

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Simulation and Optimization Based Approach for Last mile Network Design

Inês Fernandes Vilarinho



Mestrado Integrado em Bioengenharia

FEUP Supervisor: Nuno Azevedo

LTPlabs Supervisor: Nuno Pedro

July 29, 2022

Simulation and Optimization Based Approach for Last mile Network Design

Inês Fernandes Vilarinho

Mestrado Integrado em Bioengenharia

July 29, 2022

Abstract

The main objective of this work is to respond to the urgent need to create optimized distribution routes for postal services. To this end, a robust route optimization and simulation tool were created, which was divided into three parts: i) two interfaces for parameterization and configuration of the values that will be used to run the ii) algorithm that encompasses the generation of ideal routes and whose main parameters will be presented in iii) three reports that will serve as a decision support tool. Furthermore, this tool should address the main weaknesses of the process currently used by the postal service company, namely, the automation of information in the system.

The parameterization and optimization interfaces were created in Excel and are linked to the database containing all the project's essential information to vary and define crucial parameters for the model's run.

Moreover, the algorithm has three main modules, namely the generation of clusters, which organizes the traffic, the simulation of routes using various scenarios, and, finally, the optimization, whose output will be both the paths to be covered and evaluation metrics.

On the other hand, the designed reports aim to analyze the results graphically obtained by the model, making decisions more straightforward and precise.

Finally, the model outputs were compared with values currently obtained through the postal company's current process to validate. Thus, by comparing the metrics of the current studies with the outputs of the model, it was noticed that the latter manages to obtain a smaller number of routes and time used in the distribution despite having a greater traffic volume. On the other hand, from the comparison with RouteSmart, it was only possible to remove the potential of using the model since it generates a smaller number of routes for more traffic. However, this comparison should be performed with the same traffic value to obtain concrete conclusions.

Finally, the evaluation of metrics to define network integration and routing strategies led to the conclusion that the InTrevo strategy is mainly chosen as well as the multiple network. Furthermore, the annual assessment of the savings potential for distribution centers shows that this value could reach 106 688 €.

Keywords: Last mile distribution, simulation, optimization and clustering

Resumo

O presente trabalho tem como principal objetivo corresponder à necessidade urgente de criação de rotas de distribuição otimizadas. Para tal, foi criada uma ferramenta robusta de otimização e simulação de rotas que foi dividida em três partes: i) duas interfaces de parametrização e configuração dos valores que serão utilizados para correr o ii) algoritmo que engloba a geração de rotas ideais e cujos principais parâmetros serão apresentados em iii) três relatórios que servirão como uma ferramenta de apoio à decisão. Esta ferramenta deve colmatar as principais fragilidades do processo atualmente utilizado pela empresa de serviço postal, nomeadamente, a automatização da informação no sistema.

As interfaces de parametrização e otimização foram criadas em Excel e apresenta-se ligadas à base de dados que contém toda a informação essencial do projeto de forma a poder variar e definir parâmetros essenciais para a corrida do modelo.

Para além disso, o algoritmo apresenta três módulos principais, sendo eles a geração de clusters, que organiza o tráfego o restante processo, a simulação de rotas utilizando vários cenários e, por fim, a otimização cujo output serão tanto os percursos a percorrer como métricas de avaliação dos mesmos.

Por outro lado, os relatórios desenhados tem como objetivo a análise dos resultados obtidos pelo modelo de uma forma gráfica, tornando as decisões mais claras e precisas.

Por fim, os outputs do modelo foram comparados com valores atualmente obtidos através do processo atual da empresa postal de modo a validá-lo. Assim, através da comparação das métricas obtidas através dos atuais estudos com os outputs do modelo, percebeu-se que o último consegue obter um menor número de rotas e de tempo utilizado na distribuição apesar de ter um maior volume de tráfego. Por outro lado, da comparação com a RouteSmart foi apenas possível retirar o potencial de utilização do modelo uma vez que para um maior tráfego este gera um menor número de rotas. No entanto, esta comparação deverá ser realizada com o mesmo valor de tráfego para obter conclusões concretas.

Por fim, a avaliação das métricas para definição de estratégias de integração de redes e roteamento levou à conclusão de que a estratégia InTrevo é maioritariamente a escolhida bem como a rede múltipla. A avaliação anual do potencial de poupança para os centros de distribuição mostram que este valor poderá chegar aos 106 688 €.

Palavras-chave: Distribuição *last mile*, simulação, otimização e *clustering*

Acknowledgements

Em primeiro lugar gostaria de agradecer à LTPlabs, a empresa que me recebeu com toda a hospitalidade e que tanto me ensinou durante este percurso, e a todos que nela trabalham e que me receberam de braços abertos. Em especial à equipa com quem trabalhei, o Pedro, o Nuno, o Diogo e a Carolina que sempre estiveram disponíveis para me ajudar e ensinar ao longo desta jornada. Tal como não podia deixar de ser, destaco o ótimo trabalho do Nuno Pedro enquanto meu orientador e agradeço-lhe muito por toda a prontidão e partilha de conhecimentos que teve para comigo ao longo de todo o processo.

Gostaria também de deixar um especial agradecimento ao professor Nuno Azevedo que desde o início se disponibilizou para embarcar neste percurso comigo e sempre demonstrou, simpaticamente, a sua disponibilidade e interesse no projeto. Deixo também um grande beijinho a todos os que acompanharam o meu percurso de início ao fim, que o percorreram comigo e que me motivaram muito para chegar ao final desta meta.

Por fim, gostaria de agradecer à minha família. Aos meus pais que tem sempre uma palavra amiga, aos meus irmãos que, independentemente do momento, me conseguem arrancar sempre uma gargalhada, aos meus tios, primos e avós pela força, coragem e motivação para ser uma pessoa melhor e uma boa profissional. Não poderia não referir o João Afonso, aquele que desde início me encoraja a remar novos mares e me apoia incondicionalmente, quer profissional como pessoalmente. Um grande obrigado por todos os ensinamentos, por toda a ajuda que me dás e simplesmente por seres quem és.

Inês Fernandes Vilarinho

This work was financially supported by LA/P/0045/2020 (ALiCE), UIDB/00511/2020 and UIDP/00511/2020 (LEPABE), funded by national funds through FCT/MCTES (PIDDAC).

*“There are no secrets to success.
It is the result of preparation, hard work, learning from failure.”*

Colin Powell

Contents

1	Introduction	1
1.1	Context	1
1.2	Motivation	1
1.3	Objectives	2
1.4	Project organization and timeline	2
2	State of the art	5
2.1	Last mile distribution	5
2.2	Otimization: vehicle routing problem	6
2.2.1	Capacitated vehicle routing problem	7
2.3	Route simulation	9
2.4	Machine learning for simulation and optimization	10
2.5	Main metrics to evaluate last mile distribution	12
3	Companies contextualization	13
3.1	Contextualization of the consultant company	13
3.2	Postal company context	13
3.2.1	Mail type	14
3.2.2	Product matrix	14
3.2.3	Operational infrastructure	15
3.2.4	Postal code organization	16
3.2.5	General tasks' view	16
3.2.6	Types of delivery point	16
3.2.7	Distribution strategy	16
3.2.8	Implementation of distribution strategies - Current process	18
4	Methodology	21
4.1	Database structure	21
4.2	Process	22
4.2.1	Use cases	22
4.2.2	Future process	23
4.3	Traffic	23
4.4	Tasks and times	24
4.5	Implementation	25
4.6	Restrictions imposed by the company	25
4.7	Clustering, simulation and optimization approach	26
4.7.1	Clustering	27
4.7.2	Simulation	27

4.7.3	Optimization	27
4.8	Deep dive C# model	27
4.9	Validation and tests	28
4.9.1	Comparison with RouteSmart	28
4.9.2	Comparison with current study	29
4.10	Pilot project	29
5	Results and Discussion	31
5.1	Interface development results	31
5.1.1	Parametrization interface	32
5.1.2	Configuration interface	33
5.2	Model development results	33
5.2.1	Descriptive analysis of results	34
5.2.2	Comparison and validation of results	36
5.2.3	Strategies optimization	39
5.3	Report development results	45
6	Conclusions and Future Work	47
6.1	Conclusions	47
6.2	Further Work	48
A	Data collection and project planning	51
A.1	Current process and its weaknesses	51
A.2	Micro task times	51
B	Interface	55
B.1	VBA programation	55
C	C# model	57
C.1	Clustering	57
C.2	Optimization	59
C.3	Simulation	59

List of Figures

1.1	Project's timeline.	3
2.1	Vehicle Routing Problem [6]. In the left panel, it is possible to observe seven different delivery points necessary to visit by the routes and in the right one appears the generation of bearing in mind routes the best organization found.	8
2.2	Clustering operation [17].	11
3.1	Product matrix.	14
3.2	Product matrix and the appropriate vehicle restrictions.	14
3.3	Operational infrastructure.	15
3.4	Relation between different infrastructures.	15
3.5	Scheme of route organization strategies.	17
3.6	Scheme of EMS integration.	18
4.1	Method organization.	21
4.2	Example of the relationship between different tables.	22
4.3	Future process.	23
4.4	Data infrastructure.	25
4.5	Modular structure of the algorithm showing the logic implemented.	28
5.1	Parameterization interface menu.	31
5.2	Configuration interface menu.	32
5.3	Example sheet of the parameterization interface features.	32
5.4	Example sheet of the configuration interface.	33
5.5	Example of a form created in the configurations interface.	33
5.6	Route overview.	39
5.7	Detailed view of each part (A, B and C) of the route shown in Figure 5.6.	39
5.8	Print example of the evaluation report.	46
5.9	Print example of the study report.	46
A.1	Scheme for prioritizing studies in the current process.	51
A.2	Scheme of data collection in the current process.	52
A.3	Scheme of the start of the optimization in the current process.	52
A.4	Scheme of the end of the optimization in the current process.	52
A.5	Scheme of the information contained in the final reports of the current process.	53
B.1	"Update" button programming code.	55
B.2	"Save" button programming code.	55
B.3	"Add" button programming code.	56

B.4	"Delete" button programming code.	56
C.1	Generating clusters in the C# model.	57
C.2	Initialize clusters function.	58
C.3	Run clusters function.	58
C.4	Upload clusters function.	59
C.5	Optimization function.	60
C.6	Simulation function.	61

List of Tables

3.1	Description of all delivery points defined by the postal company.	17
5.1	Comparison of current As-Is and Optimized As-Is metrics.	34
5.2	Comparison of strategy results.	35
5.3	Comparison between the values obtained by the studies and the optimized model.	37
5.4	Comparison between RouteSmart and the optimized model.	37
5.5	Number of delivery points for each segment of the presented route.	38
5.6	Comparison between single and multiple networks for each of the pilot CDPs.	40
5.7	Differences found between the values presented in the Table 5.6.	40
5.8	Presentation of the values of current capacity and extra capacity needed for the implementation of the single network.	41
5.9	Comparison between segmented and InTrevo strategies for each of the pilot CDPs.	42
5.10	Differences found between the values presented in the Table 5.9.	42
5.11	Predicted gains for each of the macro scenarios tested.	44
5.12	Maximum and minimum earnings for each pilot CDP.	45
A.1	Examples of micro tasks and respective execution times.	53

Abbreviations

CDP	Postal distribution centers
CLD	Logistics and distribution centers
CO	Operational centers
CPL	Production and logistics centers
CVRP	Capacitated vehicle routing problem
DV	Specialized network
EMS	Express mail
ETI	Number of workers
LLE	Letters and large envelopes
LNP	Large and non standard packages
MP	Medium packages
P	Packages
SP	Small packages
T&T	Track and trace UEC
Equivalent mail units	
VBA	Visual basic for applications
VRP	Vehicle routing problem

Chapter 1

Introduction

1.1 Context

The greater ease of access to the online market combined with the pandemic currently experienced worldwide promoted the increase in E-commerce, with a greater volume of online purchases. E-commerce highly depends on the delivery service that will ensure the product reaches the final customer, and the last mile distribution, the part of the route from the previous distribution point to the delivery point, is the most time-consuming and costly step [1]. Thus, companies that practice this service face great competition among themselves so that, corresponding to the needs of the consumer, they can plan deliveries in an optimized way with the main objectives of reducing the overall cost of the process and improving the customer's experience.

In addition to the increase in packages traffic, it is also essential to consider the constant need for letter delivery by the postal service and thus optimize the last mile distribution to evaluate all products for distribution. Currently, studies are carried out at each distribution center to predict the best method of acting in the distribution from that point.

1.2 Motivation

Each stage of postal delivery has an associated cost for the company that performs this service, and this value depends on the type of product in question. However, it is known that the last mile distribution is the one that entails the highest global cost, corresponding to about 50% of the total value of the operation [1]. Furthermore, as this stage involves direct contact with the final customer, it is crucial to promote compliance with delivery deadlines to maintain the performance levels of each distribution center.

Thus, the primary motivation for carrying out this project is to organize routes to find the minimum possible operating cost value for the last mile distribution to different distribution points to maintain or improve performance levels and deadlines of delivery.

1.3 Objectives

This work was developed bearing two main objectives in mind, namely:

- The creation of a robust tool that allows the postal company to access information that assists strategic (routing and distribution) and tactical (resources) decisions. This tool will be associated with a rigorous simulation and optimization model that allow the calculation of important metrics to support the evaluation and prioritization of the target distribution centers;
- Propose improvements to the current study process by automating information, promoting communication between different teams (operational and planning), and designing a future process that integrates the tool's simulation outputs.

1.4 Project organization and timeline

The project presented comprises six general tasks:

- **Preparation:** This step is based on collecting the necessary data as well as evaluating and setting up the IT infrastructure;
- **Deep-dive in the process and analysis of postal distribution centers (CDPs):** It aims to evaluate the current metrics and methods used by the CDPs for strategical decisions in detail and systematize the learning and requirements for the tool. In addition, this point also covers the design of the future evaluation process of CDPs;
- **Model preparation:** Model preparation involves technology migration and preparation of an algorithm for detailed analysis of each CDP;
- **Planning tool:** At this point, the planning simulation tool is developed, as well as the execution of reports of the model outputs (dashboards). Finally, the integration of the model in the company's infrastructure is also relative to this point;
- **Tests:** Tests and validation of the tool;
- **Implementation and monitoring of studies.**

The schedule shown in Figure 1.1 shows the temporal distribution of each project task over the weeks dedicated to its completion.

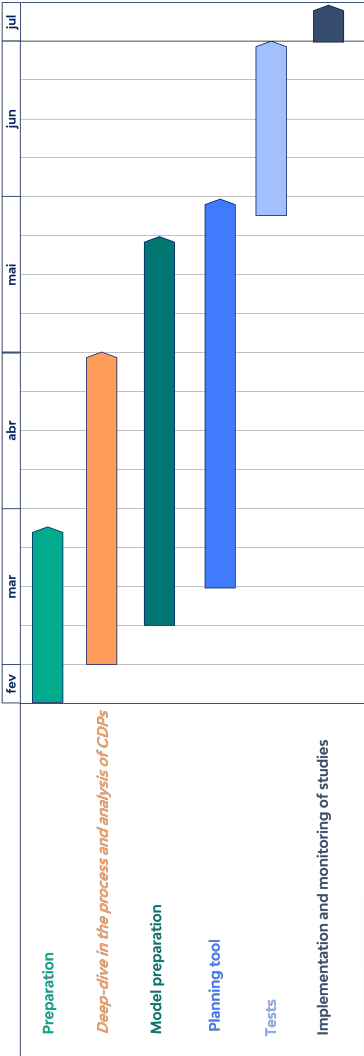


Figure 1.1: Project’s timeline.

Chapter 2

State of the art

The main goal of this work is to create a tool to assist strategical and tactical decisions regarding distribution routes of a specific postal service, with special focus on the last stage of the delivery process: the last mile distribution. Having this in mind, the current chapter comprises the definition of this final step alongside different methods and algorithms that allow the development of an end-to-end solution to effectively obtain an optimal network of deliveries.

2.1 Last mile distribution

When an order is shipped, it goes through an extended processing network that will ensure that it arrives at the receiver. First, the package will pass through sorting machines that recognize the shipping address. Then, it will be sent to a distribution center closer to the sender's address, which will be in charge of delivering at the final destination.. Finally, the last stage of the product delivery process corresponds to the last mile distribution, which refers to the last kilometers of the long journey traveled by the package. Therefore, the final step of the last mile distribution corresponds to the delivering of the product at the destination address.

The last mile distribution is the culmination of a vast process, and precisely because it is the stage presented to the consumer, it is the most impactful phase since it is the only point of service that contacts directly with the consumer [2]. Thus, to guarantee that this delivery is carried out efficiently, both in terms of speed of delivery and the costs it entails, it is necessary to have a very well-organized network in advance [2]. However, the characteristic volatility related to the number of objects, their volume, and the staticity of the delivery points can transform this exercise into a very challenging problem.

The last mile stage is considered one of the most expensive stages of the entire logistical process of sending a product and can correspond to about 50 % of the total cost of shipping the product [1].

The two most common problems with last mile deliveries are the absence of customers at the place of delivery and poor route planning. The lack of customers at the delivery point is problematic since it means a possible adjustment of routes, time spent going to the door and

ringing the bell, time to fill in the notice, and even makes the package has to return to be delivered or raised afterward. However, this problem is easily solved by promoting communication between the customer and the distributor, ensuring that the customer knows the day of the time interval the product will be delivered.

On the other hand, the lack of good route management is a more complex problem to solve, and whose impact will be considerably more remarkable since it will decrease delivery efficiency, the number of ports visited will be smaller, or the number of kilometers traveled will be higher, which could mean an unnecessary increase in fuel. In addition, inefficient route planning may not comply with certain limits that must be stipulated, such as the maximum capacity of a vehicle.

That said, it is crucial to develop a process of simulation and optimization of routes, making them more agile and efficient, guaranteeing customer satisfaction, and reducing company costs. This way, networks can be calculated regularly, considering their limitations and the number of objects arriving for distribution in each day.

2.2 Otimization: vehicle routing problem

The vehicle routing problem (VRP) is already old and its formulation was first defined when Danzig and Ramser arrived at a mathematical programming and algorithmic approach to solve the problem of delivering gasoline to the respective stations [3]. Since then, there has been a great deal of research on this subject. VRP is an optimization problem that intends to define an ideal set of routes for a certain fleet of vehicles in order to deliver products to a certain set of customers [4], having as main objective the minimization of the total costs of each route [5].

Generally, due to the difficulty of finding an optimal solution in more complex problem sets, with a large number of customers to serve, mathematical programming and combinatorial optimization are limited, so it is necessary to resort to heuristics that can cover all the complexity of the networks, both in terms of the large route area and the frequency of passage [6].

Solving this type of problem is based on determining a set of routes - a route must be associated with a vehicle and it must start and end at the same point (it will be a distribution point of the postal company), ensuring that all customer requirements and operational constraints are met and minimizing the overall cost of transportation [5]. The objective function of a VRP can be very different depending on the specific application of the result, however, some of the most common objectives are [6]:

- Minimize the overall cost of transport;
- Minimize the number of vehicles needed;
- Increase the quality of service;
- Minimize the distance travelled.

There is a wide range of vehicle routing problems, some of which are presented in the following points:

- **VRP with profits:** In this type of problem, the set of customers to be served is not provided, so the decision involves defining the customers to be done and grouping the customers to be performed by the same route [7].
- **VRP with pickup and delivery:** In this problem, it intends to determine optimal routes that include both the pick-up task from the customer to the postal distribution center and delivery from the distribution center to the customer [8].
- **VRP with time windows:** This problem is based on planning routes that must deliver particular objects to certain locations in a specific time window [9].
- **Capacitated VRP:** Intends to determine optimal routes with vehicles considering their limited capacity [6].

Knowing that there is a need to know the necessary number of resources to carry out the generated routes, it is vital to consider the limits of vehicles and sized routes. Having this in mind, the Capacitated Vehicle Routing Problem was chosen as the best approach to address the needs of the postal services [6].

2.2.1 Capacitated vehicle routing problem

The capacitated vehicle routing problem is a VRP-based problem, but, as the name implies, it adds the constraint on vehicle load. With this addition, the optimization needs to consider more broadly the mutual coupling relationship of the number of vehicles, load distribution, and route planning to achieve a shorter distance traveled [6].

2.2.1.1 CVRP mathematical modeling

In order to formulate a capable vehicle routing problem, it is vital to know in advance several aspects such as the location of the starting point of distribution, commonly known as the "depot", the location of the delivery point, what traffic there is, the number of vehicles available for carrying out the route, as well as their capacity. With access to all the necessary data already mentioned, it is possible to design a distribution route that meets all customers' needs by minimizing the operation's total cost. There are two ways to reduce the overall costs: shorter distances of routes and reducing the time spent in the process. Figure 2.1 presents an example of routes from a CVRP model [6].

The CVRP can be described through a mathematical model where n vehicles from distribution center 0 serve m customers. Let g_i be the demand for customer i , the cost of transport from customer i to customer j , and b_s the load of vehicle s .

$$x_{ij}^s = \begin{cases} 1, & \text{vehicle } s \text{ serves costumer } i \text{ followed by customer } j \\ 0, & \text{else} \end{cases} \quad (2.1)$$

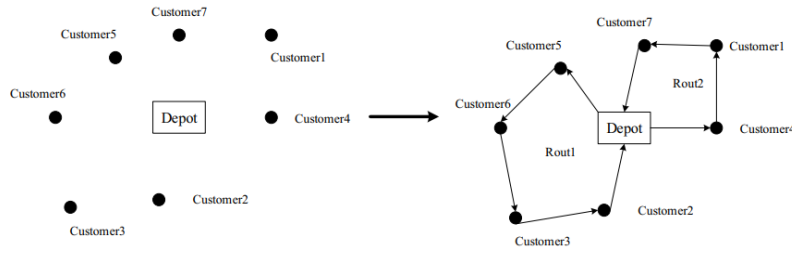


Figure 2.1: Vehicle Routing Problem [6]. In the left panel, it is possible to observe seven different delivery points necessary to visit by the routes and in the right one appears the generation of bearing in mind routes the best organization found.

The main objective of CVRP is to find the optimal route so that the transport cost, J , is minimized (Equation 2.2), that is:

$$J = \min \sum_{s=1}^n \sum_{i,j=0}^m c_{ij} x_{ij}^s \quad (2.2)$$

S.t.:

$$\sum_{j=1}^m x_{j0}^s = \sum_{j=1}^m x_{0j}^s \leq 1, s \in (1, 2, \dots, n) \quad (2.3)$$

$$\sum_{s=1}^n \sum_{j=0, j \neq i}^m x_{ij}^s = 1, i \in (1, 2, \dots, m) \quad (2.4)$$

$$\sum_{s=1}^n \sum_{i=0, i \neq j}^m x_{ij}^s = 1, j \in (1, 2, \dots, m) \quad (2.5)$$

$$\sum_{i=1}^m g_i \sum_{j=0, j \neq 1}^m x_{ij}^s \leq b_s, s \in (1, 2, \dots, n) \quad (2.6)$$

Equation 2.3 constrains that the departure and arrival points are the same and must be the depot. Equations 2.4 and 2.5 indicates that each customer is only visited once. Finally, Equation 2.6 constrains the maximum vehicle capacity [6].

The vehicle routing problem approach allows the creation of optimal distribution routes for the necessary deliveries, however, it does not take into consideration different possible object traffic scenarios. Therefore, it is essential to test the optimization model in different traffic situations in order to understand the impact of this factor on the final results, which can be achieved through simulation techniques.

2.3 Route simulation

Simulation is the art of designing a real system on a virtual environment and understanding what impact variables have on it. The simulation is a robust process that involves several steps, listed below [10]:

1. **Understanding the system:** the goal of the first step is to understand which system is going to be simulated. If the system under study is not known correctly, the variables defined may be wrong, as well as the chosen simulation methodology.
2. **Establish objectives:** it is necessary to define the goal of the simulation and to identify the parameters that measure the performance of the model.
3. **Data collection:** it is necessary to collect all relevant input data, which may or may not be stationary.
4. **Model representation:** detailed planning of the simulation model to be developed. Specialized software (such as Arena and AnyLogic) or custom-built solutions in a programming language like C# or Java can be used to develop a simulation model. Thus, choosing the software and applying the model elements is essential.
5. **Develop the model:** the actual simulation model is developed
6. **Verification of the developed model:** in this step, the developed model is verified to see if it effectively represents the real system. If it turns out that the model does not correspond to what was expected, it is necessary to go back to the beginning and develop all the points mentioned above.
7. **Model validation:** validation is the most challenging stage, as it verifies if the simulation results match those of the real system. If the expected match is not found, it is necessary to review the process to identify the issues.
8. **Tests:** in this stage, tests are performed repeatedly in order to eliminate the effects of uncertainty and increase the confidence level.
9. **Simulation model execution and analysis:** the model is executed and the outcome is analyzed.

Simulation models allow systems to be easily analyzed without real life observations, promoting monetary savings. Furthermore, as the data used for the simulation is often accurate, the output of the simulation will be very similar to what it would be in reality [11]. Another significant advantage of using simulation is the easy change of variables, allowing to obtain more results and to study the system in a longer term. Finally, applying simulation to risky situations promotes their study by excluding the dangerous fact since it will not be tested in a real environment [11].

However, some disadvantages are associated with these methods, which are mainly linked to the possibility of a high cost of installing a simulation model and the need for resources that

require a deep knowledge of the target system [11]. Nevertheless, simulation is always preferable to direct observation of the real physical system, as the simulation is risk-free.

Simulation can be different approaches, and there are three main types of simulation: continuous systems simulation, discrete systems simulation, and agent-based simulation [12].

Continuous event simulation: The continuous event simulation intends to simulate and continuously monitor a given system over time, with no time breaks in it [13].

Discrete event simulation: Discrete event simulation aims to represent a real system, and the modeled events are discrete in time. In contradiction to continuous event simulation, events occur at a given instant of time that marks a state change in the system. In this type of simulation, it is considered that the system does not change consecutive events. As a result, this simulation generally runs faster than continuous simulation [13].

Agent-based simulation: Currently, the use of agent-based simulation facilitates the validation of a concept since the degree of complexity of the process is increasing. The agents contemplated in this type of modeling act autonomously and have the function of fulfilling established objectives, being able to interact with the environment, and establishing contact with other agents. Furthermore, this type of simulation is advantageous as it promotes excellent dynamics between the different agents since each agent will adapt its behavior in response to stimuli from other agents [14]. For the computational recreation of postal services, the type of simulation that best fits is the agent-based simulation since it is necessary to be aware of vehicle capacities.

Having a way of obtaining optimal routes for different traffic scenarios, it is now important to know what values should be included in the simulation. For this, it is necessary to analyze the historical traffic data and define different profiles for the CDPs regarding this factor, which can be achieved through machine learning techniques.

2.4 Machine learning for simulation and optimization

The optimization and simulation techniques work with a large amount of data that must be previously treated and filtered. One of the data processing techniques that allow drawing conclusions about the data is clustering. Clustering is an unsupervised machine learning technique that analyzes data and, through the similarities found between different characteristics, creates a certain number of clusters (groups), ensuring that the instances from each one is more similar to each other than when compared with the ones from another cluster [15]. Depending on the purpose, the number of clusters can be pre-defined by the user or unknown.

There are several ways to formulate a clustering problem, and some differences between different cluster formations can be related, for example, to the uniqueness of the data. For example, it is not uncommon to observe data points that can belong to more than one cluster at the same time,

and it can be an advantage, or not, to consider them as part of both groups. With this in mind, it is important to carefully decide the algorithms to use [16].

One of the most popular clustering methods is the K-means algorithm, which consists on computing the average of different groups of data points iteratively until reaching an optimal distribution of clusters. This model starts by defining randomly the centroids of all clusters, whose quantity can be manually or automatically defined. Then, each data point is defined as part of the cluster with the closer centroid, with the distances being measured through typical equations such as the Euclidean formula. After this, the centroids of each cluster will be redefined by the average of all points that were defined as part of it. Having a new distribution of centroids, the data points will be reorganized to be part of the closest cluster, which can be same as before or not. This process is repeated until there are not points changing to a new cluster, meaning that the data was correctly organized. An example of this process is represented in Figure 2.2 [16] [17].

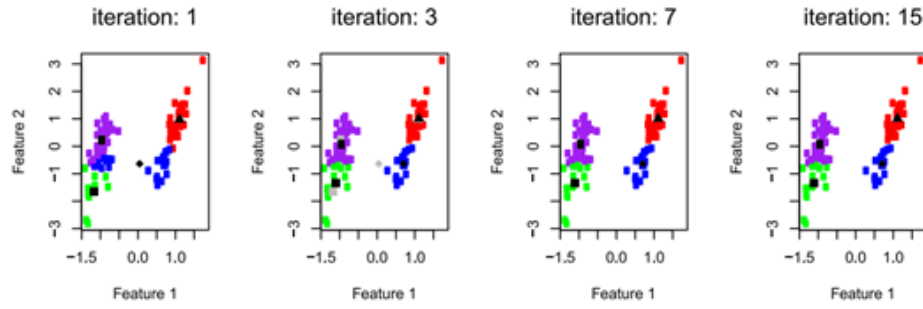


Figure 2.2: Clustering operation [17].

Mathematical model

The mathematical model behind the K-means algorithm is explained in this section.

Considering that $X = x_1, \dots, x_n$ is a dataset of a d-dimensional euclidian space, $A = a_1, \dots, a_c$ the cluster centers and $z = [z_{ij}]_{n \times c}$ where z_{ij} is a variable binary that indicates whether point x_i belongs to the k -th cluster, $k=1, \dots, c$ [17].

Equation 2.7 defines the objective function of K-means [17].

$$J(z, A) = \sum_{i=1}^n \sum_{k=1}^c x_{ij} \|x_i - a_k\|^2 \quad (2.7)$$

This algorithm is iterated through necessary conditions to minimize the objective function with equations that update cluster centers (Equation 2.8) and memberships (Equation 2.9) [17].

$$a_k = \frac{\sum_{i=1}^n z_{ik} x_{ik}}{\sum_{i=1}^n z_{ik}} \quad (2.8)$$

$$z_{ik} = \begin{cases} 1, & \text{if } \|x_i - a_k\|^2 = \min_{1 \leq k \leq c} \|x_i - a_k\|^2 \\ 0, & \text{otherwise} \end{cases} \quad (2.9)$$

Where $\|x_i - a_k\|$ represents the Euclidean distance between the data point x_i and the cluster center a_k [17].

2.5 Main metrics to evaluate last mile distribution

In order to evaluate the opportunities for improvements in the last mile distribution, it is necessary to have metrics capable of assessing its performance. As already mentioned, last mile distribution is the fraction of all distribution that has direct contact with the customer and carries the highest cost percentage. Close contact with the customer entails the responsibility of meeting the defined delivery deadlines. If the customer orders a next-day delivery service, this must be fulfilled to ensure the company's positive performance [1].

On the other hand, and at a more internal level, it is necessary to evaluate the distribution concerning its costs. These costs can be related to the delivery process but also to the used resources. Additionally, the distance and execution time can also be analyzed to understand the performance of the distribution, since these parameters are expected to be as small as possible [1]. Finally, the total number of objects should also be studied, as the delivery of more orders in the same route can be considered a point of improvement in the last mile distribution.

Finally, it is essential to realize that the metrics must be balanced as there is a trade-off between the cost of operation and the level of service. More clearly, the isolated increase in the level of service implies the existence of a more significant number of routes so that failures in delivery times are minimized. However, the necessary increase in routes to increase the level of service leads to higher costs. Bearing this in mind, paying attention to the balance of both metrics is essential to arriving at an optimal solution.

Chapter 3

Companies contextualization

3.1 Contextualization of the consultant company

The project was developed by a consulting firm external to the postal company. The company performs analytical management consulting to provide highly detailed and customized solutions according to the client's needs. It stands out for its:

- **Analytical proficiency** since the combination of advanced analytics with business experience promotes significant improvements in the profitability of results;
- **Holistic approach** through complex and extensive work with data that allows overcoming complex challenges with a holistic approach;
- **Delivery capacity** as it proposes a reduced implementation lead time;
- **Flexibility and adaptation** ensuring that solutions are fully designed for the customer, guaranteeing context-oriented customization.

3.2 Postal company context

The human being is a social species that, from an early age, works toward establishing closer relationships with other people. The letter is one of the oldest methods of approximating people physically far away. With the establishment of increasingly widened communication networks, the need to create a postal service specialized in the distribution of mail increased. In addition, and with the emergence of this need, postal delivery was adapted to include the distribution of larger packages.

A postal service can be defined as an organization responsible for the distribution and delivery of correspondence between different entities, from letters to large packages, throughout an entire specific region, guaranteeing a certain level of quality and affordable prices to meet the communicative needs of the population. Both private individuals and companies can use this service. This work aimed to optimize and simulate routes in order to ensure that the level of service and cost is reached.

3.2.1 Mail type

To match the needs of customers' delivery times, the postal company created a faster delivery service that guarantees the mail arrives at its destination within one working day. Higher costs are associated with this service to prioritize its delivery over others with a less urgent and longer delivery time. Therefore, it is essential to distinguish between the product type in regular mail and express mail (EMS).

3.2.2 Product matrix

To promote a consistent organization, the postal company defines a product matrix, shown in Figure 3.1, to include all the distributed objects in one category.

	LLE (Letters & large envelopes)	SP (Small packages)	MP (Medium packages)	P (packages)	LNP (Large & non standard packages)
Dimensions (mm)	330x250x12	400x300x100	500x300x300	1280x800x700	1600x1000x1000
Weight (kg)	0.35	2	5	30	60

Figure 3.1: Product matrix.

As can be seen in Figure 3.1, the products are divided accordingly to their dimensions. They can fit into any of the five existing categories, with the products belonging to the LLE (letters large envelopes) group being the smallest and the products belonging to the group of the LNP (large non standard packages) being considered the largest. This product matrix is essential to define what type of product can be allocated to each type of vehicle (walking, bicycle, motorbike, vehicle), with smaller products only being assigned to vehicles with a smaller capacity, as well as the opposite, as explicit in Figure 3.2.

	LLE (Letters & large envelopes)	SP (Small packages)	MP (Medium packages)	P (packages)	LNP (Large & non standard packages)
Walking (A)	✓	✓	✗	✗	✗
Bicycle (B)	✓	✓	✗	✗	✗
Motorbike (C)	✓	✓	✗	✗	✗
Vehicle (V)	✓	✓	✓	✓	✓

Figure 3.2: Product matrix and the appropriate vehicle restrictions.

In addition to the above, the products are also classified according to their trace and pattern. The trace is related to the possibility of the customer knowing the product's location. Therefore, this state is binary, which means that the product either has a trace or it does not have a trace.

On the other hand, the pattern dictates the maximum number of days the object has to be delivered, which varies between one and five days. The standard product is the traceless LLE with a three-day pattern, commonly called a letter.

3.2.3 Operational infrastructure

Currently, the postal service presents a vertical segmentation of operational infrastructures, shown in Figure 3.3, which has different functionalities.

Operational infrastructure	Number in Portugal	Product types	Localization
Production and Logistics Centers (CPL)	2	Mail	1 (CPL + CO) per forwarding region (RE)
Operational Centers (CO)	2	Express	
Logistics and Distribution Centers (CLD)	12	Mail and express	1 per forwarding group
Postal Distribution Centers (CDP)	217	Mail and express	Distributed throughout the country

Figure 3.3: Operational infrastructure.

As observed in Figure 3.3, the number of infrastructures in each division level is more significant the more proximal to the influence area, and Figure 3.4 shows the relationship between each group.

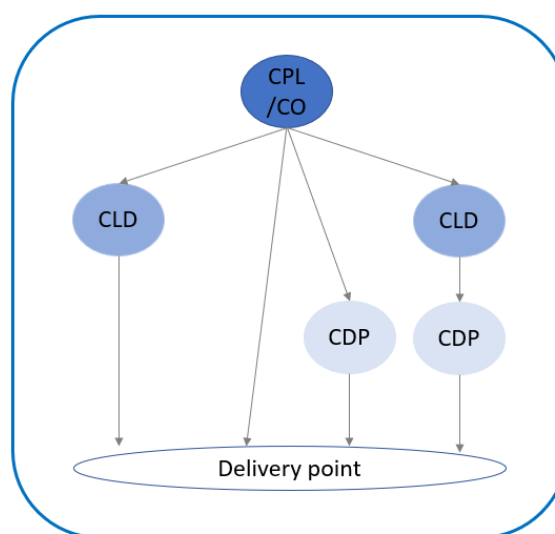


Figure 3.4: Relation between different infrastructures.

In each general distribution center, the least comprehensive distribution level, different rounds are defined, that is, the routes covered by the postman.

3.2.4 Postal code organization

The postal code was created to facilitate a postal address's logistical organization and location during deliveries. The Portuguese territory can be divided into 14 regions, 507 areas, and 184,803 arteries. The Portuguese postal code consists of 7 numbers. The first indicates the region, and the first four indicate one country area. The addition of the last three digits corresponds to the defined arteries, which, roughly speaking, can be said that the seven postal code numbers indicate the street. Finally, accurately adding each door number to its postal code means the delivery door.

As a rule, the postal organization uses a specific name to refer to an area (CP4), a street (CP7), and the complete information that includes the door number (CP10).

3.2.5 General tasks' view

For the complex functioning of a postal network, all the processes involved must be rigorous and organized. In general, the whole process can be divided into three parts:

- Reception: The object to be sent is received, and this reception can happen in different ways and is separated by format and divided to the CDP (postal distribution center)
- Division: The object undergoes a general separation
- Forwarding: Objects are forwarded to the location to be distributed.

Last mile distribution, the final stage of the delivery process, takes place mainly in postal distribution centers.

3.2.6 Types of delivery point

The objects transported by the postal company can be directed to delivery points with very different characteristics. The distinction between these points is essential since each one will be associated with different delivery times. Thus, the various points are shown in Table 3.1, with their respective description.

As is clear from Table 3.1, the points labeled “unico” refer to CP10, which contains only one door. On the contrary, the name “panel” refers to a CP10 with multiple doors, such as a building. Moreover, the denomination “residential” refers to residences, while “functional” refers to businesses, such as shops and cafes.

3.2.7 Distribution strategy

The organization of the last mile network is crucial to ensure efficient delivery of the product to the final customer. Thus, currently, two different strategies are used for this distribution that depends from distribution point to distribution point: the segmented strategy and the InTrevó strategy.

Table 3.1: Description of all delivery points defined by the postal company.

Delivery points	Definition
Residencial_RPD_unico	Distribution point refers to dwellings, residences with only one door
Functional_RPD_unico	Distribution point refers to commerce that has only one door
Post station	Stores and points associated with the postal company
Customers	Companies that receive large volumes of mail
Residencial_RPD_panel	Distribution point refers to residences contained in buildings, that is, delivery points with multiple doors
Functional_RPD_panel	Distribution point refers to buildings that have apartments dedicated to commerce, that is, delivery points with multiple doors

The first (segmented strategy) is organized by dividing the route into two distinct paths, the slow way (the one that takes both regular mail and priority mail) and the fast way, which is characterized by only stopping at delivery points where priority mail has to be delivered. This type of organization was designed considering the two routes are carried out by the same worker.

On the other hand, the InTrevò strategy divides the route into three different segments, making it possible to consolidate a more significant amount of non-priority mail. Each day, one of the three segments created is considered the slow route since it will be the one that will take the priority and non-priority products and by two fast routes. Only one worker also carries out this strategy.

Both segmented and InTrevò strategies are summarized in Figure 3.5.

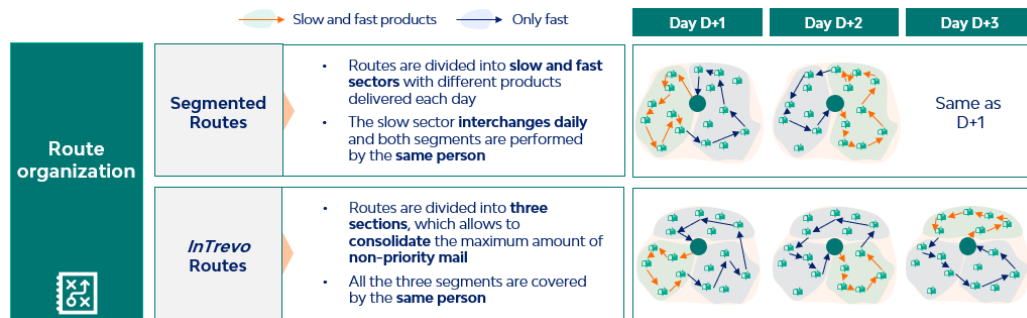


Figure 3.5: Scheme of route organization strategies.

In addition to the routing strategies mentioned, evaluating the integration of EMS (currently distributed by a specialized network - DV) in the postal service's base network is crucial. Thus, two distinct integration scenarios are evaluated: i) single network, which considers that all EMS mail is delivered from the postal distribution center, that is, by the base network, and ii) multiple network considers that EMS is delivered from the postal distribution center until it can do so and the rest is forwarded to the CO to be distributed by a specialized network.

The difference between single and multiple network are division is explicit in Figure 3.6.



Figure 3.6: Scheme of EMS integration.

3.2.8 Implementation of distribution strategies - Current process

Currently, to understand whether the strategy applied to a particular distribution center is the best or whether it is necessary to implement another strategy, a vast process consisting of four phases is essential:

- Prioritization of the study;
- Data collection;
- Optimization and validation;
- Study report.

The prioritization phase is the one that will define which of the postal distribution centers need a more urgent intervention through the evaluation of some general factors, such as for requests from local management or low productivity rates. Once the CDP that requires a more urgent reassessment is defined, the study process is then carried out, which encompasses the remaining three phases, in other words, data collection, optimization and validation, and the final report of the studies.

The data collection process is still very manual and time-consuming. At this point, the goal is to update traffic and service time per door. Firstly, it is necessary to extract the location of georeferenced doors and load this data into RouteSmart, which is a software used by the postal company whose objective is to define ideal routes for each distribution center, followed by loading a collection of motorized and non-motorized streets, roads, or paths, along with allowed traffic direction and location from the CDP and limits from the CP4. These data will then be added to the remaining information collected to achieve the objective of updating the previous values.

On the other hand, it is also necessary to extract historical traffic per object type (along the product matrix detailed above) from the last 12 months, which will be treated in two different ways: the three most significant outliers will be removed to define the traffic to optimize, and the traffic associated with unrecognized doors to the CP10 or the doors to be georeferenced will be distributed by CP7 doors and by georeferenced doors, respectively, to obtain the percentage of traffic per door. Additionally, the times of internal and delivery tasks are reviewed and adjusted according to on-site measurements.

Finally, the union of all the information will allow updating the traffic and the service time per door currently considered.

A routing strategy is initially defined in the optimization and validation process, which may be the InTrevo or the segmented. Afterward, all the parameters necessary for the optimization are configured, such as the time and maximum capacity for each route) and a parameterization of the reloading stops is carried out so that the optimization considers this need. After these steps are completed, the first process of optimization and simulation of the CDP routes under study is started. The output obtained through the optimization is analyzed and adjusted, for example, in the georeferencing data, the sequence of the route, or the points visited by it. After these adjustments, the model enters an iterative process generating new optimizations and route simulations until the ideal distribution is found.

Lastly, the results obtained are included in a complex study report that should present all the revised points and suggest a new strategy with the expected metrics.

Chapter 4

Methodology

As already mentioned, the main goal of this project is to generate optimized delivery routes for the postal service. To achieve this, it is necessary to collect and organize the data that will be used in the algorithm, as well as define the project goals. Then, the data will be processed in order to better understand the current state of the delivery system and also to prepare it for the next steps. Moreover, it is important to know and study the different restrictions imposed by the postal service.

After the preparation steps described above, the final results will be obtained through an approach that can be divided into three main stages: traffic clustering, route optimization, and route simulation for different traffic scenarios. This entire process will be leveraged by an interface that allows the parameterization of the models, and its resulting reports will be presented as dashboards for a more efficient analysis and interpretation. Finally, the results will be validated, and a pilot project will be implemented in a specific distribution center.

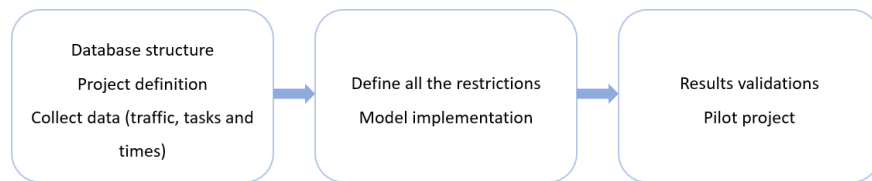


Figure 4.1: Method organization.

4.1 Database structure

One of the main requirements of developing and running an optimization and simulation model is having extensive network points, traffic and cost data correctly structured and updated. Therefore, it is necessary to explain the database structure where the data is stored. The realization of the project was based on the use of Microsoft SQL Server, a source capable of managing the database. This databased comprises a total of 190 tables populated by a wide variety of data provided by

the postal entity. The data in each table is considerable different from each other, and includes information related, for example, to the different CDPs or to the traffic of each delivered product.

The most generic tables are, for example, the *nodes* table that presents the data of all CDPs, such as the geographic location and the associated CO. In addition, the *products* table is also widely used because it is this one that presents the data of all products delivered by the postal company.

Figure 4.2 demonstrates the relationship between some tables that belong to the database.

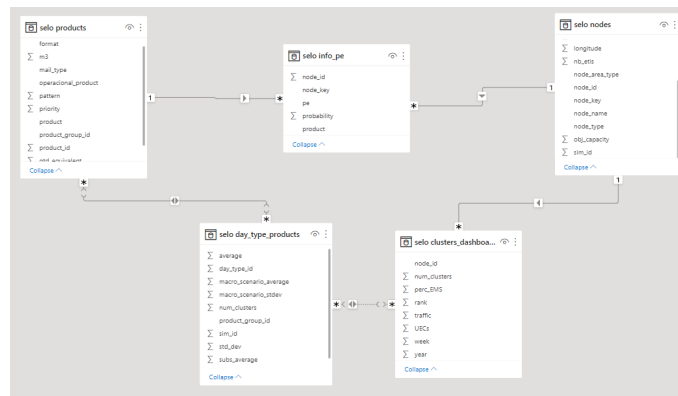


Figure 4.2: Example of the relationship between different tables.

4.2 Process

Before starting the project, it was necessary to understand the current weaknesses and improvement opportunities on the current process, how these problems could solve, and how a decision support system and tool could address these issues.

4.2.1 Use cases

Through the analysis of the current study processes, some weak points were found that led to the definition of three primary use cases:

- **Strategy evaluation:** This point emerges after perceiving the difficulty in quantifying the potential gains in implementing a new strategy. For the assessment, several traffic scenarios were considered (such as the baseline scenario, the optimistic scenario, and the pessimistic scenario) to be able to evaluate possible strategies for dividing products, distribution models, and what is the integration potential of the base network and the express network for each long-term CDP.
- **Support for studies:** It was found that the traffic considered for the current studies is static, with no consideration of the impact of variability. Furthermore, the current route generator (RouteSmart) has shortcomings, such as not considering the DVs and the limitation of

simulating only two types of speeds (walking and vehicle). Thus, this point should promote optimization of the organization of the last mile network for each CDP that is intended to be studied, as well as define and clarify the resource needs and productivity objectives to be met for each distribution center.

- **Follow-up and prioritization:** The lack of an agile calculation process that indicates which CDPs have the greatest need for review and the lack of a general metric comparison between each distribution point led to the creation of the third use case that aimed to monitor the efficiency of CDPs and to promote a continuous simulation of the optimization potential of each network.

4.2.2 Future process

The definition of the primary use cases led to the projection of a more optimized future process whose main objective is to respond to the weaknesses of the current study process. Thus, it was necessary to schematize how it will be carried out, being schematized in Figure 4.3 the logical sequence of this future projection.

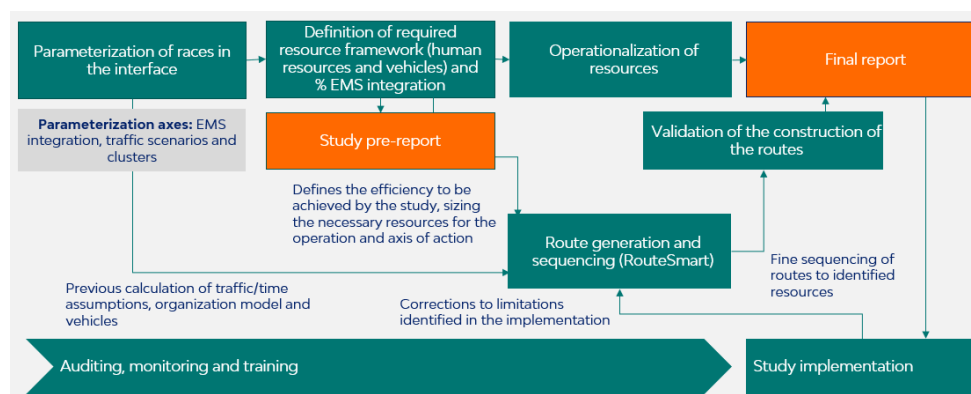


Figure 4.3: Future process.

4.3 Traffic

The historical traffic of all distribution centers was provided for the realization of the project. Firstly, it was necessary to process the data so it could respond to the needs of the model.

The first approach taken to group the data consists of grouping the traffic by product, that is, adding the amount of dispatched of each type of object per year. Subsequently, factors were applied to calculate the different macro scenarios to be considered.

Then, it was necessary to process these data to be used for both simulation and optimization. As the data source is the sorting machines that read the mail addresses, and since there is traffic that has to be manually sorted and does not have a more granular level (CP10) reading, several

histograms are needed at different levels of aggregation. So, four histograms were created to define different probabilities, referred to below:

- Probability with which each CP7 is visited, this probability being realized by CP4;
- Probability that, for each CP4, delivery is to one of the six different types of distribution points;
- Probability of delivery having a certain number of objects, by CP4;
- Probability of each CP10 being visited, filtering by CP7.

Thus, the model will use the cumulative probabilities taken from the historical data to define which CP7 and CP10 to visit, the number of objects delivered to each door, and the type of delivery point to which it refers.

Furthermore, to carry out the simulation and optimization stages, it was necessary to group the historical daily traffic data into clusters. Three types of clusters were generated to identify weeks along the year with similar traffic patterns (number of clusters = 1, 2, and 3) using a K-means algorithm, which the company's employees can analyze to draw relevant conclusions, for example, distinguishing times of greater flow of traffic (such as christmas) and the usual traffic volume. With the generation of clusters, it is possible to define the type of cluster to be optimized, and to simulate the process for various scenarios.

4.4 Tasks and times

The collection and time analysis of all tasks included in the last mile distribution was an important step in obtaining the necessary inputs for the simulation and optimization model.

The times that were most important in the process are:

- Node time: time taken to complete all tasks, from reception to dispatch, at the distribution point;
- Stem time: time taken by the postman to reach the first delivery point + time taken by the postman to return to the starting point since the last delivery point;
- Inter CP7 time: travel time between two consecutive CP7s;
- Delivery time: time that encompasses all stages of product delivery to the customer, considering actions such as ringing the bell when necessary.

On the other hand, it was essential to distinguish the delivery time of each of the different products of the operational matrix (Subsection 3.2.2) since the time considerations will change depending on the constraints of each object type. Appendix A.2 contains a detailed table of micro-tasks corresponding to the last mile distribution.

Finally, bearing in mind the standardization of the effort spent by each type of object, the UECs (Equivalent Mail Units) are calculated. This metric divides each object's unit time by the unit time of an LLE type, without a trace and with a pattern of three days.

4.5 Implementation

The model to be implemented consists of three segments: i) parameterization interface, created in Excel and VBA (Visual Basic for Applications), ii) the model in C# which, reading from the project database data and parameters, simulate and optimize the routes and finally, iii) a results dashboard in PowerBI that allows the visualization of the results obtained.

The data architecture to be used is presented in Figure 4.4.

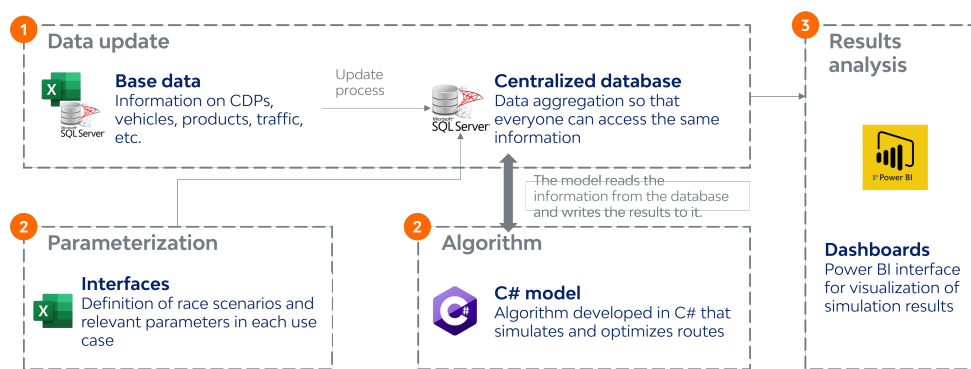


Figure 4.4: Data infrastructure.

As seen in the Figure 4.4, the project depends on base data contained in Excel files and company source databases, which are updated in a central database. Furthermore, the database connects the parameterization files and the model in C#. Finally, the information generated by the model can be exported to a dashboard to facilitate the visualization of the results.

Parameterization interface: This interface defines the different run scenarios and some of the relevant parameters for each of the use cases.

Model in C#: Algorithm developed whose function is the simulation and optimization of routes, considering the previously defined parameterizations.

Dashboard: For a more accessible and visual analysis of the results, reports are created and should show in detail the information provided by the simulation and optimization models.

4.6 Restrictions imposed by the company

The creation of ideal routes through a computational model is a robust process that must contemplate the real restrictions with which the fulfillment of the routes is faced. Thus, this section serves

to enumerate the primary limits imposed by the postal organization, as well as the assumptions made to create the model. The main limitations imposed are presented below.

- For journeys without a vehicle in larger urban and suburban CDPs, it is defined that if the distance to the first delivery point is greater than 6 km, the employee must use public transport;
- The distance between different points is calculated by multiplying a pre-calculated factor by the straight line distance between these two points and adding a constant distance per delivery point;
- Delivery time depends on the trace, the type of product (if it must be delivered by hand), and the delivery point;
- The route is designed considering an operating time of 7.5h. The modules are defined to receive regular mail and twice the priority mail in your area of influence;
- Motorization is assigned to minimize the cost of the unit object, ensuring capacity constraints;
- Engine costs include fuel, depreciation, and repairs for each of them: Motorcycles – 0.268 €/km and Vehicle – 0.313 €/km;
- Vehicle routes have no capacity restriction. However, bicycles and motorcycles have a restriction of 0.06 m^3 , and postman carts must not exceed 0.07 m^3 ;
- Walking and cycling routes are limited to a maximum of 13 km and 17.5 km, respectively;
- Redistribution: Addition of distribution points to vehicle routes and extra time factors when exceeding capacity;
- Replenishment: Initially, routes without vehicles have a capacity of 0.07 m^3 . However, when resupplying, it cannot exceed 0.06 m^3 as the carry bag may still contain some mail when resupplying is performed;
- Objects are forwarded to the next delivery day when the route exceeds the daily limit, according to i) clearance until the delivery deadline; ii) priority.

4.7 Clustering, simulation and optimization approach

Simulation and optimization are the critical points for carrying out the project, and then the logical structure of their operation will be presented. Therefore, it is crucial to highlight that both the optimization and the simulation generated, taking into account the definition of parameters in the tool developed in Excel and VBA, are explained below.

Both simulation and optimization need the clusters to run, so it is necessary to run the clustering algorithm previously.

4.7.1 Clustering

The clusters are created to group the weeks with similar traffic. Thus, the input variables the model needs to evaluate will be UECs and the percentage of EMS sent. With this data, the model can iteratively group the most similar weeks. For each distribution center, a different number of clusters are defined, and the algorithm is run considering the number of clusters equal to one, two, and three. Thus, the operator can adjust, in the optimization, what type of traffic he wants to consider.

4.7.2 Simulation

The simulation is performed through the model, so it is necessary to have the traffic considered and the number of days to be simulated as inputs. By default, the number of days to be simulated is 10, a strategic number chosen to be able to simulate each branch of each strategy more than once. It is important to note that the first three days of the simulation must be disregarded as they do not consider the undelivered mail from the previous day.

Furthermore, to carry out the simulation, it is also necessary to have access to the histograms created so that the random distribution of orders follows the historical record.

4.7.3 Optimization

For optimization, the model must have access to traffic data, the optimization percentile, the generated histograms, and the clusters. The user can manually define the optimization percentile, however, a default value of 80% will be defined.

The total traffic will be divided through histograms so that the routes connecting the different delivery points are optimal.

4.8 Deep dive C# model

The model developed in C# was structured modularly, with each module corresponding to one of the different use cases: the evaluation of strategies, support for studies, and follow-up and prioritization. All use cases are organized similarly, running the optimization function first, followed by the optimization function. Although the run functions are the same for all use cases and, consequently, the detail is the same in all runs, the results are saved in different ways to respond to each defined use case's objectives.

The optimization function consists of three main parts: the function to load the necessary data in the model, the function that runs the optimization, and the function that updates and stores the values generated by the run. The optimization process is iterative and progressively adds the closest CP7s to the previous point, creating a logical path of sequentially close points. The maximum CP7s ceiling is given when the total time reaches the maximum route time (7.5 h). This process considers the various vehicles available for carrying out the route, and the optimal route chosen will be the one with the lowest total cost.

With the assignment of CP7s to each route, the simulation is performed. This function is structured similarly to optimization since it presents a part of load parameters, including the optimization results, the simulation run, and the loading and updating of the developments in tables contained in the database. In addition to those mentioned above, the simulation allows analyzing the aggregation of objects from late deliveries to the routes of the following days.

Figure 4.5 shows the algorithm modules structure.

```

1 Optimization and Simulations Model (pseudo-code)
2 --Modular division
3 1. Strategy evaluation
4   Optimization
5   Simulation
6   Save results
7 2. Support to studies
8   Optimization
9   Simulation
10  Save results
11 3. Follow up and prioritization
12   Optimization
13   Simulation
14   Save results
15
16 Optimization function - Optimization function is divided also in three times:
17   Load_data
18   Run_optimization
19   Save_data
20
21 Simulation function - Simulation function is quite similar with optimization:
22   Load_data
23   Run_simulation
24   Save_data

```

Figure 4.5: Modular structure of the algorithm showing the logic implemented.

4.9 Validation and tests

The validation of the results obtained by the model through comparison with current methods has the principle of guaranteeing the model's reliability and providing confidence in the use of the tool when it is implemented in the company.

4.9.1 Comparison with RouteSmart

RouteSmart is a powerful software as it contains highly detailed geographic information, allowing the user to generate routes based on actual streets, roads and pathways. However is unable to map doors that are not georeferenced, and fails to consider operational challenges such as the need to have resupply points, other non-delivery tasks. Configuring it is also very time-consuming, which means generating different routing strategy or traffic scenarios is not practical. The developed model intends to abstract from the highly detailed geographic routing by applying distance factors that - on a broad analysis of results - would approximate more accurate models with a low margin of error. The developed model must be compared with this standard software in order to understand if general metrics, such as the number of stops, the number of objects delivered, the time, and the distance of each route, make sense. This comparison will be one of the most critical outputs for validating the generated model.

It is important to note that the model's objective is not to replace RouteSmart but to generate outputs that can serve as input for this software, such as the number of vehicles to be used.

However, although this platform's outputs help validate the model, a critical analysis of the results is necessary since RouteSmart also has some weaknesses, such as the impossibility of incorporating the DVs.

4.9.2 Comparison with current study

The validation of the results obtained was carried out in comparison with the latest studies by the company.

Comparing some metrics make it possible to understand how the model approaches the results previously obtained through carrying out studies of different distribution centers. The main metrics to evaluate may be the number of routes, the type of vehicles that each one needs, as well as the time, the distance traveled, the number of objects, and the number of workers needed (ETI). Thus, if the results obtained are similar or even better, the trust in the tool will be high.

4.10 Pilot project

Finally, a pilot project was implemented in four previously defined postal distribution centers (CDP) spread across the country. The CDPs chosen were A (rural), B (urban), C, and D (suburban).

The main objective of this project is to follow the evolution of metrics and implement the tool. Through this pilot project, it is possible to draw any conclusions. However, it was implemented at the beginning of July, so there will be no presentable results related to this part in Chapter 5 of the dissertation.

Chapter 5

Results and Discussion

5.1 Interface development results

The parameterization process is divided into two interfaces created in Excel, with a similar visual structure but with different end applications and target users. In this way, a parameterization interface was designed to be available for use by people from the company's central team. This interface will define general parameters related to the times associated with each task, network points, and traffic, as seen in the initial menu shown in Figure 5.1.

SIMULATION ENVIRONMENT FOR LAST MILE OPTIMIZATION		
Tempo de Tarefas	Pontos de Rede / Geografia	Produto / Tráfego
- Pressupostos Gerais - Pressupostos Conversores - Coefficientes Produção tarefas internas - Coefficientes Produção tarefas externas - Coefficientes Entrega agregados - Coefficientes Produção Global - Separação Geral Volumosos PMI - Tempos Padrão Tarefas Internas - Tarefas Internas - Formato - Tarefas Interfas - CDP Rastreabilidade Formato - Tarefas Internas - Rastreabilidade Formato Veículo Avisado - Tarefas Internas - CDP Rastreabilidade Formato Nivel de Divisão a Montante - Tarefas Internas - Formato Nivel de Divisão a Montante - Tarefas Internas - Formato Nivel de Divisão a Montante PE	- Importar Coordenadas CP7 - Pontos de rede - CDP Orgânico - Limitador Servidor de Decisões - Custos - Relação CP7 Giro	- Importar Dados RMS - Importar Dados MMS - Importar Tráfego DOL - Produtos - Tráfego Tipologia PE - Recolhas programadas - Produto por veículo - Taxa de avisados - Importar Recolhas pontuais - Parâmetro R - Parâmetro N - Parâmetro S/T - Parâmetro U

Figure 5.1: Parameterization interface menu.

On the other hand, the second interface generated is designed to allow configuration of each of the strategic/tactical use cases, whose menu is shown in Figure 5.2, developed for use by CDP teams. As seen in Figure 5.2, this interface will allow the definition of parameters related to information management, evaluation of strategies, CDP study, and performance simulation.

Both interfaces are linked to the database that contains the necessary information, with a bidirectional flow between the information in the database and that can be modified by the general team in Excel.

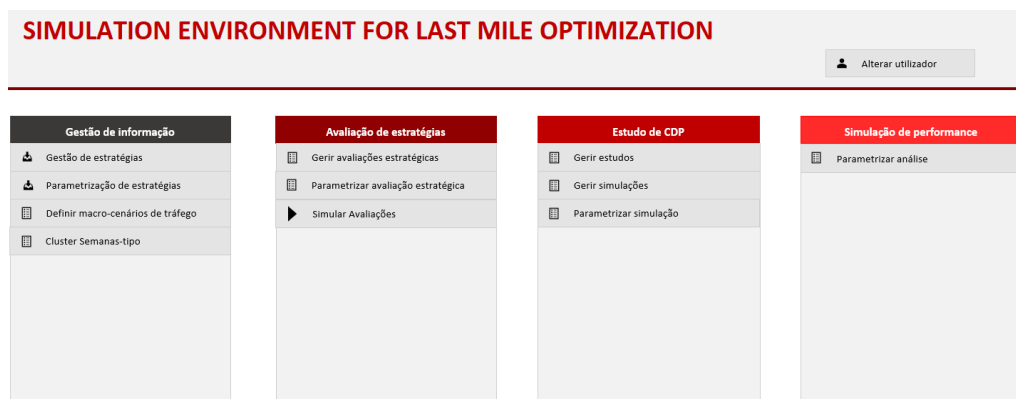


Figure 5.2: Configuration interface menu.

5.1.1 Parametrization interface

The parameterization interface is built to be an intuitive and easy-to-use tool containing the postal company's general parameters, which only the central management team will be able to access and modify. However, this is a central interface, and therefore, changes will impact all postal points in the country. For example, if the operational matrix of products, represented in a table in this interface, is changed, all distribution centers will have to adapt.

The tool was designed to present tables that can be changed and the assumptions used. In addition, the tabular representation is an advantage since it allows more flexibility analyzing the data in general and, consequently, identifying points to be changed.

As already mentioned, the interface was created to promote a bidirectional flow with the database and ensure information's easy and intuitive updating.

As can be seen in Figure 5.1, the menu of this interface is divided into three large groups: "Task times", "Network/Geography Points" and "Product/Traffic".

In general, the sheets in each group present the structure shown in Figure 5.3.

Produtos						
Atualizar produtos		Guardar alteração de produtos				
Produto	Rastreabilidade	Formato	Padrão de Entrega	Prioridade	Grupo de Produto	Volume
c/trace_LLE_d+1	c/trace	LLE	1	2	LLE_Prioritario	0.000219721
c/trace_LLE_d+2	c/trace	LLE	2	2	LLE_Prioritario	0.000219721
c/trace_LNP_d+1	c/trace	LNP	1	2	P/LNP	0.829166667
c/trace_LNP_d+2	c/trace	LNP	2	2	P/LNP	0.829166667
c/trace_MP_d+1	c/trace	MP	1	2	MP	0.023198946
c/trace_MP_d+2	c/trace	MP	2	2	MP	0.023198946
c/trace_P_d+1	c/trace	P	1	2	P/LNP	0.086431232
c/trace_P_d+2	c/trace	P	2	2	P/LNP	0.086431232

Figure 5.3: Example sheet of the parameterization interface features.

Using Figure 5.3 as an example, the sheet has embedded a function in VBA (Visual Basic for Applications) so that it can change the values of the last column through the button "Save product changes", automatically updating the table saved in the database. On the other hand, the "Update products" button imports and updates the values present in the database for the interface.

5.1.2 Configuration interface

The configuration interface structure is very similar to the previous one. However, while the parameterization interface is dedicated to the definition and modification of general parameters, the configuration interface was built to be used by the team responsible for the studies, and is oriented to the configuration of the 3 use cases.

The menu of this interface is shown in Figure 5.2. It presents four main groups: “Information management”, “Strategy evaluation”, “CDP Study” and “Performance simulation”, the last three referring to each of the three use cases defined at the beginning of the work.

The interface is very sequential, and, for example, before being able to parameterize the “Evaluation Parameterization” sheet, it is necessary to create and define strategies using the “Evaluation of Strategies” options. Figures 5.4 and 5.5 shows an example sheet and an example of a form that must be filled out to create the simulation, respectively. The primary purpose of presenting this page is to mention that default values were designed for certain variables but that these can be changed and reset by the user.

Gestão de simulações										
Estudo		Teste Oufvelas	Atualizar simulações	Adicionar simulação	Duplicar simulação	Apagar simulação				
ID simulação	Descrição	Estado	Data Início Corrida	Data Fim Corrida	Criado por	Macro-cenário	Dias simulados	Percentil de Otimização	Fator de tráfego	Número de clusters
76	Simulação 1	0	6/9/2022	6/9/2022	SELO	Baseline	10	0.8	1	3
77	Simulação 2	0	6/20/2022	6/20/2022	SELO	Baseline	10	0.8	1	3
84	Simulação 3	0	6/28/2022	6/28/2022	SELO	Baseline	10	0.8	1	1
85	Simulação 4	0	6/28/2022	6/28/2022	SELO	Baseline	10	0.8	1	2

Figure 5.4: Example sheet of the configuration interface.

Adicionar Simulação

Descrição:

Macro-Cenário:

Fator de Tráfego:

1

Dias a simular:

10

Percentil de otimização:

0.8

Número de clusters:

Adicionar Simulação

Figure 5.5: Example of a form created in the configurations interface.

5.2 Model development results

This section presents the results obtained by the model, both from a more general descriptive overview and a detailed analysis of the outcome. In addition, this section also explains the model validation results.

5.2.1 Descriptive analysis of results

In this section, generic results obtained by the model will be presented. This analysis is essential to get an overview of the actual impact that the implementation of the model may have on the postal company.

5.2.1.1 Current As-Is e Optimized As-Is

The first results are related to the generality of the postal organization. They aim to compare the metrics obtained by the current generation of routes, current As-Is, and the optimized generation of routes. The main metrics analyzed are:

- The number of routes;
- The percentage of routes through delivery (walking, bicycle, motorbike, and vehicle);
- The cost per object;
- The number of hours spent;
- The number of kilometers required to fulfill the deliveries.

Table 5.1 presents the results of the primary metrics listed above for the current process, and the model created.

Table 5.1: Comparison of current As-Is and Optimized As-Is metrics.

	Current As-Is	Optimized As-Is	Difference
Number of routes	5418	4740	-13%
Total Time (h)	-	32.6k	-
Total kilometers (km)	-	264.5k	-
Price per object (€/object)	-	0.229	-
Percentage of routes by walking (%)	10	20	100%
Percentage of routes by bicycle (%)	8	2	300%
Percentage of routes by motorbike (%)	35	38	-8%
Percentage of routes by vehicle (%)	46	40	-13%

As can be seen in Table 5.1, the number of routes obtained through the model is considerably lower than the number of routes generated by the current process, corresponding to a percentage decrease in the number of routes of about 13%.

The percentage of routes by vehicle was also a studied factor. There is an increase in the use of pedestrians and a decrease in the use of vehicles. These observations can be justified by the lower cost of using a pedestrian since the vehicle implies the extra cost of fuel, depreciation, and repairs. Although this last metric cannot be evaluated as a decrease in the overall cost, it serves as a good indicator that the routes generated by the model can guarantee some economic advantage for the postal service without a significant increase in number of resources needed.

Although there is a clear need to compare the largest number of metrics listed above and removed by the model, it was impossible to evaluate any more due to the lack of these very detailed parameters by the postal company.

5.2.1.2 Strategy results

In this section, each of the strategies was analyzed individually, that is, the segmented strategy and the InTrevo strategy with and without considering the CDP capacity for EMS integration, with the single network referring to disregarding the CDP capacity. In contrast, the multiple networks consider CDP capacity, with the remainder distributed through the nearest CO. Table 5.2 presents the results obtained for the main metrics studied for each tested scenario.

Table 5.2: Comparison of strategy results.

	Single		Multiple	
	Segmented	InTrevo	Segmented	InTrevo
Number of routes	4259	3934	4738	4413
Total time (h)	29.2k	26.8k	32.6k	30.1k
Total kilometers (km)	168.7k	164.9	265.9k	263.3k
Price per object (€/ object)	0.194	0.181	0.229	0.217
Percentage of routes by walking (%)	15	14	11	11
Percentage of routes by bicycle (%)	11	11	10	8
Percentage of routes by motorbike (%)	48	46	41	39
Percentage of routes by vehicle (%)	27	29	38	41
CDP - Mail	1726k	1705k	1714k	1709k
CDP - Packages	195k	196k	100k	101k
CO - Mail	0	0	0	0
CDP - Packages	0	0	96k	96k

It is important to keep in mind that the price per object is calculated by weighing all the factors that include monetary costs to the process, such as the number of people to be hired, the number of vehicles used as well as their typology [18]. The values expressed in Table 5.2 make it possible to perceive that there is a clear advantage in cases that apply the single network strategy. In general, all metrics analyzed are smaller when compared to the same metrics obtained by the strategy that involves EMS integration with capacity constraints. In particular, the value obtained for each object is considerably lower. For example, there is a difference of 0.035 € per object delivered for the segmented route strategy, and in the InTrevo strategy, the difference is 0.036 € per object. As it can be observed, the value appears constant when comparing both EMS integration strategies. On the other hand, the lowest price obtained through the single network may, in general, not prove to be advantageous since the value presented does not consider the costs of adjusting the installation capacity, and this conclusion should be studied more deeply for the application of the strategy to each postal distribution center.

Furthermore, focusing on the choice of the routing strategy, the InTrevo distribution presents itself as the most advantageous for both the single and multiple networks since all the metrics obtained are substantially smaller than those related to the segmented strategy. However, this analysis cannot be generalized to all postal distribution centers, so it is wrong to apply the strategy that presents the best metrics at the national level without the existence of a detailed study carried out in advance.

5.2.2 Comparison and validation of results

This section exposes the results that validate the model created by comparing it with the RouteSmart platform and the studies carried out to date. For this comparison to be valid, it is necessary to ensure that all inputs are consistent with those used for the studies and generating routes.

5.2.2.1 Comparison with studies

Comparing the results obtained by the model with those obtained by the studies is a good way of validating the tool since this is the method currently applied to define the implementation of new strategies in different distribution centers. However, for the comparison to be fair, it is necessary to bear in mind that the basis of the study routing strategy must be the same, ensuring that the results obtained are analyzable and concrete.

In this way, through the configuration tool, a simulation was created for one of the four distribution centers that will be studied, and, after running the model, three essential metrics were taken as output: the traffic considered for the simulation, the number of hours totals used daily and the number of workers (ETIs) required to fulfill these routes, the latter metric being directly linked to the number of routes (1 ETI = 1 route). The results obtained by both the current process and the optimized process, as well as the difference between them, are shown in Table 5.3.

Table 5.3: Comparison between the values obtained by the studies and the optimized model.

	Current Process	Optimized Process	Difference
ETIs	39	33	-6
Total time (h)	312	264	-48
Traffic ($\times 10^3$ objects)	28	32	4

As seen in Table 5.3, the traffic considered for the simulation is slightly higher than the traffic considered in the study carried out by the company. Therefore, the optimized route would guarantee the use of over 4000 objects. However, the metrics of the total daily time for carrying out the routes and the metrics of the number of ETIs are smaller for the optimized solution. Thus, it is possible to affirm that the optimized generation of the routes allows greater efficiency in the distribution of the products since in a shorter period and with a lower cost of hiring workers, it can deliver a greater volume of objects.

5.2.2.2 Comparison with RouteSmart

After comparing the simulations with the studies, it was necessary to compare with RouteSmart. This comparison consists of measuring the difference between this software's results and the model's results by evaluating the traffic considered, the number of stops, the percentage of CP7 visited, and the number of routes required. Thus, data related to a distribution center were extracted, and the values are shown in Table 5.4.

Table 5.4: Comparison between RouteSmart and the optimized model.

	RouteSmart	Optimized process
Traffic (nr objects)	28000	32000
Number of routes	35	33
Visited CP10 (%)	53	99
Number of stops	6600	9800

As can be seen through the analysis of Table 5.4, there is a considerable difference in the traffic to be analyzed, with the traffic used for route optimization being 15% higher than the traffic used by RouteSmart. This difference should be as slight as possible for the conclusions to be linear. However, obtaining the exact traffic data was impossible. Therefore, the values obtained must be subject to a new comparison. A clear difference is evident in the percentage of CP10 visited, as the optimized model assumes that 99% of CP10 are visited, unlike RouteSmart. This difference can be

explained by the model's limitation of considering all the existing doors, unlike the currently-used software.

On the other hand, the distance traveled associated with the optimized solution is much greater (606 km) compared to the solution obtained through RouteSmart (494 km). This difference can be explained due to the different traffic considered. However, as the distance traveled does not vary linearly with traffic, it is impossible to reach a conclusion using this metric. Furthermore, the behavior of the number of stops is identical to that of the distance traveled, being considerably higher for the optimized solution. Thus, it becomes impossible to make some conclusions regarding these two metrics.

Finally, comparing the number of routes is the only parameter that can be analyzed since the model with a greater volume of traffic generates a smaller number of routes (33 compared to the 35 of RouteSmart), implying a greater efficiency in the delivery of objects in the optimized process. However, the analysis of this parameter alone does not allow a concrete conclusion, making it necessary to carry out new comparisons to fix the traffic parameter.

5.2.2.3 Route visualization

In addition to the comparisons with the studies and with RouteSmart, it was also necessary to understand if the formation of the routes was geographically compatible with the expected, that is, the different parts of the route should be balanced with each other and the delivery points of each of the different parts of the route should be relatively close together. Thus, it was necessary to make a geographical plot to mark the delivery points, and it was required to evaluate, in a little analytical way, if the construction of the route followed the indicated parameters. Figure 5.6 shows a randomly chosen route and Figure 5.7 refers to the delivery points for each one of the segments (A, B, and C). The route represented in Figure 5.6 presents an InTrevio configuration, therefore, it has three routes (Figure 5.7). For each of them, the distribution of delivery points and the number of doors that must be visited are analyzed. At the geographic level, each route appears to include doors considerably close to each other, so the route is successfully validated. On the other hand, the number of doors associated with each of the three routes is shown in Table 5.5 and, as observed, that parameter varies only by three doors, with an average of 13 doors visited by each route. Thus, it is possible to conclude that the routes are being generated correctly, validating their possible implementation.

Table 5.5: Number of delivery points for each segment of the presented route.

Route	Number of delivery points
A	14
B	11
C	13

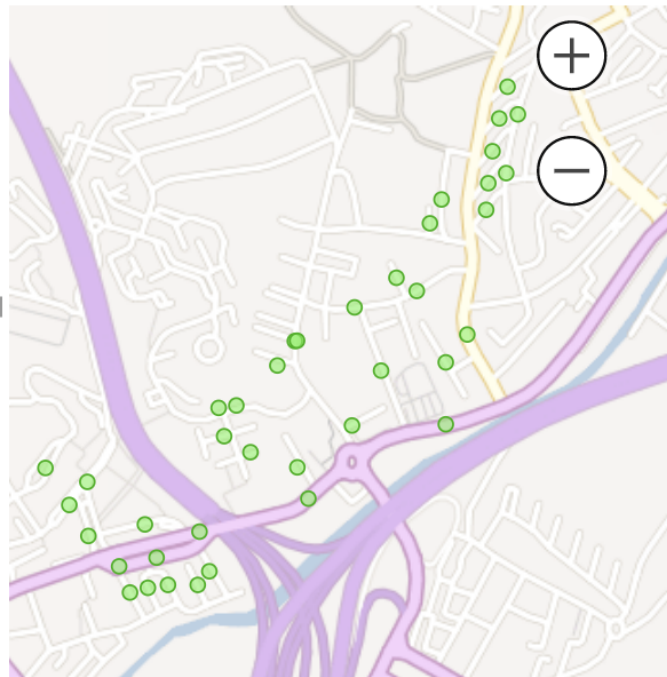


Figure 5.6: Route overview.

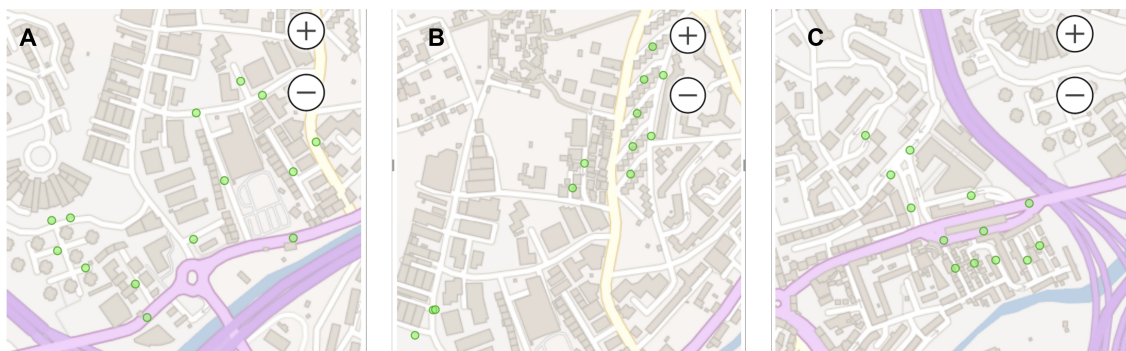


Figure 5.7: Detailed view of each part (A, B and C) of the route shown in Figure 5.6.

5.2.3 Strategies optimization

This section intends to discuss the results obtained for each pilot CDP in a deep study regarding optimization. In this way, points such as the strategies that best fit each CDP will be discussed.

5.2.3.1 Network integration strategy

The definition of the network integration strategy for each postal distribution center is only possible through a detailed analysis of the metrics obtained for each tested option (single and multiple networks). The results obtained regarding the number of routes, the total distance traveled, the

entire time to complete the route, and the average cost of delivery of each object are presented in detail in Table 5.6.

Table 5.6: Comparison between single and multiple networks for each of the pilot CDPs.

	Multiple Network				Single Network			
	C	B	A	D	C	B	A	D
Number of routes	55	71	19	30	43	54	13	19
Total distance (km)	1278	2073	2273	2498	704	949	1122	480
Total time h)	334	382	117	172	296	360	86	113
Cost per object (€/object)	0.16	0.20	0.49	0.37	0.14	0.17	0.32	0.22

To facilitate the analysis of the values presented in Table 5.6, the difference between both strategies was calculated for each metric presented through Equation 5.1.

$$\text{Difference} = \text{multiple network metric values} - \text{single network metric values} \quad (5.1)$$

Table 5.7: Differences found between the values presented in the Table 5.6.

	C	B	A	D
Number of routes	12	17	6	11
Total distance (km)	574	1124	1151	2018
Total time h)	38	22	31	59
Cost per object (€/object)	0.02	0.03	0.17	0.15

In the first analysis, and only through observing the positive sign, it is possible to see that all values obtained for the single network strategy are smaller compared to the strategy that includes the load constraint. However, as mentioned in subsection 5.2.1.2, although the single network presents itself as the most promising strategy, the costs of increasing the required capacity can be very high. Therefore, the balance of both may prefer using the multiple network.

Having this in mind, the capacity increase data, presented in Table 5.8, which would be necessary for the implementation of the single network, was collected to understand more closely whether this strategy may be applicable or not.

A: The analysis obtained for the A distribution center makes it possible to perceive a negligible difference in the number of routes (the use of the multiple network strategy only presents six more

Table 5.8: Presentation of the values of current capacity and extra capacity needed for the implementation of the single network.

	C	B	A	D
Current capacity (nr objects)	1000	800	250	600
Extra capacity needed (nr objects)	1781	1372	62	353
Increase (%)	178%	172%	25%	59%

routes than the single network). However, the difference in the distance traveled is very considerable, with a difference of more than 1000 km. The relationship of these metrics may indicate that the efficiency of the routes is considerably lower in the routes that consider the maximum capacity of the CDP since, for the same delivery points and the same number of routes, it is necessary to travel a much greater distance.

On the other hand, the time is taken to complete the routes also presents a more negligible difference than expected for the disparity of kilometers traveled, so it can be deduced that the routes idealized for complete integration of EMS are much more time-consuming. Furthermore, the difference between the costs per object is quite considerable (0.17 €/object), giving a clear advantage to using the single network.

Finally, analyzing the table that demonstrates the increase in capacity required, it is possible to see that the capacity to be increased is relatively low, and applying this strategy may be advantageous. However, it is difficult to withdraw final decisions without carrying out a financial analysis by the postal company.

B: Regarding the results obtained for the B CDP, it is possible to perceive that the application of the multiple network strategy presents 17 more routes than the single network. On the other hand, the distance traveled and the total time of completion of the routes is 1124 km and 38 h, respectively, presenting relatively high values. By merely analyzing these metrics, the strategy that best suits this CDP would be the single network. However, the difference between the costs per object is only 0.03 €/object, the need to expand the capacity is 1372 objects, and the space would have to increase the area substantially. Thus, considering that the price per object is not considerably different, it is concluded that the best integration strategy is the multiple network.

C: The difference between the metrics of the different types of integration for the C distribution center is quite similar to that obtained for the B distribution center. However, the difference between the total distance traveled is more negligible, compensating for the longest full-time completion of the routes.

On the other hand, applying the single network strategy is associated with a capacity increase of 1781 objects, which corresponds to more than double the current capacity. Thus, the strategy chosen for implementation in this CDP is the multiple network.

D: Finally, when comparing the different metrics for the CDP of D, the high difference in both the time of completion of the route (59h) and the distance traveled (2018 km) is highlighted. Furthermore, the route number difference is 11, and the price per object is 0.15 €/object. Analyzing the capacity that will have to be increased for full integration, the postal distribution center should be able to accommodate another 353 objects, about half of CDP's current capacity. Thus, similarly to the situation of the postal distribution center, it is necessary to assess more deeply the economic expense that this installation would imply.

5.2.3.2 Route strategy

After choosing the integration of networks, it is essential to analyze which strategies best adapt to the CDPs selected for carrying out the pilot project. Thus, metrics related to the number of routes generated, the total distance traveled daily, the total daily time traveled, and the delivery cost per object was extracted and compared between the two routing strategies. Tables 5.9 and 5.10, following what was done to define the best network integration strategy, present the values obtained for each strategy and the difference found between them, respectively. The difference calculation is performed by applying Equation 5.2.

$$\text{Difference} = \text{segmented strategy metric values} - \text{InTrevo strategy metric values} \quad (5.2)$$

Table 5.9: Comparison between segmented and InTrevo strategies for each of the pilot CDPs.

	Segmented				InTrevo			
	C	B	A	D	C	B	A	D
Number of routes	36	32	9	11	33	30	8	11
Total distance (km)	629	596	1003	229	606	600	985	226
Total time (h)	284	250	72	88	265	238	63	88
Cost per object (€/object)	0.037	0.047	0.245	0.044	0.036	0.044	0.250	0.039

Table 5.10: Differences found between the values presented in the Table 5.9.

	C	B	A	D
Number of routes	3	2	1	0
Total distance (km)	23	-4	18	3
Total time h)	19	12	9	0
Cost per object (€/object)	0.001	0.003	-0.005	0.005

In a first analysis, it is possible to perceive the differences found between the strategies are considerably low. However, as the price per object does not consider traffic, this metric will be the most critical decision on which routing strategy to apply to each distribution center.

A: As it is possible to analyze through Table 5.9, this distribution center is the one that presents the most significant number of kilometers traveled daily, justified by the low population density of rural areas [19], the smallest number of routes, and the daily distribution time. With this analysis, it is possible to perceive that most routes are carried out using a motor vehicle. Thus, the greater use of this type of vehicle influences the increase in the price per object compared to the values of other distribution centers. In a more comparative view between strategies, it is possible to perceive that although the InTrevo strategy uses a smaller number of routes, which are carried out in a shorter distance (18 km difference) and less daily time (9 h difference), the amount spent per object is 0.005 € higher. So while some metrics apply to the advantage of using the InTrevo strategy, the existing difference is not considerable enough when compared to the difference in monetary value. Therefore, the strategy to be applied to the A distribution center is the segmented strategy.

B: In this distribution center, although the distance traveled is slightly greater (4 km) for the InTrevo strategy, the metrics of the number of routes and the time taken to carry them out are smaller in this strategy, with a difference of 2 and 12 hours, respectively. In addition, applying this tripartite distribution promotes a gain of 0.003 € per object delivered, making the strategy globally more adjusted for the distribution center.

C: The postal distribution center in C has the most significant variation between the metrics of the number of routes, daily distance traveled, and daily time required to fulfill the routes for each of the strategies studied. Thus, the InTrevo strategy presents a variation of 3 routes compared to the segmented strategy and a variation of the distance traveled daily of 23 km, promoting an increase of 19 hours in the total time of completion of the routes. However, although the cost difference was expected to be equally higher, this was the most negligible difference found at the study points. This metric may indicate that this distribution center will have many routes associated with pedestrians and bicycles. Finally, although the cost difference is the smallest found compared to the metrics evaluated from the other distribution centers, it is favorable for applying the InTrevo strategy, promoting 0.001 € savings per object.

D: Finally, the metrics collected for this distribution center are very close between the two strategies. The only variation is in the total distance traveled, which is only 3 km and in the cost per object, demonstrating that the application of the InTrevo strategy in the center distribution center in D should save the postal company 0.005 € for each item delivered.

5.2.3.3 Benefit estimation by traffic scenario

By choosing the best routing strategy for each postal distribution center under study and considering the limitation of the EMS integration capacity, since, in the absence of sufficient data, it is assumed that there is no financial capacity for the physical increase of CDPs, the gain potential was collected and calculated for each distribution point throughout the year, and for each macro scenario tested, with the detailed values being explicit in Table 5.11.

Table 5.11: Predicted gains for each of the macro scenarios tested.

Macro scenario	Name	Daily cost As-Is	Optimized daily cost (multiple network)	Daily savings	Annual savings
Baseline	C	€ 4,302.17	€ 4,103.65	€ 198.52	€ 51,616
	B	€ 5,368.48	€ 5,211.89	€ 156.59	€ 40,712
	A	€ 2,023.48	€ 1,989.35	€ 34.13	€ 8,874
	D	€ 2,914.27	€ 2,745.58	€ 168.69	€ 43,860
Pessimistic	C	€ 3,807.63	€ 3,673.56	€ 134.07	€ 34,858
	B	€ 5,024.62	€ 4,799.13	€ 225.49	€ 58,627
	A	€ 2,033.82	€ 1,810.15	€ 223.67	€ 58,155
	D	€ 2,716.47	€ 2,509.42	€ 207.05	€ 53,834
Conservative	C	€ 4,044.61	€ 3,858.98	€ 185.63	€ 48,265
	B	€ 5,263.65	€ 4,853.32	€ 410.34	€ 106,688
	A	€ 2,094.57	€ 1,970.19	€ 124.37	€ 32,337
	D	€ 2,648.21	€ 2,582.93	€ 65.28	€ 16,972
Optimistic	C	€ 4,434.76	€ 4,352.08	€ 82.68	€ 21,497
	B	€ 5,898.12	€ 5,577.84	€ 320.28	€ 83,272
	A	€ 2,100.85	€ 1,953.45	€ 147.41	€ 38,326
	D	€ 2,845.31	€ 2,842.13	€ 3.18	€ 826
Best Case	C	€ 5,310.50	€ 5,175.42	€ 135.07	€ 35,119
	B	€ 6,723.28	€ 6,524.15	€ 199.13	€ 51,773
	A	€ 2,302.38	€ 2,163.49	€ 138.90	€ 36,113
	D	€ 4,282.90	€ 4,282.90	€ 0	€ 0

As can be seen, the application of the model will guarantee gains for the postal company. Through the analysis of Table 5.11, the only exception found is related to the costs obtained by the optimization and the As-Is costs for the D CDP, using a “Best Case” traffic scenario. However, although the model implementation does not present a clear savings advantage, it is also essential to focus on the simplicity and ease guaranteed by the model.

In Table 5.12, the values associated with the worst gain scenarios for each of the four studied CDPs are marked in red, and those with the most significant economic advantage in green. As can be seen in Table 5.12, which summarizes the maximum and minimum savings values of each CDP, the B postal distribution center is the one with the most remarkable minimum (40 712 €)

and maximum (106 688 €) earning potential, and the implementation of the model should give priority to the points distribution systems that have the most advantages. Thus, the highest priority CDP is that of B and the last one is that of D, which is the one that can have a zero return.

Table 5.12: Maximum and minimum earnings for each pilot CDP.

	Maximum savings	Minimum savings
A	€ 58,155	€ 8,874
B	€ 106,688	€ 40,712
C	€ 51,616	€ 21,497
D	€ 53,834	€ 0

5.3 Report development results

The reports were created in PowerBi to become a decision support tool since its analysis facilitates the study of the main metrics. In addition, these reports add significant value to the final solution as they are interactive and changeable depending on what the user wants to analyze.

Three different reports were generated, one for evaluating strategies, another for carrying out studies, and the last for assessing performance. The main function of the report on evaluating strategies is to analyze the best method to adopt for each distribution center. In this way, the interface will present several metrics to measure for each strategy and guarantee a long-term vision, allowing a thoughtful choice of the best strategy.

Figure 5.8 shows some prints of the strategy evaluation report.

The report for carrying out the studies has two aspects: a more general view and a more detailed view of the simulation. In the broadest perspective, for a given study, it is possible to compare the simulation results for the two different organization strategies and each of the different clusters generated. Metrics analyzed in this more general interface are, for example, total daily cost, service failure risk, and predicted efficiency. Finally, and as a summary, a table succinctly presents all the daily indicators calculated.

On the other hand, the detailed view is placed on three different pages, which are related to the time spent in carrying out the tasks, allowing to analyze the times calculated in detail for each job and each simulation and cluster strategy to the simulation so that it will enable the detailed analysis of each metric extracted by the simulation of each strategy and each cluster, and, finally, a sheet that allows a detailed analysis of the routes, both at the geographic level and at the level of the number of objects and the time needed to the same.

Figure 5.9 shows screenshots of the interface, where it can visualize all the points mentioned above.



Figure 5.8: Print example of the evaluation report.

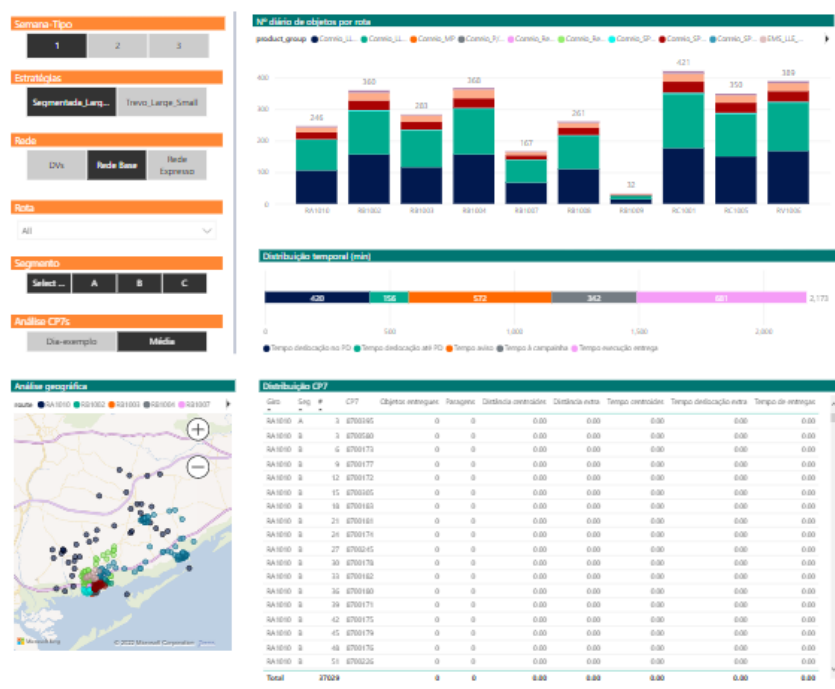


Figure 5.9: Print example of the study report.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

The analysis of the results obtained, explained in chapter 5, allowed us to draw some conclusions and outline the next steps that should be taken to arrive at a more sophisticated tool.

In this chapter, conclusions regarding the model in C# can only be presented since the results obtained both for the parameterization and configuration interface and for the reports are merely demonstrative. However, it is essential to mention that both platforms are fully functional and ready to be applied in the postal company.

Regarding the model, several conclusions can be drawn that will be separated by: general findings, comparisons and validations, and, finally, optimization of strategies, as divided in the presentation of the results.

Thus, by generally evaluating the performance metrics and the process currently implemented, it is possible to denote a clear advantage in applying the created model since it enhances the use of a smaller number of routes. In addition, the model generates an output with a more significant number of pedestrian routes and a smaller number of routes that use a vehicle for their realization compared to the current process. Additionally, the integration and routing strategies that would promote the best evaluation metrics in the company's global were compared. It was concluded that the single network strategy guarantees more significant savings. However, this also implies costs that increase capacity that may not compensate for the amount saved overall. On the other hand, the InTrevro strategy was the most advantageous. However, it is necessary to deeply evaluate each distribution point for more concrete conclusions and adjust to the needs.

On the other hand, the comparison of the results obtained with those of the current studies and RouteSmart for the same distribution point made it possible to perceive that there is potential for improvement through the use of the model, however, and specifically for the results compared with that of RouteSmart, did not allow for concrete conclusions to be reached. It is necessary to use the same traffic data for this purpose.

Finally, four specific distribution centers were studied, and the best strategies for each were analyzed, as well as the monetary benefit that the model implementation in these centers could

bring. The analysis of the network integration method led to tenuous conclusions since the variable costs of implementing more capacity are lacking. Thus, for the calculation of the monetary benefit, it was considered that all distribution centers under study are limited by their maximum capacity. Furthermore, regarding the routing strategies, only the A distribution center has advantages in using the segmented strategy, and all others should be limited to tripartite routing. Finally, excluding the clear advantage of automating the entire process, the monetary gain from applying the model in these four distribution centers could reach up to 106 688 €.

6.2 Further Work

After validating the model's outputs and obtaining satisfactory results, a pilot project will be carried out to monitor the model's implementation and evaluate metrics for the four defined postal distribution centers. The main goal of this project is to assess the impact of the strategies applied in different distribution centers. In addition, this project could serve as a mean to find small details and points to improve the model developed.

After carrying out the pilot project and confirming that the model's expected results were achieved, the project should be progressively expanded to the rest of the country. The application of the model to other distribution centers should be analyzed, and the distribution centers with the most significant savings potential should be the first to be modified.

Finally, and in the long-term view, the model should maintain time-scheduled maintenance to adapt to the small changes made by the postal company.

Bibliography

- [1] Gayani Bopage, Julian Nanayakkara, and Kasuni Vidanagamachchi. “A Strategic Model to Improve the Last Mile Delivery Performance in E-commerce Parcel Delivery”. In: Mar. 2019.
- [2] Tauseef Aized and Jagjit Srari. “Hierarchical modelling of Last Mile logistic distribution system”. In: *The International Journal of Advanced Manufacturing Technology* 70 (Feb. 2014). DOI: [10.1007/s00170-013-5349-3](https://doi.org/10.1007/s00170-013-5349-3).
- [3] Ján Dižo et al. “A mathematical model of operation of a semi-trailer tractorpowertrain”. In: *Komunikácie* (June 2022).
- [4] Stefan Irnich, Paolo Toth, and Daniele Vigo. “Chapter 1: The family of vehicle routing problems”. In: *Vehicle Routing*. Philadelphia, PA: Society for Industrial and Applied Mathematics, Nov. 2014, pp. 1–33.
- [5] Burak Eksioglu, Arif Volkan Vural, and Arnold Reisman. “The vehicle routing problem: A taxonomic review”. In: *Computers Industrial Engineering* 57.4 (2009).
- [6] Ján Dižo et al. “A mathematical model of operation of a semi-trailer tractor powertrain”. In: *Komunikácie* (June 2022).
- [7] Claudia Archetti, M. Grazia Speranza, and Daniele Vigo. “Chapter 10: Vehicle Routing Problems with Profits”. In: *Vehicle Routing*, pp. 273–297. DOI: [10.1137/1.9781611973594.ch10](https://doi.org/10.1137/1.9781611973594.ch10). eprint: <https://epubs.siam.org/doi/pdf/10.1137/1.9781611973594.ch10>. URL: <https://epubs.siam.org/doi/abs/10.1137/1.9781611973594.ch10>.
- [8] Guy Desaulniers et al. “VRP with Pickup and Delivery”. In: Jan. 2002, pp. 225–242. ISBN: 978-0-89871-498-2. DOI: [10.1137/1.9780898718515.ch9](https://doi.org/10.1137/1.9780898718515.ch9).
- [9] Brian Kallehauge et al. “Vehicle Routing Problem with Time Windows”. In: *Column Generation*. Ed. by Guy Desaulniers, Jacques Desrosiers, and Marius M. Solomon. Boston, MA: Springer US, 2005, pp. 67–98. ISBN: 978-0-387-25486-9. DOI: [10.1007/0-387-25486-2_3](https://doi.org/10.1007/0-387-25486-2_3). URL: https://doi.org/10.1007/0-387-25486-2_3.
- [10] Ján Dižo et al. “A mathematical model of operation of a semi-trailer tractor powertrain”. In: *Komunikácie* (June 2022).

- [11] Susmita Bandyopadhyay and Ranjan Bhattacharya. *Discrete and Continuous Simulation: Theory and Practice*. June 2014, pp. 1–346. ISBN: 9780429073472. DOI: [10 . 1201 / b17127](https://doi.org/10.1201/b17127).
- [12] Francis Lau and Craig Kuziemsky. *Handbook of eHealth Evaluation: An Evidence-based Approach*. Feb. 2017. ISBN: 978-1-55058-603-9.
- [13] Onur Özgün and Yaman Barlas. *Discrete vs. Continuous Simulation: When Does It Matter?* 2009.
- [14] Peer-Olaf Siebers et al. “Discrete-event simulation is dead, long live agent-based simulation!” In: *J. Simulation* 4.3 (2010), pp. 204–210. URL: <http://dblp.uni-trier.de/db/journals/jos/jos4.html#SiebersMGBP10>.
- [15] Glenn Fung. “A Comprehensive Overview of Basic Clustering Algorithms”. In: *Citeseer* (Jan. 2001).
- [16] Mayra Rodriguez et al. “Clustering Algorithms: A Comparative Approach”. In: *PLOS ONE* 14 (Dec. 2016). DOI: [10.1371/journal.pone.0210236](https://doi.org/10.1371/journal.pone.0210236).
- [17] Kristina P. Sinaga and Miin-Shen Yang. “Unsupervised K-Means Clustering Algorithm”. In: *IEEE Access* 8 (2020), pp. 80716–80727. DOI: [10.1109/ACCESS.2020.2988796](https://doi.org/10.1109/ACCESS.2020.2988796).
- [18] Seung Yoon Ko, Sung Won Cho, and Chulung Lee. “Pricing and Collaboration in Last Mile Delivery Services”. In: *Sustainability* 10.12 (2018). ISSN: 2071-1050. DOI: [10 . 3390 / su10124560](https://doi.org/10.3390/su10124560). URL: <https://www.mdpi.com/2071-1050/10/12/4560>.
- [19] *Pordate - Estatísticas sobre Portugal e a Europa*. <https://www.pordata.pt/Municipios/Densidade+populacional-452>. Accessed: 2022-07-04.

Appendix A

Data collection and project planning

A.1 Current process and its weaknesses

The steps of the current process are presented in detail (in the green blocks), and the weaknesses found at each process point are shown in the yellow blocks. It is important to note that from Figure A.1 to Figure A.5 is represented the logical path currently applied in the postal company

Prioritization of studies: Figure A.1 presents the phases of the prioritization process, as well as the points to improve.

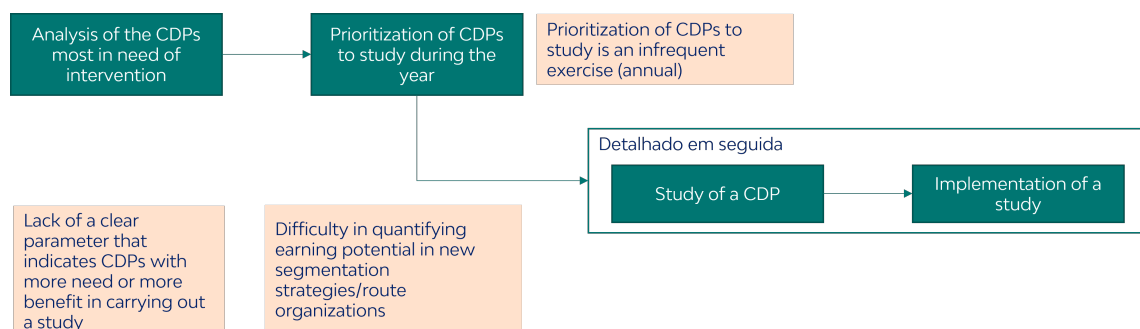


Figure A.1: Scheme for prioritizing studies in the current process.

Data collection: Figure A.2 presents the data collection process, and it was in this process that more points for improvement were found.

Otimization: The iterative optimization process is shown in Figures A.3 and A.4.

Study report: Finally, Figure A.5 presents the points presented in the study reports, as well as the points that should be improved.

A.2 Micro task times

A total of 154 different micro tasks time were collected. However, only a few examples are presented in Table A.1.

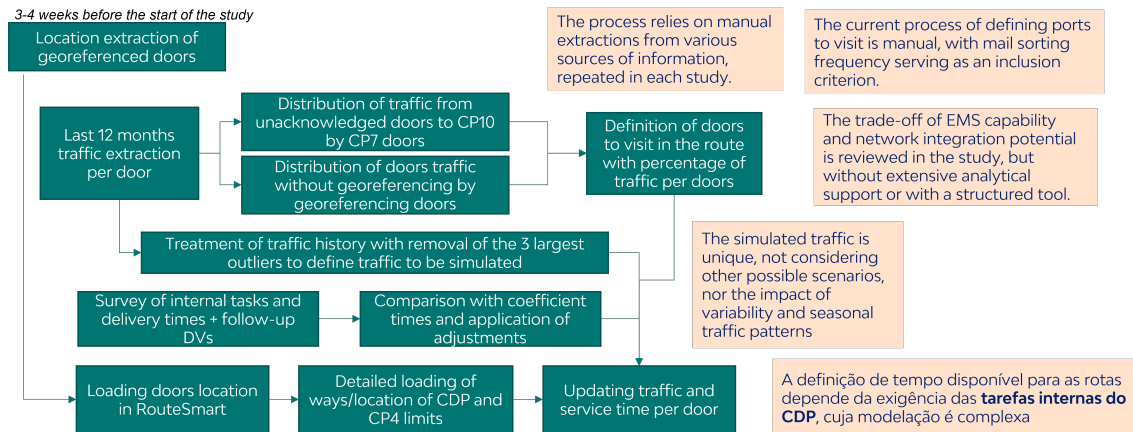


Figure A.2: Scheme of data collection in the current process.

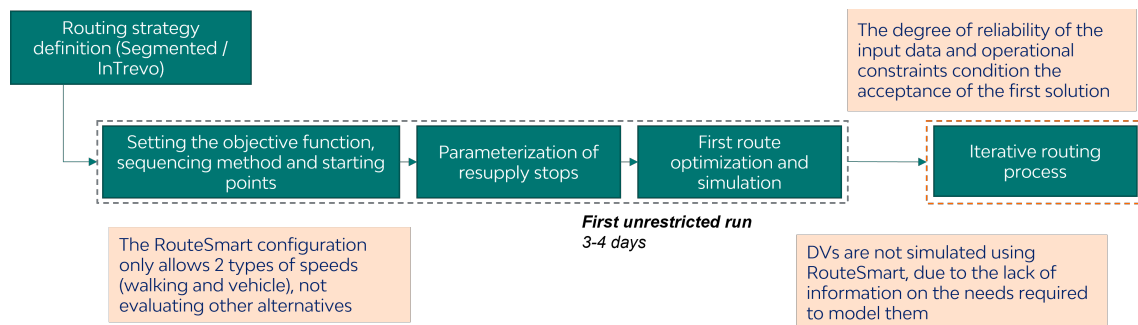


Figure A.3: Scheme of the start of the optimization in the current process.

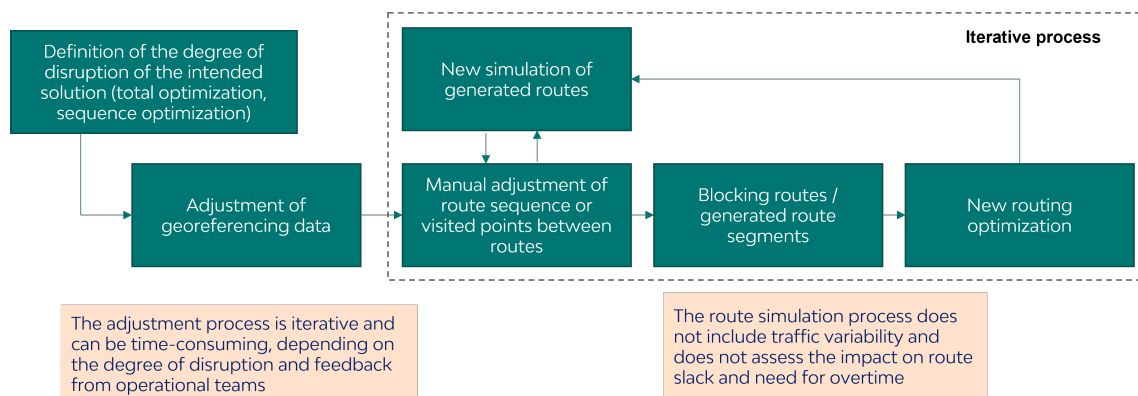


Figure A.4: Scheme of the end of the optimization in the current process.

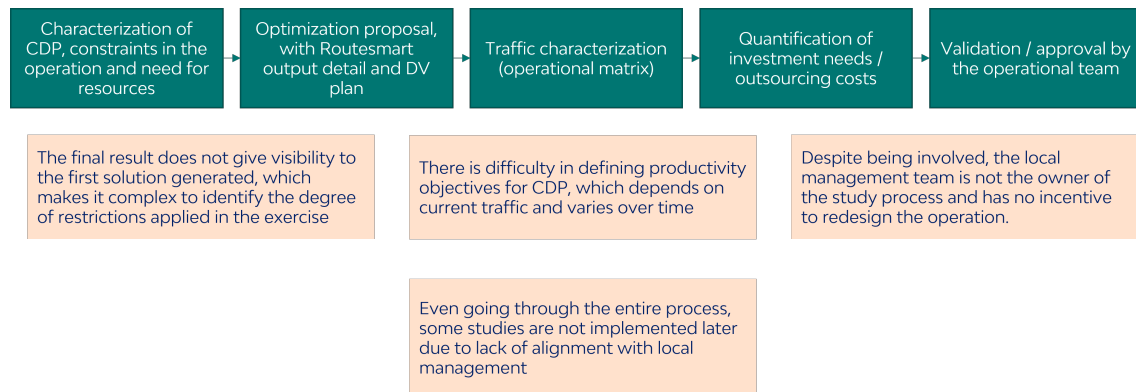


Figure A.5: Scheme of the information contained in the final reports of the current process.

Table A.1: Examples of micro tasks and respective execution times.

Macro task	Micro task	Time (min)
Discharge	Unloading vehicle r2 [container]	0.691
	Unloading vehicle r2 [pallet]	2.287
Deconsolidation	Fix empty uat after deconsolidation [container]	0.462
	Deconsolidate uat box and prepare sg and sv [box]	0.151
General Separation	SP screening without trace	0.048
	SP screening with trace	0.091
Track and Trace (T&T)	Collection and conference records [LLE]	0.045
	Record pistol [SP]	0.119
Vertical Sequencing	Collection of objects [box display] - SP	0.024
	Mail sequencing for Postal Boxes	0.039
Delivery	PE_Funcional_RPD_Panel - MP/P/LNP With T&T	2.255
	PE_Residencial_RPD_Unico - LLE Without T&T	0.056
Accountability	Accountability of objects warned with T&T - LLE	0.539
	Conference of objects reported with T& [checker]	0.284

Appendix B

Interface

B.1 VBA programming

The programming of the parameterization and configuration interfaces is carried out in VBA (Visual Basic for Applications). Its main objective is the connection between it and the database used to carry out the work. This connection is carried out through four main buttons that allow the upload of information from the database tables to Excel sheets, save, add and delete information by creating temporary tables and stored procedures. Figures B.1, B.2, B.3, and B.4 show, respectively, an example code of how the programming of these buttons works.

```
Public Sub Load()  
  
    Dim conn As New ADODB.Connection  
    conn.Open connection_string()  
    conn.CommandTimeout = timeout  
  
    Load_data conn, "Sheet_name", "selo.vw_table_name", "B", 4, "M", "Sheet_name", showMsg  
  
End Sub
```

Figure B.1: "Update" button programming code.

– Botão SAVE

```
Public Sub SAVE()  
  
    Dim conn As New ADODB.Connection  
    conn.Open connection_string()  
    conn.CommandTimeout = timeout  
  
    CreateTemporaryTable "staging_table_name", conn  
  
    ImportTableToDB conn, "Sheet Name", "staging_table_name", "B", 4, 4, "selo.sp_table_name", , , "B"  
  
    MsgBox ("As alterações foram gravadas com sucesso")  
  
End Sub
```

Figure B.2: "Save" button programming code.

```

Private Sub Add()

    Dim conn As New ADODB.Connection
    conn.Open connection_string()
    conn.CommandTimeout = timeout

    Call CorrerQuery("DELETE FROM bd.staging_table_name", conn)
    Call CorrerQuery("INSERT INTO bd.staging_table_name (column_name) VALUES ( ' " & variable & " ');", conn)

    Call CorrerQuery("exec selo.sp_table_name", conn)

    MsgBox ("A adição terminou.")
    Unload Me

End Sub

```

Figure B.3: "Add" button programming code.

```

Private Sub Delete()

    Dim conn As New ADODB.Connection
    conn.Open connection_string()
    conn.CommandTimeout = timeout

    Dim mbResult As Integer

    mbResult = MsgBox("Esta ação vai apagar as células selecionadas. Continuar?", vbYesNoCancel)

    If mbResult = vbYes Then

        Call CorrerQuery("DELETE FROM selo.staging_table_name", conn)

        For I = 0 To Lb_Cell.ListCount - 1
            If Lb_Cell.Selected(I) Then
                Call CorrerQuery("INSERT INTO selo.staging_table_name (column_name) VALUES ( ' " & variable & " ');", conn)
            End If
        Next I

        Call CorrerQuery("exec selo.sp_table_name", conn)

        MsgBox ("Células eliminadas.")
        Unload Me

    End If

End Sub

```

Figure B.4: "Delete" button programming code.

Appendix C

C# model

The present appendix demonstrates parts of the code implemented in the construction of the model. Thus, three sections are presented referring to the creation of clusters, optimization, and simulation of routes.

C.1 Clustering

The code presented in Figure C.1 shows three functions that must be executed to obtain the clusters (INITIALIZE_Clusters, RUN_Clusters, and UPLOAD_Clusters).

```
case "8":
    Console.WriteLine("Clustering");
    Simulation clusters = new Simulation(conn);

    clusters.INITIALIZE_Clusters(conn);

    int[] num = { 1, 2, 3 };
    var numClustersValues = new List<int>(num);

    foreach (var numClusters in numClustersValues)
    {
        clusters.RUN_Clusters(conn, numClusters);
        clusters.UPLOAD_Clusters(conn, numClusters);
    }
    break;
```

Figure C.1: Generating clusters in the C# model.

The initialization function is briefly shown in Figure C.2. This function intends to prepare the data for the clusters race. Thus, the model is connected to the database and collects the variables necessary for creating the model and the date range that limits the traffic of the clusters. In addition,

it is required to use functions such as the data normalization function to handle the input values correctly.

```
public void INITIALIZE_Clusters(SqlConnection conn)
{
    using (SqlCommand cmd = new SqlCommand("EXEC selo.sp_clusters_variables", conn))...

    using (SqlCommand cmd = new SqlCommand(" SELECT distinct node_id FROM selo.clusters_variables ", conn))...

    using (SqlCommand cmd = new SqlCommand("WITH date_restriction as ( " +
        "select cast(year(start_date) as varchar) as start_year, cast(datepart(week, start_date) as varchar) as start_week," +
        "cast(year(end_date) as varchar) as end_year, cast(datepart(week, end_date) as varchar) as end_week " +
        "from cluster_date ) " +
        ",concat_dates as ( " +
        "select concat(start_year, '-', right(concat('0', start_week), 2)) as initial_date," +
        "concat(end_year, '-', right(concat('0', end_week), 2)) as finish_date " +
        "from date_restriction ) " +
        "SELECT node_id, year, week, UECs, perc_EMS " +
        "FROM selo.clusters_variables v " +
        "Where concat(cast(year as varchar), '-', right(concat('0', cast(week as varchar)), 2)) between " +
        "(select initial_date from concat_dates) and(select finish_date from concat_dates)", conn))...
}
```

Figure C.2: Initialize clusters function.

On the other hand, the Run_Cluster function, shown in Figure C.3, aims to prepare and run the data in the K-means algorithm.

```
public void RUN_Clusters(SqlConnection conn, int numClusters)
{
    Console.WriteLine("Setting numClusters to " + numClusters.ToString());

    foreach (var nodeID in nodes)
    {
        if (nodeID.Value.traffic.yearWeek.Count() > 0)
        {
            //Prepare data for Kmeans
            double[][] rawData = new double[nodeID.Value.traffic.yearWeek.Count()][];

            for (int i = 0; i < nodeID.Value.traffic.yearWeek.Count(); i++)
            {
                string yearWeek = nodeID.Value.traffic.yearWeek[i];
                rawData[i] = new double[] { nodeID.Value.traffic.UECs[yearWeek], nodeID.Value.traffic.percEMS[yearWeek] };
            }

            var cluster = Cluster(rawData, numClusters);

            for (int i = 0; i < nodeID.Value.traffic.yearWeek.Count(); i++)
            {
                string yearWeek = nodeID.Value.traffic.yearWeek[i];
                if (!nodes[nodeID.Key].traffic.cluster.ContainsKey(yearWeek))
                {
                    nodes[nodeID.Key].traffic.cluster.Add(yearWeek, 0);
                }
                nodes[nodeID.Key].traffic.cluster[yearWeek] = cluster[i];
            }
        }
    }
}
```

Figure C.3: Run clusters function.

Finally, uploading clusters (Figure C.4) aims to save the results in database tables and, if applicable, replace the previously existing data.

```

public void UPLOAD_Clusters(SqlConnection conn, int numClusters)
{
    Program.Logger(timer, "Saving clustering results");

    var tableName = "selo.clusters_results";

    DataTable ClustersResults = new DataTable(tableName);

    ClustersResults.Columns.Add(new DataColumn("node_id", typeof(int)));
    ClustersResults.Columns.Add(new DataColumn("ano", typeof(int)));
    ClustersResults.Columns.Add(new DataColumn("semana", typeof(int)));
    ClustersResults.Columns.Add(new DataColumn("cluster", typeof(int)));
    ClustersResults.Columns.Add(new DataColumn("num_clusters", typeof(int)));

    foreach (var nodeID in nodes) ...

    Program.Logger(timer, $"Deleting previous data in {tableName}");

    // BUILD-ERRORS-NP

    using (SqlCommand cmd = new SqlCommand(string.Format(
        "DELETE FROM {0} WHERE num_clusters = {1}", tableName, numClusters), conn)) ...

    using (SqlBulkCopy s = new SqlBulkCopy(conn))
    {
        s.DestinationTableName = ClustersResults.TableName;
        s.WriteToServer(ClustersResults);
    }

    using (SqlCommand cmd = new SqlCommand(string.Format(
        "exec sp_clusters_run"), conn))
    {
        cmd.ExecuteNonQuery();
    }
}

```

Figure C.4: Upload clusters function.

C.2 Optimization

Figure C.5 presents the route optimization function and all the subfunctions used for this generation.

C.3 Simulation

Figure C.6 presents a small part of the simulation function created in the model, as well as the functions that make it up. Thus, it is possible to perceive that the information needed for this race is quite wide and diverse.

```

internal void RUN_RoutePlanOptimization(SqlConnection conn)
{
    Random rnd = new Random();

    Program.Logger(timer, string.Format("***** STARTING TO GENERATE OPTIMIZATION TRAFFIC SCENARIOS *****"));

    PREPARE_OptimizationScenarioTraffic(rnd);

    int rsCounter = 0;
    foreach (var ss in simulationStrategies.Values)
    {
        private int FindNearestCP7(int lastCP7, List<int> remainingCPs)
        {
            3 references | Diogo Correia, 27 days ago | 1 author, 1 change
        }

        private void CALCULATE_DailyTraffic(string dayTrafficType, DailyTraffic dt, DayType dayType, Random rnd)
        {
            2 references | Diogo Correia, 2 days ago | 3 authors, 12 changes
        }

        private void PREPARE_OptimizationScenarioTraffic(Random rnd)
        {
            1 reference | Diogo Correia, 12 days ago | 3 authors, 23 changes
        }

        private int evaluateEmsIntegration(List<EmsIntegration> emsIntegration, ObjectDelivery obj)
        {
            2 references | Diogo Correia, 12 days ago | 1 author, 2 changes
        }

        private void LOAD_OptimizationStrategyInputs(SqlConnection conn)
        {
            1 reference | Diogo Correia, 5 days ago | 3 authors, 11 changes
        }

        private void LOAD_Evaluation_OptimizationStrategyInputs(SqlConnection conn)
        {
            1 reference | Diogo Correia, 5 days ago | 2 authors, 2 changes
        }

        public void LOAD_Performance_OptimizationStrategyInputs(SqlConnection conn)
        {
            1 reference | Diogo Correia, 5 days ago | 2 authors, 2 changes
        }
    }
}

```

Figure C.5: Optimization function.

```
public void INITIALIZE_PrepareOptimizedNetworkSimulation(SqlConnection conn)...  
  
1 reference | Diogo Correia, 24 days ago | 2 authors, 7 changes  
private void PREPARE_RouteProductSegmentDayIndexes(SqlConnection conn)...  
  
1 reference | Diogo Correia, 24 days ago | 3 authors, 11 changes  
private void PREPARE_RouteMaps(SqlConnection conn)...  
  
3 references | Diogo Correia, 2 days ago | 1 author, 1 change  
private void LOAD_CdpInfluence(SqlConnection conn, string loadType)...  
  
1 reference | Diogo Correia, 25 days ago | 4 authors, 13 changes  
public void LOAD_NetworkNodes(SqlConnection conn)...  
  
1 reference | Diogo Correia, 12 days ago | 3 authors, 6 changes  
public void LOAD_RouteCosts(SqlConnection conn)...  
  
1 reference | carolinacouto, 2 days ago | 2 authors, 3 changes  
private void LOAD_EvaluationCapacity(SqlConnection conn)...  
  
1 reference | Diogo Correia, 12 days ago | 2 authors, 2 changes  
private void LOAD_PerformanceCapacity(SqlConnection conn)...  
  
1 reference | Diogo Correia, 12 days ago | 2 authors, 7 changes  
private void PREPARE_RouteSegmentMap(SqlConnection conn)...  
1 reference | Diogo Correia, 37 days ago | 3 authors, 13 changes  
private void LOAD_InternalTimes(SqlConnection conn)...  
1 reference | Diogo Correia, 12 days ago | 5 authors, 18 changes  
private void LOAD_ProductTypes(SqlConnection conn)...  
  
1 reference | Diogo Correia, 34 days ago | 1 author, 2 changes  
private void LOAD_ExternalTimes(SqlConnection conn)...  
  
1 reference | Pedro Campelo, 55 days ago | 2 authors, 5 changes  
private void PREPARE_SegmentsForCP7Selection(SqlConnection conn)...
```

Figure C.6: Simulation function.