FEASIBILITY ANALYSIS AND EFFICIENT ROUTING FOR A PARTIALLY AUTOMATED DELIVERY SYSTEM WITHIN CHALMERS CAMPUS

Master's thesis in Department of Technology Management and Economics

Can Zhang
Bingcheng Wu

Department of Technology Management
Division of Service Management and Logistic
CHALMERS UNIVERSITY OF TECHNOLOGY
Department of Mathematics
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2021
Report No. E2021:137

REPORT NO. E2021:137

FEASIBILITY ANALYSIS AND EFFICIENT ROUTING FOR A PARTIALLY AUTOMATED DELIVERY SYSTEM WITHIN CHALMERS CAMPUS

Can Zhang
Bingcheng Wu

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Technology Management and Economics
Division of Service Management and Logistic
Chalmers University of Technology
Gothenburg, Sweden 2021

Department of Mathematical Sciences
University of Gothenburg
Gothenburg, Sweden 2021

FEASIBILITY ANALYSIS AND EFFICIENT ROUTING FOR A PARTIALLY AUTOMATED DELIVERY SYSTEM WITHIN CHALMERS CAMPUS

Can Zhang Bingcheng Wu

Supervisor: Lokesh Kumar Kalahasthi, Department of Technology Management and Economics

Examiner: Ivan Sanchez-Diaz, Department of Technoloogy Management and Economics

Report no.E2021:137
Master's Thesis 2021
Department of Mathematics
Department of Technology Management and Economics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Gothenburg, Sweden 2021

FEASIBILITY ANALYSIS AND EFFICIENT ROUTING FOR A PARTIALLY AUTOMATED DELIVERY SYSTEM WITHIN CHALMERS CAMPUS

Can Zhang
Bingcheng Wu

Department of Technology Management and Economics
Department of Mathematics
Chalmers University of Technology and University of Gothenburg

Abstract

The use of autonomous driving robot transportation technology to solve the last mile delivery problem is a research hot spot in today's logistics industry. This thesis aims to analyze the feasibility of applying an Automated Delivery Robot designed by Hugo company to load different sizes of packages and the efficient planning and effectiveness evaluation for the robot. In this thesis, these problems were solved in three steps. First, the GLM model is used to fit the size data of packages shipped by the Chalmers Transportation Center within one year. Secondly, this thesis proposes the concept of package unit to help solve the vehicle routing problem by unifying the volume of the packages. The package unit of each location is calculated through the 3D knapsack problem by the simulated annealing algorithm. Finally, a mixed integer linear programming model was created to optimize the total travel distance of the robot and the related energy consumption was calculated. A case study was conducted by inputting the one day data collected by us into the MILP model, the energy consumption on that day was obtained. The results of the case study shows the use of automated robot for package delivery in university campus is feasible and efficient. This thesis provides suggestions and inspiration for the practical application of automatic transportation on university campus. This thesis also focuses on the energy consumption of automated robots and calculates the approximate energy consumed by automated robots during operation.

Keywords: Last mile delivery,Robot Delivery, GLM regression, Knapsack problem, Heuristic algorithm, TSP, Mixed-integer linear programming, Energy consumption

Acknowledgements

We would like to express our very appreciation to our thesis mentor and examiner Ivan Sanchez-Diaz from Chalmers University of Technology and our supervisor Lokesh Kumar Kalahasthi from Chalmers University of Technology for their valuable and constructive suggestions and guidance during the whole project.

We would also like to thank Shruthik Krishnan for his work and data that is important to the completion of this thesis and Hugo company for the information of their automated robots. Many thanks to the working staff of Chalmars Transportation Center for their help in data collecting process, especially the protection against the Covid-19 situation.

We also want to appreciate our family and friends for their support and encouragement.

Can Zhang & Bingcheng Wu, Gothenburg, October 2021

Contents

List of Figures

List of Tables

1 Introduction

1.1 Research Background

The last mile delivery problem has always been a hot spot in the field of logistics and transportation research. Last mile delivery comprises up to 28% of the total delivery cost [69]. Solving last mile problem is considered as an isolated topic from increasing the overall efficiency of the functioning of logistics systems [1]. From the perspective of sustainability, the growth of urban population raises the demand of goods transportation, which causes problems in negative externalities. The negative externalites are considered as cost generated by air pollution, accidents and congestion, and is generally not borne by the transport users[40]. The innovation and development of last mile logistics in recent years can significantly reduce those negative externalities produced by last mile delivery activities in the urban area [53].

With the development of Information and Communication Technologies (ICT), industry 4.0 (A heading brought into the world by German government in 2011, aims at working with a higher level of automatization achieving a higher level of operational productivity and efficiency [3]) and autonomous driving technology around the world, driving robots are considered as an alternative solution to the last mile delivery problem. The flexibility and high efficiency of autonomous driving robots give it unique advantages in package delivery. In indoor environment like factories and warehouses, autonomous driving robots such as Automated Guided Vehicles (AGVs) have been widely used. However, in outdoor environments such as urban area and campus, there are still very few automated delivery robots actually being put into use [1].

Delivery robots are small in dimensions and must maintain a low speed to guarantee pedestrians' safety. These robots have to share their space with other transport devices or pedestrians and thus prefer to operate in suburbs and areas with comparatively low traffic [29]. The short battery life makes this type of robot unable to travel a long distance. These characteristics limit the use of delivery robots in a context of small area in which the packages delivered are small in both volume and weight [14]. Cargo transportation in small-scale operations such as universities, shopping malls or residential communities can fully meet this scenario. This research aims to analyse the feasibility of using Automated Delivery Robots for the last mile delivery inside the Chalmers University campus. Currently, the transportation mode used in the Chalmers University campus for package delivery among different destinations is a combined transportation mode of trucks and pickup electric vehicles. The truck used in the campus is a type of Heavy Duty Vehicle (HDV), which is the main consumer of fuel and also the main factor of CO2 emissions in road transportation. Figure 1.1 below shows the feature of the HDV.

Figure 1.1: The Heavy Duty Vehicle used in Chalmers Transportation Centre

The electric vehicles are generally used for ordinary packages with relatively small size or weight, while HDVs are used for large packages which are loaded with pallets. Both transportation methods require specialized personnel to load and drive and currently no optimized route planning methods are applied to reduce the cost or improve the efficiency. In our search, an autonomous delivery robot designed by *HUGO Transportation AB* is used to solve the last mile delivery problem from Chalmers Transportation Centre to other destinations in Chalmers Johanneberg campus. *HUGO Transportation AB* is a technology company which provides autonomous and contact free delivery solution that increases service quality, lowers costs and contributes to liming the spread of Pandemic [27]. The vehicle presents the following characteristics:

- Maximum load volume $500(L) \times 500(W) \times 600(H)$ mm
- Maximum load capacity 100 kg
- Maximum speed 8 km/h
- Maximum driving distance 20 km
- Lithium-ion battery capacity 731 Wh

- It can overcome obstacles up to 15 cm high and incident angle less than 50°
- The ground clearance of the robot is 5 cm, and it can be used in the temperature range of -5°C-30°C

Figure 1.2: The Autonomous Delivery Robot designed by Hugo

The above characteristics of the robot make it perfectly suitable for traveling within campus. Firstly, the allowed delivery time in one day is approximately 2 hours (starting from 14:00 to 16:00), and at the same time, the robot can travel for 2.5 hours at the maximum speed of 8 km/h (20/8 = 2.5 h). Secondly, the maximum load capacity of 100 kilograms can guarantee that even if the robot is fully loaded, the maximum loading capacity of the robot will not be exceeded. From the data collected, the average fit-rate calculated by the volume of a delivery package is 13.64 % and the average weight of a package is 3.615 kilograms. By rough estimation, a robot can load about 8 packages, and the total weight of these 8 packages is about 30 kilograms. Therefore, it is far less than the maximum load capacity of the robot. Accordingly, in the route planning, the weight of the package will not become a constraint. Thirdly, the geographical environment of the campus is simpler than in densely populated places such as urban centers, so the robot which can cross a height of 15 centimeters and can climb a steep slope of 50 degrees can transport smoothly on the campus.

1.2 Objectives and research questions

The application of Automated Delivery Robots (ADR) to partially replace traditional HDV transportation and pickup EV transportation will undoubtedly improve overall transportation efficiency, reduce greenhouse gas emissions, and comply with the concept of sustainable development. The feasibility of using these ADRs inside the scope of Chalmers University campus is studied in this research including two main topics:

1. The feasibility of applying Automated Delivery Robots to load different sizes of packages

Chalmers transportation center receives many packages and is responsible for delivering these packages to the destinations everyday. Among all the packages, some of them are obviously too large to be transported while some of them are hard to determine. Therefore, the study of the feasibility of package loading should be conducted.

2. The efficient planning and effectiveness evaluation for Automated Delivery Robot on the campus of Chalmers University of Technology

Efficient planning and effectiveness evaluation refer to calculating the minimum distance that the robot can run every day to meet delivery requirements.

These ADRs using lithium batteries as energy sources will not produce any carbon emissions. In order to further verify the advantages of using robots to replace the traditional transportation modes, the energy consumption of these robots was studied. It was calculated based on the actual operation data and route plan of these robots, and became an important basis for evaluating the effectiveness of this new model in the campus environment.

1.3 Software applied in this research

The programming language that used in the process of fitting the dimensions of one year packages is R. The specific software used is *R* studio.

The programming language that used in the process of analyzing loading problems and routing planning is *Python.* The specific software used in these processes is *PyCharm* which is an integrated development environment used in computer programming and *LindoApi* that is used to solve optimization problems.

Figure 3.1, 3.2 and 3.3 were generated by *Tableau.* Figure 3.4, 3.5 and 3.6 were drawn by the geographic information system *QGIS*. Figure 4.1 was generated by *R studio.* The flowchart figure 4.2 was generated by a free drawing software *draw.io.* Figure 5.1, 5.2, 5.3, 5.4, 5.5 and 5.6 were generated by *R studio* too.

Microsoft Excel was also used to record and organize data.

1.4 Outline

This thesis starts from a literature review of last mile delivery in the section 2. It summarizes the recent research on the last mile delivery problem, especially the use of robots to solve it. The research gap is also included in this section.

In section 3, we explained the data we collected and used in our research. In section 4, we introduced our approaches to solve the questions we faced in our research and their related theories. Section 5 includes our research results and related discussions. The conclusion as well as the limitations and future work are illustrated in section 6.

2

Literature Review

In this chapter, we have compiled the articles we have read about last mile delivery, especially the use of robots to solve last mile delivery. We also explain our research gap here.

2.1 Last mile delivery

2.1.1 The exploration of improving last-mile delivery

The concept of last-mile delivery is a central issue in logistics industry. In order to improve last-mile delivery, it requires the shorter delivery time and the lower cost; while another concept 'e-commerce fulfillment' can be used to determine the efficiency of last-mile delivery. Hau L. Lee and Seungjin Whang [33] indicated that there are two core concepts for making e-fulfillment efficient which are improving the use of information and capitalizing on current physical pope lines and infrastructures. Nils Boysen, Stefan Fedtke and Stefan Schwerdfeger [8] pointed out in their survey that the popularity of last-mile delivery was brought by the following developments and challenges including increasing volume, sustainability, costs, time pressure, and aging workforce.

Xuping Wang et al. [68] explored the competitiveness of three last mile delivery models including attended home delivery (AHD), reception box (RB), and collection-and-delivery points (CDPs). They used the vehicle routing problem model and genetic algorithm to solve the efficiency of the models and calculated the total cost of each mode according to operational efficiency and cost structure before coming to the conclusion that these three modes have their own advantages in different scenarios. Mikko Punakivi et al.[50] researched on the method of unattended reception of goods including the reception box and delivery box and indicated that the unattended reception of goods can reduce the home delivery cost considerable by up to 60%. They also pointed that the reason why the unattended delivery has not been widely used was because of the fact that it requires investment and commitment from the customer. Y Wang et al. [69] provided another way to solve the last mile delivery issue and proposed an effective large-scale mobile crowd-tasking model in which a large pool of citizen workers were used to perform the last-mile delivery that had the advantages including the highly parallel and independent delivery, one-to-one communication in crowd delivery, and the promotion of green supply chain and environmental protection. VE Castillo,JE Bell and WJ Rose [10] introduced the Crowdsourced Logistics (CSL) to solve the last-mile issue in which a shipper procured transportation services via a mobile or computer application directly from members of the crowd who provide those services as an independent contractor using a personally owned vehicle asset.

2.1.2 Using autonomously driving robots in last-mile delivery

Thomas Hoffmann and Gunnar Prause [29] indicated that Industry 4.0 leaded to the development of autonomously driving delivery robots and the robots were used for intro-supply chain transport in Industry 4.0 networks as well as for the delivery to the client on the last mile. Moritz Poeting et al. [48] indicated that because of the increasing shortage of space caused by urbanization, the increase in the number of packages and the demand for environmental protection, the parcel industry was seeking new concepts, such as drones, self-driving cars, sharing economy, delivery to car trunks and parcel robots; among all of these, the application of robots in the last-mile delivery represented a promising research field, not only in logistics, but also in mathematical optimization and simulation. Kottasova [32] reported that robots could provide a cheaper and safer solution to last-mile delivery as delivery trucks exacerbate traffic congestion and their drives always found it hard to park. Wenmin Wang et al. [67] demonstrated that last mile indoor operation is reliable by processing an advanced low-cost and accurate intelligent localization and mapping algorithm that combined the IMU sensor and ORB-SLAM. A white paper presented by STANFORD VALUE CHAIN INNOVATION INITIATIVE [9] indicated that customers desired for flexible, fast, and cheap or free delivery. Delivery robots could meet the requirements of customers for fast delivery and flexibility in choosing convenient delivery time, could reduce environment impact as delivery robots did not exhaust carbon dioxide; had regulatory advantages as they were designed to take on pedestrian lanes and travel at low speeds; cost lower than drones, etc. N Boysen et al. [7] introduced the autonomous delivery robots launched from trucks (small autonomous robots on board) to be dedicated to a single customer launched from the truck and return. They applied the Mixed-Integer Programming (MIP) model to get the efficient approach to minimize the number of late delivery. M Ostermeier et al. [44] also analyzed the truck-and-robot deliveries in their paper. They transformed the problem into mixed integer planning and decomposed it into vehicle path planning and robot scheduling and found that the concept of trucks and robots can reduce the cost of the last mile by 68% compared to just trucking. Chen Cheng et al. [12] also analyzed the delivery robots and the vehicle routing problem by a mixed-integer linear programming mode and indicated that delivery robots provide cheaper, safer, and more environmentally friendly solutions to the current unsustainable last-mile challenge. Dylan Jennings and Miguel Figliozzi [31] introduced the sidewalk automated (or autonomous) delivery robots (SADRs) that delivered items to customers without the intervention of a delivery person and came to a conclusion that it could provide substantial cost and time savings in some scenarios and could significantly reduce on-road travel per package delivered by using continuous approximations. Sonneberg et al. [62] described how to solve the last-mile delivery operations with autonomous unmanned ground vehicles by a mixed-integer linear problem and developed the optimization model to establish a delivery network of stations to determine the optimal assignment of customers to these stations as well as calculating the amount of vehicles, number of customer orders, and the number of driving trips. Their research showed that the delivery robots could save personnel expenses, road space, emissions and noise. According to the search of Pani Agnivesh and his group [46], the analysis of consumer preference data showed that 61.28% of consumers are willing to pay extra to receive deliveries using autonomous delivery robot technology. Miguel Figliozzi and Dylan Jennings [22] indicated that utilizing autonomous delivery robots has significant potential to reduce energy consumption and carbon dioxide emissions in urban areas analyzed by using derived formulas based on continuous approximations of distribution problems. Farah Samouh et al. [55] indicated that applying autonomous robots for last-mile delivery helped alleviate urban congestion. Miguel A. Figliozzi [21] indicated that automated delivery robots had great potential to reduce energy consumption and reduce carbon dioxide emissions. Michele D. Simoni et al. [60] indicated that using delivery robots in urban areas could reduce waiting time and have a chance to alleviate congestion.

Delivery robots have already begun to be applied in real life. Vincent [66] reported in the twenty seventh of February, 2019 that FedEx launched their delivery robot named FedEx SameDay Bot and many startups and large firms have begun trials with similar technology. Diza [18] introduced eight leading delivery robots; including SEGWAY LOOMO DELIVERY that was designed to be used in office environment; ANYBOTICS AND CONTINENTAL; POSTMATES SERVE that could carry 50 pounds within a 30-mile radius and was deployed in Los Angeles; MARBLE that had been delivering food through a partnership with Yelp Eat24 in San Francisco since 2017; BOXBOT; NURO that was designed to delivery food and could be equipped with both refrigerated and heated compartments as needed; KIWIBOT that were 65% faster than human couriers and was specialized in personal food deliveryl; and STARSHIP ROBOTS.

Moritz Poeting et al. [48] presented a simulation model to study the parcel delivery in urban area and proved their model is useful in the allocation and scheduling of the delivery robots that reflected in higher service levels and lower delay times.

Although using robots for last-mile delivery is the general trend of technology, a lot of disputation occurred. Francis [24] reported that autonomous vehicles might will be allowed to share the pavements with humans although it might face some resistance like if the surrounding humans getting enough consideration. Julia [71] reported that delivery robots occupied the sidewalks in San Francisco while supporters and opponents had a lot of arguments about this issue. Simon [59] indicated in his report that delivery robots had received strict restrictions in San Francisco.

2.2 Research Gap

From the literature mentioned above, it can be noticed that they did not try to focus on the use of delivery robots in the university campus environment which can be considered as a special environment that has a complex flow of people and less congested traffic. Few papers discussed the exact energy consumption although it was mentioned that using autonomous robots could help save energy.

Our research focuses on the use of autonomous delivery robots in the campus of Chalmers University of Technology. The research will cover the analysis of daily information of packages and the exact energy consumption of robots operated. A study on knapsack problem was also conducted to explore the placement of packages and a model of vehicle routing problem was made to analyze the best solution of routing. It may provide a good example for the use of autonomous delivery robots within campus, and study how much energy the robot saves compared to ordinary transportation.

3
Data

The data used in the research includes three main parts: the one year dataset of package information provided by Chalmers Transportation Center; one week dataset of package information collected by our research team; and one week dataset of robot operation information collected by another research team. How the data was collected and the structure of the data will be discussed below.

3.1 One Year Dataset of Packages

The one year dataset was provided by the Chalmers Transportation Center. This dataset was extracted from the delivery system operated by Chalmers Transportation Center which contained detailed delivery information of packages that shipped from the Transportation Center between March 18th, 2019 and March 16th, 2020 including their IDs, conveyors, senders, destinations, receivers, package types, delivery time and date. The significant information used in the research were senders, destinations and package types, which were used to fit the dimensional data of the packages. Table 3.1 below shows all the destinations and table 3.2 shows the explanation for the abbreviation of each Package Type.

Table 3.1: Destinations

No.	Destination
1	ACE
2	Akademiska hus
3	Biblioteket
4	CFAB
5	Chalmers verksamhetsstöd
6	CTC
7	E-huset
8	Fysik
9	GMV
10	Kansli I
11	Karen/Cremona
12	Kemi
13	Lindholmen
14	Maskin
15	Mattematiska Vetenskaper
16	Mistra Urban Futures
17	Postrum SB3
18	SB3
19	SGI
20	Stena center
21	Stiftelsen
22	Teknikparken
23	Vasa 11 V3
24	Vasa 11 V4
25	Vasa 15
26	Vasa 2-3
27	Vasa 7

Table 3.2: Package Types

No.	Abbr.	Explanation
1	EXP	Packages from DHL, UPS, Fedex, delivered when arrived
2	FRYS	Cold packages with a fix dimension
3	INT	Internal letters
4	PALL	Packages Loaded to pallets
5	PKG	Ordinary Packages (small<=20kg and big >20kg)
6	REK	Personal Letters Registered Mail Domestic

There are a total of 33,731 records in the entire data set. In order to get an intuitive insight of the one year dataset, a stack column chart was plotted and shown below:

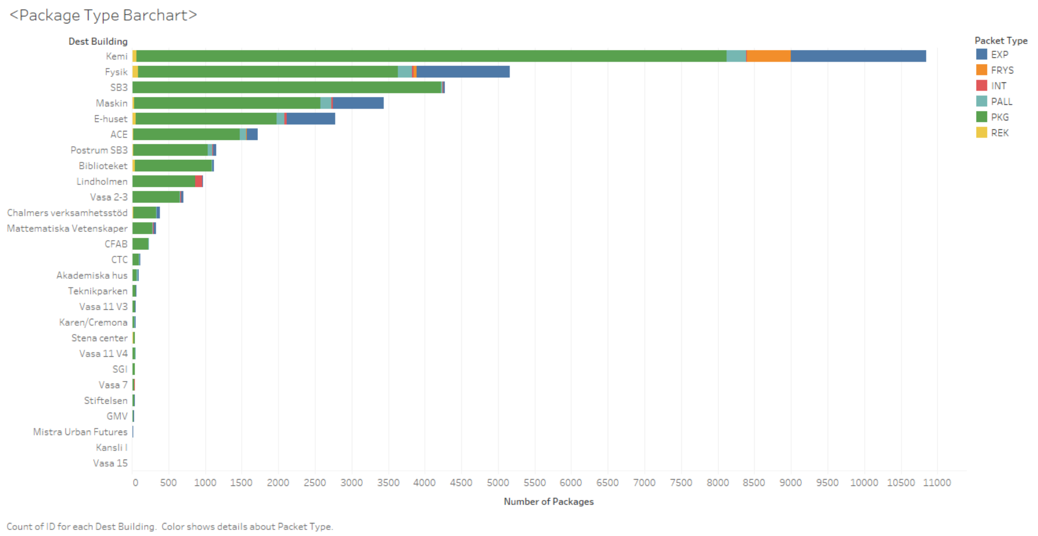

Figure 3.1: Destinations and Package types for one-year dataset

It can be directly summarized from the chart 3.1 that most of the packages were shipped to Kemi, Fysik, and SB3, and the most common types were PKG and EXP. To explore the findings, the pie charts for destinations and package types were plotted and shown below:

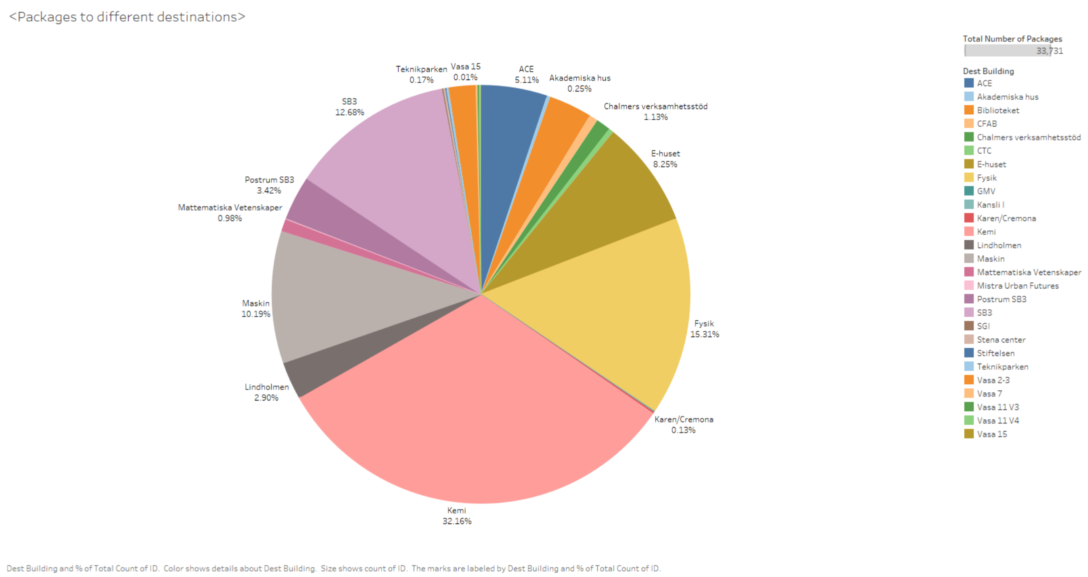

Figure 3.2: Packages to different destinations

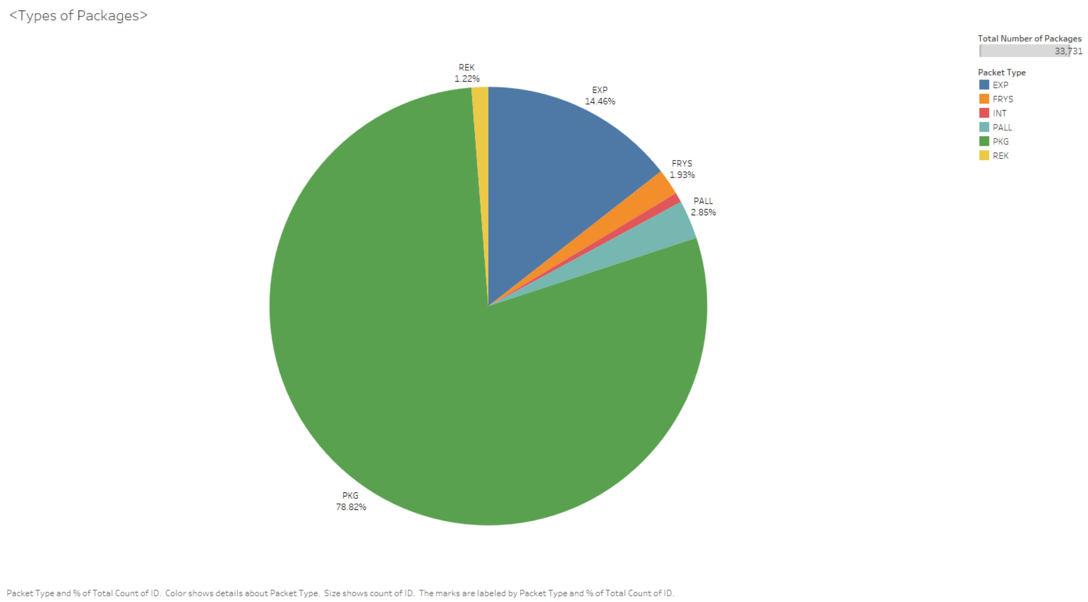

Figure 3.3: Package Types

Figure 3.2 shows that about 32% of the packages were shipped to Kemi, 15% of them were shipped to Fysik and 13% were shipped to SB3, which indicates that the first observation of the destination is correct. Meanwhile, figure 3.3 shows that, 79% of the packages were PKG and 14% of the packages were EXP.

3.2 One Week Dataset of Pacakges

The one year dataset did not contain information about the dimensions of those packages, thus a further study to fit dimensions according to the given information was required. In order to build models to obtain the dimensional data, a data collection of delivery packages within one week was conducted in Chalmers Transportation Center from February 8th, 2021 to February 13th, 2021.

During the collecting process, the dimensions including length, width, height and the weight were measured and recorded manually. The dimensions were recorded in centimeters and the weights were recorded in grams. The corresponding destinations and package types were also recorded through the labels attached to the package. The information about the senders of the packages were recorded to infer the commodity types of those packages. Commodity types include clothes, daily necessities, electronic product, food, games, IT infrastructures, laboratory equipment, machinery, medical supply, rubber products, tools and supplies, and unknown. Unknown denotes those packages cannot be inferred by its sender.

The commodities were inferred according to the information of senders. Firstly, we made an assumption about the senders that each sender corresponds to only one commodity because we could not get any other information about commodities beside senders from packages and its would help analyze. A total of 2669 senders were recorded in the one year data; however, we chose the top seventy senders that contains 74%(24943/33731) of all the packages to represent all of them. Among these seventy senders, some of them had to be ignored because they did not provide enough information or it was impossible to infer the commodity from the sender's name. For example, we could not get any information of commodities from the senders Germany, China, United Kingdom, and Unknown .etc. Some of them had to be combined as they were actually the same company such as Sigma-Aldrich Chemie and Sigma. The information of senders were searched online mainly from their official website. Some senders and commodities were easy to infer like Adlibris and'Bokus as they are bookstores so the commodity could be inferred as books. Some senders like VWR international had to be analyzed with destination. VWR international is a company of life sciences, biotechnology, pharma, etc. and its corresponding destinations were all Kemi which is the chemistry building so the commodity was inferred as Laboratory equipment.

3.3 One week Dataset of Robots

The one week robot operation data that included GPS data, position data, and energy consumption of left and right motors was provided by another research team. The data was used in the route planning and efficiency evaluation of the robots. It was collected by both manually manipulating the robots, and automatically running of these robots for 7 days. In the day 1,2 and 6, there was no GPS data available while in the other four days, the GPS was on and thus the data for these four days included the latitude and longitude. In order to plot those geographic data, a software called *QGIS* was used in the research.

QGIS is a a Free and Open source Geographic Information System (FOSS)[70]. Its main functions include browsing data and designing maps; creating, editing, managing and exporting data; and data analysis.

The figure 3.4 shows the locations of these destinations in an open street map generated by *QGIS* and figure 3.5 shows the network built by the geographic data which was collected by the robots.

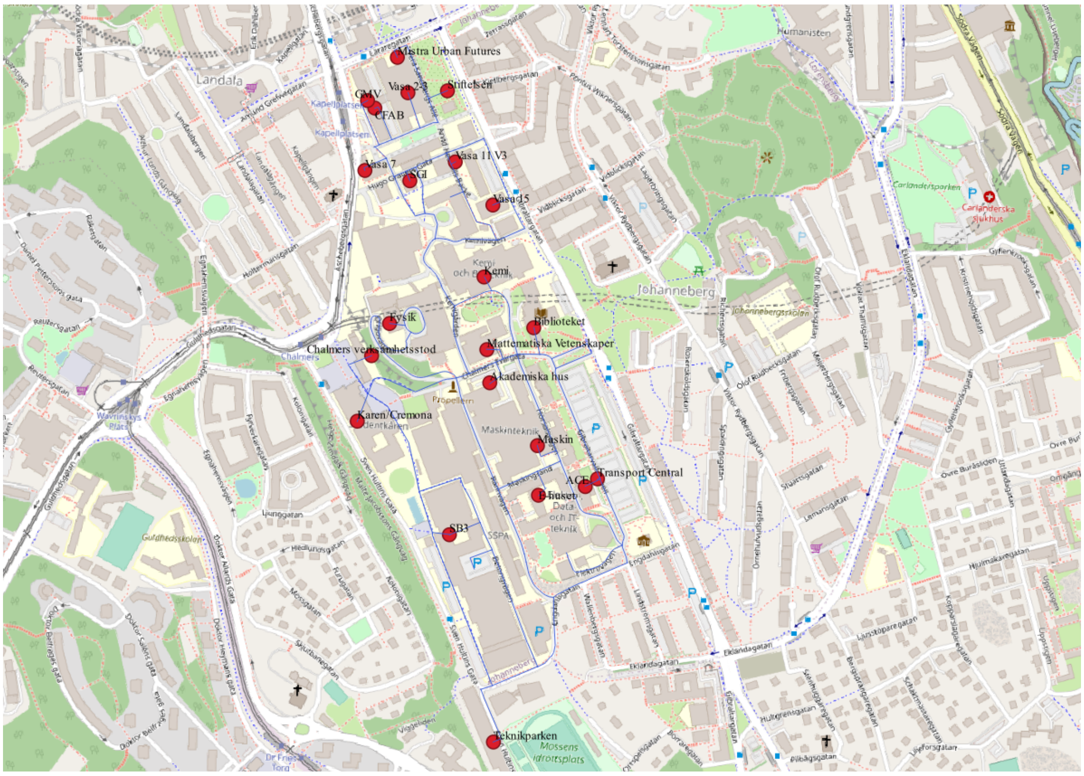

Figure 3.4: Map of destinations

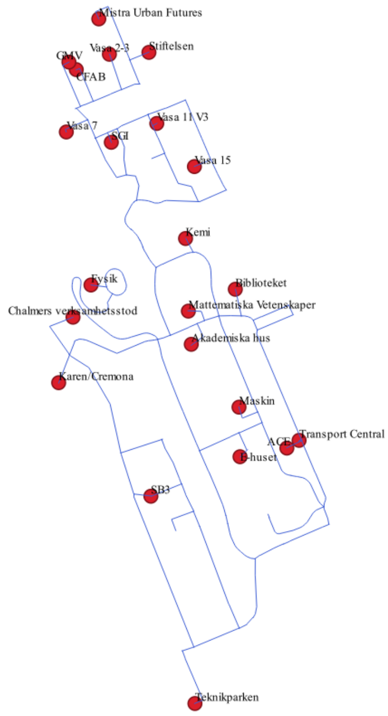

Figure 3.5: Network of destinations

QGIS can also be used to draw the trajectory of the robot. Figure 3.6 shows the path of the robot on the 7th day of the operation between various buildings on the campus, where each node is separated by 1 second. Since the position of the robot is known every second, the speed of the robot can be calculated accordingly.

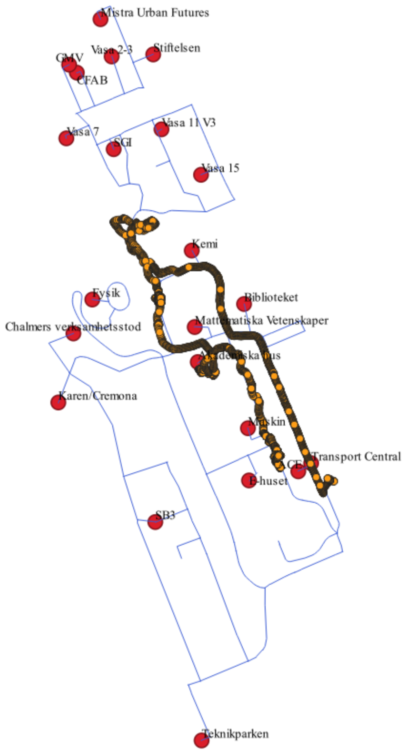

Figure 3.6: Trajectory of the robot in one day

4
Methods

In this chapter, we introduce how we conduct our research that can be divided into three parts and their related theories. Firstly, we built regression models to fit the dimensional data of packages. Secondly, we analyzed the knapsack problem and proposed the concept of package unit for the route planning. Lastly, we built a mixed-integer linear programming (MILP) model to solve the route planning problem and computed the energy consumption.

4.1 Fit the dimensional data

As mentioned in Section 3.1 and Section 3.2, not dimensional data was concluded in the one year data of packages. We have to built models to fit these dimensional data for our steps.

In this process, regression models were build to fit the dimensional data. According to the results of analyzing the one year data that will be shown on the results and discussion section of this thesis, we found that the Gamma GLM is the most critical method used to fit the dimension data of one year dataset. In this section, the related theory about generalized linear model and the generalized linear model with a Gamma-distributed dependent variable are introduced.

In statistics, one of the most commonly used methods for modeling and analyzing the relationship between specific variables is regression analysis. The output variable of regression analysis is usually denoted as Y, which is also named as dependent variable, response variable, explained variable, predicted variable or regressand variable. The input variable of regression analysis is usually denoted as $x_1, x_2, ..., x_p$, which is also named as independent variable, explanatory variable, predictor variable, or regressor variable. The most common regression analysis method is General linear model where the expression can be written as:

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{4.1}$$

β_0 in this equation represents the intercept that refers to where the graph of equation $y = \beta_0 + \beta_1 x$ will meet at y-axis when $x = 0$. β_1 represents the slope that is the change in $y = \beta_0 + \beta_1 x$ when x is changed to $x + 1$. ε represents the errors that is the deviation between actual observations and their estimated values.

A linear model can have many predictors, so the general linear model with n predictors can be expressed as:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \tag{4.2}$$

$\beta_i, i = 1, 2, \cdots, n$ here are called unknown parameters.

The characteristics of General Linear Model can be summarized as follows:

• The response variable Y and the error ε obey the normal distribution. The variance in the ordinary linear model does not change with the value of the independent variable x.

• The predictive quantity x_n is non-randomness, it is measurable and there is no measurement error. The unknown parameter β_i is considered to be an unknown but not random constant.

• The research object is always the expected value that can be denoted as $E[Y]$. $E[Y]$ can be written as:

$$E[Y] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{4.3}$$

It can be observed from equation 3.3 that the expected value and the linear combination of predictors can be connected by identity. A new method can be applied to represent this connection after introducing a new concept of link function. A link function $f(x) = x$ can be applied here and equation 3.3 will be expressed as:

$$E[Y] = f(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{4.4}$$

More information about link function will be explained later when the Generalized Linear Model is introduced.

4.1.1 GLM model

Although the General Linear Model is useful, the Generalized Linear Models (GLM) still need to be applied to extend the linear modelling framework to variables that are not normally distributed. Generalized linear models were formulated by John Nelder and Robert Wedderburn [42]. It allows the response variable to have an error distribution model other than the normal distribution. The distribution of the response variable is extended to the exponentially dispersed family including Poisson distribution, Binomial distribution, negative Binomial distribution, Gamma distribution, inverse Gaussian distribution, etc.

A GLM model is made up of four parts:

1. The probability distribution of the response variable such as the normal distribution for Y in the linear regression, the binomial distribution for Y in the binary logistic regression, etc.

2. A linear predictor:

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} \tag{4.5}$$

3. A link function:
It describes how the mean: $E(Y_i) = \mu_i$ depends on the linear predictor.

$$g(\mu_i) = \eta_i \tag{4.6}$$

4. A variance function:
It describes how the variance $Var(Y_i)$ depends on the mean.

$$Var(Y_i) = \phi V(\mu) \tag{4.7}$$

As mentioned before, the General Linear Model can be represented with a link function $f(x) = x$ which means that the general linear model can be seen as a special case of the Generalized Linear Model where $\varepsilon \sim N(0, \sigma^2)$. The linear predictor is $\eta_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$, the link function is $g(\mu_i) = \mu_i$, and the variance function is $V(\mu_i) = 1$.

The link function is an important element for the Generalized Linear Model. It is a function of the dependent variable y, which connects the result of the previous linear prediction with the value of the dependent variable y. It is y itself in the ordinary linear regression. The table below shows the link functions of common distributions.

In general, the relationship between the population mean and the linear predictor is determined by a link function.The link function converts the probability of the level of the categorical response variable into an unbounded continuous scale. After the conversion is complete, the relationship between the predictors and the response

Table 4.1: Link Functions of common distributions

Distributions	Link name	Canonical link functions
Binomial	Logit	$g(\mu) = ln(\frac{\mu}{n-\mu})$
Gaussian	Identity	$g(\mu) = \mu$
Gamma	Inverse	$g(\mu) = \mu^{-1}$
inverse.Gaussian	Inverse Squared	$g(\mu) = \mu^{-2}$
Poisson	Logit	$g(\mu) = ln(\mu)$

can be modeled with linear regression.

The choice of the link function determines the nature of the model used. The appropriate link function may not be known at the beginning, and in the case of standard homogeneous variance, people may not know the appropriate model. Indeed, various link functions may be investigated through modern computer software just like trying various models in standard regression methods. However, depending on the distribution under consideration, the natural link function can be used. In fact, GLM modeling should be regarded as a method of choosing distribution and linking. For each distribution, there is a natural link, which is derived by setting the natural location parameter equal to the linear predictor variable. That is, the link called the canonical link is derived from:

$$\theta = x'\beta. \tag{4.8}$$

Given a response y, the generalized linear model (GLM) is:

$$f(y) = c(y, \phi)exp\left\{\frac{y\theta - a(\theta)}{\phi}\right\}, \tag{4.9}$$

$$g(\mu) = x'\beta. \tag{4.10}$$

The equation for $f(y)$ specifies that the distribution of the response is in the exponential family. The second equation specifies that a transformation of the mean, $g(\mu)$, is linearly related to explanatory variables contained in x.

4.1.2 GLM model with Gamma distribution

As mentioned in the previous chapter, the Generalized Linear Models extend to the exponential family. In this research, the exponential distribution used is Gamma distribution.

Gamma distribution can be analyzed from two aspects; the first one is from exponential distribution and the second one is from the Chi-Square distribution.

Gamma distribution is the distribution of the sum of multiple independent and identically distributed (iid) exponential distribution variables where the distribution can be expressed as:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases} \tag{4.11}$$

The rate parameter $\lambda > 0$ is the parameter of the distribution.

The gamma distribution can be parameterized in terms of a shape parameter $\alpha = k$ and an inverse scale parameter $\beta = \frac{1}{\theta}$ and it is the rate parameter of gamma distribution. Therefore, a random variable X that is gamma distributed with the shape parameter α and rate parameter β can be expressed as:

$$X \sim \Gamma(\alpha, \beta) \equiv Gamma(\alpha, \beta). \tag{4.12}$$

The corresponding probability density function can be expressed as:

$$f(x;\alpha,\beta) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \tag{4.13}$$

where $x > 0$, $\alpha and \beta > 0$, the $\Gamma(\alpha)$ is the gamma function, if α is positive integers, $\Gamma(\alpha) = (\alpha = 1)!$. If α is complex numbers with a positive real part, $\Gamma(\alpha) = \int_0^{\infty} x^{z-1} e^{-x} dx$. If $\alpha = 1$, the equation then will be $f(x) = \beta e^{-\beta x}$ and this is the expression of exponential distribution.

Gamma distribution can also be seen as the extension of Chi-Square Distribution. The chi-square distribution is the distribution of the sum of squares of independent $N(0, 1)$ random variables, denoted as $y2$. where the parameter is called the degrees of freedom. Chi-squared random variables are non negative, and their distribution is skewed to the right. The mean and variance are and 2, respectively. For large , y is approximately normal. The chi-square distribution is also defined for non-integral > 0 degrees of freedom; and this distribution is conveniently thought of as intermediate between the two chi-square distributions with integer degrees of freedom which bracket .

Multiplying a 22 random variable by ţ/(2) yields a gamma random variable with parameters ţ and , denoted G(ţ,). The Chi-square distribution with n degrees of freedom can be denoted as $gamma(n/2, 2)$.

A gamma GLM is of the form:

$$y \sim G(\mu, v), \tag{4.14}$$

$$g(\mu) = x'\beta. \tag{4.15}$$

The canonical link for the gamma distribution is the inverse function. Since parameters from a model with inverse link are difficult to interpret, the log link is usually regarded as more useful.

According to *Generalized linear models for insurance data*[16], *An introduction to generalized linear models*[17] and many other papers or books, Gamma GLMs are used to model the data of non-negative continuous random variables with a long right tail.

4.1.3 GLM model and R

The GLM can be fitted by glm() function in R. R is also the software used in this research to fit the GLM model. The glm() function in R is described as:

```
glm(formula, family = gaussian, data, weights, subset,
    na.action, start = NULL, etastart, mustart, offset,
    control = list(...), model = TRUE, method = "glm.fit",
    x = FALSE, y = TRUE, singular.ok = TRUE, contrasts = NULL, ...)
```

'Formula' here refers to the symbolic description of the model to be fitted. 'Family' is a description of the error distribution and link function to be used in the model. The commonly used distribution and its canonical link function with R code are shown on the table below.

Table 4.2: Distribution and Link Function

Family	Canonical Link Function	R code
Binomial	(link = "logit")	family = binomial(link="logit")
Gaussian	(link = "identity")	family = gaussian(link = "identity")
Gamma	(link = "inverse")	family = gamma (link = "inverse")
inverse.Gaussian	(link = "1/mu2̂")	family = inverse.gaussian(link = "1/mu2̂")
Poisson	(link = "log")	family = poisson (link = "log")

'Data' is the optional data frame, list or environment that contains the variables in the model. 'Weights' should be Null or a numeric vector that represents an optional vector of 'prior weights' to be used in the fitting process. 'Subset' is an optional vector, that is used to specify a subset of observations to be used in the fitting process. 'Na.action' is a function that represents what happens when NA is included in the data. The default of 'Na.action' is set by the 'na.action' setting of 'options'; and if it is not set, it is 'na.fail'. When this value is Null, it means nothing will happen. 'Start' represents the starting values for the parameters in the linear predictor and the default is Null. 'Etastart' and 'mustart' refer to the starting values for the linear predictor and the vector of means. 'Offset' can be used to specify an a priori known component to be included in the linear predictor during fitting and it should be NULL or a numeric vector of length equal to the number of cases. 'Control' is a list of parameters used to control the fitting process such as parameters to control algorithm error and maximum number of iterations. 'Model' is a logical value to indicate whether the model frame should be used as part of the return value. 'Method' means the method to be used in fitting the model and the default method 'glm.fit' refers to use Iterative Weighted Least Square(IWLS). 'X' and 'y' are the logical values refer to whether the response vector and model matrix used in the fitting process should be used as part of the return values. 'Singular.o'k is also a logical item. 'Contrasts' are the optional lists. Generally, 'formula', 'family', and 'data' are the most important parameters and should be specified in use.

Some auxiliary functions like summary() and coef() .etc can be used to analyze the generalized linear model created by function glm(). Figure 4.1 below is the result after applying glm() function to the cleaned iris dataset contained by R. All the results of the function are shown in the figure.

```
Call:
glm(formula = Species ~ ., family = binomial(link = "logit"),
    data = iris)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.01105  -0.00065   0.00000   0.00048   1.78065

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -42.638     25.708  -1.659   0.0972 .
Sepal.Length   -2.465      2.394  -1.030   0.3032
Sepal.Width    -6.681      4.480  -1.491   0.1359
Petal.Length    9.429      4.737   1.990   0.0465 *
Petal.Width    18.286      9.743   1.877   0.0605 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 190.954  on 149  degrees of freedom
Residual deviance:  11.899  on 145  degrees of freedom
AIC: 21.899

Number of Fisher Scoring iterations: 12
```

Figure 4.1: An example of the result of glm() in R

The coefficients 'Estimate', 'Std. Error', 'z value' and Pr(>|z|) are the main results about the relationship between independent variables and dependent variables.

• Estimate refers to the intercept which equals to the β coefficient associated with each predictor.

• Std.Error refers to the standard error of coefficient estimate which represents the accuracy of the coefficient. The accuracy of the estimated value is inversely related to Std.Error.

• z value is equal to the quotient of Estimate divided by Std.Error. This represents the relationship between each independent variable and the dependent variable. When the absolute value of z value is larger, the corresponding independent variable becomes more important. Generally speaking, when the absolute value of z value is greater than 2.0, the variable is significant which means that there is statistical evidence that it is related to the dependent variable.

• Pr(>|z|) is also named as p value. The P-value is a probability that is used to measure the evidence that negates the null hypothesis or to test for whether the coefficient point estimate is different from 0. When the p value is low, it is sufficient to deny the null hypothesis. Usually, alpha or α is used to represent the significance level and $\alpha = 0.05$ is always used which means the risk of correlation is 5% when there is actually no correlation. If the P value is less than or equal to the significance level, it can be concluded that the association between the response variable and the independent variable is statistically significant; otherwise, the association is not significant and a model without his term should be refit.

Deviance Residuals are also shown in the results section to represent the contributions of individual samples to the deviance and are defined as the signed square roots of the unit deviance.

Some other information are also included in the results including null deviance, residual deviance, AIC and number of fisher scoring interactions.

Null deviance: Deviance is a measure of goodness of fit of a generalized linear model, it is a measure of badness of fit–higher numbers indicate worse fit. The null deviance shows how well the response variable is predicted by a model that includes only the intercept (grand mean).

Residual deviance: In R, the deviance residuals represent the contributions of individual samples to the deviance, they are defined as the signed square roots of the unit deviance.

AIC: AIC is the abbreviation of Akaike Information Criteria (AIC), which provides a method to evaluate model quality by comparing related models. It is based on the deviance, but will be punished due to the complexity of the model. Its purpose is to prevent from including irrelevant predictors, but the numbers themselves are meaningless. The model with the smallest AIC should be selected among all the similar candidate models. Therefore, it is useful for comparing models, but cannot be explained separately.

Number of Fisher Scoring iterations: Fisher's scoring algorithm is a derivative of Newton's method for solving maximum likelihood problems numerically. This does not really give a lot of information, other than the fact that the model did indeed converge.

4.2 Package loading analysis

The dimensional data was fitted in the process above and these data was then be used to analyze the package loading problem. This problem can be regarded as a classic NP-Complete problem, the Knapsack problem. We used the simulated annealing algorithm to solve this problem.

4.2.1 Literature Review of Loading efficiency

The knapsack problem has a long history of over one hundred years and can be tracked back to the problem of partition of numbers that Mathews [41] indicated and explained in 1897. Ross Keith and Danny Tsang [54] defined the classical knapsack problem as packing a knapsack of integer volume F with objects from K different classes in order to maximize profit; They also indicated that when the knapsack volume F is an integer multiple of the object volumes, the problem has a simple solution that fill the highest profit to volume ratio; meanwhile, if the knapsack volume ratio is not an integer multiple of the object volumes, then the problem can still be solved in O(FK) time with dynamic programming. WEI SHIH [57] presented an efficient solution algorithm based on the branch and bound search process to solve the multiconstraint zero-one knapsack problem and showed that this method was more efficient than the original Balas and improved Balas additive algorithms. The basic feature of the branch and bound method is two decision rules; one of which provides a process for estimating the upper limit of the objective function on the node, and the other specifies the selection criteria for the selection of branch variables on the node for further partitioning. Dudziński and Walukiewicz [19] introduced how to solve the binary knapsack problem and stressed the importance of dual methods for solving linear programming relaxations of the considered problems. They described two ways of generalization of the knapsack problem, of which the first one was to obtain multiple-choice knapsack problem if the special ordered sets are added and the second one was to get the nested knapsack problem if the the constraints have the nested structure. Hochbaum [28] introduced the nonlinear Knapsack problem as to maximize the separable concave objective function constrained by a single "package" under the condition of non-negative variables and introduced a fully polynomial approximation scheme to solve it. Kaiping Luo and Qiuhong Zhao [38] introduced a new method called Grey Wolf Optimizer (GWO) that is a meta-heuristic that mimics the leadership hierarchy and group hunting mechanism of grey wolves in nature and developed a binary version to tackle the multidimensional knapsack problem which has an extensive engineering background. Their experimental results statistically show the effectiveness of the new optimizer and the superiority of the proposed algorithm in solving the multidimensional knapsack problem, especially the large-scale problem. Sinha Prabhakant and Andris Zoltners [61] indicated that the multiple-choice knapnack problem is defined as a binary knapsack problem with the addition of disjoint multiple-choice constraints and extracted that the advantage of the branch and bound algorithm we proposed for this problem lies in the fast solution of linear programming relaxation and its effective subsequent optimization as a branch result. J Puchinger and GR Raidl

[49] studied the multidimensional knapsack problem and presented some theoretical and empirical results about its structure. They evaluated the different integer linear programming-based, meta-heuristic, and collaborative approaches for it and took advantage of the empirical analysis to develop new concepts for solving the MKP using integer linear programming-based and memetic algorithms. They also conducted further computational experiments with longer running times to compare the solution of their method with the most famous solution of another leading method to date using the MKP benchmark example. The results they obtained prove the effectiveness of the proposed method, and compared with the previously described method, their method has a shorter running time.

4.2.2 The simulated annealing algorithm

We used simulated annealing algorithm to solve this problem. The simulated annealing algorithm was firstly invented by S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi in 1983 [56]. It is an algorithm based on Monte-Carlo iterative solution strategy. Its principle is to start from a certain higher initial temperature, with the continuous decrease of temperature parameters, combined with the probability of sudden jump characteristics to randomly find the global optimal solution of the objective function in the solution space; that is, the local optimal solution can probabilistically jump out and eventually tend to the global optimal.

The simulated annealing algorithm can be described as follow. The first step is to randomly generate an initial solution S and bring it into the objective function $f(x)$ and define a large enough value T as the initial temperature. Then a new solution S' in the solution space will be generated by a generating function from the current solution S and the difference of the objective function $\Delta t' = f(S') - f(S)$ corresponding to the new solution will be calculated. The temperature T will also be reduced. Next, it is judged whether the new solution is accepted. The judgment is based on an acceptance criterion and the most commonly used acceptance criterion is the Metropolis criterion: if $\Delta t' < 0$ then S' will be accepted as the new current solution S, otherwise the probability $p = exp(-\Delta t'/T)$ will be used to accept S' as the new current solution S. When the new solution is determined to be accepted, replace the current solution with the new solution. When the temperature T drops to a certain minimum value, or the new solution cannot be accepted after completing the given number of iterations, the iteration is stopped, and the currently sought optimal solution is accepted as the final solution.

The objective function $f(x)$ receives a *sequence triple* of packages and estimate the number of packages that can be loaded in the robot and return it. The solution S is the *sequence triple* which determines the order of packages loaded into the robot [73]. The loading of the ordered packages is called *robot packing*.A *robot packing* is a packing which can be achieved by successively placing boxes starting from the bottom-left- behind corner, and such that each box is in-front of, right of, or over each of the previously placed boxes [74]. A new solution S' can be generated by changing elements in the *sequence triple* [73].

Algorithm 1 Simulated annealing algorithm for package loading

```
generate an initial solution r ∈ R;
choose initial time t_0
choose time step t_s
number of accepted answers k := 0
repeat
    generate r' ∈ N(r)
    if f(r') ≥ f(r) then
        accept := true
    else
        p := rand(0, 1)
        T := 1 / (t_0 + t_s · k)
        Δ := (f(r) − f(r')) / f(r)
        if p < e^(−Δ/T) then
            accept := true
        end if
    end if
    if accept then
        r := r'
        k := k + 1
    end if
until stopping-criteria
return result
```

4.2.3 Package unit

A concept of package unit was proposed in this research. The package unit is calculated by solving the package loading problem. This concept is proposed to solve the following vehicle routing problem. The sizes of packages going to different locations are different, so the number of packages carried by the robot are different. The concept of package unit is used to to unify the volume of the packages. The package unit is the smallest unit used to define a delivery requirement. It contains different number of packages which vary from destination to destination. By defining the package unit for each destination, the daily delivery requirement can thus be expressed as number of package units. The robot has a maximum capacity measured by the package unit.

4.2.4 Flowchart of loading packages

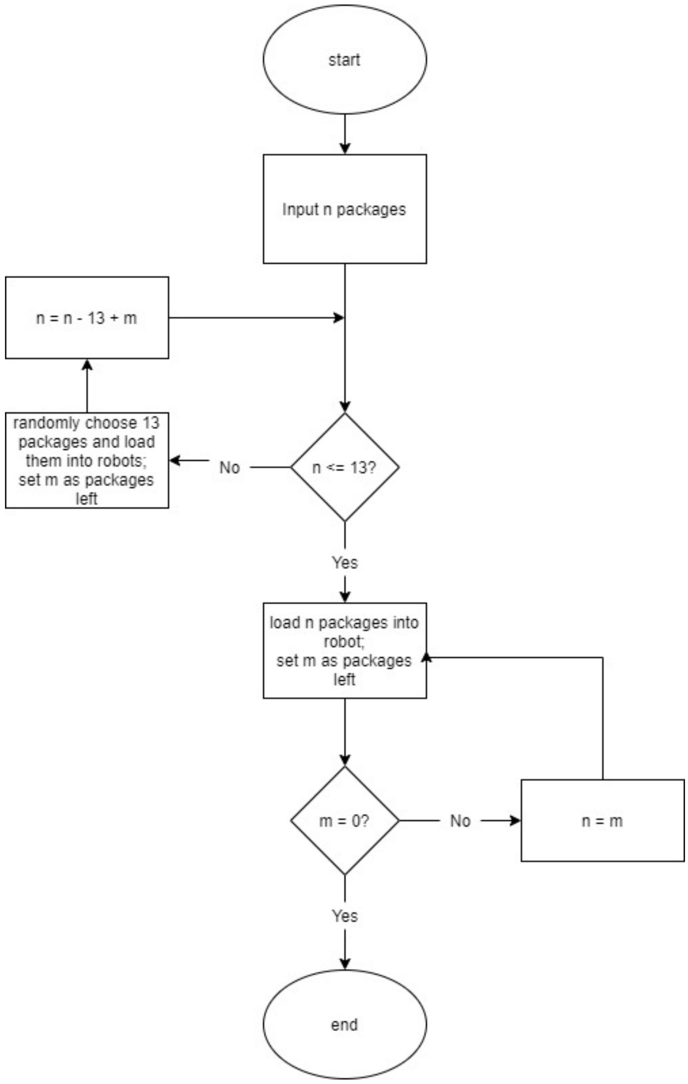

Figure 4.2: Flowchart of loading packages

Figure 4.2 describes the logic of loading packages to different destinations. The first step is to enter the overall one-year n package to one destination into the programme. The second step is to determine the size of the n. If n is smaller than 13, which is the number of packages over the average loading capacity of the robots, these n packages will be loaded into a robot and m is set as the packages left. In this situation, if m equals to 0, it means that all packages have been loaded and the programme ends; if m is greater than 0, the n packages will be updated to the left m packages and the programme returns to step 2. If n is larger than 13, 13 packages will be chosen randomly from n and will be loaded into a robot and m is set as the packages left. n is then updated to m plus $n - 13$, then go back to step 2. The programme will keep running until all the packages to certain destination are loaded into the robot. This programme will run for all destinations and package unit will be calculated accordingly.

4.3 Route planning problem

This step is to make models to analyze the route planning problem. The route planning problem is derived from the tsp problem which is a NP-Hard problem. The related literature is shown below and our model and result will be on displayed in Section 5.3.

4.3.1 Literature Review of Vehicle routing

Merrill M. Flood [23] indicated that the traveling salesman problem was posed in 1934 by Hassler Whitnery in a seminar talk at Princeton University and pointed out that there are no acceptable computational methods yet. Flood [23] also described the traveling salesman problem as finding a permutation $P = (1 i_1 i_2 i_3 \cdots i_n)$ of the integers from 1 through n that minimizes the quantity $a_{1i_2} + a_{i_2i_3} + a_{i_3i_4} + \cdots + a_{i_n1}$ and the problem can be solved by finding an efficient method for choosing a minimizing permutation because there are only $(n - l)!$ possibilities to consider. J. K. Lenstra and Kan, A H G Rinnooy [34] indicated that vehicle routing problems can be formulated as a travelling salesman problem (TSP) and the TSP can be the simplest way to solve it. Michael Hahsler and Kurt Hornik [26] introduced the R package TSP, which provides a basic structure for dealing with and solving travel salesperson problems that provided the S3 class to specify TSP and its solutions, and provided some heuristics to find a good solution. In addition, it also provided an interface with Concorde, which is one of the most accurate TSP solvers currently available. Alessandro Bertagnon and Marco Gavanelli [5]introduced the Euclidean TSP that each node is identified by its coordinates on the plane and the Euclidean distance is used as cost function. Pieter Leyman and Patrick De Causmaecker [35] discussed the intermittent traveling salesperson problem (ITSP) which is an extension of to the well-known traveling salesperson problem and explained that the difference was that each node required some processing time and the allowable consecutive processing time of a node is limited which results in the introduction of waiting time and/or multiple visits. They [35] gereralized the underlying model that determines the maximum consecutive node processing time by proposing a metaheruristic algorithm for this extended ITSP and performed computational experiments to allow for meaningful insights into each algorithm component's performance. Pandiri Venkatesh and Alok Singh [65] introduced the multiple traveling salesperson problem (MTSP) that is similar to the famous Travel Salesperson Problem (TSP), except that more than one salesperson visits the city, although only one salesperson must visit each city once. They [65] proposed two meta-heuristic approaches for the MTSP of which the first one was based on artificial bee colony algorithm, while the second one was based on invasive weed optimization algorithm and applied a local search to further improve the solution obtained through their approaches.

Chetan Chauhan et al. [11] introduced the various methods/techniques available to solve traveling salesman problem including Branch and Bound, the Cutting Plane, Branch and Cut, Dynamic Programming, Brute-force method to achieve exact solutions and Christofides' Algorithm, Clarke-Wright Algorithm, NearestNeighbour Algorithm, Insertion Heuristics, the Greedy Heuristic, Gutin and Yeo Algorithm, Hill Climbing (HC), Lin-Kernighan Algorithm, the Metropolis Algorithm, Simulated Annealing (SA) Algorithm, Tabu Search (TS), Ant Colony Optimization (ACO), and Genetic Algorithms (GAs) Algorithm. Chryssi Malandraki and Robert B. Dial [39] introduced how to solve the traveling salesman problem by using dynamic programming and indicated that dynamic programming can solve only very small problems because of the fact that the precise dynamic programming algorithm of TSP has exponential storage and calculation time requirements. They also presented a restricted DP heuristic to improve the situation. Paul Bouman et al. [?] presented the dynamic programming approaches based on Bellman-Held-Karp dynamic programming algorithm to solve TSP and stated that their approach can solve larger problems. Thomas Stutzle and Holger Hoos [63] introduced the Ant System as a cooperative search algorithm inspired by the behavior of real ants and indicated that it could be applied to the solution of combinatorial optimization problems. They also introduced an improved version of basic Ant System called MAX-MIN Ant System in order to solve the traveling salesman problem. Ismail Ellabib et al. [20] introduced the Vehicle Routing Problem with Time Windows and presented a model of an Ant Colony System to solve this problem. Leonora Bianchi et al. [6] proposed a probabilistic ant colony system (pACS) which was an ant based priori tour construction heuristic that was derived from the similar heuristic algorithm ACS that was previously designed for the TSP problem. They also showed that pACS could find better solutions for a wide range of homogeneous customer probabilities and ACS was better than pACS for high customers probabilities. Marcin L. Pilat and Tony White [47] proposed to add Genetic Algorithm to Ant Colony System (ACS) to improve performance by two modifications; of which the first one was to combine ACS and Genetic Algorithm that encodes experimental variables in ants while the second one was to use a Genetic Algorithm to evolve experimental variable values used in ACS. Grefenstette et al. [25] introduced the Genetic Algorithm and how it can be used to solve TSP problems. Fei Liu and Guangzhou Zeng [37] introduced an improved genetic algorithm with reinforcement mutation called RMGA to solve TSP problems and got the best tour in a reasonable time. The main ideas were using heterogeneous pair selection instead of random pair selection in edge assembly crossover and structuring reinforcement mutation operator by modifying the Q-learning algorithm and applying it to those individual generated from modified edge assembly crossover. Noraini Mohd Razali and John Geraghty [52] indicated that the performance of genetic algorithm can be improved by modifying the genetic operators including parent selection, crossover and mutation and showed that tournament selection strategy performed better than proportional roulette wheel and rank-based roulette wheel selections because it achieved the best solution with low computing times. They also detected that tournament and proportional roulette wheel could be better than the rank-based roulette wheel selection for smaller problems.

5 Results and Discussion

5.1 Regression Model

5.1.1 Dependent variables

Based on the primary data collected from the Chalmers Transportation Centre, the distributions of the three dimensional variables along with the Volume variable were studied first to choose the suitable regression models for application. Since histogram is a simple and versatile way to show the frequency distribution of a data set, it is applied here first in order to get an intuitive insight of the distributions of the above 4 variables:

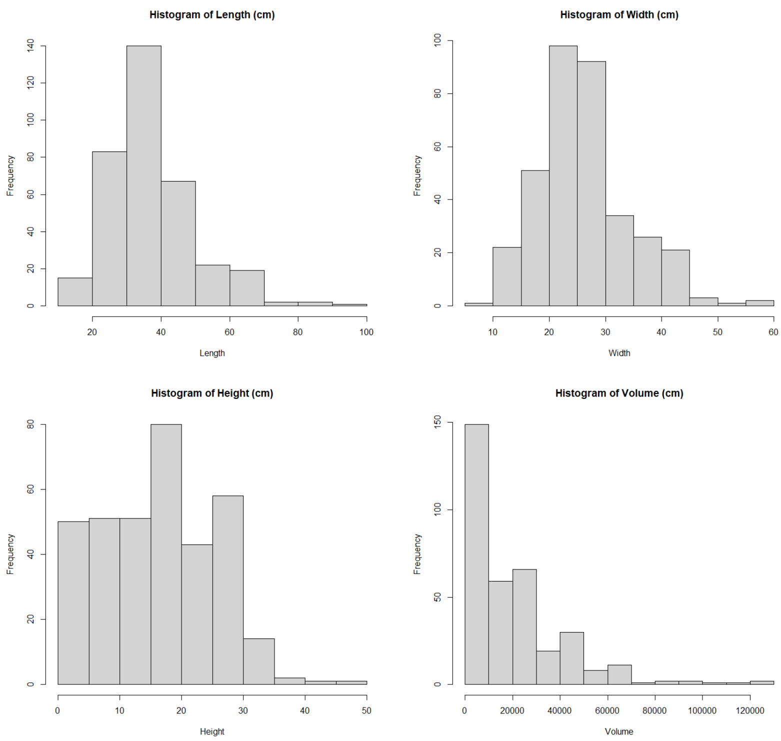

Figure 5.1: The histograms of the dimension variables

It can be directly addressed from the figure that most of the distributions of the four random variables are not normally distributed. To be more specifically, those variables have some asymmetric properties in the probability distributions. Such property can be measured by Skewness, which has values of zero, positive or negative [58].

Apart from other three variables, the distribution of height shows a characteristic of stabilization, which indicates the mass of the distribution is a constant. Data with constant mass function can not be modeled by linear regression models. On the other hand, the Height can be calculated directly by:

$$H = V/(W \times L) \tag{5.1}$$

(H is the height of the package, V denotes the volume and W denotes the width.) Accordingly, Height is removed from the modeling and volume is predicted instead.

The table below shows the statistic descriptions of the three random variables:

Table 5.1: Statistic Descriptions of Length,Width and Volume

Variable	n_Obs	Mean	SD	Median	MAD	Min	Max	Skewness	Kurtosis	percentage_Missing
Length	350	37.46	12.35	35	8.90	16	83	0.99	1.04	0
Width	350	26.64	8.27	26	6.67	8.50	60	0.71	0.94	0
Volume	350	20, 272.04	20, 778.53	13, 699.69	14, 764.75	376.25	127, 906.00	2.00	5.75	0

The negative skew indicates that the mass of the distribution is concentrated on the right of the figure. The distribution is said to be left-skewed, left-tailed, or skewed to the left. While the positive skew, on the contrary, indicates the data whose right tail is longer. The mass of this right-tailed distribution is concentrated on the left of the figure. Data with zero skewness have a property of symmetry, and generally its mass distribution is normal distribution.

The skewness of *Length*, *Width* and *Volume* are 0.986, 0.713, 2.002 respectively. The positive values of skewness indicate that these variables are all skewed to the right. In other words, the probability distributions of the three dependent variables are non-normal distribution. General linear model can not be applied accordingly. Generalize linear Model, instead, was applied due to its flexibility in the assumptions of the probability distribution of the dependent variables.

As previous described, for continues variables with positive skewness, Gamma GLM model is suitable to be used for data fitting. The main feature of a generalized Gamma distribution looks like this [43]:

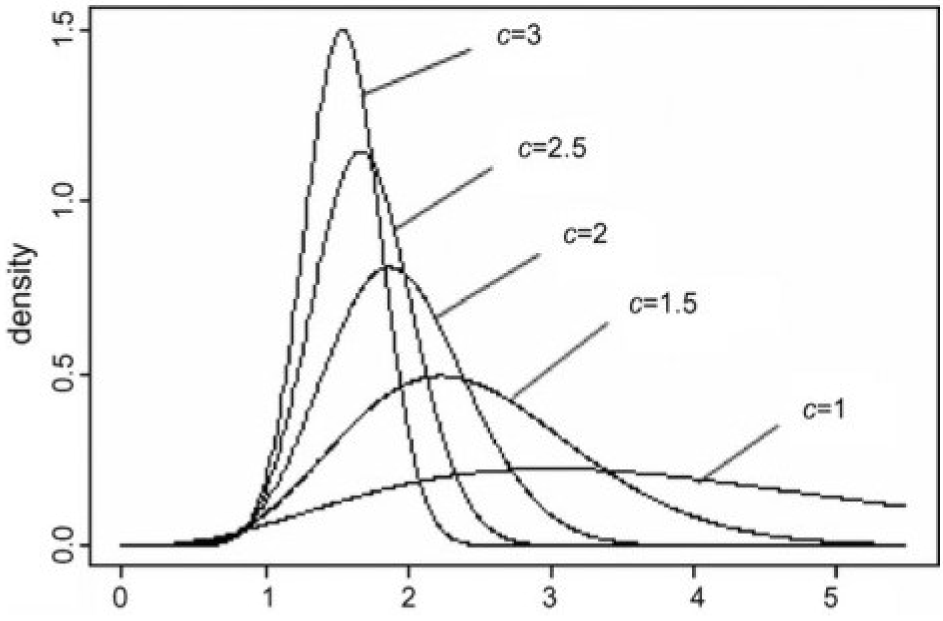

Figure 5.2: Generized Gamma distribution with different parameters c = 1, 1.5 ,2 ,2.5 , 3 from right to left

c is the shape parameter. With different values of c, the generalized Gamma distribution can fit a variety of right-skewed data. The link function chosen in the study are log link and identity link. The log link function can explain the changing rate of dependent variables with one unit change of the independent variables and the identity link can explain how much the dependent variables change according to a unit change from the independent term. In this case, since independent variables are binary variables (The independent variables will be elaborated in the next section), the model can be used to explain how the existence of an independent variable affects the rate of change of the dependent variables or the change in value of the dependent variables.

5.1.2 Independent variables

As three models are needed for the prediction of dependent variables, independent variables should be specified into three groups: *Volume*, *Length* and *Width*. The data collected from the Transportation Centre has the bellowing structure:

Table 5.2: Five records of data collected from Chalmers Transportation Centre

destination	length	width	height	volume	commodity	type
Fysik	20.50	15	14	$4,305$	Unknown	0
Stena center	27	20	1	540	Electronic product	0
Stena center	31	24.50	1	759.50	Machinary	0
Stena center	43	30	1	$1,290$	Electronic product	0
Stena center	33.50	23	10.50	$8,090.25$	Laboratory equipment	0
Stena center	24	18.50	8.50	$3,774$	Laboratory equipment	0

Week is excluded from the model since the period for collecting data is only one week, the sample is not sufficient to accept week as an independent variable. Weight is also excluded even if it could be a significant term in the model. The main reason is that the secondary data of one year does not include weight. Type is a binary variable with 0 representing the package is EXP and 1 representing it is PKG. Destination is a categorical variables with 13 values: Kemi, Fysik, E-huset, Maskin,SB3, ACE, Vasa 2-3, Biblioteket, Stena center,Chalmers verksamhetsstöd, Postrum SB3, Akademiskahus and Vasa 7. And commodity is also a categorical variable with 13 values: Laboratory equipment, IT infrastructure, Electronic product, Machinary, Tools and supplies, Books, Clothes, Daily necessities, Rubber products, Food, Games, Medical supply and Unknown. The detailed information about these two categorical variables can be found in the table below:

The categorical variables can not be used directly in regression models, thus these variables need to be encoded by some tactics. One way to encoding them is to use *Label Encoding* directly. It consists of substituting each group with a corresponding number and keeping such numbering consistent throughout the feature. However, by applying this method, the relationships among each number remains, which means different categories holding different distances with each other. In this case, such characteristics should not exist between destinations or commodity types. As a result, another encoding method, *Dummy Encoding*, is used to generate the explanatory variables required by the model. Such encoding method can eliminate the problem raised by *Label Encoding*. In *Dummy Encoding*, the additional features are created based on the number of unique values in the categorical feature. The overall additional number of features equal to that unique value minus 1. The table below shows the encoding result of the Commodities with the "Unknown" feature excluded from the model:

Value zero in the commodity column denotes that the package does not belong to that commodity type and one on the contrary. Type *Unknown* was removed from

Table 5.3: Counting of categorical variables of primary data

Destination	Count	Commodity	Count
Kemi	87	Unknown	114
Fysik	75	Laboratory equipment	72
E-huset	45	IT infrastructure	47
Maskin	44	Electronic product	33
SB3	38	Machinary	32
ACE	24	Tools and supplies	20
Vasa 2-3	16	Books	15
Biblioteket	7	Clothes	8
Stena center	7	Daily necessities	3
Chalmers verksamhetsstöd	4	Rubber products	3
Postrum SB3	2	Food	2
Akademiska hus	1	Games	1
Vasa 7	1	Medical supply	1

Table 5.4: Five records for the encoding of Commodity using *Dummy Encoding*

Clothes	Daily necessities	Electronic product	Food	Games	IT infrastructure	Laboratory equipment	Machinary	Medical supply	Rubber products	Tools and supplies
0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0
0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0

the model because it can be expressed as $(0,0,0,0,0,0,0,0,0,0,0,0)$. The sequence means the package does not belong to any of the other 12 types, thus can only be classified as *Unknown*. The variables that are removed from the encoded matrix are called *Base Variables*. For commodity types, *Unknown* is chosen as the base variable because of its dominance in quantities and information containing of errors. Some packages' type can not be obtained by observation or inferred from their senders were recorded as *Unknown*. Accordingly, *Unknown* contains errors of the model and can not be used as the explanation term. The base variable chosen for destinations is based on the similar idea. Among all 27 destinations in Chalmers, only 13 were observed in one week. Thus to deal with the one year dataset, the unobserved destinations have to be classified as one destination in the observed dataset. In this case, *Kemi* was chosen because of its large transportation flow. Deliveries sent to *Kemi*, due to the large quantities, contains less information that can be used to predict the size of the package. Similar to *Unknown*, it is chosen as the base variable of destinations.

After encoding the categorical variables, the explanation terms library is expanded to 25 terms: 12 destinations, 12 commodities, and 1 package type. As discussed above, three regression models are needed for the prediction of *Volumes*, *Length* and *Width* respectively. *Volume* is modeled first. In addition to the 25 explanation terms, the interaction terms are also studied in the regression model. These interaction items include the result of multiplying a column of commodity type and a column of destination, and the result of multiplying a commodity type, a destination, and a package type. The combination of commodity type and destinations produces 144 interaction items, which are also combined with the package type to form a new ternary interaction item. The total explanatory variable library of the model is expanded to 313 variables.

The values of these interaction terms are also very easy to interpret the model. Since the commodity type and destinations are both binary variables, their interaction item is also a binary variable. When the value is one, it represents a specific commodity type of a delivery to a certain destination and zero on the contrary. As for the three terms interaction, one represents not only the certain commodity type of the delivery is going to a certain building, but also has the package type **PKG**.

The model building for *Volume* uses all 313 terms in the term library. While the regression model for length also uses the same library. Unlike the other two models, the prediction for the width adds the length to the term library.

5.1.3 Models and Results

5.1.3.1 Volume

The Gamma regression model built for *Volume* taken the 313 terms as the independent variables and step-wise regression method was used to eliminate the multicollinearity and regression terms with insignificant coefficients. The link function chosen for *Volume* is the log link. The reason for choosing this link function is because *Volume* has an obvious right skew feature with value 2.00, and the value range is relatively large. The table below shows the final model for *Volume*:

Table 5.5: Gamma Regression Model for Volume with log link function

Term	**Coefficient**	**P-value**
(Intercept)	7.71	0.00
Type		
PKG	1.99	0.00
Destination		
Akademiska hus	2.40	0.01
Biblioteket	-0.81	0.03
E-huset	-3.84	0.00
Maskin	0.42	0.02
Postrum SB3	3.10	0.00
Vasa 2-3	-0.74	0.01
Vasa 7	-2.32	0.02
Commodity		
Clothes	0.89	0.02
Electronic product	1.60	0.00
IT infrastructure	1.06	0.00
Laboratory equipment	1.45	0.00
Interactions (Commodity : Destination)		
Electronic product : E-huset	3.10	0.00
Electronic product : SB3	4.48	0.00
IT infrastructure : Fysik	-1.37	0.01
IT infrastructure : SB3	-0.98	0.02
Machinary : Akademiska hus	-2.94	0.03
Interactions (Type : Commodity|Type : Destination)		
PKG : Electronic product	-4.73	0.00
PKG : E-huset	3.52	0.00
PKG : Fysik	0.44	0.00
PKG : Laboratory equipment	-1.02	0.03
PKG : Electronic product : Fysik	3.36	0.00

The terms with P-value less than 0.05 are chosen as significant terms. As discussed above, Type, Destination, Commodity and interaction terms are all binary variables.

In order to explain the model, the coefficient indicates the degree of influence of the explanatory term on the rate of change of the Volume value. The larger the value, the greater the degree of influence, and vice versa. For a separate explanatory term, it indicates that the how the Type, Destination, and Commodity type affect the value of the volume. For the interaction item, because the model only extracts the interaction of two items, it shows the influence of any two combinations of the above three independent items on the volume value. This model will be used to predict the volume of the one year dataset, and finally calculate the value of height.

In order to evaluate the goodness of fit for the gamma regression model, the Root Mean Square Error (Rmse) of *Volume* was calculated and the scatter plot of predicted and true values is shown below:

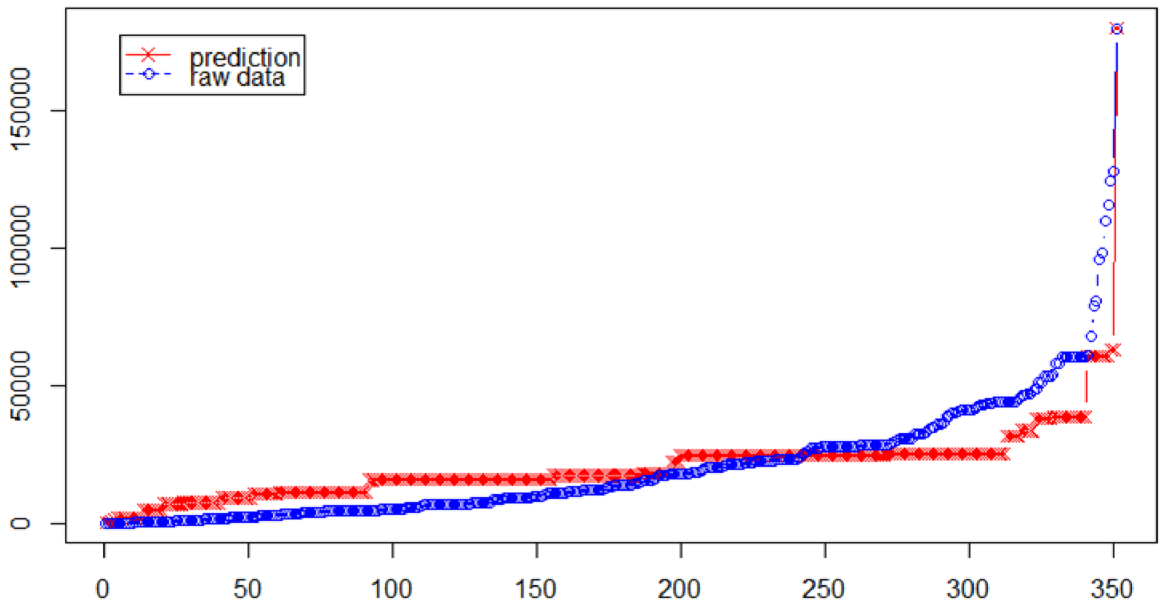

Figure 5.3: Predicting Volume VS. Volumes from raw data

The x-axis of the scatter plot shows the index value of the sample, the blue line represents the sample data, which has been sorted according to the value from small to large, and the red line represents the value predicted by the model. From the plot, the model has a good fitting effect on the data. The Rmse of the Gamma Regression Model for *Volume* is $0.018m^3$, while the standard error of *Volume* from the collected dataset is $0.021m^3$. This indicates that the prediction result of the model is more accurate than just guessing a value of *Volume*.

5.1.3.2 Length

Two Gamma regression models for length using identity link and log link respectively were built and evaluated. For each model, the independent variables were chosen by the stepwise regression terminology as introduced above. In the study, Akaike Information Criterion (AIC) was used to simplify the model. AIC is an estimator of relative quality of different statistical models introduced by Hirotugu Akike in 1974, [2]. The stepwise regression terminoloy applied in this research took AIC as the criterial for simplify the model, that is, during the stepwise process, only models having lower AIC were kept step by step. Rmse of the two models were calculated as above and the scatter plot were plotted and shown below:

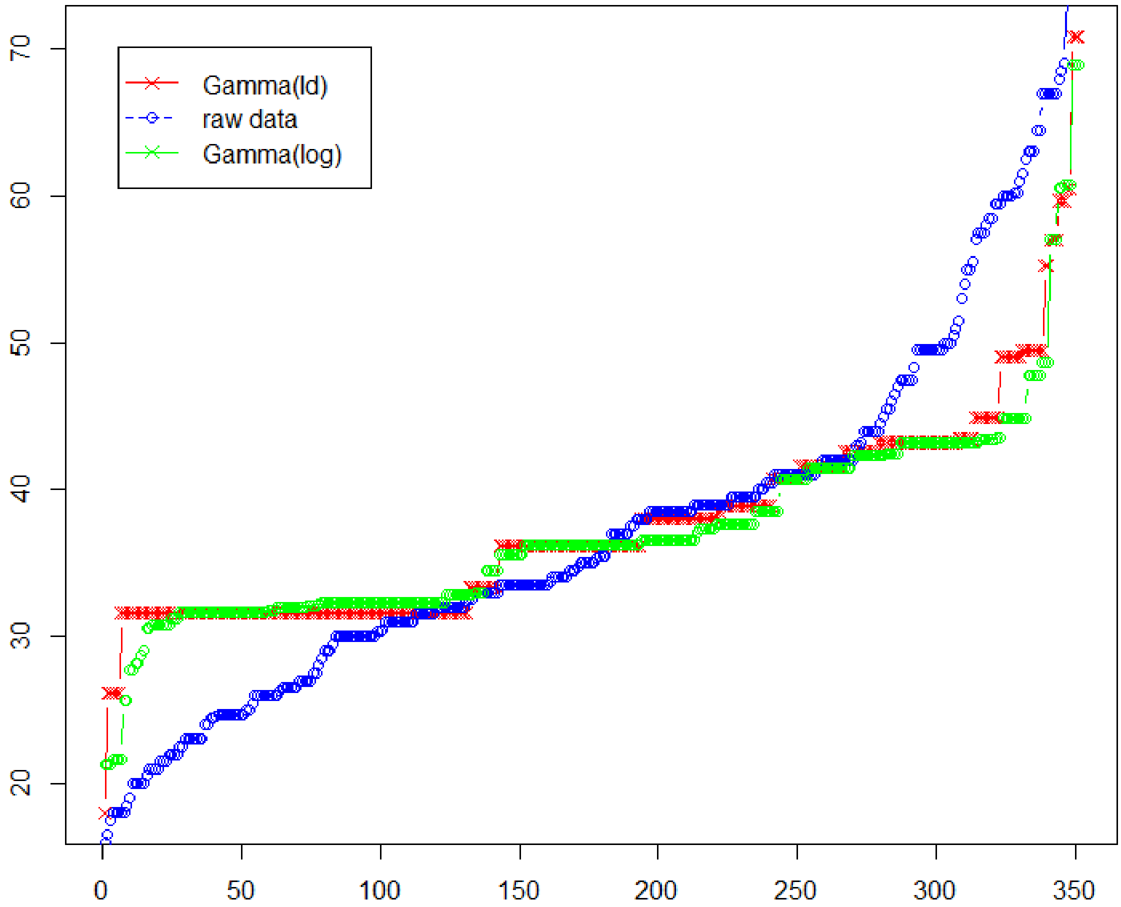

Figure 5.4: Comparison for modeling Length

The prediction results of the two regression model are similar from the scatter plot. Rmse of the model with identity link function is 10.40 cm and that of log link is 10.70 cm. The standard deviation of length is 12.70 cm, thus both model have certain predictive effect. The Gamma Regression Model with identity link function was chosen for length because of its lower Rmse. The table below shows the result of the model:

Table 5.6: Gamma Regression Model for Length with identity link function

Term	Coefficient	P-value
(Intercept)	31.63	0.00
Destination		
Chalmers verksamhetsstad	39.20	0.00
Maskin	7.24	0.00
Postrum SB3	23.62	0.03
SB3	13.24	0.00
Commodity		
Clothes	10.63	0.04
Daily Necessities	21.67	0.02
IT infrastructure	17.37	0.00
Laboratory Equipment	4.56	0.00
Machinary	-5.53	0.02
Interactions (Commodity : Destination)		
Books : Chalmers verksamhetsstad	-32.33	0.04
Electronic product : Maskin	-20.87	0.00
IT infrastructure : Fysik	-14.63	0.02
IT infrastructure : SB3	-19.03	0.00
Laboratory Equipment : ACE	20.81	0.02
Machinary : Fysik	9.10	0.03
Tools and Supplies : ACE	28.87	0.01
Interactions (Type : Commodity|Type : Destination)		
Fysik : PKG	6.36	0.00

As discussed above, significant terms such as destinations, commodity types and interaction terms were extracted from the original model. The model indicates that for some destinations and commodity types, they do have an influence on the size (length) of the package. For example, the Laboratory equipment going to Fysik building is often large and thus have a higher coefficient.

5.1.3.3 Width

Similar to length, the modeling for width has also considered the two different link function: log and identity. Rmse was calculated and the comparison scatter plot is shown below:

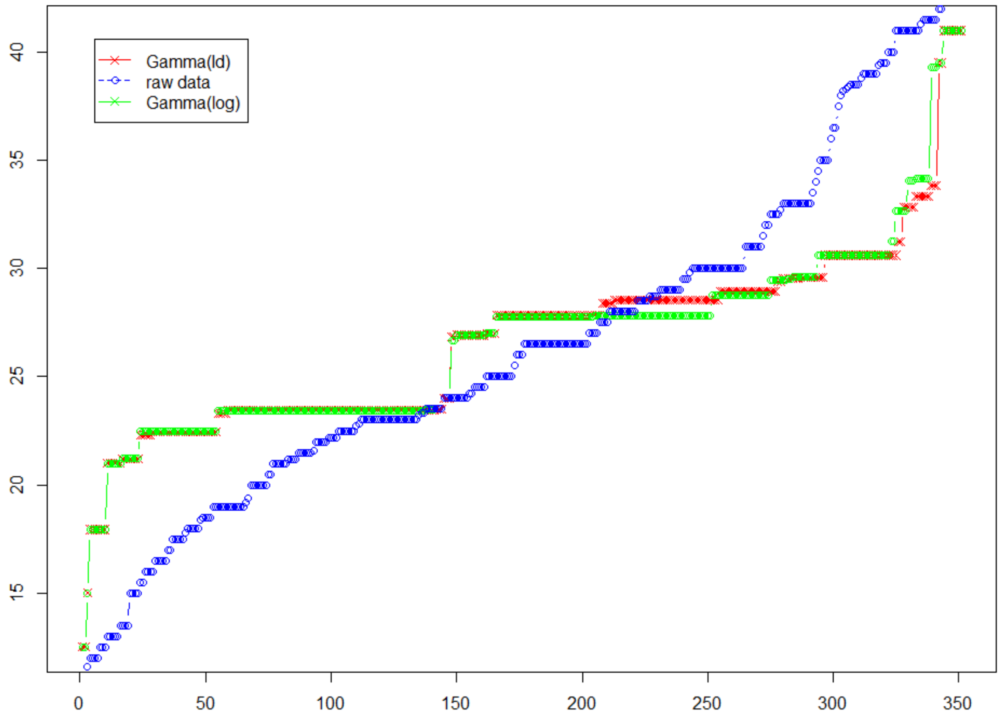

Figure 5.5: Comparison for modeling Width

The prediction results of the two regression model are similar as that of length, two lines are relatively close in many places. The Rmse of the Gamma Regression model for width with identity link function is 7.03 cm, while the Rmse of the log link is 6.95 cm. The standard deviation of *Width* is 8.28 cm. As previously discussed, the model with lower Rmse is chosen, which is the Gamma Regression Model with log link function.The chosen model is shown below:

Table 5.7: Gamma Regression Model for Width with log link function

Term	Coefficient	P-value
(Intercept)	3.15	0.00
Destination		
E-huset	-0.83	0.02
Maskin	0.21	0.00
SB3	0.23	0.02
Commodity		
Clothes	0.35	0.00
Daily necessities	0.35	0.03
Electronic product	0.73	0.03
IT infrastructure	0.37	0.00
Laboratory equipment	-0.27	0.01
Interactions (Commodity : Destination)		
Electronic product : Maskin	-0.84	0.00
IT infrastructure : Fysik	-0.31	0.02
IT infrastructure : SB3	-0.34	0.02
Laboratory equipment : Fysik	0.72	0.03
Machinary : ACE	-0.63	0.02
Tools and Supplies:ACE	0.52	0.00
Tools and Supplies:Maskin	-0.32	0.01
Interactions (Type:Commodity|Type:Destination)		
PKG : Electronic product	-0.73	0.03
PKG : E-huset	0.79	0.02
PKG : Fysik	0.62	0.02
PKG : Laboratory equipment	0.44	0.00
Interactions (Type:Commodity:Destination)		
PKG:Laboratory equipment:Fysik	-0.92	0.01

5.1.3.4 Height

Height is calculated by the prediction of *Volume*, *Length* and *Width* as discussed above. The model chosen for *Volume* is the Gamma Regression model with log link function, the model chosen for *Length* is the Gamma Regression model with identity link function, and the model chosen for *Width* is the Gamma Regression model with log link function.

5.1.3.5 One year prediction

Since all three model has been built and selected, the prediction of the one year dimension data was conducted. The packages of type REK and INT are personal or internal letters, they tend to have little effect on loading, not only because they have no height (or small height), but also because of their softness. In order to simplify the model, those packages are assumed as standard A4 size: 29.7 cm in length, 21 cm in width. The height was assumed as 1 cm in order to keep the packages as the three dimension objects. PALL will not go to robots and thus were not been predicted. FRYS has a standard size: 41 cm in length, 32.5 cm in width and 31.5 cm in height. PKG and Exp were predicted using the model. The figure below shows the modeling result of the one year dataset:

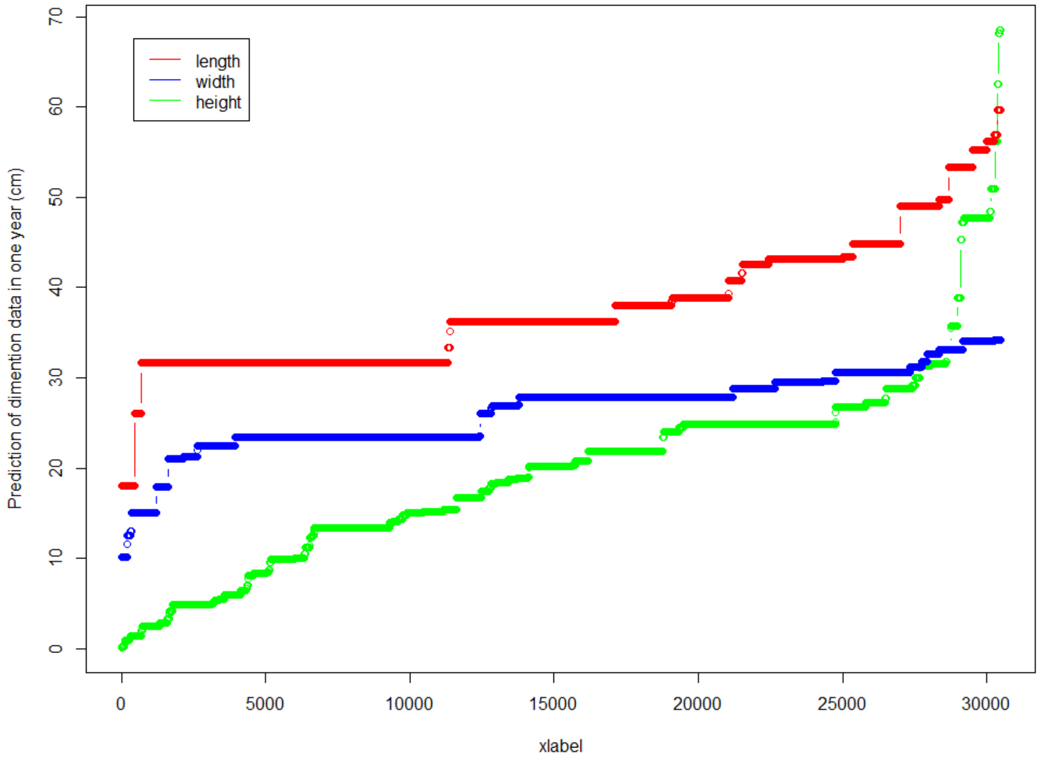

Figure 5.6: Dimension prediction for one year dataset

It can be easily observed from the figure that the prediction of the dimensions for the one year dataset is between 0 cm and 60 cm, which means for any package, the volume of the robot is enough to fit it. The dimension data predicted for one year dataset was used in the later model building and analysis.

5.2 Package loading

5.2.1 Selection of destinations

For destinations with a relatively large number of packages delivered per day, due to the collection of more sample data, the data fitting to their package size is relatively accurate, so their respective package units are calculated separately. However, for destinations with a small average daily package delivery volume, due to the lack of sample data, the fitting error of the size of their packages is large, so the package unit calculation for these destinations is based on the total number of packages in these destination, which means those destinations have a unified package unit.

The average daily package delivery volume is calculated from the one year dataset. For each destination, the total number of packages was divided by 248 (recorded work days in one year). The destinations with average daily packages over one were selected to calculate their separate package unit while other destinations with packages less one shared the same package unit. The table below shows the destinations and their average number of packages:

Table 5.8: Daily average package for different destinations

Destination	Daily Average Package
ACE	6.45
Biblioteket	4.34
Chalmers verksamhetsstöd	1.38
E-huset	10.42
Fysik	19.48
Kemi	39.96
Maskin	13.07
Mattematiska Vetenskaper	1.20
SB3	21.35
Vasa 2-3	2.74
Others	2.98

5.2.2 Knapsack loading

The packages to these destinations were loaded by the simulated annealing algorithm using the process mentioned in section 4.2. Thirteen packages were randomly selected and loaded until all the packages were completely loaded or the remaining packages could not be loaded anymore. The figure below shows 9 examples of packages being loaded to Kemi:

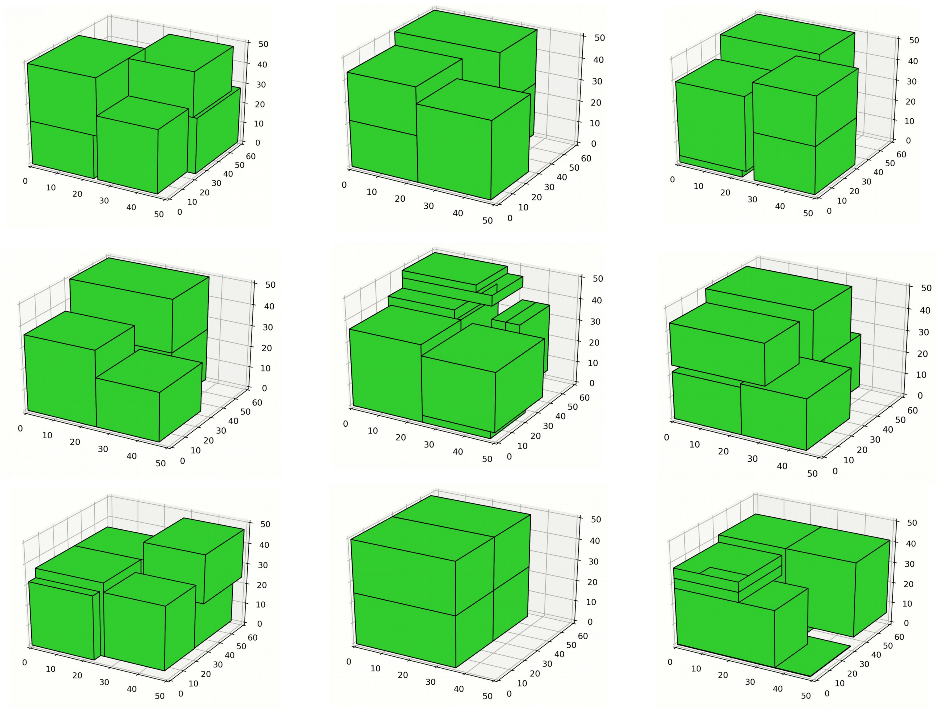

Figure 5.7: Nine examples of packages loaded to Kemi

Of all destinations with daily average packages over one, Chalmers verksamhetsstöd has most of its packages with a length over 60 cm, and thus was excluded from the model (it exceed the maximum height of the robot). Most of the packages to other destinations can be loaded to the robots, and the result of their average loading quantity as well as the package unit was shown in the table below:

Table 5.9: Package Unit for different destinations

Destination	Average loading quantity	Package Unit
ACE	7.0	1.40
Biblioteket	11.1	2.22
E-huset	8.2	1.64
Fysik	5.8	1.16
Kemi	5.4	1.08
Maskin	7.4	1.48
Mattematiska Vetenskaper	8.4	1.68
SB3	5.0	1.00
Vasa 2-3	10.0	2.00
Others	7.2	1.44

The number in the package unit column indicates how many packages are counted as a package unit for each destination, for example, 1.4 for ACE means that 1.4 packages to the ACE building represents a package unit. The standard volume of a package unit is determined by the minimum average number of loads, and the purpose of doing so is to enable the packages of other destinations to be loaded in a way to ensure that big packages will not be separated when loading. In this study, 1/5 of the total robot volume as a package unit obtained from Kemi building is used.

The capacity of the robot was determined also by the destination with the smallest average loading quantity, which is 5 package units.

5.3 Analysis of routing

5.3.1 Assumptions and notation

This section introduces a mixed-integer linear programming (MILP) model created to minimize the energy consumption of using autonomous robots for delivery in Chalmers campus. The robots load packages from the Chalmers Transportation Center and delivery these packages to assigned destinations and return to transportation center after delivering all the packages.

The goal of building the model is to minimize the total energy consumption of operation robots everyday. In order to solve the model, the capacity of robots, distance between each destinations including the transportation center, numbers of packages, and the average consumption per unit distance that was calculated by the collected data are considered. Some assumptions in the following should be made to optimize the model:

- A set of packages were packed together into an unit package. The volume of package unit transported by the robots to different destinations maintains constant.

- The number of packages loaded in an unit package transported to different destinations changes with changes in demand of each destination. The number of packages loaded in an unit package is determined by the knapsack problem mentioned early.

- Each robot has enough electricity to complete the daily task.

- Each tour starts and ends at the same destination (the transportation center).

- The demand of packages that will be transported to each destination is a known quantity.

- The weight of packages is not considered so it will not affect the energy consumption.

- No time window is considered in this research.

Some facts are also considered in the model:

- The energy consumption per unit distance traveled by the robot and the capacity of the robot is known.

- The serial number of each destination and the location of each destination is known; hence the distances between each destinations is known too.

The explanation of sets, parameters, and decision variables are are summarized in the following:

Sets	
$d, h \in G$	Set of destinations/graph nodes
$p \in P$	Set of packages
$k \in K$	Set of tours
Parameters	
C	The transport cost per distance unit
l_{dh}	The distance between two destinations/two nodes
β_{dp}	1, if the package p belongs to destination d; 0, otherwise
m	The capacity of the delivery robot
N	Number of destinations
Decision variables	
α_{dhk}	1, if tour k leads from location d to location h; 0, otherwise
γ_{pk}	1, if tour k contains package p; 0, otherwise
y_{dk}	1, if tour k contains location d; 0, otherwise
u_{dk}	Relative node visiting order

5.3.2 Mathematical model

$$min\ W = \sum_h \sum_d \sum_k l_{dh} \cdot C \cdot \alpha_{dhk} \tag{1}$$

$$\sum_k \gamma_{pk} = 1 \tag{2}$$

$$\sum_p \gamma_{pk} \leq m \tag{3}$$

$$y_{dk} - \gamma_{pk} \geq \beta_{pd} - 1 \tag{4}$$

$$\sum_h \alpha_{dhk} = y_{dk} \qquad d \neq 0, d \neq N-1 \tag{5}$$

$$\sum_h \alpha_{0hk} = 1 \tag{6}$$

$$\sum_h \alpha_{(N-1)hk} = 0 \tag{7}$$

$$\sum_d \alpha_{dhk} = y_{hk} \qquad h \neq 0, h \neq N-1 \tag{8}$$

$$\sum_d \alpha_{d0k} = 0 \tag{9}$$

$$\sum_d \alpha_{d(N-1)k} = 1 \tag{10}$$

$$u_{dk} - u_{hk} + N * \alpha_{dhk} \leq N-1 \tag{11}$$

$$\alpha_{ddk} = 0 \tag{12}$$

$$u_{dk} \geq 0 \tag{13}$$

$$\alpha_{dhk}, \gamma_{pk}, y_{dk} \in \{0, 1\} \tag{14}$$

The objective function (1) minimizes the total transportation cost per day. The total costs were represented in terms of the total energy consumed everyday when the robots operated and were calculated by multiplying the total distance traveled by the robots per day and the electric energy consumed per distance unit.

Constraint (2) ensures every package will be transported once and will be transported only once. Constraint (3) ensures that the volume of packages transported each tour is less than or equal to the capacity of the robot. Constraint (4) ensures that if the package is delivered in a certain tour, it will go to its destination in that tour. Constraint (5) ensures that at most one route departs from point d (except for the start and end nodes). Constraint (6) ensures that the tour must start from the destination zero (CTC) and constraint (7) ensures that no tour starts from the end point $N-1$. Constraint (8) ensures that destination h can only be visited zero time or once (except for the start and end nodes). Constraint (9) ensures that the start node can not be visited by other nodes and constraint (10) ensures that the end node must be visited. Constraint (11) is the Miller-Tucker-Zemlin(MTZ) formulation used to solve the sub-loop problem[72]. Constraint (12) ensures that each robot would not depart and arrive at the same destination.

5.3.3 Case Study

A case study was conducted to analyze the model we built. In the following case study, the dataset of packages on the 12th of February, 2021 was chosen from the one week data we collected. The table 5.10 showed the dataset used in the case study. A total number of 31 packages would be transported to 5 different destinations which are ACE, Biblioteket, E-huset, Fysik, Kemi, Maskin, SB3 and Vasa 2-3. Among all the packages, 3 packages would be transported to ACE, 1 would be transported to Biblioteket, 1 would be transported to E-huset, 14 would be transported to Fysik, and 12 would be transported to Kemi. These packages are transferred to standard number of package units in the following table:

Table 5.10: Dataset used for Case Study

Destination	Destination code	Number of packages	Package Unit	Number of Package Units
ACE	1	3	1.40	3
Biblioteket	3	1	2.22	1
E-huset	7	1	1.64	1
Fysik	8	14	1.16	13
Kemi	12	12	1.08	12

The packages units were than turned into the package ids from 0 to 29.

LINDO API 11.0 was used to compute the result. LINDO API with the interfaces to Python is used in this case study. The default MIP solver is the branch-and-cut method which is an iterative method that uses linear or non-linear solvers as sub-solvers, depending on the nature of the problem.

In this case, 6 routes were used in order to deliver the total packages, the detail of the routes are showed in the following figures:

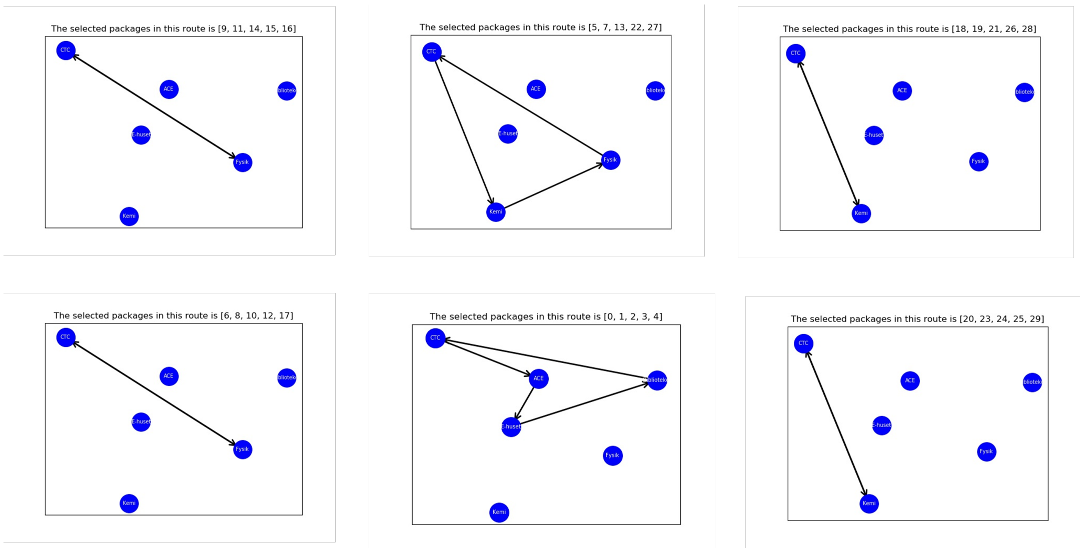

Figure 5.8: All routes for package units going to their destinations

The double arrow means the route start from CTC, going to the destination, and then return to CTC. The optimized minimum distance is 4427.43 meters. As the energy consumption per distance is 0.01475 Wh/m, the total energy consumption on that day is 65.30 Wh. As the battery capacity of one robot is 731 Wh and its maximum driving distance is 20 Km, one robot is enough for the daily task of delivery. Since time window is not considered in this case study, Chalmers transportation center can use the same robot to transport all the packages for all 6 routes.

6

Conclusion

6.1 Conclusion

This thesis focuses on the study of the efficient planning of automated delivery robot and the evaluation of the effectiveness of the automated delivery robots considering energy consumption. The automated delivery robots using lithium batteries as the energy source are planned to be used in Chalmers University of Technology campus to replace the the traditional transportation modes to delivery packages to different buildings. In our research, the package loading problem and the transportation route planning problem should be considered to solve the efficient planning. The energy consumption issue was then analyzed according to the predicted routing and electricity consumption per unit distance of the automated delivery robot.

The dimensions of packages are indispensable conditions for analyzing the package loading problem. In fact, the one year data provided by Chalmers transportation center lacks the dimensional data of packages. It is also impossible for us to collect these in a sufficiently long time interval like one year. Therefore, the fit of dimensional data using regression model was conducted in this research. The regression model fitted the length, width and height of each package in the one year data. The table A.1 contains ten packages with these data is shown in the Appendix below.

The package loading problem was also analyzed and the results of package unit which is an important element in the routing planning were given. The package unit to the most visited destinations are shown on the table 5.9.

The energy consumption based on the data of 12th of February and the predicted routes were computed from the case study. The energy consumption we predicted on the 12th of February was 65.30 Wh.

Based on the results of Section 4.1 and Section 4.2, we conclude that the use of automated robots is feasible, because most packages are within the size of the robot, and one robot can load five to eleven packages.

Based on the results of Section 4.3, we conclude that the use of robots for package delivery is very efficient. Only one robot can meet daily delivery requirement. The overall energy consumption is less than 50% of the total power of a robot.

6.2 Limitations and Future Work

All analyses in this thesis were conducted in an ideal situation. The one year data provided by Chalmers transportation center did not contain the dimensions of each package so we have to fit these data. If the actual dimensional data was given, the constraints of dimensions should be included in the MIP model. Time window is not considered in this thesis either; however, the time limit should always be considered in practical situations. More research may be conduced in the future to include the time window in the MIP model.

The distance between every destinations were calculated by the computer software QGIS. We did not measure the distance between each location on the spot. We also did not consider possible unexpected situations, such as detours caused by road repairs.

We only consider the consumption in operation while robots consume power when they stop. The energy consumption in this thesis was computed by multiplying the energy consumption of the robot per unit distance by the distance they traveled. However, the situation may be different every day, so a better method is needed to measure the actual energy consumption of the robot.

Bibliography

[1] A. Buchegger, K. Lassnig, S. Loigge, C. Mühlbacher and G. Steinbauer, *"An Autonomous Vehicle for Parcel Delivery in Urban Areas,"* 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 2018, pp. 2961-2967, doi: 10.1109/ITSC.2018.8569339.

[2] Akaike, H. (1974). *"A new look at the statistical model identification."* IEEE Transactions on Automatic Control, 19(6), 716–723. doi:10.1109/tac.1974.1100705

[3] Alcácer V, Cruz-Machado V. Scanning the industry 4.0: A literature review on technologies for manufacturing systems[J]. Engineering science and technology, an international journal, 2019, 22(3): 899-919.

[4] Andy M. Ham. *"Integrated scheduling of m-truck, m-drone, and m-depot constrained by time-window, drop-pickup, and m-visit using constraint programming,"* Transportation Research Part C: Emerging Technologies, Volume 91, 2018, Pages 1-14, ISSN 0968-090X, https://doi.org/10.1016/j.trc.2018.03.025.

[5] Bertagnon, Alessandro, and Marco Gavanelli. *"Improved Filtering for the Euclidean Traveling Salesperson Problem in CLP (FD)."* In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 02, pp. 1412-1419. 2020.

[6] Bianchi Leonora, Luca Maria Gambardella, and Marco Dorigo. *"Solving the homogeneous probabilistic traveling salesman problem by the ACO metaheuristic."* In International Workshop on Ant Algorithms, pp. 176-187. Springer, Berlin, Heidelberg, 2002.

[7] Boysen Nils, Stefan Schwerdfeger, and Felix Weidinger. *"Scheduling last-mile deliveries with truck-based autonomous robots."* European Journal of Operational Research 271, no. 3 (2018): 1085-1099.

[8] Boysen Nils, Stefan Fedtke, and Stefan Schwerdfeger. *"Last-mile delivery concepts: a survey from an operational research perspective."* OR Spectrum (2020): 1-58.

[9] BUSINESS, STANFORD. *"Technological disruption and innovation in last mile delivery."* White paper (2016).

[10] Castillo, Vincent E., John E. Bell, William J. Rose, and Alexandre M. Rodrigues. *"Crowdsourcing last mile delivery: strategic implications and future research directions."* Journal of Business Logistics 39, no. 1 (2018): 7-25.

[11] Chauhan, Chetan, Ravindra Gupta, and Kshitij Pathak. *"Survey of methods of solving tsp along with its implementation using dynamic programming approach."* International journal of computer applications 52, no. 4 (2012).

[12] Chen, Cheng, Emrah Demir, Yuan Huang, and Rongzu Qiu. *"The adoption of self-driving delivery robots in last mile logistics."* Transportation research part E: logistics and transportation review 146 (2021): 102214.

[13] Chu, Paul C., and John E. Beasley. *"A genetic algorithm for the multidimensional knapsack problem."* Journal of heuristics 4, no. 1 (1998): 63-86.

[14] De CAPITANI Da VIMERCATE A. Autonomous robots in food delivery: a simulation study[J]. 2019.

[15] De Vries, Andrie, and Joris Meys. *R for Dummies.* John Wiley Sons, 2015.

[16] De Jong, Piet, and Gillian Z. Heller. *"Generalized linear models for insurance data."* Cambridge Books (2008).

[17] Dobson, Annette J., and Adrian G. Barnett. *"An introduction to generalized linear models."* CRC press, 2018.

[18] Diaz, J. (2019, January 15). *"8 robots racing to win the delivery wars of 2019."* Fast Company. https://www.fastcompany.com/90291820/8-robots-racing-to-win-the-delivery-wars

[19] Dudziński, Krzysztof, and Stanisław Walukiewicz. *"Exact methods for the knapsack problem and its generalizations."* European Journal of Operational Research 28, no. 1 (1987): 3-21.

[20] Ellabib, Ismail, Otman A. Basir, and Paul Calamai. *"An experimental study of a simple ant colony system for the vehicle routing problem with time windows."* In International Workshop on Ant Algorithms, pp. 53-64. Springer, Berlin, Heidelberg, 2002.

[21] Figliozzi, Miguel A. *"Carbon emissions reductions in last mile and grocery deliveries utilizing air and ground autonomous vehicles."* Transportation Research Part D: Transport and Environment 85 (2020): 102443.

[22] Figliozzi, Miguel, and Dylan Jennings. *"Autonomous delivery robots and their potential impacts on urban freight energy consumption and emissions."* Transportation Research Procedia 46 (2020): 21-28.

[23] Flood, Merrill M. *"The traveling-salesman problem."* Operations research 4, no. 1 (1956): 61-75.

[24] Francis, S. (2019, May 16). *"Solving the tricky last-mile: Delivery robots to rescue e-commerce."* Robotics Automation News. https://roboticsandautomationnews.com/2019/01/25/solving-the-tricky-last-mile-delivery-robots-to-rescue-e-commerce/20707/

[25] Grefenstette, John, Rajeev Gopal, Brian Rosmaita, and Dirk Van Gucht. *"Genetic algorithms for the traveling salesman problem."* In Proceedings of the first International Conference on Genetic Algorithms and their Applications, vol. 160, no. 168, pp. 160-168. Lawrence Erlbaum, 1985.

[26] Hahsler, Michael, and Kurt Hornik. *"TSP-Infrastructure for the traveling salesperson problem."* Journal of Statistical Software 23, no. 2 (2007): 1-21.

[27] H. Deliveries. (2020). *"Last mile delivery,"* URL: https://hugodelivery.com

[28] Hochbaum, Dorit S. *"A nonlinear knapsack problem."* Operations Research Letters 17, no. 3 (1995): 103-110.

[29] Hoffmann, T., Prause, G., 2018. *"On the regulatory framework for last-mile delivery robots."* Machines 6. doi:10.3390/machines6030033

[30] *"Index | TIOBE - The Software Quality Company".* 2021. Tiobe.Com. https://www.tiobe.com/tiobe-index/.

[31] Jennings, Dylan, and Miguel Figliozzi. *"Study of sidewalk autonomous delivery robots and their potential impacts on freight efficiency and travel."* Transportation Research Record 2673, no. 6 (2019): 317-326.

[32] Kottasova, I. (2015, November 3). *"Forget drones, here come delivery robots."* CNNMoney. https://money.cnn.com/2015/11/03/technology/starship-delivery-robots/

[33] Lee, Hau L., and Seungjin Whang. *"Winning the last mile of e-commerce."* MIT Sloan management review 42, no. 4 (2001): 54-62.

[34] Lenstra, Jan Karel, and AHG Rinnooy Kan. *"Some simple applications of the travelling salesman problem."* Journal of the Operational Research Society 26, no. 4 (1975): 717-733.

[35] Leyman, Pieter, and Patrick De Causmaecker. *"A metaheuristic approach for an intermittent traveling salesperson problem extension."* In Proceedings of the 13th International Conference on the Practice and Theory of Automated Timetabling-PATAT, vol. 1. 2021.

[36] LINDO Systems, Inc. *"LINDO API 11.0 User Manual."*

[37] Liu, Fei, and Guangzhou Zeng. *"Study of genetic algorithm with reinforcement learning to solve the TSP."* Expert Systems with Applications 36, no. 3 (2009): 6995-7001.

[38] Luo, Kaiping, and Qiuhong Zhao. *"A binary grey wolf optimizer for the multi-dimensional knapsack problem."* Applied Soft Computing 83 (2019): 105645.

[39] Malandraki, Chryssi, and Robert B. Dial. *"A restricted dynamic programming heuristic algorithm for the time dependent traveling salesman problem."* European Journal of Operational Research 90, no. 1 (1996): 45-55.

[40] Maibach M, Schreyer C, Sutter D, et al. Handbook on estimation of external costs in the transport sector[J]. Ce Delft, 2008, 336.

[41] Mathews, G. B. (25 June 1897). *"On the partition of numbers"* Proceedings of the London Mathematical Society. 28: 486–490. doi:10.1112/plms/s1-28.1.486.

[42] Nelder, John Ashworth, and Robert WM Wedderburn. *"Generalized linear models."* Journal of the Royal Statistical Society: Series A (General) 135, no. 3 (1972): 370-384.

[43] O. Gomes, C. Combes, A. Dussauchoy, *"Parameter estimation of the generalized gamma distribution",* Mathematics and Computers in Simulation, Volume 79, Issue 4, 2008, Pages 955-963, ISSN 0378-4754, https://doi.org/10.1016/j.matcom.2008.02.006.

[44] Ostermeier, Manuel, Andreas Heimfarth, and Alexander Hübner. *"Cost-optimal truck-and-robot routing for last-mile delivery."* Networks (2021).

[45] Ouyang, Jun, and Gui-Rong Yan. *"A multi-group ant colony system algorithm for TSP."* In Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 04EX826), vol. 1, pp. 117-121. IEEE, 2004.

[46] Pani,Agnivesh, Sabya Mishra, Mihalis Golias, and Miguel Figliozzi. *"Evaluating public acceptance of autonomous delivery robots during COVID-19 pandemic."* Transportation research part D: transport and environment 89 (2020): 102600.

[47] Pilat, Marcin L., and Tony White. *"Using genetic algorithms to optimize ACS-TSP."* In International workshop on ant algorithms, pp. 282-287. Springer, Berlin, Heidelberg, 2002.

[48] Poeting, Moritz, Stefan Schaudt, and Uwe Clausen. *"A comprehensive case study in last-mile delivery concepts for parcel robots."* In 2019 Winter Simulation Conference (WSC), pp. 1779-1788. IEEE, 2019.

[49] Puchinger, Jakob, Günther R. Raidl, and Ulrich Pferschy. *"The multidimensional knapsack problem: Structure and algorithms."* INFORMS Journal on Computing 22, no. 2 (2010): 250-265.

[50] Punakivi, M., Yrjölä, H. and Holmström, J. (2001) *"Solving the last mile issue: reception box or delivery box?"* International Journal of Physical Distribution Logistics Management, Vol. 31 No. 6, pp. 427-439. https://doi.org/10.1108/09600030110399423

[51] *"R: What Is R?".* 2021. R-Project.Org. https://www.r-project.org/about.html.

[52] Razali, Noraini Mohd, and John Geraghty. *"Genetic algorithm performance with different selection strategies in solving TSP."* In Proceedings of the world congress on engineering, vol. 2, no. 1, pp. 1-6. Hong Kong: International Association of Engineers, 2011.

[53] Ranieri, L.; Digiesi, S.; Silvestri, B.; Roccotelli, M. *"A Review of Last Mile Logistics Innovations in an Externalities Cost Reduction Vision."* Sustainability 2018, 10, 782. https://doi.org/10.3390/su10030782

[54] Ross, Keith W., and Danny HK Tsang. *"The stochastic knapsack problem."* IEEE Transactions on communications 37, no. 7 (1989): 740-747.

[55] Samouh, Farah, Veronica Gluza, Shadi Djavadian, SeyedMehdi Meshkani, and Bilal Farooq. *"Multimodal Autonomous Last-Mile Delivery System Design and Application."* In 2020 IEEE International Smart Cities Conference (ISC2), pp. 1-7. IEEE, 2020.

[56] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi. *"Optimization by Simulated Annealing"* Science, 13 May 1983

[57] Shih, Wei. *"A branch and bound method for the multiconstraint zero-one knapsack problem."* Journal of the Operational Research Society 30, no. 4 (1979): 369-378.

[58] Susan Dean, Barbara Illowsky *"Descriptive Statistics: Skewness and the Mean, Median, and Mode"* Connexions website

[59] Simon, M. (2017, December 6). *"San Francisco Just Put the Brakes on Delivery Robots."* Wired. https://www.wired.com/story/san-francisco-just-put-the-brakes-on-delivery-robots/

[60] Simoni, Michele D., Erhan Kutanoglu, and Christian G. Claudel. *"Optimization and analysis of a robot-assisted last mile delivery system."* Transportation Research Part E: Logistics and Transportation Review 142 (2020): 102049.

[61] Sinha, Prabhakant, and Andris A. Zoltners. *"The multiple-choice knapsack problem."* Operations Research 27, no. 3 (1979): 503-515.

[62] Sonneberg, Marc-Oliver, Max Leyerer, Agathe Kleinschmidt, Florian Knigge, and Michael H. Breitner. *"Autonomous unmanned ground vehicles for urban logistics: optimization of last mile delivery operations."* In Proceedings of the 52nd Hawaii International Conference on System Sciences. 2019.

[63] Stutzle, Thomas, and Holger Hoos. *"MAX-MIN ant system and local search for the traveling salesman problem."* In Proceedings of 1997 IEEE international conference on evolutionary computation (ICEC'97), pp. 309-314. IEEE, 1997.

[64] Vance, Ashlee (6 January 2009). *"Data Analysts Captivated by R's Power"*. New York Times.

[65] Venkatesh, Pandiri, and Alok Singh. *"Two metaheuristic approaches for the multiple traveling salesperson problem."* Applied Soft Computing 26 (2015): 74-89.

[66] Vincent, J. (2019, February 27). *"FedEx unveils autonomous delivery robot."* The Verge. https://www.theverge.com/2019/2/27/18242834/delivery-robot-fedex-sameday-bot-autonomous-trials

[67] Wang, Wenmin, Wei Zhao, Xiaohan Wang, Zhihong Jin, Yuanchen Li, and Troy Runge. *"A low-cost simultaneous localization and mapping algorithm for last-mile indoor delivery."* In 2019 5th International Conference on Transportation Information and Safety (ICTIS), pp. 329-336. IEEE, 2019.

[68] Wang, Xuping, Linmin Zhan, Junhu Ruan, and Jun Zhang. *"How to choose "last mile" delivery modes for e-fulfillment."* Mathematical Problems in Engineering 2014 (2014).

[69] Wang, Yuan, Dongxiang Zhang, Qing Liu, Fumin Shen, and Loo Hay Lee. *"Towards enhancing the last-mile delivery: An effective crowd-tasking model with scalable solutions."* Transportation Research Part E: Logistics and Transportation Review 93 (2016): 279-293

[70] Welcome to the QGIS project! (2021). QGIS. https://qgis.org/en/site/

[71] Wong, J. C. (2017, December 10). *"San Francisco sours on rampant delivery robots: "Not every innovation is great.""* The Guardian. https://www.theguardian.com/us-news/2017/dec/10/san-francisco-delivery-robots-laws

[72] Miller C E, Tucker A W, Zemlin R A. *"Integer programming formulation of traveling salesman problems"* [J]. Journal of the ACM (JACM), 1960, 7(4): 326-329.

[73] Egeblad J, Pisinger D. "Heuristic approaches for the two-and three-dimensional knapsack packing problem[J]." Computers Operations Research, 2009, 36(4): 1026-1049.

[74] Martello S, Pisinger D, Vigo D, et al. *"Algorithm 864: General and robot-packable variants of the three-dimensional bin packing problem[J]."* ACM Transactions on Mathematical Software (TOMS), 2007, 33(1): 7-es.

A
Appendix 1

Table A.1: Example of predicted dimensional data

Destination_no	Length	Width	Height
7	8.35	21.21	8.12
7	31.63	10.18	4.82
12	31.63	23.44	18.76
12	31.63	23.44	15.04
12	36.19	17.93	4.82
12	31.63	23.44	4.82
12	31.63	23.44	21.83
12	31.63	23.44	4.82
12	31.63	23.44	15.04
12	36.19	17.93	4.82
12	31.63	23.44	4.82
12	31.63	23.44	15.04
12	36.19	17.93	4.82
12	31.63	23.44	20.76
7	49.00	32.65	1.36
14	18.00	26.01	1.36
14	18.00	26.01	4.82
12	31.63	23.44	15.04
12	36.19	17.93	0.22
14	38.87	28.79	26.80

The table A.1 shows the predicted dimension of twenty packages. The unit of length, width and height is centimeters.

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
Division of Service Management and Logistic
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY