and the possibility of mitigating its negative effects through operational decision-making. The problem statement is formulated as:

How can we simultaneously optimise worker fatigue and productivity in the job-shop scheduling problem for cobotic order picking systems?

We hypothesise that fatigue consideration in the JSP can be achieved by implementing a quantitative fatigue model in the JSP formulation and setting up a bi-objective optimisation with fatigue and productivity indicators. We expect predictive fatigue models to show some inaccuracies but still allow for appropriate fatigue mitigation in scheduling. Also, we expect fatigue-conscious schedules to lead to lower measured and perceived fatigue in real operations.

The remainder of this paper is structured as follows. Section 2 overviews the techniques used for fatigue quantification and human factors implementation in JSPs. Our proposed JSP formulation and experimental approach are presented in Sections 3 and 4, respectively. Section 5 presents the results of virtual problem instances and the experiments. Section 6 discusses these results, our study limitations and recommendations. Finally we make concluding remarks in Section 7.

2. Background and literature review

This thesis combines two branches of literature: human fatigue quantification and modelling and human factors in job-shop scheduling, which we will review separately in this section.

2.1. Fatigue models

Occupational fatigue can be classified into physiological and psychological elements. Given the similarities in cognitive and environmental task characteristics at Picnic's pick stations, quantifying physiological fatigue is more relevant. By doing so, we can gain valuable insights into the ways in which physical workload and fatigue can impact worker performance and identify potential solutions to improve their working conditions. Physiological fatigue describes the reduction in muscle force generation or muscle response resulting from preceding physical exertion, also called performance fatigability [14]. A measure that concerns performance fatigability in terms of mechanical output is the maximum endurance time.

2.1.1. Maximum endurance time-based fatigue models

Physical fatigue can be categorised into static and dynamic fatigue. As the name suggests, static fatigue results from holding a particular body orientation for a prolonged period, while dynamic fatigue results from movements. The maximum time a muscle can exert a static force level is called the maximum endurance time (MET). The MET is often calculated as a function of the percentage of maximum voluntary contraction (%MVC or fMVC). This is the maximum force the muscles can exert [15, 16, 17]. Many models exist to calculate the MET, with predominantly linear and power forms [16, 18]. MET-based fatigue models have some fundamental flaws for application in a scheduling context:

1. The MET alone has no real value in a scheduling context other than setting a limit on the time on task.
2. The MET gives no information about the shape of the fatigue development function between $t = 0$ and $t = MET$ [16].
3. Because the MVC is different for each muscle (pair), the MET determination will automatically be muscle-specific rather than full-body. In jobs comprising multiple sub-tasks with different muscles, MET estimation is laborious. This is amplified by individual differences, especially when dealing with a large employee pool.
4. MVC is measured in a static load scenario and, therefore, only valid for static exertion. This is not a valid assumption for most physical tasks, where we must look at dynamic fatigue.

Although the first two flaws can be addressed by the use of a fatigue level [19, 20, 21] or by calculating the rest or relaxation allowance (RA) [17, 22, 23, 16] - the fractional period that is needed on break to recover from the load - full-body dynamic fatigue evaluation remains problematic.

2.1.2. Energy expenditure-based fatigue models

Attempting to mitigate these shortcomings is [24], who use an adaptation of the RA fraction formula by [25]:

$$RA = \max \left\{ 0; \frac{\dot{EE} - MAEE}{MAEE - \dot{EE}_R} \right\} \quad (1)$$

, with \dot{EE} the average energy expenditure (EE) rate during a period, \dot{EE}_R the EE rate during rest and $MAEE$ the maximum allowable energy expenditure rate, which is the maximum level an individual can work without fatigue effects. Figure 1 shows what this relationship would look like in practice.

The work by [25] defines MAEE as 4.3 kcal/min, but others argue it depends on sex, age and body weight [26]. The same is true for \dot{EE}_R , which [25] estimates as 1.86 or 1.50 kcal/min for standing or sitting breaks, respectively, while [27] describes it as a function of the activity and body weight. Also, some works debate the direct transitions of \dot{EE} levels in Figure 1 [28].

Although EE-based fatigue models rarely output a fatigue level - which would not translate to the definition of performance fatigability - they allow for integration in a scheduling algorithm without the flaws common for MET-based models. Another significant advantage is that they require no data gathering on fatigue or

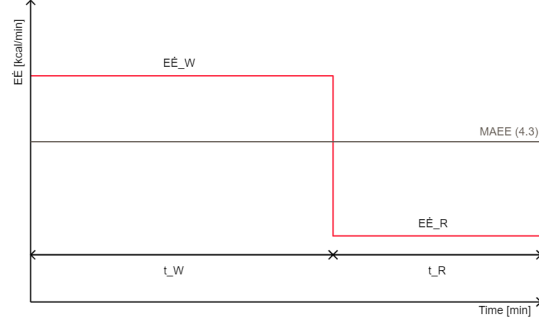


Figure 1: Graph relating EE and RA, with EE_W denoting the energy expenditure rate during the work period, t_W the working time and t_R the required resting time.

recovery rates for each individual but can work with personal characteristics ($MAEE$ and EE_R) as the model input. However, they require EE estimation before application is possible. Two main approaches exist for EE estimation: bodily measurements and mathematical task modelling.

The first approach comprises different measurement methods, such as electromyography (EMG) [29, 30], oxygen uptake (VO2) [31], indirect VO2 through heart rate (HR) [32, 17, 33, 28], direct HR data [34, 35, 36] and combined machine learning methods [37, 38]. All of these measurement methods require constant monitoring for every individual, so they are cumbersome when applied to a large workforce. Direct HR data is the least intrusive and complicated measurement method, but this comes at the cost of estimation accuracy.

The second requires an estimation of the activity intensity through the metabolic task equivalent [39], the basal metabolic rate [31] or task decomposition into single movements [27] - taking into account personal characteristics like body weight or age. None of these methods is fit for live evaluation [30], and only the last is elemental and applicable to activities that have not been previously evaluated. The method is applied in some existing works, but these lack applicative detail or do not consider both worker and task heterogeneity [40, 41, 28, 42].

2.2. Human factors in job-shop scheduling

2.2.1. Job-shop scheduling problems

In the job-shop scheduling problem, there is a set M of m machines and a set J of n jobs consisting of a sequence of N operations O to be processed. The required operations can differ per job, and precedence relations dictate the sequence in which these must be processed. The JSP applies to various production systems where production cycles differ between products. Also, each operation i can be processed only on a subset of machines (M_i) and has a specific processing time p_i . In the JSP, the decision variable is the starting time of each operation t_i , to be optimised for minimal makespan C_{max} [43]:

$$\min C_{max} = \max_{i \in O} (t_i + p_i) \quad (2a)$$

$$\text{s.t. } t_i \geq 0, \quad \forall i \in O, \quad (2b)$$

$$t_j \geq t_i + p_i, \quad \forall i, j \in O, i \rightarrow j, \quad (2c)$$

$$(t_j \geq t_i + p_i) \vee (t_j \geq t_i + p_j), \quad \forall i, j \in O, i \neq j, M_i = M_j. \quad (2d)$$

2a shows the objective to minimise the makespan, defined as the maximum of the starting + processing time for all operations i . This starting time is strictly positive, as ensured by 2b. 2c shows the constraint from the precedence relation between operations i and j , which is indicated by $i \rightarrow j$. Finally, constraint 2d ensures that no unique operations i and j can be simultaneously processed on the same machine.

Because systems the JSP applies to are so widespread, we find a variety of JSP in literature. A survey by [5] lists the common assumptions in JSP models and then gives an overview of the common sub-types from this basic model. Examples of assumptions are:

1. The set of jobs is predetermined, static and known
2. The sequence of operations in a job is predetermined
3. Operations have a unique and predetermined processing time
4. Operations are pre-assigned to only one machine
5. No worker constraints are considered

The common basic sub-types of JSP are variants of the basic formulation with one or some assumptions altered. For example, there is the dynamic JSP, which does not assume all attributes of the system or inputs are unchanged during the scheduling process, thereby altering assumptions 1 and 2, among others. Another sub-type is the flexible JSP, which would add flexibility to assumptions 2 and 4. Furthermore, regarding HF, one would consider JSP with nonconstant or nondeterministic processing times (assumption 3) or with dual-resource constraints implemented (assumption 5).

Still, the general formulation of these JSP sub-types would be similar to the one in 2. Although the formulation may look simple because it contains few constraints, the JSP is "one of the hardest combinatorial optimisation problems" [43]. The JSP is known to be NP-hard, as no algorithms exist that could solve it optimally in polynomial time. In practice, this means that exact solving methods could solve only small problem instances, even for the basic sub-types of the JSP [44].

2.2.2. Implementation of fatigue in the job-shop scheduling problem

Implementing human factors in the JSP requires the relaxation of multiple assumptions from the basic model. Although this JSP variant has not been addressed explicitly by [5], an increasing volume of literature has been published in recent years. In Table 1, we present an overview of the implementation of HF in scheduling problems in the existing literature. The work by [45] combines learning, fatigue, recovery and motivation in a virtual job-shop scheduling problem with a combined relative goal deviation objective. The authors make many simplifications in implementing the HF models, resulting in an unrealistic version of the JSP. However, to our knowledge, they are the first to suggest multiple objectives for JSP optimisation.

Article	OPF	JSP	WTHE	AV
[45]	✓	✓	✗	✗
[40]	✓	✗	✗	✗
[28]	✗	✗	✓	✗
[24]	✓	✗	✗	✓
[42]	✗	✓	✓	✓
Our Study	✓	✓	✓	✓

Table 1: Literature table of applicative HF in scheduling studies. OPF=Optimise Productivity and Fatigue, JSP=Job-Shop Scheduling Problem, WTHE=Worker and Task Heterogeneity, AV=Applicative Validation.

[40] dive into fatigue quantification using the principle of energy expenditure and review its usability in a storage location assignment problem where both makespan and fatigue alleviation are considered. Their model does not apply worker heterogeneity but only differentiates EE rates between activities. Their approach looks promising for an extension towards scheduling but requires real-life validation before application.

[28] and [24] both incorporate a combined fatigue-recovery model into a scheduling problem focusing on this practical application, of which the former includes worker and task heterogeneity and the latter considers multiple objectives in a real case study. Finally, [42] incorporate fatiguing, heterogeneous worker (MAEE) and task attributes and three different RA scheduling methods into a heuristic for DRC JSPs. This paper's approach differentiates operations and jobs regarding the fatiguing process. Since the individual, heterogeneous fatiguing process is captured in the recovery time parameter, and there is no notion of workload or fatigue level, the paper does not allow employers to apply these heuristics to their JSP with specific task and workforce characteristics.

2.3. Existing research gaps

The literature review shows two main research gaps in the area:

- The existing methods that can estimate fatigue through energy expenditure require live bodily measurements or have not been previously applied to or validated in settings with both worker and task heterogeneity.
- No existing works combine bi-objective optimisation of productivity and fatigue in a DRC JSP with worker and task heterogeneity.

This thesis fills these research gaps by applying a personalised, predictive fatigue model to a mathematical JSP formulation where real-life DRC OP system constraints are considered. Because this fatigue model has not yet been validated in a similar setting, nor have they been applied to similar scheduling problems, there is a need for real-life applicative validation of the fatigue model and resulting schedules.

3. Modelling approach

In this section, we present the selected fatigue model, followed by the mathematical JSP formulation that it was applied to. Then, we discuss the optimisation approach.

3.1. Fatigue model

Before applying the RA model in 1, we perform the task decomposition method by [27] that was introduced in Section 2.1.2. For the highly similar tasks that shoppers do, this detailed task decomposition method is suitable for estimating the average EE rate during picking, though this is still a simplification of the real world. Each body movement is coupled to a specific equation dictating the energy expenditure of that task. For example, a two-arm lift with load L is represented by:

$$EE = 10^{-2} [0.062BW(h_2 - 0.81) + (3.19L - 0.52S \cdot L)(h_2 - h_1)] \text{ for } 0.81 < h_1 < h_2 \quad (3)$$

, with BW the body weight of the shopper, h_2 the endpoint height of the lift, S the sex of the shopper and h_1 the starting point height of the lift. The formulae from [27] have been adjusted to fit our specific situation, with standard assumptions for h_1 , h_2 and the pick displacement distance. Furthermore, we assume the floor is level, no pull or pushing force is required during picking or placing, and the picking height is constant. Station switching energy and energy spent waiting between picks are only considered in the experiments.

Each elemental movement is also estimated in terms of time, therefore allowing us to take the average \dot{EE} during the pick action. The personal characteristics, a pick speed parameter, different item weights and walking distances ensure this calculation embodies both worker and task heterogeneity. The worker heterogeneity is also considered in the calculation of \dot{EE}_R , also from [27], and MAEE calculation from [26] (see 4).

$$\begin{aligned} MAEE = & 0.0016((60 - 0.55 \cdot AGE) \cdot BW) \quad \text{Men} \\ & 0.0016((48 - 0.37 \cdot AGE) \cdot BW) \quad \text{Women} \end{aligned} \quad (4)$$

3.2. Mathematical job-shop scheduling formulation

In this section, we present a mathematical formulation for the scheduling problem. We use different terminology than we have seen in Section 2.2.1 to correspond to the specific situation at Picnic. This entails the following comparison:

Literature term	Picnic term
Job	(Order) tote
Operation	Order line
Machine	(Picking) station
Worker	Shopper

Our mathematical model is based on the work by [42] and [46], but adds decision variables and constraints to improve the completeness of their formulations. The resulting attributes and assumptions of this model are:

1. Set of *shoppers* process a unique set of *order lines* for a set of *totes* on a set of *stations*.
2. All information on shoppers, stations, totes and order lines is pre-known and fixed.
3. There are no due dates or different priorities between totes or order lines.
4. Order lines are specific items to be picked per tote, which can be any item in the assortment.
5. Any shopper can pick any order line, but only on a subset of stations.
6. Totes are independent of each other; order lines are not: they are subject to precedence constraints based on pre-known item fragility categories.
7. An order line can be picked once, on only one station at a time; there is no interruption or preemption for picking.
8. Waiting is permitted between order lines.
9. Each order line has a specific processing time and rest allowance, both item- and shopper-dependent.
10. Order lines of the same tote cannot be picked simultaneously.
11. Rest allowance must be scheduled directly after picking an order line.
12. Shoppers can switch stations, which costs (travel) time.
13. Stations and shoppers only process one order line at a time and are independent.
14. No buffer, setup time, warehouse or machine availability constraints are considered.

Following the notation in Table 2, the mathematical model can be defined as follows:

$$\min x \frac{C_{max} - C_0}{C_0} + (1 - x) \frac{EE_{max} - EE_0}{EE_0} \quad (5a)$$

$$\text{s.t. } C_{max} = \max_{i \in I; j \in J_i} (C_{i,j}), \quad (5b)$$

$$EE_{max} = \max_{l \in L} \left(\sum_{k \in K} \sum_{j \in J_i} \sum_{i \in I} EE_{i,j,l} \cdot \alpha_{i,j,k,l} \right), \quad (5c)$$

$$C_{i,j} = S_{i,j} + \sum_{l \in L} \sum_{k \in K} (p_{i,j,l} + r_{i,j,l}) \alpha_{i,j,k,l}, \quad (5d)$$

$$\sum_{l \in L} \sum_{k \in K} \alpha_{i,j,k,l} = 1, \quad \forall i \in I; j \in J_i, \quad (5e)$$

$$\sum_{l \in L} \sum_{k \in K_{i,j}} \alpha_{i,j,k,l} = 1, \quad \forall i \in I; j \in J_i, \quad (5f)$$

$$S_{i,j'} \geq C_{i,j} \cdot \beta_{TO:i,j,j'}, \quad \forall i \in I; j \in J_i; j' \in J_i \setminus \{j\}, \quad (5g)$$

$$\beta_{TO:i,j,j'} + \beta_{TO:i,j',j} \leq 1, \quad \forall i \in I; j \in J_i; j' \in J_i \setminus \{j\}, \quad (5h)$$

$$\beta_{TO:i,j,j'} = 0, \quad \forall i \in I; j = j' \in J_i, \quad (5i)$$

$$\sum_{j' \in J_i} \sum_{j \in J_i} [\beta_{TO:i,j,j'}] + 1 = \sum_{l \in L} \sum_{k \in K} \sum_{j \in J_i} \alpha_{i,j,k,l}, \quad \forall i \in I, \quad (5j)$$

$$\sum_{j' \in J_i} \beta_{TO:i,j,j'} \leq 1, \quad \forall i \in I; j \in J_i, \quad (5k)$$

$$\sum_{j' \in J_i} \beta_{TO:i,j',j} \leq 1, \quad \forall i \in I; j \in J_i, \quad (5l)$$

$$S_{i',j'} \geq C_{i,j} \cdot \beta_{ST:i,j,i',j',k}, \quad \forall i, i' \in I; j \in J_i; j' \in J_{i'}; k \in K, \quad (5m)$$

$$\beta_{ST:i,j,i',j',k} + \beta_{ST:i',j',i,j,k} \leq \sum_{l \in L} \alpha_{i,j,k,l} \cdot \sum_{l \in L} \alpha_{i',j',k,l}, \quad \forall i \in I; j \in J_i; [i', j'] \in J \setminus \{[i, j]\}; k \in K, \quad (5n)$$

$$\beta_{ST:i,j,i',j',k} = 0, \quad \forall [i, j] = [i', j'] \in J, \quad (5o)$$

$$\sum_{j' \in J_{i'}} \sum_{i' \in I} \sum_{j \in J_i} \sum_{i \in I} [\beta_{ST:i,j,i',j',k}] + 1 = \sum_{l \in L} \sum_{j \in J_i} \sum_{i \in I} \alpha_{i,j,k,l}, \quad \forall k \in K, \quad (5p)$$

$$\sum_{k \in K} \sum_{j' \in J_{i'}} \sum_{i' \in I} \beta_{ST:i,j,i',j',k} \leq 1, \quad \forall i \in I; j \in J_i, \quad (5q)$$

$$\sum_{k \in K} \sum_{j' \in J_{i'}} \sum_{i' \in I} \beta_{ST:i',j',i,j,k} \leq 1, \quad \forall i \in I; j \in J_i, \quad (5r)$$

$$S_{i',j'} \geq \left(C_{i,j} + \sum_{k \in K} \sum_{k' \in K} s_{k,k'} \cdot \gamma_{i,j,i',j',k,k',l} \right) \cdot \beta_{SH:i,j,i',j',l}, \quad \forall i, i' \in I; j \in J_i; j' \in J_{i'}; l \in L, \quad (5s)$$

$$\beta_{SH:i,j,i',j',l} + \beta_{SH:i',j',i,j,l} \leq \sum_{k \in K} \alpha_{i,j,k,l} \cdot \sum_{k \in K} \alpha_{i',j',k,l}, \quad \forall i \in I; j \in J_i; [i', j'] \in J \setminus \{[i, j]\}; l \in L, \quad (5t)$$

$$\beta_{SH:i,j,i',j',l} = 0, \quad \forall [i, j] = [i', j'] \in J, \quad (5u)$$

$$\sum_{j' \in J_{i'}} \sum_{i' \in I} \sum_{j \in J_i} \sum_{i \in I} [\beta_{SH:i,j,i',j',l}] + 1 = \sum_{k \in K} \sum_{j \in J_i} \sum_{i \in I} \alpha_{i,j,k,l}, \quad \forall l \in L, \quad (5v)$$

$$\sum_{k \in K} \sum_{j' \in J_{i'}} \sum_{i' \in I} \beta_{SH:i,j,i',j',l} \leq 1, \quad \forall i \in I; j \in J_i, \quad (5w)$$

$$\sum_{k \in K} \sum_{j' \in J_{i'}} \sum_{i' \in I} \beta_{SH:i',j',i,j,l} \leq 1, \quad \forall i \in I; j \in J_i, \quad (5x)$$

$$\gamma_{i,j,i',j',k,k',l} = \sum_{j \in J_i} \sum_{i \in I} [\beta_{SH:i,j,i',j',l} \cdot \alpha_{i,j,k,l} \cdot \alpha_{i',j',k',l}], \quad \forall i' \in I; j' \in J_{i'}; k \in K; k' \in K \setminus \{k\}; l \in L, \quad (5y)$$

$$\gamma_{i,j,i',j',k,k',l} = 0, \quad \forall k = k' \in K, \quad (5z)$$

$$\beta_{TO:i,j,j'} \cdot f_{i,j} \leq f_{i,j'}, \quad \forall i \in I; j, j' \in J_i, \quad (5aa)$$

$$S_{i,j} \geq 0, \quad \forall i \in I; j \in J_i, \quad (5ab)$$

$$\alpha_{i,j,k,l}, \beta_{TO:i,j,j'}, \beta_{ST:i,j,i',j',k}, \beta_{SH:i,j,i',j',l}, \gamma_{i,j,i',j',k,k',l} \in \{0, 1\}, \quad \forall i, i' \in I; j \in J_i; j' \in J_{i'}; k, k' \in K; l \in L. \quad (5ac)$$

Here (5a) is the bi-objective to minimise, a combination of C_{max} and EE_{max} coupled by weight parameter x . This objective is a weighted sum with fractional single-objective deviation. There are better methods to investigate the trade-off between two objectives, but this method allows estimation of a Pareto front without

Notation	Description
i, i', i''	Tote index
j, j', j''	Order line index
k, k', k''	Station index
l, l'	Shopper index
I	Set of totes
J	Set of order lines per tote, J_i is a set of all order lines for tote i
$J_{i,j}$	j th order line of tote i
K	Set of stations
$K_{i,j}$	Set of stations where order line $J_{i,j}$ can be picked
L	Set of shoppers
$C_{i,j}$	Completion time of order line $J_{i,j}$
C_{max}	Maximum completion time, makespan
EE_{max}	Maximum sum of required energy expenditure for all individuals
Parameters	
C_0	Optimal value for C_{max} if only optimising for C_{max} (or $x = 1$)
EE_0	Optimal value for EE_{max} if only optimising for EE_{max} (or $x = 0$)
$EE_{i,j,l}$	Energy expenditure for order line $J_{i,j}$ when shopper l is executing it
$f_{i,j}$	Item fragility category of order line $J_{i,j}$
$p_{i,j,l}$	Processing (picking) time of order line $J_{i,j}$ when shopper l is executing it
$r_{i,j,l}$	Recovery time for shopper l after picking order line $J_{i,j}$
$s_{k,k'}$	Switching time between station k and k'
x	Bi-objective weight parameter
Decision variables	
$S_{i,j}$	Starting time of order line $J_{i,j}$ (sec.) measured from start of operation ($t = 0$)
$\alpha_{i,j,k,l} \in \{0, 1\}$	1 if station k and shopper l are selected to process order line $J_{i,j}$, 0 otherwise
$\beta_{TO:i,j,j'} \in \{0, 1\}$	1 if tote i order line $J_{i,j}$ is performed right before $J_{i,j'}$, 0 otherwise
$\beta_{ST:i,j,j',k} \in \{0, 1\}$	1 if order line $J_{i,j}$ is performed right before $J_{i,j'}$ on station k , 0 otherwise
$\beta_{SH:i,j,j',l} \in \{0, 1\}$	1 if order line $J_{i,j}$ is performed before $J_{i,j'}$ by shopper l , 0 otherwise
$\gamma_{i,i',j',k,k',l} \in \{0, 1\}$	1 if shopper l moves from station k to k' to pick order line $J_{i',j'}$ after picking $J_{i,j}$, 0 otherwise

Table 2: Notations, parameters and decision variables for the DRC JSP formulation with fatigue consideration.

many repetitions of the optimisation, which is especially relevant for NP-hard problems (see also Section 3.3) [47, 48]. The two objectives C_{max} and EE_{max} are defined in (5b) and (5c). (5d) calculates the value of auxiliary variable $C_{i,j}$. Constraint (5e) ensures every order line is executed once by one shopper and on one station, while (5f) specifies the subset of stations where this can be done.

Constraints (5g)-(5l), (5m)-(5r) and (5s)-(5x) are sequencing constraint sets on three different levels with high similarity: tote, station and shopper. First, (5g) ensures no two order lines belonging to the same tote can be picked simultaneously, no matter which station and shopper are involved. Therefore, the starting time of that tote's order line picked next is at least the completion time of the previous one. Constraints (5l)-(5l) then define sequence decision variable β_{TO} . Second, (5m)-(5r) ensure the same as (5g)-(5l), but on the station level and with the notion that the order lines are indeed picked at that station. Third, (5s)-(5x) do this on the shopper level, thereby switching the roles of k , K and l , L . One extra addition is the station switching time $s_{k,k'}$, for which we require (5y) and (5z) to define switch decision variable γ . Finally, (5aa) ensures that order lines within one tote are picked in the order of the fragility category, and (5ab) and (5ac) define the types of the decision variables.

3.3. Model optimisation

Unlike the works by [42] and [46] that it was derived from, our formulation is not an example of a mixed-integer linear programming (MILP) problem. These formulations are common in JSP literature and can be solved by most exact solvers. Due to constraints like (5s) and (5y), where we multiply decision variables, not all constraints are linear, and our formulation categorises as a mixed-integer nonlinear programming (MINLP) problem. Although other solution techniques exist and analytical solution techniques may only be suitable for small problem instances [49, 50, 5], we use an exact solver to demonstrate the maximum potential benefit from fatigue consideration in a DRC OP system because constraint and model verification is easier when comparing globally optimal solutions.

Because of the presence of binary variables, the MINLP is non-convex by definition, which leaves us with few options when looking for a solver [51]. The two best-performing options are SCIP and BARON, but open-source academic solver SCIP is the only one that is freely available without restrictions. Other advantages are SCIP's integrated implementation in Python through PySCIP0pt [52], easy iterative constraint generation, insights in the solving process and a wide variety of solver parameter settings. Therefore, we use the PySCIP0pt interface to solve the scheduling problem in SCIP.

SCIP, or Solving Complex Integer Problems, uses a spatial branch-and-bound algorithm with linear relaxations. This allows the solver to find global optima for non-convex MINLPs [53]. SCIP was installed on Windows Subsystem for Linux and executed on an Intel 10700F octa-core 2.9/4.7GHz processor with 16GB DDR4 RAM.

The number of decision variables and constraints is roughly equal to $N_{totes}^2 N_{orderlines}^2 N_{stations}^2 N_{shoppers}$, meaning that these values blow up even for small problem instances. To maximise the problem instance sizes that we can run, we impose three solver speed improvement measures.

First, we run the optimisation process in the SCIP interactive shell to be able to use its concurrent solve option. This means that SCIP uses multiple solvers concurrently, each with a different random seed for the branch-and-bound process. Because the separate solvers can communicate with each other, this can speed up the process because they can traverse the solution space more quickly. Also, with random solver instance creation, the dependency on a favourable set of initial values is decreased by using multiple instances. On the contrary, because this requires the concurrent creation of nodes in the branch-and-bound process, memory is a restricting factor when using many processor cores. We found decent results for low memory usage with four concurrent solvers.

Second, we impose a stopping criterion based on the run time of the process (6000 seconds per optimisation run). As this early stopping does not guarantee a feasible solution, we set feasibility emphasis arguments and check the validity of the solution before saving. If no feasible solution is found within the time limit, the program reruns the optimisation with double the time limit and different solver instances. This is useful when selecting an appropriate run time setting but often unnecessary when a sufficiently long run time was initially set. Generally speaking, SCIP can find near-optimal solutions quickly but cannot guarantee a global optimum when problem instances become larger.

Third, we impose (6) as an additional constraint to limit the search space. This assumes that all order lines are picked by one shopper, for which the longest possible processing time and RA are assumed, plus the maximum travel time after each pick.

$$S_{i,j} \leq \left(\max(p_{i,j,l}) + \max(r_{i,j,l}) + \max(s_{k,k'}) \right) \cdot \sum_{i \in I} |J_i| \quad \forall i \in I; j \in J_i \quad (6)$$

After successful constraint and model output verification, we optimise a series of virtual problem instances. To judge the effectiveness of our bi-objective optimisation approach over the traditional single-objective function with makespan, we estimate the Pareto front for each of these instances. The Pareto front can be represented as a plot where multi-objective optimisation is visualised (see Figure 2). One objective is on the x-axis, and the other is on the y-axis, and each dot represents a single solution. In a minimisation

problem like ours, an ideal solution would be as far as possible in the bottom left corner. However, we are constrained by the feasible solution space, limiting the feasible solutions to the dots in the figure.

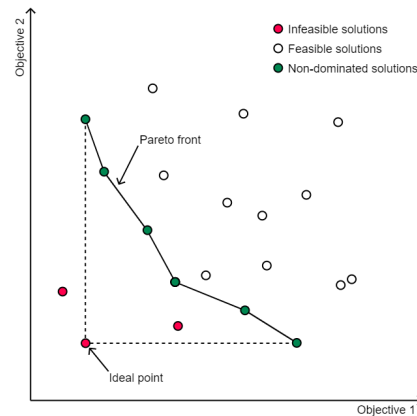


Figure 2: Visualisation of a Pareto front with objectives f_1 and f_2 . Each dot represents a solution to the bi-objective optimisation problem.

If no other solutions exist that, relative to some solution, improve one objective without hurting the other, we call this a non-dominated solution. A line through all non-dominated solutions is called the Pareto front, indicating Pareto-optimal solutions. The Pareto front allows management to weigh the importance of the two objectives and select a solution accordingly. Depending on the priority of the two objectives, the overall preferred solution could be any of the points on the Pareto front.

Our weighted objective is not the perfect way to estimate the Pareto front, as it does not allow finding non-dominated solutions when the Pareto front is concave [48]. However, alternative methods like the ϵ -constraint method do not pose a real advantage when using an exact solver where time constraints limit the solver's capability, as they would require many optimisations to be run.

For each problem instance we run, we first run the two single-objective versions of objective function (5a) to get C_0 and EE_0 . Then, we use these values and vary the bi-objective weight x in 25 equally-spaced steps between 0 and 1 and optimise the resulting MINLP instances. Since we receive global optima from the solver, any solution where x is neither 1 nor 0 will return a Pareto-optimal schedule, allowing us to estimate the Pareto front.

4. Empirical study

We perform experiments to validate the fatigue model and the resulting schedules simultaneously. The experiments are done during normal picking operations in Picnic's OP system, a real-life facility still in ramp-up. This can cause demand fluctuations and disturbances that require constant monitoring, therefore allowing only two participants to be tested at once.

The participants are healthy Picnic employees familiar with the OP system (> 1 month of experience), so learning effects are minimised. The mathematical model includes worker heterogeneity, so we try to include this in the data as much as possible. Therefore, we randomly select shoppers from the working schedule a day before the tests and ask them to participate. We expect 20 participants to be sufficient to represent the variance in the entire employee pool.

4.1. Fatigue model validation

As mentioned in Section 2.3, our method for predictive fatigue estimation requires validation if they are applied to a setting where worker and task heterogeneity are concerned. The task decomposition method from Section 3.1 is no exception; we describe its experimental validation in this section. In Section 2.1, we explained that HR-based $\dot{E}E$ estimation is most suitable for application in real life. Using personal data from participants and mathematical models from [54, 36], we can calculate $\dot{E}E_{HR}$ from the momentary heart rate.

Previous research has shown that the Polar H10 sensor is the best-performing commercially available device to measure heart rate in dynamic activities [55, 56]. This device has electrodes embedded in a chest strap that output an electrocardiogram (ECG) and has been supplied by TU Delft’s TEL research group. The H10 finds the QRS complex of the person’s heartbeat using a detection algorithm on the filtered ECG signal. From consecutive QRS complexes, one can determine the R-R interval, which indicates the time between two heartbeats. This output is filtered once again according to “physiological criteria” before being converted to a heart rate in beats per minute (BPM) [57].

An open-source program called Cardia¹ logs the received BPM on a laptop. The BPM data is preprocessed to return one value per second, thereby using a centred moving average of 5 seconds following the sampling rates used by [58, 59]. The resulting values for $\dot{E}E_{HR}$ form our ground truth for fatigue model evaluation, even though it itself is an estimation of the true $\dot{E}E$. Given the application in a real-life OP system, these inaccuracies are unavoidable.

The fatigue model in Section 3.1 is used to give two different estimates of $\dot{E}E$: $\dot{E}E_{orders}$ and $\dot{E}E_{picks}$. The first is based on the order data for the entire day of the experiment and can be calculated before the start of the experiment, as soon as the participant has filled in some personal information. However, shoppers work only a portion of the day and station switching is common, meaning that the order data does not reflect the actually performed picks for a shopper. The second method, therefore, uses the recorded pick events during the experiment to calculate $\dot{E}E_{picks}$. Other than the input data, the two estimates use the same approach.²

4.2. Schedule validation

With $\dot{E}E_{orders}$ known before the start of the experiments, we can calculate the appropriate RA using (1). To see the effects of this application, we divide the experiments into two blocks of approximately 1.5 hours each. Both scenarios are under everyday circumstances, with the exception of the heart rate monitoring and qualitative fatigue ratings during the experiments. Scenario 2 differs from scenario 1 because it applies the RA principle after every 20 minutes of picking. The RA break is held standing still at the station. Half of the participants perform scenario 1 before scenario 2, and the other half vice versa.

The qualitative fatigue ratings are expressed in the RPE scale by [60], a standard score to evaluate perceived exertion. We gather this data every 10 minutes. At the end of the experiment, participants fill in a NASA TLX form [61], which is another qualitative rating of the perceived effort and workload. We compare the outcomes of these qualitative ratings and the three $\dot{E}E$ estimates between the two scenarios for each individual.

5. Results

In this section, we present the results from the modelling study and empirical study separately.

¹Cardia version V1.2.0.1 <https://github.com/uwburn/cardia>

²More extensive explanation of the data recording and analysis steps can be found in Section 4.3 of this paper’s accompanying thesis report.

5.1. Modelling study

Because of our exact solver, we only run small virtual problem instances through our scheduling model. We find globally optimal solutions within 6000 seconds of run time until the number of decision variables exceeds roughly 2,500. This is already the case for problem instances with 3 totes, 3 order lines per tote, 3 stations and 2 shoppers, a problem size tiny in comparison to real-life problem sizes. However, even for these small problems, we can identify some key outcomes of our developed approach.

Our bi-objective optimisation is able to find non-dominated solutions, although only a few for each problem instance. This can provide granular estimations of the Pareto front in the discrete solution space. The exact shape remains unknown, but the handful of points in the Pareto front plots can already assist managerial decision-making. Note that one point in the Pareto front can be the solution space representation of several unique schedules with the same objective 'value pairs'.

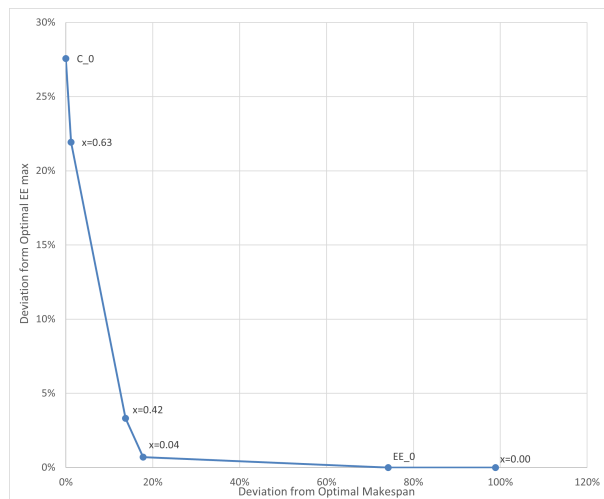


Figure 3: Estimated Pareto front with only globally optimal solutions. Each dot may represent multiple schedules, but only one objective value pair.

In Table 3 we show the results for a problem instance with 3 totes, 3 order lines per tote, 3 stations and 2 shoppers, leading to 1995 decision variables and 2449 constraints. We get 7 unique objective value pairs and a maximum run time of 168 seconds. Four points are confirmed Pareto-optimal because they represent globally optimal solutions that are non-dominated.

Run	C_{max} [s]	EE_{max} [kcal]	Pareto-optimal
C_0	131.57	74.22	✓
EE_0	229.20	58.18	✗
0.00	261.68	58.18	✗
0.04	154.94	58.59	✓
0.42	149.63	60.11	✓
0.63	133.09	70.94	✓

Table 3: Unique solution value pairs.

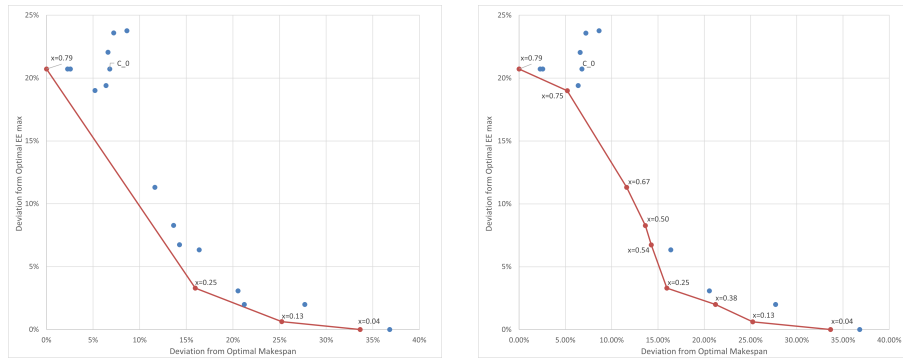


Figure 4: Two different shapes for the Pareto front estimate without globally optimal solutions.

Figure 3 shows the resulting estimate of the Pareto front, with the relative deviation of C_0 on the x-axis and that of EE_0 on the y-axis. Solutions as close to 0% deviation are preferred. In comparison to single-objective optimisation for the makespan, run C_0 , we can see that we are now able to find some alternative solutions. This allows us to make improvements in terms of EE distribution that single-objective optimisation would not have allowed for, such as a 10% better EE distribution against a 1.5% makespan increase. Using weight parameter x , management could decide upon the relative priority of these two objectives. We see comparable results for other small problem instances, with different shapes of the estimated Pareto front.

If we compare the first Pareto-optimal solution in these instances to the original, single-objective optimum (C_0), we can calculate the steepness of the line piece between these two solutions (e.g., ' C_0 ' - ' $x = 0.63$ ' in Figure 3). For 53% of all problem instances with purely global optima we considered, this line piece has a steepness of 4 or higher, meaning the fractional decrease of EE_{max} is more than four times the fractional increase of C_{max} that accompanies it. The steepness number is even as high as 700 in one case, where a 49% decrease in EE_{max} required a 0.07% increase of the makespan.

Larger problem instances, which have more decision variables and thus more degrees of freedom in scheduling, are seen to deliver more feasible solutions in optimisation, but these solutions are not guaranteed to be globally optimal. In general, this means there are no solutions that can be guaranteed to be Pareto-optimal. If we only look at the solutions that are not dominated by any other known solution - meaning they could be non-dominated, but cannot be guaranteed as such - we can also estimate the Pareto front in this case, as shown in Figure 4. However, the front can be drawn in multiple ways, even in a concave shape.

5.2. Empirical study

First of all, men and women of different ages are evenly represented in our population sample. Also, their heart rate records show no remarkable anomalies in comparison to population databases. Predicted $\dot{E}E_R$ and $MAEE$ values vary between participants. We also see that $\dot{E}E_R$ and $\dot{E}E_{orders}$ (for the same order data) increase slightly with the $MAEE$, but that the $MAEE$ values tend to relatively exceed these by a larger margin for higher $MAEE$ and that these relations differ per individual.

Figure 5 shows visually the distribution and relative relation between the different estimates for $\dot{E}E$. Generally, we see that the individuals that are below the median in one of the box plots are also below the median in the others and vice versa, especially when comparing $\dot{E}E_{picks}$ and $\dot{E}E_{orders}$.

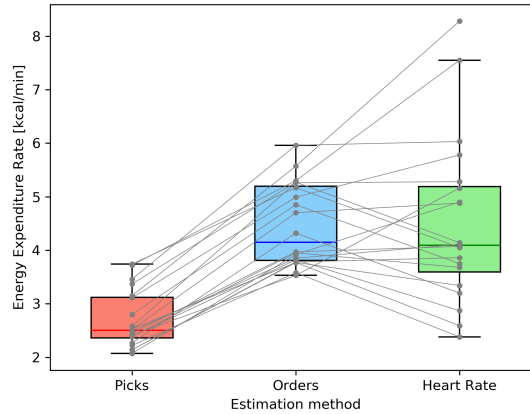


Figure 5: Three interconnected box plots representing the average values from the different estimation methods. Each line represents one individual during the tests.

As we treat $\dot{E}E_{HR}$ as the most accurate estimate of the energy expenditure rate in these circumstances, we evaluate the accuracy of $\dot{E}E_{orders}$ and $\dot{E}E_{picks}$ by comparing their estimates to this value. For $\dot{E}E_{orders}$, this results in a root root-mean-square error (RMSE) of 1.14 kcal/min, while $\dot{E}E_{picks}$ has an RMSE of 2.20 kcal/min. These values are in agreement with the visual representation in Figure 5.

This means that the least accurate pick data leads to a more accurate $\dot{E}E$ estimation. To find out if a general bias causes these inaccurate estimations, we look at the Pearson correlation coefficient between $\dot{E}E_{HR}$ and $\dot{E}E_{picks}$. The Pearson correlation coefficient is a number between 1 and -1, indicating how one function's behaviour relates to another. If two functions show highly similar behaviour, the correlation coefficient will be close to 1, while opposite behaviour results in values close to -1. If there is no relation between two functions and they behave irrespective of each other, the correlation coefficient is close to 0. On a shift level, we see a very low correlation for all participants, some even slightly negative. This means that on a shift level, $\dot{E}E_{picks}$ is a bad estimator of $\dot{E}E_{HR}$.

If we look at a smaller time span, the results relate more to the time span in which we would consider RA calculations during application. In Figure 6 we plot the deviation of the average values of $\dot{E}E_{HR}$ and $\dot{E}E_{picks}$ and the correlation between the two for each 10 minute interval in the measurement data for all participants. Although some good estimates exist, the vast majority has a negative deviation of more than 1 kcal/min and we also see a high number of intervals with low or negative correlation coefficients.

A closer look at the fluctuation in pick demand per 10 minute interval reveals a possible explanation for this behaviour (see Figure 7). The demand fluctuation exists between participants, within the experimental data of a single participant, and between the two scenarios for a given participant. We see varying numbers, but most values are below 60 picks per 10 minutes. According to the estimated picking time of 7.9 to 12.7 seconds from our model, this would be the average number of picks per 10 minutes that a shopper could process. However, if we categorise the data in Figure 6 for each of these pick demand bins, we see no significant differences in the locations of the points.

The high demand fluctuation makes a comparison between scenarios 1 and 2 very difficult, as demand has not been consistent during the experiment. Pick demand is the basis for $\dot{E}E_{picks}$ calculation, therefore making this measure irrelevant to the schedule validation. However, neither $\dot{E}E_{HR}$ nor the qualitative

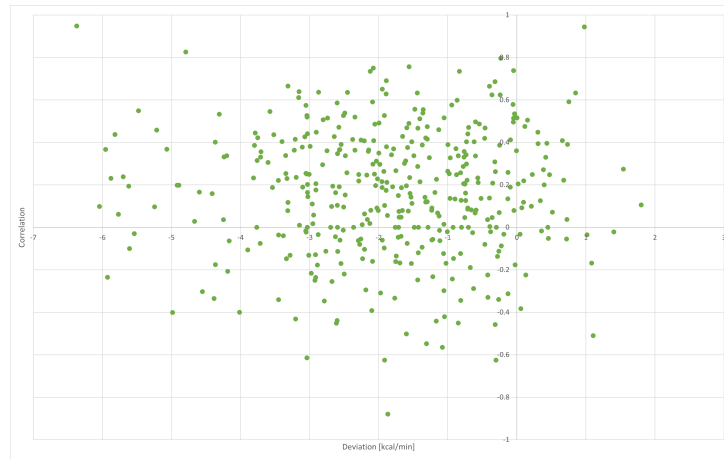


Figure 6: Overview of all 10 minute intervals for all participants. Each dot represents one measurement interval, with the average $\overline{EE}_{HR} - \overline{EE}_{picks}$ deviation on the x-axis and the $\overline{EE}_{HR} - \overline{EE}_{picks}$ correlation coefficient on the y-axis. Good predictions are at the top of the figure and close to the y-axis.

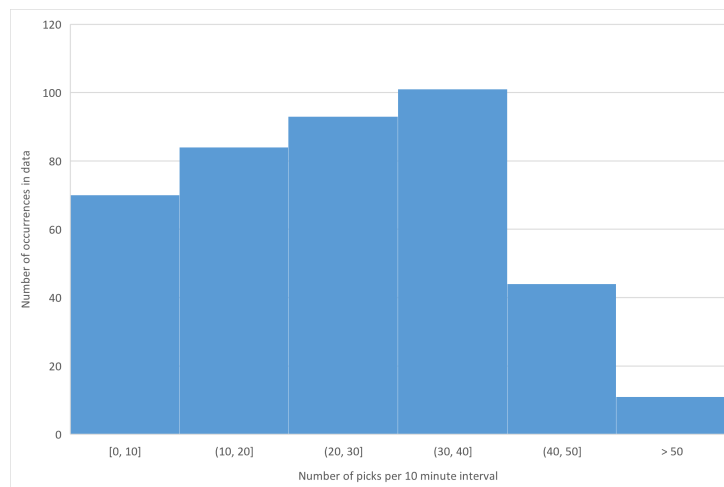


Figure 7: Overview of all pick demands per 10 minute interval for all participants. The data has been filtered for periods where there was constant heart rate registration for 10 minutes, then categorised into bins of picks per 10 minute intervals.

fatigue ratings show significant differences between the two scenarios.³

³More extensive and insightful results relating to individual event response comparison can be found in Sections 5.2.2, 5.5.3 and 5.2.4 of this paper's accompanying thesis report.

6. Discussion

Section 5.1 shows that we are able to draw a rough estimate of the Pareto front for our bi-objective optimisation problem, as long as the solver returns a guaranteed global optimum. This is an expected result given our weighted sum approach [48]. Even though our methods only allow for a few unique non-dominated solutions to be found, we are able to improve the energy expenditure distribution in comparison to single-objective optimisation solution ' C_0 '. This choice would depend on managerial insights, but the steep Pareto front line pieces that we found indicate considerable potential for implementation. This is in accordance with our hypothesis.

The results of the empirical study in Section 5.2 show a good selection of participants for this study. Also, the differences and relations between $MAEE$, $\dot{E}E_R$ and $\dot{E}E_{orders}$ show the value of fatigue-conscious scheduling, also highlighting the importance of our personalised fatigue modelling approach over the assumptions made by [40]. However, our predictive fatigue model shows large inaccuracies in estimating the fatigue throughout all experiments. This is the case on a shift level but also for 10 minute intervals. A first valid suspicion could relate these inaccuracies to the highly volatile pick demand, but our results show no clear relation between the pick demand and estimation accuracy. This contradicts our expectations of the estimation method's performance.

Shoppers do much more than just order picking, such as pallet loading, cleaning and opening boxes. We estimate that about 25% of the shift is not spent according to our task decomposition description. However, the task intensity of the other tasks, we believe, is similar to that of picking. This would explain the better estimations of $\dot{E}E_{orders}$, as this assumes a continuous picking workload. Furthermore, lots of operational challenges occurred during the experiments, such as system downtime and shopper reassignment, troubling the experimental circumstances.

These factors could also have influenced the schedule validation results, where we see no clear differences between scenarios or the two groups of participants in terms of $\dot{E}E_{HR}$ or the qualitative fatigue ratings. Therefore, we cannot conclude that the positive results from our modelling study translate to real-life applications.

This study has five key limitations, which we summarise with accompanying suggestions for future research:

- First, our bi-objective optimisation approach is unsuitable for finding a good Pareto front estimate that is fit for providing managerial insights in application. Given the limitations of our exact solver, this is not a problem that another bi-objective optimisation approach would have solved. Now that the principle behind our mathematical formulation has shown to be valid, future work could improve on the practical applicability by considering a different solving approach that can exhibit the model's full potential in realistic problem instances. Other extensions could include further model detailing, different optimisation objectives, fatigue-dependent productivity and long-term studies excluding external factors.
- Also, we choose to model energy expenditure, the precursor of fatigue, and therefore not fatigue itself. The model's inputs are based on population data from 1981 which shows inherent flaws, and we look at its output as an absolute value. Looking at energy expenditure or fatigue overload may be more respective to an individual's well-being than the total sum we included in the bi-objective. This relates closely to an ethical question regarding equity and equality: do we really want to schedule heavier tasks to people that can handle them better? And, if so, do we then look at absolute or relative fatigue overload? Both questions are relevant for implementation and offer an interesting direction for future work.
- FCA is by no means a lab environment, and it is difficult to predict or even measure the actual work done by shoppers; station switching is common, pick demand fluctuates a lot, and there are long

periods of inactivity due to operational issues. This also influences the capabilities of $\dot{E}E_{orders}$, as this calculation takes single station demand data, while it was rare for shoppers to only service one station during their work. Future research could capitalise on the testing environment that FCA offers by taking these external influences into account in a long-term study.

- Fourth, the influence of the demand fluctuation cannot be separated from the possible influences of the breaks or different scenario order, especially if there is so little pick demand that there is often involuntary resting outside the scheduled RA breaks. Also, given the unpredictable nature of the demand, there is a low likelihood that the RA estimations could be accurate enough to create the working circumstances that scenario 2 requires. In a more mature OP system, the same tests could lead to different results.
- Finally, our fatigue modelling approach is susceptible to external influences on the heart rate, too high $\dot{E}E_R$ calculation [27], too low $MAEE$ values [26], too high HR_{max} [54], or too high calculation of $\dot{E}E_{HR}$ [36]. These four models rely on population averages that may not apply to a small sample of individuals [62] or may have become outdated. With only 20 participants partaking in the experiments, this could have led to inappropriate assumptions and calculations, leading to inaccuracies in not only the predicted $\dot{E}E$ but also the measured values. Future studies could benefit from having more participants on a longer term, thereby being able to filter out any anomalies in the data.

Apart from the future work opportunities in these bullet points, we share some managerial recommendations. Fatigue in the workplace is a multidimensional construct and partially results from the work environment. Occupational fatigue has proven links to productivity and a decrease in employee well-being, in both the short and the long term. Our work shows that occupational fatigue can be considered in operational decision-making and that there are real benefits in doing so. The implementation of our methods allows for improvements at virtually no cost.

Although we identify multiple opportunities for further research in human fatigue modelling, measurement and incorporation in scheduling problems, we do not have to wait for this research to be finished. The substantial potential benefits of human factors consideration in the working environment motivate employers to start applying the learnings from the current state of research sooner rather than later.

We have seen that pick- and order-based $\dot{E}E$ estimations do not allow for accurate predictions or fast application to other activity types, especially for the type of environment we considered. Instead, heart rate measurements allow operations management to quantify the relative physical workload per type of task in the fulfilment centre. This can provide quantitative insights to consider during workforce planning. A good start may be to do this on a granular level as opposed to the meticulous level that our JSP formulation describes. Given the current operational challenges, pick-level task assignment is out of reach for application, while activity-level job rotation is a realistic option. Job rotation is not actively applied in present-day operations, but it could have benefits that exceed fatigue alleviation: worker learning, ordered break schedules, and more diverse work. We think these benefits can improve job satisfaction and employee retention rates.

The finer scope of our developed JSP formulation is not fit-for-purpose in the current circumstances but may become a relevant topic in future operational decision-making. In conjunction with the knowledge that academic research has gathered in the meantime, a more mature OP system could benefit from a detailed approach towards scheduling with human factors. In that case, ethical aspects of the implementation should also be taken into account. Our fatigue quantification method through pick and order data, or heart rate, in particular, requires gathering and storing sensitive personal information from employees. Although this information can be normalised and anonymised, we believe that the potential risks for individuals may not be taken into consideration by every employer.

7. Conclusion

Increasing automation in the work environment can offer new solutions to the operational challenges in partially automated OP systems. Although methods exist to quantify fatigue and consider it in a scheduling

context, these still need to find their way into real-life applications. This thesis presents a detailed formulation for the job-shop scheduling problem in a cobotic order picking system, considering physical workload distribution and total makespan simultaneously. Thereby we bridge the gap between theoretical scheduling problem approaches and operational reality in the context of human fatigue. Our work establishes the first steps towards human fatigue consideration in scheduling for real-life operations.

Based on the results in sections 5.1 and 5.2, this study makes four main contributions:

- First, we combine existing energy expenditure and rest allowance models into a hybrid energy expenditure quantification model that can predict resting requirements. This means that this model can be used in a DRC job-shop scheduling problem without live measurements. With a detailed task element description, this model can be applied to a multitude of scheduling problems. The level of detail and worker heterogeneity consideration that our approach possesses has not been shown prior in the scheduling literature.
- Second, we propose a detailed mathematical formulation for the JSP in a real-life order picking system, incorporating realistic system constraints and a bi-objective that allows for simultaneous consideration of makespan and physical workload. This approach allows us to find schedules that would not have been found using single-objective optimisation which is common in existing JSP literature. These alternative solutions to the scheduling problem can offer better distribution of physical workload without hurting the total production time too much. For scheduling problems in DRC system contexts, existing literature has not shown our level of detail and bi-objective approach towards the scheduling problem.
- Third, we evaluate the practical applicability of our energy expenditure quantification approach in a real-life OP operation. Our results show that neither our order- and pick-based EE estimations provide accurate predictions compared to the heart rate-based estimation. This may partly be caused by the large variation in pick demand and personal characteristics for the 20 participants, but could also suggest that our estimation method is unsuitable for accurate prediction. Given the experimental difficulties and missing elements in our task decomposition, we are inconclusive about the potential of our energy expenditure quantification approach in a more structured environment.
- Fourth, we investigate the effects that rest allowance application during work has on the energy expenditure and subjective fatigue ratings. We cannot find significant differences due to external influences during the experiments. This remains an open topic for research, as it could quantify the potential benefits of appropriate break schedules.

From these contributions, we can draw the following conclusions:

- We find that energy expenditure-based fatigue models can best be applied in scheduling problem formulations. In combination with the rest allowance principle, this allows for a personalised approach to fatigue in a scheduling context. The applied predictive fatigue model allows for detailed task decomposition without prior measurements, able to evaluate energy expenditure requirements for various tasks. This is in accordance with our hypothesis.
- Our modelling study shows that the proposed bi-objective JSP formulation can successfully generate schedules for partially automated OP systems while considering fatigue and productivity. As expected, the results are promising for application, although our results are limited to small problem instances. We argue that the mathematical formulation could benefit from a different solution approach, without the need for reformulation.
- Our empirical study shows that the applied predictive energy expenditure model was inaccurate in estimating energy expenditure rates in a real operation and that heart rate sensors are a more viable

approach at this time. Also, we find no significant changes in measured or qualitative fatigue after applying the generated fatigue-conscious schedules, in contrast with the results from the modelling study. Both results are in contrast with our expectations prior to this research and offer opportunities for further research.

7.1. Final remarks

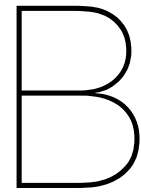
In conclusion, this thesis presents a detailed formulation for the job-shop scheduling problem in a dual-resource constrained order picking system, considering physical workload distribution and total makespan simultaneously. Our approach provides an answer to the question of how human fatigue models can be implemented in the scheduling problem and highlights the importance of verifying these models in real-life situations. Our results also highlight the potential discrepancies in rest allowance calculation through pick data and heart rate. These findings have practical implications for scheduling in real-life operations, providing management with alternative schedules that consider productivity and employee physical strain. This thesis sets the first steps towards human fatigue consideration in scheduling for real-life operations and opens up avenues for further research in this field.

References

- [1] D. Coppola, E-commerce worldwide - statistics & facts, Statista (10 2021).
URL <https://tinyurl.com/5n8vjfb4>
- [2] R. de Koster, T. Le-Duc, K. J. Roodbergen, Design and control of warehouse order picking: A literature review, *European Journal of Operational Research* 182 (2) (2007) 481–501. doi:10.1016/j.ejor.2006.07.009.
- [3] C. Verschuere, D. Lübbli, F. Laizet, G. D’Auria, T. Wachinger, V. Simmons, J. Nyssens, R. Vallöf, *Disruption and Uncertainty: The State of Grocery Retail 2021: Europe*, Tech. Rep. November, McKinsey & Company (3 2021).
- [4] M. Thüner, H. Zhang, M. Stevenson, F. Costa, L. Ma, Worker assignment in dual resource constrained assembly job shops with worker heterogeneity: an assessment by simulation, *International Journal of Production Research* 58 (20) (2020) 6336–6349. doi:10.1080/00207543.2019.1677963.
- [5] H. Xiong, S. Shi, D. Ren, J. Hu, A survey of job shop scheduling problem: The types and models, *Computers & Operations Research* 142 (2022) 105731. doi:10.1016/j.cor.2022.105731.
- [6] National Safety Council, What is Fatigue Costing Your Company? - National Safety Council (2022).
URL <https://www.nsc.org/work-safety/safety-topics/fatigue/cost>
- [7] Z. Sedighi Maman, M. A. Alandari Yazdi, L. A. Cavuoto, F. M. Megahed, A data-driven approach to modeling physical fatigue in the workplace using wearable sensors, *Applied Ergonomics* 65 (2017) 515–529. doi:10.1016/j.apergo.2017.02.001.
- [8] B. A. Kadir, O. Broberg, C. S. d. Conceição, Current research and future perspectives on human factors and ergonomics in Industry 4.0, *Computers and Industrial Engineering* 137 (2019) 106004. doi:10.1016/j.cie.2019.106004.
- [9] S. Winkelhaus, E. H. Grosse, S. Morana, Towards a conceptualisation of Order Picking 4.0, *Computers and Industrial Engineering* 159 (2021) 107511. doi:10.1016/j.cie.2021.107511.
- [10] W. P. Neumann, J. Dul, Human factors: Spanning the gap between OM and HRM, *International Journal of Operations and Production Management* 30 (9) (2010) 923–950. doi:10.1108/01443571011075056.
- [11] V. Vijayakumar, F. Sgarbossa, A literature review on the level of automation in picker-to-parts order picking system: Research opportunities, *IFAC-PapersOnLine* 54 (1) (2021) 438–443. doi:10.1016/j.ifacol.2021.08.050.
- [12] E. H. Grosse, C. H. Glock, W. P. Neumann, Human factors in order picking: a content analysis of the literature, *International Journal of Production Research* 55 (5) (2017) 1260–1276. doi:10.1080/00207543.2016.1186296.
- [13] R. de Koster, *Warehousing 2030*, in: R. Merkert, K. Hoberg (Eds.), *Global Logistics and Supply Chain Strategies for the 2020s: Vital Skills for the Next Generation*, Springer International Publishing, Cham, 2023, pp. 243–260. doi:10.1007/978-3-030-95764-3_14.
- [14] K. Jongbloed, E. Van der Kruk, R. De Vos, J. Harlaar, Defining and measuring fatigue in endurance sports, Ph.D. thesis, Delft University of Technology, Delft (1 2022).
- [15] K. El Ahrache, D. Imbeau, B. Farbos, Percentile values for determining maximum endurance times for static muscular work, *International Journal of Industrial Ergonomics* 36 (2) (2006) 99–108. doi:10.1016/j.ergon.2005.08.003.
- [16] M. Y. Jaber, W. P. Neumann, Modelling worker fatigue and recovery in dual-resource constrained systems, *Computers and Industrial Engineering* 59 (1) (2010) 75–84. doi:10.1016/j.cie.2010.03.001.
- [17] B. Daria, C. Martina, P. Alessandro, S. Fabio, V. Valentina, Fatigue and recovery: research opportunities in order picking systems, *IFAC-PapersOnLine* 50 (1) (2017) 6882–6887. doi:10.1016/j.ifacol.2017.08.1211.
- [18] L. Rose, M. Ericson, B. Glimskär, B. Nordgren, Ergo-index-development of a model to determine pause needs after fatigue and pain reactions during work, *Computer Applications in Ergonomics, Occupational Safety and Health* (1992) 461–468.
- [19] M. A. Abdous, X. Delorme, D. Battini, F. Sgarbossa, S. Berger-Douce, Multi-objective optimization of assembly lines with workers’ fatigue consideration, *IFAC-PapersOnLine* 51 (11) (2018) 698–703. doi:10.1016/j.ifacol.2018.08.400.

- [20] S. Konz, Work/rest: Part II - The scientific basis (knowledge base) for the guide, *International Journal of Industrial Ergonomics* 22 (1-2) (1998) 73–99. doi:10.1016/S0169-8141(97)00069-3.
- [21] M. Y. Jaber, Z. S. Givi, W. P. Neumann, Incorporating human fatigue and recovery into the learning-forgetting process, *Applied Mathematical Modelling* 37 (12-13) (2013) 7287–7299. doi:10.1016/j.apm.2013.02.028.
- [22] K. El ahrache, D. Imbeau, Comparison of rest allowance models for static muscular work, *International Journal of Industrial Ergonomics* 39 (1) (2009) 73–80. doi:10.1016/j.ergon.2008.10.012.
- [23] F. Fruggiero, S. Riemma, Y. Ouazene, R. Macchiaroli, V. Guglielmi, Incorporating the Human Factor within Manufacturing Dynamics, *IFAC-PapersOnLine* 49 (12) (2016) 1691–1696. doi:10.1016/j.ifacol.2016.07.825.
- [24] S. Finco, M. Calzavara, F. Sgarbossa, I. Zennaro, Including rest allowance in mixed-model assembly lines, *International Journal of Production Research* 59 (24) (2021) 7468–7490. doi:10.1080/00207543.2020.1843731.
- [25] A. D. Price, Calculating relaxation allowances for construction operatives - Part 1: Metabolic cost, *Applied Ergonomics* 21 (4) (1990) 311–317. doi:10.1016/0003-6870(90)90202-9.
- [26] C. G. De Souza E Silva, B. A. Franklin, D. E. Forman, C. Gil, S. Araújo, Influence of age in estimating maximal oxygen uptake, *Journal of Geriatric Cardiology* 13 (2016) 126–131. doi:10.11909/j.issn.1671-5411.2016.02.010.
- [27] A. Garg, D. B. Chaffin, G. D. Herrin, Prediction of metabolic rates for manual materials handling jobs, *American Industrial Hygiene Association Journal* 39 (8) (1978) 661–674. doi:10.1080/0002889778507831.
- [28] M. Calzavara, A. Persona, F. Sgarbossa, V. Visentin, A model for rest allowance estimation to improve tasks assignment to operators, *International Journal of Production Research* 57 (3) (2019) 948–962. doi:10.1080/00207543.2018.1497816.
- [29] T. Steinebach, J. Wakula, A. Mehmedovic, The Influence of an Ergonomic Storage Location Assignment on Human Strain in Manual Order Picking, in: N. L. Black, W. P. Neumann, I. Noy (Eds.), *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021)*, Springer International Publishing, Cham, 2021, pp. 511–521.
- [30] D. Battini, S. Finco, F. Sgarbossa, Human-oriented assembly line balancing and sequencing model in the industry 4.0 era, *International Series in Operations Research and Management Science* 289 (2020) 141–165. doi:10.1007/978-3-030-43177-8_8.
- [31] A. C. Hackney, Measurement Techniques for Energy Expenditure, Exercise, Sport, and Bioanalytical Chemistry (2016) 33–42. doi:10.1016/b978-0-12-809206-4.00013-5.
- [32] M. Calzavara, A. Persona, F. Sgarbossa, V. Visentin, A device to monitor fatigue level in order-picking, *Industrial Management and Data Systems* 118 (4) (2018) 714–727. doi:10.1108/IMDS-05-2017-0182.
- [33] R. Li, P. Deurenberg, J. G. Hautvast, A critical evaluation of heart rate monitoring to assess energy expenditure in individuals, *The American Journal of Clinical Nutrition* 58 (5) (1993) 602–607. doi:10.1093/AJCN/58.5.602.
- [34] J. de la Riva, A. I. Garcia, R. M. Reyes, A. Woocay, Methodology to Determine Time Allowance by Work Sampling Using Heart Rate, *Procedia Manufacturing* 3 (2015) 6490–6497. doi:10.1016/J.PROMFG.2015.07.934.
- [35] Z. Sedighi Maman, Y. J. Chen, A. Baghdadi, S. Lombardo, L. A. Cavuoto, F. M. Megahed, A data analytic framework for physical fatigue management using wearable sensors, *Expert Systems with Applications* 155 (2020) 113405. doi:10.1016/J.ESWA.2020.113405.
- [36] H. K. Hilloskorpi, M. E. Pasanen, M. G. Fogelholm, R. M. Laukkanen, A. T. Mänttari, Use of heart rate to predict energy expenditure from low to high activity levels, *International Journal of Sports Medicine* 24 (5) (2003) 332–336. doi:10.1055/S-2003-40701/ID/25.
- [37] S. Hajifar, H. Sun, F. M. Megahed, L. A. Jones-Farmer, E. Rashedi, L. A. Cavuoto, A forecasting framework for predicting perceived fatigue: Using time series methods to forecast ratings of perceived exertion with features from wearable sensors, *Applied Ergonomics* 90 (2021). doi:10.1016/j.apergo.2020.103262.
- [38] A. Aryal, A. Ghahramani, B. Becerik-Gerber, Monitoring fatigue in construction workers using physiological measurements, *Automation in Construction* 82 (2017) 154–165. doi:10.1016/J.AUTCON.2017.03.003.
- [39] J. A. Álvarez-García, B. Cvetkovic, M. Lustrek, A survey on energy expenditure estimation using wearable devices, *ACM Computing Surveys* 53 (5) (2020). doi:10.1145/3404482.
- [40] D. Battini, C. H. Glock, E. H. Grosse, A. Persona, F. Sgarbossa, Human energy expenditure in order picking storage assignment: A bi-objective method, *Computers & Industrial Engineering* 94 (2016) 147–157. doi:10.1016/J.CIE.2016.01.020.
- [41] M. Calzavara, C. H. Glock, E. H. Grosse, F. Sgarbossa, An integrated storage assignment method for manual order picking warehouses considering cost, workload and posture, <https://doi.org/10.1080/00207543.2018.1518609>.
- [42] N. Berti, S. Finco, O. Battaia, X. Delorme, Ageing workforce effects in Dual-Resource Constrained job-shop scheduling, *International Journal of Production Economics* 237 (2021) 108151. doi:10.1016/j.ijpe.2021.108151.
- [43] M. Šeda, Mathematical Models of Flow Shop and Job Shop Scheduling Problems, *International Journal of Physical and Mathematical Sciences* 1 (7) (2007) 307–312.
- [44] D. J. Hoitomt, P. B. Luh, K. R. Pattipati, A Practical Approach to Job-Shop Scheduling Problems, *IEEE Transactions on Robotics and Automation* 9 (1) (1993) 1–13. doi:10.1109/70.210791.
- [45] M. Othman, G. J. Gouw, N. Bhuiyan, Workforce scheduling: A new model incorporating human factors, *Journal of Industrial Engineering and Management* 5 (2) (2012) 259–284. doi:10.3926/jiem.451.
- [46] M. Yazdani, M. Zandieh, R. Tavakkoli-Moghaddam, F. Jolai, Two meta-heuristic algorithms for the dual-resource constrained flexible job-shop scheduling problem, *Scientia Iranica E* 22 (3) (2015) 1242–1257.
- [47] R. T. Marler, J. S. Arora, Survey of multi-objective optimization methods for engineering, *Structural and Multidisciplinary Optimization* 26 (6) (2004) 369–395. doi:10.1007/S00158-003-0368-6/METRICS.
- [48] R. L. Becerra, C. A. Coello Coello, Solving hard multiobjective optimization problems using ϵ -constraint with cultured differential evolution, in: *Parallel Problem Solving from Nature-PPSN IX*, Vol. 4193 LNCS, Springer, Berlin, Heidelberg, 2006, pp.

- 543–552. doi:10.1007/11844297_55/COVER.
- [49] D. W. Sellers, Survey of approaches to the job shop scheduling problem, *Proceedings of the Annual Southeastern Symposium on System Theory* (1996) 396–400doi:10.1109/SSST.1996.493536.
- [50] J. Xu, X. Xu, S. Q. Xie, Recent developments in Dual Resource Constrained (DRC) system research, *European Journal of Operational Research* 215 (2) (2011) 309–318. doi:10.1016/j.ejor.2011.03.004.
- [51] J. Kronqvist, D. E. Bernal, A. Lundell, I. E. Grossmann, A review and comparison of solvers for convex MINLP, *Optimization and Engineering* 20 (2019) 397–455. doi:10.1007/s11081-018-9411-8.
- [52] S. J. Maher, M. Miltenberger, J. P. Pedroso, D. Rehfeldt, R. Schwarz, F. Serrano, PySCIPOpt: Mathematical Programming in Python with the SCIP Optimization Suite, in: *Mathematical Software - ICMS 2016*, Vol. 9725, 2016, p. 301 – 307. doi:10.1007/978-3-319-42432-3_37.
- [53] K. Bestuzheva, M. Besançon, W.-K. Chen, A. Chmiela, T. Donkiewicz, J. Van Doormalen, L. Eifler, O. Gaul, G. Gamrath, A. Gleixner, L. Gottwald, C. Graczyk, K. Halbig, A. Hoen, C. Hojny, R. Van Der Hulst, T. Koch, M. Lübbecke, S. J. Maher, F. Matter, E. Mühmer, B. Müller, M. E. Pfetsch, D. Rehfeldt, S. Schlein, F. Schlösser, F. Serrano, Y. Shinano, B. Sofranac, M. Turner, S. Vigerske, F. Wegscheider, P. Wellner, D. Weninger, J. Witzig, *The SCIP Optimization Suite 8.0* (2021).
- [54] H. Tanaka, K. D. Monahan, D. R. Seals, Age-predicted maximal heart rate revisited, *Journal of the American College of Cardiology* 37 (1) (2001) 153–156. doi:10.1016/S0735-1097(00)01054-8.
- [55] S. R. Pasadyn, M. Soudan, M. Gillinov, P. Houghtaling, D. Phelan, N. Gillinov, B. Bittel, M. Y. Desai, Accuracy of commercially available heart rate monitors in athletes: a prospective study, *Cardiovascular Diagnosis and Therapy* 9 (4) (2019) 379. doi:10.21037/CDT.2019.06.05.
- [56] J. D. Stone, H. K. Ulman, K. Tran, A. G. Thompson, M. D. Halter, J. H. Ramadan, M. Stephenson, V. S. Finomore, S. M. Galster, A. R. Rezaei, J. A. Hagen, Assessing the Accuracy of Popular Commercial Technologies That Measure Resting Heart Rate and Heart Rate Variability, *Frontiers in Sports and Active Living* — www.frontiersin.org 3 (585870) (3 2021). doi:10.3389/fspor.2021.585870.
- [57] Polar Research and Technology, Polar H10 Heart Rate Sensor System [White Paper], Tech. rep., Polar Research and Technology, Kempele (11 2019).
- [58] J. C. Barbero-Alvarez, V. M. Soto, V. Barbero-Alvarez, J. Granda-Vera, Match analysis and heart rate of futsal players during competition, <https://doi.org/10.1080/02640410701287289> 26 (1) (2008) 63–73. doi:10.1080/02640410701287289.
- [59] J. Ingebrigtsen, M. Bendiksen, M. B. Randers, C. Castagna, P. Krstrup, A. Holtermann, Yo-Yo IR2 testing of elite and sub-elite soccer players: Performance, heart rate response and correlations to other interval tests, <https://doi.org/10.1080/02640414.2012.711484> 30 (13) (2012) 1337–1345. doi:10.1080/02640414.2012.711484.
- [60] G. A. Borg, Psychophysical Bases of Perceived Exertion, *Medicine & Science in Sports & Exercise* 14 (5) (1982) 377–381.
- [61] S. G. Hart, L. E. Staveland, Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, *Advances in Psychology* 52 (C) (1988) 139–183. doi:10.1016/S0166-4115(08)62386-9.
- [62] J. Achten, A. E. Jeukendrup, Heart Rate Monitoring Applications and Limitations, *Sports Med* 33 (7) (2003) 517–538.



Workload questionnaire report

To get an indication of the workload and fatigue experienced by shoppers, a questionnaire was set up for the shoppers at FCA's zone pick during normal circumstances. This document presents and discusses the findings of this study.

B.1. Methods

During normal picking operations, shoppers were asked to fill in a questionnaire consisting of three parts. This is supposed to be done before starting their shift or after taking a break of 15-20 minutes [Aryal et al. \(2017\)](#).

1. Before they start picking, shoppers filled in a list of questions regarding their age, gender and experience at Picnic, followed by a selection of questions from the validated questionnaire set of COPSOQ ([Llorens-Serrano et al., 2019](#)). The wording of these questions is set up in such a way that they do not impute any bias in the responses. These questions regard the workload that shoppers generally experience in their current job.
2. While they were picking, shoppers rated their perceived fatigue level according to the RPE (Rating of Perceived Exertion) scale of ([Borg, 1982](#)) approximately every 5 minutes. This is supposed to only shortly disrupt their workflow because longer intermission could influence the results of this study.
3. After they finished their tasks, shoppers filled in the NASA TLX questionnaire ([Hart and Staveland, 1988](#)). These 6 questions consider this task specifically, but again with a focus on the workload instead of the fatigue. Together with the COPSOQ questions, this gives an indication of how the tasks during this experiment compared to normal working conditions. This could give some context for the outcomes of the experiment.

B.2. Key takeaways

- Most shoppers find the work at zone pick physically tiring, but not all.
- The workload at stations varies considerably, even during shifts. This can largely be explained by the slow supply of order totes to stations; this causes system-induced idle time at the station and hinders productivity. This also adds moments of rest to the picking task, albeit uncontrolled and undosed.
- The experienced workload by shoppers differs a lot between stations, as clear from shoppers' remarks and notes during the tests.
- Physical demand is generally found to be high at zone pick.
- In almost all subjects the RPE ratings were higher during or right after periods in which there were more picks done.
- It is difficult to track a single shopper's performance and ratings in the current setting. This has to do with them doing different tasks and switching stations. To the best of our knowledge, the figures in this report represent the work done at the stations that were manned by the corresponding shopper. However, these observations could be incorrect and influence the validity of the data comparison.
- The RPE scores could serve as a good indication of the perceived fatigue in the validation phase of my thesis. The ratings often correspond with the verbal and written notes by the shoppers during testing and were understood quickly. The ratings are, however, sensitive to environmental factors and subjectivity.

Physical measurements (heart rate) could help counter these negative aspects. This will still be subject to discussion in the coming weeks.

B.3. Results and discussion

Shopper 5 had a clear increase in their RPE score throughout the experiment (see Figure B.1). They were also consistent at filling in the form. The shopper had just started their shift at a period of relatively constant workload at their station. At around 9:15, the shopper was assigned to another task (decanting) but kept filling in the ratings. From their COPSOQ answers the shopper indicates finding the work physically demanding, which was also the result of the TLX questions.

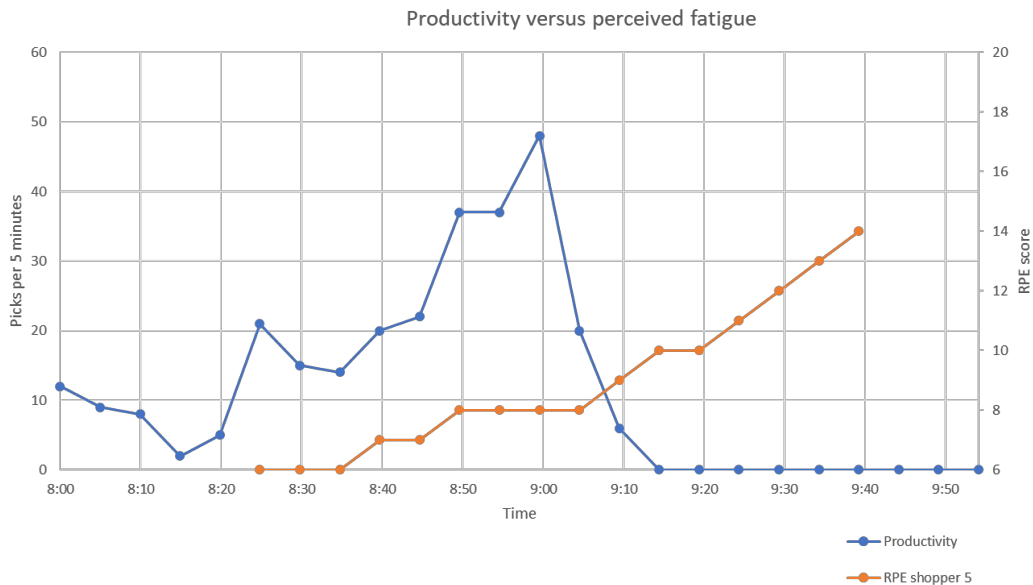


Figure B.1: Productivity and RPE scores as recorded for shopper 5

Shopper 2 had already worked since 6:00 and showed a general increase in their fatigue level up to the highest recorded RPE score (17) of all shoppers in the series of tests (see Figure B.2). They noted that the workload at their stations (blue line) was a bit lower around 14:40 and then took control of a total of 8 stations (yellow line) for the remainder of their shift. This increase in workload concurs with an increase in their RPE score, while the decreased workload at their own stations could correspond to the dip in the perceived fatigue just before that. The shopper stated that zone picking can be physically tiring or exhausting sometimes, more so than GTP picking. Considering today’s task, they stated that the physical demand was not too high, but the pace and effort were very high.

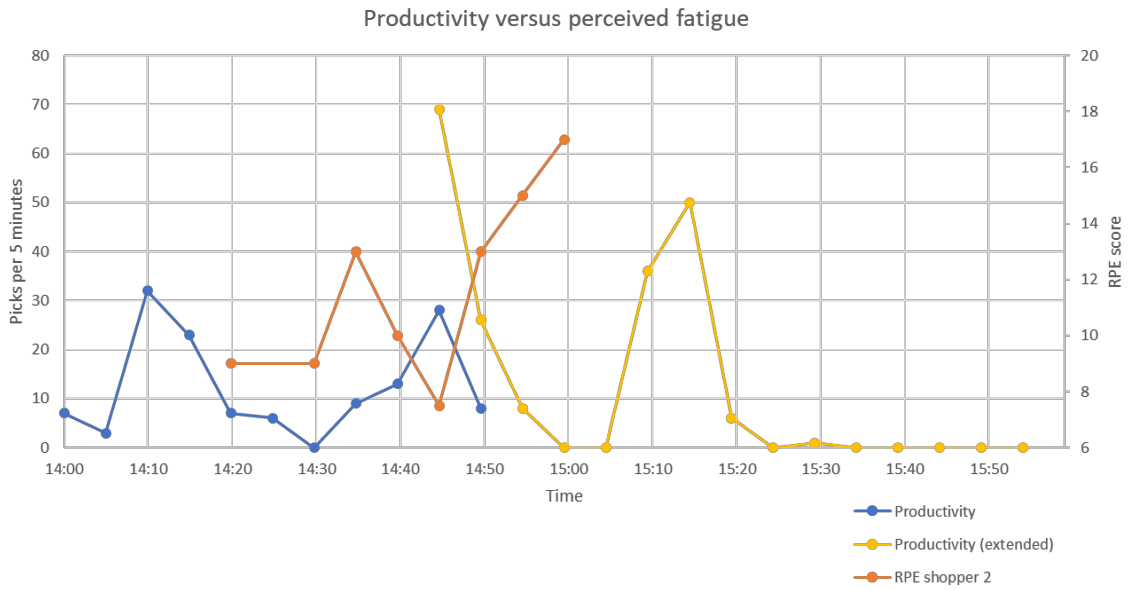


Figure B.2: Productivity and RPE scores as recorded for shopper 2

The graphs from shopper 11 show different behaviour (see Figure B.3). Although their RPE scores are slightly higher during higher demand periods, their ratings were very low nonetheless. This shopper just started picking and had "very little to do" while working a station that mostly required picking single bell peppers - a light item. They found both the work in general and the tasks today to be low in physical demand but were mostly frustrated with the lack of work and boredom. The latter was reflected by their NASA TLX rating and their notes in the RPE scores form. Looking at the picks per 5 minutes, this shopper had very little to do in comparison to the previous two shoppers.

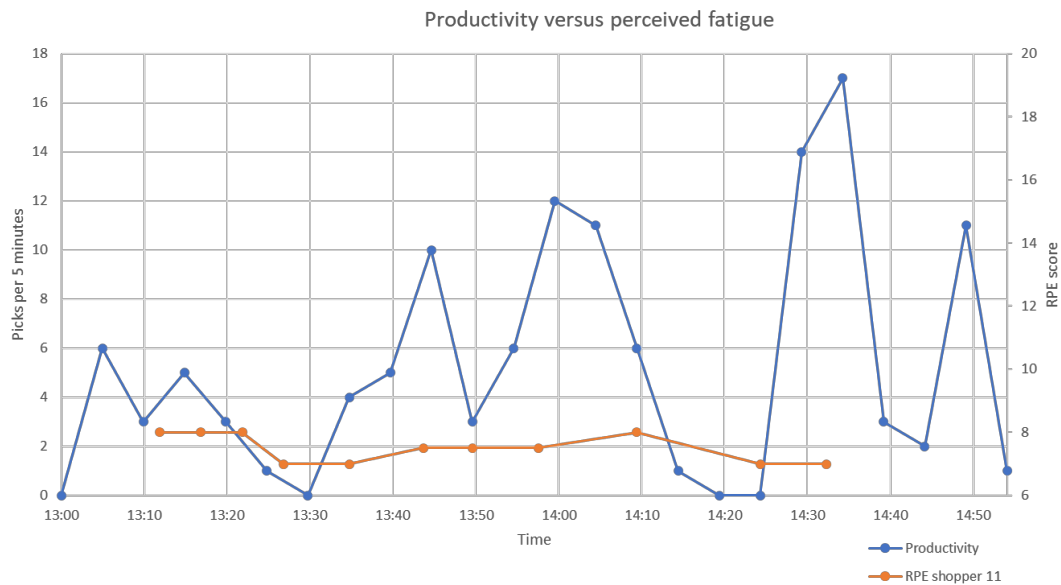


Figure B.3: Productivity and RPE scores as recorded for shopper 11

Less clear results can look like those of shoppers 3 and 4, starting at 8:00 and 6:00, respectively (see Figure B.4). Both filled in the questionnaires at the end of their day and in the middle of their shift, thereby not satisfying the conditions suggested by Aryal et al. (2017). The shoppers were working stations 1, 2 and 3 together so are difficult to distinguish in the data. Shopper 3 indicated this was their first workday and they were very tired. This is reflected in both their COPSOQ and TLX feedback. They show no variation in

their fatigue level but have a constant rating of 14. Contrarily, shopper 4 had 5.5 years of experience and regarded this as a light shift in comparison to the general situation. They stated, "Fatigue really depends on the station/items to pick. Right now I'm only picking cartons of 6 eggs, while other days I'm dragging heavy bags of cat litter.". This could explain their constantly low RPE scores.

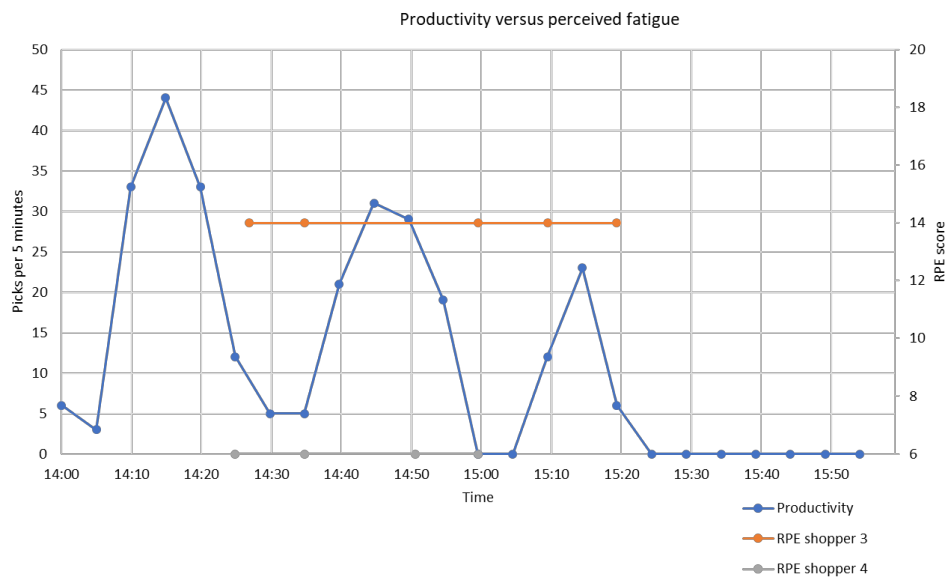
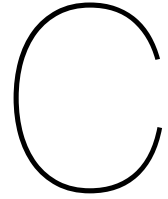


Figure B.4: Productivity and RPE scores as recorded for shoppers 3 and 4



Generation of model inputs

In this section, we will describe the full process of input generation in `Optimize SCIP.py`, function `_init_inputs`. Problem instances are generated using 5 hyperparameters, as shown in an overview in Table C.1. Each problem instance run is given a random seed to use when generating and sampling input values, meaning that multiple different solver runs for the same problem instance will deal with the exact same inputs.

Hyperparameter	Short name	Definition
n_{totes}	I	The total number of totes
$n_{orderlines}$	J	The number of order lines per tote
$n_{stations}$	K	The total number of stations
$n_{shoppers}$	L	The number of shoppers
n_{items}	N	The number of different items to consider

Table C.1: Problem instance hyperparameters and their definition.

C.1. Shopper properties

In order to use the task decomposition method by Garg et al. (1978), we need three different shopper properties: gender, body weight and age. Gender is a binary parameter generated by `random.randint(0,1)`, 0 for female, 1 for male. Body weight is a float and follows different normal distributions for adult men and women according to the population data of Millar (1986). We generate these numbers using `gender*np.random.normal(76.7, 12.1) + (1-gender)*np.random.normal(61.5, 11.1)`. Finally, aligned with the population data for the body weight, age is a random integer between 20 and 69 and is generated using `random.randint(20,69)`.

Another shopper property is the picking speed that dictates the shopper-dependent item processing time $p_{j,l}$. For each task element, we assume non-novice shoppers to vary between 90% and 110% of the normal processing time. Shopper speed is therefore generated using `random.randint(90, 110)`. All four parameters are then saved in the global Python dictionary `shopper_properties`.

C.2. Inventory

For all optimisation runs we want to use the same inventory and slotting. As most of our problem instances are small, we create an inventory file for 6 different stations and 8 slots per station in `Inventory_Setup.py`. Each item in these 48 slots receives a station number, slot walking distance, item weight and a fragility category. For walking distance and item weight, we want values that lead to differences in the energy expenditure of the activity, so we sampled from two extreme groups of real slot walking distances and item weights. The following code snippet generates the dictionary `inventory`, which is used to sample items for each unique problem instance.

```
# Predefined slot walking distances (4 slots within 2m, 4 slots above 6m)
walking_distances = [1.65, 0.70, 1.01, 1.96, 6.36, 7.41, 8.50, 9.62]

# Item weight categories, 0=low, 1=high
weight_cats = [0,0,0,0,1,1,1,1]
```

```

inventory = {'Station': [], 'Distance': [], 'Weight': [], 'Fragility': []}
# Total number of stations to consider is max 6
for station in range(6):
    # Each station has 4 light and 4 heavy items
    weight_cat=random.sample(weight_cats, 8)
    # Total number of slots per station is 8
    for slot in range(8):
        inventory['Station'].append(station)
        inventory['Distance'].append(walking_distances[slot])
        # Low category 100-1000 grams
        if weight_cat[slot] == 0:
            inventory['Weight'].append(random.random()*0.9+0.1) # [kg]
        # High category 4000-7000 grams
        else:
            inventory['Weight'].append(random.random()*3+4) # [kg]
        # Add fragility category between 1 and 5
        inventory['Fragility'].append(random.randint(1, 5))

```

After importing the inventory list we create a subset of the inventory based on K , so that we do not sample order lines from more stations than we consider in our problem.

C.3. Item properties

From the inventory subset we sample, without replacement, N unique items. If the number of stations times the 8 slots per station is lower than N , we sample $8K$ items and then sample, with replacement, N items from that list. We then generate dictionary `item_properties` with our sample and the items' corresponding properties from the inventory, using the following code snippet:

```

# Item properties
self.item_properties = {'Station': [], 'Distance': [], 'Weight': [], 'Fragility': []}
# Sample a random, non-replacing set of n_items from the inventory
for item in random.sample(inventory['Index'].tolist(), min(n_items, len(inventory))):
    # Returns the value in column 'station' and row 'item' in the inventory dataframe
    self.item_properties['Station'].append(inventory.loc[item, 'Station'])
    self.item_properties['Distance'].append(inventory.loc[item, 'Distance'])
    self.item_properties['Weight'].append(inventory.loc[item, 'Weight'])
    self.item_properties['Fragility'].append(inventory.loc[item, 'Fragility'])

```

C.4. Sets and parameters

The sets and parameters of the mathematical model in Section 3.2 are now created from the above inputs. Sets I , K and L are simply lists of I , K and L indices, respectively. Set J is made by sampling, with replacement, an index from our `item_properties` dictionary for all J order lines per I totes. Any value $J_{i,j}$ will therefore be an integer from `range(min(n_items, len(inventory)))`. Finally, set $K_{i,j}$ is a single value in our case, as each item has only one location in the inventory. If we were to add item multiplicity in the system, this requires changing the setup of `Inventory_Setup.py` and giving items a tuple or list of integers in the `Station` column.

Parameter $f_{i,j}$ is simply taking the `Fragility` column of `item_properties`, which can later be accessed by slicing element $J_{i,j}$ from it. Parameter $s_{k,k'}$ is estimated by assuming all stations are equidistant and station switching between two adjacent stations takes 7 seconds. Parameters p , r , and EE are determined for each separate order line $J_{i,j}$ in a separate task decomposition function that takes arguments `gender`, `body_weight`, `age`, `speed`, `walking_dist` and `item_weight`; all extracted from `shopper_properties` and `item_properties` for all combinations of $J_{i,j} \in J$ and $l \in L$.

C.4.1. Task decomposition

In function `find_pick_info` from `Helper_Functions.py` we perform the elemental task decomposition by Garg et al. (1978) that was introduced in Section 2.1.2. The function is used in both the input generation for the mathematical model in Section 3.2 and the calculation of the energy expenditure and pick time for individual item picks in Section 4. We identify 5 different phases of each item pick and the corresponding

formulae from [Garg et al. \(1978\)](#). In [Table C.2](#), an overview of the used symbols and their definition for these formulae.

Symbol	Unit	Definition
ΔEE	kcal	Energy expenditure for activity
\dot{EE}	kcal/min	Energy expenditure rate for activity
BW	kg	Body weight of shopper
L	kg	Weight of the load
S	-	Gender; 0 for females, 1 for males
h_1	m	Height from floor; starting point lift, endpoint lower
h_2	m	Height from floor; endpoint lift, starting point lower
X	m	Horizontal movement of workpiece
V	m/s	Walking speed of shopper

Table C.2: Symbols used in the formulae by [Garg et al. \(1978\)](#) and their definition.

The formulae from [Garg et al. \(1978\)](#) have been adjusted to fit our specific situation. This entails making some assumptions about the values of the above symbols. First of all, we assume BW , L and S to be given from the item and shopper properties for both the mathematical model and the experiments. The load's weight depends on the item weight and the number of items to be picked per order line. In the mathematical model, all order lines contain only a single item, but in the experiments they may contain several counts of the same item. We assume this only to affect the weight of the load - which is then the sum of all item weights in the order line - and the picking task is the same otherwise. The value of h_1 and h_2 is different for each movement where they play a role. In general, we assume the arm resting position to be at 1.2m, the pick height at 0.5m, the item scan and pick confirm height at 1.6m and the placing height at 1.2m from the floor. Furthermore, the arm extension during picking or placing, or X , is assumed to be 0.5m and the walking speed V is 1m/s. Both the walking speed and all pick time estimates are influenced by the picking speed of the shopper ($X = 1m/s \times Speed_{pick}$). The corresponding sections below give the time estimates of each different phase.

Finally, we assume the floor is level, there is no pull or pushing force required during picking or placing, and the picking height is constant. Station switching energy and energy spent waiting between picks are only considered in the experiments because the mathematical model distributes station switching and waiting time equally between shoppers. Also, station switching is much more common in the experiments than in the simulation.

Read item info

The start of a pick requires shoppers to stand at the station and read the item info on the screen. On the screen, the shopper is informed of the item to pick, the slot letter code, the number of items to pick and the compartment of the tote in which the item must be placed. We estimate this action to take 0.5 seconds. During this time, the standing energy expenditure rate is required:

$$\begin{aligned} &\text{Standing (kcal/min):} \\ &\dot{EE} = 0.024BW \end{aligned} \tag{C.1}$$

Walk to slot

Once informed of the item location and the number of items to pick, the shopper walks to the assigned slot. The time this takes depends on the walking speed V and the distance, which is item-dependent (see [Figure C.1](#)).

$$\begin{aligned} &\text{Walking without load (kcal/min):} \\ &\dot{EE} = 10^{-2} (51 + 2.54BW \cdot V^2) \end{aligned} \tag{C.2}$$



Figure C.1: Slots at varying distances from the station.

Pick item

The item pick consists of seven elements. First, the shopper lowers their arms from their resting position to the lowest possible position. Then, they stoop to the level of the pick. While the pick happens, they stand in a bent position for about 1 second. Then, the actual pick happens by extending and then retracting the arms (see Figure C.2) Finally, they stoop back up and put their arms in the resting position again. During these last three steps, there is an added load, as the item(s) are now in their hands. The total pick time is estimated to take 3 seconds.

Arm lower (kcal/lower):

$$\Delta EE = 10^{-2} [0.093BW (h_2 - 0.81) + (1.02L + 0.37S \cdot L) (h_2 - h_1)] \text{ for } 0.81 < h_1 < h_2$$

Stoop lower (kcal/lower):

$$\Delta EE = 10^{-2} [0.268BW (0.81 - h_1) + 5.22S (0.81 - h_1)] \text{ for } h_1 < h_2 < 0.81$$

Standing in bent position while picking (kcal/min):

$$\dot{EE} = 0.028BW$$

Extending both arms (kcal/movement):

$$\Delta 0.0357X$$

Retracting both arms with load (kcal/movement):

$$\Delta EE = 10^{-2} X (3.57 + 1.23L)$$

Stoop lift with load (kcal/lift):

$$\Delta EE = 10^{-2} [0.325BW (0.81 - h_1) + (1.411 + 0.76S \cdot L) (h_2 - h_1)] \text{ for } h_1 < h_2 \leq 0.81$$

Arm lift with load (kcal/lift):

$$\Delta EE = 10^{-2} [0.062BW (h_2 - 0.81) + (3.19L - 0.52S \cdot L) (h_2 - h_1)] \text{ for } 0.81 < h_1 < h_2$$

(C.3)



Figure C.2: Shopper picking a cucumber.

Walk to station

The shopper walks back to the station with the items in their hands. Again, the time taken depends on the walking speed V and the distance.

Carrying, loads held against thighs or against waist (kcal/min):

$$\dot{E}E = 10^{-2} [68 + 2.54BW \cdot V^2 + 4.08L \cdot V^2 + 11.4L + 0.379(L + BW)G \cdot V] \quad (\text{C.4})$$

Place item

The item placement in the tote consists of seven steps. First, the shopper raises their arms to the scanning height, after which they extend their arms forward to scan the item (see Figure C.3). While the scanning takes place, the shopper stands still at the station for about 1 second. After scanning, shoppers lower their arms to place the item(s) (see Figure C.4). Then, they raise their arms to confirm the pick on the screen, after which they retract and lower their arms back to the resting position. The total place time is estimated to be 3 seconds.



Figure C.3: Scanning items before placing them in the tote.

Arm lift with load (kcal/lift):

$$\Delta EE = 10^{-2} [0.062BW (h_2 - 0.81) + (3.19L - 0.52S \cdot L) (h_2 - h_1)] \text{ for } 0.81 < h_1 < h_2$$

Extending both arms with load (kcal/movement):

$$\Delta EE = 10^{-2} X (3.57 + 1.23L)$$

Standing while placing (kcal/min):

$$\dot{EE} = 0.024BW$$

Arm lower with load (kcal/lower):

$$\Delta EE = 10^{-2} [0.093BW (h_2 - 0.81) + (1.02L + 0.37S \cdot L) (h_2 - h_1)] \text{ for } 0.81 < h_1 < h_2$$

Arm lift without load (kcal/lift):

$$\Delta EE = 10^{-2} [0.062BW (h_2 - 0.81)]$$

Retracting both arms (kcal/movement):

$$\Delta EE = 0.0357X$$

Arm lower (kcal/lower):

$$\Delta EE = 10^{-2} [0.093BW (h_2 - 0.81)] \text{ for } 0.81 < h_1 < h_2$$

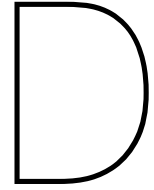
(C.5)



Figure C.4: A shopper placing an item in a tote.

C.4.2. From task decomposition to pick info

After calculating the time taken and energy expenditure of each picking phase, we can sum them for a single pick. This gives a total energy expenditure and time taken for the pick, which gives us the average energy expenditure rate \overline{EE} during the pick. We calculate the required recovery time $r_{i,j,l}$ using (2.6), (2.9) and (2.8). For the experiments, this process is a little different, as explained in Section 4.



Additional results from the mathematical optimisation

For a problem instance with four totes, two order lines per tote, two stations and two shoppers, we get 826 variables and 1162 constraints. The 27 optimisation runs result in 6 unique objective 'value pairs' (C_{max} , EE_{max}) (see Table D.1). All individual optimisation runs returned their global optimum, with run times up to 813 seconds. Four points (including C_0) are confirmed Pareto-optimal because they represent globally optimal solutions that are non-dominated. The corresponding Pareto front estimate is presented in Figure D.1.

Run	C_{max} [s]	EE_{max} [kcal]	Pareto-optimal
C_0	187.95	117.83	✓
EE_0	318.21	63.40	✗
$x = 0.00$	327.58	63.40	✗
$x = 0.04$	215.79	63.40	✓
$x = 0.46$	208.83	65.12	✓
$x = 0.88$	206.28	70.28	✓

Table D.1: Unique solution value pairs.

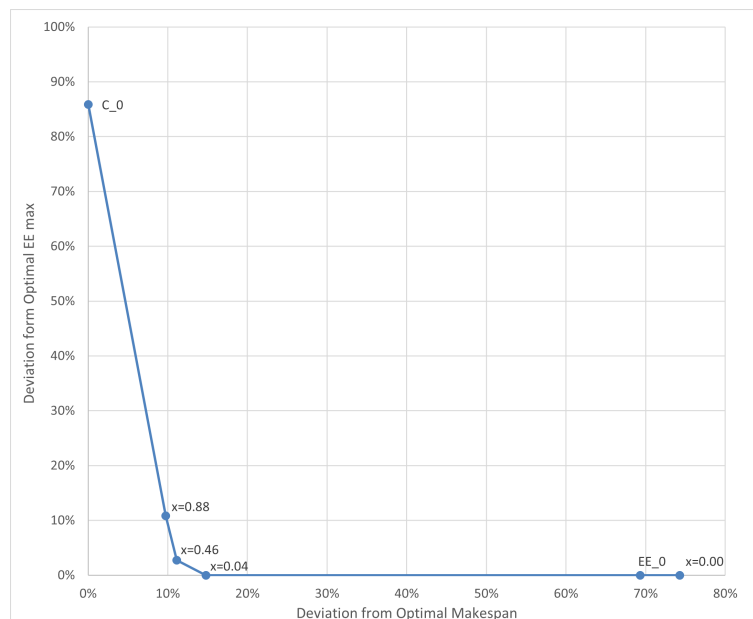


Figure D.1: Estimated Pareto front with only globally optimal solutions. Each dot may represent multiple solutions but only one objective value pair.

Table D.2 and Figure D.2 show the results for a problem instance with 3 totes, 3 order lines per tote, 2 stations and 2 shoppers that has 1046 variables and 1467 constraints. The maximum run time is 111 seconds and we see 5 unique objective value pairs, of which 3 Pareto-optimal points. Unlike previous problem instances, run ' C_0 ' does not represent a Pareto-optimal solution here, meaning that there may be a solution possible that dominates it. The steep line piece between ' C_0 ' and ' $x = 0.71$ ' is not certain to represent a trade-off decision, although this would be the steepest of all results.

Because the C_{max} value corresponding to run ' EE_0 ' is very high and this run is dominated by ' $x = 0.04$ ', we do not show it in the figure for clarity.

Run	C_{max} [s]	EE_{max} [kcal]	Pareto-optimal
C_0	62.99	35.70	✗
EE_0	298.92	25.72	✗
0.04	67.47	25.72	✓
0.13	63.35	25.92	✓
0.71	63.08	26.16	✓

Table D.2: Unique solution value pairs.

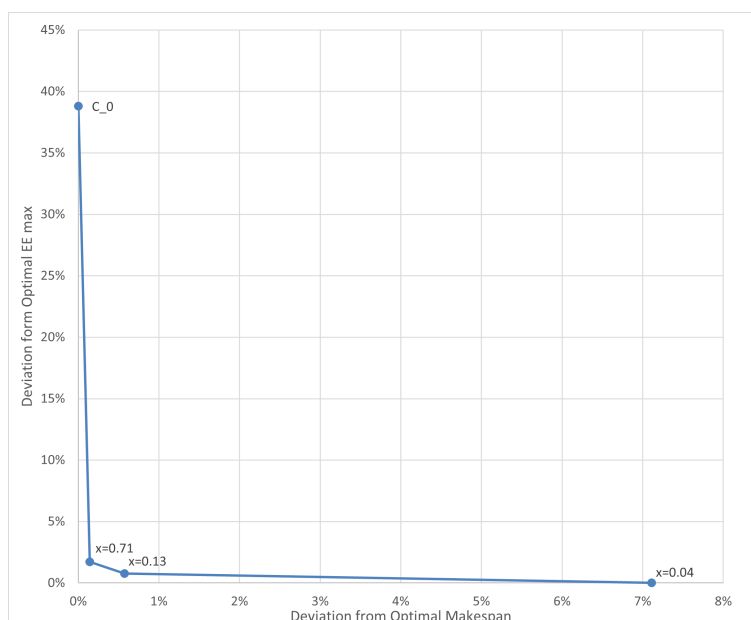


Figure D.2: Estimated Pareto front with only globally optimal solutions. Each dot may represent multiple solutions but only one objective value pair.

The final problem instance has 3 totes, 2 order lines per tote, 3 stations and 2 shoppers and is presented in Table D.3 and Figure D.3. With 884 decision variables and 1114 constraints, we get only globally optimal solutions, very short run times of under 2 seconds, 4 unique objective value pairs and 2 non-dominated solutions. Although the Pareto front estimate looks steep, the scale of the figure is very askew. Also, run ' C_0 ' looks to be dominated by ' $x = 0.04$ ' but we can see in the table that this is not the case, even though it is close.

Run	C_{max} [s]	EE_{max} [kcal]	Pareto-optimal
C_0	40.60	15.01	✓
EE_0	65.26	14.97	✗
0.00	237.14	14.97	✗
0.04	40.65	14.99	✓

Table D.3: Unique solution value pairs.

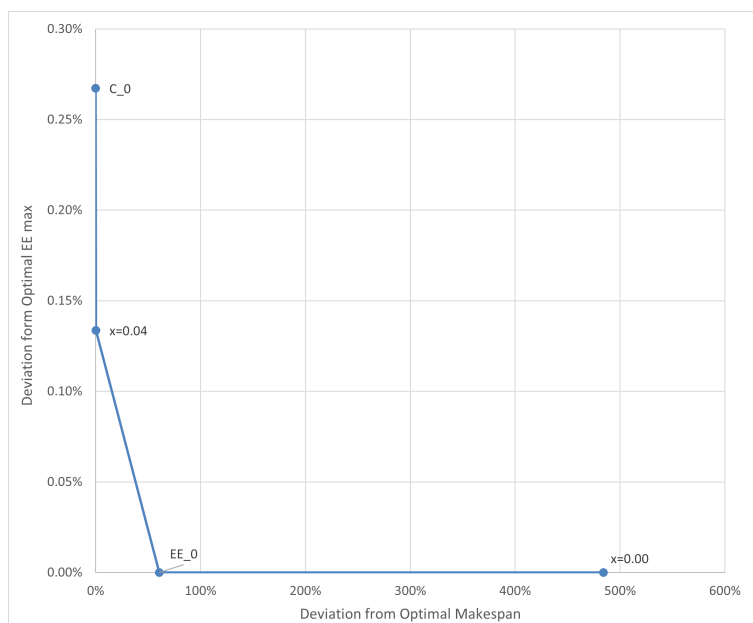
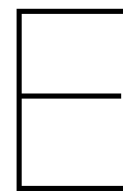


Figure D.3: Estimated Pareto front with only globally optimal solutions. Each dot may represent multiple solutions but only one objective value pair.

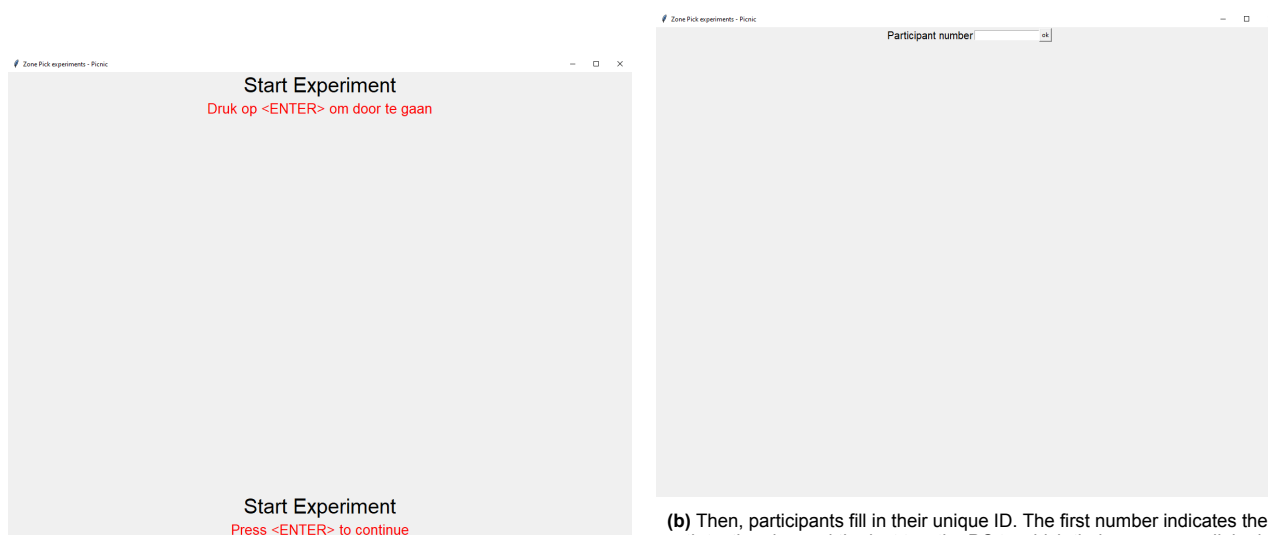
$-\%EE_{max}$	$+\%C_{max}$	Steepness
0.00	0.00	0.00
0.00	0.00	0.00
1.16	75.86	0.02
2.87	51.23	0.06
1.03	9.30	0.11
1.35	11.18	0.12
2.72	13.54	0.20
4.45	7.31	0.61
4.37	5.50	0.79
0.14	0.12	1.17
0.07	0.06	1.17
8.23	6.43	1.28
4.93	3.20	1.54
1.55	0.67	2.31
16.16	3.40	4.75
10.97	2.29	4.79
5.64	1.16	4.86
17.45	3.04	5.74
75.00	9.75	7.69
30.33	1.68	18.05
23.79	1.22	19.50
44.86	1.05	42.72
39.11	0.67	58.37
41.53	0.70	59.33
30.09	0.50	60.18
30.19	0.44	68.61
35.91	0.32	112.22
37.09	0.14	264.93
40.70	0.11	370.00
48.91	0.07	698.71

Table D.4: Optional trade-off decisions that could be made by choosing the first non-dominated solution over ' C_0 ', for all problem instances that returned strictly global optima from optimisation. The first column indicates the fractional improvement in energy expenditure distribution, the second column indicates the corresponding increase in makespan. The third column indicates the steepness of the line piece between the two considered solutions, thereby quantifying the trade-off. The first two rows did not return alternative solutions from optimisation.



Graphical user interface for experiments

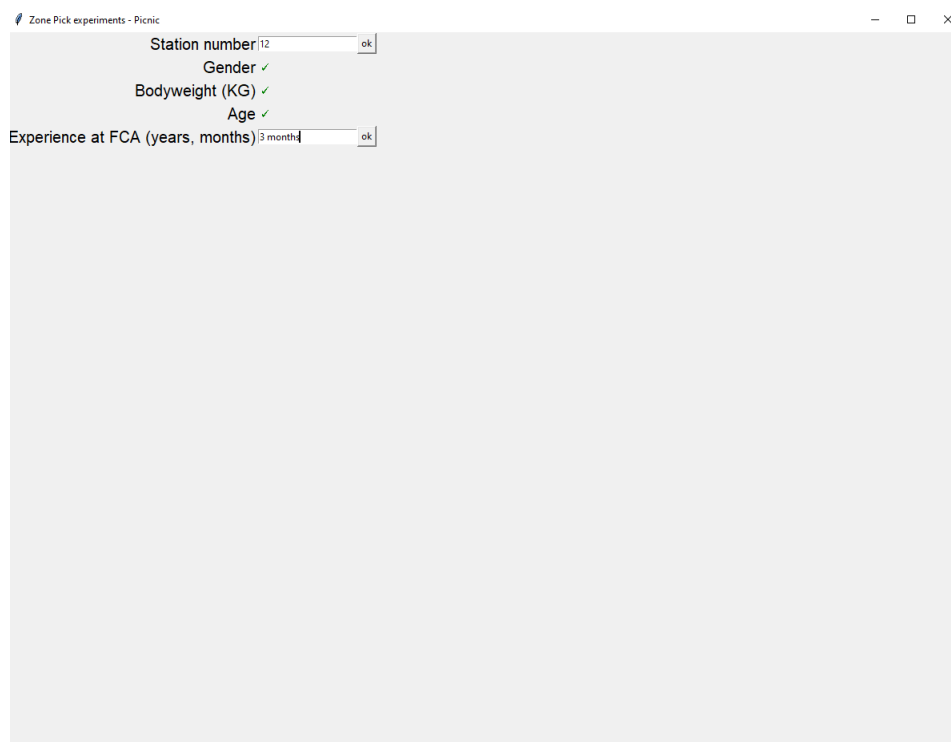
In this section, we will go over the workflow in the GUI during the experiments. Each step is shown and explained in the figures. Note that these steps are in a different order for participant group B.



(a) Startup up screen of the GUI. After hitting `Enter` participants can choose between the Dutch and English version of the GUI.

(b) Then, participants fill in their unique ID. The first number indicates the x -th testing day and the last two the PC to which their sensor was linked (01-04). The link between their name and participant number is recorded nowhere.

Figure E.1: Standard GUI startup.



Zone Pick experiments - Picnic

Station number

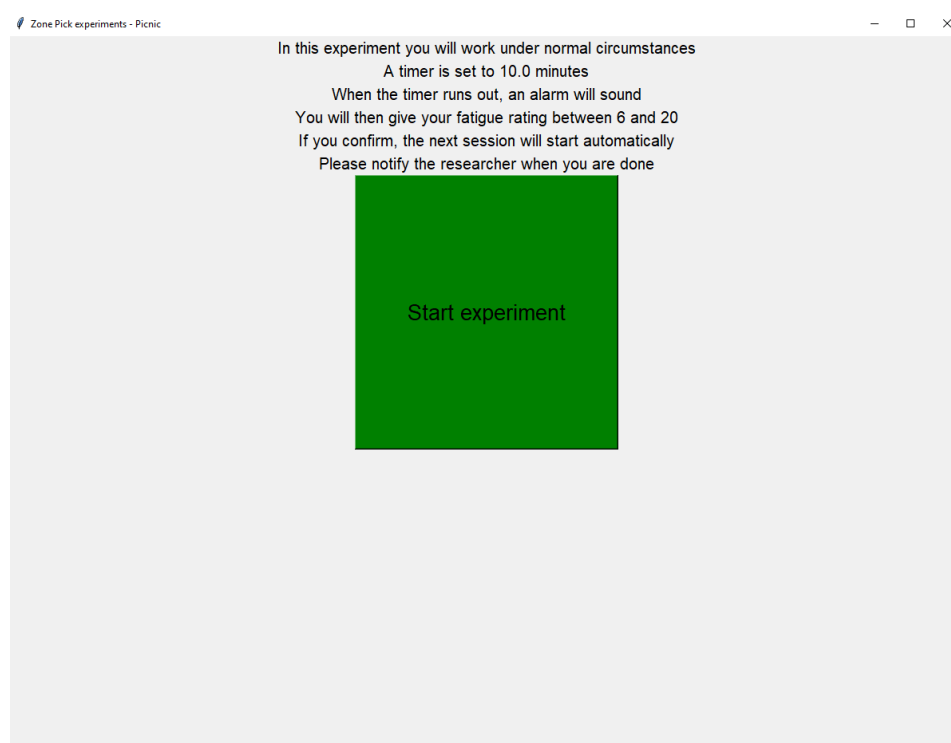
Gender

Bodyweight (KG)

Age

Experience at FCA (years, months)

Figure E.2: This is where participants fill in their personal information. This is required for the calculation of RA during the experiments and for analysis after the experiments. The experience is not crucial for the experiments but could serve as a sidenote during analysis. All info must be confirmed and is type checked by the program, after which it is stored in a .csv file linked to the participant ID.



Zone Pick experiments - Picnic

In this experiment you will work under normal circumstances
A timer is set to 10.0 minutes
When the timer runs out, an alarm will sound
You will then give your fatigue rating between 6 and 20
If you confirm, the next session will start automatically
Please notify the researcher when you are done

Figure E.3: Before the start of the first experiment scenario, shoppers are informed of the key elements of the experiment again. Once they hit 'start experiment', the experiment logfile is created.

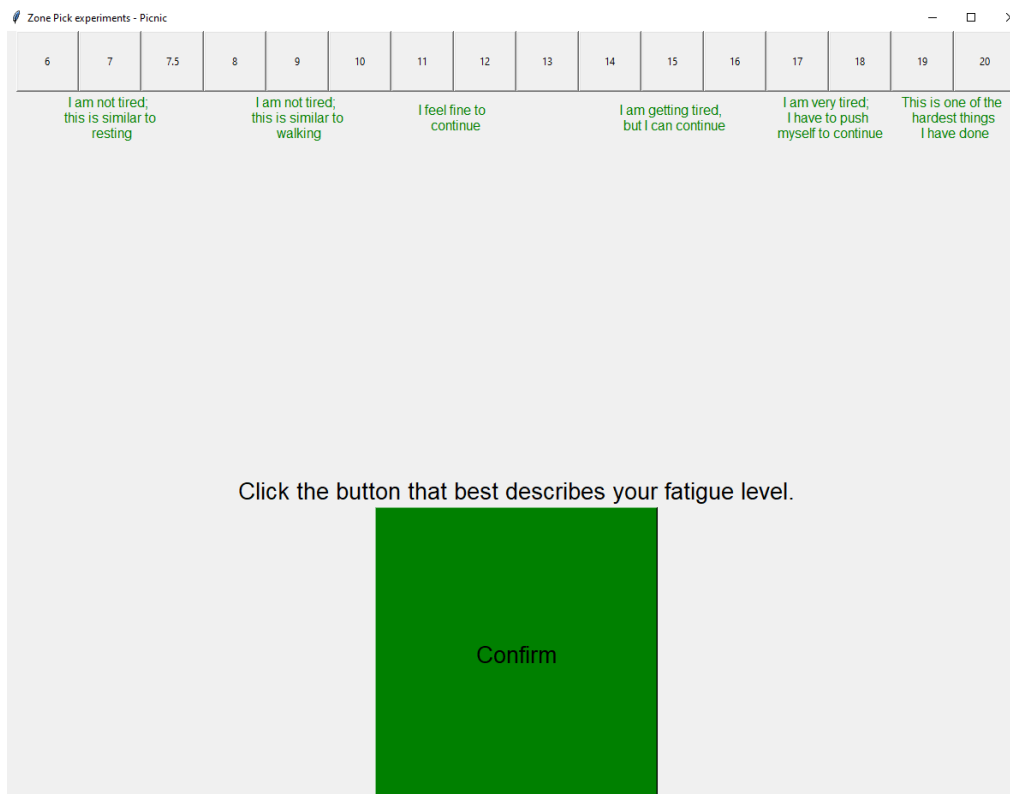
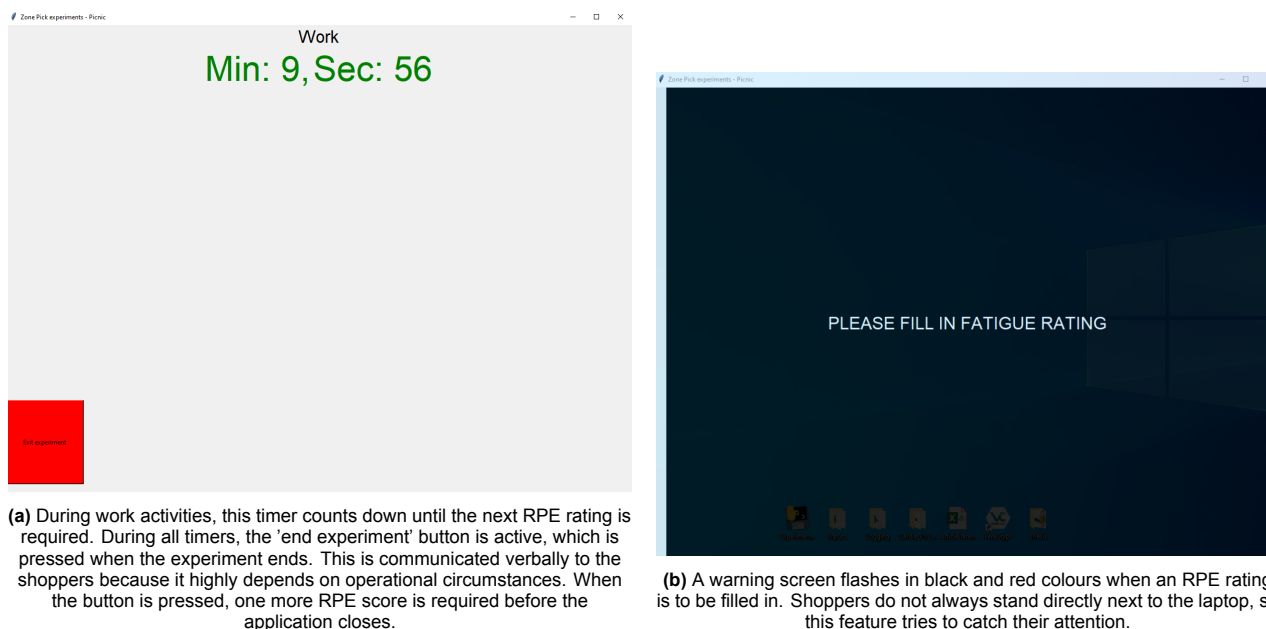


Figure E.4: At the start of the scenario, after every 10 minutes and at the end of the scenario this window opens. The RPE scale is taken from Borg (1982) and buttons are linked to their values. The descriptive elements are shown to provide reference to shoppers and are copied from the original work by Borg et al. In theory, the ratings between 6-20 should relate to heart rates of 60-200 BPM, but this is not a proven correspondence.



(a) During work activities, this timer counts down until the next RPE rating is required. During all timers, the 'end experiment' button is active, which is pressed when the experiment ends. This is communicated verbally to the shoppers because it highly depends on operational circumstances. When the button is pressed, one more RPE score is required before the application closes.

(b) A warning screen flashes in black and red colours when an RPE rating is to be filled in. Shoppers do not always stand directly next to the laptop, so this feature tries to catch their attention.

Figure E.5: Start and end of working time.

After the RPE window is shown again, the above steps repeat (timer, warning, RPE) until the scenario ends. Upon starting the GUI for the second scenario, the shopper sees the first three windows (startup, language selection and participant number). If the same participant number is entered as before, the system will recognise

the ID in the logs and skip the personal info window.

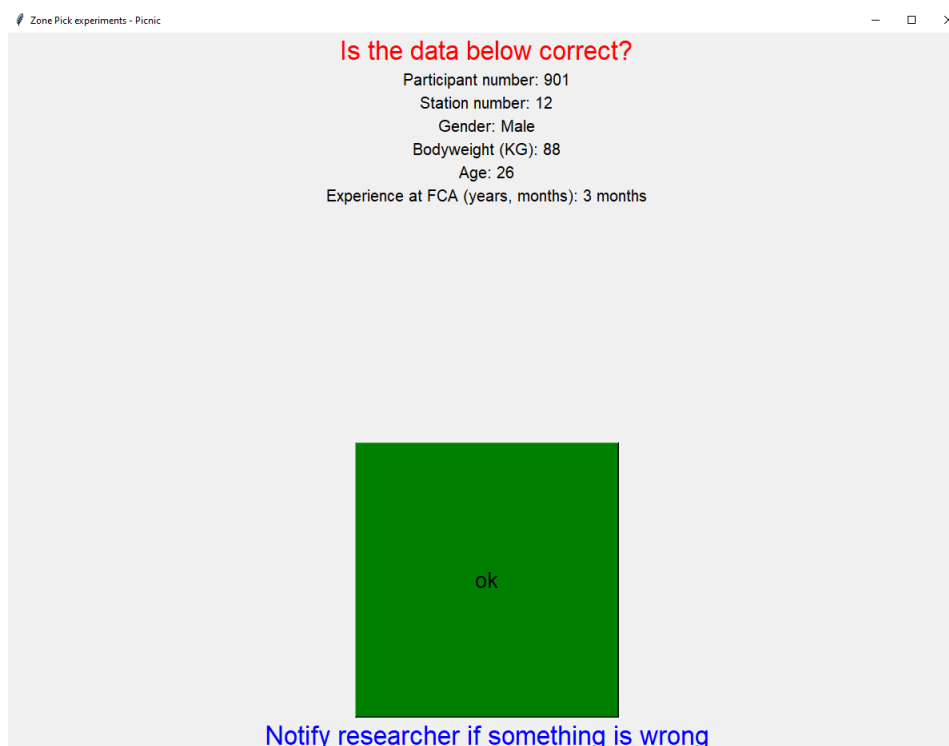


Figure E.6: After the start of the second scenario and matching the participant number from before, this window asks shoppers to confirm that the previously filled-in info has not changed. This is mostly important to prevent a mix-up or unwanted station switching.

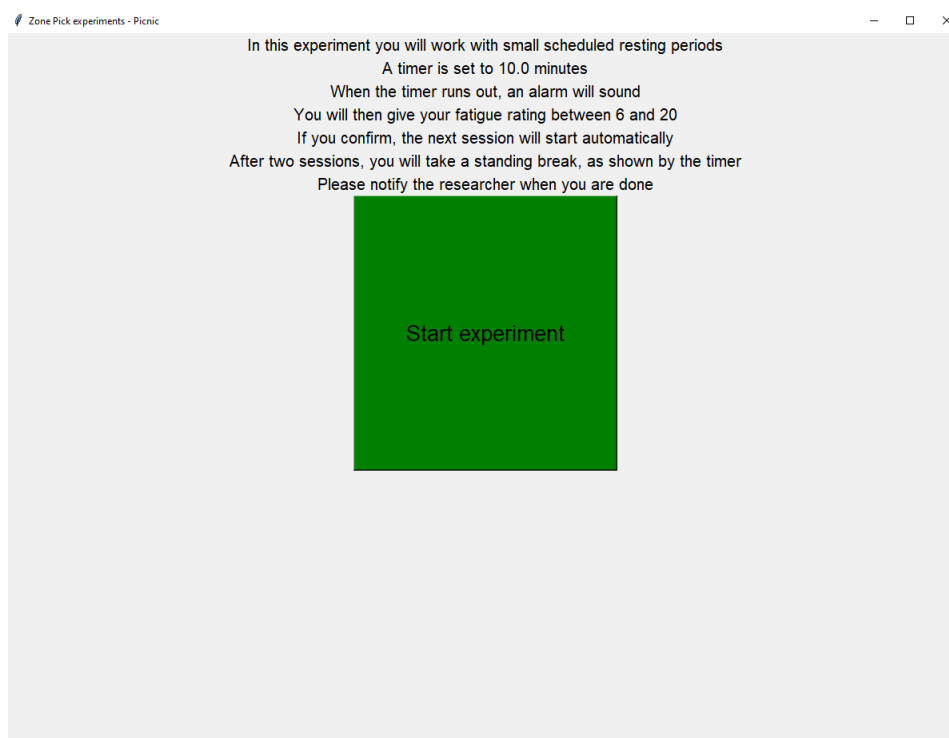


Figure E.7: The second scenario starts with an instruction too. Now, additional breaks will be triggered after each second RPE window. When the 'start experiment' button is pressed, the GUI will calculate the appropriate RA using the order data from WMS, the personal info filled in before and the station number.

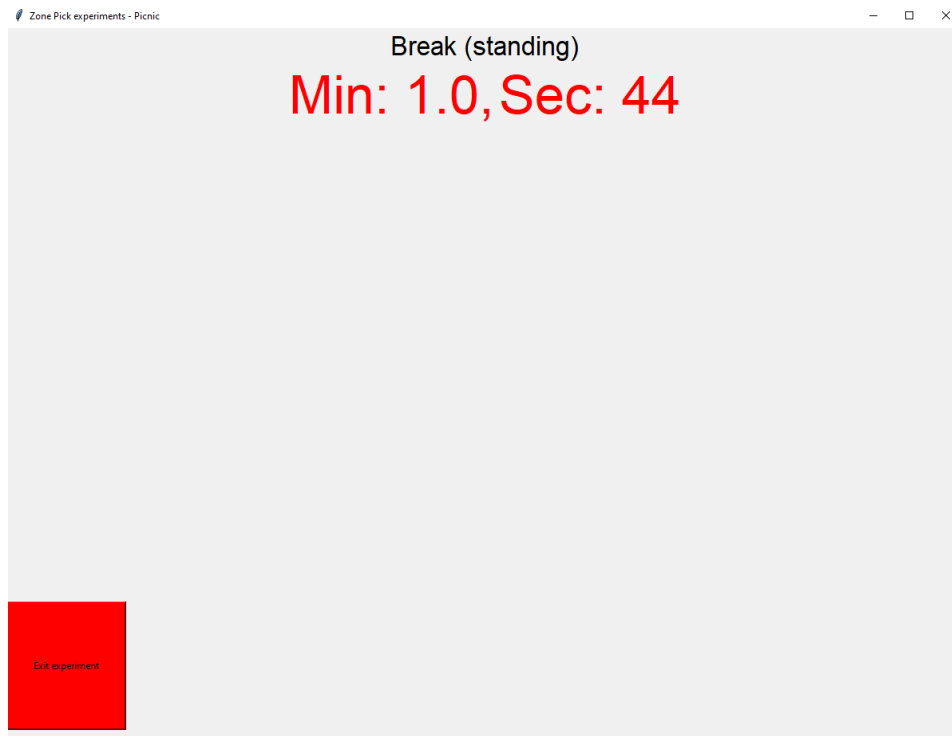


Figure E.8: When a break is triggered, this screen is shown. The timer is now red and the text has also changed, indicating that the timer has a different meaning than previously. The normal break timer starts again after this one runs out.

Click the button that best describes your opinion

Mental demand: How mentally demanding was the task?
Very Low Very High

Physical demand: How physically demanding was the task?
Very Low Very High

Temporal demand: How hurried or rushed was the pace of the task?
Very Low Very High

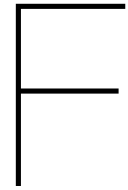
Performance: How successful were you in accomplishing what you were asked to do?
Perfect Failure

Effort: How hard did you have to work to accomplish your level of performance?
Very Low Very High

Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?
Very Low Very High

Confirm

Figure E.9: Finally, when both scenarios have started and the 'end experiment' button is pressed in the second scenario, shoppers are asked to fill in the standard NASA TLX form. Once all questions have been answered and confirmed, the GUI closes and the experiment is done.



Additional experimental results

Participant	Correlation	p-value	Correlation +3min	Correlation +4min	Correlation +5min
1	0.20	0.000	0.28	0.21	0.26
2	0.24	0.000	0.19	0.18	0.11
3	0.13	0.000	0.01	0.02	-0.03
4	0.10	0.000	0.09	0.04	0.06
5	0.11	0.000	-0.12	-0.09	-0.14
6	0.21	0.000	0.10	0.12	0.15
7	0.05	0.000	0.15	0.10	0.14
8	-0.02	0.988	0.02	0.01	-0.04
9	0.15	0.000	0.11	0.14	0.12
10	0.23	0.000	0.17	0.22	0.20
11	0.47	0.000	0.22	0.17	0.24
12	0.01	0.079	-0.01	-0.01	0.04
13	0.32	0.000	0.18	0.16	0.06
14	0.15	0.000	0.01	-0.02	-0.07
15	0.38	0.000	0.30	0.31	0.27
16	0.28	0.000	0.02	-0.06	-0.09
17	0.00	0.612	-0.03	0.01	0.03
18	-0.05	1.000	-0.11	-0.14	-0.11
19	0.29	0.000	0.10	0.17	0.20
20	0.33	0.000	0.13	0.23	0.19

Table F.1: Correlation between the $\dot{E}E_{HR}$ and $\dot{E}E_{picks}$ values during the experiments for the 20 participants. Column 2 shows the Pearson correlation coefficient and column 3 the corresponding one-sided p-value indicating its significance. Columns 4 to 6 show the correlation coefficients for the $\dot{E}E_{HR}$ respectively 3, 4 and 5 minutes ahead, following [Achten and Jeukendrup \(2003\)](#). The coloured cells indicate the highest value for the correlation coefficient, which shows no clear benefit of the time displacement.

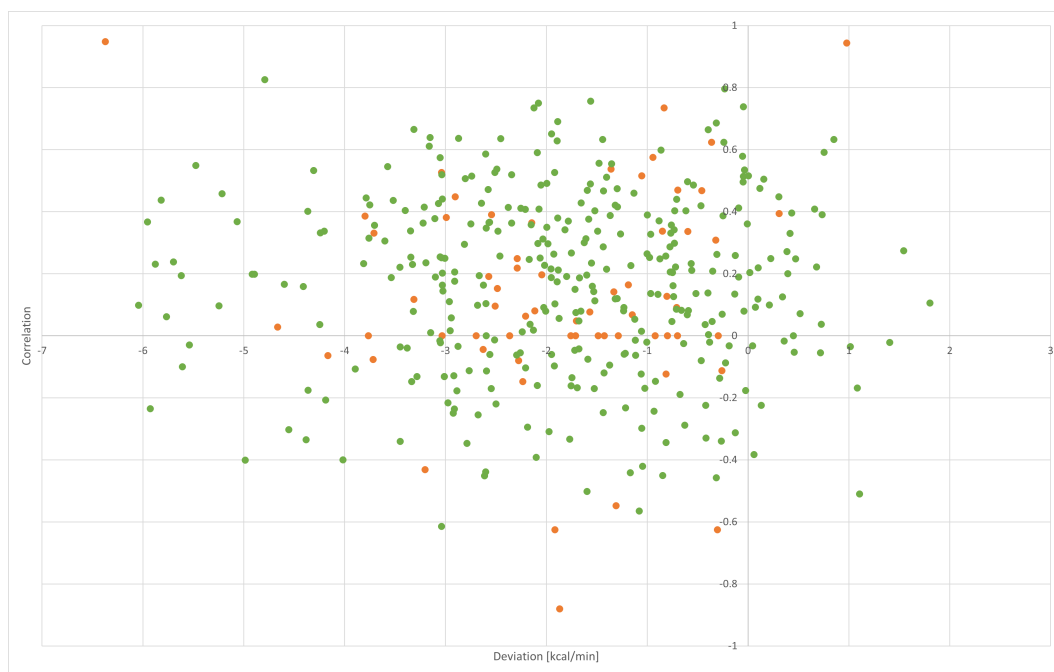


Figure F.1: Overview of all 10 minute intervals for all participants. Each dot represents one measurement interval, with the average $\dot{E}E_{HR} - \dot{E}E_{picks}$ deviation on the x-axis and the $\dot{E}E_{HR} - \dot{E}E_{picks}$ correlation coefficient on the y-axis. The orange dots represent intervals with less than 10 picks per 10 minutes, the green dots all other intervals.

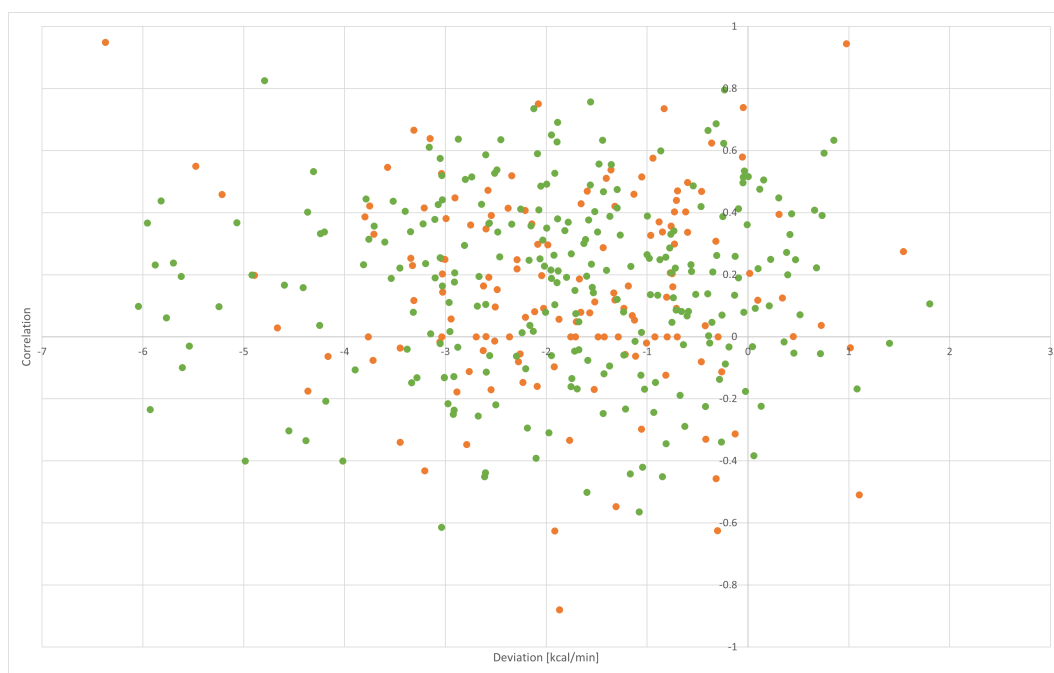


Figure F.2: Overview of all 10 minute intervals for all participants. Each dot represents one measurement interval, with the average $\dot{E}E_{HR} - \dot{E}E_{picks}$ deviation on the x-axis and the $\dot{E}E_{HR} - \dot{E}E_{picks}$ correlation coefficient on the y-axis. The orange dots represent intervals with less than 20 picks per 10 minutes, the green dots all other intervals.

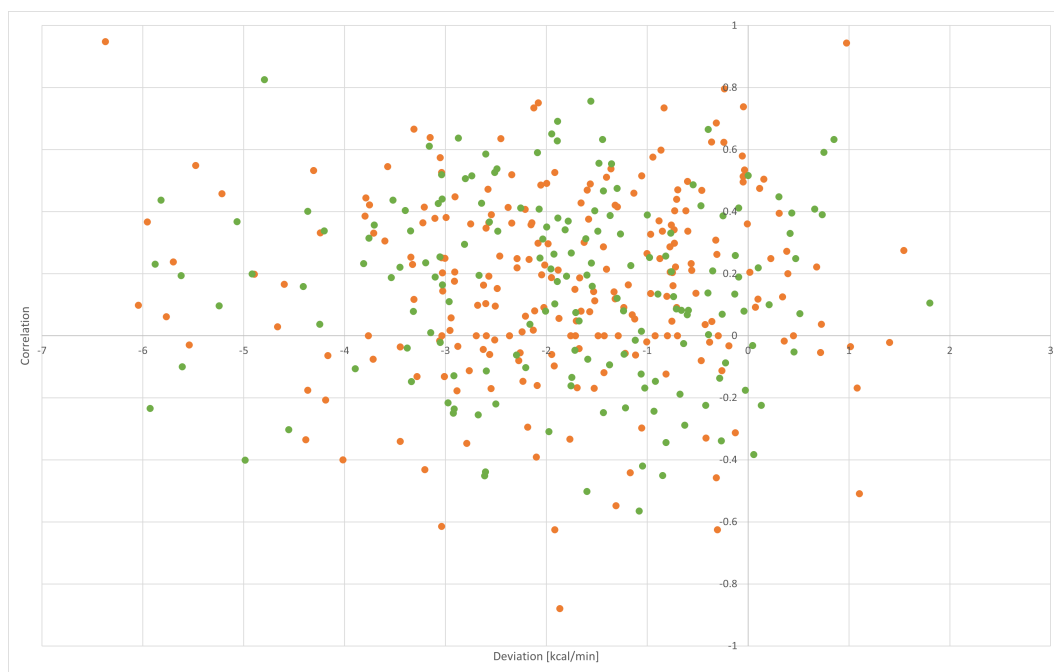


Figure F.3: Overview of all 10 minute intervals for all participants. Each dot represents one measurement interval, with the average $\overline{EE}_{HR} - \overline{EE}_{picks}$ deviation on the x-axis and the $\overline{EE}_{HR} - \overline{EE}_{picks}$ correlation coefficient on the y-axis. The orange dots represent intervals with less than 30 picks per 10 minutes, the green dots all other intervals.

Participant	RA - orders	RA - picks	RA - HR
1	72.48%	0.00%	56.81%
2	23.23%	0.00%	124.69%
3	28.55%	0.00%	42.35%
4	15.35%	0.00%	0.00%
5	0.00%	0.00%	0.00%
6	17.60%	0.00%	22.26%
7	11.80%	0.00%	59.35%
8	3.93%	0.00%	8.02%
9	0.10%	0.00%	0.00%
10	0.00%	0.00%	0.00%
11	4.38%	0.00%	11.13%
12	3.60%	0.00%	85.45%
13	10.75%	0.00%	101.01%
14	0.00%	0.00%	13.79%
15	0.00%	0.00%	0.00%
16	0.00%	0.00%	0.00%
17	0.00%	0.00%	0.00%
18	0.00%	0.00%	0.00%
19	0.00%	0.00%	0.00%
20	0.00%	0.00%	0.00%

Table F.2: Overview of the predicted rest allowance requirement to meet the MAEE levels. The percentages in the table indicate the required break length as a percentage of the working time. A task of 10 minutes with an RA of 20% will take 12 minutes including the RA to meet the MAEE requirement.

Participant	Average RPE	NASA TLX Score	RA - HR
1	6.43	0	56.81%
2	10.28	22	124.69%
3	8.84	20	42.35%
4	6.14	3	0.00%
5	7.98	4	0.00%
6	9.24	28	22.26%
7	12.71	15	59.35%
8	6.17	6	8.02%
9	8.80	20	0.00%
10	7.24	7	0.00%
11	6.88	3	11.13%
12	9.14	25	85.45%
13	10.64	4	101.01%
14	9.31	7	13.79%
15	7.00	43	0.00%
16	6.14	4	0.00%
17	7.17	9	0.00%
18	10.33	30	0.00%
19	9.09	10	0.00%
20	8.48	27	0.00%

Table F.3: Overview of personal qualitative fatigue ratings in comparison to the RA requirement for all participants. RPE scores range from 6 to 20. The NASA TLX score is a combination of scores that relate fatigue and effort, with a minimum of 0 and a maximum of 60.



Figure F.4: Shoppers create an item slotting that decreases their walking distances. We see one slot 'D' and three items in its location, while their actual location is barely visible in the top right corner. The walking difference is not very big but can lead to different EE values during a shift.



Figure F.5: Another example of manual item reslotting. Shoppers place one box of mangoes closer to the station, resulting in virtually no walking time for every 10 mangoes that are picked. This figure also shows large height differences between item pick locations. This is not taken into account in the task decomposition.

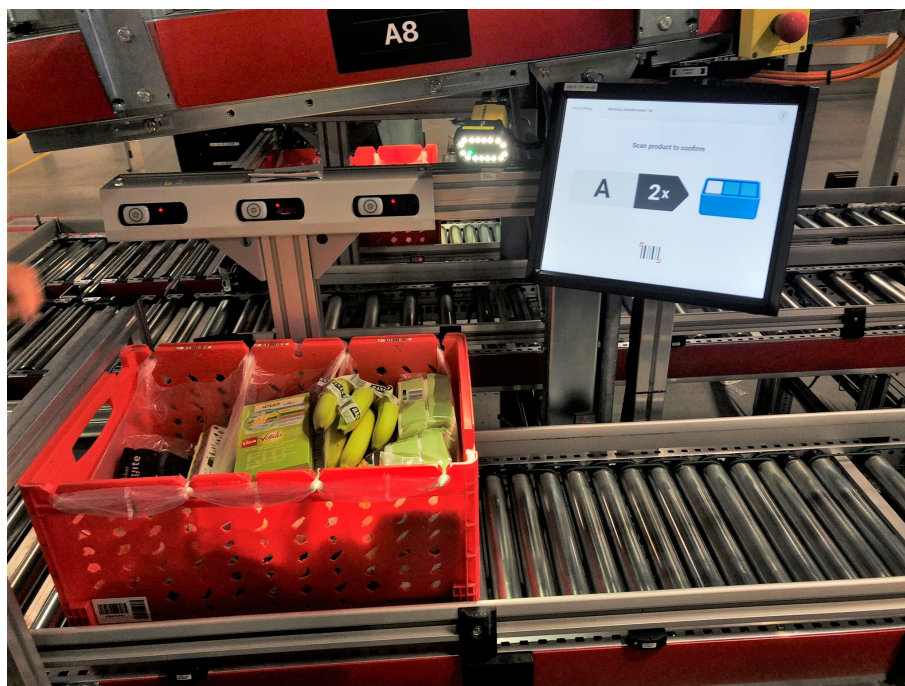


Figure F.6: When totes are very full and items do not fit easily anymore, shoppers need to repack the tote in a more efficient manner. This takes considerable time and effort.



Figure F.7: Shoppers need to get rid of empty crates, boxes and plastic packaging. This is usually stored or discarded close to the station, but can take around 20% of the total working time.



Figure F.8: Some items, like toilet paper, needs to be unpacked from a pallet before picking. This adds picking time and energy.