Universidade do Minho
School of Engineering

Hugo Daniel Silva Carvalho

An Intelligent Decision Support System for the Freight Transport Sector

October 2022

An Intelligent Decision Support System for the Freight Transport Sector
Hugo Silva Carvalho
UMinho | 2022

Universidade do Minho
School of Engineering

Hugo Daniel Silva Carvalho

An Intelligent Decision Support System for the Freight Transport Sector

Master Thesis
Master in Engineering and Management of Information Systems

Work developed under the supervision of:
Prof. Dr. Paulo Alexandre Ribeiro Cortez

October, 2022

Acknowledgements

This is far more than a simple document or even a dissertation. This is the climax of a lot of work, effort, crossroads, joy and sorrow, loyalty and betrayal, but, especially, of many people. To all of you, I dedicate this work and express my most tremendous and purest gratitude for what you have moulded me into and everything you have made possible for me to achieve.

To my parents, who, despite all the adversities that life holds, always supported me, always provided me everything to live and grow according to my potential, to get to this very moment, to the person I have become.

To Maria, my everyday safe haven, my better half, my dearest long-time best friend and whom life has decided to reunite, my greatest praise for all the affection, understanding, companionship, and inspiration.

To my long-time friends, David and Nuno, for all the encouragement, for representing my safeness, for helping me to escape from the responsibility and effort that this path required, for helping me to celebrate the small victories but, mainly, for always being there to help me rise up in all the defeats.

To all the people this journey has presented me with, fellow classmates or professors. Yet, my greatest gratefulness goes to the special twelve that I have surrounded myself with most closely and have walked alongside me throughout this journey. For pushing this experience to the best, the most exciting book, the best tale life could have told me. I will be eternally grateful to you for all the growth, comradeship, devotion, support, and laughs, but also for all the drama, misfortunes, and sorrows, because much of what I am today is thanks to you.

I also would like to express a particular recognition to my supervisor Prof. Dr. Paulo Cortez for all the experience and availability that he offered me throughout the development of this work, as well as the trust and opportunities provided.

Finally, I also owe some special words to all my research colleagues in Machine Learning, with whom I have worked during the last two years, for all the experience and knowledge they have shared with me, for the work environment, and for their contribution to my professional growth.

To all the people I have and have not mentioned, there are no words to thank you for everything you have made me experience and how much you have made me grow at all levels, especially as a person.

To this *Acknowledgements* section for making me realize how lucky I am for all of you.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the Universidade do Minho.

________________________, ____________________________________

(Place) (Date)

____

(Hugo Daniel Silva Carvalho)

"There is only one good, knowledge, and one evil, ignorance" (Socrates)

Abstract

An Intelligent Decision Support System for the Freight Transport Sector

The Freight Transport Sector is under immense pressure. With globalisation, world population growth, migration of people to urban areas, exponential technological development, and growing environmental concerns, efficient freight distribution can be challenging. An inefficient movement of goods may lead to economically nonviable businesses and adverse social and climate effects. Even though technological development may represent one of the stated variables causing this pressure on the sector, it may also stand as one of the solutions in mitigating these effects and a mobilising agent for its progressive evolution. An adequately defined strategy supported by an intelligent decision support system can hold a value-added impact on the Freight Transport Sector, making it less expensive, safer, cleaner, sustainable, and more efficient.

This dissertation work employs software development, Machine Learning (ML) techniques, and optimisation approaches on several fronts of the Freight Transport Sector. This includes resource optimisation, driver suggestion and management according to legal regulations, transport loads optimisation, and, finally, route sequencing and optimisation. Due to the globalisation effect on the sector and its logical impact on increased international and intercontinental transport, this thesis also addresses, studies and provides a solution for multimodal transport optimisation.

This work is part of the Research and Development (R&D) project "*aDyTrans – Dynamic Transportations Platform*", resulting from the collaboration between a technological interface centre, the Centro de Computação Gráfica (CCG), and *abmn.pt - Business Solutions*, a technology consultancy company operating in the Freight Transport Sector.

Keywords: Freight Transport Sector, Geographic Information System, Machine Learning, Geospatial Clustering, Transport Optimisation, Resource Optimisation, Load Optimisation, Multimodal Transport Optimisation

Resumo

Um Sistema Inteligente de Suporte à Decisão para o Setor dos Transportes de Mercadorias

O setor dos transportes de mercadorias está sob uma pressão imensa. Perante a globalização, o crescimento da população mundial, a migração de pessoas para áreas urbanas, o exponencial desenvolvimento tecnológico sentido atualmente e as crescentes preocupações ambientais, a distribuição eficiente do transporte de mercadorias pode ser um desafio. Uma circulação ineficiente de mercadorias pode levar a negócios economicamente inviáveis e a efeitos sociais e climáticos adversos. Embora o desenvolvimento tecnológico possa constituir uma das variáveis anteriormente mencionadas, e causadoras desta pressão sob o setor, pode também representar uma das soluções para mitigar estes efeitos e, simultaneamente, um agente mobilizador para a sua evolução progressiva. Desta forma, uma estratégia devidamente concebida e apoiada por um sistema inteligente de apoio à decisão pode exercer um impacto de valor acrescentado no setor do transporte de mercadorias, tornando-o menos dispendioso, mais seguro, mais limpo, mais sustentável e mais eficiente.

Esta tese de mestrado emprega desenvolvimento de software, técnicas de ML e abordagens de otimização nas várias frentes do setor do transporte de mercadorias, incluindo otimização de recursos, sugestão e gestão de motoristas de acordo com regulamentos legais, planeamento, alocação e otimização de cargas de transporte, e, finalmente, sequenciação e otimização de rotas. Devido ao efeito da globalização no setor e ao seu consequente impacto lógico no aumento do transporte internacional e intercontinental, esta tese também aborda, estuda e fornece uma solução para a otimização do transporte multimodal.

Este trabalho faz parte de um projecto de Investigação e Desenvolvimento (I&D) "*aDyTrans – Dynamic Transportations Platform*", resultante da colaboração entre um centro de interface tecnológica, o CCG, e a *abmn.pt - Business Solutions*, uma empresa de consultoria tecnológica que opera no setor do transporte de mercadorias.

Palavras-chave: Setor dos Transportes de Mercadorias, Sistema de Informação Geográfico, Machine Learning, Agrupamento Geográfico, Otimização do Transporte, Otimização de Recursos, Otimização da Carga, Otimização do Transporte Multimodal

Contents

List of Figures

List of Tables

List of Algorithms

List of Listings

Acronyms

ACO	Ant Colony Optimisation
AI	Artificial Intelligence
API	Application Programming Interface
BD	Big Data
CBC	COIN-OR Branch and Cut
CCG	Centro de Computação Gráfica
CICD	Continuous Integration and Continuous Delivery
CP	Constraint Programming
CRISP-DM	CRoss Industry Standard Process for Data Mining
CSV	Comma Separated Values
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DcGC	Distance-constrained Geospatial Clustering
DM	Data Mining
DS	Design Science
DSRM	Design Science Research Methodology
DSS	Decision Support System
E2ES	End-to-End Solution
EC	European Commission
EDA	Exploratory Data Analysis
EU	European Union
G-CARP	Green Capacitated Arc Routing Problem

GA	Genetic Algorithm
GDPR	General Data Protection Regulation
GeoJSON	Geographic JavaScript Object Notation
GHG	Greenhouse Gas
GIS	Geographic Information System
HGA	Hybrid Genetic Algorithm
HOS	Hours-of-Service
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
I&D	Investigação e Desenvolvimento
IDSS	Intelligent Decision Support System
IS	Information Systems
IT	Information Technology
JPG	Joint Photographic Experts Group
JSON	JavaScript Object Notation
JWT	JSON Web Token
KDD	Knowledge Discovery in Databases
KML	Keyhole Markup Language
LM	Levenberg-Marquardt
MDS	Multidimensional Scaling
MILP	Mixed Integer Linear Programming
ML	Machine Learning
MPEC	Mathematical Programming with Equilibrium Constraints
MTO	Multimodal Transport Operator
NN	Neural Network
NP	Nondeterministic Polynomial Time
OR	Operations Research
ORM	Object Relational Mapping
OSM	OpenStreetMap

OSRM	Open Source Routing Machine
PCA	Principal Component Analysis
PNG	Portable Network Graphics
PyPI	Python Package Index
R&D	Research and Development
REST	Representational State Transfer
RL	Reinforcement Learning
SCIP	Solving Constraint Integer Programs
SQL	Structured Query Language
SVG	Scalable Vector Graphics
TAM	Technology Acceptance Model
TCP	Transit Consolidation Problem
TSP	Travelling Salesman Problem
TXT	Text File
URL	Uniform Resource Locator
US	United States
VRP	Vehicle Routing Problem
WSGI	Web Server Gateway Interface
XLSX	Microsoft Excel Spreadsheet

Chemical Symbols

CO_2 Carbon Dioxide

1

Introduction

1.1 Context and Motivation

The exponential development of areas such as Artificial Intelligence (AI), ML and Big Data (BD) is having a profound impact on the way we interact with the world around us. Representing a robust set of technologies that can help humans solve everyday problems, their applications are transforming the various activity sectors. In particular, the transport sector benefits the most from this technological revolution, making it safer, cleaner, brighter, and more efficient.

This dissertation is part of the R&D project "*aDyTrans – Dynamic Transportations Platform*" resulting from the partnership between CCG and *abmn.pt - Business Solutions*, striving to provide a privileged space for research and development activities within advanced transport planning and optimisation. The project aims to predict and define the allocation of resources for the transport of goods based on the pursuit of the objectives set by the European Commission (EC) in the White Paper "Roadmap to a Single European Transport Area - Towards a competitive and resource efficient transport system", for the time horizon 2010-2050. Sustainability strategies in the transport sector, dynamic operations planning, vehicle routing algorithms and methods for selecting and combining transport modes are the backdrops of this project.

The company has been in the market since 2001. Its mission is to develop solutions and provide qualified and professional services to support organisations in controlling management activities and processing business information. It intends to affirm itself as a true partner in the success of its clients, recognizing them as its main asset. In the context of this project, the main purpose intends to evolve its product *"aTrans - Logistic Software"* with the empowerment of an intelligent system to support the decision-making process. The current solution lies within the Freight Transport Sector. It is a solution that allows not only the control of all costs inherent to the business - tolls, taxes, mobile communications, or others - but also effective route planning, reducing to low percentages the empty kilometre. It contemplates all vehicle maintenance and the management of auxiliary equipment.

1.2 Objectives and Expected Results

Given the wide range of variables that can affect the Freight Transport Sector, the challenges associated with the development of this project focused primarily on the type of operations and resources used by freight transport companies that are clients of *abmn.pt - Business Solutions*. As the main problems are centred on freight orders and the respective service value to be estimated, we considered the following essential related characteristics:

- Transport services origin and destination based on GPS position.
- Vehicle fleet including all expenses history and associated costs.
- Seasonality of transport services and opportunities of grouping loads by zones.
- The entire driver circumstances, primarily logistical and legal.
- Network of logistic infrastructure partners (*i.e.*, Logistic Hubs, Ports, Airports, and Railway Stations).

The mentioned characteristics intend to answer the following set of pretensions of this project:

1. Support decision-making in the allocation of human resources to the variety of transport services to be provided based on the European Union (EU)'s set of regulations for the maximum daily and fortnightly driving times as well as the required daily and weekly rest periods for all drivers of road haulage and passenger vehicles as outlined in the EC Regulation No. 561/2006. These guidelines essentially state that:
   - The daily driving period cannot exceed 9 hours, except twice a week when it can be extended to 10 hours.
   - Total weekly driving time is limited to 56 hours, and total fortnightly driving time is limited to 90 hours.
   - The daily rest period must be at least 11 hours, with a maximum of 9 hours allowed thrice per week. Daily rest can be divided into 3 hours of rest followed by 9 hours of rest for a total of 12 hours.
   - After 4.5 hours, breaks of at least 45 minutes (separable into 15 minutes and 30 minutes) should be taken.
   - Weekly rest is 45 continuous hours, which can be reduced to 24 hours every other week. Compensation arrangements are in place for a shorter weekly rest period. Except for coach drivers engaged in a single occasional service of international passenger transport, who may postpone their weekly rest period after 12 days to accommodate coach holidays, weekly rest is to be taken after 6 days of working.

- In exceptional circumstances, daily and weekly driving times may be extended by up to 1 hour to allow the driver to reach his/her residence or the employer's operational centre to take a weekly rest period. Exceeding the daily and weekly driving times by up to 2 hours is also permitted to allow the driver to arrive at his/her residence or the employer's operational centre to take a regular weekly rest period.

Driver regulations are a fundamental component of transport management. Its administration is also of high complexity, given the number of parameters to be considered. With this component, business agents will be able to obtain greater efficiency and profitability in the use of their resources.

2. Plan a future transport service based on preliminary plannings. Predictability is beneficial in any organisational process as it allows easier management of resources and services. This tool will improve the quality and efficiency of traffic operators' work, which is often stressful because they must fulfil contractual delivery and collection requirements.

3. Allow for multimodal distribution in long-distance services planning. In order to move goods from one geographical location to another, a service frequently requires the use of more than one transport modality. For example, in an intercontinental transport context, ship or plane transport will be required in addition to pick up and delivery by road transport. On the other hand, a service carried by rail transport can represent a more practical solution, either financially or operationally. The system will assist the operator by recommending the most suitable transport modalities.

4. Provide the various options for delivering a service and the associated costs. Optimising resources is becoming increasingly important in organisations, which is no different in the transport sector. This tool will assist business agents in achieving better resource optimisation and, as a result, increased business profitability. The system will inform the operator of the various service options.

5. Facilitate the planning of services using various resources (drivers and vehicles). The resources employed for the transport service are defined in terms of the characteristics of the transport. To accomplish this, it is frequently necessary to mix various resources. For instance, the driver picks up a different tractor or exchanges trailers. These are only a few activities that may be performed on this platform to ensure that the best resources are used to deliver the service. On the other hand, the system will propose to the operator which resources are most appropriate for the service, considering history and patterns.

6. Identify groupage opportunities as new services are available. Volume, weight, and floor meters must be optimised for profitability. With the help of this tool, the operator will be notified of services with comparable features at the loading and unloading sites so that, if desired, he or she can associate them and monetise the resources.

7. Optimise deliveries according to their physical attributes and the legal restrictions on the vehicles required to transport them. This optimisation makes it feasible to sequence, assign, and route a given set of transport services and related human resources efficiently. In addition to being profitable, this optimisation allows a decrease in carbon footprint. Even if a few extra kilometres on a given route might seem insignificant, when this optimisation is applied consistently on a large fleet of vehicles, the cumulative reduction in kilometres and time can significantly save costs and promote a sustainable growth and evolution of the Freight Transport Sector.

1.3 Research Methodology

Considering the nature of this project as a R&D resulting from the partnership between CCG and *abmn.pt - Business Solutions*, which expects the development of an artefact - a freight transport IDSS - to address a specific set of problems, we adopted Design Science (DS) as the research paradigm. A Design Science Research Methodology (DSRM) for Information Systems (IS) research was chosen for the project, which incorporates principles, practices, and procedures required to carry out such research. The presented DS process, depicted in Fig. 1, includes six steps: problem identification and motivation, the definition of the objectives for a solution, design and development, demonstration, evaluation, and communication.

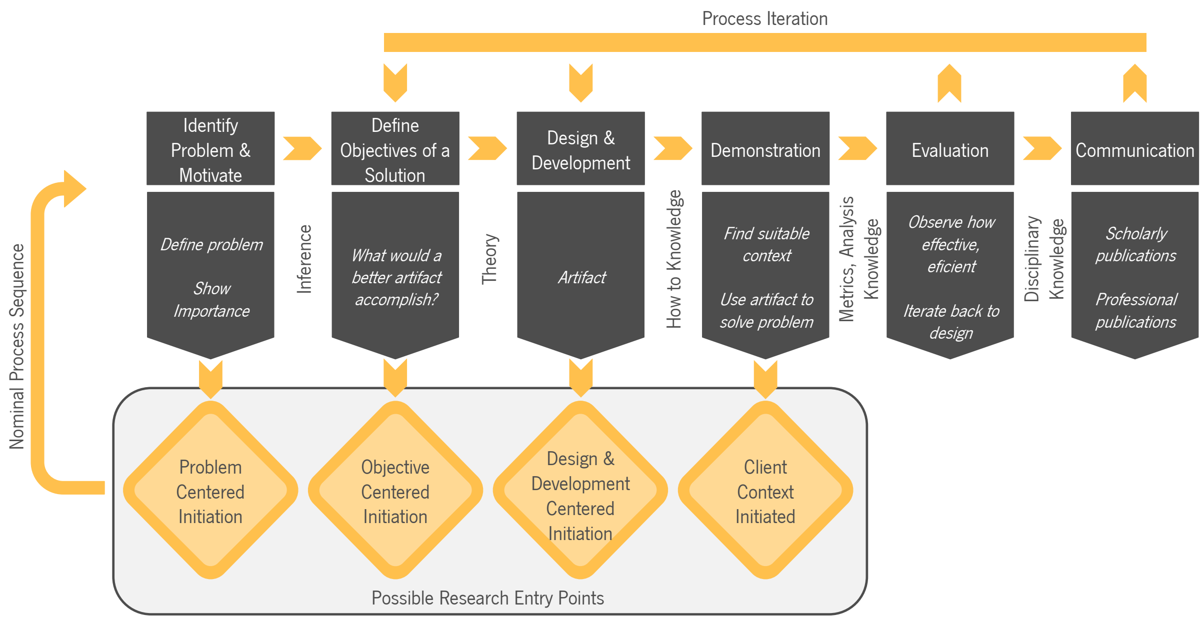


Figure 1: DSRM Process Model. Adapted from Peffers et al. (2007) [84].

The first step of the methodology consists of defining the specific research issue and establishing the significance of a solution. It is helpful to conceptually break down the problem so that the solution can adequately reflect its complexity since the problem description is crucial to developing an artefact that can effectively provide a solution.

In the next step, using the problem definition and knowledge of what is feasible and possible, the objective lies within the inference of the goals of a solution. The objectives can be qualitative, such as describing how a new artefact is intended to enable solutions to problems not previously addressed, or quantitative, such as terms in which a desirable solution would be better than current ones. This set of goals should be logically deduced from the problem description.

Regarding the *Design and Development*, the third step represents the actual artefact's development. Such artefacts could be instantiations, models, constructions, procedures, new attributes, or instantiations of technological, social, and informational resources. A design research artefact is any design incorporating a research input into the design.

The next step, the *Demonstration*, shows how the artefact can be used to resolve one or more instances of the issue. This could entail applying it to an experiment, simulation, case study, piece of evidence, or other applicable tasks. One of the resources needed for this activity is an adequate understanding of how to use the artefact to solve the problem.

In the *Evaluation* activity, we assess how effectively the artefact contributes to a solution that addresses the problem. Comparing a solution's objectives to the results that were obtained when the artefact was used in the Demonstration is the focus of this activity. It requires expertise with pertinent measurements and analysis methods. Evaluation can take many formats depending on the artefact and the problem venue.

The last step of the methodology consists of informing researchers and other pertinent audiences, such as working professionals, about the problem and its significance, the artefact, its uniqueness and utility, the rigour of its design, and its effectiveness. Similar to how the nominal structure of an empirical research process is typical for empirical research papers, researchers may use the format of this process to structure scholarly research publications. An understanding of the discipline culture is necessary for communication.

1.4 Thesis Organisation

This master's thesis is structured in four main parts, including seven chapters. Its contents and integrity lead to achieving and revealing the objectives and expected results mentioned above. The first part comprises Chapter 1 and Chapter 2, providing a brief overview of the issues addressed in this thesis and relevant current state of the art, respectively. In particular, Chapter 2 is divided into two main ranges. The first covers a theoretical introduction of a set of meaningful addressed concepts, namely, *"Big Data"*, *"Data Mining"*, *"Artificial Intelligence"*, *"Machine Learning"*, *"Decision Support Systems"* and *"Geographic Information Systems"*. On the other hand, the second range of this chapter follows a different approach. In addition to presenting the sector under study, it provides a critical practical state-of-the-art analysis of historical and modern studies related to the application of computational support in decision-making in the Freight Transport Sector. Thus, it consists of the *"Freight Transport Sector"*, *"Vehicle Routing Problem"*, and *"Decision Support Systems in Freight Transport Sector"*.

The second part of this document comprises Chapter 3, Chapter 4, Chapter 5 and Chapter 6, consisting of the main experimental work developed during this thesis. Chapter 3 presents an IDSS to optimise transport and logistics activities in the Freight Transport Sector. This IDSS comprises three main modules that can be used individually or chained: a geographic clustering detection of transport services, a transport driver suggestion, and a route and truckload optimisation. It was entirely designed and developed to support real-time data. It consists of an End-to-End Solution (E2ES), which covers all the main transport and logistics activities from the registration in the database to the optimised transport plan. Chapter 4, based on the requested requirement from the company to isolate the IDSS presented in the previous chapter, offers an additional artefact regarding multimodal transport optimisation. The solution models the essential freight transport scenario using mathematical programming to find a solution that minimises the total costs of combinatorial routing plans, including multiple transport modalities such as air, sea, road and rail. Chapter 5 presents an additional development of this dissertation project, an open-source Python package that allows a quick generation of random coordinates within a set of geographic boundaries. The framework offers methods to generate, save and plot a determined number of random coordinates generated from a location polygon or a simple name/address through its geocoding. Such geographic methods are well-suited for research and development activities in transport and logistics, such as the Travelling Salesman Problem (TSP) and the Vehicle Routing Problem (VRP). Chapter 6 presents the knowledge and source code transfer environment of the work developed within the three previously mentioned chapters. The third part comprehends Chapter 7. The main conclusions of this dissertation, limitations and potential gaps for future research are addressed in this chapter. Finally, the last part (Appendix A) includes a user manual for the developed Python package regarding the work embedded in Chapter 5.

2

Background

2.1 Introduction

A literature review denotes the foundation for all types of research. In addition to knowledge transfer, they are a means of establishing proper guidelines for the timely response to questions posed by a specific problem. If done correctly, they can devise new ideas and directions for the research question [98]. As a result, it is critical to conduct a literature review with the utmost care and rigour to develop increasingly better and more precise hypotheses and research questions, thus improving the quality of the research conducted on a given topic.

2.2 Literature Review Strategy

The present chapter reviews the relevant state-of-the-art works concerning the usage of Software Development, ML, and optimisation approaches applied to the Freight Transport Sector. Thus, we explain the literature search strategy, followed by a theoretical introduction of a set of meaningful addressed concepts. Then, we break down the sector under study, followed by the analysis of practical case studies developed in the area, aiming to detect possible research gaps that this dissertation intends to prosecute and add value to it. Lastly, the end of this chapter culminates in a brief enumerated description of the tools and technologies we resort to in response to the proposed problem and objectives. As a result, this literature review was divided into two major stages: a theoretical introduction to the subject, in which main concepts are presented, and practical works, describing relevant case studies.

Regarding the theoretical introduction, most of the references used concern books found in online databases provided by services such as Google Scholar and Google Books. This literature search included the terms *"Big Data"*, *"Data Mining"*, *"Artificial Intelligence"*, *"Machine Learning"*, *"Decision Support Systems"*, *"Intelligent Decision Support Systems"*, *"Geographic Information Systems"*, *"Geographic Information Systems for Transportation"*, among others. The references found in this search were used in the first stage of the literature review, namely: Big Data (Section 2.3), Data Mining (Section 2.4), Artificial Intelligence (Section 2.5), Machine Learning (Section 2.6), Decision Support Systems (Section 2.7) and Geographic Information Systems (Section 2.8). Despite not being directly reflected in the mentioned content, some online courses were attended as a supplement for this first stage, aiming to establish and harden essential insights into a few relevant subjects such as BD and ML. Coursera and Udacity were the platforms used to take these courses.

For the second stage of this literature review, the goal was to locate real-world case studies applied to the activity sector under consideration. In addition to books from the mentioned online databases, the literature content from this stage included journal and conference articles, and master and doctoral dissertations. Google Scholar, Google Books, Scopus, IEEE Xplore, and Science Direct, were the databases used in this search. Among the terms used were *"Freight Transport"*, *"Freight Optimisation"*, *"Route Optimisation"*, *"Truck Consolidation"*, *"Transport Groupage"*, *"Transport Clustering"*, *"Vehicle Routing Problem (VRP)"*, *"Travelling Salesman Problem (TSP)"*, *"Freight Drivers Optimisation"*, *"Multimodal Transport Optimisation"*, and *"Intermodal Transport Optimisation"*. While conducting this literature review component, it is crucial to note that, whenever possible, filters based on the year of publication were applied to each search to prioritise the most recent studies, predominantly published in the last ten years.

Regarding papers, priority was established to the journal type, and the SCImago Journal Rank was an invaluable tool to select papers from the most prestigious journals. Finally, it should be highlighted that the process of selecting or discarding a paper was based on the analysis of the content provided in its abstract, results and conclusions sections. We found case studies on optimising the various fronts that comprise the Freight Transport Sector thanks to the knowledge transferred and resulting from this search. These insights were covered in Section 2.9, Section 2.10, and Section 2.11, the components of the second phase of the current chapter. Section 2.12 contains a brief list of the tools and technologies we used in response to the proposed problem and objectives. The chapter is then concluded in Section 2.13 based on the analysed case studies and the set of technologies described.

2.3 Big Data

BD refers to massive amounts of data. BD is a relatively new concept, but data storage and analysis have been a theory for many years. Before the advent of computers and databases, data was presented on paper as registers [10].

A large amount of data is produced daily due to the advent of computers, the internet, and mobile technology. Nowadays, every digital action leads to an increase of data [104]. Social media, internet purchases, network gadgets, and academic records represent examples of data generators. In 2017, Herschel and Miori (2017) [54] expected 50 zettabytes of data to be generated by 2020. Recent research indicates that this prediction exceeded the previous year and that by 2025, it will have increased more than four times. Due to the COVID-19 pandemic's increased demand, which resulted from more individuals working and studying from home and using home entertainment alternatives more frequently, the growth has been more significant than initially anticipated [50].

In the past, data storage was limited to databases and spreadsheets, leading to discarding the majority of data due to the challenging task of organising it into rows and columns. BD includes both structured (*e.g.*, databases), unstructured (*e.g.*, videos, audio, documents, photos, comments, followers, likes, tags, tweets, clicks, and chats), and semi-structured data (*e.g.*, currency conversion. weather, XML, graph or text data, and e-commerce) [38].

BD is further characterised as 3V's, Volume, Velocity, and Veracity [24]. Demchenko et al. (2013) [32] classified BD into 5V's, which are Volume, Velocity, Variety, Veracity, and Value. However, later in 2018, Saggi and Jain (2018) [92] characterised BD into 7V's, namely Volume, Velocity, Variety, Valence, Veracity, Variability, and Value.

2.4 Data Mining

Ongoing advancements in Information Technology (IT) have allowed the storage of large and diverse databases. Given the extensive amount of data available today, it is critical to transfer meaning to it and extract patterns, relationships, and anomalies that can lead to practical and valuable information and knowledge. Even though there is no broadly agreed and accepted definition, Data Mining (DM) is commonly defined as the process of extracting knowledge from a set of data [52, 51, 79, 45]. For this reason, DM is considered a common synonym for Knowledge Discovery in Databases (KDD) and, also, ML [45]. In this dissertation, DM and KDD will be regarded as synonyms for extracting meaningful and valid patterns from a given dataset that may help decision-making, and we will apply the conceptualisation of DM. ML, on the other hand, will be defined as a set of learning techniques employed in the process of accomplishing the objectives mentioned above, therefore making the decision-making process intelligent.

The term DM first appeared in the late 1980s and represents an analogy to gold mining, with knowledge representing gold and data representing rocks [51]. In 1990, the CRoss Industry Standard Process for Data Mining (CRISP-DM) arose as a standardised approach to DM [79]. As depicted in Fig. 2 it consists of the following steps: business understanding, data understanding, data preparation, modelling, evaluation, and deployment.

2.5 Artificial Intelligence

AI has a much longer history than is generally believed in areas ranging from science to philosophy. Its idealisation can even be found in Ancient Greece. In the epic poem Iliad by Homer, dating from the 8th century BC, air bellows and automatic doors are mentioned, which seem to be able to anticipate the intentions and wishes of their users and perform a set of tasks spontaneously and with a moderate degree of autonomy. In the same work, Hephaestus, the god of fire, is mentioned as having created artificial assistants to help him in his activities after being expelled from Olympus [22].

Figure 2: CRISP-DM methodology.

Its current notability is primarily due to Alan Turing. The latter, in 1950, proposed a variation of a simple parlour game as a means of identifying a machine that can think: A human judge interacts with two computer terminals, one controlled by a computer and the other by a person, but the judge does not know which is which, in order to test the ability of a machine to exhibit intelligent behaviour equivalent to or indistinguishable from a human [105].

However, the birth of AI dates back to 1956, at the famous Dartmouth Summer Research Project on Artificial Intelligence, where, despite resistance from scientists, the term was approved and described by John McCarthy as *"the science and engineering of making intelligent machines"* [63].

Today, AI is an exponentially growing field with many different disciplines, ranging from mathematics, economics, neuroscience, psychology, and computer engineering. With the variety of separate opinions on what AI stands for, the lack of agreement on a standard (*i.e.*, criteria, benchmark tests, goals, regulations) makes its healthy growth extremely challenging.

2.6 Machine Learning

Despite the lack of an universally agreed definition of ML, Samuel (1959) [94] used the term for the first time to characterise the field of research that allows computers to learn without being explicitly programmed. ML is similar to the scientific process but on a much larger scale. It goes through the same steps of creating, testing, and dismissing or refining hypotheses. A scientist may spend his whole life developing and testing hundreds of theories, while a ML system can accomplish the same in a fraction of a second. It is the domain of discovery automation and is transforming numerous activity sectors [35]. ML algorithms have as their raw material data to fulfil the stated goals. This data is categorised according to its type and learning purpose. The field is divided into three main types: Supervised, Unsupervised, and Reinforcement Learning (RL).

2.6.1 Supervised Learning

This learning model is designed to make predictions from an unforeseen input. Given a set of input data and its corresponding target values, it is possible to train a model (*Regression/Classification*) to predict the targeted value of a new set of data. Thus, supervised learning uses classification algorithms and regression techniques to develop predictive models. Classification techniques predict discrete values, and it is recommended that the data is categorised, labelled, or separated into specific groups. On the other hand, regression techniques predict continuous values. A linear regression, for example, aims to model the relationship between two variables by fitting the linear equation to the observed data [96].

2.6.2 Unsupervised Learning

In Unsupervised Learning, the goal is to find the regularities of the input data, understand how often a particular event happens and learn the pattern of what usually occurs and what does not [96]. This approach does not provide the ML algorithm with any tagged data. In this type of learning, data collection is often provided without any ground truth labels. The ML algorithm attempts to discover hidden/semantic structures in the data or create relationships between different features. Principal Component Analysis (PCA), K-Means, and Multidimensional Scaling (MDS) are examples of unsupervised learning methods [80].

2.6.3 Reinforcement Learning

RL tackles the issue of how an autonomous agent that detects and acts in its environment could make the best decisions to attain its objective while taking into account the effects of its actions. This approach presupposes that there is interaction with the environment. In this technique, a reward system is utilised to imply success. Therefore no processing is required because the algorithm knows the reward it expects to receive for each sort of action it does [80].

According to Ravichandiran (2018) [88], the following are the steps in a typical RL algorithm:

1. The agent interacts with its environment by performing an action.

2. The agent accomplishes an action and transitions from one state to another.

3. Finally, the agent will be rewarded for the activity it took.
4. The agent will determine whether the activity was good or poor based on the reward.
5. If the action was excellent, that is, if the agent earned a positive reward, the agent chooses to perform that action; otherwise, the agent will attempt another action that results in a positive reward. As a result, it is essentially a trial and error learning process.

2.7 Decision Support Systems

One way to get an advantage when facing new challenges in today's globalised and competitive world is by using digital data to support decision-making. A Decision Support System (DSS) is a computer application that supports the decision-making process through the usage of IT [86]. DSSs can improve decision quality, speed up decision-making, support better information management, and reduce costs and uncertainty by helping decision makers interpret data more effectively [49]. Additionally, they provide a comprehensive view of the decision-making process and facilitate communication between employees and managers [78]. Currently, DSSs are used in various fields, including to support the management of human resources, finance, logistics, marketing, and manufacturing. In the three decades of its existence, DSSs have evolved from a radical movement that altered how businesses regarded information systems to a widespread commercial IT movement in which all enterprises participate. DSSs have been an important area of study within IS.

2.7.1 Intelligent Decision Support Systems

An IDSS goes beyond traditional DSSs by integrating AI technologies such as ML, DM, knowledge management, and Metaheuristics [9, 29]. These technologies can help identify patterns and relationships in data that were not previously detectable or understood. As a result, these systems have the potential to make more accurate predictions than standalone DSSs [62]. In particular, the ML algorithms used in IDSSs can learn from experience and dynamically improve their performance in the course of time, using their knowledge to provide more accurate predictions and predict future events based on observed trends.

2.8 Geographic Information Systems

A GIS is a framework that collects, organises, analyses, and displays spatially related data. It combines software, applications, and data to uncover patterns, correlations, and combinations that can help users make better decisions. GIS outputs include print maps and charts, spatial analyses, and, more recently, interactive and dynamic information systems and applications that are accessible online. Satellite photos and other remotely sensed data, as well as computerised data capture from fieldwork, have advanced in recent years, reducing the need for manual digitising [89].

GIS has been used in a variety of ways since its original conception. Combining different data sources and transforming them into geospatial datasets allows users to monitor changes in an area, understand trends, identify problems, create corresponding solutions (*e.g.*, transport planning), manage and respond to events (e.g., extreme weather events and disaster response), forecast future scenarios, and set priorities [113]. Geographic data is stored in GIS as thematic layers as exemplified in Fig. 3.

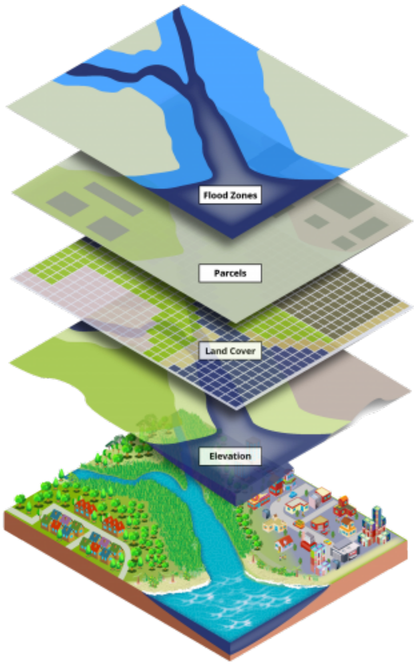


Figure 3: GIS layering example. Adapted from GISGeography.

2.9 Freight Transport Sector

The Freight Transport Sector represents an essential element of the global economy, as it assumes a unique service within supply chains, linking spatially separated supply and demand locations. The accessibility of sites for transporting goods is a vital variable in the economic development of society [93].

Freight transport flows have been steadily increasing, much due to population growth, lower trade barriers, and lower transport costs. In addition, the development of transport flows is driven by rising consumption levels and the customisation of products and services. This growth has been facilitated by large expanses of infrastructures, including roads, railroads, waterways, airports, seaports, warehousing, and transhipment activities. However, in recent decades, flows of goods have also become an area of concern for public policy in a very different way. These concerns relate to protecting the environment from harmful side effects (*e.g.*, local health-related emissions, Greenhouse Gas (GHG) emissions, traffic accidents) resulting from the ripening of freight transport [101].

International trade has increased steadily since the advent of globalisation, and demand for transport services has increased along with it. Multimodal transport is a logical development of traditional unimodal road transport and represents a must-have solution for intercontinental shipments. Multimodal transport is considered the movement of goods using two or more different transport modalities, such as road, rail, air, or inland waterway, as well as short or deep-sea shipping, as part of a contract. In this type of contract, a Multimodal Transport Operator (MTO) oversees the entire haulage contract, from shipping to the destination. Goods movement could be domestic or international; in this case, additional procedures, such as customs clearance, may be required [107]. According to EUROSTAT, in 2018, 76.50% of freight was transported by road, 18% using rail, and 5.50% via inland waterways across the 28 EU member countries [7].

2.10 Vehicle Routing Problem

The VRP is often associated with the distribution of goods between warehouses and customers. VRP was first proposed in 1959 as a generalisation of the TSP, focusing on the optimal routing of a fleet of fuel delivery trucks between a given terminal and a large number of service stations. It was referred as the "*Truck Dispatching Problem*" [31].

VRP is defined as the problem of devising least-cost delivery routes from a depot to a set of geographically dispersed customers, subject to severe constraints. In practice, numerous variants of the problem exist due to the diversity of operating rules and constraints found in different industries. Over time, real-life complexities have been incorporated into the problem, such as travel times, time windows for delivery and collection of goods, and even legislation for drivers' working hours [16], a topic addressed in this dissertation.

2.11 Decision Support Systems in the Freight Transport Sector

Freight transport optimisation is one of the most studied topics in the scientific logistics literature. This proportion is, in its majority, justified by the fact that the Freight Transport Sector is deeply connected to global economic development, representing one of the cornerstones of today's fierce market competition and customers' high demands [17, 13].

The extent of this Freight Transport Sector does justice to the multiple areas of study within it in applying intelligent systems that can optimise and evolve the industry. In addition to the actual methods and circumstances of transport itself [1, 68, 117], a complex collection of factors come into the equation that can directly or indirectly influence the efficiency of the movement of goods. The combination of types of goods [67], geospatial clustering of transport sites [36, 40, 115] and the consolidation of loads [6, 118, 46] are some of the most relevant covered subjects. However, it is extended to the time windows for delivery and collection of goods [15, 5, 111, 39, 43], traffic and bottlenecks in logistics networks and infrastructures [55, 11] and the weather conditions [4, 116]. Recently, the focus has also been the optimal location of warehouses, logistics centres and hubs [59, 114, 112], the optimal combination of carriers and modes to be used in a given transport network [119, 58, 30, 66] and, even the legal regulations of the sector, with, in the case of road freight transport, driving hours and rest periods for drivers being a variable [102, 110, 44] with an increasing importance at the time of decision-making.

Regarding the most studied subfield of this sector, route optimisation, Gayialis et al. (2022) [41] proposes an information system that promotes the effective delivery of commodities inside metropolitan regions. The system uses a collection of Operations Research (OR) algorithms to support logistics operations efficiently. It primarily focuses on optimising vehicle routes and schedules while taking into account delivery time windows, client-specific needs, the peculiarities of the street network, the necessity for dynamic routing and rerouting, and traffic congestion problems in cities. Following a different approach, to minimise overall cost, which includes the cost of generation and emission of GHG, the cost of vehicle usage, and the cost of routing, Tirkolaee et al. (2018) [103] developed a unique model for the multi-trip Green Capacitated Arc Routing Problem (G-CARP). Due to the Nondeterministic Polynomial Time (NP)-hardness of the problem, the authors designed a Hybrid Genetic Algorithm (HGA), in which simulated annealing and heuristics are employed to generate preliminary solutions before a Genetic Algorithm (GA) finds the ideal solution. In the particular subfield of hazardous materials freight transport, Ma et al. (2018) [68] develops a road screening and route optimisation algorithm for dangerous materials transport based on a GA and Levenberg-Marquardt (LM) Neural Network (NN) (GA-LM-NN). For the case study of a freight transport and logistics company in South Italy, Calabrò et al. (2020) [20] used a unique Ant Colony Optimisation (ACO) algorithm to solve a VRP for incoming logistics. The goal was to determine the best possible routes to carry palletised fruit and vegetables from various farms to the main depot while minimising the overall distance travelled by trucks.

Regarding logistical and pedestrian geospatial clustering, Luo et al. (2017) [64] present a *k-means* based station aggregation method that can quantitatively evaluate partitioning by considering flow and geographical distance information. With the specific goal of increasing the ratio of average intra-cluster flow to average inter-cluster flow while retaining the spatial compactness of all clusters, the method was applied to a case study in Haaglanden, Netherlands. In two segments, Chhetri et al. (2014) [26] present a geospatial clustering approach to represent the underlying regional geography of the logistics landscape. First, identifies economic activities and broader spatial logistics functions that characterise an urban setting and then delineates significant spatial logistics employment clusters. Gavrilenko (2020) [40] outlines a clustering algorithm for identifying two groups of regions in the Russian Federation's freight sector. The first group consists of innovators or areas with the best opportunities for innovation implementation and dissemination. The areas of this group are cluster cores. The second group consists of imitators or areas with limited innovative potential. Geographically and by the dominant mode of transport, the areas were organised around the cores. The increase in the efficiency of management of road freight transport at the meso-level is ensured by forming seven clusters in the Russian Federation. Ducret et al. (2016) [37]

create a prototype of a decision-making tool using spatial modelling and clustering to help organise delivery based on city characteristics. The framework's primary goal is to assist distribution providers in achieving appropriate and complete territorial diagnosis before establishing efficient logistics organisations tailored to the parts of cities. This work demonstrates that data that is not necessarily freight data, such as spatial-oriented data, can provide relevant information for urban logistics because the data has been treated operationally, under urban freight issues, and has been confronted with the field during various evaluation steps. Mesa-Arango and Ukkusuri (2015) [75] present a novel approach to clustering demand lanes. Community detection is used to cluster emergent networks and identify profitable demand collections. The authors emphasise the importance of freight logistics network clustering as a strategic decision for carriers. It integrates new businesses into their networks, identifies potential economies, optimises operations, and develops revenue management strategies. The author presents a novel algorithmic approach for clustering demand lanes that is based on dependent sampling over historical data and a series of network transformations. In short, the Latin-hypercube technique is used to collect price and volume samples. To determine the optimal distribution of trucks associated with each sample, a profit maximisation linear program is solved. Trip chains are mined based on these flows to assess the bilateral utility of synergetic lanes. Finally, these utilities are used to populate an interconnectivity network, which is investigated using a community detection algorithm to cluster demand lanes. Letnik et al. (2018) [61] propose a dynamic loading bay assignment model for urban last-mile deliveries. It seeks to address the issue of determining the optimal number and location of loading bays and their management for energy-efficient urban freight deliveries. In conjunction with a routing algorithm, optimisation is based on fuzzy *k-means* clustering of receivers to select the best possible loading bay dynamically. The model is validated using actual delivery data from the historic city centre of Lucca, Italy. Compared to the current situation, simulation results show significant savings in time and distance travelled by freight vehicles, CO_2 emissions and fuel.

One of the other topics studied in the area, as aforementioned, concerns optimal cargo packing. Martello et al. (2000) [71] discusse the issue of finding the least amount of three-dimensional rectangular bins that can be orthogonally packed with a given set of rectangular items. In practice, the problem is extremely difficult to solve because it is strongly NP-hard. Lower bounds are discussed, and it is demonstrated that the continuous lower bound's asymptotic worst-case performance ratio is one-eighth. The development of an exact algorithm for filling a single bin leads to the definition of an exact branch-and-bound algorithm for the three-dimensional bin packing problem, which also incorporates original approximation algorithms. Extensive computational results are presented, involving instances with up to 90 items. It is demonstrated that many instances can reach their optimality in a reasonable amount of time. Bu and Yin (2007) [19] take advantage of a three-dimensional bin packing optimisation to load a set of different size regular goods in containers with given quantities at the container transfer station, with the goal of minimising the number of containers required. The authors developed the mathematical model by taking into account the general characteristics of container multimodal transport. To enforce constringency, the authors published an immune algorithm for the problem. In this algorithm, effective antibody coding is designed with load capacity, loading mass, placement direction and sequence in mind, and affinity function is designed with loading stability and freight transport time window in mind. Computer simulation demonstrates the model's and algorithm's efficiency and validity.

On multimodal transport optimisation, Ziaei and Jabbarzadeh (2021) [119] present a multi-objective model for locating transfer points and routing in a multimodal network of hazardous materials while accounting for multiple different uncertainties and minimising carbon emissions, risk, and transport costs. With polyhedral uncertainty sets, this study models uncertainties in accident probabilities, emission factors, and costs of establishing transfer points and employs robust counterpart optimisation to address these uncertainties. The proposed approach is applied in a case study of petroleum product transport. Findings confirm the effectiveness and robustness of the developed model. It is discovered that for every additional dollar spent, the proposed model can reduce carbon emissions by 873 grams and individuals at risk by 0.28. Within the same context, Jiang et al. (2020) [58] explore an intriguing regional multimodal logistics network design problem in the context of urban cluster development, with a CO_2 emission reduction target and uncertain demands. A regional logistics network, from the standpoint of system optimisation, consists of a centralised logistics authority and a centralised carrier, where the logistics authority is responsible for the regional logistics network configuration in terms of the number, location, and scale of logistics parks, as well as the subsidies of rail transport links, and the carrier's decisions include the choice of transport route for each logistics demand. This multi-stakeholder decision-making problem is first formulated as a bi-level programming model, followed by its equivalent Mathematical Programming with Equilibrium Constraints (MPEC) to depict the leader-follower behaviors, given a determined logistics demand pattern. The case study demonstrates that the proposed robust optimisation method is a viable solution to the problem of designing a regional multimodal logistics network with uncertain demand. Lv et al. (2019) [66] approach multimodal transport optimisation as a Transit Consolidation Problem (TCP), an issue that considers simultaneous loading and unloading of cargo to/from containers. A GA-based technique for addressing TCP was developed in this work because of GAs's effectiveness in handling complex optimisation issues. The result makes two distinct contributions to the literature. Before considering limits on container capacity and transport optimisation, this paper provides a detailed description and quantification of the TCP process. Secondly, this study employs a genetic algorithm to analyse the possible advantages of transit consolidation for freight forwarders, including the prudent use of resources and the optimal mode and transport route. Using a genetic algorithm with the double-layer chromosomes coding, the TCP can effectively solve and obtain a satisfactory solution. Moreover, the load coefficient can reduce the number of containers; the fixed cost of containers in different modes of transport can limit the number of containers and obtain the optimal layout of cargo; the average dynamic tonnage utilisation ratio improved by 15.86% by comparing with establishing the hub. Udomwannakhet et al. (2018) [108] discuss the various multimodal transport planning methodologies available. There is a need for network modelling approaches to optimise people and freight transport using multiple techniques. A multimodal transport cost model demonstrates and clarifies multimodal transport routing alternatives while selecting the most efficient transport modes. Risk factors are also taken into account for each route, mode of transport, and node link. The purpose of this paper is to conduct a literature review on a generic multi-transport network model based on the use of computational techniques to improve efficiency.

Hours-of-Service (HOS) regulations represent one of the variables recently added to the problem of freight transport, and they should be considered when generating long-haul routes for several days or weeks. Transport companies, particulary, are responsible for ensuring that driving plans can be safely carried out while complying with regulatory break and rest periods. HOS regulations in the United States (US), the EU, Canada, and Australia typically impose daily and weekly rest periods and driving and working hours limits. Tilk and Goel (2020) [102] investigate the vehicle routing and truck driver scheduling problem, in which routes and schedules must comply with truck driver hours of service regulations. It describes a backward labelling method for producing feasible time schedules. It demonstrates how the labels produced by the backward method can be combined with labels created by a forward labelling method. The bidirectional labelling is embedded in a branch-and-price-and-cut approach and is being evaluated for hours of service regulations in the US and the EU.

2.12 Tools and Technologies

Table 1 presents the diverse technological tools that were adopted in this work during the research and development of the IDSS.

Table 1: Technological tools used throughout the development and research of the IDSS.

Scientific Field	**Tool**	**Description**
Data Extraction	Microsoft Structured Query Language (SQL) Server	Microsoft SQL Server is one of the most powerful and advanced database management systems. The tool is a relational database management system designed to perform business-critical applications. It provides industry-leading tools to optimise and manage information, including DM and reporting, analysis, mobile devices and workgroup synchronisation [12].
Exploratory Data Analysis (EDA)	SweetViz	Sweetviz is an open-source Python library that generates appealing aesthetic exploratory data analysis for various applications in business analytics, data science and AI [14].
	Pandas Profilling	Pandas Profiling is an open-source Python module for analysing datasets and examining the efficiency of ML algorithms. It employs the pandas' library and works with Python-based data analysis tools. Straightforwardly, Pandas Profiling can extract summaries of datasets [18].

Table 1: Technological tools used throughout the development and research of the IDSS.

Scientific Field	Tool	Description
Development	Python	Python is a high-level, all-purpose programming language. The language is designed for clarification, focusing on code readability. Python uses garbage collection and has dynamic typing, supporting a variety of paradigms for programming, including functional, object-oriented, and structured programming. The language was put forth as a link between shell programming and C. Over the last decade, it has become one of the most popular languages for software development in academic research settings [91].
	pandas	Based on the Python programming language, pandas is an open-source data analysis and manipulation tool that is fast, strong, flexible, and straightforward. It was created for quantitative finance applications and is considered the essential high-level building block for performing realistic, real-world data analysis in Python [73].
	NumPy	NumPy is a package for scientific computing in Python. It provides array handling, data processing and simple linear algebra with efficient multi-dimensional arrays. It speeds up basic operations and has been used to perform functional numerical tasks, including signal processing, image processing, matrix computations and data analysis [53].
	PTV Developer	PTV Developer is a mobility application that uses location service APIs to address the VRP. It facilitates freight transport, taking into account the physical restrictions imposed by the road network and the dimensions of the freight vehicles, in contrast to other routing tools like Google Maps or OpenStreetMap (OSM). PTV Developer has millions of roads, addresses, and tourist attractions and is accessible in over 240 countries. Its capability to consider the legal constraints placed on drivers is also crucial to note [48].

Table 1: Technological tools used throughout the development and research of the IDSS.

Scientific Field	Tool	Description
	RanCoord	RanCoord is an open-source Python package that makes it simple to create random coordinates inside a given set of geographic borders. Through its geocoding, the framework provides ways to create, save, and plot a predetermined number of random coordinates derived from a location polygon or a short name or address. TSP and VRP are two development and research initiatives in logistics and transport that are ideally suited to and focused on using this package [21].
	GeoPy	GeoPy is a Python package for location and geographic coordinates. It offers a variety of geospatial data manipulation tools, such as pointwise variance, distance-to-centre computations, spatial joins and aggregations, and dealing with raster maps [76].
	Shapely	Shapely is a Python tool for manipulating and analysing geographic data. It offers resources for topologies creation, algebraic operations, and geographic description [42].
Modelling	scikit-learn	scikit-learn is a powerful package for ML. Through a consistent Python interface, it offers a variety of practical techniques for statistical modelling and ML, such as dimensionality reduction, clustering, and classification. NumPy, SciPy, and Matplotlib represent the foundation for this library [83].
	OR-Tools	OR-Tools is an open-source optimisation software suite designed to tackle the world's most challenging problems such as the VRP, flows, Mixed Integer Linear Programming (MILP), and Constraint Programming (CP) [85].
	CVXPY	CVXPY is an open-source Python-based modelling language for convex optimisation issues. Instead of stating optimisation issues in the limited standard form demanded by solvers, it enables a natural expression that follows the mathematics [34].

Table 1: Technological tools used throughout the development and research of the IDSS.

Scientific Field	**Tool**	**Description**
	COIN-OR Branch and Cut (CBC)	CBC is a popular open-source MILP solver. It was created with various significant applications, including state estimation, designing chemical reaction networks, simulating biomedical models, and structural optimisation [27].
Routing	OSM	OSM is a collaborative initiative to build a free, editable global geographic database. Its users create and upgrade its content through various activities, including installing map tiles, uploading edits and corrections, and undertaking tagging work and mapping features [28].
	Open Source Routing Machine (OSRM)	OSRM is a C++ implementation of a high-performance routing engine for optimised paths in road networks. It combines the OSM project's open and accessible road network with complex routing algorithms [65].
Data Visualisation	Kepler.gl	The Kepler.gl tool is a high-performance web-based application for visually explore massive geolocation data independent of the data type. Several open source technology platforms, including one of the most extensive geographic databases in the world and specialist visualisations tools like d3.js, Google Chart Tools, and OGR, are used to build the Kepler project [106].
	Mapbox	Mapbox provides location information, including maps and directions to developers in applications, mobile devices, websites and more. Its robust platform and easy-to-use tools allow users to create unique products that combine data with maps to help people get around safely and efficiently [70].
	Plotly	Plotly is a Python package for scientific data execution and analysis. It offers a set of tools for creating and sharing interactive visualisations that make it simple to explore, comprehend, and communicate data across multiple domains [56].

Table 1: Technological tools used throughout the development and research of the IDSS.

Scientific Field	**Tool**	**Description**
Containerisation	Docker	Docker is an open-source platform for developers that allows them to run and manage applications in isolated containers. Its particularities enable its users to run multiple applications on a single machine, test them in isolation, and share resources like storage and networks [74].
Deployment	Flask	Flask is a Python-based web framework that uses tools, libraries, and technologies to create web applications. It allows rapid and easy development of REST APIs and can be expanded with as many third-party packages as necessary [47].
Data Encryption	JSON Web Token (JWT)	JWT is a small, Uniform Resource Locator (URL) safe, representation of a signed or encrypted data object or message digest. The data object may include any content and be either compressible material, cryptographically hashed content, or certificate signatures [95].
RESTful API	Swagger	Swagger is an open-source set of tools for creating REST-based APIs. It streamlines the API development process by defining standards and providing the tools to create beautiful, safe, performant, and scalable APIs [99].
	Postman	Postman is an API development tool that aids in the creation, testing, and modification of APIs. This tool contains almost all of the functionalities that any developer could require [57].
WSGI HTTP Serving	Gunicorn	Gunicorn is a UNIX Python Web Server Gateway Interface (WSGI) HyperText Transfer Protocol (HTTP) Server. It gives the developer an elegant and straightforward way to deliver production-ready applications to users without having to worry about the server's internal workings [25].
	Waitress	Waitress is an HTTP server for Python. It is a production-quality pure-Python WSGI server with very acceptable performance and is compliant with the most current standards [72].

Table 1: Technological tools used throughout the development and research of the IDSS.

Scientific Field	Tool	Description
Data Warehousing	SQLAlchemy	SQLAlchemy is a Python SQL toolkit and Object Relational Mapping (ORM) tool that provides application developers complete SQL capability and flexibility. It offers a complete set of well-known enterprise-level persistence patterns created for fast, adequate database access and translated into a straightforward, Pythonic domain language [12].
Code Versioning	Git	Git is a free and open-source distributed version control system designed to handle projects of all sizes with speed and efficiency [23].

2.13 Summary

The current chapter reviewed the most recent state-of-the-art works on the application of software development, ML techniques, and optimisation approaches to the Freight Transport Sector. Thus, we explained the strategy for searching the literature, followed by a theoretical introduction to a set of relevant addressed concepts. The sector under research was then presented, followed by an analysis of practical case studies developed in the area, with the goal of detecting potential research gaps that this dissertation intends to pursue and add value to. Finally, a brief enumerated description of the tools and technologies used in response to the proposed problem and objectives was presented.

3

An Intelligent Decision Support System for the Road Freight Transport

3.1 Introduction

Due to the global market competition, current world market dynamics (*e.g.*, effect of the COVID-19 pandemic) and other issues (*e.g.*, sustainability), there is an increased pressure to improve the road Freight Transport Sector. The current exponential commercial, industrial and technological development experienced today, combined with the obligation to fulfill the escalating consumer needs in this increasingly fierce market competition, makes the road Freight Transport Sector a key player in global economic development. In contrast, this uncontrollable development and consumerism lead to unsustainability and harmful effects on the environment and the health of society in general. Until recently, the planning of transport activities was mainly focused on cost minimisation [33]. However, the transport sector represents a substantial source of GHG emissions. As a result, one of the main goals of supply chain management programs is to develop more sustainable transport solutions [82]. Under this context, an IDSS represents a potential valuable solution [8] , since they can be used to extract valuable insights from the freight transport data, allowing to address several objectives (*e.g.*, reduce costs and gas emissions).

In this chapter, we propose an E2ES that consists of an IDSS that targets the main activities and tasks involved in planning freight transport. This work covers all the objectives and expected results stated in Section 1.2 of Chapter 1, excluding the one responsible for the multimodal transport optimisation, given the *abmn.pt - Business Solutions* requirement to isolate that module. The isolated module is presented in a later stage of this dissertation, namely in Chapter 4. The IDSS execution process was designed in conjunction with a consultancy company that is responsible for the transport management of several Portuguese freight transport companies. Moreover, the business knowledge and data obtained proved to be crucial throughout the development, testing, and validation of the IDSS functionalities. In particular, the IDSS includes: an adaptive distance-constrained geospatial clustering algorithm, for the detection of possible clusters of loading or unloading locations; a tool for ranking drivers in their allocation to a transport service by considering their distance to a loading location; an overall routing optimisation process that considers the rules stipulated by the EC in Regulation No. 561/2006 regarding driving time and rest periods, and includes truckload and sequencing operations. In order to evaluate the proposed IDSS, a questionnaire was assembled using the TAM 3 [109] and open interviews.

3.2 Materials and Methods

3.2.1 Problem Formulation

The set of Portuguese companies covered in the development of this IDSS is currently operating in the Freight Transport Sector and is primarily composed of the following components: the main logistics centre, a fleet of vehicles, and drivers. The main logistics centre represents the company headquarters and it is also used as a warehouse and distribution centre. The primary role is operationalising logistics activities and freight transport. We highlight that the transport planning and resource allocation are performed using a software system in which the proposed IDSS will be incorporated. This system works as the cornerstone of the entire transport and logistics business.

Regarding the drivers, these represent the human resources that perform transport activities. In addition to being essential to the execution of the transport services, these resources require a particular organisation and management to comply with the requirements set forth by the EC Regulation No. 561/2006, which establishes driving and rest periods.

Each business that holds a fleet of vehicles can diverge in their types, models, sizes, capacities, and purposes. The digital transition philosophy in these Portuguese freight transport companies has led to data storage of the mentioned tasks and components. However, data analysis, pattern recognition, and automation are not optimised or even implemented. The range of transport planning activities is managed and arranged based on intuition and human experience of the transport software system users. Often, this manual approach is non optimal. The lack of interoperability and divergent (ofen *ad-hoc*) human logistics decisions lead to a misuse of resources and, consequently, to an unsustainable development.

3.2.2 Proposed IDSS

In this work, we propose an IDSS that follows a three-tier architecture that assumes three logical and physical computing elements, as presented in Fig. 4. The Data Layer is accountable for extracting and processing data from the different databases regarding the surrounding factors involved in freight transport planning, including data on freight services to be planned, data associated with the various drivers (relevant to comply with EU rules) and data on the physical properties of trailers. In addition, there is a middleware with PTV [77], a mobility software endoweded with the capability of providing distances and routes that consider the truck dimensions and legal restrictions. Its outputs are sent to the Processing Layer.

The Processing Layer of the IDSS is composed of three modules: Groupage Detection Module (GpDM), Resource Optimisation Module (RsOM), and Route Optimisation Module (RtOM).

Currently, the company plans routes based on specific Portuguese districts (*e.g.*, Viseu, Coimbra, Castelo Branco), creating long and more expensive routes, thus spending more resources. To tackle this issue, the GpDM assumes an unsupervised ML algorithm in charge of planning clusters of route points to travel based on a maximum distance, thus producing better route planning and cost management (*e.g.*, time and fuel). The developed Distance-constrained Geospatial Clustering (DcGC) algorithm can group geographic locations based on their geographical positional similarities and a maximum adjustable distance between transports inserted in the same cluster. This way, it is possible to improve the current approach implemented with an adaptive method to create dynamic zones according to the reality of each moment of the component execution. This functionality was designed to address the inefficiency of the current grouping of transports that, despite belonging to the same zone, are highly dispersed in the vertical or horizontal geographical spectrum. It is also important to note the particularity of its performance in cross-border areas where the problem of geographic coordinates separation from very close locations is addressed. The DcGC pseudocode is presented in Alg. 1.

Algorithm 1: DcGC algorithm.

Inputs : A dataset D consisting of geographic coordinates $\{latitude, longitude\}$ and a maximum great-circle distance $maxDistance$ between coordinates part of the same cluster.

Outputs: A N set of clusters that satisfy the maximum distance between geographic coordinates.

```
initialisation;
distances = Empty list;
for iteration ∈ [1, coordinatesNumber] do
    clusters = Fit K-means model of N = iteration centers;
    for c ∈ clusters do
        dm = Distance matrix calculation of the geographic coordinates part of c using the
          great-circle distance;
        append dm to distances
    end
    if distances contains a distance > maxDistance then
        continue
    else
        return clusters
    end
    DcGC
end
```

Afterwards, a Time Difference (TD) leaderboard of possible drivers is calculated and provided by the Resource Optimisation Module (RsOM) module. The TD leaderboard is calculated using the following equation:

$$TD = |PID - AD_d| \tag{3.1}$$

where PID represents the expected start date by the company, and AD_d represents the driver availability date. This ranking is then presented to the driver manager, who will choose the most suitable driver to travel the route provided by the GpDM module.

The last module is the Route Optimisation Module (RtOM), which is responsible for creating and optimising the route to be covered based on two objectives: sequencing the road to be covered in the shortest possible distance; maximising the load carried by the various vehicles on these routes. Furthermore, this module provides a better selection of vehicles depending on each route created. One of the main advantages lies within the ability of each module to be executed individually, and if requested, the outputs will be returned to the IDSS user.

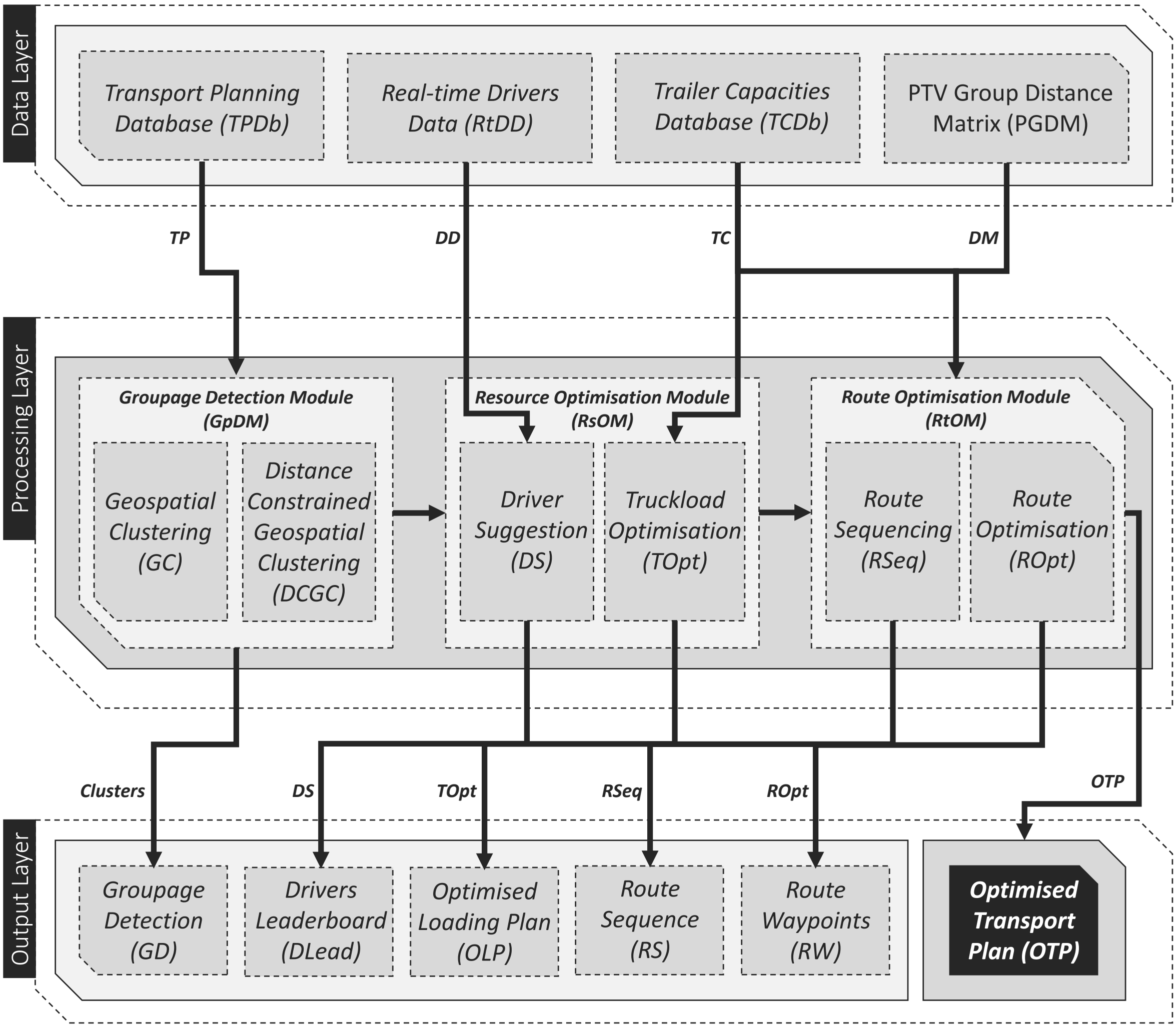


Figure 4: Proposed IDSS architecture.

3.2.3 Evaluation Methodology

The proposed IDSS was evaluated by a research team comprising AI, Software Development, and freight transport experts. Several individual components of the proposed IDSS (*e.g.*, GpDM module) are evaluated by considering specific logistics scenarios. To globally validate the full IDSS, we adopted the TAM 3, which allowed us to develop a questionnaire with ten questions to assess the system acceptance and evaluation. TAM 3 is composed of six constructs: Perceived Usefulness (PU); Perceived Ease of Use (PEOU); Perception of External Control (PEC); Job Relevance (REL); Output Quality (OUT); and Behavioral Intention (BI). Each question included a 5-point Likert scale [90] choice with a range of 1 (strongly disagree) to 5 (strongly agree) for each response. Direct feedback from the consultancy company responsible for transport management was also captured. Moreover, we also map the IDSS tool functionalities, comparing them with current logistics practices and procedures (termed here as the AS-IS approach).

3.3 Results

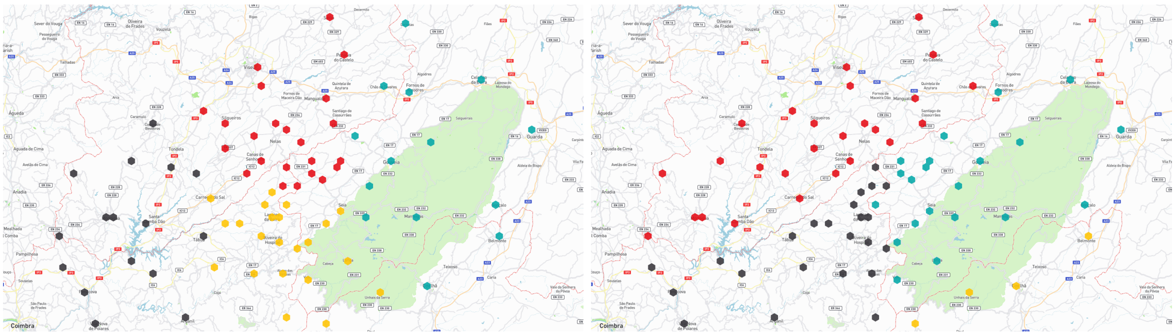
3.3.1 Developed IDSS Execution Process

The designed IDSS was exclusively developed using the Python language. Prototyping and deployment were accomplished using `Flask`. Finally, for the set of critical and proposed features in this chapter, namely ML and optimisation techniques, `scikit-learn`[1], `ortools`, and `PTV Developer`[2], were the tools used. It is also important to note that, at a preliminary stage of development, to overcome the lack of data derived from General Data Protection Regulation (GDPR) bureaucratic issues, the `RanCoord`[3] package was used. The package offers an easy procedure to generate random coordinates within a set of geographic boundaries [21].

Groupage Detection Module (GpDM) is responsible for detecting possible load groupages using zone grouping and clustering techniques. Fig. 5(a) and Fig. 5(b) depicts the transition from static to adaptive and dynamic zones, elevating the compactness of the clusters. This illustrates the grouping of 71 transport load zones to be carried out, spread over 4 districts within the Portuguese territory. The AS-IS approach considers the district borders, thus obtaining 4 grouping zones. These zones are divided into 26, 23, 22, and 4 loading locations respectively.

Coincidentally, the proposed approach using the DcGC algorithm, executed with a standard maximum distance used by the different transport companies of 55 kilometres, also clustered the data into 4 clustering zones, consisting of 23, 14, 16, and 22 locations. The equal number of clusters makes it possible to compare the efficiency and primacy of DcGC. The enhancement of transport groupage can be verified through each approach's distance matrix of the different clusters. The AS-IS method obtained an average distance between geographic points of the same cluster of 19.45 kilometres, composed of 23.14, 19.04,

[1] https://scikit-learn.org/stable/index.html
[2] https://developer.myptv.com/
[3] https://github.com/hugodscarvalho/rancoord



(a) "AS-IS" method. (b) DcGC method.

Figure 5: Comparison of the "AS-IS" method and the geographic coordinates clustering using DcGC.

20.21, and 15.44 kilometres. On the other hand, the DcGC algorithm allowed a remarkable reduction of this distance to 15.42 kilometres, composed of averages of 13.98, 18.76, 15.71, and 13.23 kilometres. The contribution of the algorithm extends to balancing the number of locations grouped in the same cluster. The standard deviation of the number of sites to be grouped in each cluster was reduced from 8.60 to 3.80.

The **Resource Optimisation Module (RsOM)** includes the driver suggestion (DS) and the optimisation of truckload allocation (TOpt). Table 2 illustrates how this component behaves, including the set of input attributes in the Transport and Drivers sections. The Leaderboard section represents the result of this micro-framework, assembling qualified drivers to execute a transport arranged by TD. The resulting leaderboard (DLead) relevance lies within the decision-making support throughout the driver selection process, enhancing its efficiency. The information regarding driving hours will allow the manager to comply with the EU standards. The second component of the module relates to truckload optimisation (TOpt) by formulating the problem as a bin packing problem. The goal of issues of this nature is to find the best combination to pack a set of items of given sizes or weights into containers with fixed capacities. Using the Solving Constraint Integer Programs (SCIP) for MILP solver [3, 2], we developed an algorithm capable of optimising the allocation of loads in different vehicles to maximise the occupancy of the vehicle and minimise the number of vehicles needed to pack and transport them, turning it into a multi-objective optimisation. Regarding a set of 11 loads stored in a specific warehouse of one of the freight transport companies, it was possible to find the best combination of loads with a minimised solution of 4 vehicles and maximised occupation of the different vehicles of 99.60%, 100.00%, 77.10%, and 98.80% with a capacity established at 25 tons, the standard legal net load value. This optimisation process is depicted in Fig. 6.

The last module of the IDSS represents the **Route Optimisation Module (RtOM)**. As exemplified in Table 3 and Fig. 7, the component takes as input parameters a set of loads to be transported to 13 geographically separated hypermarkets (HM) from the Central Region of Portugal from a logistics centre and, as a second input, a set of 7 available vehicles and their respective physical profiles. This component of the IDSS generates a load allocation to the set of vehicles and the sequencing of hypermarkets to be carried out by each one. The solution presented in Table 3 and illustrated in Fig. 7 is a multi-objective optimisation that takes into account: (i) minimising the distance traveled by each vehicle; (ii) maximising the occupancy of each vehicle. Objective (ii) maximises the occupancy of each vehicle, allowing the usage of only 5 vehicles of the 7 available. Finally, the second component of the module only increases the value of the solution provided by the previous one. Rather than just returning the sequence of locations to cross, it returns the complete route to be taken by each vehicle through a set of waypoints depicted in Fig. 7. The user of the IDSS can specify whether to avoid tolls or not.

Table 2: Driver Suggestion (DS) module attributes and output

Context	**Attribute**	**Description**
Transport	Transport ID	Transport identification
	Expected start datetime	Expected start date
	Start latitude	Initial transport latitude point
	Start longitude	Initial transport longitude point
	Expected driving time	Estimated driving hours
Driver	Driver ID	Driver identification
	Availability datetime	Availability date and time
	Day driving hours	Expected daily driving time
	Week driving hours	Expected weekly driving time
	Fortnight driving hours	Expected fortnightly driving time
	Availability latitude	Latitude geographical point
	Availability longitude	Longitude geographical point
Leaderboard	Driver ID	Driver identification
	Time window	Time duration to reach transport location
	Distance	Distance to transport location
	Day driving hours	Expected daily driving time
	Week driving hours	Expected weekly driving time
	Fortnight driving hours	Expected fortnightly driving time

3.3.2 Evaluation

The designed TAM 3 questionnaire is presented in Table 4. The *Results* columns include the average of five freight transport managers responses based on a Likert scale and their corresponding standard deviation (SD). It must be underlined that, in order to avoid bias, these managers correspond to people from the

Table 3: Route Optimisation Module (RtOM) execution.

Vehicle ID	Load	Occupancy	Sequence
1	19	95.45%	LC - HM4 - HM8 - LC
2	25	100.0%	LC - HM3 - HM9 - HM13 - HM12 - LC
3	23	92.0%	LC - HM7 - HM6 - LC
4	24	96.0%	LC - HM2 - HM10 - HM11 - LC
5	23	92.0%	LC - HM1 -HM5 - LC

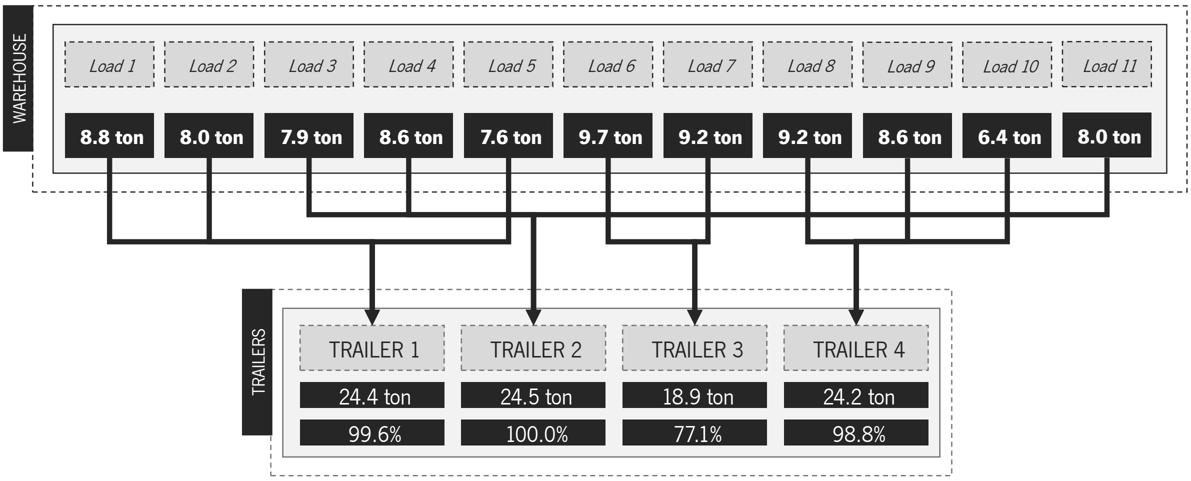


Figure 6: Truckload optimisation functionality usage.

different transport companies who were not directly involved in the research, design, or development process of the proposed IDSS. The average responses range of 3.6 to 4.6 and the standard deviation (SD) is comprised between ± 0.16 and ± 0.40, revealing a positive evaluation and acceptance of the proposed IDSS and compliance among the various people involved in the questionnaire. The most positive responses are inserted in Job Relevance (REL1 and REL2) and Behavioral Intention (BI1) with an average rating of 4.6. In order to increase the value of this evaluation, the given feedback during all technical sessions of design, presentation, and enrichment of the IDSS was also aggregated. Regarding the GpDM, the feedback was unanimously positive with these managers pointing out a remarkable optimisation of the clustering process of transport services evidenced by the examples of comparison between the AS-IS process and the developed DcGC algorithm. Regarding RsOM, the feedback on the truckload optimisation component was particularly emphasised based on the transition between intuition and human experience to an optimal combinatorial truckload optimisation algorithm. On the other hand, the driver suggestion component, despite its value recognition to the decision-making process and its clear relevance, the suggestion of its improvement regarding increased automation and inclusion of the driving rules and rest times stipulated by the EC was given by 4 of the 5 managers. Lastly, RtOM was the best verbally rated module with a

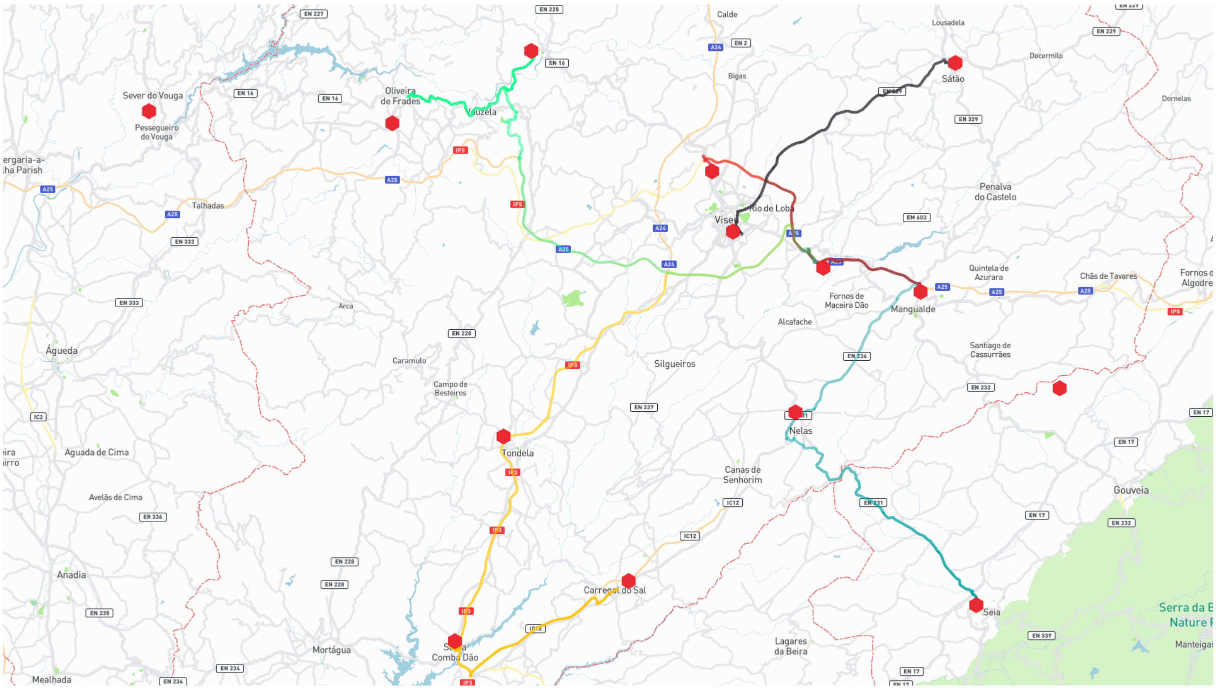


Figure 7: Route Opitimisation Module (RtOM) explanatory execution.

clear satisfaction of the complexity and benefits provided by its sequencing and detailed routing of an optimised transport plan for each required vehicle. Overall, all respondents concluded that the proposed IDSS (including its three modules), is valuable to optimise transport and logistics activities and represents a possible positive digital disruption of the companies. Table 5 summarises the main features introduced by the proposed IDSS, which substantially enhances the current activities and procedures (AS-IS).

3.4 Summary

This chapter presented an IDSS to optimise transport and logistics activities as an E2ES in a set of Portuguese companies currently operating in the Freight Transport Sector and clients of *abmn.pt - Business Solutions*. The IDSS functionalities meet the set of the objectives and expected results stated in Section 1.2 of Chapter 1, excluding the one responsible for the multimodal transport optimisation, given the *abmn.pt - Business Solutions* requirement to isolate that module. The isolated module is presented in a later stage of this document, in Chapter 4. The proposed IDSS follows a three-tier architecture: Data Layer – responsible for extracting and processing data from different data sources accordingly to the respective component to be addressed; Processing Layer – which represents the core of the proposed IDSS, consisting of the set of freight transport crucial procedures and tasks, namely: (a) Groupage Detection Module (GpDM); (b) Resource Optimisation Module (RsOM); (c) Route Optimisation Module (RtOM); and Output Layer – which includes the results and outputs from the middle layer.

The execution process was evaluated by five managers who were not in the research, design, or development process of the proposed IDSS to avoid bias. The adopted evaluation methodology was TAM 3,

Table 4: The adopted TAM 3 questionnaire and results (average of five responses).

Construct	Items	Question	Results	
			Mean	SD
Perceived Usefulness (PU)	PU1	Using the IDSS improves my performance in my job.	4.0	± 0.40
	PU2	The IDSS is (potentially) useful in my job.	4.4	± 0.24
Perceived Ease of Use (PEOU)	PEOU1	I find the IDSS to be very easy to use.	3.6	± 0.24
	PEOU2	It is easy to get the information that I want from the IDSS.	3.6	± 0.24
Perceptions of External Control (PEC)	PEC1	I have the knowledge to use the IDSS.	3.8	± 0.16
Job Relevance (REL)	REL1	In my job, the usage of the IDSS is important.	4.6	± 0.24
	REL2	The use of the IDSS is pertinent to my various job-related task.	4.6	± 0.24
Output Quality (OUT)	OUT1	The quality of the output I get from the IDSS is high.	4.4	± 0.24
	OUT2	I have no difficulty telling others about the results of the IDSS.	4.0	± 0.4
Behavioral Intention (BI)	BI1	Assuming I had access to the IDSS, I intend to use it.	4.6	± 0.24

Table 5: Comparison between the current functionalities (AS-IS) and proposed IDSS tasks and procedures.

Functionalities	**AS-IS**	**IDSS**
Group transport services by zone	✓	✓(GC)
Cluster transport services by geographic radius		✓(DcGC)
Driver suggestion		✓(DS)
Optimised truckload allocation		✓(TOpt)
Optimised route sequencing		✓(RSeq)
Route	✓	✓(ROpt)
Route optimisation		✓(ROpt)

comprising questionnaires and feedback gathered throughout technical sessions of the design, presentation, and enrichment of the IDSS. Overall, a very positive evaluation and acceptance were obtained. In future work, we intend to add new modules and components to the IDSS articulated with the suggestions and the expected evolution of the Freight Transport Sector.

4

Multimodal Transport Optimisation

4.1 Introduction

International trade has increased steadily since the advent of globalisation, and demand for transport services has grown along with it. Intercontinental shipments and multimodal transport represent a development of traditional unimodal road transport. The transport industry has always provided a challenging but stimulating environment for applying optimisation techniques. The volume of goods transported worldwide has increased interest and utility of operational research methodologies required to manage complex transport systems effectively.

Numerous transport modalities, including trucks, planes, and ships, are used in delivery services. Costs will vary depending on the routes and transport modes. Considerations for cost reduction include transport costs, delivery time limits, and commodities consolidation. This project models the essential scenario using mathematical programming and finds a solution that minimises total costs. Fig. 8 and Fig. 9 present simple examples of multimodal transport optimisation combinations.

Figure 8: Multimodal Transport demonstrative example 1.

Figure 9: Multimodal Transport demonstrative example 2.

4.2 Materials and Methods

4.2.1 Proposed Multimodal Transport Optimisation

In this dissertation, we propose an additional module for the intelligent decision support system for **Multimodal Transport Optimisation**.

The problem uses the framework **CVXPY**, a modeling language for convex optimisation problems. Instead of the constrictive standards demanded by mathematical solvers, the tool allows a natural style that follows mathematics. It uses a linear and mixed integer programming solver called **CBC**.

4.2.2 Presuppositions

(a) The delivery method is **deterministic**; therefore, random events will not affect delivery time or cost.

(b) The optimisation will only consider **regular containers**.

(c) The containers' constraints only relate to the **good's volumes**.

(d) The model solely assesses the **main carriageways**. There is no consideration for the first and last miles between the end user and the origin or destination delivery site.

(e) Overall cost considers **transport costs**, **warehouse cost** and **goods taxes**.

(f) The lowest unit of time is **day**.

4.2.3 Objective

The objective is to minimise the total cost, which comprises three components: transport costs, storage costs and taxes. Transport cost includes fixed costs of the route and the containers. Route cost covers fixed costs of all used routes, while container cost is obtained through the multiplication of the number of containers used in each route and the cost of each container. The storage costs are determined by multiplying the products' volume, number of storage days, and the warehouse fee per cubic meter per day. The total of all goods' transit and import taxes makes up the tax cost. Eq. (4.1) formulates the objective function.

$$\underset{X,Y,Z}{\text{minimise}} \quad \textit{Transport Cost(X)} + \textit{Storage Cost(Y)} + \textit{Tax Cost(Z)} \tag{4.1}$$

4.2.4 Constraints

1. Each good must ship-out from its origin to another logistic infrastructure in the network.

2. Each good must, eventually, ship-in at its destination.
3. Each good cannot ship-out to its origin.
4. Each good cannot ship-in from its destination.
5. Each good must have the same number of ship-ins and ship-outs.
6. Each good can only be transported from a logistic infrastructure once.
7. Each good can only be transported to a logistics infrastructure once.
8. In a transition logistic infrastructure (neither origin nor destination), the ship-in date and time should be before the ship-out date and time.
9. At the origin, the ship-out date and time should be the same or after the order date.
10. At the destination, the ship-in date and time should be the same or before the required delivery date and time.

4.3 Results

4.3.1 Developed Multimodal Transport Module

In our case, we used real-world data from one of the transport companies under study to select logistic infrastructures spread over 11 cities in 5 different sovereign countries in **Western Europe**, namely **Portugal, Spain, France, the United Kingdom** and the **Netherlands**. The transport network created by these logistics infrastructures assembles 100 routes connecting them. The considered transport modalities include **truck, sea, rail** and **air**. Every route has a unique **transport modality**, **cost**, **travel time**, and **weekly time windows** for transport. Each city has a logistics hub where items can be stored for a period to accommodate specific transport time windows or wait for other goods to be delivered simultaneously. Table 6 comprises the considered logistic infrastructures network information, and Fig. 10 depicts them geographically.

Given the routes that make up our logistics infrastructure network, we assembled a set of load orders to demonstrate the execution and possible solutions suggested by this multimodal transport optimisation component. Table 7 presents the set of load orders, including their identification, origin, destination, type of commodity, order and required delivery dates, and the type of freight transport (*i.e.*, International or Domestic).

Table 6: Logistic Infrastructures Network

Sovereign Country	City	Type	Transport Mode	Description	Latitude	Longitude
Portugal	Porto	Logistics hub	Truck	Porto Logistics Hub	41.176663	-8.648467
		Railway facility	Rail	Porto Station	41.149409	-8.585082
	Maia	Airport	Air	Francisco Sa Carneiro Airport	41.238853	-8.671588
	Matosinhos	Maritime facility	Sea	Port of Leixoes	41.185741	-8.701646
Spain	Barcelona	Logistics hub	Truck	Barcelona Logistics Hub	41.503361	2.190335
		Railway facility	Rail	Barcelona Sants Railway Station	41.378759	2.139764
		Airport	Air	Josep Tarradellas Barcelona El Prat Airport	41.300735	2.082033
	Bilbao	Maritime facility	Sea	Port of Bilbao	43.342888	-3.045569
France	Paris	Logistics hub	Truck	Paris Logistics Hub	48.881541	2.324011
		Railway facility	Rail	Paris Nord Station	48.881932	2.357039
		Airport	Air	Charles de Gaulle Airport	49.006732	2.551259
	Le Havre	Maritime facility	Sea	Port of Le Havre	49.476341	0.149223
United Kingdom	London	Logistics hub	Truck	London Logistics Hub	51.488791	-0.222977
		Railway facility	Rail	London Station	51.531169	-0.125276
		Airport	Air	Heathrow Airport	51.470970	-0.450544
	Felixstowe	Maritime facility	Sea	Port of Felixstowe	51.948508	1.323143
Netherlands	Amsterdam	Logistics hub	Truck	Amsterdam Logistics Hub	52.416401	4.836212
		Railway facility	Rail	Amsterdam Central Station	52.377637	4.902190
		Airport	Air	Amsterdam Airport Schiphol	52.309154	4.764017
	Rotterdam	Maritime facility	Sea	Port of Rotterdam	51.904612	4.485924

Table 7: Load orders for Multimodal Transport Optimisation

Order ID	Origin	Destination	Commodity	Order Date	Delivery Date	Freight Type
1	London Logistics Hub	Porto Logistics Hub	Gold	01-09-2022	25-09-2022	International
2	Amsterdam Logistics Hub	Barcelona Logistics Hub	Vehicle Parts	09-09-2022	23-09-2022	International
3	Barcelona Logistics Hub	London Logistics Hub	Pharmaceutics	03-09-2022	24-09-2022	International
4	Amsterdam Logistics Hub	Porto Logistics Hub	Medical Instruments	09-09-2022	27-09-2022	International
5	Paris Logistics Hub	Amsterdam Logistics Hub	Jewelry	04-09-2022	27-09-2022	International
6	Porto Logistics Hub	Barcelona Logistics Hub	Corn	05-09-2022	23-09-2022	International
7	Paris Logistics Hub	Barcelona Logistics Hub	Aluminium	03-09-2022	24-09-2022	International
8	Porto Logistics Hub	London Logistics Hub	Integrated Circuits	05-09-2022	27-09-2022	International
9	Barcelona Logistics Hub	Amsterdam Logistics Hub	Steel Pipes	03-09-2022	27-09-2022	International
10	London Logistics Hub	Amsterdam Logistics Hub	Lithium	06-09-2022	01-10-2022	International

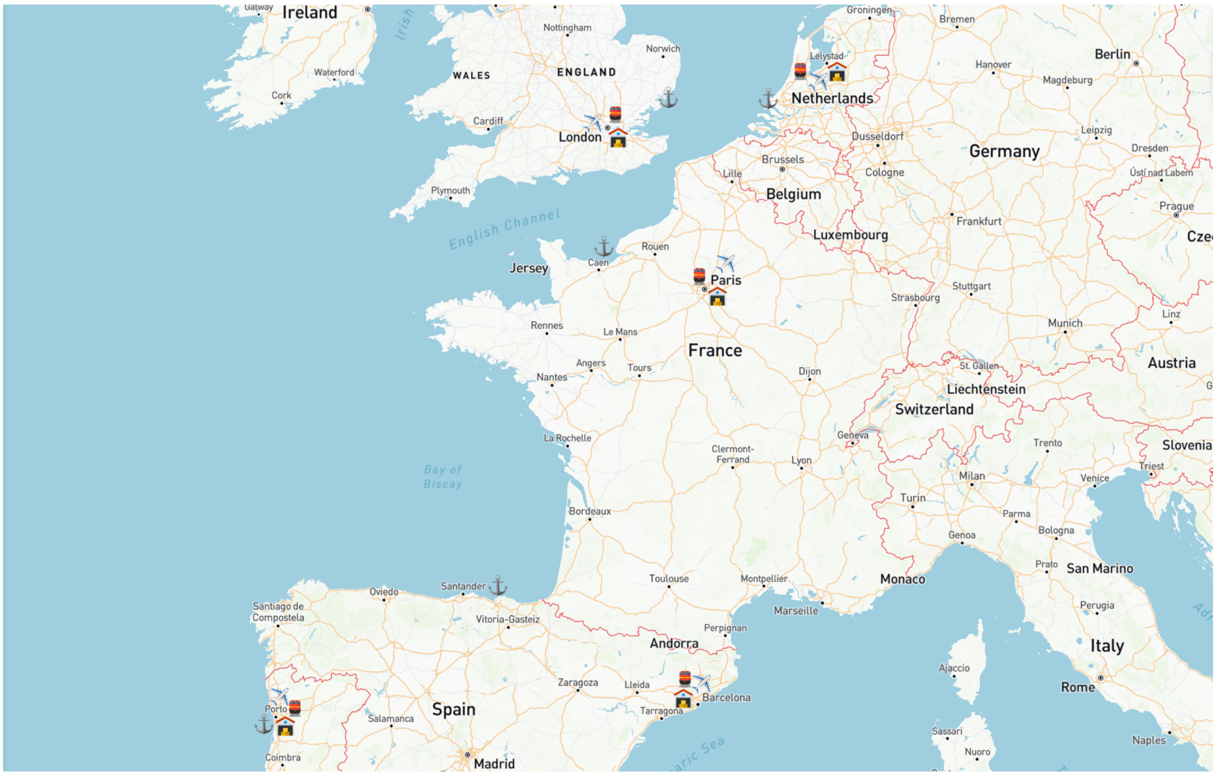


Figure 10: Geographic distribution of the Logistic Infrastructures.

4.3.2 Evaluation

The module developed within the scope of multimodal transport optimisation represents a late addition to the project's expected results and, consequently, to this dissertation. Thus, its embryonic character leads to the scarcity of results and evaluation methods. However, given the increasing relevance of the multimodal transport process, we believe, together with managers from the Freight Transport Sector, *abmn.pt - Business Solutions*, and its clients, that this represents a critical module in the objectives and expected results of the project. It should also be mentioned that the non-existence of such a process in the companies covered by this project will not allow a comparison of results. The entire design and development process follows the freight transport companies' business context and the multimodality ideologies presented in the practical case studies of the literature review. According to *abmn.pt - Business Solutions* and its clients, this will represent a fundamental evolution in their strategies, in terms of innovation, financial return and environmental sustainability.

4.4 Summary

This capter presented an additional module for the intelligent decision support system for multimodal transport optimisation.

The goal was to reduce the total cost, including transport costs, storage costs, and taxes. The cost of transport includes the fixed costs of the route and the containers. Route cost consists of the fixed costs of all used courses, whereas container cost is calculated by multiplying the number of containers used in each route by the cost of each container. The storage costs are calculated by multiplying the volume of the products, the number of storage days, and the warehouse fee per cubic meter per day. The total of all transit and import taxes on all goods constitutes the tax cost.

This late addition to the project in conjunction with the inexistence of such a process in the companies covered by this project has not allowed a reliable evaluation of the results and the innovation brought by the module. However, according to the company and its clients, this module will mark a significant shift in their strategies in terms of innovation, financial return, and environmental sustainability.

5

RanCoord – Random Geographic Coordinates Generator for Mobility, Transport, and Logistics Research and Development Activities

5.1 Introduction

Data has become a valuable asset. However, while it can potentially provide the extraction of useful insights, it also can be associated with privacy issues. Geospatial data is no different as it may represent the identifier of an individual. Because of this, users of geospatial data must consider privacy and confidentiality implications. In addition to its privacy concerns, it should also be noted that geospatial data is often a scarce resource [60].

This chapter presents an additional development of this dissertation project, an open-source Python package that offers a straightforward, quick, and intuitive strategy to circumvent these concerns and availability barriers by generating random geographic coordinates to suit the user needs, both in terms of their quantity and their global geographic position. These outputted geographic coordinates can be used for any of its user's purposes. The package scope arose from both development and research in areas related to transport and logistics. Fig. 11 depicts the package architecture. The pseudocode presented in Alg. 2 complements the information presented in Fig. 11, by presenting step by step, the set of tasks performed.

5.2 Materials and Methods

5.2.1 Proposed Package

The framework follows a three-tier architecture, which separates the application into 3 logical and physical computing tiers. The core lies within the Processing Layer, composed of 5 modules whose different funcionalities can be used either chained together or individually.

Algorithm 2: RanCoord algorithm.

Inputs : A name/address $nameAddress$ or a geographic polygon $geoPoly$; a number of geographic coordinates num; booleans $dmBool$, $saveBool$, and $mapBool$, to calculate distance matrix dm, save the data in a $file$, and generate the map, respectively

Outputs: A set of num geographic coordinates $\{Latitude, Longitude\}$; a distance matrix dm; a file $file$; a map map

```
initialisation;
if geoPoly is empty then
  | polygon = geocode nameAddress
else
  | polygon = geoPoly
end
minLat, minLon, maxLat, maxLon = polygon.bounds
coordinates = Empty list;
while length of coordinates < num do
  | randomCoordinate = random.uniform( minLat, maxLat ), random.uniform(
  |   minLon, maxLon )
  | if randomCoordinate within polygon then
  |   | append randomCoordinate to points
  | else
  |   | continue
  | end
end
if dmBool then
  | dm = distanceMatrix( coordinates )
else
  | continue
end
if saveBool then
  | multipleFormatsSaver( coordinates )
else
  | continue
end
if mapBool then
  | plotCoordinates( coordinates )
else
  | continue
end
return coordinates, dm, file, map
```

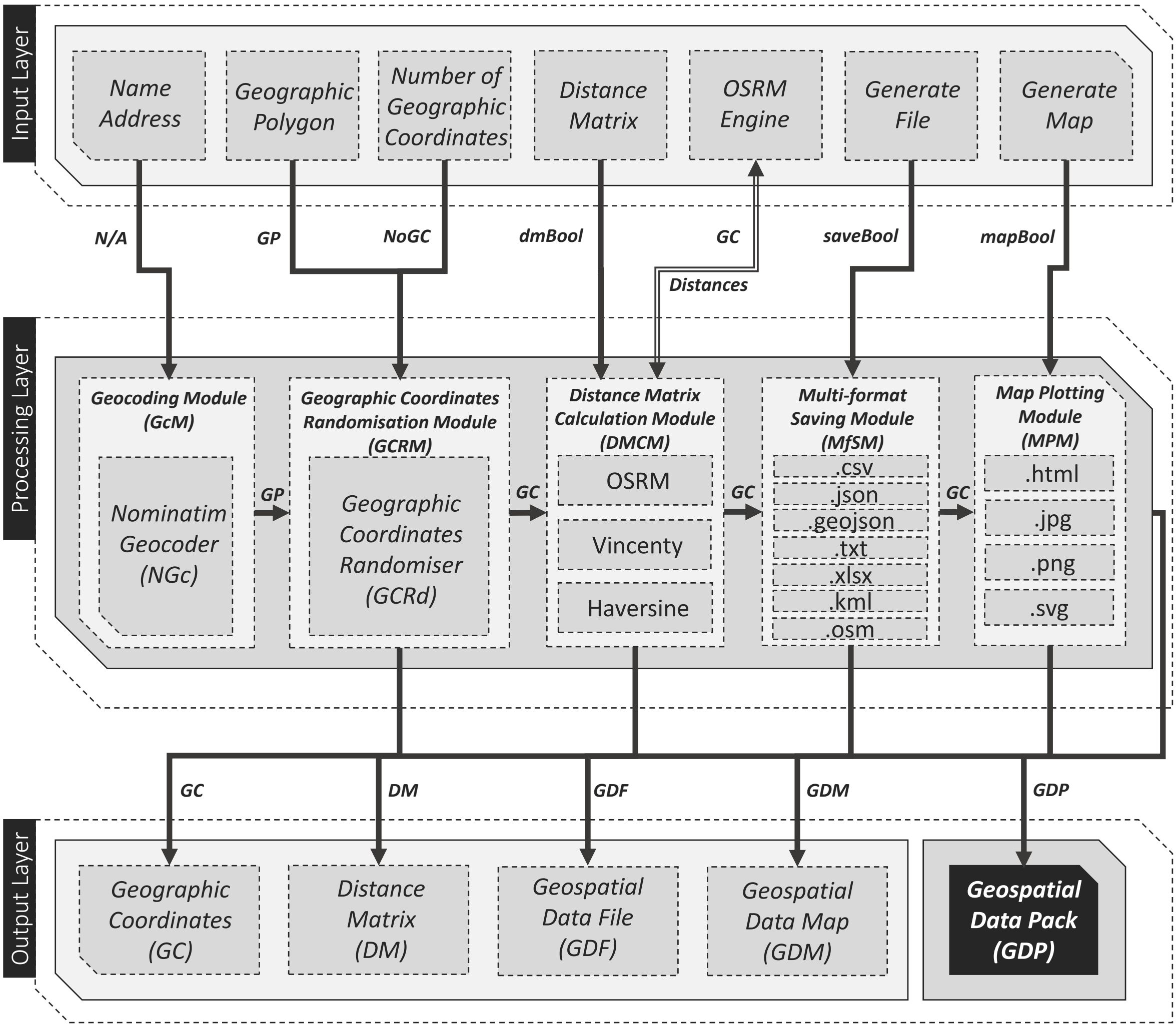


Figure 11: RanCoord architecture.

5.2.1.1 Input Layer

The **Input Layer** is the communication layer of the application, including the set of parameters for the geographic coordinates generation that meets the user's needs and a possible data source for the distance matrix calculation, the OSRM Engine, a modern open source C++ routing engine for shortest paths in road networks. Regarding the submission parameters, these include:

(a) **Name/Address (N/A)** or a **Geographic Polygon (GP)** to be considered.

(b) **Number of Geographic Coordinates (NoGC)** to be generated.

(c) The option to return the **distance matrix (dmBool)** of the geographic coordinates to be generated using the default distance calculation method (*Haversine*).

(d) The option to save or not to **save the geographic coordinates (saveBool)** using the standard format (JavaScript Object Notation (JSON)).

(e) The option to **plot and save the resulting map (mapBool)** of the geographic coordinates to be generated.

It is relevant to highlight that the input parameter **(a)** can be filled and executed using two options of the user's choice:

(i) **Name/Address –** The name, address, query, or a structured query the user wishes to use as the geographical position to be considered. A street, city, county, state, country, or postcode should be provided.

(ii) **Geographic Polygon (GP) –** object that represents a filled area consisting of a list of at least three coordinate tuples that forms the outer ring and a (possible) list of hole polygons [42].

5.2.1.2 Processing Layer

The **Processing Layer** represents the main phase of the framework where the random generation of geographic coordinates occurs. Coordinates that meet the set of parameters specified by the user in the Inputs phase. If the user has chosen the first option of execution using the Name/Address option in the Input Layer, the process will go through a geocoding stage of the corresponding name or address, where its bounding box[1] will be requested to a third party. Accordingly, the behavior of the components that make up this phase can be described as follows:

1. **Geocoding Module (GcM)** - geocoding stage of the specified name or address requesting its bounding box to *Nominatim*. This geocoder uses OSM data to find locations on Earth by name and address [81]. This set of coordinates will be converted into a Shapely Polygon and then passed to the randomisation component.

2. **Geographic Coordinates Randomisation Module (GCRM)** - the core component of the current phase and the whole framework. Using the Shapely Polygon from the previous component, depending on the user's choice and the number of locations to be generated, it generates a random set of pairs containing latitude and longitude following a continuous uniform distribution. The distribution formulated in Eq. (5.1) and presented in Fig. 12 describes an experiment with an arbitrary outcome that lies between certain bounds. The bounds are defined by the parameters, *a* and *b*, which are the minimum and maximum values.

$$fx(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

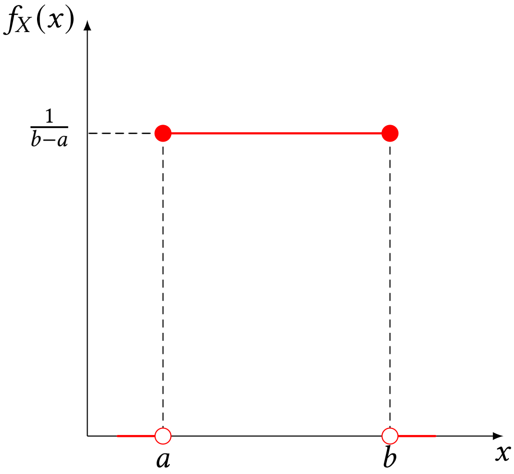


Figure 12: Probability density function of the continuous uniform distribution

3. **Distance Matrix Calculation Module (DMCM)** - This component allows the user to get the distance matrix of the generated geographic coordinates. This matrix will have a size of $N \times N$ where N is the number of geographic coordinates specified by the user and, consequently, generated. This two-dimensional data structure is often required to solve transport and logistics problems such as the TSP and VRP. Thus, using RanCoord, there is no longer a need for a third party or user implementation to obtain the distance matrix.

Table 8: $N \times N$ distance matrix

$$D = \left(\begin{array}{c|ccccc} & (lat, lon)_1 & (lat, lon)_2 & (lat, lon)_3 & \dots & (lat, lon)_d \\ \hline (lat, lon)_1 & 0 & d_{12} & d_{13} & \dots & d_{1d} \\ (lat, lon)_2 & d_{21} & 0 & d_{23} & \dots & d_{2d} \\ (lat, lon)_3 & d_{31} & d_{32} & 0 & \dots & d_{3d} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (lat, lon)_n & d_{n1} & d_{n2} & d_{n3} & \dots & 0 \end{array} \right)$$

As shown in Table 8, the main diagonal of the square matrix will be null since the distance between the same point is zero. The remaining values will be filled using the method selected by the user, where *d* represents the distance between the respective pairs of coordinates. The tool allows the calculation of this distance matrix using three different methods:

(a) **Haversine** – Determines the **great-circle distance**[2] between two points on a sphere given their longitudes and latitudes [69].

[1] set of minimum and maximum longitudes and latitudes of a geographical area.
[2] shortest distance between two points on the surface of a sphere.

(b) **Vincenty** – Calculates geodesic distances between a pair of latitude/longitude points on an ellipsoidal model of the Earth. Unlike the Haversine method for calculating distance on a sphere, this formula is an iterative method and assumes the Earth is an ellipsoid[3] [69].

(c) **Open Source Routing Machine (OSRM)** – High-performance routing engine for shortest paths in road networks. It combines sophisticated routing algorithms with the open and free road network data of the OSM project [65]. Despite its high performance, the fact that it represents a solution that verifies road networks and not just a mathematical distance between two coordinates makes this the most computationally expensive method but also the most accurate.

4. **Multi-format Saving Module (MfSM)** - Save the generated geographic coordinates in the following standard formats: Comma Separated Values (CSV), JSON, Geographic JavaScript Object Notation (GeoJSON), Text File (TXT), Microsoft Excel Spreadsheet (XLSX), Keyhole Markup Language (KML), and OSM.

5. **Map Plotting Module (MPM)** - Plot the generated geographic coordinates on a map using the Python Folium package, which is powered by Leaflet.js [87]. The map can be saved in the following formats: HyperText Markup Language (HTML), Joint Photographic Experts Group (JPG), Portable Network Graphics (PNG), and Scalable Vector Graphics (SVG).

5.2.1.3 Output Layer

The Output Layer represents the final stage of the tool and comprises the output of the different modules of the Processing Layer. Thus, a Geospatial Data Pack (GDP) can be generated if the framework is used in a chained approach or the following single outputs:

(a) **Geographic Coordinates (GC)** – Set of requested geographic coordinates that meet the parameters submitted by the user.

(b) **Distance Matrix (DM)** – Distance matrix of the generated geographic coordinates.

(c) **Geospatial Data File (GDF)** – A file containing the generated geographic coordinates.

(d) **Geospatial Data Map (GDM)** – A map containing the generated geographic coordinates.

Fig. 13 shows the plotted result of a RanCoord execution with the specification of 50 geographic coordinates located in the city limits of London, United Kingdom.

[3] surface that may be obtained from a sphere by deforming it by means of directional scalings, or more generally, of an affine transformation.

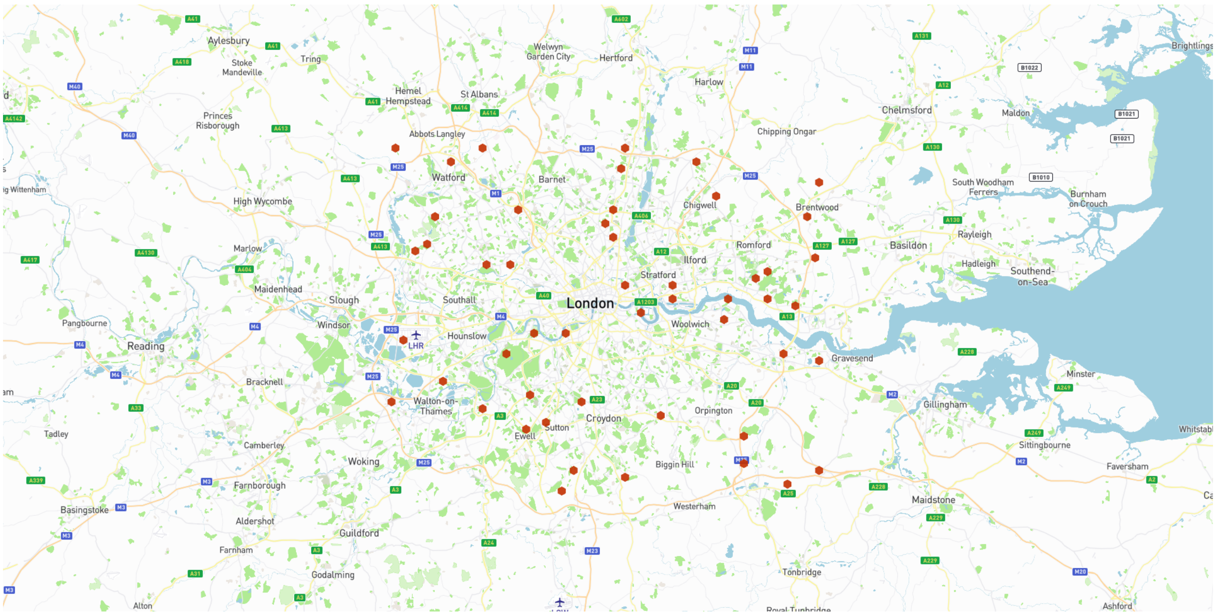


Figure 13: Generation of 50 random coordinates within the geographic limits of the city of London, United Kingdom.

5.3 Impact on Academic Research

RanCoord provides a straightforward, fast, and intuitive way to generate random coordinates within a set of boundaries. A growing number of researchers are studying TSP, VRP, and their variants, considering real-life applications and scenarios [100]. The present framework reduces the time and effort required to acquire geographic coordinates within a specific location, eliminates the need to use geographic data that does not fit a particular problem to be addressed, and provides geographic coordinates in an effortless, fast, and customized way to the user's needs. Moreover, since it generates random coordinates, the module does not rise privacy or confidentiality issues.

RanCoord was used in several scientific studies and development activities. As a means of overcoming the lack of data derived from GDPR issues, RanCoord was adopted to generate vehicle routing problem instances within locations that match the geographic logistics context of several freight transport companies, thereby improving the communication of results with the managers of the various companies and, at the same time, accelerating the research and development of optimisation methods in months [97].

5.4 Future Work

RanCoord is a relatively new package, having its lifetime spanned across since April 2022. The future work's main focus is to optimise its performance in terms of execution time, where several changes in its source code will be foreseen, however, maintaining its core functionalities presented here. New methods of generating geospatial data beyond continuous uniform distribution can be expected. Finally, the ultimate aspiration is to evolve this framework by integrating routing optimisation methods, mainly Metaheuristics such as genetic algorithms, ACO algorithms, simulated annealing, and tabu search [29]. It is also expected a new module that provides functionalities associated with geospatial clustering using methods such as K-Means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). The DcGC algorithm presented in Chapter 3 may also be integrated into the framework.

5.5 Summary

This chapter described an extension of this dissertation project, an open-source Python package that provides a simple, quick, and intuitive strategy for overcoming these concerns and availability barriers by generating random geographic coordinates to suit the user's needs, both in terms of quantity and global geographic position. These outputted geographic coordinates can be used for any purpose desired by the user. The scope of the package arose from both development and research in transport and logistics.

6

Prototype

After the work carried out and presented in Chapter 3, Chapter 4, and Chapter 5, the focus transited to the development of instruments endowed with the ability of technology and knowledge transfer of the entire set of case studies covered and the range of functionalities implemented in the development of components of an IDSS for the Freight Transport Sector.

Two APIs were developed as a requirement of *abmn.pt - Business Solutions*, given its acquaintance with the integration of REST APIs. The first one, depicted in figure Fig. 14, transfers the developments mentioned in Chapter 3 and Chapter 5. In order not to compromise the functioning of the previous tool, *abmn.pt - Business Solutions* requested the isolation of the multimodal transport optimisation module, represented in the second API developed, shown in Fig. 15. The endpoints for each API are already being tested and validated by *abmn.pt - Business Solutions*. According to *abmn.pt - Business Solutions*'s work methodology, these functionalities are being integrated into *abmn.pt - Business Solutions*'s proprietary software entitled aTrans as they are validated.

This deployment work used the `Flask` tool and its respective containerisation using `Docker` containers. `JWT` was used to ensure the integrity and security of the work developed. The APIs were developed and adapted for two operating systems, Linux and Windows. The only divergence between the two versions is related to the serving framework used. For Linux, `Gunicorn` was used, and for Windows, due to operating system conflicts, the alternative was `Waitress`.

Figure 14: IDSS API.

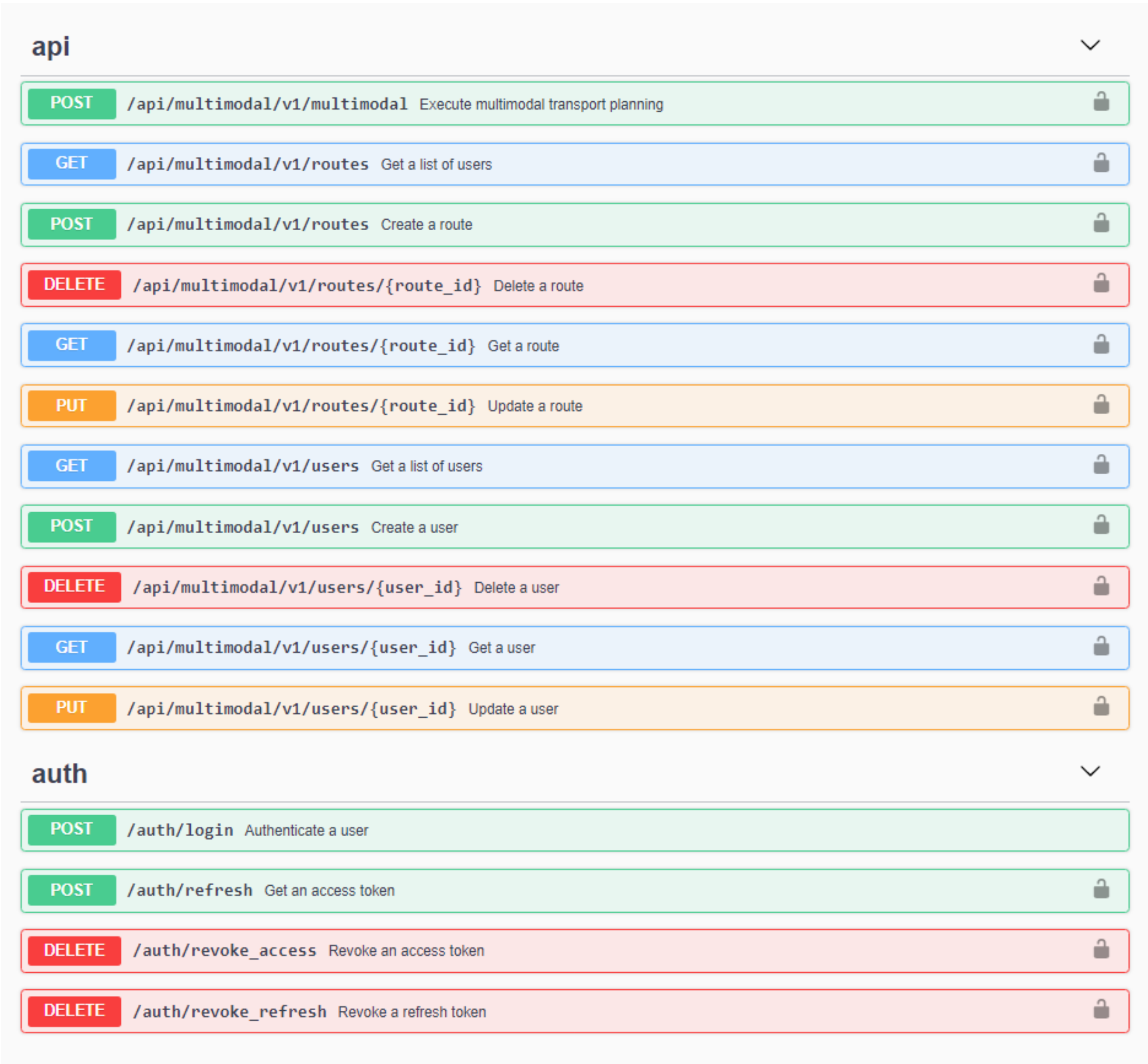


Figure 15: Multimodal transport optimisation module REST API.

7 Conclusions

7.1 Overview

There is increased pressure to improve the Freight Transport Sector as a result of global market competition, current world market dynamics (*e.g.*, the effect of the COVID-19 pandemic), and other issues (*e.g.*, sustainability). The current exponential commercial, industrial, and technological development, combined with the responsibility to satisfy escalating consumer needs in this increasingly competitive market, makes road freight transport a key player in global economic development. In contrast, uncontrollable development and consumerism have unsustainable and harmful effects on the environment and the general health of society.

This dissertation work was developed under a R&D project designated "*aDyTrans – Dynamic Transportations Platform*", resulting from the collaboration between a technological interface centre, the CCG and *abmn.pt - Business Solutions*, a technology consultancy company operating in the Freight Transport Sector. The work developed was essentially based on the implementation of software development, ML techniques, and optimisation approaches on several fronts of the Freight Transport Sector, including resource optimisation, driver suggestion and management according to legal regulations, transport loads planning, allocation and optimisation, route sequencing and optimisation, and, finally, multimodal transport optimisation. However, additional work has been incorporated with the development and scientific contribution of an open-source Python package that provides a straightforward, fast, and intuitive way to generate random coordinates within a set of geographic boundaries. This work suits mobility, transport and logistics research and development activities.

This document was divided into four major parts, comprising seven chapters. The first section includes Chapter 1 and Chapter 2, which provided a brief overview of the issues addressed in this thesis and the relevant current state of the art. Specifically, Chapter 2 performs a theoretical introduction to a set of relevant addressed concepts. Then, it provides a critical state-of-the-art analysis of studies concerning the use of computational support in decision-making in the Freight Transport Sector.

The second part of the document comprised the primary experimental work developed during this thesis, including Chapter 3, Chapter 4, Chapter 5, and Chapter 6. Chapter 3 presented a IDSS for optimising transport and logistics activities in the Freight Transport Sector. This IDSS consists of three main modules that can be used independently or in combination: geographic clustering detection of transport services, transport driver suggestion, and route and truckload optimisation. It was built from the ground up to support real-time data. It consists of a E2ES, which covers all of the main transport and logistics processes, from database registration to optimised transport plan. Chapter 4 provided an additional artefact regarding multimodal transport optimisation. The solution uses mathematical programming to model the essential freight transport scenario to find a solution that minimises the total costs of combinatorial routing plans that include multiple modes of transport, such as air, sea, road, and rail. The mentioned dissertation project's additional development, Chapter 5, focused on an open-source Python package that allows quickly generating random coordinates within a set of geographic boundaries. Through its geocoding, the framework provides methods for generating, saving, and plotting a predetermined number of random coordinates generated from a location polygon or a simple name/address. Such geographic methods, such as the TSP and VRP, are well-suited for research and development activities in transport and logistics. Chapter 6 presents the knowledge and source code transfer environment of the work developed within the three previously mentioned chapters. The third part included Chapter 7, which presents the main findings of this dissertation, its limitations and potential gaps for future research. Finally, the final section (Appendix A) details a user manual for the developed Python package.

7.2 Discussion

Overall, it is possible to conclude that the IDSS implementation was successful, in that all the proposed objectives were achieved and that the goal of supporting decision-making in the freight transport sector was also completed. Besides the stipulated requirements, it was possible to agglomerate the whole set of functionalities developed in a prototype that contains two APIs already being integrated into *abmn.pt - Business Solutions* proprietary software. In the Freight Transport Sector domain, the literature review combined with the imposed requirements made it possible to detect flaws and opportunities in the development of solutions for the optimisation and evolution of the sector. Thus, besides representing an added value solution for the set of *abmn.pt - Business Solutions* clients, it is possible to state that the work developed also brings a substantial scientific contribution to the sector by developing innovative methods and techniques.

In terms of scientific contributions, the work developed during this dissertation resulted in the publication of a conference paper at the *"23rd International Conference on Intelligent Data Engineering and Automated Learning (IDEAL2022)"* entitled *"An Intelligent Decision Support System for Road Freight Transport"*, with its presentation scheduled for November 24-26, 2022 in Manchester, United Kingdom:

- Silva Carvalho, H., Pilastri, A., Matos, L. M., Matta, A., Novais, R., & Cortez, P. (2022). **An Intelligent Decision Support System for Road Freight Transport**. *Intelligent Data Engineering* *and Automated Learning – IDEAL 2022*. Cham: Springer International Publishing.

In addition, an open-access Python package was developed and made available on the GitHub and Python Package Index (PyPI) platforms, and its reproducibility is stored on the Code Ocean platform. With the development of this package comes the submission and acceptance of a journal paper named *"RanCoord - A random geographic coordinates generator for transport and logistics research and development activities"* in the journal Software Impacts:

- Silva Carvalho, H., Pilastri, A., Novais, R., & Cortez, P. (2022). **RanCoord – A random geographic coordinates generator for transport and logistics research and development activities**. *Software Impacts, 14.*

Finally, it is also worth mentioning that all the work developed has received positive feedback regarding its performance and optimisation in the Freight Transport Sector, directly and indirectly leading to a reduction in costs and time and increasing sustainability.

7.3 Future Work

Due to technological developments, it is necessary to pay special attention to the entire set of updates in the surrounding variables associated with the analysed industry. Indeed, recent developments indicate that the evolution of the Freight Transport Sector is heading towards carbon neutrality with the use of sustainable transport modes, such as, for example, electric vehicles. Regarding the work developed under this dissertation, a Continuous Integration and Continuous Delivery (CICD) approach will be adopted to keep up with technological developments and keep the contributions introduced in this dissertation updated.

Bibliography

[1] Rusul Abduljabbar, Hussein Dia, Sohani Liyanage, and Saeed Asadi Bagloee. "Applications of Artificial Intelligence in Transport: An Overview". In: *Sustainability* 11.1 (2019-01), p. 189. doi: 10.3390/su11010189. url: https://doi.org/10.3390/su11010189 (cit. on p. 14).

[2] Tobias Achterberg. "SCIP: solving constraint integer programs". In: *Mathematical Programming Computation* 1.1 (2009-07), pp. 1–41. issn: 1867-2957. doi: 10.1007/s12532-008-0001-1. url: https://doi.org/10.1007/s12532-008-0001-1 (cit. on p. 29).

[3] Tobias Achterberg, Timo Berthold, Thorsten Koch, and Kati Wolter. "Constraint Integer Programming: A New Approach to Integrate CP and MIP". In: *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*. Ed. by Laurent Perron and Michael A. Trick. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 6–20. isbn: 978-3-540-68155-7. doi: https://doi.org/10.1007/978-3-540-68155-7_4 (cit. on p. 29).

[4] Taslima Akter, Suman Kumar Mitra, Sarah Hernandez, and Karla Corro-Diaz. "A spatial panel regression model to measure the effect of weather events on freight truck traffic". In: *Transportmetrica A: Transport Science* 16.3 (2020-01), pp. 910–929. doi: 10.1080/23249935.2020.1719552. url: https://doi.org/10.1080/23249935.2020.1719552 (cit. on p. 15).

[5] Derya Eren Akyol and René B.M. De Koster. "Determining time windows in urban freight transport: A city cooperative approach". In: *Transportation Research Part E: Logistics and Transportation Review* 118 (2018-10), pp. 34–50. doi: 10.1016/j.tre.2018.07.004. url: https://doi.org/10.1016/j.tre.2018.07.004 (cit. on p. 15).

[6] Julian Allen, Michael Browne, Allan Woodburn, and Jacques Leonardi. "The Role of Urban Consolidation Centres in Sustainable Freight Transport". In: *Transport Reviews* 32.4 (2012-07), pp. 473–490. doi: 10.1080/01441647.2012.688074. url: https://doi.org/10.1080/01441647.2012.688074 (cit. on p. 14).

[7] Claudia Archetti, Lorenzo Peirano, and M. Grazia Speranza. "Optimization in multimodal freight transportation problems: A Survey". In: *European Journal of Operational Research* 299.1 (2022), pp. 1–20. issn: 0377-2217. doi: https://doi.org/10.1016/j.ejor.2021.07.031. url: https://www.sciencedirect.com/science/article/pii/S0377221721006263 (cit. on p. 14).

[8] David Arnott and Graham Pervan. "A critical analysis of decision support systems research". In: *Formulating Research Methods for Information Systems: Volume 2*. London: Palgrave Macmillan UK, 2015, pp. 127–168. isbn: 978-1-137-50988-8. doi: 10.1057/9781137509888_5. url: https://doi.org/10.1057/9781137509888_5 (cit. on p. 24).

[9] David Arnott and Graham Pervan. "A Critical Analysis of Decision Support Systems Research Revisited: The Rise of Design Science". In: *Journal of Information Technology* 29.4 (2014-12), pp. 269–293. doi: 10.1057/jit.2014.16. url: https://doi.org/10.1057/jit.2014.16 (cit. on p. 12).

[10] Maria Ijaz Baig, Liyana Shuib, and Elaheh Yadegaridehkordi. "Big data adoption: State of the art and research challenges". In: *Information Processing & Management* 56.6 (2019), p. 102095. doi: https://doi.org/10.1016/j.ipm.2019.102095. url: https://www.sciencedirect.com/science/article/pii/S0306457319301773 (cit. on p. 8).

[11] E Y Balaev, T V Konovalova, A E Litvinov, S L Nadiryan, and I S Senin. "Optimization of freight traffic in cities using the forecast transport model of freight traffic". In: *IOP Conference Series: Materials Science and Engineering* 918.1 (2020-09), p. 012045. doi: 10.1088/1757-899x/918/1/012045. url: https://doi.org/10.1088/1757-899x/918/1/012045 (cit. on p. 15).

[12] Michael Bayer. "SQLAlchemy". In: *The Architecture of Open Source Applications Volume II: Structure, Scale, and a Few More Fearless Hacks*. Ed. by Amy Brown and Greg Wilson. aosabook.org, 2012. url: http://aosabook.org/en/sqlalchemy.html (cit. on pp. 18, 23).

[13] Tolga Bektaş, Jan Fabian Ehmke, Harilaos N. Psaraftis, and Jakob Puchinger. "The role of operational research in green freight transportation". In: *European Journal of Operational Research* 274.3 (2019-05), pp. 807–823. doi: 10.1016/j.ejor.2018.06.001. url: https://doi.org/10.1016/j.ejor.2018.06.001 (cit. on p. 14).

[14] Francois Bertrand. *Sweetviz: Automated EDA in Python*. https://github.com/fbdesignpro/sweetviz. Accessed: 07/07/2022. 2020 (cit. on p. 18).

[15] Narath Bhusiri, Ali Gul Qureshi, and Eiichi Taniguchi. "Application of the Just-In-Time Concept in Urban Freight Transport". In: *Procedia - Social and Behavioral Sciences* 125 (2014-03), pp. 171–185. doi: 10.1016/j.sbspro.2014.01.1465. url: https://doi.org/10.1016/j.sbspro.2014.01.1465 (cit. on p. 15).

[16] Kris Braekers, Katrien Ramaekers, and Inneke Van Nieuwenhuyse. "The vehicle routing problem: State of the art classification and review". In: *Computers & Industrial Engineering* 99 (2016-09), pp. 300–313. doi: 10.1016/j.cie.2015.12.007. url: https://doi.org/10.1016/j.cie.2015.12.007 (cit. on p. 14).

[17] Juan José Bravo and Carlos Julio Vidal. "Freight transportation function in supply chain optimization models: A critical review of recent trends". In: *Expert Systems with Applications* 40.17 (2013-12), pp. 6742–6757. doi: 10.1016/j.eswa.2013.06.015. url: https://doi.org/10.1016/j.eswa.2013.06.015 (cit. on p. 14).

[18] Simon Brugman. *pandas-profiling: Exploratory Data Analysis for Python*. https://github.com/pandas-profiling/pandas-profiling. Accessed: 07/07/2022. 2019 (cit. on p. 18).

[19] Lei Bu and Chuan-zhong Yin. "Three-Dimensional Loading Optimization Problem of Freight Transshipment in Multimodal Transportation". In: *International Conference on Transportation Engineering 2007*. American Society of Civil Engineers, 2007-07. doi: 10.1061/40932(246)533. url: https://doi.org/10.1061/40932(246)533 (cit. on p. 16).

[20] Giovanni Calabrò, Vincenza Torrisi, Giuseppe Inturri, and Matteo Ignaccolo. "Improving inbound logistic planning for large-scale real-world routing problems: a novel ant-colony simulation-based optimization". In: *European Transport Research Review* 12.1 (2020-04). doi: 10.1186/s12544-020-00409-7. url: https://doi.org/10.1186/s12544-020-00409-7 (cit. on p. 15).

[21] Hugo Silva Carvalho, André Pilastri, Rui Novais, and Paulo Cortez. "RanCoord – A random geographic coordinates generator for transport and logistics research and development activities". In: *Software Impacts* (2022), p. 100428. issn: 2665-9638. doi: https://doi.org/10.1016/j.simpa.2022.100428. url: https://www.sciencedirect.com/science/article/pii/S2665963822001129 (cit. on pp. 20, 28).

[22] Stephen Cave, Kanta Dihal, and Sarah Dillon, eds. *AI Narratives*. Oxford University Press, 2020-02. doi: 10.1093/oso/9780198846666.001.0001. url: https://doi.org/10.1093/oso/9780198846666.001.0001 (cit. on p. 9).

[23] Scott Chacon and Ben Straub. *Pro git*. Apress, 2014 (cit. on p. 23).

[24] Jinchuan Chen, Yueguo Chen, Xiaoyong Du, Cuiping Li, Jiaheng Lu, Suyun Zhao, and Xuan Zhou. "Big data challenge: a data management perspective". In: *Frontiers of Computer Science* 7.2 (2013-04), pp. 157–164. doi: 10.1007/s11704-013-3903-7. url: https://doi.org/10.1007/s11704-013-3903-7 (cit. on p. 9).

[25] Benoit Chesneau. *Gunicorn*. https://gunicorn.org/. Accessed: 15/09/2022. 2010 (cit. on p. 22).

[26] Prem Chhetri, Tim Butcher, and Brian Corbitt. "Characterising spatial logistics employment clusters". In: *International Journal of Physical Distribution & Logistics Management* 44.3 (2014-04), pp. 221–241. doi: 10.1108/ijpdlm-03-2012-0086. url: https://doi.org/10.1108/ijpdlm-03-2012-0086 (cit. on p. 15).

[27] COIN-OR. *COIN-OR Branch and Cut*. https://projects.coin-or.org/BC. Accessed: 01/09/2022. 2000 (cit. on p. 21).

[28] OpenStreetMap contributors. *OpenStreetMap*. https://www.openstreetmap.org/. Accessed: 05/05/2022. 2017 (cit. on p. 21).

[29] Paulo Cortez. *Modern Optimization with R*. Springer International Publishing, 2021. doi: 10.1007/978-3-030-72819-9. url: https://doi.org/10.1007/978-3-030-72819-9 (cit. on pp. 12, 47).

[30] Takao Dantsuji, Daisuke Fukuda, and Nan Zheng. "Simulation-based joint optimization framework for congestion mitigation in multimodal urban network: a macroscopic approach". In: *Transportation* 48.2 (2019-12), pp. 673–697. doi: 10.1007/s11116-019-10074-y. url: https://doi.org/10.1007/s11116-019-10074-y (cit. on p. 15).

[31] G. B. Dantzig and J. H. Ramser. "The Truck Dispatching Problem". In: *Management Science* 6.1 (1959-10), pp. 80–91. doi: 10.1287/mnsc.6.1.80. url: https://doi.org/10.1287/mnsc.6.1.80 (cit. on p. 14).

[32] Yuri Demchenko, Paola Grosso, Cees de Laat, and Peter Membrey. "Addressing big data issues in Scientific Data Infrastructure". In: *2013 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, 2013-05. doi: 10.1109/cts.2013.6567203. url: https://doi.org/10.1109/cts.2013.6567203 (cit. on p. 9).

[33] Emrah Demir, Tolga Bektaş, and Gilbert Laporte. "A review of recent research on green road freight transportation". In: *European Journal of Operational Research* 237.3 (2014), pp. 775–793. issn: 0377-2217. doi: https://doi.org/10.1016/j.ejor.2013.12.033. url: https://www.sciencedirect.com/science/article/pii/S0377221713010175 (cit. on p. 24).

[34] Steven Diamond and Stephen Boyd. "CVXPY: A Python-embedded modeling language for convex optimization". In: *Journal of Machine Learning Research* 17.83 (2016), pp. 1–5 (cit. on p. 20).

[35] Pedro Domingos. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. USA: Basic Books, Inc., 2018. isbn: 0465094279 (cit. on p. 11).

[36] Raphaëlle Ducret, Bernard Lemarié, and Alain Roset. "Cluster Analysis and Spatial Modeling for Urban Freight. Identifying Homogeneous Urban Zones Based on Urban Form and Logistics Characteristics". In: *Transportation Research Procedia* 12 (2016), pp. 301–313. doi: 10.1016

/j.trpro.2016.02.067. url: https://doi.org/10.1016/j.trpro.2016.02.067 (cit. on p. 14).

[37] Raphaëlle Ducret, Bernard Lemarié, and Alain Roset. "Cluster Analysis and Spatial Modeling for Urban Freight. Identifying Homogeneous Urban Zones Based on Urban Form and Logistics Characteristics". In: *Transportation Research Procedia* 12 (2016), pp. 301–313. doi: 10.1016/j.trpro.2016.02.067. url: https://doi.org/10.1016/j.trpro.2016.02.067 (cit. on p. 15).

[38] Nada Elgendy and Ahmed Elragal. "Big Data Analytics: A Literature Review Paper". In: *Advances in Data Mining. Applications and Theoretical Aspects*. Springer International Publishing, 2014, pp. 214–227. doi: 10.1007/978-3-319-08976-8_16. url: https://doi.org/10.1007/978-3-319-08976-8_16 (cit. on p. 9).

[39] Saeed Fazayeli, Alireza Eydi, and Isa Nakhai Kamalabadi. "Location-routing problem in multimodal transportation network with time windows and fuzzy demands: Presenting a two-part genetic algorithm". In: *Computers & Industrial Engineering* 119 (2018-05), pp. 233–246. doi: 10.1016/j.cie.2018.03.041. url: https://doi.org/10.1016/j.cie.2018.03.041 (cit. on p. 15).

[40] N G Gavrilenko. "Algorithm for regional clustering of road freight transport in the Russian Federation". In: *IOP Conference Series: Materials Science and Engineering* 953.1 (2020-11), p. 012081. doi: 10.1088/1757-899x/953/1/012081. url: https://doi.org/10.1088/1757-899x/953/1/012081 (cit. on pp. 14, 15).

[41] Sotiris P. Gayialis, Evripidis P. Kechagias, and Grigorios D. Konstantakopoulos. "A city logistics system for freight transportation: integrating information technology and operational research". In: *Operational Research* (2022-03). doi: 10.1007/s12351-022-00695-0. url: https://doi.org/10.1007/s12351-022-00695-0 (cit. on p. 15).

[42] Sean Gillies. *Shapely: manipulation and analysis of geometric objects*. toblerity.org, 2007–. url: https://github.com/Toblerity/Shapely (cit. on pp. 20, 43).

[43] Dominik Goeke. "Granular tabu search for the pickup and delivery problem with time windows and electric vehicles". In: *European Journal of Operational Research* 278.3 (2019-11), pp. 821–836. doi: 10.1016/j.ejor.2019.05.010. url: https://doi.org/10.1016/j.ejor.2019.05.010 (cit. on p. 15).

[44] Asvin Goel, Thibaut Vidal, and Adrianus Leendert Kok. "To team up or not: single versus team driving in European road freight transport". In: *Flexible Services and Manufacturing Journal* 33.4 (2020-12), pp. 879–913. doi: 10.1007/s10696-020-09398-0. url: https://doi.org/10.1007/s10696-020-09398-0 (cit. on p. 15).

[45] Ronaldo Goldschmidt, Emmanuel Passos, and Eduardo Bezerra. *Data Mining*. 2015-05. isbn: 9788535278231 (cit. on p. 9).

[46] Axel Grimault, Nathalie Bostel, and Fabien Lehuédé. "An adaptive large neighborhood search for the full truckload pickup and delivery problem with resource synchronization". In: *Computers & Operations Research* 88 (2017-12), pp. 1–14. doi: 10.1016/j.cor.2017.06.012. url: https://doi.org/10.1016/j.cor.2017.06.012 (cit. on p. 14).

[47] Miguel Grinberg. *Flask Web Development - Developing Web Applications with Python*. O'Reilly, 2014. isbn: 978-1-4493-7262-0. url: http://shop.oreilly.com/product/0636920031116.do (cit. on p. 22).

[48] PTV Group. *PTV Developer: A powerful set of APIs for transport planning*. https://developer.myptv.com/. Accessed: 09/08/2022. 2020 (cit. on p. 19).

[49] Ashkan Hafezalkotob and Arian Hafezalkotob. "A novel approach for combination of individual and group decisions based on fuzzy best-worst method". In: *Applied Soft Computing* 59 (2017-10), pp. 316–325. doi: 10.1016/j.asoc.2017.05.036. url: https://doi.org/10.1016/j.asoc.2017.05.036 (cit. on p. 12).

[50] Abid Haleem, Mohd. Javaid, Ibrahim Haleem Khan, and Raju Vaishya. "Significant Applications of Big Data in COVID-19 Pandemic". In: *Indian Journal of Orthopaedics* 54.4 (2020-05), pp. 526–528. doi: 10.1007/s43465-020-00129-z. url: https://doi.org/10.1007/s43465-020-00129-z (cit. on p. 8).

[51] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques*. 2012. doi: 10.1016/C2009-0-61819-5 (cit. on p. 9).

[52] David Hand, Heikki Mannila, and Padhraic Smyth. *Principles of Data Mining Cambridge*. Vol. 2001. 2001, pp. 175–187. isbn: 026208290X. url: http://link.springer.com/10.1007/978-1-4471-4884-5 (cit. on p. 9).

[53] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. "Array programming with NumPy". In: *Nature* 585.7825 (2020-09), pp. 357–362. doi: 10.1038/s41586-020-2649-2. url: https://doi.org/10.1038/s41586-020-2649-2 (cit. on p. 19).

[54] Richard Herschel and Virginia M. Miori. "Ethics & Big Data". In: *Technology in Society* 49 (2017-05), pp. 31–36. doi: 10.1016/j.techsoc.2017.03.003. url: https://doi.org/10.1016/j.techsoc.2017.03.003 (cit. on p. 8).

[55] Seyedmohsen Hosseini and Abdullah Al Khaled. "Freight flow optimization to evaluate the criticality of intermodal surface transportation system infrastructures". In: *Computers & Industrial Engineering* 159 (2021-09), p. 107522. doi: 10.1016/j.cie.2021.107522. url: https://doi.org/10.1016/j.cie.2021.107522 (cit. on p. 15).

[56] Plotly Technologies Inc. *Collaborative data science*. https://plotly.com/. Accessed: 15/09/2022. 2015 (cit. on p. 21).

[57] Postman Inc. *Postman*. https://www.postman.com/. Accessed: 15/09/2022. 2012 (cit. on p. 22).

[58] Jiehui Jiang, Dezhi Zhang, Qiang Meng, and Yajie Liu. "Regional multimodal logistics network design considering demand uncertainty and CO2 emission reduction target: A system-optimization approach". In: *Journal of Cleaner Production* 248 (2020-03), p. 119304. doi: 10.1016/j.jclepro.2019.119304. url: https://doi.org/10.1016/j.jclepro.2019.119304 (cit. on pp. 15, 17).

[59] Eva Kostikov, Petra Jilkova, and Pavla Kotatkova Stranska. "Optimization of e-commerce distribution center location". In: *Marketing and Management of Innovations* 5.2 (2021), pp. 166–178. doi: 10.21272/mmi.2021.2-14. url: https://doi.org/10.21272/mmi.2021.2-14 (cit. on p. 15).

[60] Jae-Gil Lee and Minseo Kang. "Geospatial Big Data: Challenges and Opportunities". In: *Big Data Research* 2.2 (2015-06), pp. 74–81. doi: 10.1016/j.bdr.2015.01.003. url: https://doi.org/10.1016/j.bdr.2015.01.003 (cit. on p. 40).

[61] Tomislav Letnik, Alessandro Farina, Matej Mencinger, Marino Lupi, and Stane Božičnik. "Dynamic management of loading bays for energy efficient urban freight deliveries". In: *Energy* 159 (2018-09), pp. 916–928. doi: 10.1016/j.energy.2018.06.125. url: https://doi.org/10.1016/j.energy.2018.06.125 (cit. on p. 16).

[62] BiaoJun Li, Chuantao Yao, Feng Zheng, Lin Wang, Jiashui Dai, and Quanzhou Xiang. "Intelligent Decision Support System for Business Forecasting Using Artificial Intelligence". In: *Arabian Journal for Science and Engineering* (2021-07). doi: 10.1007/s13369-021-05886-z. url: https://doi.org/10.1007/s13369-021-05886-z (cit. on p. 12).

[63] Vladimir Lifschitz. "John McCarthy (1927–2011)". In: *Nature* 480.7375 (2011-11), pp. 40–40. doi: 10.1038/480040a. url: https://doi.org/10.1038/480040a (cit. on p. 10).

[64] Ding Luo, Oded Cats, and Hans van Lint. "Constructing Transit Origin - Destination Matrices with Spatial Clustering". In: *Transportation Research Record: Journal of the Transportation Research Board* 2652.1 (2017-01), pp. 39–49. doi: 10.3141/2652-05. url: https://doi.org/10.3141/2652-05 (cit. on p. 15).

[65] Dennis Luxen and Christian Vetter. "Real-time routing with OpenStreetMap data". In: *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. GIS '11. Chicago, Illinois: ACM, 2011, pp. 513–516. isbn: 978-1-4503-1031-4. doi: 10.1145/2093973.2094062. url: http://doi.acm.org/10.1145/2093973.2094062 (cit. on pp. 21, 45).

[66] Bowen Lv, Bin Yang, Xiaolin Zhu, and Jun Li. “Operational optimization of transit consolidation in multimodal transport”. In: *Computers & Industrial Engineering* 129 (2019-03), pp. 454–464. doi: 10.1016/j.cie.2019.02.001. url: https://doi.org/10.1016/j.cie.2019.02.001 (cit. on pp. 15, 17).

[67] Zichong Lyu, Dirk Pons, Yilei Zhang, and Zuzhen Ji. “Freight Operations Modelling for Urban Delivery and Pickup with Flexible Routing: Cluster Transport Modelling Incorporating Discrete-Event Simulation and GIS”. In: *Infrastructures* 6.12 (2021-12), p. 180. doi: 10.3390/infrastructures6120180. url: https://doi.org/10.3390/infrastructures6120180 (cit. on p. 14).

[68] Changxi Ma, Wei Hao, Fuquan Pan, and Wang Xiang. “Road screening and distribution route multi-objective robust optimization for hazardous materials based on neural network and genetic algorithm”. In: *PLOS ONE* 13.6 (2018-06). Ed. by Xiaosong Hu, e0198931. doi: 10.1371/journal.pone.0198931. url: https://doi.org/10.1371/journal.pone.0198931 (cit. on pp. 14, 15).

[69] Hagar Mahmoud and Nadine Akkari. “Shortest Path Calculation: A Comparative Study for Location-Based Recommender System”. In: *2016 World Symposium on Computer Applications & Research (WSCAR)*. 2016, pp. 1–5. doi: 10.1109/WSCAR.2016.16 (cit. on pp. 44, 45).

[70] Mapbox. *Mapbox*. https://www.mapbox.com/. Accessed: 15/08/2022. 2019 (cit. on p. 21).

[71] Silvano Martello, David Pisinger, and Daniele Vigo. “The Three-Dimensional Bin Packing Problem”. In: *Operations Research* 48.2 (2000-04), pp. 256–267. doi: 10.1287/opre.48.2.256.12386. url: https://doi.org/10.1287/opre.48.2.256.12386 (cit. on p. 16).

[72] Chris McDonough. *Waitress*. https://docs.pylonsproject.org/projects/waitress/en/stable/. Accessed: 15/09/2022. 2010 (cit. on p. 22).

[73] Wes McKinney. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56–61. doi: 10.25080/Majora-92bf1922-00a (cit. on p. 19).

[74] Dirk Merkel. “Docker: lightweight linux containers for consistent development and deployment”. In: *Linux journal* 2014.239 (2014), p. 2 (cit. on p. 22).

[75] Rodrigo Mesa-Arango and Satish V. Ukkusuri. “Demand clustering in freight logistics networks”. In: *Transportation Research Part E: Logistics and Transportation Review* 81 (2015-09), pp. 36–51. doi: 10.1016/j.tre.2015.06.002. url: https://doi.org/10.1016/j.tre.2015.06.002 (cit. on p. 16).

[76] Michal Migurski. *Geopy: geocoding library*. geopy.org, 2007 (cit. on p. 20).

[77] PTV Group: Mobility and Transportation Software. *PTV Route Optimiser*. PTV Group: Mobility and Transportation Software. url: https://www.myptv.com/en/logistics-software/ptv-route-optimiser (cit. on p. 25).

[78] J.A. Morente-Molinera, G. Kou, I.J. Pérez, K. Samuylov, A. Selamat, and E. Herrera-Viedma. "A group decision making support system for the Web: How to work in environments with a high number of participants and alternatives". In: *Applied Soft Computing* 68 (2018-07), pp. 191–201. doi: 10.1016/j.asoc.2018.03.047. url: https://doi.org/10.1016/j.asoc.2018.03.047 (cit. on p. 12).

[79] Matthew North. *Data Mining for the Masses*. 2012, p. 264. isbn: 9780615684376. url: http://1xltkxylmzx3z8gd647akcdvov.wpengine.netdna-cdn.com/wp-content/uploads/2013/10/DataMiningForTheMasses.pdf%5Cnhttps://sites.google.com/site/dataminingforthemasses/ (cit. on p. 9).

[80] Taiwo Oladipupo. "Machine Learning Overview". In: *New Advances in Machine Learning*. InTech, 2010-02. doi: 10.5772/9374. url: https://doi.org/10.5772/9374 (cit. on p. 11).

[81] OpenStreetMap contributors. *Planet dump retrieved from https://planet.osm.org*. https://www.openstreetmap.org. 2017. (Visited on 2022-07-04) (cit. on p. 43).

[82] Shenle Pan, Eric Ballot, and Frédéric Fontane. "The reduction of greenhouse gas emissions from freight transport by pooling supply chains". In: *International Journal of Production Economics* 143.1 (2013), pp. 86–94. issn: 0925-5273. doi: https://doi.org/10.1016/j.ijpe.2010.10.023. url: https://www.sciencedirect.com/science/article/pii/S0925527310004160 (cit. on p. 24).

[83] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on p. 20).

[84] Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger, and Samir Chatterjee. "A Design Science Research Methodology for Information Systems Research". In: *Journal of Management Information Systems* 24.3 (2007-12), pp. 45–77. doi: 10.2753/mis0742-1222240302. url: https://doi.org/10.2753/mis0742-1222240302 (cit. on p. 4).

[85] Laurent Perron and Vincent Furnon. *OR-Tools*. Version 9.3. Google. url: https://developers.google.com/optimization/ (cit. on p. 20).

[86] Gloria Phillips-Wren. "Intelligent Decision Support Systems". In: *Multicriteria Decision Aid and Artificial Intelligence*. John Wiley & Sons, Ltd, 2013-02, pp. 25–44. doi: 10.1002/9781118522516.ch2. url: https://doi.org/10.1002/9781118522516.ch2 (cit. on p. 12).

[87] Python-Visualization. *Folium*. https://python-visualization.github.io/folium/. Version 0.11.0. 2020-12-28. url: https://python-visualization.github.io/folium/ (cit. on p. 45).

[88] Sudharsan Ravichandiran. *Hands-On Reinforcement Learning with Python*. Birmingham, England: Packt Publishing, 2018-06 (cit. on p. 11).

[89] D. Rhind, T. Connolly, Calif.) Environmental Systems Research Institute (Redlands, and Birkbeck College. *Understanding GIS: The ARC/INFO Method*. Self-study workbook. Environmental Systems Research Institute, 1990. isbn: 9781879102002. url: https://books.google.pt/books?id=2EfeAQAACAAJ (cit. on p. 12).

[90] John Robinson. "Likert Scale". In: *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht: Springer Netherlands, 2014, pp. 3620–3621. isbn: 978-94-007-0753-5. doi: 10.1007/978-94-007-0753-5_1654. url: https://doi.org/10.1007/978-94-007-0753-5_1654 (cit. on p. 28).

[91] G. van Rossum. *Python tutorial*. Tech. rep. CS-R9526. Amsterdam: Centrum voor Wiskunde en Informatica (CWI), 1995-05 (cit. on p. 19).

[92] Mandeep Kaur Saggi and Sushma Jain. "A survey towards an integration of big data analytics to big insights for value-creation". In: *Information Processing Management* 54.5 (2018-09), pp. 758–790. doi: 10.1016/j.ipm.2018.01.010. url: https://doi.org/10.1016/j.ipm.2018.01.010 (cit. on p. 9).

[93] Samir Saidi and Sami Hammami. "Modeling the causal linkages between transport, economic growth and environmental degradation for 75 countries". In: *Transportation Research Part D: Transport and Environment* 53 (2017), pp. 415–427. issn: 1361-9209. doi: https://doi.org/10.1016/j.trd.2017.04.031. url: https://www.sciencedirect.com/science/article/pii/S1361920916307696 (cit. on p. 13).

[94] Arthur L. Samuel. "Some Studies in Machine Learning Using the Game of Checkers". In: *IBM J. Res. Dev.* 3.3 (1959), pp. 210–229. doi: 10.1147/rd.33.0210. url: https://doi.org/10.1147/rd.33.0210 (cit. on p. 10).

[95] Krishna Shingala. "JSON Web Token (JWT) based client authentication in Message Queuing Telemetry Transport (MQTT)". In: *CoRR* abs/1903.02895 (2019). arXiv: 1903.02895. url: http://arxiv.org/abs/1903.02895 (cit. on p. 22).

[96] Gangadhar Shobha and Shanta Rangaswamy. "Machine Learning". In: *Handbook of Statistics*. Elsevier, 2018, pp. 197–228. doi: 10.1016/bs.host.2018.07.004. url: https://doi.org/10.1016/bs.host.2018.07.004 (cit. on p. 11).

[97] Hugo Silva Carvalho, André Pilastri, Luís Miguel Matos, Arthur Matta, Rui Novais, and Paulo Cortez. "An Intelligent Decision Support System for Road Freight Transport". In: *Intelligent Data Engineering and Automated Learning – IDEAL 2022*. Cham: Springer International Publishing, 2022 (cit. on p. 46).

[98] Hannah Snyder. "Literature review as a research methodology: An overview and guidelines". In: *Journal of Business Research* 104 (2019-11), pp. 333–339. doi: 10.1016/j.jbusres.2019.07.039. url: https://doi.org/10.1016/j.jbusres.2019.07.039 (cit. on p. 7).

[99] Smartbear Software. *Swagger*. https://swagger.io/. Accessed: 15/09/2022. 2011 (cit. on p. 22).

[100] Shi-Yi Tan and Wei-Chang Yeh. "The Vehicle Routing Problem: State-of-the-Art Classification and Review". In: *Applied Sciences* 11.21 (2021). issn: 2076-3417. doi: 10.3390/app112110295. url: https://www.mdpi.com/2076-3417/11/21/10295 (cit. on p. 46).

[101] Lóránt Tavasszy and Gerard de Jong. "Introduction". In: *Modelling Freight Transport*. Elsevier, 2014, pp. 1–12. doi: 10.1016/b978-0-12-410400-6.00001-x. url: https://doi.org/10.1016/b978-0-12-410400-6.00001-x (cit. on p. 13).

[102] Christian Tilk and Asvin Goel. "Bidirectional labeling for solving vehicle routing and truck driver scheduling problems". In: *European Journal of Operational Research* 283.1 (2020), pp. 108–124 (cit. on pp. 15, 18).

[103] Erfan Tirkolaee, Ali Hosseinabadi, Mehdi Soltani, Arun Sangaiah, and Jin Wang. "A Hybrid Genetic Algorithm for Multi-Trip Green Capacitated Arc Routing Problem in the Scope of Urban Services". In: *Sustainability* 10.5 (2018-04), p. 1366. doi: 10.3390/su10051366. url: https://doi.org/10.3390/su10051366 (cit. on p. 15).

[104] Ming-Hsiang Tsou. "Research challenges and opportunities in mapping social media and Big Data". In: *Cartography and Geographic Information Science* 42.sup1 (2015), pp. 70–74. doi: 10.1080/15230406.2015.1059251. url: https://doi.org/10.1080/15230406.2015.1059251 (cit. on p. 8).

[105] A. M. Turing. "I.—COMPUTING MACHINERY AND INTELLIGENCE". In: *Mind* LIX.236 (1950-10), pp. 433–460. doi: 10.1093/mind/lix.236.433. url: https://doi.org/10.1093/mind/lix.236.433 (cit. on p. 10).

[106] Uber. *Kepler.gl*. https://kepler.gl/. Accessed: 15/08/2022. 2020 (cit. on p. 21).

[107] Jirapha Udomwannakhet, Pimolpatchara Vajarodaya, Saengkarn Manicho, Kwanjira Kaewfak, Josep Blay Ruiz, and Veeris Ammarapala. "A review of multimodal transportation optimization model". In: *2018 5th International Conference on Business and Industrial Research (ICBIR)*. 2018, pp. 333–338. doi: 10.1109/ICBIR.2018.8391217 (cit. on p. 14).

[108] Jirapha Udomwannakhet, Pimolpatchara Vajarodaya, Saengkarn Manicho, Kwanjira Kaewfak, Josep Blay Ruiz, and Veeris Ammarapala. "A review of multimodal transportation optimization model". In: *2018 5th International Conference on Business and Industrial Research (ICBIR)*. IEEE, 2018-05. doi: 10.1109/icbir.2018.8391217. url: https://doi.org/10.1109/icbir.2018.8391217 (cit. on p. 17).

[109] Viswanath Venkatesh and Hillol Bala. "Technology Acceptance Model 3 and a Research Agenda on Interventions". In: *Decis. Sci.* 39.2 (2008), pp. 273–315. doi: 10.1111/j.1540-5915.2008.00192.x. url: https://doi.org/10.1111/j.1540-5915.2008.00192.x (cit. on p. 25).

[110] Thibaut Vidal, Gilbert Laporte, and Piotr Matl. "A concise guide to existing and emerging vehicle routing problem variants". In: *European Journal of Operational Research* 286.2 (2020-10), pp. 401–416. doi: 10.1016/j.ejor.2019.10.010. url: https://doi.org/10.1016/j.ejor.2019.10.010 (cit. on p. 15).

[111] Jiahai Wang, Yuyan Sun, Zizhen Zhang, and Shangce Gao. "Solving multitrip pickup and delivery problem with time windows and manpower planning using multiobjective algorithms". In: *IEEE/-CAA Journal of Automatica Sinica* 7.4 (2020), pp. 1134–1153 (cit. on p. 15).

[112] Dong Yang, Yaohua Wu, and Wenkai Ma. "Optimization of storage location assignment in automated warehouse". In: *Microprocessors and Microsystems* 80 (2021-02), p. 103356. doi: 10.1016/j.micpro.2020.103356. url: https://doi.org/10.1016/j.micpro.2020.103356 (cit. on p. 15).

[113] Toshiyuki Yokota and Homer Pagkalinawan. "Application of geographic information systems in impact evaluation and geospatial portfolio analysis of transport projects". en. In: *Transp. Res. Rec.* (2022-06), p. 036119812210920 (cit. on p. 13).

[114] Meng You, Yiyong Xiao, Siyue Zhang, Pei Yang, and Shenghan Zhou. "Optimal mathematical programming for the warehouse location problem with Euclidean distance linearization". In: *Computers & Industrial Engineering* 136 (2019-10), pp. 70–79. doi: 10.1016/j.cie.2019.07.020. url: https://doi.org/10.1016/j.cie.2019.07.020 (cit. on p. 15).

[115] Nannan Yu, Martin De Jong, Servaas Storm, and Jianing Mi. "Transport Infrastructure, Spatial Clusters and Regional Economic Growth in China". In: *Transport Reviews* 32.1 (2012-01), pp. 3–28. doi: 10.1080/01441647.2011.603104. url: https://doi.org/10.1080/01441647.2011.603104 (cit. on p. 14).

[116] Ghazal Zakeri and Nils O. E. Olsson. "Investigating the effect of weather on punctuality of Norwegian railways: a case study of the Nordland Line". In: *Journal of Modern Transportation* 26.4 (2018-06), pp. 255–267. doi: 10.1007/s40534-018-0169-7. url: https://doi.org/10.1007/s40534-018-0169-7 (cit. on p. 15).

[117] Fotios Zantalis, Grigorios Koulouras, Sotiris Karabetsos, and Dionisis Kandris. "A Review of Machine Learning and IoT in Smart Transportation". In: *Future Internet* 11.4 (2019-04), p. 94. doi: 10.3390/fi11040094. url: https://doi.org/10.3390/fi11040094 (cit. on p. 14).

[118] Mo Zhang and Qian Hu. “An effective evolutionary algorithm for the biobjective full truckload transportation service procurement problem”. In: *Computers & Industrial Engineering* 127 (2019-01), pp. 1012–1023. doi: 10.1016/j.cie.2018.11.036. url: https://doi.org/10.1016/j.cie.2018.11.036 (cit. on p. 14).

[119] Zahra Ziaei and Armin Jabbarzadeh. “A multi-objective robust optimization approach for green location-routing planning of multi-modal transportation systems under uncertainty”. In: *Journal of Cleaner Production* 291 (2021-04), p. 125293. doi: 10.1016/j.jclepro.2020.125293. url: https://doi.org/10.1016/j.jclepro.2020.125293 (cit. on pp. 15, 17).

A

RanCoord: User Manual

A.1 Overview

RanCoord is an open-source Python package that allows quickly generating random coordinates within a set of geographic boundaries. The framework offers methods to generate, save and plot a determined number of random coordinates generated from a location polygon or a simple name/address through its geocoding. Such geographic methods are well-suited for research and development activities in transport and logistics, such as the TSP and the VRP.

Figure 16: RanCoord Python package logo.

A.2 Installing

Using pip:

Listing A.1: Installing RanCoord directly from pip.

```
pip install rancoord
```

A.3 Importing

To import the package, simply use the following command:

Listing A.2: Importing RanCoord.

```
import rancoord as rc
```

A.4 Quick Start

RanCoord, as mentioned before, allows generating random coordinates within a set of geographic boundaries. However, it provides a set of extra functionalities that any researcher or developer can use when manipulating geographic data. Some of its main features are revealed in the following short guide:

Listing A.3: RanCoord quick start execution.

```
# Get the geographic bounding box of the city of London
bounding_box = rc.nominatim_geocoder(address = London, United Kingdom)

# Convert the bounding box to a geographic polygon
geographic_polygon = rc.polygon_from_boundingbox(
    boundingbox = bounding_box
)

# Gererate the random geographic coordinates and distance
# matrix of the polygon
lat, lon, distance_matrix = rc.coordinates_randomizer(
    polygon = geographic_polygon,
    num_locations = 50,
    plot = True,
    save = True,
    matrix = True
)

# Save the geographic coordinates in a different format
# from default JSON (.csv, .txt, .xlsx)

rc.multiple_formats_saver(
    lat = lat,
    lon = lon,
    columns = ['latitude', 'longitude'],
    file_format = 'csv',
    file_name = 'csv_coordinates',
    dir_name = 'coordinates'
)
```

```
## Generate the distance matrix using a different distance
## calculation algorithm from default Vincenty's (OSRM, Haversine)

rc.osrm_matrix(
    lat = lat,
    lon = lon,
    method = 'osrm'
)
```

A.5 Outputs

The following content presents the outputs of the **RanCoord** package execution presented in Section A.4.

Listing A.4: RanCoord execution outputs.

```
# Get the first 5 elements of latitudes
lat[:5]

[51.34884228427017,
 51.314982703733385,
 51.46469502323293,
 51.35482857606807,
 51.446114342064]

# Get the first 5 elements of longitudes
lon[:5]

[0.2673346385105909,
 -0.20491384181826444,
 0.18433999944047685,
 0.2359266566281253,
 0.18395932809122217]

# Get the first 5 elements of the auto distance matrix
# using Vincenty's method for distance calculation

distance_matrix[:5]

array([[ 0.    , 33.128, 14.124,  2.287, 12.28 , 38.731, 29.632, 39.984,
         2.876, 60.651, 36.74 , 28.226, 40.957, 14.685, 19.437, 18.681,
        13.689, 49.847, 31.444, 46.175, 48.022, 41.785, 22.789, 62.57 ,
        16.293, 12.463, 20.933, 38.687, 36.758, 47.154, 27.527, 33.268,
        37.488, 45.539, 41.599,  1.23 , 40.233, 40.39 , 49.901, 33.231,
        35.813, 53.639, 43.277, 49.787, 39.102, 24.735, 52.008, 28.636,
```

```
        37.864, 29.383],
       [33.128,  0.    , 31.805, 31.041, 30.755, 11.442, 10.615, 31.939,
        32.411, 41.679,  9.175, 19.059, 41.79 , 18.86 , 15.537, 15.443,
        20.348, 40.346, 10.527, 25.001, 14.944, 22.814, 10.411, 42.405,
        18.492, 30.053, 41.896, 27.183, 38.583, 22.577, 13.181, 36.393,
        29.619, 12.661, 28.102, 34.354, 10.726, 13.924, 35.322, 37.941,
         8.18 , 27.254, 19.225, 34.032, 12.652, 10.11 , 37.537, 36.777,
        51.298, 22.477],
       [14.124, 31.805,  0.    , 12.739,  2.067, 33.037, 24.158, 27.336,
        11.248, 49.145, 31.699, 19.001, 26.9  , 14.845, 16.344, 22.382,
        13.066, 36.73 , 25.741, 36.623, 45.302, 32.122, 22.259, 51.318,
        13.779,  2.368, 10.497, 27.143, 22.738, 38.781, 21.255, 19.268,
        25.119, 42.458, 30.193, 14.643, 35.259, 33.956, 37.906, 19.131,
        31.089, 45.345, 35.269, 38.115, 32.951, 21.725, 39.831, 14.513,
        25.068, 18.673],
       [ 2.287, 31.041, 12.739,  0.    , 10.781, 36.463, 27.353, 37.969,
         2.24 , 58.475, 34.485, 25.992, 39.356, 12.467, 17.167, 16.881,
        11.423, 47.889, 29.166, 43.913, 45.902, 39.532, 20.666, 60.374,
        14.013, 10.861, 20.493, 36.551, 35.134, 44.869, 25.24 , 31.639,
        35.444, 43.391, 39.444,  3.476, 37.988, 38.111, 47.782, 31.699,
        33.566, 51.352, 40.99 , 47.638, 36.826, 22.503, 49.906, 27.203,
        37.106, 27.222],
       [12.28 , 30.755,  2.067, 10.781,  0.    , 32.672, 23.639, 28.506,
         9.418, 50.058, 31.209, 19.156, 28.678, 13.227, 15.221, 20.644,
        11.444, 38.08 , 25.289, 37.037, 44.553, 32.533, 20.958, 52.173,
        12.447,  0.721, 12.165, 27.986, 24.484, 38.96 , 20.827, 20.997,
        26.19 , 41.746, 31.016, 12.881, 34.792, 33.731, 38.919, 20.959,
        30.539, 45.54 , 35.353, 39.041, 32.665, 20.753, 40.899, 16.423,
        27.124, 19.203]])
```

A.6 Directory and Files Structure

So far, the set of executions performed using the package should have created the following directory and file structure:

```
├── coordinates
│   ├── coordinates_date_time.csv
│   └── coordinates_date_time.json
└── maps
    └── map_date_time.html
```

UMinho | 2022 Hugo Silva Carvalho **An Intelligent Decision Support System for the Freight Transport Sector**