INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

MESTRADO EM ENGENHARIA MECÂNICA

MÉTODOS DE ANÁLISE DA COMPLEXIDADE NO PROBLEMA DE EMPACOTAMENTO DE PALETES DO DISTRIBUIDOR

HUGO ANDRÉ DE ALMEIDA BARROS

novembro de 2018

POLITÉCNICO
DO PORTO

Instituto Superior de Engenharia do Porto

COMPLEXITY ANALYSIS IN THE DISTRIBUTOR PALLET LOADING PROBLEM

Hugo André Almeida Barros
Nº 1131227

2018

School of Engineering, Polytechnic Institute of Porto
Department of Mechanical Engineering

POLITÉCNICO DO PORTO

POLITÉCNICO
DO PORTO

COMPLEXITY ANALYSIS IN THE DISTRIBUTOR PALLET LOADING PROBLEM

Hugo André Almeida Barros
Nº 1131227

Dissertation presented to ISEP – School of Engineering, Polytechnic Institute of Porto to fulfill the requirements necessary to obtain a Master's degree in Mechanical Engineering, carried out under the guidance of Engineer António Ramos

2018

ISEP – School of Engineering, Polytechnic Institute of Porto
Department of Mechanical Engineering

POLITÉCNICO
DO PORTO

POLITÉCNICO
DO PORTO

JURY

President
Doctor Manuel Dores de Castro
Adjunct Professor, ISEP
Supervisor
Doctor António Galrão Ramos
Adjunct Professor, ISEP
Second supervisor
Doctor Maria Teresa Pereira
Adjunct Professor, ISEP
Examiner
Doctor Fernanda Amélia Ferreira
Adjunct Professor, ESHT

ACKNOWLEDGEMENTS

I’d like to thank the following people, for all the help given during the time spent researching and writing this thesis:

My thesis supervisor, the professor António Galrão Ramos, for the continuous guidance that allowed me to finish this report;

The professor Maria Teresa Pereira, for introducing me to the company “Luis Simões” and helping me with the usage of the software called SPSS;

The Luis Simões workers, with a special thanks to António Lacerda and Hélder Sousa for arranging the visits to the company’s warehouse. To Joana Ribeiro for helping me during the field work and introducing me to the warehouse pickers, who deserve special thanks as well for helping me during their labour time.

My university colleagues and friends, who shared with me the experience of making a thesis, resulting in a brainstorm of ideas.

My international buddies, who gave me a helping hand mostly in the grammatical component of this report.

Last but not least, my family, specially my parents, who were always there giving their support during such hard times.

KEYWORDS

Pallet loading problem, trim-loss, complexity, principal component analysis, multiple linear regression

ABSTRACT

In the pallet loading problem, one of the main goals is to allocate the highest number of boxes as possible, to minimize empty spaces in the pallet. Those empty spaces are called trim-loss. If all boxes have a rectangular shape, which is the most common one, it is possible to pack them so that their faces are coincident with themselves. By doing that, the trim-loss can be minimized. Although loading a pallet may seem linear to most people, some customers impose restrictions that increase the complexity of the pallet loading. Due to that, to evaluate the complexity of a packed pallet, some metrics were created. They consist in an evaluation of a set of parameters that are inherent to the pallet loading process and affect its complexity. After analysing some of those constraints and loading methods enforced by some pickers in a real company, it was possible to obtain samples where the metrics were applied to learn which parameters add the most complexity in the pallet loading process. In the future, after knowing the relevancy of each parameter, the metrics can be used in pallet generation tools to learn how complex is the loading of a certain pallet and study new and easier ways to load the boxes that reduce the complexity of such process.

Two statistical tests were then used to analyse the samples retrieved: the principal components analysis and the multiple linear regression. The first is used to combine multiple variables into a smaller set that represents the most relevant information, while the multiple linear regression uses the variables and respective observations to calculate a model that can predict the value of the complexity of a packed pallet in given circumstances. In the first one, it was learned that three principal components were extracted, but since the third one explained a small percentage of the total data variance, it was decided to retain only two components: the box quantities, which explains 41% of the total variance, followed by the box dimensions, explaining 28% of the total variance. The multiple linear regression revealed that the component representing the box quantities, which contains the Number of Box Types, Number of Column Piles, Number of Boxes, Time Spent Packing, and Percentage of Fragile Boxes variables is the component that mostly increase the complexity of pallet cargo arrangements. Although the model can predict the data that was obtained with an average accuracy, some of the coefficients ended up being small, those being related to the components Box Dimensions, which has the Number of Heavy Boxes, Average Box Weight, Average Maximum Width variables, and Height Between Pile and Worker and Number variables, meaning that they aren't very significant towards evaluating the complexity of a pallet loading process. Using a multiple linear regression with the 9 variables showed that the variable who adds more complexity is the Number of

Column Piles. Overall, the results obtained were acceptable, and showed that the variable that adds more complexity is the ones that the pickers see as adding more complexity, and also that the results of the multiple regression with the components match the one using the original variables. It is worth noting that this variable is subjective, meaning that one worker's perception on the complexity may not match others' perception. Despite having obtained only one variable being considered as statistically significant towards explaining the complexity in the pallet loading problem, it doesn't mean it's the only one that adds complexity.

PALAVRAS CHAVE

Problema de carregamento de paletes, *trim-loss*, complexidade, análise de componentes principais, regressão linear múltipla

RESUMO

No problema de carregamento de paletes, um dos grandes objetivos é alocar o maior número de caixas possível, visando minimizar espaços vazios conhecidos por *trim-loss*. Se todas as caixas possuírem um formato retangular, que é o formato mais comum, é possível arrumá-las de forma que as suas faces fiquem encostadas entre si, minimizando assim o *trim-loss*. No entanto, apesar do empacotamento de caixas em paletes parecer linear para a maioria das pessoas, certos clientes impõem restrições que aumentam a complexidade do empacotamento. Como tal, para avaliar a complexidade de um arranjo de paletes, criaram-se métricas, que consistem na avaliação de um conjunto de parâmetros inerentes ao processo ou às características do carregamento de paletes que afetam a sua complexidade. Após analisar numa empresa real as restrições e os métodos de empacotamento usados pelos operadores, foi possível obter amostras onde as métricas são aplicadas para tentar saber quais as mais relevantes no processo, para assim futuramente estas métricas serem aplicadas em ferramentas de geração de paletes para poder analisar os resultados obtidos e estudar maneiras onde estas sejam carregadas mais facilmente.

Posteriormente, dois testes estatísticos foram aplicados aos dados recolhidos: uma análise de componentes principais e a regressão linear múltipla. O primeiro usa-se para combinar várias variáveis e formar um conjunto mais pequeno que represente a informação mais relevante, enquanto a regressão linear múltipla usa as variáveis e respetivas observações para calcular um modelo que consiga prever valores de complexidade do carregamento de paletes em quaisquer circunstâncias. No primeiro, verificou-se a existência de três componentes principais, mas dado que o terceiro componente explica uma percentagem da variância total dos dados pequena, decidiu-se extrair apenas dois componentes: as quantidades das caixas é o componente que explica maiores valores de variância nos dados (41%), seguido pelas dimensões das caixas, explicando 28% da variância total dos dados. A regressão linear múltipla revelou que o componente que representa as quantidades das caixas, que contém as variáveis Número de Tipos de Caixa, Número de Colunas, Número de Caixas, Tempo Despendido a Carregar Caixas e Percentagem de Caixas Frágeis, é aquele que faz crescer mais substancialmente a complexidade do carregamento de caixas em paletes. Com os vários testes, verificou-se que os componentes Dimensões das Caixas, que possui as variáveis Número de Caixas Pesadas Carregadas, Peso Médio das Caixas, Largura Máxima Média, e a diferença de alturas entre pilhas de caixas e o operador, não acrescentam muita significância na explicação da avaliação da complexidade no problema de carregamento de paletes. A regressão linear múltipla com as variáveis originais mostrou que o Número de Colunas é a variável que adiciona mais complexidade. Apesar do modelo obtido ter significância, quase todos os coeficientes obtidos acabaram por ser baixos e com valores Significância (sig.) acima de 0,05, não sendo essas variáveis relevantes no modelo.

Valores baixos de Cronbach’s Alpha e R^2 ajustado evidenciam a suscetibilidade da aparição destes valores. No geral, os resultados obtidos nesta dissertação foram satisfatórios, mas os coeficientes baixos da regressão linear múltipla não foram bons. O número de observações retido e o escalamento das variáveis são causas possíveis para esta discrepância de valores ter acontecido. Vale a pena referir que a variável que avalia a complexidade é uma variável subjetiva, pelo que o que um picker considera como sendo complexo pode não corresponder ao que outros trabalhadores pensem. Apesar de, estatisticamente, apenas uma variável ter significância na explicação da complexidade, na realidade todas as variáveis têm alguma influência na complexidade do carregamento de caixas em paletes. No geral, a perceção dos trabalhadores tem semelhanças com aquilo que se obteve nos resultados das regressões lineares.

LIST OF SYMBOLS AND ABBREVIATIONS

List of abbreviations

2D	two-dimensional
3D	three-dimensional
3PL	third-party logistics
C&P	Cutting and Packing
e. g.	for example
ISO	International Organization for Standardization
KMO	Kaiser-Meyer-Olkin
LIFO	Last In First Out
PCA	Principal Components Analysis

List of units

°	Degrees
cm	Centimetres
kg	Kilograms
m	Meters
mins	Minutes
rad	Radians

List of symbols

&	and
%	percentage
>	higher than
≤	lesser or equal than
=	equal to
+	plus
-	minus

GLOSSARY OF TERMS

Input value minimization	Situation where there's no need to use all large objects to accommodate the small objects, therefore the minimum number of large objects must be used
Large object	Platform where the small items are piled at, and is commonly either a container or pallet
Output value maximization	Situation where all large objects available must be used to accommodate the small objects to pack
Pallet loading problem	Problem whose focus is to find space optimization to load the biggest number of boxes in a pallet as possible
Pallet shuttle	Mechanism installed in warehouse racks that moves pallets with products into the edge of the rack, making it easier to lower it to the ground
Picker	The worker who places the boxes in a pallet, according to the requirements of a client's order
Small object	It's the package containing the product that will be shipped and placed in a large object
Supply chain	Network between companies and its suppliers that produces and distributes a certain product
Trim-loss	Empty spaces located between the pallet's surface and the top of the pile
UT	Number used by Luis Simões to distinguish each packed pallet and each order

FIGURES INDEX

TABLES INDEX

INDEX

INTRODUCTION

1 INTRODUCTION

1.1 CONTEXTUALIZATION

In the industry, there are plenty of activities that most companies must do to organize their products, which includes the storage activities, one of the most common. Since the competitiveness is usually high, the companies try to save the most money as possible in storage and transportation expenses, but usually, the variables involved in the process make it quite complicated and tough to handle. Many companies end up having a deficient logistic system that can't deal with the needs of the costumers when it comes to receiving the order quickly enough. With that in mind, the companies look for technologically advanced systems to help solve those problems. With the aid of mathematical models and automatized systems, the supply chain can flow with greater performance, leading to bigger money savings and customer satisfaction.

In a warehousing system, order picking is one of the most important activities. That's because it is generally laborious and costly. It is estimated that picking represents around 55% of all warehouse-related expenses (Murray, 2018). As such, optimizing the way that order picking works can be considered as a priority for any company.

Picking can be done either automatically or manually. The former requires an automatized system of machines as well as good warehouse organization to optimize the flow of the packing. The latter has at least one picker placing the boxes in either a container or a pallet, depending on what each company decides to use. The way the boxes are placed in the recipient can have a huge impact in the loading/unloading operations. Quantifying how impactful those box arrangements are for human and machine labour is important, so they can be rearranged to optimize the timing of the operations, both for the client and for the working personnel.

Over time, many authors have been writing about the subject mentioned above, but they have a wide range of disagreements between one another. They have different perspectives on the constraints, the objectives, and the complexity of arrangements, so the authors do not have a consensus on what is the best approach. It is also worth noting that older papers have solutions that are more limited than newer ones. This thesis will look mostly into the latest literature released about the subject and analyse the approaches made to the problem to come up with a solution that can have a better practical use.

1.2 OBJECTIVES

The main goal of this report is to analyse the complexity of pallet arrangements. To achieve that, the sequence below was followed:

- Review the literature and analyse the multiple approaches given by several authors to this problem;

- Characterize the problem in terms of what the authors refer to as complexity of arrangements, constraints, and objectives;
- Create metrics to quantify the complexity of a cargo arrangement;
- Gather samples;
- Validate the data obtained with computational tests.

1.3 WORK METHODOLOGY

In order to analyse this problem, the first approach was to find out the characteristics of this type of problem in a generic manner. After doing so, a more detailed investigation of the pallet loading problem was made in the literature to obtain all data needed to comprehend it, its constraints, objectives, and other characteristics. Afterwards, a 3PL (third-party logistics) company was visited to see their pallet loading system and visualize their approach to the problem. After analysing the pallet loading problem in theory and in a practical way, a mathematical model was developed to read the complexity of real-life situations about the pallet loading problem.

1.4 DISSERTATION STRUCTURE

This dissertation is divided into 4 different chapters. In this chapter, it's made an introduction to the problem, which references how it will be approached, what goals need to be satisfied to reach an acceptable solution and the steps taken to get there.

The second chapter contains a review of what is written in the literature about the pallet loading problem, including a more overall insight on how this type of problem is classified, what makes this such a big issue when it comes to an industrial environment and what are the singularities of this problem.

In the third chapter, there's a description of the company where the field work took place and what was done during that time. Also, it contains information about how the data was processed, what methodologies were used and what results came out of it.

Finally, in chapter 4, there's an overview of the results and conclusions taken from all the data gathered and methods used that were written in this report. There's also a critical review of the methods applied and what could be modified to improve the results.

BIBLIOGRAPHIC WORK

2 BIBLIOGRAPHIC WORK

2.1 A TYPOLOGY FOR C&P (CUTTING AND PACKING) PROBLEMS BY WÄSCHER

The problem to be analysed in this thesis can be classified as a cutting and packing problem. In the literature, it is known by many other names, such as cutting stock or trim loss problem. Although the focus will be mainly on the packing part of the problem, due to its complexity, some authors proposed typologies of the problem so that it's possible to define it better. Originally, Dyckhoff (1990) wrote a typology of the problem that was many years later improved by Wäscher, Haußner, & Schumann (2007). In the following paragraphs, all the categories considered by the author will be detailed to have a better insight of all the characteristics of the problem.

To understand the basics of the logical structure of the problem, it is easier to visualize an example. The figure 1 exhibits a container loading problem. In it, there's a large container where some boxes have the same width and variable heights and lengths. Note that the last two parameters aren't random, but instead have specific measures.

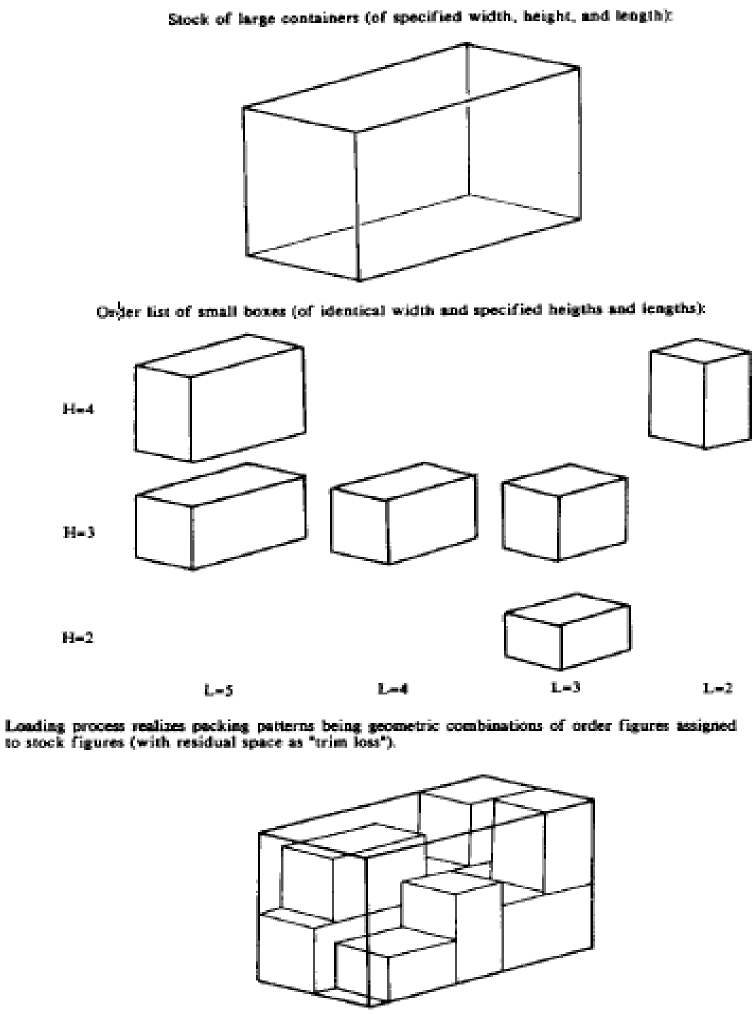

Figure 1: Representation of a container loading problem (Dyckhoff, 1990)

As seen in Figure 1, it is possible to identify two groups of objects to consider in the problem. One is a bunch of large objects, which represent containers, the other is an aggregation of smaller objects, the boxes that will be packed in the items of the other group. Besides some situational constraints and objectives, the aspect that is given more emphasis on is the geometrical combinations of the smaller objects into packing patterns that can be allocated to the container. Knowing these two groups that form geometric bodies, it is possible to add a third one to the mix. After having a container filled with boxes, the patterns formed inside the parts of the container representing the empty space are known as “trim loss”. That unused space can be seen in Figure 2 in 2D. This third group is only considered by Dyckhoff (1990) and not Wäscher et al. (2007), even though the latter acknowledges its importance in his article.

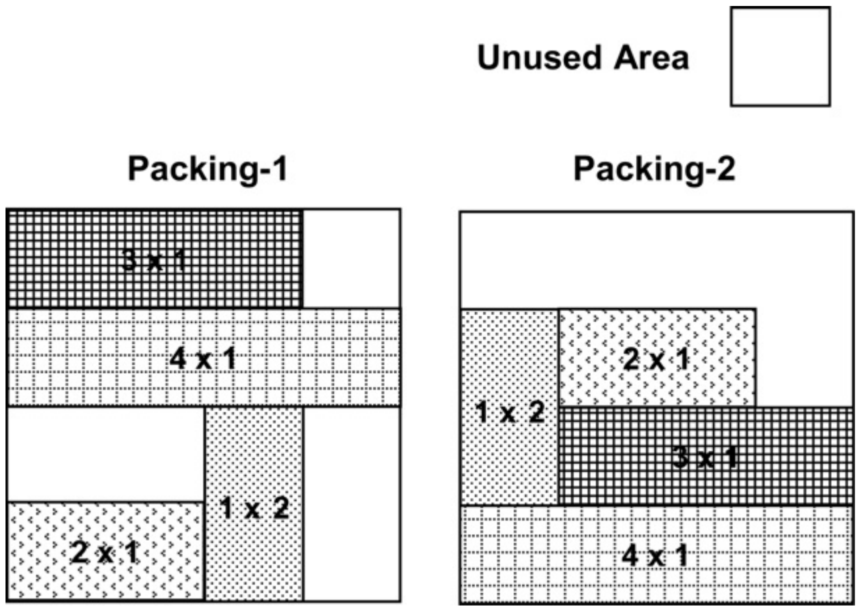

Figure 2: 2D representation of the trim-loss (Gonçalves, 2007)

Wäscher et al. (2007) mentions five of most important aspects of the C&P problem:

Dimensionality: In this component, the author considers problems with three different dimensions: one-, two-, and three-dimensional problems. In the literature, problems with more than 3 dimensions are also referred, for example by Lins, Lins, & Morabito (2002), although that isn’t very common.

Kind of assortment: Wäscher et al. (2007) refers to two situations for this component. One is the output value maximization and the other input value minimization. The first refers to a situation where a set of large objects isn’t enough to accommodate a set of small items ready to be packed. Therefore, all large objects must be used, which means there’s no need to select them. This forces to maximize the value of small items to be assigned to each of the big ones. When it comes to input minimization, once again there’s a set of small items and another set of big items, but this time the larger ones can accommodate the smaller ones. The goal now is to minimize the value of large objects to be used. The author explains that in this matter, the concept of value means something generic and depends on the problem being analysed. It could be a material quantity or a cost. He also mentions that output value maximization and input value minimization can be translated into waste minimization, which means that the goal is to minimize the unused space of the larger objects as in minimizing trim-loss. Although he only considers these two situations, he acknowledges that both may simultaneously occur in real life practice.

Assortment of small items: There are three separate cases according to Wäscher et al. (2007). They are “identical small items”, “weakly heterogeneous assortment of small items”, and “strongly heterogeneous assortment of small items”. All of these are related to the dimensions of the items: length, width, and height. In the first case mentioned above, all items have the same shape and size. Because of that, this scenario can be considered the simplest of all three. In weakly heterogeneous assortments, the tinier items can be grouped into a smaller number of classes compared to the total number of items. Overall, they are similar to themselves in terms of shape and size. Items with same size and shape but different orientations are treated as a different type of item. In strongly heterogeneous assortments, only a few elements are identical in terms of shape and size. That means each item will be treated as an individual entity. This is the most complex scenario due to the differences in all the items.

Assortment of large objects: The author refers to two major scenarios. One is where there’s just one large object, the other is where many large objects are used. In the first case, the set of large objects is made of only one element. Wäscher et al. (2007) divides this into two other situations. The first has the large object possess only fixed dimensions and the other has at least one dimension of the object is variable in its extension. The latter may be harder to deal with due to the dimension or dimensions that are variable, which forces a different approach than an object with constant dimensions. When it comes to the usage of multiple objects, the author also created three subcases that are identical to the ones seen above in the assortment of small objects. They are “identical large objects”, “weakly heterogeneous assortment”, and “strongly heterogeneous assortment”. He also treats these items as if they always have fixed dimensions and a rectangular shape. Objects different that those are treated as problem variants.

Shape of small items: The author distinguishes two types of items for 2D (two dimensional) and 3D (three dimensional) problems: regular and irregular items. Regular items have well known geometrical shapes such as rectangles and cylinders, among others, while irregular ones do not. Wäscher et al. (2007) also says that in 2D problems, the regular objects have further distinctions into circular items, rectangular items, among others.

After describing all the characteristics inherent to C&P problems as told by Wäscher et al. (2007), it is possible to see how complex this kind of problem can be. The Figure 3 sums up what was said in the previous paragraphs of this section and each categorization can be compared to Dyckhoff (1990) by looking at Figure 4.

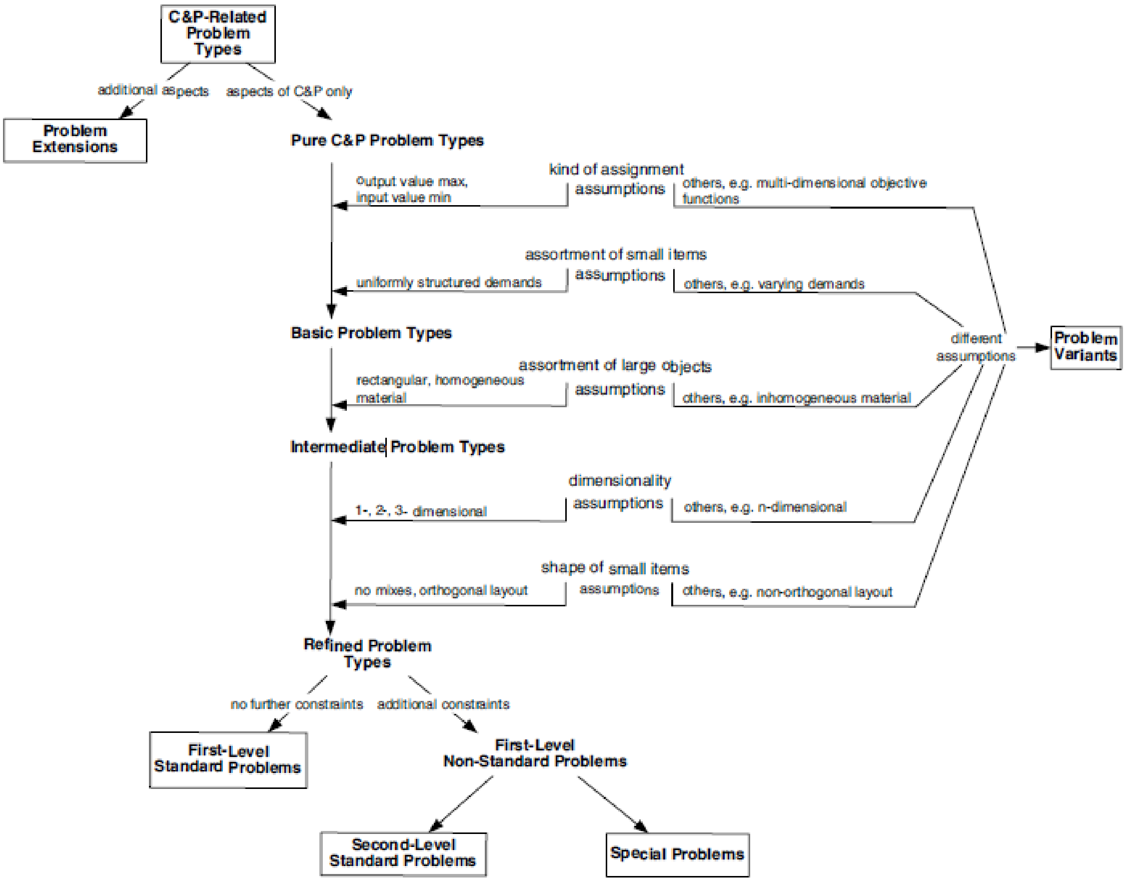

Figure 3: Schematization of the C&P problem (Wäscher et al., 2007)

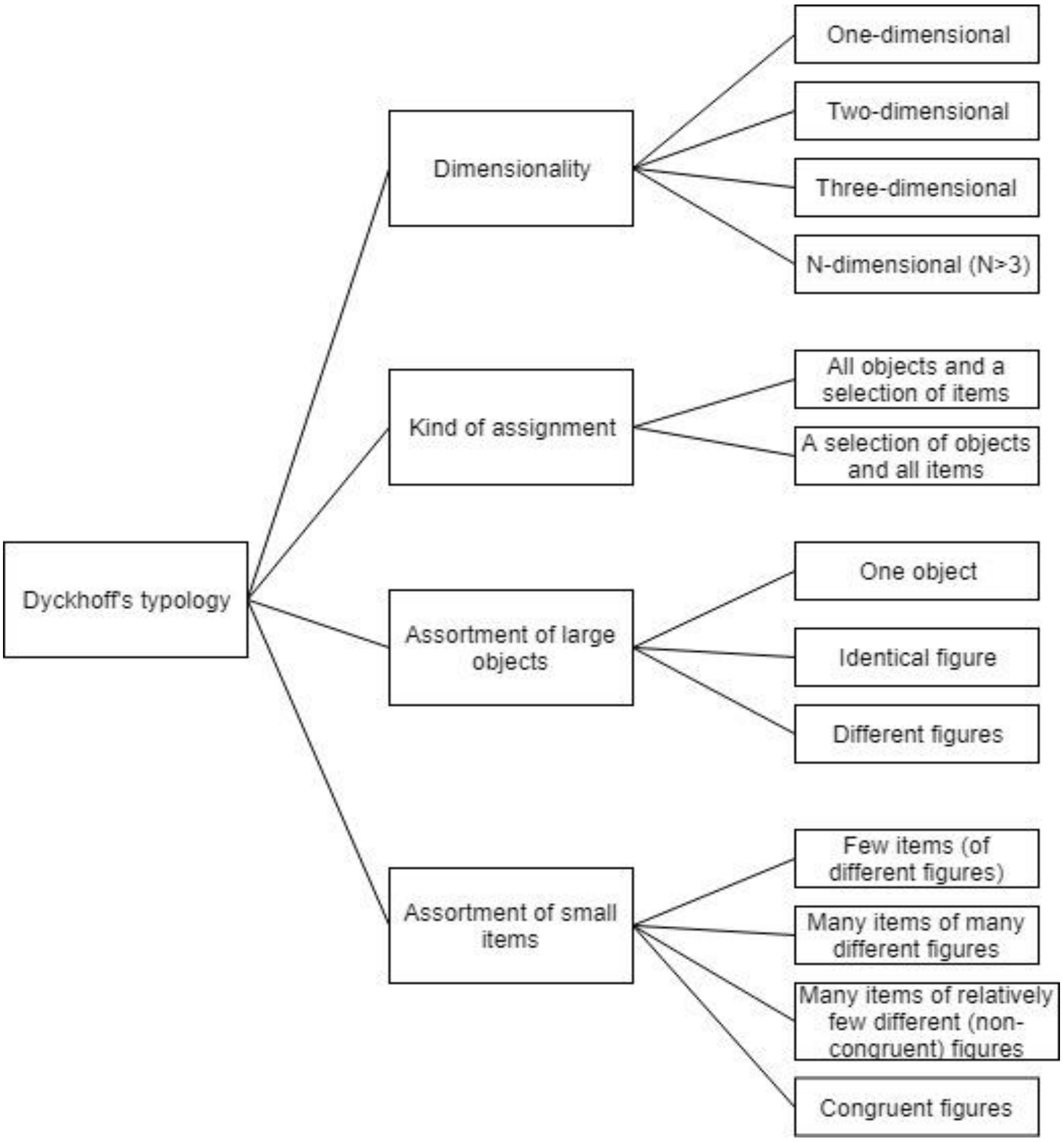

Figure 4: Schematization of Dyckhoff's typology

2.2 THE PALLET LOADING PROBLEM

2.2.1 PROBLEM DESCRIPTION

After exhibiting the types of C&P problems as referred by Wäscher et al. (2007) in section 2.1, in this sub-section that knowledge will be applied to the pallet loading problem to characterize it and showcase all parcels involved in it.

While explaining the basis of a C&P problem, an example of a container loading problem was shown by Wäscher et al. (2007). That problem doesn't differ much from the pallet loading problem. The main difference is that the packing platform to be used to carry the boxes is a pallet instead of a container. From a practical viewpoint, the container is safer than a pallet. While a pallet is simply a platform where objects can be placed orthogonally to its basis, the container can be considered as a pallet with walls. That allows the pile to be placed in the object to be higher than in a pallet because the walls of the container can ensure a bigger cargo stability that is lacking in a pallet, although some mechanics can be used to increase safety in cargo transportation by stabilizing the position of the boxes.

The pallet loading problem, dimension-wise, although it can be considered a 2D problem by some authors, here it will be treated as a 3D problem. Packing boxes into a pallet sometimes may not be viewed as a 3D problem if the arrangement doesn't involve the placement of boxes on top of other boxes. Usually, that doesn't occur because companies want to use the least number of pallets as they can to transport the cargo, therefore filling them with boxes is usually the way to go.

In the pallet loading problem, both scenarios (output value maximization and input value minimization) classified by Wäscher et al. (2007) suit this problem. That's because in practical situations, having a very limited number of pallets to carry the cargo and having a number of pallets great enough to consider it infinite are both possible to occur. As seen in section 2.1, the first situation requires a maximization of the "value" of small objects. If that number represents packing of boxes per pallet, then that's the case to consider for this problem, since the bigger the number of boxes packed in one pallet, the more complex the arrangement of the packing can get. That also occurs when having an elevated heterogeneous assortment, because it makes the act of packing less linear due to the increase of available options to sort the items. This type of assortment is rather common in companies that deal with multiple types of items, such as Amazon, and it is the one considered in this pallet loading problem.

The large object to consider for this problem is the pallet. This object is very common worldwide to transport cargo in warehouses and can be made of many materials, such as wood, metal, and plastic. Multiple worldwide organizations have created standards for the dimensions of the pallets, such as ISO (International Organization for Standardization), which means that there's a large diversity of pallets when it comes to size, although some have similar dimensions. For this problem, it should be considered the existence of multiple large objects with either heterogeneous or homogeneous assortment of large objects. That's because in a large warehouse environment there's usually a mass unload or load of boxes from different destinations where each of them uses different pallets.

The small items in this problem will be considered as being rectangular boxes. That's the most common box format and is the best one to use when attempting to minimize trim-loss.

2.2.2 COMPLEXITY CONSTRAINTS

In the pallet loading problem, the complexity constraints are some of the hardest to work around. The reasons for this are that the manual pallet loading may not be acceptable due to the possibility of the patterns being misunderstood by the loading personnel, and the operation may take longer to perform. Also, the usage of automatized mechanisms for packing may not be the most suitable to do complex cargo arrangements. Besides all these issues, such operations may cost a lot of money. These situations highlight the existing limitations in human and technological resources. The increase of the complexity in packing patterns usually translates into a greater materials handling effort. That effort is more significant if the complexity of the pattern causes changes in the way the box loading is done. Instead of being able to load the boxes with clamp or forklift trucks, the complexity of the pattern may force a manual load of the boxes, which makes the process much more laborious. In case there's no alternative box loading methods that optimize the process, the pattern must conform the limitations of the technology used (Bischoff, EE; Ratcliff, 1995).

In the literature, when studying complexity constraints, the concepts of guillotine patterns and robot-packable pattern are often mentioned. A loading pattern is said to be guillotineable if, for example, a parallelepiped is being transported and that object is subjected to multiple cuts parallel to its faces. By doing so, smaller parallelepipeds are obtained, which represent the multiple stacked boxes. This pattern can be visualized in Figure 5 in both 2D (on the left) and 3D (on the right). Although this arrangement is easy to pack and describe, it isn't always the proper option to load pallets, due to the instability of the cargo during transportation. This situation is more critical in pallets, where other operations are needed to restrain the boxes and make the transportation safe, such as thrink-wrapping or interlocking (Bortfeldt & Wäscher, 2013).

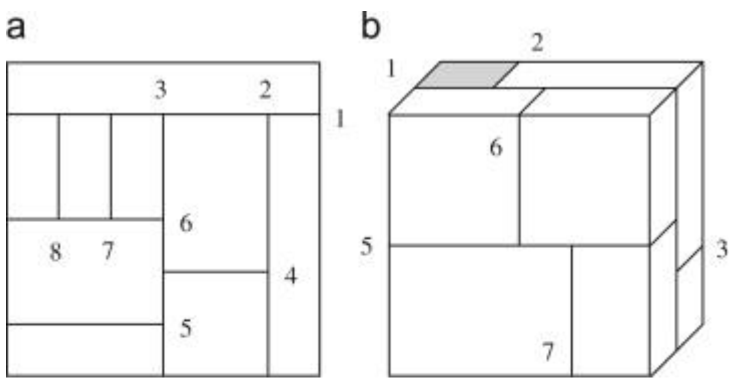

Figure 5: Example of (a) 2D and (b) 3D guillotineable pattern (Amossen & Pisinger, 2010)

The other pattern, the robot packable pattern, is done according to Martello, Pisinger, Vigo, Boef, & Korst (2007), Egeblad & Pisinger (2009), Amossen & Pisinger (2010) by placing the first box in the bottom left behind corner of the container and then placing the remaining boxes either on the right, front or on top of the first one. A guillotineable pattern is also a robot packable pattern due to the way the boxes faces coincide with each other and the container or pallet, but the opposite situation doesn't occur, as shown in Figure 6. The boxes are packed by a robot equipped with an elevating mechanism parallel to the base of the container or pallet. Each box is lifted and released in the correct position using vacuum cells. Due to the nature of this pattern, robots who are used to pack boxes possess extra constraints so that there are no collisions between previously packed boxes and the boxes to pack next. Because of this, there can't be any boxes packed in front of, to the right of, or above the destination of the boxes the robot is currently placing. They can be also placed manually in the pallet, although that operation will likely take more time to perform than with an automatized system.

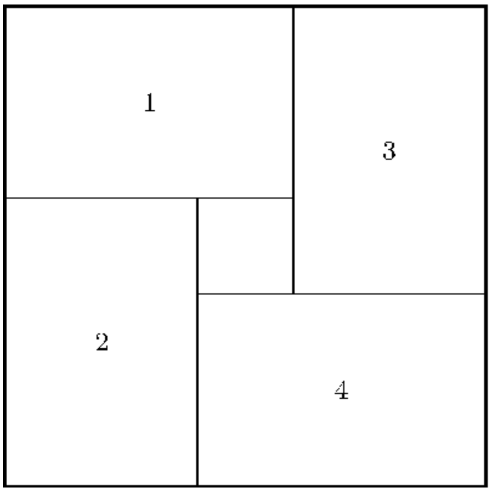

Figure 6: 2D representation of a robot packable pattern (Martello, Pisinger, Vigo, Boef, & Korst, 2007)

To handle these constraints, multiple authors present different solutions. Morabito & Arenalest, (1994) proposes a AND/OR graph approach, which consists in an algorithm combining two basic strategies: backtracking, which chooses all available non-final paths to be explored, and hill-climbing, which chooses the optimal path and keeps discarding the remaining ones. Hifi, (2002) uses an approximate algorithm containing additional constraints, such as stability constraints, which is referred to in the next section. Egeblad & Pisinger, (2009) propose a local search algorithm, which searches for potential solutions by doing local changes to the current solution, that until a solution considered as optimal is found or when a time bound is elapsed. The author refers that this solution performs better in medium sized instances.

In the end, the robot packable pattern can be considered as an automatized variation of the guillotineable pattern, due to its usage in robotic systems, while the guillotineable pattern is more adequate in manual loading because it's easy to understand and pack in the picker's perspective.

2.2.3 OTHER CONSTRAINTS

The pallet loading problem possesses multiple constraints other than the complexity of the packing arrangement. Some only exist under certain circumstances, as seen in the following paragraphs, and others appear all the time when loading the boxes into the pallet. In the literature, most authors describe constraints that are mostly used in a mathematical viewpoint and not in real life practice. Below there are multiple examples of constraints shown in the literature.

Vertical and horizontal stability (dynamic and static): When dealing with a pallet loaded with cargo, one must pay attention to the weight that is being supported by both the small and the large objects. From a vertical standpoint, the pallet must be able to hold the cargo without breaking or suffering any type of damage. Also, the distribution of the boxes must the stable enough not to fall nor, when handling fragile content, balance in the pile. The small objects must also be able to withstand the weight of boxes stacked on top of them. In the case of fragile content, it's strongly advised to place them on top of the pile to avoid damage. These situations only refer to a static position, when the pallet loaded with boxes isn't moving. In a dynamic standpoint, as in when the objects are being transported, they must be packed in a way that assures no boxes will fall or balance much due to speed differences related to the acceleration of the movement. When slowing down or speeding up the movement of the pallet, the boxes will tend to keep their current velocity, which will cause the stack to balance. When the pile's size is too big when compared to the size of the base, the boxes will tend to fall. This horizontal component also affects the interaction between boxes. Some small items who possess fragility in horizontal orientations must be placed in a way that limits their interactions with other boxes and the movements in the stack. All these situations referred above are unsafe and the packing must be carefully planned to avoid such scenarios (Bischoff, EE; Ratcliff, 1995; Bortfeldt & Wäscher, 2013; Elhedhli, Gzara, & Yan, 2017; Junqueira, Morabito, & Sato Yamashita, 2012; Lin, Kang, Liu, & Li, 2016; Morabito & Arenalest, 1994; Ramos, Oliveira, & Lopes, 2016)

Ramos et al. (2016) refers to a variation of the static stability constraint known as "full base support constraint". It imposes that the entire base of a box must be in contact with the base of either the pallet or another box. That way, the whole pile would be stable and no box would balance much.

Box overlap: This is a common constraint in C&P problems. Although it has no practical use, it is mandatory when modelling the problem. This constraint imposes that no small item will occupy the position of an already placed box. Without this constraint, the modelled problem wouldn't be functional because of the eventual overlap of the boxes, creating an unrealistic pile that wouldn't match a real stack (Amossen & Pisinger, 2010; Castro & Grossmann, 2012; Elhedhli et al., 2017; Jeong, 2016; Lin et al., 2016; Martello et al., 2007; Morabito & Arenalest, 1994; Silva, Oliveira, & Wäscher, 2016)

Orientation: When placing a box in a pallet, it is possible to place it in six different ways as exemplified in Figure 7, which correspond to each face of the small object lying in the surface of the pallet. For this, it is assumed that the edges of the pallet are parallel to the edges of the (rectangular) boxes, which means that each small object can only rotate 90° when deciding its placement. This would be translated into a parallelism relation between boxes, which minimizes trim-loss. Such an effect could be used in other polygon-shaped boxes by adding a parallelism constraint between small objects. One common situation that showcases this constraint is when some of the boxes to be placed have a “This way up” sign, which restrains the orientation in a vertical way (Bischoff, EE; Ratcliff, 1995; Bortfeldt & Wäscher, 2013; Elhedhli et al., 2017; Hifi, 2002; Jeong, 2016; Lin et al., 2016; Silva et al., 2016).

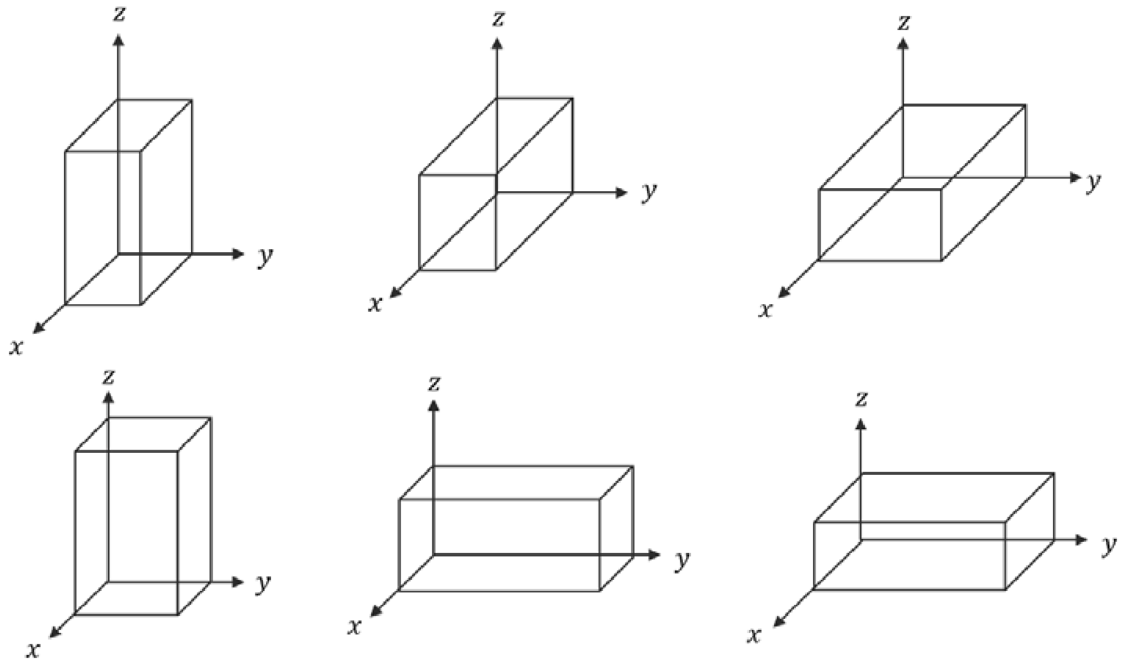

Figure 7: Different possible box placement orientations (Jeong, 2016)

LIFO (Last In First Out)/priority constraint: When a pallet has a pile with boxes that won’t be unloaded in the same destination, it is easier for the working personnel if the stack is organized in such way that the boxes meant to be unloaded at the first destinations are prioritized in the unload operation. That can be done by starting the pallet loading with the placement of the boxes to be unloaded last and progressively do this process by destination priority. In the end, the boxes meant to be unloaded in the first destination will be on top of the stack, while the ones with latter destinations will be either close or part of the base of the pile. This action will optimize the unloading operation in the multiple destinations and assure that the pile won’t need any type of rearrangement while unpacking the boxes (Bischoff, EE; Ratcliff, 1995; Bortfeldt & Wäscher, 2013; Elhedhli et al., 2017; Ramos et al., 2016).

Physical (maximum reach): If the pallet loading/unloading is done manually, the pile must be suited not to cause any trouble to the operator. In a stack made of small and light boxes, the unload operation may not cause any problems but in case heavy or big packages exist in the pile then the safety of the worker who oversees loading/unloading the pallet may be compromised due to physical limitations, such as the constant repetition of movements, fatigue, and others. This aspect can be controlled by utilizing some of the approaches to be referred next. One is to pile the boxes by placing the heaviest ones near the base and the lightest closer to the top of the stack. This would not only make the base of the pile stable but also increases the safety for the worker to unload the boxes, since the heaviest packages are the ones easier to reach and pick up. Another approach is to load the boxes in a way that they are easy to reach and identify during the unload process. That can be achieved by paying attention to the average size of a worker added to the length of the arms. If the pile is too big for that combination of measurements, then it isn't possible to reach the upper boxes. One possibility would be to lift multiple boxes at once, including the unreachable ones but that could cause the cargo to fall and damage the goods and injure the worker, so that isn't advisable (Ramos et al., 2016).

Multi-drop: This situation refers to when a pallet is carrying cargo whose constituents don't have all the same destination. For a better unload operation, it is desirable to place the boxes with the same destination close to each other in a way that avoids forcing the re-load of other boxes due to having some that are meant to unload in later destinations on top or in front of the small items meant to be unloaded earlier.

In the literature, there's also a constraint called zero unloading cost. Unloading cost refers to the effort put into unloading boxes. This concept is referred to in the context of home delivery services and is considered as an extreme case of the multi-drop situation. The unloading cost of a box is proportional to the number of boxes that haven't been delivered yet that need to be unloaded and reloaded to enable the unloading of the required box. To calculate this value, the "invisible and untouchable" rule is used. If the box to be unloaded has another small item positioned in front of it or in a place where the person can't touch it, then the unloading cost isn't zero.

In a pallet loading problem, position-related constraints may also exist. There are two types of position constraints, relative and absolute position. Relative positioning occurs between the locations of small objects. One example of this is when different boxes with food and perfumes are loaded on the same pallet. It is advisable that those items are separated in the stack. Absolute positioning is the position of the small items on the pallet. The LIFO constraint referred earlier is one example of this constraint. In the end, multi-drop constraint is basically a mix of absolute and relative positioning constraints (Bischoff, EE; Ratcliff, 1995; Bortfeldt & Wäscher, 2013; Lin et al., 2016).

Complete shipment: In some orders, the product is divided into multiple boxes, where each box contains different parts. The complete shipment constraint forces the placement of the boxes whose components belong to the same assembly in the same shipment to assure they end up in the same destination (Bischoff, EE; Ratcliff, 1995; Elhedhli et al., 2017; Jeong, 2016).

THESIS DEVELOPMENT

3 THESIS DEVELOPMENT

3.1 RESEARCH METHODOLOGY

This section aims to describe all the methods used in the research made towards the analysis of the complexity in the distributor pallet loading problem, including all the techniques used and what goals are meant to be achieved with them, which will be detailed in the following paragraphs.

In chapter 2, the literature about the complexity in the pallet loading problem was reviewed, with many authors having different considerations about this problem. But how do these perspectives fare with the perspectives of workers who have to face this problem in a daily basis? For that, a 3PL company was visited in a first instance to communicate with the workers who load pallets with boxes in a daily basis to see what they consider as complex in the pallet loading process. After learning that information, a set of metrics is created to replicate the complexity of the pallet loading problem. These consist in certain parameters related to the process and to the items used that add complexity to the packing process. A scale is created as well to identify the different complexity levels added by each parameter.

With the metrics created, the same 3PL company was visited in a second instance to obtain samples, which consist in measuring the complexity of multiple loading processes according to the set of metrics created. After each worker finishes loading its pallet, they would evaluate the complexity of the loaded pallet according to the scale created. Then, the remaining parameters would be obtained after the generation of an Excel file, containing the information about each observation. Depending on the content of that file, the metrics would eventually need adjustments, to improve the way they adapt to the information available. These observations are needed for statistical purposes, to be described next.

Having enough observations, to learn more about each para meter, a PCA (principal component analysis) is used to showcase which parameters are the most important when it comes to complexity of the pallet loading problem. At last, a multiple linear regression is used to determine a mathematical model that can predict the complexity of any given loaded pallet. The next sections will exhibit these methods in a more detailed way.

3.1.1 THE PRINCIPAL COMPONENTS ANALYSIS METHOD

The principal component analysis is a factorial analysis technique that transforms a set of correlated variables into a smaller number of independent variables, which correspond to a combination of the original ones. These new variables are known as principal components (Maroco, 2014, p. 459). This technique is often seen as a data reduction method, but another main characteristic of this method is the reduction of the presented information into a smaller number of variables, the principal components, who represent the most relevant information contained in the original variables. The principal components will be sorted by highest to lowest in terms of importance. That importance is translated into a higher variance rate of the collected data. The first component explains the highest variance rate, the second explains a rate not explained by the previous one, and so on. The most irrelevant component is the one that contributes the least to the total variance of the data (Rocha, 2015).

To validate the usage of the PCA in a set of variables, some authors propose different rules for data validation. Costa, Pereira, & Lopes (2018) refer multiple authors who recommend a minimum of 100 observations. Others, such as MacCallum, Widaman, Preacher, & Hong (2001), apply a rule that suggests a minimum of 5 observations per variable, which means that, if there are 6 variables being analyses, then at least 30 observations must be retrieved. Another rule that was created by Moreira (2007) which refers that, if the number of variables to analyse is 5 or less, then 50 observations must be retrieved. In case there's more than 5 and up until 15, then the minimum number of observations must be equal to the number of variables multiplied by 10. If there are more than 15 variables, the author applies the rule referred above, where there must be at least 5 observations per variable.

The previous paragraph only mentions data validation methods to effectively run a PCA, but there are other techniques that can be used to check how consistent the number of observations possessed is. One of them is Alpha's Cronbach value. This value has a range from 0 to 1 and evaluates the reliability of the data obtained. George & Mallery (2003) created a rule of thumb with classifications that can be viewed in Table 1. Although generally in the literature most authors only consider as acceptable a value above 0,7, only a value below 0,5 can be considered as unacceptable, being the in-between values questionable, requiring further analysis to see if the data is adequate or not. It is worth noting that this value is sensitive to the correlation and number of variables tested. Having few variables may cause a low alpha score, as well as low correlation between them. One cause for low correlation is having variables that measure irrelevant data. These issues can cause a test to be wrongly discarded, therefore the Cronbach's Alpha must be carefully used (Maroco, 2014).

Table 1: Classification for the Cronbach's alpha values (George & Mallery, 2003)

Cronbach's alpha	Internal consistency
$0{,}9 \leq \alpha$	Excellent
$0{,}8 < \alpha \leq 0{,}9$	Good
$0{,}7 < \alpha \leq 0{,}8$	Acceptable
$0{,}6 < \alpha \leq 0{,}7$	Questionable
$0{,}5 < \alpha \leq 0{,}6$	Poor
$\alpha < 0{,}5$	Unacceptable

There's also the KMO (Kaiser-Meyer-Olkin) test, which measures how suitable the obtained data is to be used in a factorial analysis. The test checks how adequate each individual variable is as well as the whole grouping of variables together. Like Cronbach's Alpha, the output of this test ranges from 0 to 1. In Table 2, it is possible to see the classifications of the ranges of values that are obtained from this test. According to (Costa et al., 2018), values above 0,8 are considered as adequate for factor analysis, while values below 0,6 show that the samples used aren't good enough to be used. The intermediate values not mentioned above indicate that the set of variables has average quality in terms of factor analysis usage.

Table 2: Classification for the KMO values (Costa et al., 2018)

KMO	Classification
≥ 0,90	Very good
Between 0,80 and 0,89	Good
Between 0,70 and 0,79	Average
Between 0,60 and 0,69	Reasonable
Between 0,50 and 0,59	Bad
< 0,50	Unacceptable

Along with the KMO test, Bartlett's test of sphericity is usually applied before the PCA. The purpose of this test is to test the hypothesis that the correlation matrix obtained from the data retrieved is an identity matrix, which possesses a diagonal of 1's and remaining values equalling 0. If the correlation values between all variables are close to zero, that means the variables are unrelated, which means they are not suited to be grouped into a principal component. If the hypothesis is rejected, that means the PCA may be useful when applied to the respective set of variables. (Maroco, 2014)

After executing the PCA, how many principal components should be retained? Although Maroco (2014) refers to four different methods to do that, only the following two will be used. First, there's the Kaiser criterion, which refers that factors that explain more variance than the variance that is explained by the original variables should be retained. The variances of the principal components are called eigenvalues. The second rule is the Scree plot criterion. A Scree plot is a graphical representation of the principal components (x-axis) and respective eigenvalues (y-azis). The number of components to retain equals the x value corresponding to the point from the graphic from where the slope becomes close to being horizontal. In Figure 8, although four components could be extracted by looking at the tendency of the line, the fourth component ends up being excluded because its eigenvalue is lower than one, so 3 principal components are retained there.

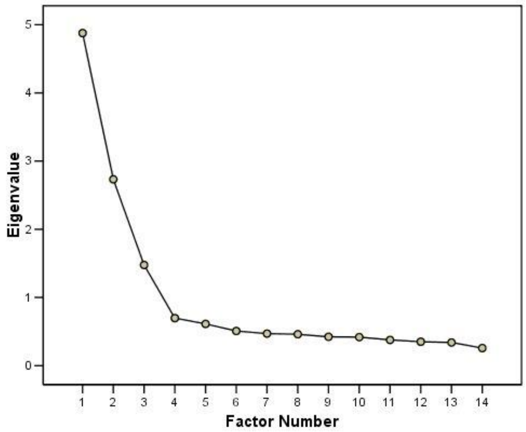

Figure 8: Example of a Scree plot (https://www.ibm.com/support/knowledgecenter/en/SSLVMB_24.0.0/spss/tutorials/fac_telco_kmo_01.html)

To sum up, each principal component is an exact linear combination (a pondered sum) of the original variables (Lattin, Carroll, & Green, 2011). According to RIBAS & Vieira (2011), each linear function is similar to the multiple linear regression, present in the next section. The weights of the principal components are useful to indicate how much variance in each of the original variables is explained by the principal components, and the square of the correlation coefficient represents the variance explaining power of the variable. (Lattin et al., 2011)

3.1.2 THE MULTIPLE LINEAR REGRESSION METHOD

The multiple linear regression is a mathematical analysis method where, given a set of data with independent variables (known as predictors) and one dependent variable (known as a criterion), the independent ones will be used to predict the value of the dependent variable (Maroco, 2014). The independent variables can be quantitative (e.g., salary income) or qualitative (e.g. ethnic group). After using this method, an equation will be generated. The values obtained are adjusted to the set of variables that were used, therefore, after applying the equation, it is expected to obtain similar values to the dependent variable. This relation between both types of variables can be seen in the equation 1. The dependent variable, the y, equals the sum of the products between each coefficient, β_i , and the respective observations related to each variable, the x. To refine the model, there's also the β_0, the y-intercept, and also the model's residuals, as e_i. (Lynn, 2007)

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + e_i \qquad (1)$$

Considering that this method considers all observations, it is important to look for outliers that may cause an incorrect adjustment to the retrieved data. With the outliers affecting the creation of the model, when comparing its predictions to the original dependent variable's value, it is expectable that the differences between both types of values in each observation end up being higher than without the outlier. Although the model will always showcase differences between the values of the original and predicted variables, it is imperative to look for outliers so that the model obtained has the best accuracy as possible. In Figure 9, the differences between a model with and without an outlier can be seen with more detail.

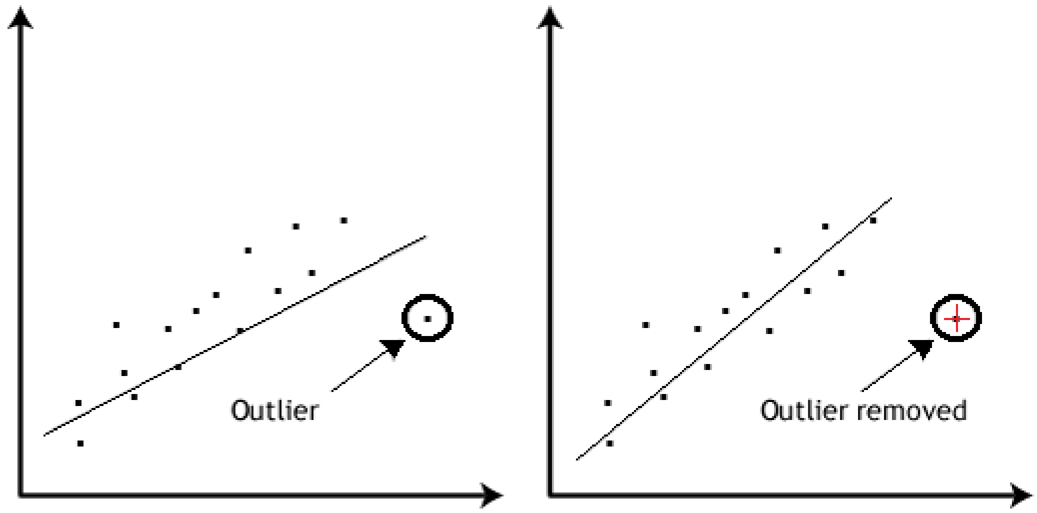

Figure 9: Scree plots showcasing the effects of an outlier (adapted from (Data Science Group, 2014))

3.2 COMPANY CHARACTERISATION

For a more practical insight on the pallet loading problem, the company known as Luís Simões was made available to help understand how the processes related to the loading of pallets work and what are the biggest difficulties that hinder the picking operations the most. They provided a tour in one of their warehouses where part of the working staff would explain all the procedures while others would execute the operations simultaneously.

Luís Simões is a company created in Portugal that operates in the Iberian Peninsula. It exists in the market for almost 70 years and is one of the best logistics-related companies in the country. It has around 2000 workers and 70 facilities built in the main regions of Portugal and Spain oriented for different purposes, such as transportation and cross-docking. ("*Alguns números que espelham a dimensão da Luís Simões*," n.d.)

The visited warehouse is located near the Leixões Port, which is a strategic location to facilitate the transportation of the containers filled with cargo into the large ships. It possesses over 30 docks prepared to quickly load the packed pallets into the containers attached to the trucks.

The warehouse has different sections to accommodate the different types of products, such as house cleaning products and drinks being packed in separate pallets. When an order is requested by one of their clients, a mechanism known as pallet shuttle is activated by a worker in order to drag a pallet filled with the wanted type of product into the edge of the rack. Then, with the aid of a forklift, the pallet is then lowered and transported to the appropriate place. If the pallet is ready to be shipped, then it's transported to the correct dock near other packed pallets that will be inserted in the same truck. If it needs to be labelled then it is transported to a section where the labels are printed, attached to the pallet or to each type of product, depending on what was requested by the client and scanned to have information on everything that is contained by the pallet and will be shipped. In case that pallet isn't ready to be shipped, then it goes to a zone dedicated for picking, where workers can carry boxes between pallets until the client's requirements are fulfilled and the pallet is ready to be labelled and shipped.

The company has records of multiple things related to the processes and the products. Each pallet has an associated sheet of paper containing the codes of all products. That way, all the content packed in each pallet can be visualized more easily with no need to analyse the pile itself. When it comes to the processes, each action performed by the working personnel is recorded and stored in a database, with data related to the type of operation that was executed, who did it and when. There's also a list with all the specifications required by each client related to the way they want the pallet to be loaded, which may include the constraints referred in a previous section. After learning how the warehousing system functions and interacting with the workers to learn their biggest struggles while packing pallets, it will be possible to build accurate metrics to estimate how difficult it is to pack a pallet according to certain specifications. In the following pages, those metrics will be explained in detail.

3.3 METRICS TO ANALYSE THE PALLET PACKING PROBLEM

Packing a pallet manually, although it may seem linear to some, it involves more variables than expected. From the weight of the boxes to the difference of heights between the pile of boxes and the worker, there are multiple aspects that may cause trouble for the employee when loading the boxes. Each parameter will be evaluated by using a scale that ranges from 1 to 10. For example, a parameter that is evaluated with a 1 out of 10 would mean that it doesn't add almost any complexity to the packing of the pallet, while a 10 would mean that such parameter causes massive trouble in the pallet loading problem. Although having one parameter classified with a 10/10 would increase the complexity of the loading process, that doesn't mean the whole process would be difficult overall, since the remaining parameters could be easy to deal with. Because of that, a similar classification will be given as well to the whole pallet packing. Each parameter will also have a different weight in the classification because they don't add the same amount of difficulty between themselves. Some can also get a score of 0, which means they don't apply to that loading process. These metrics take into consideration situations where only one person is working on one single pallet.

After explaining how this metric system works, the following paragraphs will be dedicated to classifying and detailing each parameter according to what was written earlier in this chapter.

Number of boxes: The first variable to be classified is the number of boxes to be packed on a pallet. Independently of how big and heavy each box is, the higher the number of boxes being packed is, the more time consuming it is to arrange them properly in the pile. Even though this variable solely counts the number of boxes to be packed, it can be said that the average size of the boxes being packed is related to this parameter. That's because, if the boxes to pack are huge, a much smaller number of boxes would fit the pile due to size restrictions. The same logic applies to small boxes. If the boxes to be packed are mostly small, then the pile could have dozens of boxes, making the loading operation more time-consuming. Due to this, the best way to measure the difficulty of this parameter is to check the number of boxes to be loaded. The score increases along with the increasing number of boxes to pack. The maximum difficulty attributed is set at 80 boxes or more, since it's very rare for a pallet to take so many boxes at once.

Average box weight: The second parameter to rate is the average box weight. Here, the idea is to measure the accumulated fatigue of the worker caused by the weight of the boxes. The higher the measured average weight is, the higher it goes on the scale, being a 10/10 score given to an average weight of 15 kg or more. This weight is used in the maximum rate of the scale because, according to (K-State Reseach and Extention, 2007), a worker shouldn't carry something weighting above 23 kg (without considering bending the back or knees or stretching the arms), due to a high risk of injury. Therefore, as a safety margin, 15 kg was considered instead of 23 kg. One particularity of this metric is that, hypothetically, 5 boxes weighing 10 kg each or 50 with the same weight would have the same rating. Although this wouldn't make much sense in a practical way, since packing 50 boxes with an average of 10 kg would add more complexity than 5 boxes with the same weight, this problem is eliminated by pairing this variable with the parameter mentioned previously since that's the one dealing with the number of boxes. Both are considered separately but rating the average box weight alone wouldn't make sense if the number of boxes to pack wasn't measured as well.

Percentage of fragile boxes to pack: The next variable to measure is the percentage of fragile boxes to pack. Placing this type of box in a pile is usually more restrictive than other boxes because it may have a specific orientation to be placed or may disallow the placement of other boxes on top of them. Therefore, transporting and placing one of these may require special attention, taking more time than most boxes to pack. It is considered maximum difficulty to pack over 50% of fragile boxes to pack in a pile. Just like the previous variable, this one only makes sense by pairing it with the parameter that covers the total amount of boxes to pack. In case there are no fragile boxes to pack, this variable is rated with a 1.

Average maximum box width: The fourth variable represents the average maximum box width. This concept refers to the maximum length available for the worker to pick and carry the box. Some boxes may have a huge difference of size between its own measures. The best example of a box that would have a high score of difficulty would be one with a similar size of a pallet, for example, 160x160x20 cm. If a worker tries to pick it up using the smallest length it can be a major struggle because it would tend to fall, due to the centre of mass being far from the person's hands, which are supporting its weight. Due to this, the best way to carry this type of box would be by picking it up in a way that the smallest base is perpendicular to the floor and the lengthiest part of the box is parallel to the worker's body, making it stable for transportation. The scale created for this metric is evenly separated, in 18 cm. The maximum difficulty would be something close to the size of an adult person. That's because, according to (Johnson, 2005), the arm span of an individual, when compared to its height, has a ratio of approximately 1.

Number of box types to label: Another parameter considered is the number of box types to label. The company uses a label system with a unique barcode for each type of box being packed so that, when the packed pallet reaches its destination, the clients scan the code to check what was received. To measure the difficulty of this variable, it is used the number of types of boxes to do it. Although it doesn't add much complexity to the process, it can be very time consuming according to the workers in case they need to label many types of boxes.

Number of box types to separate in the pile: The next parameter to measure is considered as one of the most complex by the working personnel at Luis Simões's warehouse. Some clients request the separation of each type of product on the pallet. That is done by storing each type of product in a segment of the surface area of a pallet, making multiple piles with variable heights. This situation can be seen in Figure 10. What makes this type of arrangement so complicated for the workers is the fact that, the more types of boxes there are to separate, the harder it is to place them in a way that the whole pile is stable enough. Since the packed pallet would be constituted by other smaller piles of products, the format of the full pile could end up being irregular. Therefore, assuring that the positioning of each product is good enough to avoid significant balancing of the boxes is particularly important in a situation like this. The maximum difficulty in this scenario is having 10 or more types of boxes to pack in the pallet. Just like the scenario where the percentage of fragile boxes is measured, this parameter may not be applied to a certain pallet packing plan because it doesn't always occur.

Figure 10: Pallet packed with each type of product in its own stack (https://www.koleimports.com/media/catalog/product/cache/1/image/360x360/9df78eab33525d08d6e5fb8d2713 6e95/Z/S/ZS025.jpg)

Number of upper boxes to make a surface for another packed pallet: When multiple pallets are packed and intended to go in the same shipment, to save some space in the truck's container, there's the possibility to stack a packed pallet on top of another one so that the container gets filled with more packed pallets. This can only apply under two conditions: first, the stack of pallets can't exceed the height limits, which may be imposed by a client or by the container's own dimensions and second, the upper packed pallet must have a stable pile as well as its surface, otherwise it may not be safe to do this operation. To measure this parameter, it is used the number of boxes in the upper layer of the pile, since they must be arranged in a more careful way to make sure the stacked packed pallet doesn't fall. The maximum difficulty for this task is considered as having over 9 boxes on the pallet.

Height difference between worker and pile: One of the major struggles of the manual pallet loading process is the maximum reach of the worker to the top of the pile. If the worker is short and the pile is big, that person will struggle to place the final boxes in the upper layer of the pile. This scenario represents the one with the highest difficulty. Numerically, this parameter is measured by the difference of heights between the worker and the pile. The maximum difficulty is represented by a difference of at least 35 cm between the worker and the pile, representative of the scenario mentioned earlier.

Number of heavy boxes to pack: The final parameter to be measured is the number of heavy boxes to pack. For this metric, a heavy box is considered as weighing at least 8 kg. Although as previously explained about the parameter that measures the average weight of the boxes, having a special focus on particularly heavy boxes is crucial because they may cause physical issues to the workers who lift them if they aren't careful. The Figure 11 showcases some examples of postures that can cause injuries during the transport and placement of the boxes. It is recommendable to bend the knees, lift the box and keep it around waist level to minimize the risk of injury. Bending the back while attempting to pick up the box, holding it without placing the hands firmly in its base and extremities, and lift such weight above shoulder level are must-avoid situations for any worker involved (K-State Reseach and Extention, 2007). Sometimes it is needed to place heavy boxes in a very low or very high spot, so under those circumstances, it may be better to get more workers to help in order to decrease the effort applied to such difficult task. A 10/10 score would be given to this parameter if the pallet is packed with 18 or more heavy boxes. If the pallet isn't packed with boxes at least as heavy as 8 kg, then this variable is classified with a 1 out of 10.

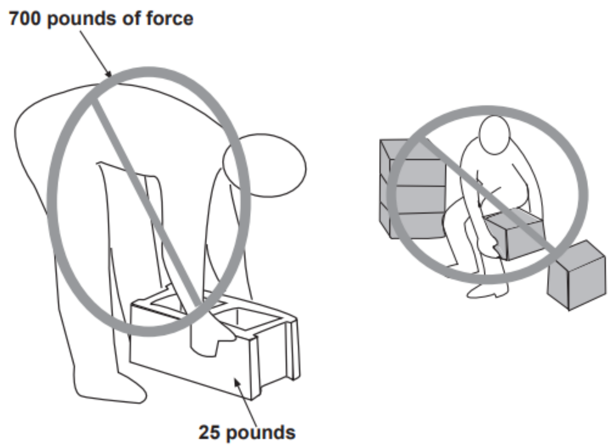

Figure 11: Some postures to avoid when lifting boxes (700 pounds is approximately 318 kg, and 25 pounds equals 11 kg) ((K-State Reseach and Extention, 2007))

After showcasing all variables that can increase the complexity of the pallet packing process, all the information referred above can be seen with more detail in Table 3. The percentages are all the same by default (the cumulative total will be equal to 1), but that will change after collecting samples from the warehouses where data about packed pallets will be gathered to balance the accuracy of those values. The quantities classified by each parameter on the image may also change, since they may not be accurate as well.

Table 3: Parameters and scale of the created set of metrics

Parameter\scores	1	2	3	4	5	6	7	8	9	10	Units
Number of boxes	≤ 5	6 to 10	11 to 15	16 to 20	21 to 30	31 to 40	41 to 50	51 to 60	61 to 80	> 80	
Average box weight	≤ 1,5	1,6 to 3	3 to 4,5	4,6 to 6	6 to 7,5	7,6 to 9	9 to 10,5	10,5 to 12	12 to 15	> 15	kg
Percentage of fragile boxes	≤ 5	6 to 10	11 to 15	16 to 20	21 to 25	26 to 30	31 to 35	36 to 40	41 to 50	> 50	%
Average box maximum width	≤ 18	19 to 36	37 to 54	55 to 72	73 to 90	91 to 108	109 to 126	127 to 144	145 to 162	> 162	cm
Number of box types to label	1 or 2	3 or 4	5 or 6	7 or 8	9 or 10	11 or 12	13 to 15	16 to 18	19 or 20	> 20	
Number of box types to separate	--	--	< 4	4	5	6	7	8	9	> 9	
No. of upper boxes to make a base	≤ 10	11 to 15	16 to 20	21 to 25	26 to 30	31 to 35	36 to 40	41 to 45	46 to 50	> 50	
Worker and pile height difference	**	***	*** to 5	6 to 10	11 to 15	16 to 20	21 to 25	26 to 30	30 to 35	> 35	cm
No. of heavy boxes to pack*	1 or 2	3 or 4	5 or 6	7 or 8	9 or 10	11 or 12	13 or 14	15 or 16	17 or 18	> 18	

* heavy box is at least 8 kg ** top of the pile below waist level *** top of the pile between waist level and worker's height

3.4 SAMPLE RETRIEVAL

With the metrics set, the next step is to apply these in real life situations to gather enough data that allows the weighing of each metric to increase the metric system's accuracy. For this, the previously referred company Luís Simões made itself available once more.

The data gathering process is the following: each worker, after finishing the loading process of a pallet, will print a sticker with a barcode and other numbers to ship alongside the packed pallet. Two of them help to differentiate each packed pallet and each order. At the end of the day, the software system would create an Excel file with all the data collected by the workers. During the box loading process, each box is scanned so that the Excel file contains information such as the time the box was picked, the quantity and its dimensions. After gathering the previously referred pair of numbers from multiple packed pallets, each picker would rate the complexity of the pallet they packed from 1 to 10.

Although most of the data needed to extract and measure according to the developed metrics can be easily obtained in that Excel file, some could only be attained via indirect methods and some couldn't be obtained at all. Due to this, some changes in the metrics were made, which will be detailed in the following paragraphs.

Before explaining how each parameter was obtained, first, it is important to see how the information is displayed in the given Excel file. In Figure 12, it is possible to see an example of what is stored in the file.

Date	Guia	Category	Place	Family	Article	Worker numb	Amount of boxes	Picking time	UT1	UT2
20180730	1150212653	MMM	BB209	VVV	23132064	343434	5	201807301146	343476155	9456048
Type	**Amount**	**Width**	**Depth**	**Height**	**Weight**	**Type**	**Amount**	**Width**	**Depth**	**Height**
P AL	160	0,8	1,2	1,03	442,44	CRT	#N/D	#N/D	#N/D	#N/D
Descript.	**Weight**	**Type**	**Amount**	**Width**	**Depth**	**Height**				
Y	#N/D	.EMB	#N/D	#N/D	#N/D	#N/D				

Figure 12: Extractable data from a loaded pallet

In the first line, on the left, there's the date of the picking process. Then there's one of the numbers that showcase what order this pallet belongs to. This is one of the two numbers that were needed to identify each sample that was evaluated. There's also the category, place (which is the location inside the warehouse), and the family of products. The amount of boxes quantifies how many boxes are piled in the pallet in each wave. In this case, 5 boxes were placed on the pallet at the time mentioned in the picking time parameter. The UT1 and UT2 together is the second number needed to identify the sample. In the second line, there is a group of duplicate parameters. The file contains data about the packed pallet where the boxes were taken from, the box and the package. It details the dimensions, the weight and how many boxes the pallet has and how many packages each box has. As the image shows, the box and package information doesn't always exist. At last, there's the "descript." parameter that shows the client that requires these products. Note that, although some numbers are representative of a real sample, some were altered to hide confidential information about the clients and products.

Now that is known the information that the Excel file contains, it was time to extract the needed data to evaluate each sample. To start, the number of boxes was extracted. As explained in this chapter, it is recorded how many boxes are loaded in each instance. After summing up the number of boxes loaded in each wave, the total number of boxes the pallet carries is obtained. The weight of the boxes needed to quantify the average box weight contained in the pile, as well as the number of heavy boxes, is in most cases easily obtained using the previously mentioned number of boxes parameter. In certain cases, the weight of some boxes had to be calculated by using the information about the pallet where the box was taken from. Knowing the number of boxes that the pallet contains and the total weight (pallet + boxes), since the pallet has a weight of 25 kg, it is possible to subtract such number to the total weight of the packed pallet and then divide that subtraction by the total number of boxes to obtain the weight per box.

When it comes to the box fragility, such information isn't obtainable directly. To distinguish fragile boxes from non-fragile boxes, each box had to be analysed one by one according to certain characteristics inherent to the materials of the product and the box itself. Four different ratings were given to each box. For a box to be considered non-fragile, both the product and the box had to be consistent and resistant. If the box itself had signs that it could rupture without damaging the goods, then it would be rated differently from the previous ones. For example, in some card boxes, there's a part where the cardboard is taped. If a significant area of the tape isn't attached to the cardboard itself, then it could be a fragile point from where the tape could be ripped off by an edge of a heavy box and slightly damage the package. Although this can be a problem under certain circumstances, it isn't significant enough compared to the other two scale units whose boxes are considered as fragile. In products where there's no cardboard revolving the product, if the upper area of contact is too small or is made by non-resistant materials, they can be considered as fragile.

Not all products are contained inside of cardboard boxes, so when the product is in direct contact with other boxes, the recipient must possess enough resistance in all points of contact not to ruin the product. The final fragility scale is when both the box and the product can be easily destroyed if a heavy box is placed on top of them. In this analysis, the fragility of the boxes was always taken into consideration how the analysed boxes would fare with heavy boxes. That's because some packages considered as fragile could take loads of boxes on top of them but that doesn't mean they would damage the lower boxes since the cumulative weight could be low enough not to cause any damage. Some boxes possess a symbol indicative of how many similar boxes it can handle in a stack, so that should be considered during the loading process.

Next, to evaluate the average maximum box width, between the width and depth of the boxes, the highest value of each box was used to calculate that average value, in a similar process when compared to the calculation of the average box weight. It is worth noting that this metric was changed. Instead of using centimetres as a unit, due to the Excel file possessing the measurements in meters, the metric was changed to meters to reflect the way the information was given. This and other changes will be exhibited in upcoming paragraphs, as well as an image comparing the old and the updated metrics.

The number of box types to label was also changed. Although the scale remains the same, it was changed to count the number of box types. That's because, although the workers have to label each pallet after being fully packed and as a request of certain clients, label each type of product as well, having a metric to quantify the heterogeneity of boxes in the pallet is more important than checking how many products are labelled, since that process only occurs under rare circumstances and the effort spent doing it mostly affects time (which is a newly created metric that will be explained in later paragraphs) than any other parameter. To quantify the number of box types packed into the pallet, the article number is used, and each different number equals to one box type. Some boxes look identical in terms of dimensions and aspect, but there are slight differences that the workers must distinguish so that the order is sent with the correct quantities of each product.

With the aid of the previous metric, it was possible to quantify the number of box types to separate in the pile, as in each product piled in a different stack within the pallet. First, a list was given of each client that requires this type of load. In samples with clients that didn't request such packing type, the variable is quantified as 0. Otherwise, the number of boxes types is equal to the number of box types to separate in the pile. Although they may be equal, each metric is scaled differently, which means those two parameters will have different ratings.

About the parameter that measures the number of upper boxes that will serve as a base to stack another packed pallet, due to the fact that this information isn't collected by the company and is also dependent on the methods applied by each worker to pile the boxes, this metric was scrapped. Instead, a new metric was created. This metric evaluates the time spent fully packing a pallet. Since the difficulty of packing a pallet is related to how long the loading process lasts, it makes sense to implement such a metric. The registered times of the first and last wave of boxes loaded were subtracted to obtain the time spent loading the pallet.

Last, to measure the height difference between the worker and the packed pallet, each worker would tell how high they are. Then, since the Excel file given doesn't directly have how high the packed pallet is, indirect methods were applied. Knowing the area of the pallet and each box, they were successively subtracted until the pallet area remaining was almost 0. Then another layer would be calculated. Before applying this method, the boxes were sorted by weight, since heavier boxes are usually placed in the base of the pallet. When all the boxes were packed, then the tallest box of each layer would be summed and the result would be the total height of the packed pallet. With this method, the more boxes the pallet has, the more inaccurate this method can be. Despite this issue, it is accurate enough for this purpose. After obtaining the height of the pallet and loaded boxes, this value is subtracted to the height of the worker who loaded the boxes there.

Before showcasing the values obtained for each sample, in Table 5, there's an overview of the updated metrics. The changes are highlighted in yellow. As seen below, the metric that evaluates the time spent packing a pallet is scaled in intervals of 5 minutes. That's because the most complex samples retrieved lasted nearly an hour to pack, matching that timing to the maximum rating of the scale.

In annexes I and II the values calculated and respective evaluations are shown. The last column of values showcases the evaluation given by each picker to the loading process of the pallet they loaded. The metrics in those tables are represented by numbers, which are decoded in Table 4.

With the 38 samples evaluated, the next step is to use this data in a statistical software known as SPSS (version 23). To know the importance of each parameter in the whole evaluation and reduce the number of variables obtained, a PCA (principal components analysis) will be done as well as a multiple linear regression. If the results aren't acceptable, more samples would be needed to complement the ones already showcased. Such information is detailed below.

Table 4: Variable decoding

Variable	Code
Packed pallet evaluation	0
Number of boxes to pack	1
Average box weight	2
Percentage of fragile boxes	3
Number of box types	4
Average maximum width	5
Number of column piles	6
Time spent loading a pallet	7
Height difference between worker and pile	8
Number of heavy boxes	9

Table 5: Updated set of metrics

Parameter\Scores	1	2	3	4	5	6	7	8	9	10	Units
Number of boxes	≤ 5	6 to 10	11 to 15	16 to 20	21 to 30	31 to 40	41 to 50	51 to 60	61 to 80	> 80	
Average box weight	≤ 1,5	1,6 to 3	3 to 4,5	4,6 to 6	6 to 7,5	7,6 to 9	9 to 10,5	10,5 to 12	12 to 15	> 15	kg
Percentage of fragile boxes	≤ 5	6 to 10	11 to 15	16 to 20	21 to 25	26 to 30	31 to 35	36 to 40	41 to 50	> 50	%
Average box maximum width	≤ 18	19 to 36	37 to 54	55 to 72	73 to 90	91 to 108	109 to 126	127 to 144	145 to 162	> 162	$x10^{-2}$ m
Number of box types	1 or 2	3 or 4	5 or 6	7 or 8	9 or 10	11 or 12	13 to 15	16 to 18	19 or 20	> 20	
Number of box types to pack in columns	--	--	< 4	4	5	6	7	8	9	> 9	
Time spent loading a pallet	≤ 5	6 to 10	11 to 15	16 to 20	21 to 25	26 to 30	31 to 35	36 to 40	41 to 45	> 45	mins
Worker and pile height difference	**	***	*** to 5	6 to 10	11 to 15	16 to 20	21 to 25	26 to 30	30 to 35	> 35	$x10^{-2}$ m
No. of heavy boxes to pack*	1 or 2	3 or 4	5 or 6	7 or 8	9 or 10	11 or 12	13 or 14	15 or 16	17 or 18	> 18	

* heavy box is at least 8 kg ** top of the pile below waist level *** top of the pile between waist level and worker's height

3.5 DATA PROCESSING

3.5.1 PRINCIPAL COMPONENTS ANALYSIS

In this section, the software known as SPSS (version 23) will be used to run statistical tests on the data retrieved. As referred in section 3.1.1, most authors recommend a minimum of 45 observations, but only 38 were obtained, which may affect the quality of the results. As previously referred, a PCA will be used to learn the importance of the gathered data in the pallet packing problem. Then, a multiple linear regression will be used to create a model to make it possible to predict the complexity if a pallet cargo arrangement. In the following paragraphs, the values obtained from the usage of those statistical methods will be detailed.

The first procedure to do is to check the value of Cronbach's Alpha. This value, as seen in section 3.1.1, showcases the reliability of the set of observations. The Table 1 located in the aforementioned section shows how to rate the value obtained. After selecting the 9 independent variables, the value obtained can be seen in Table 6:

Table 6: Reliability statistics

Cronbach's Alpha	N of Items
0,568	9

A Cronbach's Alpha of 0,568 is rated as poor. Since this result isn't considered as unacceptable, although not acceptable as well, further testing will be done with the samples gathered, since, as explained in section 3.1.1, many tests are improperly rejected when a Cronbach's Alpha value below 0,7 is obtained.

Next, it will be done the PCA analysis. As seen in section 3.1.1, this statistical test reduces the number of variables, which are combinations of the previous ones and explain the most important information. Although the rules recommended by Costa et al. (2018) and other authors referred in chapter 3.1.1 imply that this sample isn't adequate for a PCA, this doesn't mean the data isn't suited for this type of test. One of the possible outputs while doing the PCA in SPSS is the KMO and Bartlett's test of sphericity. The values obtained can be checked below in Table 7.

Table 7: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,636
Bartlett's Test of Sphericity	Approx. Chi-Square	164,344
	df	36
	Sig.	0,000

The KMO value of 0,636 is classified as mediocre, as seen in Table 2, which means the data tested is good enough for a PCA, despite that the value should be a bit higher. The sigma value of Bartlett's test of sphericity is below 0,05, which means the tested set of variables is adequate for a PCA by showcasing that the correlations matrix, referred in the next paragraph, possesses significant correlations between the nine variables, rejecting the hypothesis that the correlations matrix is an identity matrix.

In the Tables 8 to 17 there are the frequency tables for the 10 variables used. In some cases, such as the ones in Tables 8 and 13, there's a large percentage of observations concentrated in the same value. In other cases, for example in Tables 14 and 17, the data is widely dispersed between the full scale, being the complete opposite of the situation previously mentioned.

Table 8: Frequency table for the Number of Boxes variable

Value	Frequency	Percent age	Cumulative %
1	1	2,6	2,6
4	1	2,6	5,3
5	3	7,9	13,2
6	5	13,2	26,3
7	2	5,3	31,6
8	2	5,3	36,8
9	1	2,6	39,5
10	23	60,5	100,0
Total	38	100,0	

Table 9: Frequency table for the Average Weight of Packed Boxes variable

Value	Frequency	Percent age	Cumulative %
2	19	50,0	50,0
3	5	13,2	63,2
4	5	13,2	76,3
6	1	2,6	78,9
7	2	5,3	84,2
8	1	2,6	86,8
9	2	5,3	92,1
10	3	7,9	100,0
Total	38	100,0	

Table 10: Frequency table for the Percentage of Fragile Boxes variable

Value	Frequency	Percent age	Cumulative %
1	16	42,1	42,1
3	4	10,5	52,6
4	3	7,9	60,5
5	2	5,3	65,8
6	1	2,6	68,4
8	2	5,3	73,7
9	2	5,3	78,9
10	8	21,1	100,0
Total	38	100,0	

Table 11: Frequency table for the Average Maximum Width variable

Value	Frequency	Percent age	Cumulative %
1	8	21,1	21,1
2	19	50,0	71,1
3	8	21,1	92,1
4	3	7,9	100,0
Total	38	100,0	

Table 12: Frequency table for the Number of Box Types variable

Value	Frequency	Percent age	Cumulative %
1	12	31,6	31,6
2	4	10,5	42,1
3	2	5,3	47,4
4	2	5,3	52,6
5	4	10,5	63,2
6	1	2,6	65,8
7	5	13,2	78,9
8	4	10,5	89,5
10	4	10,5	100,0
Total	38	100,0	

Table 13: Frequency table for the Number of Column Piles variable

Value	Frequency	Percent age	Cumulative %
1	21	55,3	55,3
4	1	2,6	57,9
6	1	2,6	60,5
7	1	2,6	63,2
9	1	2,6	65,8
10	13	34,2	100,0
Total	38	100,0	

Table 14: Frequency table for the Time Spent Packing variable

Value	Frequency	Percent age	Cumulative %
1	9	23,7	23,7
2	9	23,7	47,4
3	5	13,2	60,5
4	6	15,8	76,3
5	4	10,5	86,8
6	1	2,6	89,5
7	1	2,6	92,1
9	1	2,6	94,7
10	2	5,3	100,0
Total	38	100,0	

Table 15: Frequency table for the Height Difference Between Worker and Pile variable

Value	Frequency	Percent age	Cumulative %
1	7	18,4	18,4
2	14	36,8	55,3
4	1	2,6	57,9
5	2	5,3	63,2
9	1	2,6	65,8
10	13	34,2	100,0
Total	38	100,0	

Table 16: Frequency table for the Number of Heavy Boxes Packed variable

Value	Frequency	Percent age	Cumulative %
1	19	50,0	50,0
3	1	2,6	52,6
6	4	10,5	63,2
7	1	2,6	65,8
8	2	5,3	71,1
9	2	5,3	76,3
10	9	23,7	100,0
Total	38	100,0	

Table 17: Frequency table for the Packed Pallet Evaluation variable

Value	Frequency	Percent age	Cumulative %
1	1	2,6	23,7
2	4	10,5	47,4
3	6	15,8	60,5
4	6	15,8	76,3
5	4	10,5	86,8
6	7	18,4	89,5
7	4	10,5	92,1
8	4	10,5	94,7
9	2	5,3	100,0
Total	38	100,0	

The Table 18 indicates the descriptive statistics for all variables. It is worth highlighting the mean of the variable representing the number of boxes, which is much higher than the other variables, and the standard deviation of the number of column piles variable, indicating that there's a large data variation in the observation retrieved for this variable. The values used in Table 18 are the scaled ones and not the original measurements. The correspondence column shows the range of values that matches the values in the mean column. For example, a mean of 8,39 for the Number of Boxes variable translates into 51 to 60 boxes, according to the scale shown in Table 5.

Table 18: Descriptive statistics

Variable	Mean	Correspondence	Std. Deviation	N
0-Packed pallet evaluation	5,03		2,16	38
1-Number of boxes	8,39	51 to 60	2,33	38
2-Average weight of packed boxes	3,92	3 to 4,5 kg	2,74	38
3-Percentage of fragile boxes	4,47	16 to 20%	3,73	38
4-Average maximum width	2,16	19 to 36 cm	0,86	38
5-Number of box types	4,39	7 or 8	3,21	38
6-Number of column piles	4,66	4	4,28	38
7-Time spent packing	3,37	11 to 15 mins	2,44	38
8-Height difference between worker and pile	4,95	6 to 10 cm	3,96	38
9-Number of heavy boxes packed	4,66	7 or 8	3,98	38

The correlation matrix in Table 19 indicates the variables with high correlation amongst themselves. The positive values close to 1 indicate a high correlation. The values in bold highlight the pairs of variables with the highest correlation. The Table 4 located in the previous section decodes the numbers representing each variable. Also note that the matrix is symmetric, being the diagonal composed by 1's the symmetry line. These correlations are between the original variables, and not after the variable transformation in the PCA.

Table 19: Correlations matrix

Variable code	1	2	3	4	5	6	7	8	9	0
1	1,000									
2	**-0,525**	1,000								
3	0,379	-0,428	1,000							
4	**0,571**	-0,458	**0,516**	1,000						
5	-0,235	**0,526**	-0,261	-0,240	1,000					
6	0,371	-0,270	0,251	**0,771**	-0,140	1,000				
7	0,458	-0,299	0,262	**0,663**	-0,080	0,480	1,000			
8	0,456	-0,108	0,015	0,280	0,314	0,055	0,460	1,000		
9	-0,282	**0,692**	-0,192	-0,133	0,453	-0,044	0,072	0,230	1,000	
0	0,416	-0,118	0,166	-0,046	0,431	**0,591**	0,356	0,155	0,029	1,000

By observing the high correlations in bold in Table 19, it is possible to see that most of them revolve around the variables representing the number of box types and average weight of boxes. The latter one has a unique correlation with the number of boxes, because it is both high and negative. This indicates that a high average box weight is a cause to a lower number of boxes, which is a logical conclusion because heavy weight is often accompanied by large box dimensions (which is also corroborated with the high correlations between average box weight and average maximum width). With big boxes, a lesser number of boxes can fit into the pallet, as shown by these correlation values. The Evaluation variable has high correlation with the Number of Column Piles, which was one of the constraints that the box pickers referred to as one that adds complexity to the pallet loading process. Two other variables who have slightly significant correlations with the Evaluation one is the Number of Boxes and Number of Box Types variables. It is worth noting that these two correlations aren't as high as the ones in bold.

Looking at the communalities table, Table 20, it is possible to see the variances explained by the variables before and after the PCA, represented by the Initial and Extraction columns, respectively. Initially, the communalities are equal to 1, and after the extraction they have values between 0 and 1. Most communalities extracted from the principal components are high, which means that the principal components represent overall all variables well. One exception may be the percentage of fragile boxes, which has a value below 0,5, meaning that this is the least important variable.

Table 20: Communalities

	Initial	Extraction
1-Number of boxes	1,000	0,740
2-Average weight of packed boxes	1,000	0,848
3-Percentage of fragile boxes	1,000	0,407
4-Average maximum width	1,000	0,638
5-Number of box types	1,000	0,903
6-Number of column piles	1,000	0,796
7-Time spent packing	1,000	0,690
8-Height difference between worker and pile	1,000	0,907
9-Number of heavy boxes packed	1,000	0,785
Extraction Method: Principal Component Analysis.		

The Table 21 with the total variance showcases how each component explains part of the total variance in the whole data. It is possible to identify 3 principal components, where the main component explains 40% of the total variance. Together, all 3 components explain 75% of the total variance, which is an acceptable value.

Table 21: Eigenvalues and total variance

Component	Initial Eigenvalues		
	Total	% of variance	Cumulative %
1	3,641	40,453	40,453
2	2,003	22,250	62,704
3	1,070	11,892	74,596
4	0,731	8,127	82,723
5	0,567	6,296	89,019
6	0,431	4,786	93,805
7	0,276	3,066	96,872
8	0,171	1,902	98,774
9	0,110	1,226	100,000

In the scree plot displayed in Figure 13, the slope of the line doesn't have a drastic shift as seen in the example of chapter 3.1.1. Although the eigenvalue of component 3 is slightly above 1, the line doesn't have that inclination that separates the principal components from the non-principal components.

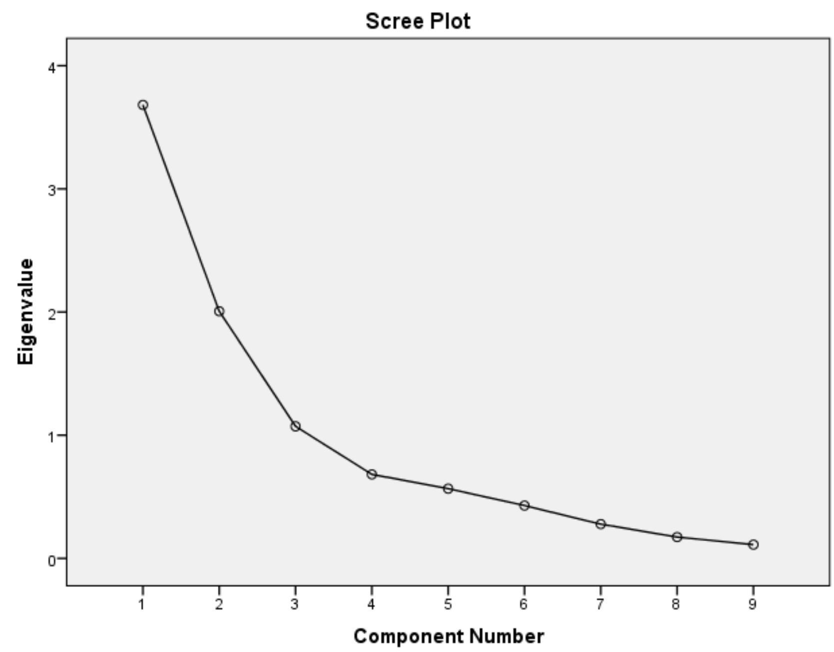

Figure 13: Scree plot obtained in the PCA

Finally, the principal components are seen below, in Table 22. The component 1, which explains 41% of the total variance, contains the number of column piles, number of box types, time spent packing, and percentage of fragile boxes. This component can be called as Box Quantities. The second component explains 22% of the total variance and is affected by the number of heavy boxes packed, average weight of packed boxes and average maximum width. This component can be named as Box Dimensions. The final principal component only covers 12% of the total variance and possesses the remaining two variables, the height difference between worker and pile and number of boxes. This component can be called Complexity.

Table 22: Rotated Component Matrix

	Component		
	1	2	3
6-Number of column piles	**0,891**	-0,023	-0,031
5-Number of box types	**0,890**	-0,213	0,256
7-Time spent packing	**0,666**	0,008	0,496
3-Percentage of fragile boxes	**0,497**	-0,397	0,053
9-Number of heavy boxes packed	0,088	**0,880**	0,048
2-Average weight of packed boxes	-0,247	**0,850**	-0,255
4-Average maximum width	-0,214	**0,713**	0,290
8-Height difference between worker and pile	0,072	0,184	**0,932**
1-Number of boxes	0,387	-0,443	**0,628**

The 3 principal components extracted according to the eigenvalue rule and scree plot explain 75% of the total variance of the original variables. To see if this percentage can increase, there's an alternative PCA that can be done, for categorical variables. Originally, the PCA was done with quantitative data, but if the type of variable is changed to "ordinal" in SPSS, this alternative PCA method can be used. Some of the outputs differ from the original PCA, but they will be explained throughout the next pages.

After running the categorical PCA in SPSS, setting the test for 9 dimensions, the following values were obtained, in Table 23. Note that dimension is the equivalent to a component. By observing the eigenvalues, it is possible to extract 3 principal components, just like the previous PCA. Together, they explain a similar variance percentage to the previous PCA analysis, around 75%. There's also the Cronbach's Alpha for each component, which was obtained according to the respective eigenvalues. The first two components have decent Alpha values, being the first one much higher due to explaining much more variance of the original values than the other ones. The third component has a very low Cronbach's Alpha, because the eigenvalue isn't that big as well as the variance percentage explained.

Table 23: Data extraction from the categorical PCA

Dimension	Cronbach's Alpha	Variance Accounted For	
		Total (Eigenvalue)	% of Variance
1	0,818	3,659	40,657
2	0,572	2,034	22,599
3	0,033	1,030	11,442
4	-0,456	0,711	7,905
5	-0,839	0,573	6,365
6	-1,423	0,442	4,907
7	-2,796	0,287	3,188
8	-5,967	0,159	1,762
9	-9,516	0,106	1,175
Total	1,000	9,000	100,000

Considering that 3 principal components were extracted, it is possible to run this PCA with the number of dimensions equalling the number of components extracted and see if significant changes occur. Below, in Table 24, there's the data extracted from this new test. Compared to the values obtained in Table 23, the eigenvalues and percentage of variance explained increased, meaning that the reliability of the 3 components increased as well. The third principal component still has a low Cronbach's Alpha but it's substantially better than the previous one.

Table 24: Rotated data extraction from the categorical PCA with 3 dimensions

Dimension	Cronbach's Alpha	Variance Accounted For	
		Total (Eigenvalue)	% of Variance
1	0,840	3,954	43,932
2	0,585	2,085	23,167
3	0,142	1,145	12,722
Total	0,968	7,184	79,822

In Table 25 it is possible to see the rotated component loadings for this situation. In each column, the variables in bold are the ones being represented by the respective component. When compared to Table 22, the loadings in the latter table are smaller than in Table 25. The percentage of variance explained is also smaller in Table 22 (75% to 80%), which indicates that this last categorical PCA is the one with higher quality.

In the end, 3 principal components were extracted, following the "eigenvalue > 1" rule, collectively explaining 80% of the total variance of the original variables. With the generation of the variables representing these 3 components, which are the Box Quantities, Box Dimensions and Complexity in the next section they will be used to run a multiple linear regression to see what component adds the most complexity in the pallet loading process.

Table 25: Rotated component loadings

	Component		
	1	2	3
5-Number of box types	**0,901**	-0,191	0,208
6-Number of column piles	**0,887**	0,004	-0,033
7-Time spent packing	**0,736**	-0,054	0,436
3-Percentage of fragile boxes	**0,711**	-0,266	-0,065
9-Number of heavy boxes packed	0,063	**0,890**	0,036
2-Average weight of packed boxes	-0,344	**0,868**	-0,136
4-Average maximum width	-0,181	**0,764**	0,325
8-Height difference between worker and pile	0,035	0,231	**0,932**
1-Number of boxes	0,446	-0,528	**0,617**
Cronbach's Alpha	0,840	0,585	0,142
% of variance	43,932	23,167	12,722

Considering that the third component explains a low percentage of the total variance, and its eigenvalue is very close to the rejection border, it is possible to force SPSS to redo the test while creating only two components. The results of that second test are shown in the tables below.

First, the Table 26 shows that the 2 dimensions have decent Cronbach's Alpha values, and together they explain around 69% of the total variance of the original variables. The next table,

Table 27, displays the two principal components, with the respective variables in bold (only variables with a non-negative component loading of at least 0,5 is determinant for that component). The first one possesses 5 variables (Number of Box Types, Time Spent Packing, Number of Column Piles, Number of Boxes, Percentage of Fragile Boxes) and can be called Box Quantities, and the second one possesses 4 variables (Average maximum width, Number of Heavy Boxes Packed, Average Weight of Packed Boxes and Height Difference Between Worker and Pile) and is known as Box Dimensions.

Table 26: Rotated data extraction for categorical PCA with 2 dimensions

Dimension	Cronbach's Alpha	Variance Accounted For	
		Total (Eigenvalue)	% of Variance
1	0,836	3,725	41,387
2	0,718	2,521	28,007
Total	0,945	6,245	69,394

Table 27: Rotated component loadings for 2 dimensions

	Component	
	1	2
5-Number of box types	**0,930**	-0,091
7-Time spent packing	**0,867**	0,144
6-Number of column piles	**0,791**	-0,042
1-Number of boxes	**0,773**	-0,286
3-Percentage of fragile boxes	**0,624**	-0,330
4-Average maximum width	-0,132	**0,827**
9-Number of heavy boxes packed	-0,133	**0,826**
2-Average weight of packed boxes	-0,535	**0,750**
8-Height difference between worker and pile	0,416	**0,608**
Rotated Cronbach's Alpha	0,836	0,718
% of variance	41,387	28,007

The Figure 14 shows the positioning of each observation in the two-dimensional map defined by the two principal components and their positioning relatively to the original variables, after running the categorical PCA. With the analysis of the map, it is possible to see that the observations 5 and 1 seem to be outliers, due to being far from the component loadings and the other observations. Since they don't seem to be severe outliers, and considering the sample size, they won't be removed from the database.

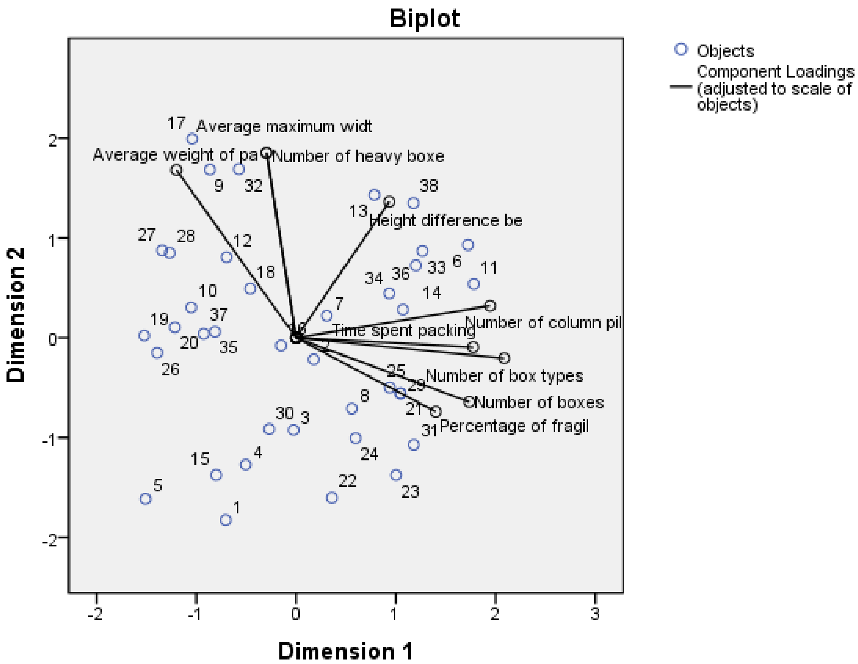

Figure 14: Biplot with the 2 principal components

3.5.2 MULTIPLE LINEAR REGRESSION

After using the PCA with the 9 collected independent variables, now a multiple linear regression will be done with the principal components extracted in the previous section. As seen in section 3.1.2, this method uses the observations to create a mathematical model that can be used to evaluate the complexity of a pallet cargo arrangement, and with that, it will be possible to conclude which principal components and variables are responsible for the addition of complexity, and if the results obtained match the perception of the workers about it.

First, a regression will be done using the 2 principal components previously extracted and the dependant variable, the Packed Pallet Evaluation variable. To do this multiple linear regression, the method chosen is called Stepwise. There are multiple iterations, starting with a model possessing only one variable and then progressively adding a new variable and removing others if they aren't significant enough to the model (Maroco, 2014).

The Tables 28, 29 and 30 show the outputs for the multiple linear regression. The Table 28 has the model summary and overall fit statistics. The adjusted R square value has a value of 0,202, meaning that this linear regression explains 20,2% of the variance in the data, which is a small value but adequate for the data. It is expected to see the R values increase with the increase of the number of variables inserted into the model.

The Table 29 shows the F-test, which tests the null hypothesis that the model explains zero variance in the dependent variable. The sig. value is below 0,05, which means the null hypothesis is rejected, meaning that this model explains a significant amount of the variance of the dependent variable.

The B values in Table 30 indicate the predictability for each variable. The stepwise method only selected one of the two components obtained, considering the Box Quantities component the only relevant component for the model. The *t* values showcase that the Box Quantities component has decent predictability power over the dependent value, although its coefficient is small when compared to the constant.

Table 28:Model summary (2 principal components, Stepwise method)

Model	R	R square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	0,473[a]	0,224	0,202	1,932	2,034

a. Predictors: (Constant), Box Quantities

Table 29:ANOVA[a] table (2 principal components, Stepwise method)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	38,666	1	38,666	10.364	0,003[b]
	Residual	134,307	36	3,731		
	Total	172,974	37			

a. Dependent Variable: Packed Pallet Evaluation
b. Predictors: (Constant), Box Quantities

Table 30: Coefficients[a] table (2 principal components, Stepwise method)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	5,026	0,313		16,042	0,000		
	Box Quantities	1,009	0,313	0,473	3,219	0,003	1,000	1,000

a. Dependent Variable: Packed pallet evaluation

To validate a multiple linear regression, some assumptions must be checked to see if the test is reliable. These are the normality, homogeneity, error independency, and multicollinearity (Maroco, 2014).

First, there's a graphical overview on the normality of the residuals, with a Predicted Probability plot. The dots in Figure 15, although they have some slight deviations from the line, overall they follow the line's tendency, which corroborates the normality of the regression. The Figure 16 is used to check homogeneity. The dots representing the residuals seem widely dispersed and not concentrated in a specific zone, which means the homogeneity exists here. Finally, to confirm error independency, there's the Durbin-Watson value. The value obtained should be higher than 1,5 and below 2,5. The ones obtained in the two regressions executed with the principal components are approximately 2,1, which means there's error independency. Finally, the multicollinearity should have a value below 10. The ones obtained are equal to 1, which means there's no multicollinearity between variables.

The equation 2 represents the predictability of the model, where x_1 represents the Box Quantities component. In the end, this regression allowed to conclude that the component Box Quantities, which contains the variables Number of Column Piles, Number of Box Types, Number of Boxes, Time Spent Packing and Percentage of Fragile Boxes, is the component that explains the variance of the Evaluation variable better, while the Box Dimensions component isn't very relevant towards explaining the dependant variable.

$$y = 5{,}026 + 1{,}009x_1 \quad (2)$$

Next, another multiple linear regression will be done but this time with the original 9 variables and not with the principal components. This will be done to see if the variables contained by the Box Quantities component are considered relevant by the multiple linear regression.

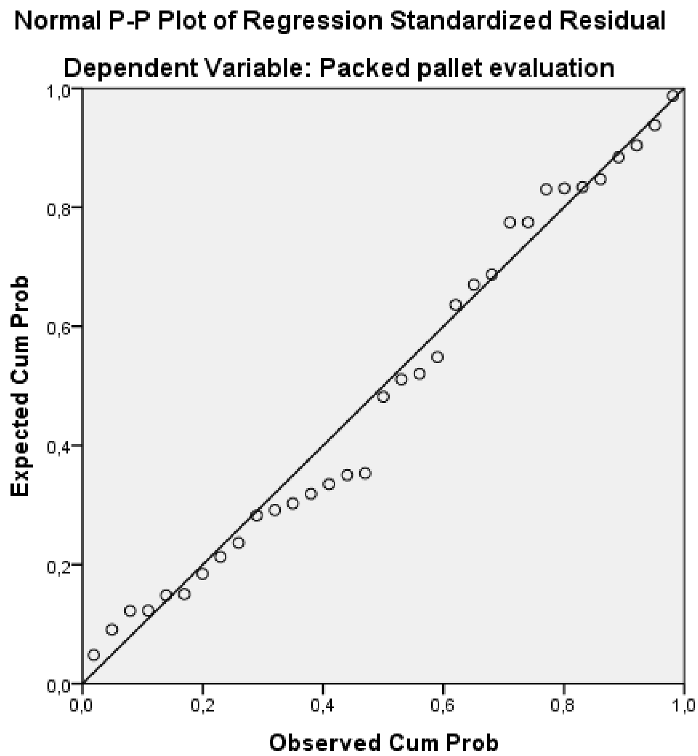

Figure 15: Predicted Probability plot (for 2 principal components)

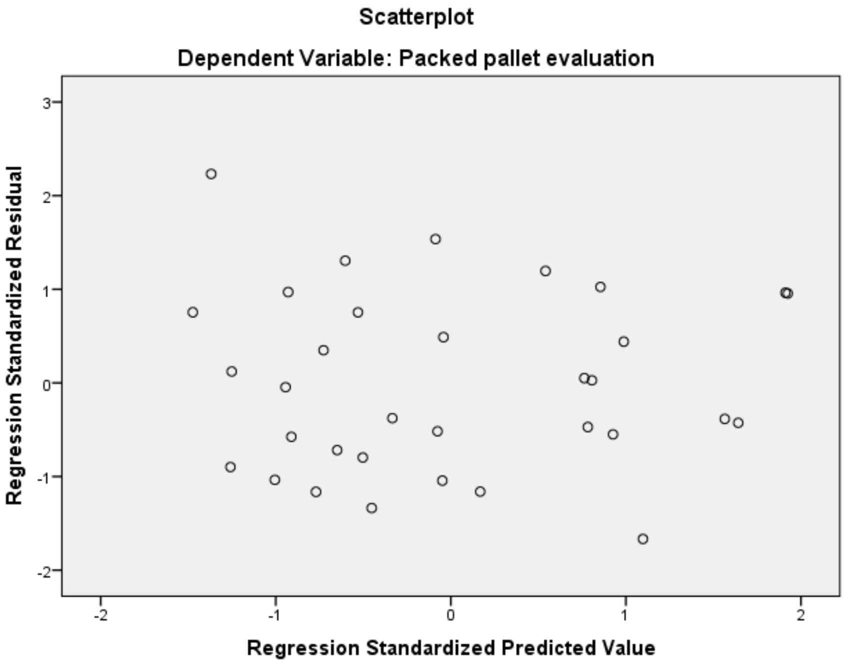

Figure 16: Scatterplot to test residuals homogeneity (for 2 principal components)

Here, Tables 31, 32 and 33 show the outputs for the multiple linear regression with the original 9 variables using the Stepwise method. The Adjusted R square is higher than the ones in previous tests, having the model explain 33,1% of the variance of the Evaluation variable, and the Durbin-Watson value is also good. The table 32 shows a high F value and low Sig. value, which is an indicator that this is a relevant model, although only 1 out of 9 variables were kept in the model using the Stepwise method. That variable is the Number of Column Piles, which is contained by the Box Quantities component. The Figure 17 shows that the residuals follow the line with no odd deviations, meaning that there's data normality. The Figure 18 shows that the dots are spread in the graphic, meaning there's data homogeneity.

$$y = 3{,}637 + 0{,}298x_1 \tag{3}$$

The equation 3 sums up the model with the variables and respective coefficients present in Table 33. The variable that affects the complexity of the pallet loading problem, according to this model, is the Number of Column Piles. This multiple linear regression matches what was seen in the principal components analysis, where it was seen that the Box Quantities component was the most significant component. Although this test showed that, statistically, only one variable is responsible towards explaining the Evaluation variable, the latter variable is subjective, which means that other parameters can affect this variable, depending on the perception of the different workers.

Table 31: Model summary (for the 9 original variables, Stepwise method)

Model	R	R square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	0,591[a]	0,349	0,331	1,769	2,376

a. Predictors: (Constant), Number of Column Piles

Table 32: ANOVA[a] table (for the 9 original variables, Stepwise method)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	60,338	1	60,338	19,285	0,000[b]
	Residual	112,636	36	3,129		
	Total	172,974	37			

a. Dependent Variable: Packed Pallet Evaluation

b. Predictors: (Constant), Number of Column Piles

Table 33: Coefficients[a] table (for the 9 original variables, Stepwise method)

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	3,637	0,427		8,517	0,000		
Number of Column Piles	0,298	0,068	0,591	4,391	0,000	1,000	1,000

a. Dependent Variable: Packed pallet evaluation

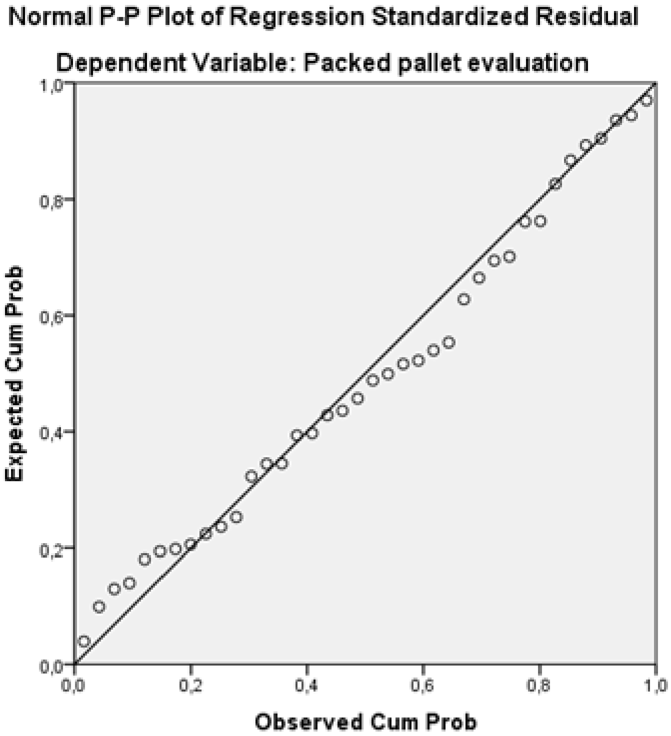

Figure 17: Predicted Probability plot (for 9 variables)

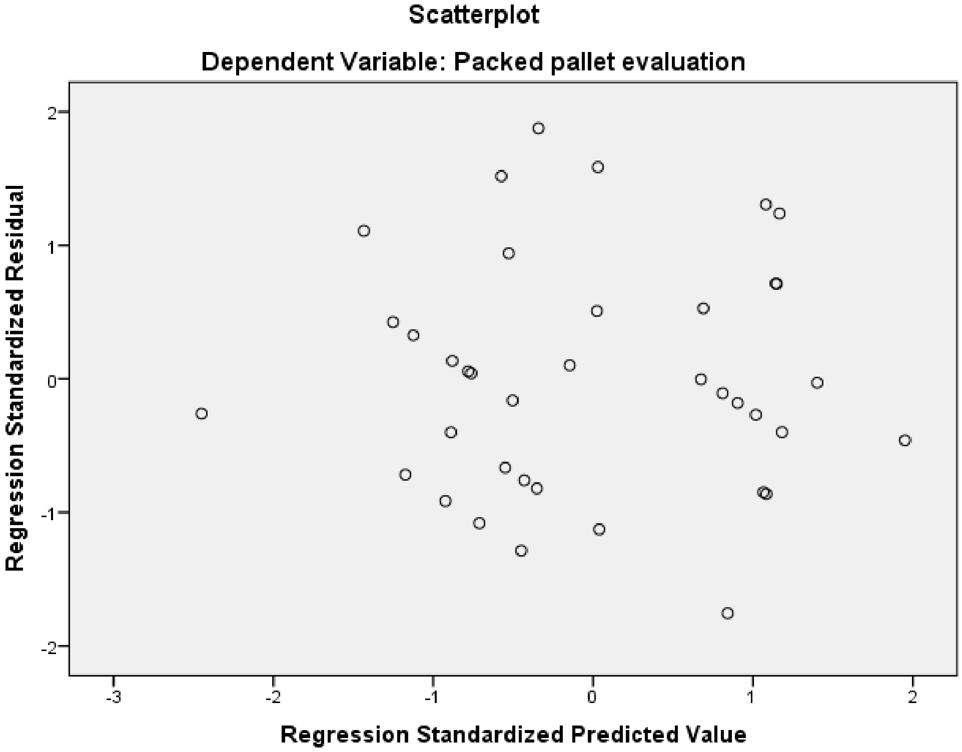

Figure 18: Scatterplot to test residuals homogeneity (for 9 variables)

CONCLUSIONS

4 CONCLUSIONS AND PROPOSALS OF FUTURE WORKS

4.1 CONCLUSIONS

The objectives referred to in section 1.2 were mostly accomplished. The literature reviewed represents a major part of what is said about the pallet packing problem, ranging from older papers with authors who are a reference in this subject to recent paper with a fresher view on this matter. The metrics created attempted to reflect the issues referred to in the literature as well as the main concerns of the picking personnel at Luis Simões about the complexity of cargo arrangements. The scaling of the parameters seemed adequate, although some measured observations possess values that are widely dispersed from the created scale. For example, rating the number of boxes to pack in a pallet requires a minimum of 80 boxes, but there are multiple packed pallets with over 100 boxes. This could mean that the scale wasn’t optimally adjusted and could require a rescale.

About the principal components analysis results, the parameters that have more variance explained by the principal components, according to the first “Communalities” table, are the number of box types, the height difference between worker and pile and the average weight of packed boxes. These are the expected results because these parameters have dispersed classifications as seen in the observations table. Most results are either close to the top of the close or near its bottom, which contributes to high variance levels. The opposite applies to the time spent packing a pallet, percentage of fragile boxes, and number of boxes packed. These variables presenting the least variance is expected, except for the number of boxes packed, since there are multiple observations all with different values utilized. Something that may explain the low variance is the scale utilized because, as explained in the previous paragraph, there are multiple arrangements that top the maximum levels of the scale. Therefore, most of those differences are eliminated and the variance explained by this variable decreases drastically.

About the principal components themselves, the obtained trio of components is acceptable, since the set of variables of each component are correlated and make sense measuring them together. They also explain a big part of the total variance of the data (79%), although it would be better if that value was a bit higher. That difference is more impactful in the second PCA, where 69% of the variance is explained by the two principal components. In the end, according to the second PCA containing two principal components, it is possible to see by looking at component 1 that the number of column piles, number of box types, number of boxes, time spent packing the pallet and the percentage of fragile boxes are the ones that cause more fluctuation of the data, which should be the variables who affect the most the evaluation of the complexity of pallet packing by the pickers. This should be confirmed by the multiple linear regression, which will be reviewed in the next paragraph.

In the multiple linear regression, it was shown that the Box Quantities component, containing the variables Number of Column Piles, Number of Box Types, Number of Boxes, Time Spent Packing and Percentage of Fragile Boxes, is the one that explains the complexity of the pallet loading process. The Box Dimensions component, containing the variable measuring the difference of heights between the stacked pile and worker, number of heavy boxes to load, average box weight and the average maximum width, wasn't deemed impactful in the complexity of the pallet loading problem. In the regression with the 9 original variables, only the variable Number of Column Piles was retained in the model when using the Stepwise method, meaning that the significance of the other variances was too low. It is worth noting that the variable retained belongs to the Box Quantities component, matching the results from the regression with the components. Considering a variable that measures something that is subjective, there can be discrepancies of opinions and the results could showcase that. For a box picker, one variable that adds much complexity may not be the case for a different worker. These perspectives may clash, but in this situation, the results showed that one of the parameters that most workers deem as complex was proven to be indeed complex. This doesn't mean other don't affect the complexity, but only that one had statistical evidence that it was increasing the complexity. Others, such as the Number of Boxes, who was expected to have big weight in the complexity of a pallet cargo arrangement due to affecting most of the other variables, wasn't very significant in the regression but it was more relevant than most, as it belongs to the Box Quantities component.

In the end, only a few variables, which are included in a single principal component, are significant enough to explain the complexity in the pallet loading problem, particularly the Number of Column Piles variable. The relatively small number of observations and the scale used may have had some effect in the tests applied, either the PCA or the multiple linear regression. The Cronbach's Alpha value for the 9 original variables denoted that the data retrieved could present some statistical flaws. The parameters that were scrapped, as seen in section 3.3, could've added a new dynamic to the results obtained, but some were either not adequate to apply or there wasn't information available to rate that. Despite these issues, the goals proposed for this report were mostly achieved.

4.2 PROPOSALS OF FUTURE WORKS

As a suggestion for future works, perhaps repeat the process executed in this report but with multiple changes. First, use a different scaling for the variables to use, or even use the raw data without any types of scales. Also, evaluate the complexity of the pallet cargo arrangements with different parameters and a different and perhaps wider set of observations. Such differences would create a new dynamic for the new model, and it would be interesting to observe which one has the best predictive power as well as which one represents better what the complexity of the pallet loading problem is about.

BIBLIOGRAPHY

5 BIBLIOGRAPHY

Alguns números que espelham a dimensão da Luís Simões. (n.d.). Retrieved July 5, 2018, from http://www.luis-simoes.pt/page/indicadores-ls

Amossen, R. R., & Pisinger, D. (2010). Multi-dimensional bin packing problems with guillotine constraints. *Computers and Operations Research*, *37*(11), 1999–2006. https://doi.org/10.1016/j.cor.2010.01.017

Bischoff, EE; Ratcliff, M. (1995). Issues in the development of approaches to container loading. *Omega*, *23*(4), 377–390.

Bortfeldt, A., & Wäscher, G. (2013). Constraints in container loading-A state-of-the-art review. *European Journal of Operational Research*, *229*(1), 1–20. https://doi.org/10.1016/j.ejor.2012.12.006

Castro, P. M., & Grossmann, I. E. (2012). From time representation in scheduling to the solution of strip packing problems. *Computers and Chemical Engineering*, *44*, 45–57. https://doi.org/10.1016/j.compchemeng.2012.05.002

Costa, R. A., Pereira, T., & Lopes, I. C. (2018). The reverse logistics of unsold medications in pharmacies in Campania, Italy. *Springer Proceedings in Mathematics and Statistics*, *223*, 25–38. https://doi.org/10.1007/978-3-319-71583-4_3

Data Science Group. (2014). Correlation Analysis. Retrieved September 19, 2018, from https://datasciencegroup.wordpress.com/2014/04/29/correlation-analysis/

Dyckhoff, H. (1990). A typology of cutting and packing problems. *European Journal of Operational Research*, *44*(2), 145–159. https://doi.org/10.1016/0377-2217(90)90350-K

Egeblad, J., & Pisinger, D. (2009). Heuristic approaches for the two- and three-dimensional knapsack packing problem. *Computers and Operations Research*, *36*(4), 1026–1049. https://doi.org/10.1016/j.cor.2007.12.004

Elhedhli, S., Gzara, F., & Yan, Y. F. (2017). A MIP-based slicing heuristic for three-dimensional bin packing. *Optimization Letters*, *11*(8), 1547–1563. https://doi.org/10.1007/s11590-017-1154-5

George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update. wps. ablongman. com/wps/media/objects/385. *George 4answers Pdf*.

Gonçalves, J. F. (2007). A hybrid genetic algorithm-heuristic for a two-dimensional orthogonal packing problem. *European Journal of Operational Research*, *183*(3), 1212–1229. https://doi.org/https://doi.org/10.1016/j.ejor.2005.11.062

Hifi, M. (2002). Approximate algorithms for the container loading problem. *International Transactions in Operational Research*, *9*(6), 747–774. https://doi.org/10.1111/1475-3995.00386

Jeong, M. (2016). *Container Packing Problem with Guillotine Cutting and Complete-Shipment Constraints*. Seoul.

Johnson, S. (2005). Size Of A Human: Body Proportions. Retrieved from https://hypertextbook.com/facts/2006/bodyproportions.shtml

Junqueira, L., Morabito, R., & Sato Yamashita, D. (2012). Three-dimensional container loading models with cargo stability and load bearing constraints. *Computers and Operations Research*, *39*(1), 74–85. https://doi.org/10.1016/j.cor.2010.07.017

K-State Reseach and Extention. (2007). Back Injury Prevention For the Landscaping and Hiorticultural Services Industry. *OSHA*. Retrieved from https://www.osha.gov/dte/grant_materials/fy06/46g6-ht22/back_injury_prevention.pdf

Lattin, J., Carroll, J. D., & Green, P. E. (2011). Análise de dados multivariados. *São Paulo: Cengage Learning*, *475*.

Lin, C. C., Kang, J. R., Liu, W. Y., & Li, C. C. (2016). On two-door three-dimensional container packing problem under home delivery service. *Journal of Industrial and Production Engineering*, *33*(3), 205–214. https://doi.org/10.1080/21681015.2015.1136705

Lins, L., Lins, S., & Morabito, R. (2002). An n-tet graph approach for non-guillotine packings of n-dimensional boxes into an n-container. *European Journal of Operational Research*, *141*(2), 421–439. https://doi.org/https://doi.org/10.1016/S0377-2217(02)00135-2

Lynn, E. (2007). Multiple Linear Regression. *Multiple Linear Regression*, 3–6. https://doi.org/10.1007/978-3-319-11997-7

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample Size in Factor Analysis: The Role of Model Error. *Multivariate Behavioral Research*, *36*(4), 611–637. https://doi.org/10.1207/S15327906MBR3604_06

Maroco, J. (2014). *Análise estatística: com utilização do SPSS*. (ReportNumber, Ed.) (6th ed.).

Martello, S., Pisinger, D., Vigo, D., Boef, E. Den, & Korst, J. (2007). Algorithm 864: General and robot-packable variants of the three-dimensional bin packing problem. *ACM Transactions on Mathematical Software*, *33*(1), 7–es. https://doi.org/10.1145/1206040.1206047

Morabito, R., & Arenalest, M. (1994). An AND / OR-Graph Approach to the Container Loading Problem. *Elsevier Science*, *I*(1), 59–73.

Moreira, A. C. (2007). Comparação da Análise de Componentes Principais e da CATPCA na Avaliação da Satisfação do Passageiro de uma Transportadora Aérea . *Investigação Operacional* . scielopt .

Murray, M. (2018). Order Picking In The Warehouse - Supply Chain / Logistics. Retrieved July 4, 2018, from https://www.thebalance.com/order-picking-in-the-warehouse-2221190

Ramos, A. G., Oliveira, J. F., & Lopes, M. P. (2016). A physical packing sequence algorithm for the container loading problem with static mechanical equilibrium conditions. *International Transactions in Operational Research*, *23*(1–2), 215–238. https://doi.org/10.1111/itor.12124

RIBAS, J. R., & Vieira, P. R. da C. (2011). Análise multivariada com o uso do SPSS. *Rio de Janeiro: Ciência Moderna*.

Rocha, J. F. S. (2015). *TRANSPORTE MARÍTIMO DE CURTA DISTÂNCIA NA UNIÃO EUROPEIA E EM PORTUGAL*.

Silva, E., Oliveira, J. F., & Wäscher, G. (2016). The pallet loading problem: A review of solution methods and computational experiments. *International Transactions in Operational Research*, *23*(1–2), 147–172. https://doi.org/10.1111/itor.12099

Wäscher, G., Haußner, H., & Schumann, H. (2007). An improved typology of cutting and packing problems. *European Journal of Operational Research*, *183*(3), 1109–1130. https://doi.org/10.1016/j.ejor.2005.12.047

ANNEXES

6 ANNEXES

6.1 ANNEX I - APPLICATION OF THE CREATED METRICS IN THE COLLECTED SAMPLES

Sample\Code	1	2	3	4	5	6	7	8	9	Evaluation
Units		Kg	%	m			mins	m		
1	55	2,38	20,0	0,170	5	0	4	-1,087	0	4
2	121	3,09	54,5	0,310	11	0	9	1,340	0	4
3	99	2,05	77,8	0,210	9	0	6	-0,652	0	4
4	174	2,36	2,3	0,170	4	0	15	-0,617	0	2
5	5	2,61	0	0,153	1	0	1	-1,294	0	1
6	87	3,61	36,8	0,361	25	> 9	42	1,310	11	9
7	199	3,79	38,7	0,290	18	0	8	1,340	11	4
8	450	2,48	13,3	0,150	8	0	31	1,340	0	3
9	16	23,40	0	0,570	2	0	16	-0,756	16	2
10	50	5,86	0	0,530	1	0	2	1,340	0	4
11	176	2,57	24,4	0,220	54	> 9	48	1,220	11	9
12	32	12,25	0	0,290	9	9	9	-0,566	22	6
13	97	9,14	1,0	0,310	10	> 9	22	1,220	32	7
14	78	2,41	3,8	0,390	13	> 9	24	0,323	0	5
15	153	2,61	0	0,150	1	0	8	-0,774	0	8
16	375	2,91	0	0,410	3	0	13	1,340	0	3
17	31	23,40	0	0,560	2	0	6	0,106	31	7
18	86	7,96	0	0,270	3	0	8	0,140	6	5
19	33	5,31	0	0,330	2	0	4	-1,126	18	2
20	25	10,60	100	0,280	1	0	3	-0,888	25	2
21	52	2,43	61,5	0,270	14	> 9	19	-0,839	0	6
22	143	1,90	21	0,160	10	> 9	13	-0,885	0	7
23	108	2,28	27,8	0,170	24	> 9	21	-0,930	0	7
24	91	1,83	49,5	0,220	15	> 9	15	-0,924	0	8
25	106	1,60	12,3	0,190	13	> 9	17	-0,865	0	6
26	24	5,86	0	0,530	1	0	3	-0,096	0	5
27	32	17,72	0	0,340	1	0	7	-0,258	32	6
28	24	14,07	0	0,390	1	0	9	-0,726	24	3

Sample\Code	1	2	3	4	5	6	7	8	9	Evaluation
Units		Kg	%	m			mins	m		
29	90	1,63	52,2	0,200	16	> 9	16	-0,770	0	5
30	162	2,23	100	0,270	1	0	12	0,051	0	6
31	130	1,89	53,8	0,160	22	> 9	25	-0,832	0	3
32	99	10	0	0,600	1	0	5	1,220	99	3
33	95	3,88	10,5	0,240	16	> 9	27	1,220	17	6
34	115	2,31	87,8	0,320	7	7	20	1,340	16	8
35	34	4,37	17,6	0,390	4	4	10	-0,672	14	3
36	1550	2,62	41,9	0,200	14	0	90	1,340	120	4
37	49	4,95	14,3	0,260	6	6	5	-0,713	12	8
38	77	5,92	15,6	0,390	18	> 9	19	1,220	38	6

6.2 ANNEX II - SCALING OF THE PARAMETERS MEASURED ACCORDING TO THE CREATED SET OF METRICS

Sample\Code	1	2	3	4	5	6	7	8	9	Evaluation
1	8	2	4	1	3	1	1	1	1	4
2	10	3	10	2	6	1	2	10	1	4
3	10	2	10	2	5	1	2	2	1	4
4	10	2	1	1	2	1	3	2	1	2
5	1	2	1	1	1	1	1	1	1	1
6	10	3	8	3	10	10	9	10	6	9
7	10	3	8	2	8	1	1	10	6	4
8	10	2	3	1	4	1	7	10	1	3
9	4	10	1	4	1	1	4	2	8	2
10	7	4	1	3	1	1	1	10	1	4
11	10	2	5	2	10	10	10	10	6	9
12	6	9	1	2	5	9	2	2	10	6
13	10	7	1	2	5	10	5	10	10	7
14	10	2	1	3	7	10	5	9	1	5
15	10	2	1	1	1	1	2	2	1	8
16	10	2	1	3	2	1	3	10	1	3
17	6	10	1	4	1	1	2	5	10	7
18	10	6	1	2	2	1	2	5	3	5
19	6	4	1	2	1	1	1	1	9	2
20	5	8	10	2	1	1	1	1	10	2
21	8	2	10	2	7	10	4	2	1	6
22	10	2	5	1	5	10	3	1	1	7
23	10	2	6	1	10	10	5	1	1	7
24	10	2	9	2	7	10	3	1	1	8
25	10	2	3	2	7	10	4	2	1	6
26	5	4	1	3	1	1	1	2	1	5
27	6	10	1	2	1	1	2	2	10	6
28	5	9	1	3	1	1	2	2	10	3
29	10	2	10	2	8	10	4	2	1	5

Sample\Code	1	2	3	4	5	6	7	8	9	Evaluation
30	10	2	10	2	1	1	3	4	1	6
31	10	2	10	1	10	10	5	2	1	3
32	10	7	1	4	1	1	1	10	10	3
33	10	3	3	2	8	10	6	10	9	6
34	10	2	10	2	4	7	4	10	8	8
35	6	3	4	3	2	4	2	2	7	3
36	10	2	9	2	7	1	10	10	10	4
37	7	4	3	2	3	6	1	2	6	8
38	9	4	4	3	8	10	4	10	10	6