OPTIMISING THE THROUGHPUT OF A COMPLEX MULTIMODAL FREIGHT NETWORK

Denesh Reddy

A thesis submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy.

Johannesburg, 2025

Declaration

I declare that this thesis is my own, unaided work, except where otherwise acknowledged. It is being submitted for the Degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination to any other University.

Signed this 24th day of April 2025.

DReddy

DENESH REDDY

Abstract

This research, situated within the logistic domain, extends its relevance to global supply chains. In a context where revenues are constrained, customer demand for capacity expansion is rising, the commodity mix is evolving, and time sensitivity is critical, relying solely on critical infrastructure expansion for capacity development is neither sustainable nor desirable. Consequently, this study focuses on unlocking latent capacity within existing infrastructure and facilities.

Initially, a strategic area for growth was identified, enabling railways to increase their market share in the container sector. The transition of goods from road to rail is a significant economic lever, particularly as the global shift away from fossil fuels continues and railways remain an efficient mode of transport. While bulk commodities are predominantly transported by rail, the Fourth Industrial Revolution presents substantial opportunities in other sectors, notably in the increasing volume of goods suitable for containerization.

To address these challenges, a mathematical model was developed to serve as a theoretical foundation for an intermodal hub-and-spoke system, offering a framework for approaching the optimal solution proposed in this thesis. This work makes a novel contribution by addressing the hub network problem with considerations for train lengths and departure decisions. The research investigates whether cost optimisation can be achieved by optimising train lengths. Although the proposed mathematical model is general, it has been specifically applied to a South African context.

The optimisation process was carried out using the MATLAB® programming language, and customizing solutions for the specific problem, thereby representing another novel contribution to the field.

In conjunction, a Simulink® model of a conceptual hub-and-spoke system was developed as a further independent contribution, incorporating key elements to facilitate the study of operations and the impact of varying train lengths. The integration of MATLAB® and Simulink® allowed for solution techniques to be developed using built in optimisation toolboxes, separate from the solution technique developed for the mathematical model. Simulation scenarios were designed to cover a range of operational environments, including considerations for double-stacked container trains. While double stacking is not feasible on narrow-gauge tracks, it provides insight into the potential benefits of transitioning to wider gauges.

Results for the South African corridor problem indicate that train lengths can be optimised to achieve lower unit costs, for both the mathematical model and the simulation.

Dedication

To my dad, Vishnu Reddy, an academic visionary who encouraged me to start this journey.

To Tharini, Jaipriya and Prenayen, for whom I wish extraordinary paths.

To Saloshna, whose love, understanding and support gave me the freedom to explore.

To my mum, for her unwavering support.

Acknowledgements

I am grateful to the following people and organisations, without whom this work would not have been possible:

Prof Beatrys Lacquet, whose quiet leadership guided this study without restrictions,

Prof Montaz Ali, whose academic brilliance and enthusiasm inspired much of the "leave your comfort zone" thinking,

and Transnet, which sponsored this study and allowed the latitude for advanced research.

Table of Contents

List of figures

List of tables

Definitions

Container corridor	Logistic route developed where containers are the dominant traffic.
Destination	Centre where goods are consumed.
Double stack	Refers to arrangement of containers where they are stacked on a wagon two levels high.
feu	Forty-foot equivalent unit defined in the logistic industry for a container size
First mile	Section of journey (any mode) from the origin to a consolidation point, e.g., a hub.
Last mile	Section of journey (any mode) from the last hub to Destination.
Mainline	Section of railway line serving as a main thoroughfare, where other railway line flows are combined.
Nodes	Point in a network at which a facility may be located, may be defined as a hub.
O-D pair	Origin-Destination pair
Origin	Centre where goods are produced.
Rail gauge	Distance between the two rails on a railway line, in South Africa this is primarily Cape gauge at 1067 mm.
Rand	South African currency (R)
Single stack	Refers to arrangement of containers where they are stacked on a wagon one level high.
Slot	Space on a defined section of track which may accommodate one train; quantity indicates the number of trains per direction per day.
Structure gauge	Definition delineating boundaries for rolling stock and infrastructure in a vertical plane.

teu	Twenty-foot equivalent unit defined in the logistic industry for a container size.
Track gauge	Distance between tracks
NatCor	Natal Corridor - Logistic corridor linking Gauteng and the Port of Durban.
Origin1	Production point in the model
Origin2	Production point in the model
O1Rate	Production rate of containers at Origin1 in teu/h
O2Rate	Production rate of containers at Origin2 in teu/h
Train	Combination of locomotives and wagons, quantities required are based on operational, environmental and technical factors.

This page left intentionally blank.

Chapter One

Introduction

1.1 An overview

The global economy consists of centres of production and of consumption, where goods are produced and consumed to satisfy some requirement. As a result, multiple economies of various sizes have developed, some with local and some with broader influence, extending to a global scale. Logistic chains develop around these economies, becoming more concentrated along corridors as demand for the goods increases, (Agamez-Arias & Moyano-Fuentes, 2017).

Over time, volumes may increase to satisfy demand leading to the development of various modes as freight is transported with ever-increasing sophistication. The evolution of supply chains is driven by the volumes required, the urgency of this demand and the requirement for more cost-efficient supply chains, (Kurtulus & Cetin, 2020). Supply chains evolve further to work together globally.

A transport infrastructure develops around logistic chains. In practice, the infrastructure should be designed to meet objectives at a planning stage. The logistic chain may be measured in terms of how it delivers capacity, where the salient parameters are cost effectiveness, efficiency and on time deliveries, (Agamez-Arias & Moyano-Fuentes, 2017). This typically leads to a combination of rail, road and pipeline networks on land. This is known as intermodal transportation.

Any transport mode has advantages and disadvantages. They are selected based on the operational and environmental requirements, (Bruns, et al., 2014).

Whereas road transport is used for the first and last miles of the logistic chain, multimodal networks have evolved to using short rail transport for these journeys. Ideally, rail transport would be used for the long journeys in-between.

The nature of the railway industry lends itself to consistent, large volumes, whereas the road trucking industry can adapt to just-in-time requirements. Rail services are inflexible and require large volumes as they have high fixed costs to contend with. This results in trucks being more competitive in an economy requiring agility. However, rail has lower externality costs, meaning it has less impact on all aspects of the environment, (Woroniuk & Marinov, 2013). Rail is also cheaper to operate than road transport vehicles (as much as 11 times per ton-km), (Kuo & Miller-Hooks, 2012).

From the foregoing discussion, it appears that the logistic environment is evolving. While there are both bulk commodities on rail and smaller parcel quantities on road, these modalities will not be sustainable for the long term. However, there is an opportunity to combine rail and road modes to complement each other.

Intermodal networks have developed because road is flexible in terms of reaching customer locations and incurs minimal start-up costs. Rail transport, on the other hand, is inflexible and has high infrastructure start-up costs. However, rail is far more cost effective when large quantities are transported over long distances. Therefore, this has led to the development of an intermodal hub and spoke model, combining rail and road to employ their respective strengths.

Based on the existing infrastructure and demand for goods, an ideal logistic chain must exist. The movements in that chain should be configured for the lowest cost, while realising a demand volume throughput. In this process, the interaction with the other parts of the network must be considered, as changes in any part of the network can have an influence across modal boundaries, (Agamez-Arias & Moyano-Fuentes, 2017).

Containers are suited to intermodal networks as they allow for the transfer between modes without handling of the goods themselves. Where transport is required for bulk liquids, pipelines would be considered (bulk liquids are not considered any further in this study).

In the field of Operation Research problems in transportation, there have been various location models developed. One such area of research revolves around the Hub Location Problem. Alumer & Kara (2008) surveyed the various formulations developed to study hubs and their purpose. Hub location models can be classified as the p-hub median problem, hub location problem with fixed costs, p-hub centre problem and hub covering problems.

Campbell (1994) presented integer programming formulations for four types of discrete hub location problems. Hubs are facilities that are used for transhipment and switching points in transportation networks. These networks may have large numbers of origins and destinations.

In Campbell (1996), the history of the p-hub median problem was analysed since its development in 1965. Most research had focussed on single hub allocations, whereas Campbell (1996) had an interest in multiple hub allocations, where any node could be allocated to multiple hubs. Various conditions were found for the treatment of hubs.

Li, et al. (2013) presented a general framework to model intermodal transport networks using a linear programming model. A generic model for container flows in nodes and links, considering time dependencies, was formulated. Li, et al. (2013) identified the requirement to expand transport services using existing infrastructure. Technology levers were used to expand and build rail market share, thus reducing the need for capital expenditure. A framework was designed for a model to accommodate rail, road and waterways. Containerised freight could be moved between the transportation modes and stored in the nodes. This model considered dynamic behaviour, including loading, unloading and storing containers. Capacity constraints were modelled for the links and nodes.

Serper & Alumur (2016) built a model to facilitate the design of hub networks. This model allowed different vehicle types and alternative transportation modes. The minimum total cost was studied, while using vehicles of different sizes to achieve economies of scale. Serper & Alumur (2016) considered trucks of different sizes to optimise the costs of transportation. This concept focussed on individual trucks of varying sizes.

Using rail and road as complementary modes would be more economical and efficient in moving significant volumes of goods than a pure road-based system, (Crainic & Kim, 2007). Some transport requirements are more suited to road. In this case, there should not be any competition. In cases where rail and road can complement each other with their respective strengths, they should be run as complementary services, as an intermodal system.

Train length is a significant lever in a railway, it affects how much volume the whole system can transport and the size of rolling stock fleets. The effects of varying train lengths are not a focus area of other studies in the logistic discipline, mathematical formulations incorporating various length trains in a hub and spoke system are not proposed for this purpose.

1.2 Nature and purpose of the study

This study was undertaken to examine the logistic chain and develop a mathematical formulation to minimise cost through the optimisation of train lengths in a hub and spoke network for container trains. The containers are moved from one train to another at the hub without handling of the goods within the container, this is regarded as a mode change, (Meng & Wang, 2011).

The mathematical formulation was tested first with a simple example, solved manually without optimisation.

A decision was taken to use a Genetic Algorithm (GA) to solve the problem as it was unique in its structure, and GA's may be applied where the functions are non-differentiable, non-continuous, comprise arbitrary types of variables, could have competing goals and may not even have any analytical expression. A bespoke GA was developed to optimise the cost using the model, a simple example was then solved to test the working of the model and the GA. A larger real-world problem was then approached, and the model was applied and solved using the GA to determine the train lengths. The focus was on developing the model, hence it was not deemed necessary to test the application on further examples.

Following this, another approach was also developed as random delays could not be handled by the mathematical formulation. The logistic system was modelled in a simulation and the solution was tested using the available software tools. The simulation uses a genetic algorithm from the Simulink® Analysis Toolbox to conduct the optimisation, separate from the solution developed for the mathematical formulation. The simulation allowed the introduction of real-world scenarios, such as weather delays.

This research considers improving the system capacity of an intermodal hub and spoke transport system, where at least one of the modes is rail. The interaction between the efficiencies of rail and the interface to the product origin (first mile) shall be studied. In optimising the interaction in this system, economies of scale can be achieved, thereby encouraging the customer to use it as intended.

1.3 Research objectives

It is proposed that running the longest possible trains does not yield the benefits which are expected. Instead, the operating cost of a multimodal system can be minimised by varying capacity usage to levels below the maximum of the individual component capacities. This represents a unique contribution in a hub and spoke system, where the train lengths can be varied in different sections of the journey.

In this research, the conditions shall be studied to establish how to minimise the cost in an intermodal system by varying the train lengths entering and leaving the hub. The effect on volume throughput and cost shall be analysed. This is stated as the following four research objectives:

i. Given the infrastructure of a hub and spoke system, develop a cost minimisation mathematical model considering train lengths and a cost objective with available constraints;
ii. Implement an algorithm to solve a corridor level variant of this model;
iii. Determine the minimum operating cost of the system and how this is influenced by train lengths; and
iv. Develop a simulation model to introduce real world complications to the problem.

1.3.1 Limitations of study

The study is constrained by the following limitations, which are not necessarily negative but serve to define the study:

i. The study considers a narrow-gauge railway system
ii. The boundaries of the study are limited to surface transport on land
iii. Optimising train lengths for the rail sections are considered, but not road traffic (although the model can accommodate this)

iv. Limited to container traffic
v. The model does not consider traction (where different locomotive types may be necessary) changes
vi. Handling equipment limitations are not considered
vii. Weather conditions and unpredictable events are not considered
viii. The simulation uses a genetic algorithm from the Simulink® Analysis Toolbox
ix. The simulation finds train lengths for the different sections and runs these train lengths for the duration of a simulation period

1.4 Organisation of the thesis

After the current introductory chapter, the thesis' chapters consist of a literature review in Chapter 2, followed by an overview of the logistic system in Chapter 3, and the development of the mathematical model in Chapter 4. The description of a corridor level problem is presented in Chapter 5. Numerical results for the mathematical model are presented in Chapter 6. The design of the simulation is presented in Chapter 7 and the development of the same in Chapter 8. Simulation results are presented in Chapter 9. This is followed by a critical evaluation in the Discussion, Chapter 10. Chapter 11 follows with the Conclusion wrapping up the thesis. The thesis includes several appendices containing further detailed information not deemed necessary in the body of the text.

The thought process in the thesis is clearly presented as follows:

Literature review: The existing research background is presented. Chapter 2 presents the existing state of knowledge, highlighting the shortcomings required to be addressed in this work.

Logistic system overview: An overview of a logistic system is presented in Chapter 3, to aid the understanding of how a hub and spoke system works, presenting critical information for how the mathematical model and simulation are designed.

Model development: The development of the mathematical formulation is presented in Chapter 4, including the thinking behind how the model will work. Chapter 4 proceeds to discuss the operations strategy. The contribution of this chapter is a significant part of the novelty in this thesis. Here the development of the mathematical formulation to optimise costs based on train lengths is presented.

Conceptualisation of problem: Chapter 4 continues to present an example to show how the mathematical model will be applied and how the constraints are satisfied. Furthermore, some manual calculations are presented using the mathematical model.

South African Corridor problem: The model is extended and applied to a more complex variant of the problem in Chapter 5. Here a transport corridor is considered with real world data.

Numerical results: The numerical experiments are presented in Chapter 6. Preliminary results are presented concerning the simple example and the setup of the model, followed by the presentation of results for a study on a South African container corridor for a cost optimisation case, where the train lengths may be varied.

Simulation development: A simulation model has been developed, with the capability of including real world considerations. The detail of how the simulation was designed and how the optimisation was performed is given in Chapters 7 and 8. This begins with the parameters chosen and the setup thereof for the simulation runs, before moving on to present the process flow for the simulation model. The simulation results are then presented in Chapter 9.

Discussion: The learnings from this work are discussed in Chapter 10, showing the novel development and contribution to knowledge.

Conclusion: The contribution of this work is presented in Chapter 11, focussing on the novel developments leading to cost optimisation, specifically relating to variable train lengths in a hub and spoke system.

Chapter Two
Literature Review

2.1 Introduction

The state of the art of intermodal systems, hub location problems and transportation literature are reviewed here, focussing on the study of intermodal traffic. There is a significant body of work in operations research, pertaining to intermodal transport. In the main, these relate to the location of hubs, cost and capacity issues. After an online literature search, it was found that the problem of varying the train lengths presented an area where the current knowledge base could be expanded.

The literature review was conducted over the full period of the study, from preparing for the proposal, to after the thesis was prepared. There was a continuous search for further information as questions arose and nuances in the understanding of the work required further validation and investigation. The library resources of the University of the Witwatersrand were used to conduct the literature review over this period. All relevant literature was considered, irrespective of the time period, some of the studies went as far back as the 1960’s, however the bulk of the literature is from more recent times.

The research conducted thus far has been categorised into broad areas: intermodal studies, the evolving economic environment, network hub location problems, capacity studies, genetic algorithms and various applications thereof, and intermodal market share, which are critically reviewed in the following sections.

2.2 Intermodal studies

A survey of the field around operations research related to hub location problems are presented. An intermodal system refers to a transport environment where the transport of goods takes place in the same loading unit or vehicle by successive modes without handling of the goods themselves, (Anon., 2003), therefore the same mode could also be deployed before and after any required handling of the goods. This technique shall be developed and explored, where the operation can be modified for some benefit in an intermodal environment, but a mode change may not be required.

A diverse range of research has been conducted, giving a broad background to the field. Bontekoning, et al. (2004) conducted a survey of intermodal studies, they considered whether a new applied transportation research field was developing. Ninety-two publications were reviewed in this paper. They revealed how intermodal rail – truck problems were being investigated. Up to that time, intermodal transport was considered a competing mode to single mode transport. Through the 1980s and 1990s, the concept of intermodalism became an issue of government policy. The intermodal transport policy was driven by efficiency, coordination benefits, environmental and volume demand issues. Bontekoning, et al. (2004) concluded that intermodal research was a new field. It should be classified as a research field as it is unique and the problems are complex, requiring new knowledge to gain insight.

Bontekoning, et al. (2004) found that intermodal research was in a pre-paradigmatic phase, and expected to evolve into a stage of normal science over a short period of time. This paper framed the research field in its philosophical space and investigated how it was evolving. It provided insight into the various areas of research, demonstrating how researchers considered their topic in a philosophical sense. It presented the various directions that research in this field was focussed on, which academic journals were of significance and who the leading researchers were.

A further literature review was conducted by Agamez-Arias & Moyano-Fuentes (2017), focusing on intermodal research. The state of intermodal research was presented. It was categorised into three broad areas, being basic principles consisting of conceptualisation of transportation systems, improvements consisting of quality of service and economic efficiency, and mathematical modelling consisting of objective functions and constraints. The paper identified the direction of research in the field. It also formulated a classification for the field and found interactions between the various research initiatives. They identified the gaps in the research which they were not aware of at the time. A proposal for future research was made. This was valuable because it gave a sense of what issues existed.

Agamez-Arias & Moyano-Fuentes (2017) conducted research in the field of intermodal transport systems and the efficiency gains over single mode transport. The basic modelling objectives were to reduce transportation costs and delivery times. However, these parameters cannot be considered in isolation. They identified that a relationship exists between the elements in the logistic chain, being the products, infrastructure locations, transport modes and various other criteria. In addition, infrastructure location influences distance, freight flows and volumes. The loading unit itself also has its own set of characteristics, that are addressed by planning models.

Agamez-Arias & Moyano-Fuentes (2017) analysed the various approaches in intermodal transport. They found that there was a developing focus on research to optimise resources where activities were connected between different parts of a logistic network. They identified that by treating the whole logistic chain, this would obtain benefits across the boundaries of a single actor's control. This was a distinct new focus area. Crainic & Kim (2007) considered that if the companies participating in a logistic chain were to co-operate on planning synchronisation, this would have benefits for the overall efficiency of the system. This is of significant interest as it relates to the current study, as it identified that improvements in one area can have interactions and benefits over the whole system. This is the principle being developed and exploited in this thesis.

Intermodal transport provides the system or levers to transport freight between two points economically, and which can also be feasible from an operational perspective, (Meng & Wang, 2011). These modes may be combined in many ways, but these combinations are to a large part dependent on the geographic layout, the distances and the type of infrastructure, (Hayuth, 1994).

2.3 Evolving economic environment

The evolving economic environment around the world threatens existing rail business models necessitating rail to evolve if it is to remain sustainable.

It is noted that the economy is evolving and changing in nature with the development of e-commerce and economic globalization that have taken hold in the past few decades, (Kuo & Miller-Hooks, 2012).

Kuo & Miller-Hooks (2012) conducted work in the operational environment, i.e., where daily reaction to customer requirements is required. A model was developed for collaboration between competing railway companies (not shown because the model was not of interest, however the difference in approaching operational and tactical decision making was). Where synergy existed between origins, destinations and delivery dates, they proposed combining goods. This would allow railway companies to use spare capacity on their competitors' trains. The mathematical formulation was developed for the problem of serving spot markets. This helped in framing how operations are set up and how there are different levels of intervention and agility which may be considered. Conversely, here the idea is not to make daily changes but to set up a system for its environment as it exists and to have it operate in that manner for a medium-term time horizon.

In addition to the spot market nature of some parts of the economy, there are also parts of the railway business that are suited to bulk material handling. Demand forecasts have begun to take note of a decline in coal requirements as the nature of the industry changes, (Nicholas & Buckley, 2019). The generation of power using coal is an established technology. As power generation develops and better options become available, the use of coal will change regardless of policy. Here, the environment and the outlook for coal were framed. It gave insight into whether this key commodity would be used in the medium term and hence whether a demand would exist for railway companies. The indications are that demand for coal volumes are in serious doubt, therefore the inference may be that rail companies should develop other areas of their businesses.

Woroniuk & Marinov (2013) found that the container market represented a significant growth opportunity for railway companies needing to attract market share from road, this is in contrast to bulk material handling. Containers are well suited to the intermodal market because the goods are protected. The hub and spoke operating concept is used to leverage efficiencies, where short trains and/or road transport from the source are combined into long trains at an intermodal hub. From this point it would usually be a long journey by train as this is needed to overcome the burden of high fixed costs and benefit from rail. The hub and spoke concept is especially suited to containers as they are easily handled. Containers are also suitable for the last mile journey from the train to the customer’s doorstep on road.

As the cost of transport of their goods to market becomes more significant, companies must consider alternative modes and combinations of modes of transport to stay competitive, (Racunica & Wynter, 2005). Railways have enjoyed an advantage over road in movement of bulk commodities. Railways can no longer rely on this type of business where they could operate without significant competition from other modes. The rail mode has specific advantages over road depending on distance, weight, parcel size and flexibility requirements. The same holds true for road regarding its specific advantages as a mode. Economic benefits accrue to the rail mode as distance and volume increase, whereas road gains an advantage where flexibility is more valued. Flexibility could be origin-destination flexibility or time flexibility, therefore perpetuating the intermodal hub and spoke model.

2.4 Network hub location problems

Network hub location problems focus on mathematical formulations for the transport network problem. Formulations have been developed in the literature to study various types of problems and for various objectives.

Several mathematical formulations are known in the literature for the hub location problem, of which the following models are relevant to the research carried out in this thesis. Hubs, their purpose and the various formulations developed to model them are defined as the ‘p-hub median’ problem (where p is the number of hubs), ‘hub location problem with fixed costs’, ‘p-hub centre’ problem and ‘hub covering’ problems, (Alumer & Kara, 2008).

The p-hub median problem is applied to a hub and spoke network structure and refers to a set of hubs between the origins and destinations, where the objective is to minimise the cost by selecting the locations of the hubs. It can be split into two versions, single allocation and multiple allocation, referring to the allocation of demand nodes to one or many hubs.

In defining the p-hub median problem, it was noted by Campbell (1996) that hub location problems have two key characteristics, viz. that origin-destination flows travel via hubs, and that interhub transportation benefits from a discount factor, $0 \leq \alpha \leq 1$. Further, hub location problems have the following characteristics:

i. n demand locations are given;
ii. the flow for the origin-destination pairs is given;
iii. the transportation cost (per unit) and the interhub discount factor is given.

Hub median location problems have an objective to minimise transportation costs by finding hub locations. This is fundamentally different from the problem under consideration in this thesis, where the hubs are given and the objective is to minimise transportation costs by finding optimal train lengths.

In the formulation, Alumer & Kara (2008) consider w_{ij} as the flow between the origin-destination pair, with nodes *i* and *j*. The unit transportation cost, c_{ij}, applies between *i* and *j*, x_{ik} is defined as 1 if node *i* is allocated to a hub at *k*, and 0 otherwise; x_{kk} is 1 if node k is a hub and 0 otherwise. Essentially, a proportion of the flow is assigned to each sector of the journey where it exists. A unit cost is applied to find the cost per sector, following which the cost function is minimised. Decisions are made about which nodes are assigned to the hubs and how many hubs exist. The single allocation p-hub median problem formulation is, (Alumer & Kara, 2008):

Minimise $\sum_i \sum_j w_{ij} \left(\sum_k x_{ik} c_{ik} + \sum_m x_{jm} c_{jm} + \alpha \sum_k \sum_m x_{ik} x_{jm} c_{km} \right)$ (1)

s.t (n-p+1)$x_{kk} - \sum_i x_{ik} \geq 0$ for all k, (2)

$$\sum_k x_{ik} = 1 \qquad \text{for all i,} \tag{3}$$

$$\sum_k x_{kk} = p, \tag{4}$$

$$x_{ik} \in \{0,1\} \qquad \text{for all i, k.} \tag{5}$$

Where *n* and *p* are the number of nodes and hubs respectively.

Here, possible hub locations are k = 1, 2,….m from which p is chosen, and the total demand points or nodes are i = 1, 2,…..n, n>m. Equation (1) minimizes the cost of the flow, with α in the third term accounting for economies of scale on the hub to hub journey. This formulation is an accepted notation describing the cost optimisation of a hub and spoke system, including intermodal operation. In Equation (1), each of the three terms applies to a journey between different nodes, which could be different transport modes and where the following apply:

1st Term: w_{ij} must equal w_{ik}

2nd Term: w_{ij} must equal w_{jm}

3rd Term: w_{ij} must equal w_{km}

The formulation is flexible in that the input, p, can be varied to consider specific hub quantities. The existing formulation for a hub model considers the flow through a system and targets many objectives, such as hub location and costs, among others; capacities of hubs can also be included in the model with additional constraints. Hub location problems are concerned with identifying hub locations and how these hubs interact with the nodes in their network. Therefore, they consider multiple locations, and the complexity generally arises from this. Where they fall short, however, is in modelling or observing how transporting the freight in combinations of vehicles (trains) has any implication or considering that the train length may be varied to minimise the cost through the system. This is a shortcoming in the existing literature and the capacity to perform this analysis will need to be developed in this study.

The second model proposed by Campbell (1996), defined a p-hub median as a discrete problem and expanded on the history of its development from 1965. While it was found that most research focused on single hub allocations, Campbell (1996) was interested in multiple allocations to more than one hub (meaning that any node may be allocated to multiple hubs). This was believed to be necessary to minimise transportation costs.

Consider that flow occurs from node *i* to *j*, via hubs *k* and *m*, in that order, the mathematical formulation is as follows, (Campbell, 1996):

Minimise	$\sum_i \sum_j \sum_k \sum_m w_{ij} x_{ijkm} c_{ijkm}$		(6)
subject to	$\sum_k \sum_m x_{ijkm} = 1$	for all i,j	(7)
$\sum_k y_k = p$			(8)
$0 \leq x_{ijkm} \leq y_k$		for all i,j,k,m	(9)
$0 \leq x_{ijkm} \leq y_m$		for all i,j,k,m	(10)
$y_k \in \{0,1\}$			(11)

where,

x_{ijkm} = fraction of flow from *i* to *j*, via *k* and *m*

y_k=1 if location k is a hub, 0 otherwise

$w_{ij} = flow\ from\ i\ to\ j$

$c_{ijkm} = c_{ik} + c_{mj} + \alpha c_{km}$

Equation (7) is a constraint ensuring that every origin-destination pair flow pass through a hub pair, equation (8) ensures there are p hubs, equations (9) and (10) ensure that the flow is routed via nodes that are hubs, equation (11) identifies if a hub is chosen or not.

In the mathematical formulation of the problem considered in this thesis, the transportation cost is minimised over all origin-destination pairs, as it is in the existing literature. For the intermodal hub and spoke model, this formulation shall be applied considering the destination as the port, which shall also be a hub. Thus, the freight flow will be routed from node *i* to *j*, via *k*, with node *j* = *m*, shown in Figure 2.1. Of significance, while the terms in the existing formulation depict flow and fractions of flow, the formulation shall be developed to consider trains and wagons instead, thus allowing these parameters to be determined when the cost is optimised, leading to the study of variable length trains. This development shall be expounded on in Chapter 4.

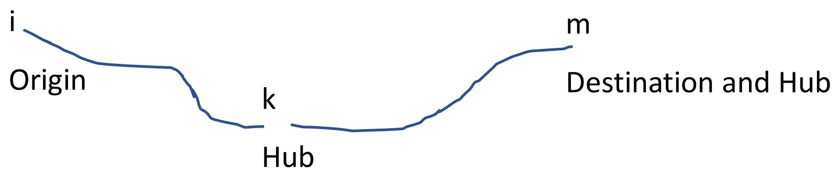


Figure 2.1 Logistic layout for two hubs

2.5 Capacity studies

Capacity studies also concern mathematical models but are presented here for the various directions of enquiry they followed. The capacity studies attempted to find optimal solutions by considering different size vehicles to achieve economies of scale. This is analogous to running longer trains as studied in this thesis.

Serper & Alumur (2016) built a model to facilitate the design of hub networks. This model allowed different vehicle types and alternative transportation modes. They were considering multiple parameters to find the locations and capacities of the hubs, to identify the transportation modes suited to these hubs and to find the number of vehicles to serve these hubs. The authors developed solutions to find the minimum total cost. They included fixed costs, capital and operational costs of vehicles, transportation and material handling costs. Serper & Alumur (2016) proposed that an operation uses different types of vehicles in their hub networks, thereby achieving economies of scale. Vehicles with larger capacities can carry more goods at a lower unit cost of transportation. This is used to decrease transportation costs where possible.

Small parcel delivery networks operate on a hub and spoke network structure. Serper & Alumur (2016) proposed that smaller vehicles would be used to transport parcels from the branch offices to the hubs. Between hubs, larger vehicles would travel; hence a consolidation occurred, and this would achieve economies of scale. Their objective was to optimise the location and the capacities of hubs and the transportation modes at the hubs. In addition, the allocation of non-hub nodes to hubs and the number of vehicles (of large and small size) required to find the lowest total cost were determined. In Serper & Alumur (2016) a mathematical formulation was presented as a mixed integer programming formulation for the single allocation capacitated intermodal hub network design problem. They developed a heuristic algorithm, as real-world problems could not be solved in a reasonable time. The problem was considered as two subproblems for the heuristic, one being a hub network design, and the second a vehicle assignment problem. In the hub network design problem, they found the locations of the hubs and determined which nodes should operate as hubs and which as spokes. The vehicle assignment problem dealt with the type and number of vehicles required and therefore also the capacities required at the hubs. They tested the sensitivity of the linear model with the Turkish transportation network. The paper demonstrated the demands on computing time and the associated quality of results.

A critical takeaway from Serper & Alumur (2016) was how the mathematical model was developed and used in finding the types and sizes of trucks to service the hubs. Although their paper had multiple contributions, as it related to this study, they determined the optimal number of vehicles to operate in their fleet. They found that larger vehicles would result in improved economies of scale. However, it was done for trucks, operating individually, not in a combined group as in a train. This differs from this research where the subject was to optimise how many vehicles should be operated together in a train. The focus is also slightly different in that they were looking for the optimum fleet size to have, not how running that fleet differently would optimise the train length. Nevertheless, this paper did have value in that it expanded on how economies of scale could be exploited in a hub and spoke system, and how the operation between hubs is fundamentally different from the travel between a node and the hub.

A general framework was presented for modelling intermodal transport networks in a linear model, (Li, et al., 2013). The notable contribution in this work was the formulation of the generic model for container flows in nodes and links which considered time dependencies. Li, et al. (2013) identified the requirement to expand transport services without necessarily building additional infrastructure. Instead, they used technology levers to build rail market share. To determine the optimum route choices over the network a linear problem was solved.

They considered an intermodal network comprising rail, road and waterways. The model allowed container freight to move between any of the modes. Containers could also be stored in the nodes. The links were designed as transport and transfer links. As a transport link, it facilitated movement between nodes. The transfer link allowed for the changing of a mode at a node. Dynamic behaviour was also considered, including loading, unloading and storing containers. Constraints on the capacities of links and nodes was also considered. The dynamics of container transport and transfer time and cost in links was considered. This extended to the transport capacity of links, and the ability of nodes to store containers. A generic intermodal transport model was proposed that could be extended to consider the effect of other modes. Li, et al. (2013) recognised that the integration and coordination of different transport modes in an intermodal network could allow for optimal use of the infrastructure. Capacity was developed, while also resulting in cost and energy efficient transport services.

2.6 Case studies

Serper & Alumur (2016) conducted a case study on a Turkish network, following the development of a model and a heuristic to establish how to use different sizes and types of vehicles to achieve economies of scale. An 81 node network was studied, where it was difficult to compute where the lower bounds were, a more relaxed solution was used after a set time period of calculation and the gaps between lower bound and the heuristic solution were determined. The study moved on to select the number of hubs initially and then started to vary the costs. It was found that there was no improvement as the number of solutions increased above a certain threshold, so a limit was set. The study then tested different values for fixed hub and vehicle costs, after which the effect on the number of hubs, their location and type was observed. Different fixed costs for establishing hubs were tested next, again these had an effect on the hubs and the resultant operation of the network, influencing the locations of hubs and the number of vehicles required. It was observed that increasing the fixed costs drove down the number of hubs, affected the concentration of the flows and varied the total cost of procuring and operating vehicles. This led to the observation that the cost of vehicles also affected the hub locations (but not the number in this case) and therefore that the hub location and vehicle assignments should be considered in parallel. The structured manner of the investigation and how observations were made and conclusions drawn from the effect on the results were noted. Regarding future work, one of the recommendations that Serper & Alumur (2016) proposed was to incorporate decisions on the frequency and schedules of vehicles, which is a key consideration in this thesis.

2.7 Planning, train scheduling and market share issues

Planning is important and must consider the positioning of different levels of work. This is critical to successfully achieve goals. The study by Assad (1980) on a hierarchical framework for the planning process of rail freight transport determined the different levels of planning involved. It was found that decisions are made at three levels – strategic, tactical and operational. Strategic decisions typically focus on a long time horizon and require large capital investments, usually involving higher levels of management. Tactical decisions have medium term time frames and usually focus on effective use of existing resources rather than new acquisitions. The time frame and level of aggregation in such a model must allow it to factor in broad changes in the environment without having to incorporate day to day changes. Operational decisions dealing with daily activities are normally addressed by lower levels of management who are directly concerned with operational issues.

Assad (1980) then formulated a general model to address the interaction between routing decisions and activities in a railway yard, where trains are compiled and decompiled. Using this model, the economics associated with consolidating blocks of traffic into a single train are studied. This work by Assad (1980) contributed to framing the positioning of this research. The thinking, the execution and the planning is different between day-to-day operations and medium to long term planning and strategy development. Hence, finding the optimal train length for a logistic system, should be an action for a medium-term framework, i.e. the system is designed based on the contributing parameters and this system is operated in this way over a medium term and not changed significantly every day, as the nature of the environment is inherently inflexible.

Jafarian-Moghaddam (2021) considered balancing speed and maintenance costs with running faster, shorter length trains, primarily for passenger rail networks. The motivation in this paper is to save energy costs and deliver a faster service, addressed from a train scheduling perspective.

Trains may also be homogeneous (carrying a single commodity) or heterogeneous (carrying multiple commodities). Jafarian-Moghaddam (2021) considered railway operations factors such as compiling and decompiling trains, and shunting movements, all in terms of the effort expended in carrying out these activities in the shortest time. This consideration is important when considering heterogeneous trains, whether additional wagons of possibly another commodity are to be added. This consideration becomes less important when considering what is essentially a unit train as contemplated in this study, using containers allows a railway to consider the container as the basic unit and not the actual commodity contained within. Jafarian-Moghaddam (2021) proposed a model with three objectives being to reduce travel time, increased train length and reduced energy consumption.

In intermodal market share research was conducted on how different modes interact. Specific objectives were explored to identify how to operate in a changing economic environment. Kurtulus & Cetin (2020) investigated the parameters affecting intermodal rail market share as compared to the road market. The parameters were transportation cost, transit time, delay percentage, frequency and free time. The following characteristics determine the migration to intermodal transport: transportation distance, packaging, value, shipment size, accessibility, freight type, perishability and fragility. Kurtulus & Cetin (2020) found that in Turkey where their study was based, that volumes could be attracted to intermodal rail transport by increasing the train frequency and reducing the transit time by 50 %. In running their simulation model, they found that they could grow the intermodal rail market share from 10 % to 29 %. Kurtulus & Cetin (2020) identified the critical parameters that play a role in rail and road market share determination. This is of value as it identifies what factors should be considered for incorporation in the model in this study.

2.8 Various applications of Genetic algorithms

Genetic algorithms are increasingly recognized as powerful optimisation techniques that draw inspiration from the principles of natural selection and genetics. Their utility spans various domains, including engineering, ecology, and computer science, making them a versatile choice for solving complex optimisation problems. One of the primary reasons for choosing genetic algorithms is their ability to effectively navigate large and complex search spaces, particularly in non-linear and multi-dimensional problems where traditional optimisation methods do not perform well. This characteristic is particularly beneficial in scenarios involving numerous decision variables and intricate relationships among them, as highlighted by Karamouz, et al., (2010), who employed genetic algorithms to optimise crop patterns while considering the non-linearity of water demand and supply.

Moreover, genetic algorithms are particularly adept at handling optimisation problems that are beyond the reach of analytical techniques. Hamblin (2013) emphasizes that genetic algorithms can relax model assumptions and evolve behaviours in individual-based models, which is crucial in fields like ecology and evolution where traditional models may not adequately capture the complexity of biological systems. This flexibility allows researchers to simulate co-evolutionary processes and adapt to dynamic environments, thus enhancing the robustness of their models.

In addition to their adaptability, genetic algorithms exhibit a strong global search capability, which is essential for avoiding local optima in complex landscapes. This feature is particularly relevant in the context of fuzzy logic controllers, where Khan, et al. (2008) demonstrated that integrating genetic algorithms can significantly enhance the search for optimal fuzzy rules and membership functions. By leveraging the global search capabilities of genetic algorithms, researchers can achieve better performance and faster convergence to optimal solutions, which is a critical advantage in real-time applications.

The hybridization of genetic algorithms with other optimisation techniques further amplifies their effectiveness. For instance, Lin & Chen (2014) proposed a hybrid algorithm that combines genetic algorithms with simulated annealing and backpropagation (BP) learning, resulting in a robust optimisation framework that capitalizes on the strengths of each method. Such hybrid approaches can address the shortcomings of individual algorithms, leading to improved performance in various applications, including neural network training and optimisation tasks.

Furthermore, the parallelization of genetic algorithms enhances their computational efficiency, allowing them to tackle larger and more complex problems. Arkhipov, et al. (2020) highlighted the application of parallel genetic algorithms in transportation planning and logistic, where the iterative nature of genetic algorithms can be exploited to generate multiple solutions simultaneously. This parallel implementation not only accelerates the search process but also improves the diversity of solutions, which is crucial for achieving optimal outcomes in competitive environments.

The versatility of genetic algorithms extends to their application in multi-objective optimisation problems, where multiple conflicting objectives must be balanced. Oviedo-Carranza & Dominguez-Navarro (2022) demonstrated the effectiveness of multi-objective genetic algorithms in optimizing generation and distribution in microgrids, where the genetic algorithms were designed to meet specific operational constraints while maximizing overall system performance. This adaptability to multi-faceted problems makes genetic algorithms a preferred choice in fields such as renewable energy management and resource allocation.

In the context of engineering applications, genetic algorithms have proven invaluable in optimizing design parameters and operational strategies. For example, the work by Li et al. (2022) illustrates how genetic algorithms can be employed to enhance energy-saving measures in building design, integrating multi-objective optimisation functions to improve energy efficiency. Such applications emphasize the practical benefits of genetic algorithms in achieving sustainable design goals while addressing complex engineering challenges.

Moreover, the robustness of genetic algorithms in the presence of noise and uncertainty is a significant advantage. Medeiros, et al. (2007) noted that genetic algorithms are inherently resilient to noise, making them suitable for applications where data may be imperfect or incomplete. This robustness is particularly advantageous in real-world scenarios where environmental factors can introduce variability into the optimisation process.

The ability of genetic algorithms to evolve solutions over successive generations allows for continuous improvement and adaptation. This evolutionary aspect is crucial in dynamic environments where conditions may change over time. The research by Sharma & Singhal (2014) highlights the potential of hybrid genetic algorithms to address space allocation problems, demonstrating how genetic algorithms can adapt to evolving constraints and objectives in complex systems.

In summary, intermodal studies, hub location studies and capacity studies, applicable to various industries and modes of transport have been reported above. The extension of those approaches and solutions to address the logistic problem have been explored.

Furthermore, the choice of genetic algorithms as an optimisation technique has been explored, emphasizing that genetic algorithms can navigate complex search spaces, handle non-linear relationships, and adapt to dynamic environments.

Their hybridization with other optimisation methods enhances their effectiveness, while parallel implementations improve computational efficiency. The versatility of genetic algorithms in addressing multi-objective problems and their robustness in the face of uncertainty further solidify their position as a preferred choice in various fields, including engineering. As research continues to evolve, the applications and methodologies surrounding genetic algorithms are likely to expand, offering new solutions to complex optimisation challenges. In the next chapter, the configuration of a logistic system is described.

Chapter Three

Logistic System Configuration

3.1 Logistic system configuration

The elements of a logistic system required for this investigation are described before they are configured in a conceptual layout. Careful consideration of the requirements to unpack the research question have led to the development of this layout.

A logistic environment is a complex mix of elements designed ultimately to move freight from an origin to a destination. These elements have varying purposes and different operational characteristics. They comprise of rail, rail vehicles, road, road vehicles, hubs with handling and storage facilities amongst other components. These elements work together in a logistic chain, and they may be operated with different strategies in place.

Various modes develop as volume and frequency requirements dictate. The available modes are each matched to different types of freight, volumes, and the frequency that transport is required, as well as external factors such as pollution, interaction with the public (in the case of interaction between freight and movement of the general public) and economic factors.

The surface modes generally available are rail, road and pipeline, supported by sea and air transport modes; shipping for global reach and air for urgency and/or global reach. There are multiple complexities related to these modes. The modes are appropriate depending on the circumstances.

Decisions on how to use various transport modes are made for the circumstances based on the required operating parameters which are briefly presented in Table 3.1.

Table 3.1 Some characteristics of transport modes[1]

Mode	**Operating parameters**	**Advantages**	**Disadvantages**
Rail	• Large volumes • Long distances	• Economy of scale • Controlled environment • Separation from the public • Low carbon footprint	• Inflexible to variable customer locations and volumes • High startup costs
Road	• First/last mile • Short distance • Small volumes	• Flexible to variable customer locations	• High operating cost • Public interaction • High carbon footprint
Pipeline	• Liquid bulk • Long/short distance	• Economy of scale • Controlled environment • Separation from the public • Low carbon footprint	• Inflexible to variable locations • High startup costs
Sea	• Global transport • Large volume	• Economy of scale • Lower carbon footprint	• High startup cost • Time (slow)
Air	• Global/local transport	• Time (fast)	• Carbon footprint • High cost

[1] Table developed from logistic industry general knowledge, multiple sources

In addition to considering the type of mode, further decisions must be made about how that mode is used. Decisions happen at various levels, being strategic, tactical and operational, all with a different focus and time horizon, (Assad, 1980).

Given the complex nature of this environment, a mathematical model was developed to observe the complexities of the logistic operation under varying circumstances. A hub and spoke model was also built in Simulink® to unpack the functioning of this system.

In the next sections, the elements of the environment are introduced, and used in these models.

3.1.1 Intermodal system

This study is conducted on intermodal logistic systems. In this section, the essential elements of such systems that are essential for the operating environment are presented.

3.1.1.1 Origins

An origin is a point of production in any of the industrial, mining or agricultural sectors. To conduct its business, the origin would produce goods and receive goods as raw materials or consumables required for operations.

The connectivity of the origin to the transport network would depend on its size, production capacity, financial resources, and the demand for its products. These goods may accumulate to considerable volumes depending on the size and the nature of the production facility and would require various external logistic solutions to move the goods outside of the facility.

Decisions about transport solutions must be made by the entity controlling the origin. A facility producing large volumes which are also transported over large distances to market would more likely have a rail siding to facilitate rail transport direct from the site. Alternatively, an origin may use road as a first mile service provider, to transport goods to a Hub where they are transferred to a train for a longer journey. This makes use of economies of scale and the advantages rail offers as a transport mode.

An origin with less demanding transport requirements may use road as the primary transport service, however, a decision was made to use the rail transportation mode in the first mile in the models and examples developed in this study.

Facilities that are taking advantage of intermodal transport will pack their goods into containers to facilitate the efficient transfer of the goods from one mode of transport to another without additional manual handling of the goods.

3.1.1.2 Transport

In this study, the transport of goods in containers is considered. Containers may be transported by rail or by road. Transport encompasses factors such as the distance traversed, speed of travel, volumes carried and delays which may be encountered.

Rail transport makes use of a special container train, requiring locomotives, wagons and crew.

Road transport of containers in this case may be applicable for the first or last mile, it may also be used for the entire journey. Road transport is more onerous than rail, requiring individual trucks and drivers for each consignment, as well as being subject to the unpredictable elements of using a public road, including traffic, capacity constraints and increased interaction with the public compared to rail. Rail transport was however used for the first mile.

Containers may also be loaded on a train in a single level or stacked two high, known as double stacking. This effectively doubles the capacity of a train. Double stacking can only be used where the track gauge is broad enough to result in a stable train when the centre of gravity is raised by double stacking. This would also require that the railway’s structure gauge allows the height of the train when double stacked. South African railway systems can only be single stacked due to the stability limitations of the rail gauge.

3.1.1.3 Hub

A hub is a critical part of an intermodal transport system, it is at the hub that goods are transferred between modes and where trains are compiled into different lengths. Hubs also store containers, acting as a buffer until loading onto the right train is possible.

The management of a hub has a significant impact on the capacity, efficiency and cost effectiveness of an intermodal system. It is a pivotal point at which the strategy of how the system is operated (in terms of train lengths, frequencies, and destination decisions) can be deployed.

3.1.1.4 Port

The port represents a modal shift between water and land. On the land side, such container terminals are generally serviced by rail and trucks. Depending on the configuration of the logistic system, a port may also be considered a hub.

The port represents a localized destination. Freight is delivered to the port where it is handled and loaded onto ships or delivered to a last mile destination near the port. Freight loaded onto a ship also continues its journey over the sea to a global destination, making use of the advantages of sea travel in an intermodal system (instead of air travel, for example).

At the port goods may also be stored until the right ship is available, or the right train or truck in the case of goods arriving by ship.

3.1.1.5 Locomotives

Locomotives provide the traction for a train. They are allocated based on the train length, terrain and speed profile of the track as these define the tractive effort required. Locomotive consideration in the study is limited to the cost contribution and the balancing of resources.

3.1.1.6 Wagons

Wagons carry the freight, typically containers in an intermodal system. A wagon can carry two teu's when single stacked. Major benefits of railway wagons over trucks are that they can be combined in a train to any length (the length is limited by external factors which are not the subject of this study), can sustain heavy axle loads and have very low rolling resistance as they run on steel wheels on steel rails.

Wagons can also be combined into longer trains or removed from a train; they can be combined with wagons that are being delivered to a common customer or a common area. These combinations are all influenced by the strategies employed by the railway. There is a negative aspect to handling the wagons as it uses time and resources.

3.1.1.7 Crew

Crew are the human elements in a train and require significant training and knowledge of the train and the terrain. Crew are deployed in various ways depending on the philosophy of the railway, typically with one train driver and one train assistant on a train.

3.1.1.8 Freight

Freight are the goods produced and moved by companies to a market to generate revenue and profit. Some goods are suited to being containerised. More and more goods are containerised to increase the efficiency of the transport. They may also be protected in this way and require fewer handling resources as the smallest unit becomes the container and only this must be handled. Refrigerated containers are available for perishable goods.

The containers can be stacked on a wagon on one level or two levels high (note that this is only possible on a broader track gauge and not in South Africa). One rail wagon can carry two teu's on each level, shown in Figure 3.1.

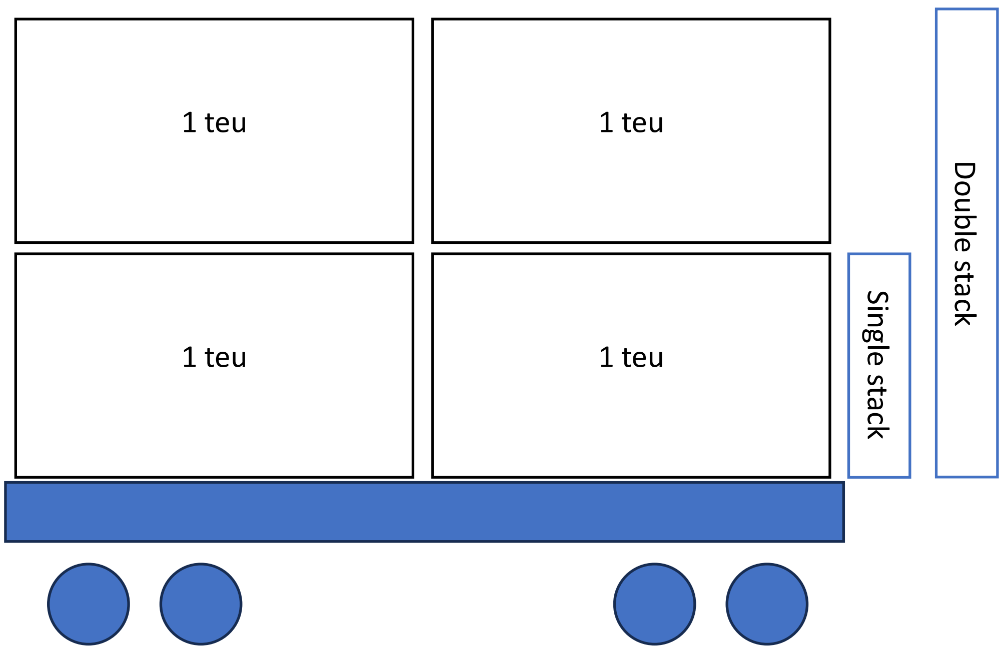


Figure 3.1 Freight wagon with containers

3.1.1.9 Inefficiencies

An intermodal system is subject to inefficiencies at every stage, costing time and money while eroding capacity. Strict management and the adherence to maintenance and operating policies are required to minimize inefficiencies.

3.1.1.10 Nonlinear delays

Nonlinear delays refer to events that are not typically planned for and usually result in large, unpredictable delays. These delays may be due to random events such as derailments, vandalism, floods etc.

3.2 Conceptual layout of the basic model

The following conceptual layout of a transport model in Figure 3.2 is proposed to study an intermodal system and generate data.

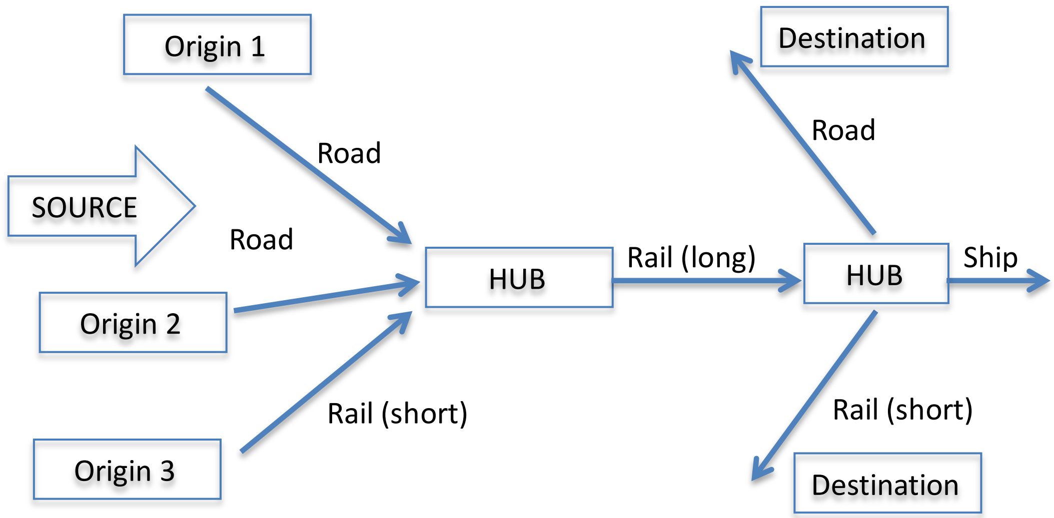


Figure 3.2 Hub and spoke conceptual layout

The transport of freight will be modelled using the basic layout of a hub and spoke network shown in Figure 3.2. The goods originate at the origins, identified as the Source on the left of the diagram. Consider the case where goods are produced at a factory and containerised. The containers may be stored temporarily at storage facilities at the factory. Production of goods will result in a set of containers at the factory, increasing at a defined production rate. These containers are transported via rail or road on the initial journey to the hub, where there are also temporary storage facilities. At the hub they may be compiled into longer trains for a rail journey to the port where they are offloaded from the train, stored for a brief period, and loaded onto a ship. If they have a local destination, they may also be transported by road, sea or rail to that destination. This layout will form the design basis of the model.

The effect on volume throughput shall be tested when costs are optimised with varying train lengths on the entry and exit side of an intermodal hub. The hub facilitates movement from road onto rail and vice versa. At the hub trains are also compiled and decompiled. Thus, at the hub there is an opportunity to combine short trains into long trains. Historically, this practice is believed to make the best use of the slot capacity on the main line. This principle of maximising slot capacity shall be studied thoroughly mathematically.

Chapter Four
The Mathematical Model

The development of a novel mathematical model for cost optimisation based on a hub and spoke network is presented in this chapter. Given a number of constraints and input values, this optimisation will be carried out with respect to train lengths and departure decisions.

4.1 Mathematical formulation

A logistic chain will be modelled mathematically, from the perspective of the transport system. The role of the optimisation is to mathematically find the optimal train lengths to transport goods at the lowest cost.

In this section, a mathematical formulation using an intermodal hub network is presented (refer to Figure 3.2 for the layout of the network). This formulation was developed from a p-hub network problem with existing and known locations and characteristics of the network, with the intention of finding the optimal train length to transport containers from an origin to a destination. Hence, the formulation has been developed to consider the number of wagons and trains. Therefore, the model is developed for an operational perspective, where an existing network is used, and its operation is optimised.

The layout and notations are presented in Figure 4.1, where the idea of connecting origin points (see O_1 and O_2) to a hub (see H_1) via a first mile transit is presented. Consider that a flow occurs from node *i* (multiple origins) to *j* (destination), via hubs at *k* and *q*, designated H_1 and H_2. H_2 is located at the port, destination *j*, therefore *j* and *q* are located at the same node. The problem is formulated to find the optimal train lengths to transport containers originating at a node *i* to a destination H_2 via H_1 over a given period of time, with a given throughput requirement. A seaport, defined conceptually as H_2, forms a boundary to the problem.

There are two types of optimisation variables: the number of trains and the train length (number of wagons in each such train), over any given period. Trains travel from the origins to H_1, then they travel from H_1 to H_2 on a main line journey, possibly as a different length train. The number of trains travelling from Origin *i* to H_1 is denoted by the variable x_{i1}, $i \in \{1, 2 \ldots n\}$, where n is the number of origins. Similarly, the number of trains travelling from H_1 to H_2 is denoted by the variable x_{12}. The corresponding train length variables are denoted as y_{i1} and y_{12}, having different length constraints before and after H_1 due to the infrastructure and/or operation configurations. The number of containers available at origin *i* is denoted by w_i and w_{H1} is the number of containers available at the hub, H_1. These are not optimisation variables since the value of w_i and w_{H1} are not determined optimally. All these variables have integer values. An illustration of running different length trains under different operating regimes is presented in Section 4.3. Cost (variable and fixed) of transporting each wagon per unit distance between the origins and the hub H_1 and between H_1 and H_2 are defined as m and m_{H1}, respectively. The overhead cost is defined as mO and applies only when a train does not run, it accounts for wasted resources if a slot is not used. The costs represent existing operational costs (typically personnel, etc.), maintenance costs and the cost of the organisation to support the railway. Note that overhead costs are also allocated when a train does run as part of that cost component.

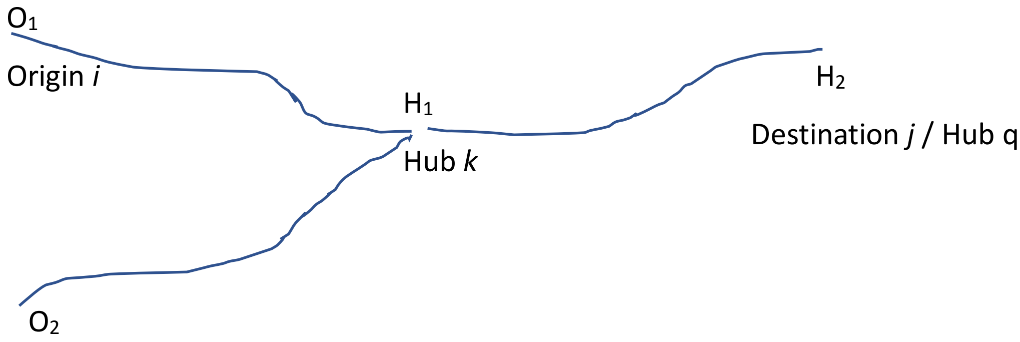


Figure 4.1 Conceptual layout of hub and spoke system

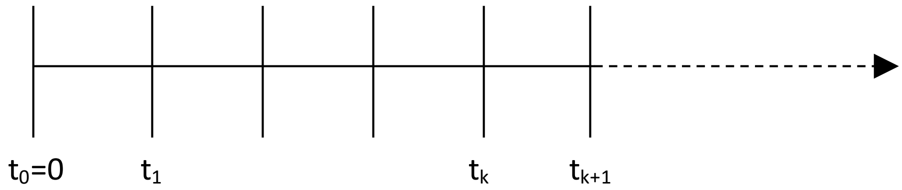


Figure 4.2 Timeline for model

Figure 4.2 depicts the timeline for the model, where t_0 represents the start of the modelling period, and where the model time progresses in discrete but equal, $t_{k+1}-t_k$ time intervals. Here, e slot time periods with discrete time points $t_0, t_1, \ldots. t_e$ were considered. For each time interval, goods are being produced and containerised, and trains are being compiled, requiring a decision on train length and departure at the following time point. Transit times are indicated by t_{i1} and t_{12}, where t_{i1} is the transit time from origin i to the hub and t_{12} is the transit time from H_1 to H_2. The optimisation and other variables will now have a time dimension associated with them. For example, the variables will be denoted as $x_{i1}(t_k)$, $y_{i1}(t_k)$, $x_{12}(t_k)$ and $y_{12}(t_k)$, k = 0,1,2,….e. The initial container values at each node, $w_i(t_0)$ and $w_{H1}(t_0)$ are given. Two binary variables are defined as follows:

$$x_{i1}(t_k) = \begin{cases} 1, & if\ a\ train\ departs\ from\ origin\ i\ at\ t_k, \\ 0, & if\ a\ train\ does\ not\ depart, \end{cases} \quad \text{for all } i, k. \qquad (1)$$

$$x_{12}(t_k) = \begin{cases} 1, & if\ a\ train\ departs\ from\ H_1\ at\ t_k, \\ 0, & if\ a\ train\ does\ not\ depart, \end{cases}$$ for all *k*. (2)

$ORateP_i(t_k)$ are the container production rate per period at each origin *i* and $ORateP_{H1}(t_k)$, the rate per period at which containers arrive at H_1. D_{Total1} and D_{Total2} are the total demand at H_1 and H_2 respectively, this is the number of teu which are required at the destination.

For the entire period, the mathematical model is defined as follows:

Minimise

$$\sum_{k=0}^{e}\sum_{i=1}^{n} my_{i1}(t_k)x_{i1}(t_k)d_{i1} + m_{H1}y_{12}(t_k)x_{12}(t_k)d_{12} + mO(-x_{i1}(t_k)+1) + mO(-x_{12}(t_k)+1) \quad (3)$$

subject to:

$\rho y_{i1}(t_k) \leq w_i(t_k)$, for all *i* , k, (4)

$y_{i1}(t_k)(1 - x_{i1}(t_k)) \leq 0$, for all *i*, *k,* (5)

$\rho y_{12}(t_k) \leq w_{H1}(t_k)$, for all *k,* (6)

$y_{12}(t_k)(1 - x_{12}(t_k)) \leq 0$, for all *k,* (7)

$LB_i <= y_{i1}(t_k) <= UB_i$, for all *i*, *k,* (8)

$LB_{12} <= y_{12}(t_k) <= UB_{12}$, for all *k.* (9)

where LB_i and UB_i are lower and upper bounds of train lengths that can be used at origin *i*. Similarly, LB_{12} and UB_{12} are respectively the lower bound and upper bound of allowable train lengths at the hub, H_1, departing on the main line towards H_2. These bounds are based on the physical characteristics of the railway and are assumed to be given for the purpose of this study.

$$\rho \sum_{k=0}^{e} \sum_{i} x_{i1}(t_k) y_{i1}(t_k) \geq D_{Total1} \tag{10}$$

$$\rho \sum_{k=0}^{e} x_{12}(t_k) y_{12}(t_k) \geq D_{Total2} \tag{11}$$

The parameters are updated as follows:

$$w_i(t_k) = w_i(t_{k-1}) - \rho x_{i1}(t_{k-1}) y_{i1}(t_{k-1}) + ORateP_i(t_{k-1}), \qquad \text{k = 1,2,...e, each } i, \tag{12}$$

$$w_{H1}(t_k) = w_{H1}(t_{k-1}) - \rho x_{12}(t_{k-1}) y_{12}(t_{k-1}) + ORateP_{H1}(t_k), \qquad \text{k = 1,2,...e,} \tag{13}$$

$$ORateP_{H1}(t_k) = \sum_{i=1}^{n} \rho x_{i1}(t_{k-1}) y_{i1}(t_{k-1}) \qquad \text{If } t_{i1} < t_{k+1} - t_k, \text{ k = 1,2,...e,} \tag{14}$$

$ORateP_{H1}(t_k)$ is updated as follows when $t_{i1} > t_{k+1} - t_k$:

$$ORateP_{H1}(t_l) = \begin{cases} \sum_{i=1}^{n} \rho x_{i1}(t_j) y_{i1}(t_j) & for\ l \in S, \\ 0 & for\ l \notin S. \end{cases} \tag{15}$$

where $l = k + \left\lceil {t_{i1}}/{(t_{k+1} - t_k)} \right\rceil$ and $j = l - \left\lceil {t_{i1}}/{(t_{k+1} - t_k)} \right\rceil$, k = 0,1,2,...,e-$\left\lceil {t_{i1}}/{(t_{k+1} - t_k)} \right\rceil$. The set S contains the indices $l \in \{0, 1, 2, ..., e\}$ such that $S \subseteq \{0,1,2,...e\}$.

Refer to Sections 4.2.3 and 4.2.4 for examples of calculations and for where these constraints are shown to be satisfied.

The full description of the variables, parameters, mathematical expressions and constraints of the above model is given below:

Eqns (1) and (2) are a binary decision for whether a train departs or not. Eqn (3) is the objective function where $y_{i1}(t_k)$ and $y_{12}(t_k)$ are the number of wagons moving from origin i to H_1 and from H_1 to H_2, respectively. In Eqn (3), the first term is the cost component from the origins to the Hub, H_1, the second term is the cost component between the hubs, H_1 and H_2, and similarly the third and fourth terms are the overhead cost components, containing a component of the costs from terms one and two which must be accounted for only when a train does not run in the slot. This accounts for the overhead costs which are incurred and would be wasted if a slot is unused. These overhead costs include maintenance, the operations and the structure of a railway company. Note that the first and second terms in Eqn (3) differ from previous formulations in that they encompass the train length and train decision variables now, instead of the hub location and flow parameters.

Binary decisions for trains departing from origin i to H_1 and from H_1 to H_2, respectively, are depicted by $x_{i1}(t_k)$ and $x_{12}(t_k)$ at each time t_k, k = 0, 1,…e.

The fixed distances between origin i and H_1 and between H_1 and H_2, respectively, are depicted by d_{i1} and d_{12}.

Fixed and variable costs per unit distance per wagon from the origins and from H_1 to H_2, respectively, are depicted by m and m_{H1}, these result from operating costs. The overhead cost, mO, only applies when a train does not run (because m and m_{H1} do include overhead costs also). The overhead costs result from the cost of maintaining an organisation to run the railway, including the maintenance, crew and personnel costs of the railway. These costs will vary between different organisations as they are dependent on the structure and purpose of each organisation.

In Eqns (4) and (6), $w_i(t_k)$ and $w_{H1}(t_k)$ are the number of containers waiting for departure at origin i and at hub H_1, respectively, at time t_k. In Eqn (4), ρ is the number of container teu's per wagon.

D_{Total1} in Eqn (10) and D_{Total2} in Eqn (11), are the total Demand at H_1 and H_2 respectively. This is the number of teu's which are specified to be transported from all the origins, t_{12} is the transit time from H_1 to H_2. The demand requirement, $D_{Total2} > D_{Total1}$, as the accumulation of containers at the destination represents the total volume throughput of the system and may include containers which were present earlier at H_1, or which originated outside of the system.

In Eqn (12), $ORateP_i(t_k)$ is the container production rate per period at each origin i. In Eqn (13), $ORateP_{H1}(t_k)$ is the rate per period at which containers arrive at H_1. The transit time from each origin i, is depicted by t_{i1}, $i \in \{1, 2 \ldots n\}$.

An explanation is provided for finding the parameter $w_{H1}(t_k)$ when $t_{i1} > t_{k+1} - t_k$. Eqn (13) is used to calculate $w_{H1}(t_k)$, and $ORateP_{H1}(t_k)$ is found using Eqn (15).

The calculation for $ORateP_{H1}(t_k)$ is demonstrated here where transit time $t_{i1} > t_{k+1} - t_k$, and $t_{i1} = 1.2$ hrs for Eqn (15). For this example, consider $t_0, t_1, \ldots t_e$, suppose that t_0 starts at 08:00 on a day and that the interval, $t_{k+1} - t_k = 1$ hrs. If $k = 0,1,\ldots 4$, then the last time point is at 12pm. Since $l = k + \left\lceil t_{i1}/(t_{k+1} - t_k) \right\rceil$, it follows that S = {1, 3, 4} for k = 0, 1, 2. Hence, for $k = 0$, $l = 2$; $k = 1$, $l = 3$; $k = 2$, $l = 4$. The $k = 2$ is the last value since $2 = l - \left\lceil t_{i1}/(t_{k+1} - t_k) \right\rceil = 4 - 2$.

The mathematical expression in Equations (1) to (15) have now been described. Eqns (1) and (2) is a binary decision for whether a train departs or not. Eqn (3) is the cost function to be optimised, where the first term is the cost over all the routes from each origin *i* to the hub and the second term is the cost between the hubs, H_1 and H_2. Eqns (4) and (6) ensure that train lengths are limited by the number of available containers as well as defining the relationship between wagons and containers. Eqns (5) and (7) do not allow wagons if there are no trains. Eqn (8) and (9) define the train length constraints of the system. Eqns (10) and (11) define the transport demand requirement of the system. Eqns (12) and (13) define how many containers are available for departure, considering how many were available at the previous time period; how many have since left on a train and how many were produced in this time period. Eqns (14) and (15) defines the arrival rate of containers, at the hub, H_1. The total production at origin *i* is given by:

$$P_{Total} = \sum_{k=0}^{e} ORateP_i(t_k). \quad (16)$$

Eqn (16) finds the total production quantity of containers.

To implement the optimisation algorithm, the inequality constraints, Eqns (4-7 and 10-11), have been handled by penalising them in a penalty function. The constraints of Eqns (8, 9) are handled by limiting the integer inputs. The penalised Objective Function is defined in Eqn (17).

Minimise F(x,y) = $\sum_k \sum_i my_{i1}(t_k)x_{i1}(t_k)d_{i1} + m_{H1}y_{12}(t_k)x_{12}(t_k)d_{12} + mO(-x_{i1}(t_k)+1) + mO(-x_{12}(t_k)+1) + R_1\phi_1(x_{i1}(t_k), y_{i1}(t_k)) + R_2\phi_2(x_{12}(t_k), y_{12}(t_k)) + R_3\phi_3\big(y_{12}(t_0), y_{12}(t_1), \dots y_{12}(t_e)\big) + R_4\phi_4(y_{i1}(t_k)|i = 1,2, \dots n; k = 0,1, \dots e) + R_5\phi_5(y_{i1}(t_k), w_i(t_k)|i = 1,2, \dots n; k = 0,1, \dots e) + R_6\phi_6(y_{12}(t_k), w_{H1}(t_k)|k = 0,1, \dots e)$ (17)

The construction details of Eqn (19) are as follows:

$$\phi_1(x_{i1}(t_k), y_{i1}(t_k)) = max\{0, p(x_{i1}(t_k), y_{i1}(t_k))\}, \quad (18)$$

$$p(x_{i1}(t_k), y_{i1}(t_k)) = y_{i1}(t_k)(1 - x_{i1}(t_k)); \quad (19)$$

$$\phi_2(x_{12}(t_k), y_{12}(t_k)) = max\{0, f(x_{12}(t_k), y_{12}(t_k))\}, \quad (20)$$
$$f(x_{12}(t_k), y_{12}(t_k)) = y_{12}(t_k)(1 - x_{12}(t_k)); \quad (21)$$

$$\phi_3(y_{12}(t_0), \dots y_{12}(t_e)) = max\{0, h(y_{12}(t_0), \dots y_{12}(t_e))\}, \quad (22)$$
$$h(y_{12}(t_0), \dots y_{12}(t_e)) = D_{Total2} - \sum_k \rho x_{12}(t_k) y_{12}(t_k); \quad (23)$$

$$\phi_4(y_{i1}(t_k) | i = 1,2, \dots n; k = 0,1, \dots e) = max\{0, g(y_{i1}(t_k) | i = 1,2, \dots n; k = 0,1, \dots e)\}, \quad (24)$$
$$g(y_{i1}(t_k) | i = 1,2, \dots n; k = 0,1, \dots e) = D_{Total1} - \sum_i \sum_k \rho x_{i1}(t_k) y_{i1}(t_k); \quad (25)$$

$$\phi_5(y_{i1}(t_k), w_i(t_k) | i = 1,2, \dots n; k = 0,1, \dots e) = max\{0, n(y_{i1}(t_k), w_i(t_k) | i = 1,2, \dots n; k = 0,1, \dots e)\}, \quad (26)$$
$$n(y_{i1}(t_k), w_i(t_k) | i = 1,2, \dots n; k = 0,1, \dots e) = \sum_i \sum_k \rho y_{i1}(t_k) - w_i(t_k); \quad (27)$$

$$\phi_6(y_{12}(t_k), w_{H1}(t_k) | k = 0,1, \dots e) = max\{0, o(y_{12}(t_k), w_{H1}(t_k) | k = 0,1, \dots e)\}, \quad (28)$$
$$o(y_{12}(t_k), w_{H1}(t_k) | k = 0,1, \dots e) = \sum_k \rho y_{12}(t_k) - w_{H1}(t_k). \quad (29)$$

Suitable penalty parameters are defined as R_1, R_2, R_3, R_4, R_5 and R_6 and used in Eqn (17), whose values will be tested numerically. The parameters R_1 and R_2 apply to Eqns (19) and (21) respectively, where no wagons are allowed to run in a slot if no train is running. The parameters R_3 and R_4 apply to Eqns (23) and (25) respectively, which require a predetermined demand to be met. The parameters R_5 and R_6 apply to Eqns (27) and (29) respectively, which require that there must be sufficient containers to make up a train.

In Eqn (17), terms 1, 2, 3 and 4 calculate the cost, while terms 5 to 10 consist of penalties to implement the constraints of the model. Using larger penalty parameters will create larger values for terms 5 to 10 of Eqn (17), where ideally these terms should be 0, if the constraints were met. Therefore, in minimising Eqn (17), the lowest cost will be determined while also implementing the constraints. The values of R_1, R_2, R_3, R_4, R_5 and R_6 must be determined so that each of terms 5 to 10 will have sufficient impact on the cumulative value of terms 1 to 4 to enable an effective minimisation.

In Eqn (17), the first two terms have the same structure of previous p-hub median formulations presented in Section 2.4. However, Eqn (17) differs from previous work in that those first two terms have parameters for train length, train decision and distance, not flow volumes and hub decisions. The remainder of the terms in Eqn (17) incorporate the constraints of the logistic operation.

The transportation cost is minimised over all combination of train lengths and number of trains. The purpose of this optimisation is to decide the values of variables $x_{i1}(t_k)$, $y_{i1}(t_k)$, $x_{12}(t_k)$ and $y_{12}(t_k)$, k=0,1,2,….e, so that the cost is minimised and demand constraints (10), (or (11)), together with other constraints are satisfied.

To motivate the mathematical model and the optimisation thereof, a small example has been formulated by choosing the input parameter values of the above model. Scenarios were then manually built to motivate the optimisation model, presented in the next section. The optimal results of this sample problem will be presented in the numerical section. The maximum train length cannot be the optimal solution because lower costs are incurred by running shorter trains and there is an overhead cost, mO, which is incurred for empty slots.

4.2 An example problem

The motivation of the optimisation model has been demonstrated here. The process starts with a manual solution, then advances to a bespoke genetic algorithm developed to solve the problem in Section 5.3.

Here the optimisation problem will be motivated, specifically why the maximum train length may not necessarily be the optimal solution. Various scenarios were defined to depict the effect of different types of train departure decisions, arising from different train departure strategies. The strategies will determine the number of wagons that run and whether a train runs in a time slot. The calculations presented here are intended to illustrate and give clarity to the various operating strategies.

A train make-up strategy is necessary to make decisions about when to release a train. While studying the calculations and operational possibilities, it can be deduced that two basic scenarios will arise. These scenarios can be described as follows:

- Full train length - The train must wait to make up the full defined length, subject to the satisfaction of the constraints at each t_k. The railway line is designed to operate with trains of a certain maximum length, based on environmental factors. In this scenario, consider the position that a railway will only operate trains of this designed maximum length to maximise the capacity of the slot. This implies that the railway may lose some slots, e.g. $x_{i1}(t_k)$=0, or $x_{12}(t_k)$=0, if there are no maximum length trains available at t_k.

- Slot priority - When the slot is open, the train can be released, irrespective of whether it is full length or shorter. This is a more flexible approach and implies that there is a greater probability that a train can be run in any given slot, albeit at a shorter length.

4.2.1 Conceptualization of problem

The example problem is developed based on the hub and spoke layout in Figure 4.1. Input values for the problem are given in Table 4.1, except for $ORateP_{H1}(t_k)$ which must be calculated for each k from the train arrivals. Values for all the variables are then found manually, presented in Table 4.2 (Full train length scenario) and Table 4.3 (Slot priority scenario). Numerical examples shall be presented here to fully explain these scenarios and to satisfy the mathematical constraints. Note that results are manually chosen which are not necessarily optimal. In this example 11-decision points, i.e. e = 10, were considered for ease of demonstration. Each time period, $t_{k+1} - t_k$, is considered a slot in which one train may run on each section, where this time period is 8 hours long. In these examples, t_{i1} and $t_{12} < t_{k+1} - t_k$. The objective function values are then calculated and evaluated for two scenarios, over an 80-hour period.

Consider that the railway runs trains according to these scenarios. Trains are run between two origins (O_1 and O_2), a hub (H_1) and a destination (H_2), refer to the layout in Figure 4.1. The trains from the two origins travel to H_1 where they may be handled and compiled into a different length train, after which they travel to H_2. The containers are produced at each origin at a fixed production rate and are the same for both scenarios.

The distances are defined as: d_{11} from O_1 to H_1, d_{21} from O_2 to H_1 and d_{12} from H_1 to H_2.

The initial container values are chosen arbitrarily, while the other values are aligned with those from the corridor study data set, obtained from Transnet (a state-owned logistic operator) with permission.

Table 4.1 Parameter values

Parameter	Value
D_{Total1} [teu]	200
D_{Total2} [teu]	400
$w_1(t_0)$ [teu]	75
$w_2(t_0)$ [teu]	70
$w_{H1}(t_0)$ [teu]	120
$ORateP_1$ [teu]	28
$ORateP_2$ [teu]	28
d_{11} [km]	40
d_{21} [km]	50
m [ZAR]	50
d_{H1} [km]	500
m_{H1} [ZAR]	40
mO [ZAR]	60 000
ρ	1
LB_i	15
UB_i	40
LB_{12}	40
UB_{12}	50

The demand for transport of goods by these transport modes, depicted by D_{Total1} and D_{Total2}, to satisfy business requirements are specified. The train travels between the origins and H_1 at 30 km/h, while it travels at 40 km/h between hubs H_1 and H_2. The costs, m and m_{H1}, were determined using the cost structure of a railway, however the values used here are nominal but acceptable to protect confidential cost information. The overhead cost, mO, was similarly determined from examining the resultant costs of the system.

The train lengths in Eqns (8) and (9) are limited by the physical characteristics of the railway operation and are given.

In the Full train length scenario, sufficient wagons must be available for the predefined train length, or the train will miss its slot. Note that no optimisation takes place in the Full train length scenario, in this example or otherwise.

In the Slot priority scenario, shorter train lengths are manually chosen. Some slots may be missed due to other circumstances, (this may be attributable to other factors such as the containers not being available).

4.2.2 Numerical testing

Numerical testing is performed to further the understanding of how using the various scenarios will affect the volume throughput and cost, using nominal data for the inputs. This data is presented in Tables 4.2 and 4.3. The second column of Tables 4.2 and 4.3 present the variables and inputs. The first row presents the time slots, t_k, while the input and output values are given below each t_k, in the corresponding row.

The data in Table 4.2 represents a case where maximum length trains are run at every opportunity. In Table 4.3, the train lengths and departure decisions are taken manually, in both cases the mathematical model is tested but no optimisation takes place.

Wagons, (y_{i1}, $i \in \{1, 2\}$ and y_{12}), are allocated to each slot at the determined train length. In Table 4.2 this is the maximum train length and in Table 4.3 shorter train lengths are manually chosen. The train (x_{i1}, $i \in \{1, 2\}$ and x_{12}) is subject to a decision and will only depart if there are more containers than wagons available, and if the minimum number of wagons for a train are available. Tables 4.2 and 4.3 show trains and wagons which run from Origin$_1$ (O_1) and Origin$_2$ (O_2) to the Hub, and then the trains and wagons which run from the Hub to the destination. Costs are calculated for each slot.

For ease of understanding the calculation of $ORateP_{H1}(t_k)$ and $w_{H1}(t_k)$ is provided, in rows 11 and 12, respectively, for slot k = 4, 5, 6, 7 in Table 4.2. Note that $ORateP_{H1}(t_k) = 0$ if $l \notin$ S. Using Eqn (14):

$w_{H1}(t_4) = w_{H1}(t_3) - \rho x_{12}(t_3) y_{12}(t_3) + ORateP_{H1}(t_3)$

= $w_{H1}(t_3) - \rho x_{12}(t_3) y_{12}(t_3) + (\rho x_{11}(t_3).\, y_{11}(t_3) + \rho x_{21}(t_3).\, y_{21}(t_3))$

= 130 – 1x1x50 + (1x1x40 + 1x1x40)

= 160 (which is the result in row 12)

$w_{H1}(t_5) = w_{H1}(t_4) - \rho x_{12}(t_4) y_{12}(t_4) + ORateP_{H1}(t_4)$

= $w_{H1}(t_4) - \rho x_{12}(t_4) y_{12}(t_4) + (\rho x_{11}(t_4).\, y_{11}(t_4) + \rho x_{21}(t_4).\, y_{21}(t_4))$

= 160 – 1x1x50 + (1x1x40 + 2x1x40)

= 190 (which is the result in row 12)

$w_{H1}(t_6) = w_{H1}(t_5) - \rho x_{12}(t_5) y_{12}(t_5) + ORateP_{H1}(t_5)$

= $w_{H1}(t_5) - \rho x_{12}(t_5) y_{12}(t_5) + (\rho x_{11}(t_5).\, y_{11}(t_5) + \rho x_{21}(t_5).\, y_{21}(t_5))$

= 190 – 1x1x50 + (1x1x40 + 1x1x40)

= 220 (which is the result in row 12)

$w_{H1}(t_7) = w_{H1}(t_6) - \rho x_{12}(t_6) y_{12}(t_6) + ORateP_{H1}(t_6)$

= $w_{H1}(t_6) - \rho x_{12}(t_6) y_{12}(t_6) + (\rho x_{11}(t_6).\, y_{11}(t_6) + \rho x_{21}(t_6).\, y_{21}(t_6))$

= 220 – 1x1x50 + (1x1x40 + 1x1x40)

= 250 (which is the result in row 12)

Only one set, S, is required and hence $ORateP_i(t_k)$, k = 0,1,…l, is compliant as long as $\rho y_{i1}(t_k) \leq w_i(t_k)$ holds in the calculations.

For $ORateP_{H1}(t_k)$, $y_{12}(t_k) = 0$ if $k \neq l$. The example is based on using $x_{12}(t_k) = 1$ and $y_{12}(t_k) > 0$.

A similar calculation applies in Table 4.3, where the calculation for t_4 is presented below:

$w_{H1}(t_4) = w_{H1}(t_3) - \rho x_{12}(t_3) y_{12}(t_3) + ORateP_{H1}(t_3)$

= $w_{H1}(t_3) - \rho x_{12}(t_3) y_{12}(t_3) + \ (\rho x_{11}(t_3). y_{11}(t_3) + \rho x_{21}(t_3). y_{21}(t_3))$

= 125 – 1x1x50 + (1x1x24 + 1x1x32)

= 131 (which is the result in row 12)

Table 4.2 Full train length scenario

Row													Total
1	Time interval (k)	0	1	2	3	4	5	6	7	8	9	10	
2	Time (t_k)	08:00	16:00	00:00	08:00	16:00	00:00	08:00	16:00	00:00	08:00	16:00	
3	Containers ($w_1(t_k)$)	75	103	91	79	67	55	43	31	19	7	35	
4	Containers ($w_2(t_k)$)	70	98	86	74	62	50	38	26	14	2	30	
5	Production rate ($ORateP_1(t_k)$)	28	28	28	28	28	28	28	28	28	28	28	308[1]
6	Production rate ($ORateP_2(t_k)$)	28	28	28	28	28	28	28	28	28	28	28	308[2]
7	Wagons ($y_{11}(t_k)$)	0	40	40	40	40	40	40	40	40	0	40	616[3]
8	Trains ($x_{11}(t_k)$)	0	1	1	1	1	1	1	1	1	0	1	720[4]
9	Wagons ($y_{21}(t_k)$)	0	40	40	40	40	40	40	40	40	0	40	
10	Trains ($x_{21}(t_k)$)	0	1	1	1	1	1	1	1	1	0	1	
11	$ORateP_{H1}$	0	0	80	80	80	80	80	80	80	80	0	
12	Containers ($w_{H1}(t_k)$)	120	70	100	130	160	190	220	250	280	310	260	
13	Wagons ($y_{12}(t_k)$)	50	50	50	50	50	50	50	50	50	50	50	
14	Trains ($x_{12}(t_k)$)	1	1	1	1	1	1	1	1	1	1	1	
15	Volume	0	0	50	50	50	50	50	50	50	50	50	450[5]
16	Cost	1060000	1240000	1240000	1240000	1240000	1240000	1240000	1240000	1240000	1060000	1240000	13 280 000

Table 4.3 Slot priority scenario

Row													Total
1	Time interval (k)	0	1	2	3	4	5	6	7	8	9	10	
2	Time (t_k)	08:00	16:00	00:00	08:00	16:00	00:00	08:00	16:00	00:00	08:00	16:00	
3	Containers ($w_1(t_k)$)	75	79	83	83	87	91	95	99	103	107	111	
4	Containers ($w_2(t_k)$)	70	66	62	75	71	67	63	59	55	51	47	
5	Production rate ($ORateP_1(t_k)$)	28	28	28	28	28	28	28	28	28	28	28	308[1]
6	Production rate ($ORateP_2(t_k)$)	28	28	28	28	28	28	28	28	28	28	28	308[2]
7	Wagons ($y_{11}(t_k)$)	24	24	28	24	24	24	24	24	24	24	24	616[3]
8	Trains ($x_{11}(t_k)$)	1	1	1	1	1	1	1	1	1	1	1	603[4]
9	Wagons ($y_{21}(t_k)$)	32	32	15	32	32	32	32	32	32	32	32	
10	Trains ($x_{21}(t_k)$)	1	1	1	1	1	1	1	1	1	1	1	
11	$ORateP_{H1}$	0	56	56	43	56	56	56	56	56	56	56	
12	Containers ($w_{H1}(t_k)$)	120	126	132	125	131	137	143	149	155	161	167	
13	Wagons ($y_{12}(t_k)$)	50	50	50	50	50	50	50	50	50	50	50	
14	Trains ($x_{12}(t_k)$)	1	1	1	1	1	1	1	1	1	1	1	
15	Volume	0	0	50	50	50	50	50	50	50	50	50	450[5]
16	Cost	1188000	1188000	1153500	1188000	1188000	1188000	1188000	1188000	1188000	1188000	1188000	13 033 500

The numbered references presented in Table 4.2 and 4.3 are addressed by the following footnotes:

1 is Total $ORateP_1$

2 is Total $ORateP_2$

3 is Total $ORateP_1 + ORateP_2$

4 is Total wagons departing O_1 and O_2

5 is Total volume transported to destination, H_2

The cost at each t_k is calculated from Eqn (3), where the total wagons travelling from the origins to H_1 are multiplied by m and the total wagons travelling from H_1 to H_2 are multiplied by m_{H1}, and the overhead cost, mO, applies only when a train does not run.

4.2.3 Satisfaction of constraints for the full train length scenario

Table 4.2 presents the data for the Full train length scenario. Data is generated for 10 time slots with two origins, a hub and a destination. Containers (w_1 and w_2) are stacked at both origins, and they are produced at a production rate, ($ORateP_1$ and $ORateP_2$), at each origin and indicated in the table.

The constraints have been satisfied for t_0, t_1,…t_{10} in Table 4.2 as follows:

The constraints (5) and (7) are fully satisfied using the data in rows 7 to 10 and rows 13 to 14, respectively. This can be checked taking the values under each k and t_k.

Constraints (4) and (6) are satisfied using the data in rows 3, 4, 7 and 9. In row 7 and 9, Eqn (4) has been satisfied for $y_{i1}(t_k)$, $w_i(t_k)$, i = 1, 2 when compared to row 3 and 4. For example for t_1, $y_{11}(t_1)$=40 and $w_1(t_1)$=103; $y_{21}(t_1)$=40 and $w_2(t_1)$=98 and for t_8, $y_{11}(t_8)$=40 and $w_1(t_8)$=19; $y_{21}(t_8)$=40 and $w_2(t_8)$=14.

Eqn (6) has been satisfied, using data in row 15, for each t_k the wagons which can be used in a train are restricted by the number of containers available. For example, for t_1, $y_{12}(t_1)$ = 50 and $w_{H1}(t_1)$ = 70 and for t_8, $y_{12}(t_8)$ = 50 and $w_{H1}(t_8)$ = 280.

Eqn (7) has been satisfied, using data in row 14, for $x_{12}(t_k)$, $y_{12}(t_k)$ by checking row 13, no wagons are assigned if no train runs in any time slot.

The result for Eqn (12) is displayed in row 11, using the train lengths and departure decisions from rows 7 to 10, also satisfying Constraint (13).

Constraint (14) has been satisfied using data in rows 13 and 14 for $y_{12}(t_k)$ and $x_{12}(t_k)$.

Constraint (15) has been satisfied using data in row 7 and 9.

Constraint (16) has been satisfied using data in row 13.

The values of $w_i(t_k)$, i = 1, 2; k = 0,1,…10, have been calculated in rows 3 and 4 using Eqn (18). For example, for t_2:

$$w_1(t_2) = w_1(t_1) - \rho x_{11}(t_1)y_{11}(t_1) + ORateP_1(t_2)$$

= 103 – 1x1x40 + 28

= 91 (which is the result in row 3)

The parameters $w_{H1}(t_k)$, k = 0,1,…10, have been calculated in row 12 using Eqn (11) and (12), incorporating data from row 11. For example, for t_2:

$$w_{H1}(t_2) = w_{H1}(t_1) - \rho x_{12}(t_1)y_{12}(t_1) + ORateP_{H1}(t_1)$$

= 70 – 1x1x50 + (1x1x40 + 1x1x40)

= 100 (which is the result in row 12 at t_2)

Constraint (13) has been satisfied using data in rows 7, 8, 9 and 10. $\sum_k \sum_i x_{i1}(t_k) y_{i1}(t_k)$ = 720, while D_{Total1} = 200. Constraint (14) has been satisfied using data in row 15. $\rho \sum_k x_{12}(t_{k-2}) y_{12}(t_{k-2})$ = 450. The parameter t_{k-2} is used because t_{12} = 15.75 h and the period $t_k - t_{k-1}$ = 8 h.

The value of the cost objective function is R 13 280 000, calculated in row 16 over the whole period.

4.2.4 Satisfaction of constraints for the Slot priority scenario

Table 4.3 presents the results for the Slot priority scenario. Data is generated for 10 time slots with two origins, a hub and a destination. Containers (w_2 and w_3) are stacked at both origins and they are produced at a production rate, ($ORateP_1$ and $ORateP_2$), at each origin and indicated in the table, these parameters are the same as the Full train length scenario indicated in Table 4.2, hence the starting conditions for both scenarios are identical.

In Table 4.3 the container levels that are available for departure are calculated (consider production and containers which have left on previous trains) and the volume of containers that arrive in each slot at the destination. Note that the containers have an initial value of 75 and 70. Note that the spreadsheet checks if sufficient containers are available (at the origins and at the hub) for the required train length and that a train cannot depart if it is not complete. All constraints of the mathematical formulation are satisfied in running trains.

The objective function is set up to calculate costs as per Section 4.1. Train lengths that are less than the maximum are chosen manually. In this manner trains are run through the system while ensuring that the constraints and parameters are used as defined. It is intended to choose wagons (y_{i1}, $i \in \{1, 2,\}$ and y_{12}) and trains (x_{i1}, $i \in \{1, 2\}$ and x_{12}) to calculate the cost objective function manually to test the mathematical formulation. Wagons are allocated to each slot at the chosen train length. The train (x_{i1}, $i \in \{1, 2\}$ and x_{12}) is subject to a decision and will only depart if there are sufficient containers for the required wagons available, and if the number of wagons for a train are available.

Table 4.3 shows trains and wagons which run from $Origin_1$ and $Origin_2$ to the Hub, H_1, and then the trains and wagons which run from the Hub to the destination. Costs are calculated for each slot.

The constraints have been satisfied in Table 4.3 as follows:

The constraints (5) and (7) are fully satisfied using the data in rows 7 to 10 and rows 13 and 14, respectively.

Constraints (4) and (6) are satisfied using the data in rows 7 and 9. Eqn (4) has been satisfied for $y_{i1}(t_k)$, $w_i(t_k)$, i = 1, 2. For example for t_2, $y_{11}(t_2) = 28$ and $w_1(t_2) = 83$; $y_{21}(t_2) = 15$ and $w_2(t_2) = 62$ and for t_8, $y_{11}(t_8) = 24$ and $w_1(t_8) = 103$; $y_{21}(t_8) = 32$ and $w_2(t_8) = 55$.

In row 13, Eqn (6) has been satisfied, using data for w_{H1} in row 12. For example, for t_2, $y_{12}(t_2) = 50$ and $w_{H1}(t_2) = 132$ and for t_8, $y_{12}(t_8) = 50$ and $w_{H1}(t_8) = 155$.

In row 13 and 14, Eqn (2) has been satisfied.

In row 14, Eqn (7) has been satisfied, using data in row 13, no wagons are assigned if no train runs in any time slot.

The result for Eqn (12) is displayed in row 11, using the train lengths and departure decisions from rows 7 to 10, also satisfying Constraint (13).

Constraint (14) has been satisfied using data in rows 13 and 14 for $y_{12}(t_k)$ and $x_{12}(t_k)$.

Constraint (15) has been satisfied using data in rows 7 and 9 for $y_{i1}(t_k)$, i = 1, 2.

Constraint (16) has been satisfied using data in row 13.

The parameters $w_i(t_k)$, i = 1, 2, k=0,1,…10, have been calculated in rows 3 and 4 using Eqn (14). For example, for t_2:

$$w_1(t_2) = w_1(t_1) - \rho x_{11}(t_1) y_{11}(t_1) + ORateP_1(t_2)$$

= 79 – 1x1x24 + 28

= 83 (which is the result in row 3)

The parameters $w_{H1}(t_k)$, k = 0,1,…10, have been calculated in row 12 using Eqn (11), incorporating data from row 11. For example, for t_2:

$$w_{H1}(t_2) = w_{H1}(t_1) - \rho x_{12}(t_1) y_{12}(t_1) + ORateP_{H1}(t_1)$$

= 126 – 1x1x50 + (1x1x24 + 1x1x32)

= 132 (which is the result in row 12)

Constraint (13) has been satisfied using data in rows 7, 8, 9 and 10. $\sum_k \sum_i x_{i1}(t_k) y_{i1}(t_k)$ = 603, while D_{Total1} = 200. Constraint (14) has been satisfied in row 14, $\rho \sum_k x_{12}(t_{k-2}) y_{12}(t_{k-2})$ = 450.

The value of the objective function is R 13 033 500 calculated in row 16 over the whole period using Eqn (3).

4.2.5 Operations strategy comparison

The important finding from this testing of the mathematical model was that the Full train length scenario delivered 450 containers at an operating cost of R 13 280 000 (thirteen million, two hundred and eighty thousand), whereas the Slot priority scenario delivered 450 containers at an operating cost of R 13 033 500 (thirteen million, thirty three thousand and five hundred). These findings are summarised in Table 4.4.

Table 4.4 Scenario comparison summary

Scenario	Containers	Cost [R]
Full train length	450	13 280 000
Slot priority	450	13 033 500

In summary, all the constraints were satisfied. Table 4.4 indicates that the Full train length scenario incurred a higher cost than the Slot priority scenario while delivering the same quantity of containers at the destination.

There is an optimisation opportunity in finding a way for these scenarios to operate. The scenarios presented above illustrate the reason for considering alternative train departure strategies. The preferred strategy will be an output of the study, derived from analysing the data.

Further study is necessary using more sophisticated techniques to understand how these scenarios are affected by the balancing of rolling stock in the system, the effect of external factors and the layout of the network.

The mathematical model describes how the system works and defines the cost objective function required to optimise it. However, it must be supplied with certain information about external factors to give a complete answer. The model can perform calculations and cost optimisation for the movement of freight through such a system to find the optimal train length in each sector for the minimum cost. The validity of the model has been interrogated, testing each constraint and ensuring that the results satisfy requirements.

Furthermore, the properties of physical assets are being evaluated in a value free manner, i.e. they are independent of any possible bias which could be introduced by the researcher. The study is conducted from a deduction standpoint, meaning that data is being evaluated to either confirm or deny the supporting hypothesis.

In the following chapter, the description of a large, complex problem and the solution thereof using the model are presented.

Chapter Five

South African Corridor Application

The application is expanded to address more complex problems and used to solve the South African corridor example in this Chapter. A description of the genetic algorithm configuration for this purpose is presented.

5.1 Description of problem

Mathematically, the South African corridor application is the same problem which was described in Chapter 4. Here a container corridor in South Africa has been examined. The problem was extended to work over a longer time period to allow a realistic situation to develop. The model was set up with pertinent data for the corridor, so that the logistic operation could be examined over a period of one week. Three train slots are available every day. Three slots per day is a realistic set up for a single flow consideration. Cost optimisation was performed on this application, by manipulating the train lengths and the number of trains. The parameters to deliver this service are based on the actual container volumes required and the layout and characteristics of the corridor, shown in Figure 5.1.

The Port of Durban is located on the eastern seaboard of South Africa. It is a major general freight interface for the African continent. The port is connected to the interior of the country by a major freight railway line (consisting of electrified, double line, narrow gauge track, considered a high-capacity line suitable for container traffic) and a highway for trucks. The interior, notably the Province of Gauteng, is a major economic centre and located along a north-south corridor connected to other African countries to the north. City Deep, Hub1, is a major Hub in the province of Gauteng, where modal interface between road and rail takes place, and freight is handled, and the movements coordinated between multiple origins and destinations.

Figure 5.1 Location of facilities in the South African model application

The service considered in this case transports containers from two production facilities, at Pretoria (Origin1) and Pyramid (Origin2), located around the economic area of Gauteng to the coast (at the Port of Durban (Hub2)) in 50 wagon trains (Upper Bound, UB_{12}). The train service also returns from the port carrying containers imported by ship.

The boundaries of this problem are the origin nodes and the port node. The problem is bounded by the Pretoria and Pyramid origins in the interior of the country and by the Port of Durban at the coast. The portion of the corridor modelled consists of a railway track connecting the Port of Durban, and the hinterland of South Africa. A container hub and rail links to the customers are used in this case.

5.2 Application of mathematical model

The mathematical model, Eqns (3) to (15), is developed to minimise cost. This is done by deciding how long the trains should be and how many will run over a period of a week, given a slot schedule which denotes available slots. The inputs are the distances between the origins, hub and destination, as well as the volumes of the goods to be transported. The production centres are denoted as Origin1 and Origin2 in the model, while the hubs are located at City Deep (Hub1) and the Port of Durban (Hub2, located at the destination). For the model, details of actual specifications are required, viz., distance and production rate. The production rate is the rate at which goods are containerised at each of the production centres or factories, see $ORateP_1$ and $ORateP_2$ in Table 5.1. These factors are used as inputs when the train service is designed to meet the customers' requirements.

Two origins are considered, one in Pretoria (Origin1) and one in Pyramid (Origin2). Freight from these places will be routed on trains to the City Deep container terminal (Hub1), and then on to the Port of Durban (Hub2). This is a hub-to-hub journey, and the destination is also denoted as a hub because economies of scale are realised on this main line journey. City Deep is a container hub complex for the railway, located in the south of the City of Johannesburg, Gauteng. At City Deep, road trucks and trains with freight originating and departing the region may be compiled. The trains will also be returning from the port with containers that are stacked at the port, destined for the hinterland of the country.

Referring to Figure 5.1, at the origins, goods are produced, which are packed in containers at these premises. The containers must be transported to the Port of Durban from where they will be exported. In line with the current practice, the train service was designed with a first mile train journey from the customer to the City Deep container terminal, this train is 40 wagons long (Upper Bound, UB_i). At City Deep the wagons are compiled into 50 wagon trains (Upper Bound, UB_{12}). This train then awaits a departure slot on the main line and departs on the container corridor to Durban, where it arrives at the Kings Rest terminal (Hub2). The Kings Rest terminal serves as a hub located at the Port. The containers are offloaded and stacked at the Port, to await departure on a ship. At the time when the containers are offloaded from the wagons, the wagons may return to the interior of the country, loaded or unloaded as the railway planned. Therefore, these containers are not considered anymore as the focus is on the logistic system itself, at this point the containers move outside the boundary of this study.

Furthermore, the maximum train lengths mentioned above are decided based on various constraints, including train technology, track infrastructure and operational limitations. These limitations are not treated any further in this study, here the only concern is the maximum train length which is given and the optimised train length which shall be determined. As there is a variable cost per container and a fixed overhead cost per slot, it will not be cost effective both to run longer trains and to leave available slots unutilised, hence the need for optimisation.

The full description of the setup parameters for the South African corridor problem is presented in Table 5.1. Here the demand, initial container stack quantities, production rates, distances, costs and container density are shown in the first column, followed by the values.

Table 5.1 Corridor problem input parameters

Parameter	Value
D_{Total1} [teu]	200
D_{Total2} [teu]	400
$w_1(t_0)$ [teu]	75
$w_2(t_0)$ [teu]	70
$w_{H1}(t_0)$ [teu]	120
$ORateP_1$ [teu]	28
$ORateP_2$ [teu]	28
d_{11} [km]	170
d_{21} [km]	158
m [ZAR]	4
d_{H1} [km]	630
m_{H1} [ZAR]	10
mO [ZAR]	30 000
ρ	2
LB_i	15
UB_i	40
LB_{12}	40
UB_{12}	50

The problem consists of six optimisation variables: y_{11}, x_{11}, y_{21}, x_{21}, y_{12}, x_{12}, where the x variables are binary, and the y variables are integer. Hence, in the description of the problem, there are six variables, and the total dimension is determined by the number of decision points. The demand requirement, $D_{Total2} > D_{Total1}$, as the accumulation of containers at the final destination represents the total volume throughput of the system.

The timeline, depicted in Figure 5.2, runs for 21 time points, so chosen because it

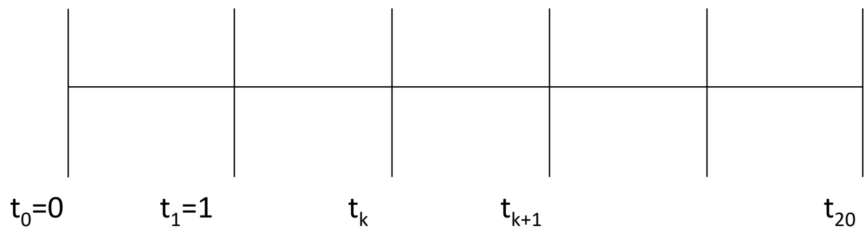


Figure 5.2 Timeline representation for South African corridor problem

is sufficient time for a railway logistic system to reach a steady state over a week. In this case, consider that three trains run in every 24-hour period, three per day. Thus, 20 periods are equivalent to a week. The problem is set up in this way because the logistic system configuration is considered and not other factors such as the railway capacity, etc. As before, in Chapter 4, decisions are taken at time t_k, $k = 0, 1,2,\ldots 20$, resulting in the total number of decision variables being 126: $y_{11}(t_k)$, $x_{11}(t_k)$, $y_{21}(t_k)$, $x_{21}(t_k)$, $y_{12}(t_k)$, $x_{12}(t_k)$.

Apart from the input presented in Table 5.1, other inputs are distances presented in Table 5.2.

Table 5.2 The distance matrix between Origins and Hubs

	H1 (City Deep)	**H2 (Durban)**	**Origin1 (Pretoria)**	**Origin2 (Pyramid)**
Origin1 (O_1)	170	-	-	-
Origin2 (O_2)	158	-	-	-
Hub1 (H_1)	-	630	170	158
Hub2 (H_2)	630	-	-	-

Table 5.2 presents the relation between the various origin and hub nodes. The notation '-' indicates that the distance is not applicable. Table 5.2 presents the distance between the origin-hub and hub-hub pairs in kilometres. Train lengths

are limited between LB_i and UB_i wagons on the origin to hub journey, and between LB_{12} and UB_{12} wagons on the main line journey between the hubs (the values of which are given in Table 5.1).

In this study, the transport of freight is first minimised by choosing the variable values to optimise the cost presented in Eqn (3) subject to the Constraints (4) – (15).

The movement of trains and freight is modelled, and the optimisation is performed based on the cost function presented in Section 4.1. The model finds the optimum train lengths and number of trains. The value of the optimisation variables $y_{11}(t_k)$, $x_{11}(t_k)$, $y_{21}(t_k)$, $x_{21}(t_k)$, $y_{12}(t_k)$ and $x_{12}(t_k)$ determine the optimal train lengths and number of trains.

5.3 Description of Genetic Algorithm

The Genetic Algorithm (GA) solution technique has been chosen for this application. The application of the GA to this specific problem is presented in this chapter.

5.3.1 Overview

Genetic algorithms are a class of biologically inspired techniques for solving optimisation problems (Harada & Alba, 2020). These algorithms are loosely based on Darwin's theory of evolution. They can be used in a wide range of applications across varied industries and disciplines and can address problems where the functions are non-differentiable, non-continuous, ill-defined, comprise arbitrary types of variables, could have competing goals and may not even have any analytical expression. The fact that the problem under consideration in this study has both binary and integer variable types makes the GA a suitable solution technique.

Genetic algorithms find applications in multiple disciplines, such as ecology, evolution, and mechanical system fault identification, solving optimisation problems beyond analytical techniques, with evolving behaviour in models, and simulating co-evolutionary processes, (Hamblin, 2013).

Genetic algorithms can be used to find global optima, making them often selected as a solution technique, (Martowibowo & Damanik, 2021). The author of this paper considered the machining of parts and the trade-off against surface roughness and the tool characteristics using experimental results.

Hamblin (2013) explored the application of genetic algorithms, finding that they combined exploration (mutation and crossover) with exploitation (selection and replacement) in the process of increasing the occurrence of good solutions in a population. Exploration refers to moving around the search space, while exploitation refers to converging on a minimum (or maximum). Two point crossover was preferred over one point crossover as cases where good combinations of alleles are further apart on the chromosome are more likely to be disrupted. There were preferences for mutation and crossover rates, however any selection that worked for the problem at hand was deemed valid.

In summary, genetic algorithms provide a powerful and flexible optimisation approach by leveraging natural selection and genetics. The approaches discussed demonstrate the adaptability and effectiveness of genetic algorithms across a wide range of applications. The set up and parameter selection for the genetic algorithm is important and will affect the performance and solution quality.

5.3.2 Application of Genetic Algorithm

A decision is motivated to use a GA to solve the hub and spoke problem of Chapter 4. The genetic algorithm is then tailored to solve this problem, developing chromosomes with binary conditions for train departure decisions and integer conditions for train lengths for each section of the journey. A solution protocol is then described using double point crossover, elitism and mutation.

The genetic algorithm was implemented to solve the simple example problem of Section 4.2 first. It was then extended to deal with this complete chromosome for the main problem.

The script to examine the hub and spoke network of the South African corridor, depicted in Figure 5.1, was coded in MATLAB® based on the mathematical formulation developed in Section 4.1. The problem was optimised using a genetic algorithm where a chromosome consisted of the full specification of a solution.

Our problem is unique in its structure and application as the problem itself is a novel formulation. It required a bespoke genetic algorithm developed to solve a problem of its structure with both integer and binary variables, the development of which is also a novel contribution.

The genetic algorithm was chosen to solve this problem because it can address problems where the functions are non-differentiable, non-continuous, comprise arbitrary types of variables, could have competing goals and may not even have any analytical expression. While this problem does have an analytical expression and has been well defined, it is still a challenging problem with one set of variables being binary and one set of variables being integer.

A genetic algorithm manages many tentative solutions. An initial population is constructed that may be random or may use some heuristics. These solutions will be evaluated based on a fitness function, such as the one in Eqn (19), Section 4.1, and tested for how well they fit the target problem. New tentative solutions are generated at each iteration/generation through a variation process, employing techniques such as selection, crossover and mutation. The next population of tentative solutions are constructed from the current population and new tentative solutions. The process is repeated until predetermined termination criteria are satisfied.

Some terminology used in GA's is introduced here, the specific application thereof follows. Genomes are sets containing characteristics of the problem. Selection and replacement refers to the process of choosing the more favorable aspects of the genome as it relates to the problem. The Crossover technique produces children from two parent solutions, it is possible that the child will be a better offspring. The fitness landscape refers to the set of solutions and how well they address the problem requirements.

Selection and replacement rates are chosen with consideration for the tradeoff between exporation and exploitation. Lower values (considered here as below 30%) of each will prioritise exploration (because they inhibit convergence), while higher values will prioritise convergence (at risk of not fully exploring the search space), (Hamblin, 2013). It is valuable in deciding on these tradeoffs to consider if the fitness landscape is easy or hard. In a hard landscape, higher selection rates will converge faster, but are at risk of missing the best solution.

Selection

Selection refers to a process of choosing fitted pairs of chromosomes which can be used to achieve a more desired outcome.

Double Point Crossover

The Crossover technique produces a child (or more) from two parent solutions, it is possible that the child will be a better offspring.

Elitism

When elitism is implemented, a chosen percentage of the fittest individuals are selected for the next generation.

Mutation

Mutation involves adding or subtracting a small step to the chromosome. As a general rule a good starting point for mutation rates was found as 1/l, where l is the length of the genome.

Search termination

The termination conditions are as follows, (Sivanandam & Deepa, 2008):

- Maximum generations (used in the model)
- Elapsed time
- No change in fitness
- Stall generations
- Stall time limit

These termination conditions will end the GA process, care must be taken that they do not result in a premature termination.

The process is expressed in the pseudo-code example, (Harada & Alba, 2020):

Initialization(P_0)

Evaluation(P_0)

While Termination criterion do

1. P_t' ←variation(P_t)
2. Evaluate(P_t')
3. P_{t+1} ←replacement(P_t,P_t')

//merge the new and old population and select a new population

Here P_t denotes the population of solutions at iteration (or generation), t. P'_t denotes the new population determined from P_t. In 1. P'_t is calculated using selection, crossover and mutation. In 2. the fitness of solutions in P'_t are found using Eqn (19). In 3. the next P_{t+1} is calculated from P'_t and P_t (elitism), this process flow is demonstrated in Figure 5.3.

Genetic algorithms will be used to find the train lengths resulting in the minimum cost for the operating environment in this study. Therefore, they have been chosen to find the optimal train length in each of the runs, for those specific conditions.

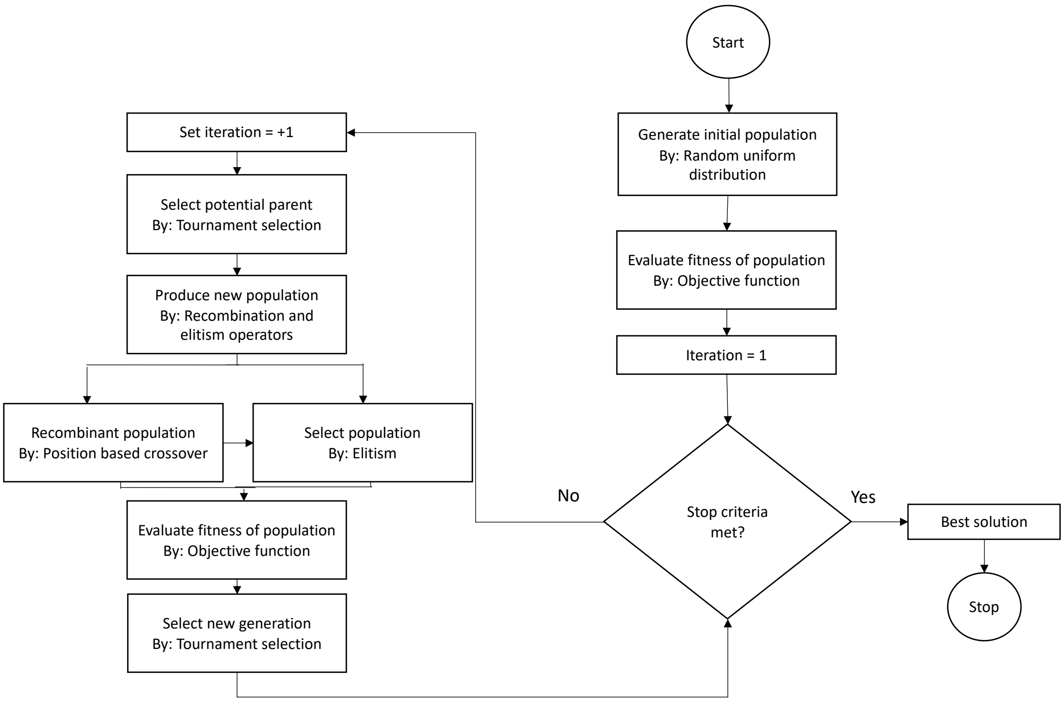


Figure 5.3 Process flow of genetic algorithm (Jafarian-Moghaddam, 2021)

Genetic algorithms can be configured in various ways, selecting and configuring the tools available in the algorithm. One example of this follows the process flow, shown in Figure 5.3, (Jafarian-Moghaddam, 2021). Therefore, each run will consist of multiple calculations and iterations in the process of finding an optimal train length at each time point, t_k, provided that a train departs at t_k.

Before solving problem (3) – (15) using GA, the implementation details for the GA need to be stated. For this problem, a penalised Objective Function Eqn (17) has been defined which has several penalty parameters. Hence, instead of optimising Eqn (3), Eqn (17) shall be optimised by choosing values of the penalty parameters R_1, R_2, R_3, R_4, R_5 and R_6 and of the variables. It is however well known that the penalty parameters are sensitive and that their values are difficult to determine. Extensive numerical studies have been conducted to determine these parameter values. Apart from this, the GA has been implemented using Crossover and Elitism as Mutation has not proved effective. Tournament selection of the parent solution has been used to select pairs on which to apply Crossover. A solution of the problem (i.e. a chromosome) has two separate parts, one consisting of binary variables and the other consisting of integer variables. A Crossover has been conducted separately on them.

The development of genetic algorithms as a solution technique is described in the next few sections. A detailed description of the implementation is provided below.

5.3.3 Search space

All feasible solutions for a problem exist in the search space, where each point is a possible solution. The genetic algorithm technique searches this space for the best solution (Sivanandam & Deepa, 2008). In this case, the search space is defined by (these parameters are defined in Section 4.1):

$$x_{i1}(t_k) = \begin{cases} 1, & if\ a\ train\ departs\ in\ origin\ at\ t_k, \\ 0, & if\ a\ train\ does\ not\ depart, \end{cases} \qquad \text{for all } i, k,$$

$$x_{12}(t_k) = \begin{cases} 1, & if\ a\ train\ departs\ in\ H1\ at\ t_k, \\ 0, & if\ a\ train\ does\ not\ depart, \end{cases} \qquad \text{for all } k,$$

$15 <= y_{i1}(t_k) <= 40$, for all *i*, *k,*

$40 <= y_{12}(t_k) <= 50$, for all *k.*

The initial population has been randomly generated from the search space.

5.3.4 Population

The population consists of several individuals or chromosomes being tested. The size of this population depends on the complexity of the problem. Ideally, the population should start as large as possible so that the whole search space may be explored. While a large population is useful, it does require more computational effort to converge. Population sizes of around 100 individuals are used practically for most problems, (Sivanandam & Deepa, 2008). Experimentation indicated that a population of 50 chromosomes was an acceptable starting point. The initial population was generated randomly where the chromosome includes the train lengths and train numbers. The genetic structure is depicted in Table 5.3 (using the nomenclature developed in Section 4.1).

Table 5.3 contains the initial p solutions, P_0 = {X_0^1, X_0^2, …. X_0^p}, where for each solution (column) the first 63 (rows) are binary and the second 63 rows are integer variables. Note that the MATLAB® code was developed for a simpler chromosome (used in the simple example problem of Section 4.2) first. It was then extended to deal with this complete chromosome with considerably more complexity and being more difficult to solve and satisfy the constraints.

The problem relates to the development of a model. This model is tested on a full-scale example in Section 5.1. As the goal was to develop the model, the corridor example implementation is deemed sufficient, it is not necessary to test the implementation thereof on multiple systems.

Table 5.3 Genetic structure

X_0^1	X_0^2	....	X_0^p
$x_{11}(t_0)$	$x_{11}(t_0)$		$x_{11}(t_0)$
⋮	⋮		⋮
$x_{11}(t_{20})$	$x_{11}(t_{20})$		$x_{11}(t_{20})$
$x_{21}(t_0)$	$x_{21}(t_0)$		$x_{21}(t_0)$
⋮	⋮		⋮
$x_{21}(t_{20})$	$x_{21}(t_{20})$		$x_{21}(t_{20})$
$x_{12}(t_0)$	$x_{12}(t_0)$		$x_{12}(t_0)$
⋮	⋮		⋮
$x_{12}(t_{20})$	$x_{12}(t_{20})$		$x_{12}(t_{20})$
$y_{11}(t_0)$	$y_{11}(t_0)$		$y_{11}(t_0)$
⋮	⋮		⋮
$y_{11}(t_{20})$	$y_{11}(t_{20})$		$y_{11}(t_{20})$
$y_{21}(t_0)$	$y_{21}(t_0)$		$y_{21}(t_0)$
⋮	⋮		⋮
$y_{21}(t_{20})$	$y_{21}(t_{20})$		$y_{21}(t_{20})$
$y_{12}(t_0)$	$y_{12}(t_0)$		$y_{12}(t_0)$
⋮	⋮		⋮
$y_{12}(t_{20})$	$y_{12}(t_{20})$		$y_{12}(t_{20})$

The first row, $P_0 = \{X_0^1, X_0^2, \ldots X_0^p\}$, denotes the initial population. The population at the t-th iteration is denoted as $P_t = \{X_t^1, X_t^2, \ldots X_t^p\}$. In Table 5.3 each column represents a chromosome of 126 optimisation variables. The *x* parameters are the train decision variables, and the *y* parameters are the train length variables. A population of *p* chromosomes is generated randomly. In the table each column represents a chromosome X_0^1 and this extends to *p* chromosomes in the population, i.e. the full table has *p* columns.

Each chromosome in the initial population P_0 in Table 5.3 has been generated randomly within the search space given in Section 5.3.3.

The first 63 values in each column are generated by flipping a coin and assigning a 1 or a 0. For example, for X_0^1, a random number generator is used, restricted to 1 or 0. Then assign $x_{11}(t_k)$ = 1 or $x_{11}(t_k)$ = 0 with equal probability. This process was followed for all binary variables. For integer values, the values were generated within the given lower and upper integer value. For example, for 15 <=$y_{i1}(t_k)$<= 40, a random number generator was used to assign an integer between the values of 15 and 40 to $y_{i1}(t_k)$. Similarly, a random integer was assigned between the values of 40 and 50 to $y_{12}(t_k)$ for all *k*.

5.3.5 Chromosome

In a genetic algorithm, the population of solutions for a problem may be specified with chromosomes defined to store the relevant characteristics of individuals in the population. Here, the algorithm is run over 21 decision time points and therefore each chromosome contains 21 variables for each of the x and y variables. Three of the variables, train decision (x) are binary and the other three, train length (y), are integer, for a total of 126 variables. The algorithm follows an evolutionary technique to find the best solution. The length of each chromosome is 126 variables which are generated randomly, initially. At each iteration after the initial P_0, e.g. at iteration P_t, each chromosome is generated using genetic operations such as selection, crossover and elitism using P_{t-1}. These are expanded below.

5.3.6 Selection

At the t-th iteration the GA has population P_t = {X_t^1, X_t^2, …. X_t^p}, where the corresponding values are chosen, using the crossover and elitism techniques. These variable values are used to calculate fitness values using Eqn (17). Hence, corresponding to P_t a fitness string {F(X_t^1), F(X_t^2), …. F(X_t^p)} exists. Tournament selection has been used to select a pair of chromosomes from P_t. The process is as follows: select two parents at random (n1 and n2), and choose the best of the two, X_t^{n1} and X_t^{n2}, by comparing F(X_t^{n1}) and F(X_t^{n2}). Then denote this solution as X^i, and repeat the process and select X^j. The pair (X^i, X^j) is then used for Crossover, where this pair generates a new pair of chromosomes. For a population p = 50, calculate 20 such pairs to generate 40 new chromosomes for P_{t+1}. The Crossover is robust due to the presence of binary and integer variables, hence a 20% Elitism (retention rate) was used.

Population P_{t+1} is generated from P_t using a double point crossover technique on the binary variables x_{11}, x_{21} and x_{12} and a real coded crossover on the integer variables y_{11}, y_{21} and y_{12}. By using Crossover 40 new chromosomes have been generated in P_{t+1}, a 20% retention rate of the best chromosomes was used, this is known as elitism.

The use of mutation with a rate of 10^{-3} hinders the convergence and constraints satisfaction in a series of runs that were tested, considering also the size of the chromosome. Hence, a decision was taken not to use the mutation.

Table 5.4 Parent chromosomes for crossover

X_t^i	X_t^j
$x_{11}(t_0)$	$x_{11}(t_0)$
⋮	⋮
$x_{11}(t_{20})$	$x_{11}(t_{20})$
$x_{21}(t_0)$	$x_{21}(t_0)$
⋮	⋮
$x_{21}(t_{20})$	$x_{21}(t_{20})$
$x_{12}(t_0)$	$x_{12}(t_0)$
⋮	⋮
$x_{12}(t_{20})$	$x_{12}(t_{20})$
$y_{11}(t_0)$	$y_{11}(t_0)$
⋮	⋮
$y_{11}(t_{20})$	$y_{11}(t_{20})$
$y_{21}(t_0)$	$y_{21}(t_0)$
⋮	⋮
$y_{21}(t_{20})$	$y_{21}(t_{20})$
$y_{12}(t_0)$	$y_{12}(t_0)$
⋮	⋮
$y_{12}(t_{20})$	$y_{12}(t_{20})$

Table 5.5 Sample crossover illustration

	Binary							Integer									
X_t^i	0	1	0	0	0	0	0	33	23	31	28	18	19	18	16	40	23
X_t^j	0	0	1	0	0	0	1	40	15	39	27	38	26	19	26	15	33
Of1	0	1	1	0	0	0	0	33	23	31	28	18	19	18	20	15	23
Of2	0	0	0	0	0	0	1	40	15	39	27	38	26	19	22	40	33

The parent chromosomes are presented in Table 5.4. It is not possible to consider the full-length chromosome of 126 variables here, hence only the concept is presented here. The crossover is illustrated in Table 5.5, where a sample of the chromosomes are presented in the rows. The chromosomes are labelled as X_t^i

(Parent1), X_t^j (Parent2), Of1 (Offspring1) and Of2 (Offspring2) for the binary and integer sections. These sections of the chromosomes are as labelled in the first row. For each section (binary and integer), two random crossover points are chosen. The values for X_t^i and X_t^j are swapped between these two points, to generate the offspring, Of1 and Of2. In the binary part of the crossover section, two integer locations are randomly generated again using the rule, 1 + floor (random x I), where I is the total length of the binary part of the chromosome and random $\in$ (0, 1). Then the double point Crossover is performed from the lower integer location to the higher integer location of the binary part.

For the integer part of the crossover section two integer locations are generated within the integer section. For each integer component, k between the two integer locations, a random number, Z_k, was generated in (-1/2, 3/2) as follows, $Z_k = -1/2 + (2 \times r_k)$, where r_k is a random number in (0,1). The following two operations are performed for two new offspring:

- Integer section of Parent 1 x Z_k + Integer section of Parent 2 x $(1-Z_k)$
- Integer section of Parent 1 x $(1-Z_k)$ + Integer section of Parent 2 x Z_k

where Z_k is generated for each component, k, as described below.

As an example, for the highlighted terms in Table 5.5, for 19 and 18 (where r_1 is randomly generated as 0.05 and Z_1 is -0.40):

$19 \times Z_1 + 18 \times (1-Z_1) = 18$,
$19 \times (1-Z_1) + 18 \times Z_1 = 19$,
where non-integer values are truncated to integers.

For 26 and 16, the following calculation is performed (where r_2 is 0.45 and Z_2 is 0.40):

$26 \times Z_2 + 16 \times (1-Z_2) = 20$,

$26 \times (1-Z_2) + 16 \times Z_2 = 22$.

For 15 and 40, the following calculation is performed (where r_3 is 0.9 and Z_3 is -1.3):

$15 \times Z_3 + 40 \times (1-Z_3) = 8$,

$15 \times (1-Z_3) + 40 \times Z_3 = 48$.

In this case, note that the results are outside of the range defined in Section 5.3.3, therefore the results are capped at 15 for the lower bound and 40 for the upper bound, as defined by Eqn (8) in Section 4.1 and referring to Table 5.1, where it was considered that the corresponding chromosome location corresponds to the integer variable y_{i1}.

Similar calculations are performed for each component that falls under crossover. This will generate integers both smaller and larger than those that are in X_t^i and X_t^j. This allows the resulting offspring to have a full representation of all integer values in the search space. This was possible due to the use of $Z_k \in (-1/2, 3/2)$. Here a clipping technique was used to keep the integer value within the given range. Non-integer values are truncated to integer values.

In Table 5.5 the values which have been swapped are colored grey for illustration.

Elitism

When calculating Population P_{t+1} from P_t Crossover and Elitism were used. For an example where $p = 50$, by using Crossover 40 new chromosomes were introduced, but the population here has 50 chromosomes. The remaining chromosomes are taken from P_t. The 10 best solutions (20% elitism) from P_t were used, copying them into P_{t+1}. This gives 50 chromosomes in P_{t+1}, which are an improvement on P_t on average.

5.3.7 Search termination

The termination conditions were limited by the maximum generations.

These termination conditions will end the GA process, care must be taken that they do not result in a premature termination.

The model has been applied to the working of a South African freight corridor. In this chapter the setup and application of the genetic algorithm to find a good solution to the problem is demonstrated.

Chapter Six
Numerical Results

Numerical data for the mathematical model is presented, first for the simple example and then for the main corridor problem. The constraints are examined, and conditions are identified in which the optimisation is successful.

6.1 Introduction

The mathematical model was developed and motivated in Chapter 4, with the use of a manual example. In Chapter 5, the description of a South African container corridor was presented. A Genetic Algorithm (GA) was also developed to perform an optimisation on this model. First, the implementation of the GA on a simple example was tested, the results are presented here.

Then the results of the model on the container corridor are presented. This is the model developed to implement on a large-scale application. The development and debugging were performed in the simple model and once it was working properly, the problem could be extended to a more complex variant to solve corridor level problems. The focus of the study is on the development and verification of the model itself, therefore other examples are not necessary.

Optimisation runs were performed to optimise costs through changing train lengths and decisions on running a train, for the corridor. This resulted in finding the minimum cost and train length. Volume throughput levels were tested in comparison to the production rates. This scheme enabled the finding of a feasible solution. These tests were conducted on an Apple MacBook Pro with an M1 pro chip and 32 GB of RAM.

6.2 Implementation of GA in Simple Problem

Here a stochastic solution is developed using MATLAB®. The simple problem is developed based on the hub and spoke layout in Figure 4.1. The problem description and input values are given in Section 4.2, while the optimisation code is listed in Appendix A. Values are found for all the variables using the Genetic Algorithm, (the development of the full solution GA is set out in Section 5.3). In this example 10 periods were considered for ease of demonstration. The simple problem has six parameters, $x_{11}(t_k)$, $y_{11}(t_k)$, $x_{21}(t_k)$, $y_{21}(t_k)$, $x_{12}(t_k)$ and $y_{12}(t_k)$, with 11 variables each, for a total of 66 variables. The fitness function values are then calculated and evaluated.

Consider that the railway runs trains as follows: Trains are run between two origins (O_1 and O_2), a hub (H_1) and a destination (H_2). Each period, $t_{k+1} – t_k$, is considered a slot in which one train may run on each section. The containers are produced at each origin at a fixed production rate.

The testing scheme involved running the optimisation with all R_i penalty values set to 10 initially and performing 100 generations, however the demand constraints were not met. The number of generations were then increased to 150 whilst monitoring the constraints. All penalty parameter values were increased to 100 after analysis of the penalised Objective Function. The values for R_1, R_2, R_5 and R_6 may remain at 100. Difficult to satisfy constraints were identified and the corresponding penalty values, R_3 and R_4, were increased to 200 and tested at 100 and 150 generations again. The demand constraints proved difficult to satisfy, while noting that the other penalty values could not be reduced. The demand constraint penalty values needed to be increased such that the result which satisfied the constraint would be the lowest cost solution. After analysis and experimentation, R_3 and R_4 were increased to 1 000 with successful outcomes. The values (R_1, R_2, R_3, R_4, R_5, R_6) = (100, 100, 1000, 1000, 100, 100) are then used for the final test runs. Since stochasticity is involved due to random number generation, three runs were conducted, and average results are also presented.

In Table 4.1, D_{Total1} and D_{Total2} are the demand for transport of goods by these transport modes and to satisfy business requirements, they are specified. The purpose is to test the optimisation, note that mathematically this is the same problem that was addressed manually in Section 4.2.

In the Slot priority scenario, shorter train lengths are chosen. Some slots may be missed due to other circumstances, this may be attributable to other factors such as the containers not being available.

The simple problem was coded into MATLAB®, this represents the first trial of the train length optimisation problem with the genetic algorithm. The pertinent results are summarised in Table 6.1, with more detailed results available in Appendix B and C.

Table 6.1 Pertinent results for Simple Problem

	Volume [container teu's]	Cost [Rand]	Cost/wagon [Rand]
Run 1	398	2 794 000	7 020
Run 2	405	2 799 500	6 912
Run 3	451	2 897 500	6 425
Average	418	2 830 333	6 786

Table 6.1 presents the results for the simple problem. In Table 6.1, the first column contains labels for three runs and the average, the train lengths are manipulated by the algorithm to find the lowest cost while satisfying (in some cases almost satisfying) the constraints. The data under the third column is due to Eqn (3). The data presented includes the volume of containers which were delivered, the cost of the operation and the resultant cost per wagon, and the average value over the

three runs. Suitable penalty values, as discussed above, have been determined for this problem above and are applied across all of these runs. Results presented in Table 6.1 are obtained using 150 generations of the GA. The volume transported does not satisfy the demand constraint in Run 1, however the result is close to the demand constraint. Note that the cost demonstrated does not include penalty costs, i.e. Eqn (17).

Comparing the average value to the best value from Run 3 depicts the progression in the quality of the solutions.

The problem has been solved here using the GA developed for the purpose. Previously, this problem was solved manually to demonstrate the mathematical model in Table 4.3. Those results were significantly different, the volume transported was 450 teu's at a cost of R 13 033 500. In Table 4.3 the author's intuition of minimising the train lengths was followed, while attempting to use every slot, to avoid wasteful overhead costs. However, it is also possible to run some longer trains and avoid certain slots, and this could result in a lower cost, this is not a simple task to accomplish manually as it is counterintuitive. However, it is clear that in this case the optimisation routine achieves a lower cost while delivering the required wagons (with their containers). During the optimisation, this was achieved by limiting the train lengths from the origins, y_{11} and y_{21}, thus saving costs. The train lengths between the hubs, y_{12}, however, was not as restricted to achieve the required volume.

Table 6.2 Optimised train lengths and decisions for simple problem

Time interval (k)	0	1	2	3	4	5	6	7	8	9	10	Total [teu]
$x_{11}(k)$	1	0	0	0	1	1	1	1	1	0	0	
y_{11} (k)	38	25	38	24	40	36	23	36	40	20	28	
Containers (O_1 to H_1)	38	0	0	0	40	36	23	36	40	0	0	213[1]
$x_{21}(k)$	0	1	1	1	1	0	0	1	1	0	0	
$y_{21}(k)$	35	20	34	23	36	34	27	37	38	40	28	
Containers (O_2 to H_1)	0	20	34	23	36	0	0	37	38	0	0	188[2]
$x_{12}(k)$	1	1	1	1	1	1	0	1	1	1	0	
$y_{12}(k)$	40	49	48	46	46	43	41	47	41	45	43	
Containers (H_1 to H_2)	40	49	48	46	46	43	0	47	41	45	0	405[3]

where:

1 indicates the total containers moved from O_1 to H_1

2 indicates the total containers moved from O_2 to H_1

3 indicates the total containers moved from H_1 to H_2

The GA was run three times with the same R_i values. Run 2 is selected from Table 6.1 for demonstration as the volume constraint, D_{Total2}, was satisfied at the lowest cost and Run 1 slightly violates the demand constraint. These optimised train lengths and decisions are presented in Table 6.2. A pattern is emerging where the first mile train lengths may be significantly shorter than the maximum while achieving the target (where it is not a stretch target). Note that the chromosome contains train lengths for all time points as it is defined in this manner, however the train length is only applicable if there is a corresponding positive train departure decision for that time point and this is enforced by the constraints of Eqns (5) and (7). For O_1, most slots are used with an average train length of 25.4 wagons. For O_2, fewer slots are used with an average train length of 24.8 wagons. For H_1, most slots are used with an average train length of 44.7 wagons. These averages are calculated only where there is a positive train departure decision.

At hub H_1, 213 containers arrive from origin O_1 and 188 containers arrive from origin O_2. The total number of containers arriving at the destination hub H_2 is 405. It should be noted that these quantities may not sum precisely due to the presence of initial containers already in the system. In this solution, all constraints—namely, demand, train capacity, and container availability—are satisfied.

The total required demand at the destination, D_{Total2}, was 400. While the chromosome representation may include wagons assigned to time slots where no train is scheduled, such wagons remain stationary. Containers must be available before a train can be assembled and dispatched.

6.3 Corridor problem application

The results for the application on the South African container corridor, described in Chapter 5, are presented in the following sections. The full description of the setup parameters for the South African corridor problem is presented in Table 5.1.

The problem consists of six optimisation variables: y_{11}, x_{11}, y_{21}, x_{21}, y_{12}, x_{12}, where the x variables are binary, and the y variables are integer, based on the nomenclature developed in Chapter 4. Hence, in the description of the problem, there are six variables, and the total dimension is determined by the number of decision points. The demand requirement, $D_{Total2} > D_{Total1}$, as the accumulation of containers at the final destination represents the total volume throughput of the system.

In this case, consider that three trains run in every 24-hour period, three slots per day. Thus, 20 periods are equivalent to a week. As before, in Chapter 4, decisions are taken at time t_k, k = 0,1,2,…20, resulting in the total number of decision variables being 126: $y_{11}(t_k)$, $x_{11}(t_k)$, $y_{21}(t_k)$, $x_{21}(t_k)$, $y_{12}(t_k)$, $x_{12}(t_k)$.

6.3.1 Sensitivity analysis of penalty parameters

A sensitivity analysis was performed by studying the effect on each term of the penalised objective function when the penalty parameter arises. The penalty parameters relate to wagon, demand and container penalty conditions. The constraints, Eqns (8) and (9), have been handled during the crossover process via clipping as explained in Section 5.3.6. In addition there are also overhead costs which are incurred when no trains run.

To begin, the GA has been run 5 times for 150 generations using the parameter values obtained for the simple problem. These values are (R_1, R_2, R_3, R_4, R_5, R_6) = (100, 100, 1000, 1000, 100, 100). These are suitable values which produced optimal value for the simple problem with satisfaction of all constraints. However, when applied to the main (larger) problem, these parameter values proved unsatisfactory with respect to constraint violations. The most troublesome parameters are R_3 and R_4. As before, these parameter values were increased by 100%, while the remaining parameters were increased by 20%, while the population size was increased to 200 and 300 generations were performed. Hence, with (R_1, R_2, R_3, R_4, R_5, R_6) = (120, 120, 2000, 2000, 120, 120), the GA was run 5 times for 300 generations. This was to determine if there are constraints that are persistently violated. Again, there were 2 sensitive values, R_3 and R_4, corresponding to constraints (10) and (11). The values of R_3 and R_4 were increased by 50% and the GA was run 5 times to observe the effects. These values along with the other parameter values provided stability in finding optimal objective values and with acceptable constraint violation. Therefore, these values, (R_1, R_2, R_3, R_4, R_5, R_6) = (120, 120, 3000, 3000, 120, 120), were used for the final numerical testing. These values provided acceptable levels of violation at each run.

Figure 6.1 presents the performance when using the three penalty regimes described above, with the y-axis being the unit cost per teu transported and the x-axis representing each run. The first penalty regime was used for runs 1 to 5 where (R_1, R_2, R_3, R_4, R_5, R_6) = (100, 100, 1000, 1000, 100, 100), the second for runs 6 to 10 where (R_1, R_2, R_3, R_4, R_5, R_6) = (120, 120, 2000, 2000, 120, 120) and the third for runs 11 to 15 where (R_1, R_2, R_3, R_4, R_5, R_6) = (120, 120, 3000, 3000, 120, 120). For runs 11 to 15, the volume target was achieved in all of the 5 runs and the average unit cost achieved was R 5 749, which was acceptable. The unit cost is defined as the total cost for that run divided by the number of teu's which were transported.

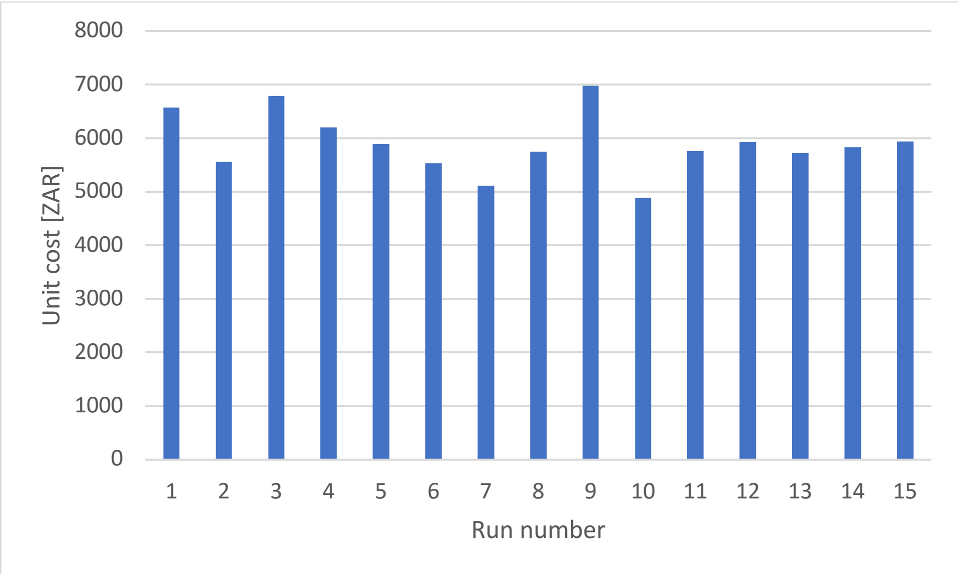


Figure 6.1 Performance testing of penalty values

Therefore, this penalty regime is chosen, where (R_1, R_2, R_3, R_4, R_5, R_6) = (120, 120, 3000, 3000, 120, 120). Runs 1 to 5 are categorised as poor, where the average unit cost is R 6 203, Runs 6 to 10 are categorised as average, where the average unit cost is R 5 652, Runs 11 to 15 are categorised as good, where the average unit cost is R 5 838, but the volume target was achieved for all the runs.

There are three groups of constraints: wagon (Eqns (5), (7)), demand (Eqns (10), (11)) and container (Eqns (4), (6)). Each presented different challenges.

The demand constraint requires that a defined demand requirement is met. The container constraint requires that there are sufficient containers available in the stack (at a particular location and time) to load the required wagons for the train.

For the wagon constraint, the initial chromosome generation results in wagon numbers being populated, however the calculation of costs only includes wagons which ran in a train and therefore the constraint is satisfied. This constraint is not difficult to satisfy given the design of the GA and the execution of the solution here. It is included for completeness in the mathematical formulation as it could be required in other solution methods.

The demand constraint was satisfied when demand was within reasonable levels compared to the production rate. One cannot move more of the goods than the quantum that is produced and that are available in the stack.

The container constraint referred to the container stacks, i.e., the accumulation of containers at the origins and at the hub. This was a significant factor and had an impact on the operation of the system. Mathematically, it was possible to satisfy this constraint. However, it had the single largest effect on volume throughput, simply put, if there are no containers available, the trains cannot run. This became more an issue of undersupply than about optimising the logistic system.

The container constraint results in a penalty value when there are insufficient containers available for the wagons that are waiting to depart. The chromosome has wagon quantities available for all time and locations, with a corresponding decision on whether the train runs. If a train is required to run, the cost will be calculated and a penalty incurred if there are insufficient containers. Therefore, this constraint is critical in ensuring that there are no erroneous cost allocations and required effective penalty parameters, R_1 and R_2. Note that the container constraint is calculated for each time period and each container stack, as there are

21 time points in this study, some of them will have positive penalty values, affecting how chromosomes are selected in the genetic algorithm.

Using the set of R_i values, (R_1, R_2, R_3, R_4, R_5, R_6) = (120, 120, 3000, 3000, 120, 120), the GA was run for 300 generations, and these results are summarized in Table 6.3. The best results are presented in the graphs on pages 113 to 115.

6.3.2 Optimisation results

Optimisation results are presented in Table 6.3. These results present the volume which was moved and the cost, as well as identifying the best and worst runs with the average over all of these runs.

Table 6.3 Optimisation results

Run	Cost	Volume	Unit cost	Comment
11	2 603 432	452	5 760	
12	3 213 972	542	5 930	Best
13	2 577 140	450	5 727	
14	3 102 680	532	5 832	
15	2 709 216	456	5 941	Worst
Average	2 841 288	486	5 837	

Run 12 is identified as the best result, as the highest volume was achieved, while satisfying the constraints. The average Cost, Volume and Unit cost are also presented in the last row, which indicates the improvement to the best result.

The best solution is now studied using the solution for Run 12. The binary variables on train decisions are first observed, this is summarised in Figure 6.2, 6.3 and 6.4. Note that these train departure decisions will apply to Figure 6.5, 6.6 and 6.7, respectively. Run 12 is considered as the best run overall because although there are lower overall costs in some other runs, they also moved less volumes.

Therefore, the lowest unit cost gives a consistent measure of which run was actually better.

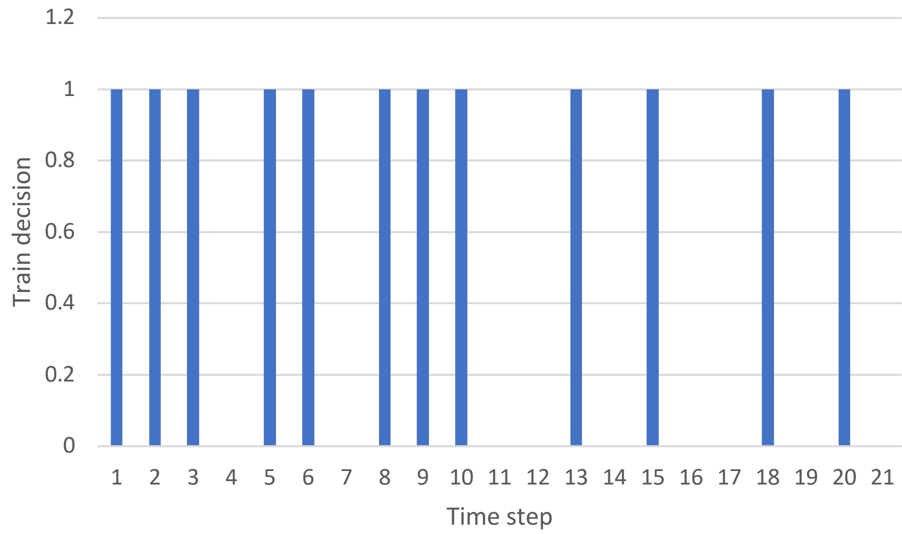


Figure 6.2 Train decisions at Origin 1

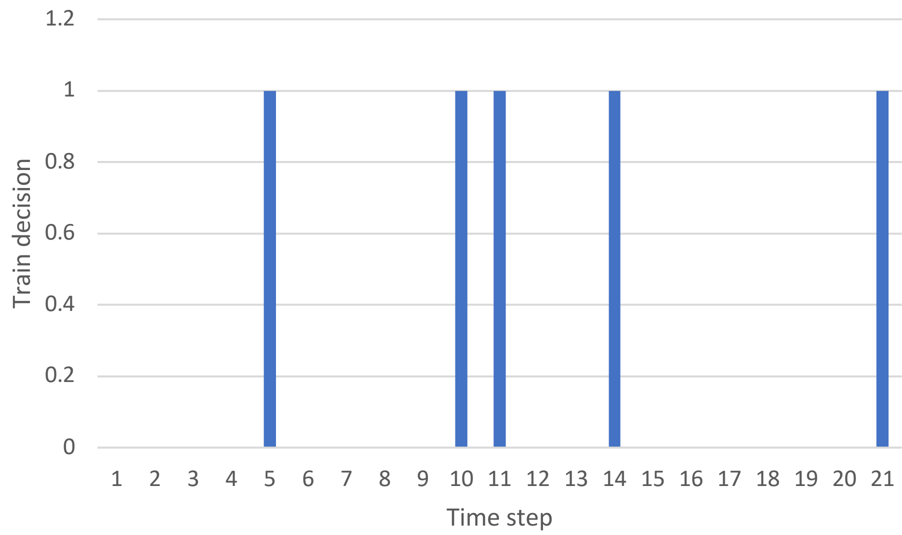


Figure 6.3 Train decisions at Origin 2

When comparing Origin 1 and Origin 2 train decisions in Figures 6.2 and 6.3 respectively, note that there are significantly more departures in Origin 2. The constraint for D_{Total1} is dependent on the total deliveries from both origins and in this case the combination of departures between the two origins allowed for the lowest cost.

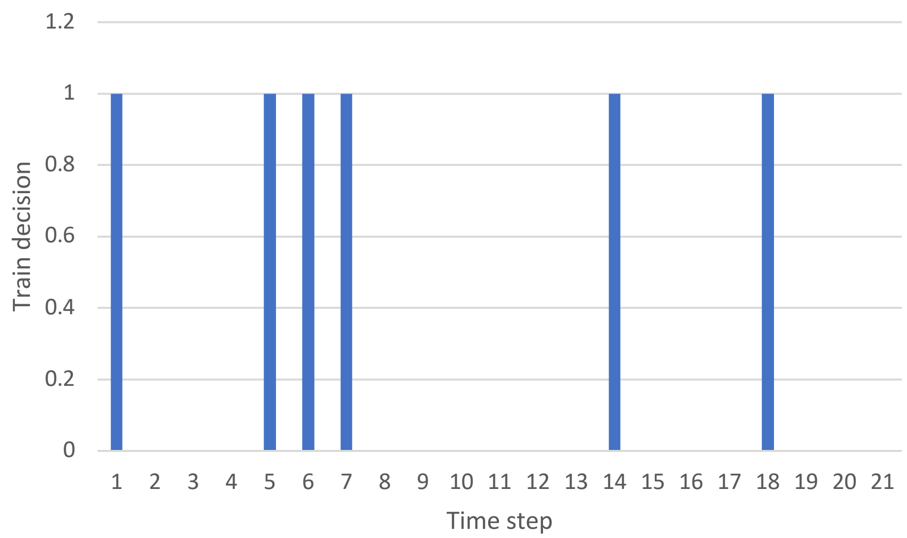


Figure 6.4 Train decisions at Hub 1

Figure 6.4 presents the train decisions taken during Run 12 at Hub 1, where only six slots were used for the period under consideration.

Next, the integer variables of the Run 12 are studied, this is summarised in Figure 6.5, 6.6 and 6.7 for train lengths departing from Origin 1, Origin 2 and Hub 1 respectively. Figure 6.5 presents y_{11}, Figure 6.6 presents y_{21} and Figure 6.7 presents y_{12}. While the chromosome contains an integer value for each run, by its definition, these values will only apply if there is a corresponding positive train departure decision in Figure 6.2, 6.3 and 6.4, as per the constraints of Eqn (5) and (7), avoiding infeasible solutions by the careful selection of penalty values. Note that in all cases the train lengths are below the maximum allowed length.

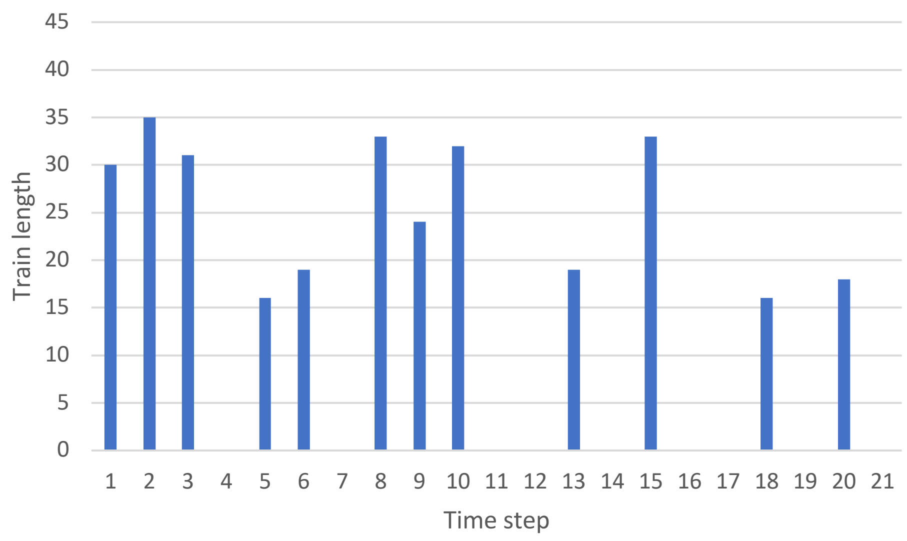


Figure 6.5 Train lengths at Origin 1

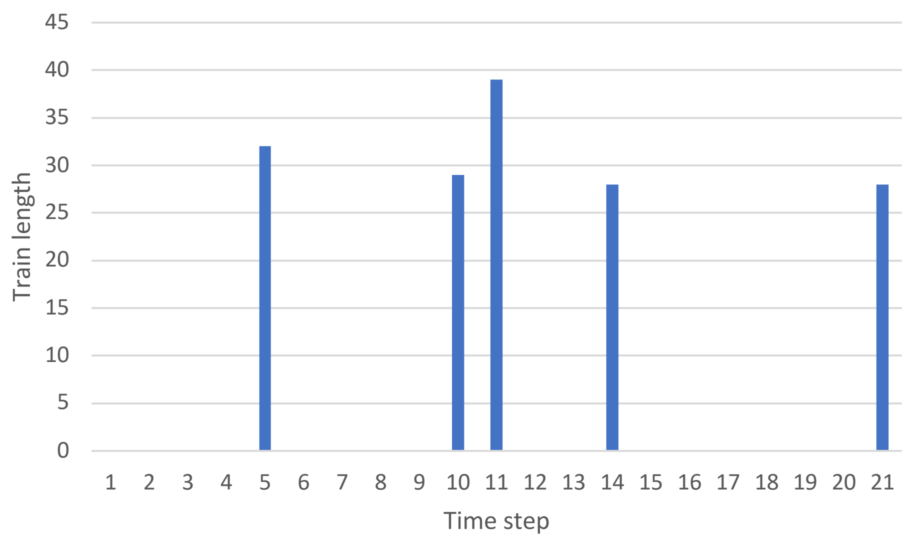


Figure 6.6 Train lengths at Origin 2

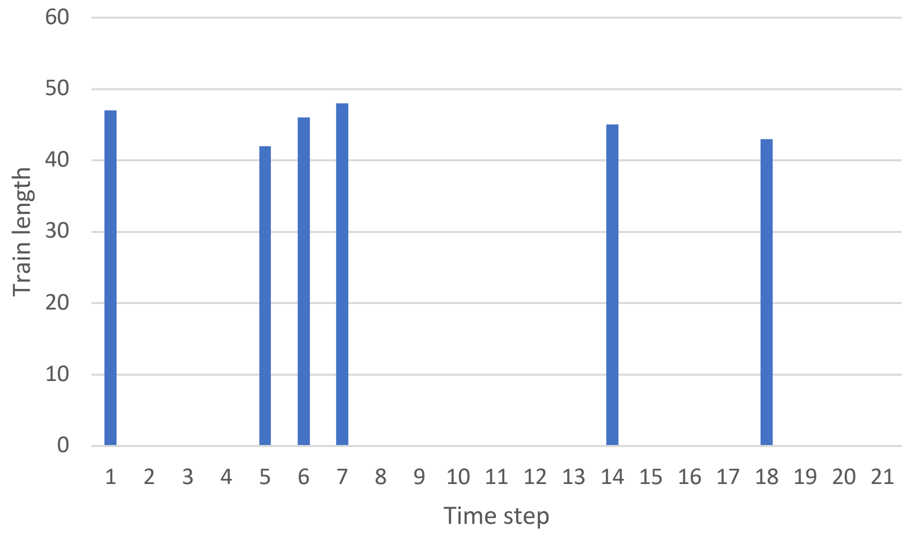


Figure 6.7 Train lengths at Hub 1

A range of train lengths were used in the various time steps spanning a large part of the available range. Note that the chromosome is generated randomly and is chosen based on cost, which is influenced by the penalty factors. Although the chromosome does have train lengths at every time step, they are only used if a train departure decision exists for the corresponding time step, the unused train lengths have been removed in the above figures.

For Run 12, the cost incurred was R 3 213 972 when transporting a volume of 542 teu's, for a unit cost of R 5 930.

6.4 Comments on the optimised results

Relating the results to the research objectives, a novel mathematical formulation was constructed and used as the theoretical base for the modelling of an intermodal hub and spoke system. In the model a cost optimisation was performed. Conducting optimisation on the logistic system led to the conclusion that train lengths could be optimised to deliver volumes at lower costs. While all train lengths were reduced, the unit costs were higher in some cases, due in part to the overhead costs when a slot is not used.

The optimisation was able to find feasible solutions where all constraints were satisfied, and the required volumes were moved. In the optimisation process train lengths were chosen which were lower than the upper limits. This is considered a successful application of the optimisation.

Further, the container stack size is a critical factor contributing to the ability to move volumes in general, and to the ability to support optimised train lengths. It was found that optimising the cost by reducing the train lengths did lead to higher volumes being delivered at a lower unit cost, if the container stack size is not a limiting factor.

Chapter Seven
Simulation Design

7.1 Simulation model overview

Following the development of the mathematical model and its solution, an alternative technique was tested using a simulation, which allows for increased capability, described here.

The simulation model was also developed to model the transport of containers in a hub and spoke logistic environment. The approach is to construct a transport environment, recreating the hub and spoke system of Figure 3.2. Volumes are transported through this system using locomotives, wagons and crew, which must then be managed. It is also possible to use road vehicles if required. The simulation implements the constraints of Section 4.1. Similarly, the fitness of the results is evaluated using the objective function of Section 4.1. The simulation allows for the introduction of real-world aspects to the considerations. The effect of these aspects can then be used to observe the effects on the operation, leading to further conclusions on what conditions are suited or not suited to optimising train lengths. The simulation allows for the consideration of the following additional aspects for possible further research:

1. Testing the effect of capacities (rail line, road, handling facilities, maintenance, effect of outside disturbances);
2. Include different capacity vehicles;
3. Include different transport speeds;
4. Testing of effect of balancing of locomotives, wagons and crew (distribution in the environment);

5. Input outside disturbance and delays and test resilience (ability to recover) of a logistic system, and
6. Simulation allows for expansion of research, other items to investigate may be included such as different handling practices, container business models etc.

These elements and others can be included in the simulation without a need for updating of the mathematical formulation, which can still be used as the cost objective function, subject to evaluation. The simulation development is presented in the following chapter, and results are presented in Chapter 9. The simulation model is a development of the work on the hub and spoke logistic system with a focus on the operational complexities. The simulation model provides a more intuitive feel of the environment and allows for manipulation of the operations to test various scenarios, allowing for determining whether an optimal solution can be practically implemented.

Here, the simulation introduces new elements in the following areas:

a) Checks for resources (crew, wagons, locomotives) before departure
b) Allows for delays both for routine events such as maintenance failures and non-routine events such as major weather disruptions (e.g. floods)

The simulation has been developed to explore the operation with varying parameters. Scenarios have been developed to model the effect of a range of factors when running the simulation, viz. distance between nodes, speed of transport, production rate at factories and density of container stacking. A run plan has been developed to incorporate all high and low combinations of these factors, presented in Table 7.1.

Various complications can be included in the model, including linear and non-linear delays. These can be used to incorporate operational delays such as maintenance delays, insufficient crew and locomotives not being in the right place. More serious non-linear delays can also be catered for such as derailments or natural disasters such as floods impacting the operation.

The simulation model focusses on the operational aspects, rather than the optimisation. Elements such as the balancing of resources, rolling stock fleet sizes tailored to the operation (production rates, travel distances, speeds and maintenance requirements) may be tested, which are required for a stable system.

The simulation was built and runs in Simulink®. Essentially, building blocks of the hub and spoke logistic operation are constructed and the objective function of the previously developed mathematical model in Section 4.1 is used to calculate the cost, which must be minimised. This is performed by a genetic algorithm from a MATLAB® toolbox which is run in a script, refer to Figure 7.1. The model is run from a MATLAB® script which sets up the parameters and calls the optimisation routine (Genetic Algorithm). This is followed by a simulation of the freight movement using the optimised train lengths which are returned to the genetic algorithm for evaluation. Further details of the simulation development are presented in Chapter 8. For each run the process is as follows:

a) Set parameters for the run (O1HDist, O2HDistl, O1Rate, O2Rate, MainlineD, described in Section 7.2)
b) Start the script in MATLAB® (see Figure 7.1)
c) Call genetic algorithm (see Figure 7.1)
d) Find new train length (using genetic algorithm process)
e) Run simulation with new train length
f) Repeat until genetic algorithm terminates

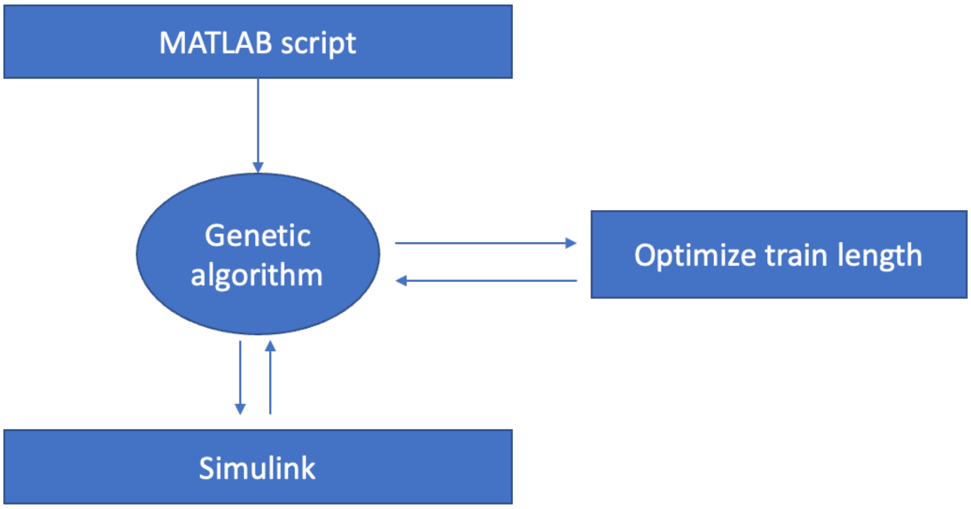


Figure 7.1 Train length optimisation process

A limitation of this process is that train lengths are determined and used for all trains which run in the period of the simulation, instead of choosing a new train length for each slot. This may also be considered a benefit from a rail company perspective as constant changes in train length are not required (which may be difficult to achieve).

7.2 Simulation model parameters

To model the movement of freight through the system, one must consider the production rates, distance of travel, train lengths and train departures. These parameters have been defined in Section 4.1, the nomenclature is applied to the code as follows:

a) TrainLength(1): Train length departing O_2 – y_{21}
b) TrainLength(2): Train length departing O_1 – y_{11}
c) TrainLength(3): Train length departing Hub – y_{12}
d) TrainLength(4): Train length departing Port, required to balance resources
e) Train departure decisions (from Origins) – x_{11} and x_{21}
f) Train departure decision from Hub – x_{12}

g) Production volume of containers (multiple Origins)

h) Freight arriving (Destination)

i) O1Rate: Production rate at O_1

j) O2Rate: Production rate at O_2

k) O1HDist: Distance from O_1 to hub – d_{11}

l) O2HDist: Distance from O_2 to hub – d_{21}

m) MainlineD: Distance from Hub to Port – d_{12}

The following run plan, listed in Table 7.1, has been developed for the simulation. Each of the parameters, distance, speed, production and density (single or double stacked) will be varied, generating data where optimum train lengths are found for each situation, allowing a theoretical framework to be developed.

Table 7.1 Run plan

Conditions	Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Distance	short	x	x			x	x			x	x			x	x		
Distance	long			x	x			x	x			x	x			x	x
Speed	slow	x		x		x		x		x		x		x		x	
Speed	fast		x		x		x		x		x		x		x		x
Production	Low	x	x	x	x					x	x	x	x				
Production	High					x	x	x	x					x	x	x	x
Density	Low	x	x	x	x	x	x	x	x								
Density	High									x	x	x	x	x	x	x	x

In Table 7.1, the parameters are listed in the first column and the conditions are defined in the second column. All combinations of these conditions are tested over sixteen runs, listed in the first row and marked with a corresponding “x”. The run plan is designed to produce output for all combinations of the various environmental parameters. In each run, the output will be at the end of an optimisation process using a genetic algorithm attempting to find the optimal train length. MATLAB® uses the genetic algorithm to vary the train length parameter, TrainLength(1), TrainLength(2), TrainLength(3) and TrainLength(4), running multiple simulations in Simulink® to minimise the objective function. These optimisations are stored in data sets.

Table 7.1 presents the parameters chosen to represent the various scenarios required by the run plan. This table identifies all combinations of the parameters Distance, Speed, Production and Density.

- Distances (short and long) are for the sections from the origins to the hub and the main line section,
- Speed (high and low) are chosen for the vehicle speed on those sections,
- Production rates (high and low) at each of the origins are defined,
- Stack density (whether containers are stacked one or two high) are also defined.

A further set of tests where no optimisation takes place are stored in a second data set. These simulations are conducted for all the Runs defined in Table 7.1. These simulations form a reference against which to evaluate the effect of the optimisations. The trains are run at the predetermined maximum lengths for each section.

Descriptions for the conditions in Table 7.1 are presented in Table 7.2.

Table 7.2 Parameters in the simulation

Condition	Distance	Speed	Production	Density	x
Description	Travel distance between nodes	Transit speed	Production rate of containers at the origin	Containers are single stacked or double stacked	Parameters selected for that run, incorporating all combinations

7.2.1 Assumptions

The demand for goods and hence the containers is supplied by the production of the factories in the origin modules. This demand is considered independent of external factors such as demand for the product, contracts and global economic trends. It is assumed that all goods produced by the factory will be transported in the logistic network.

It is assumed that a minimum demand exists and that the capacity of the wagons and of the train is not wasted.

As the capacity of the logistic chain is under consideration, the traction required is assumed and not explicitly calculated. Furthermore, train length changes below a certain threshold will not trigger a change in the number of locomotives required as this is not a focus of this study. All locomotives will be returned to the origin.

The energy saved by running shorter trains is not under consideration.

7.2.2 Testing regime

A regime for running the simulations, gathering the relevant data, and optimising this data is defined here.

The testing regime is designed to contribute to answering the research question. The simulation has evolved into a process to provide insight into how the intermodal system is operating and what the optimal way to operate it will be. To this end, the following objectives must be achieved:

1. Define various operating environments.
2. Vary train lengths to values less than the maximum.

3. Observe goods arriving at the destination.
4. Can the train lengths be optimised?
5. Under what conditions are freight arrivals maximised?

7.2.3 Data source

Train operations data has been obtained from Transnet, a South African logistic company, to calibrate the model. This data includes train transit times, distances, demand for the goods and costs relevant to railway operations.

7.2.4 Test focus

Intermodal transport refers to the movement of goods by several modes, where the goods themselves are not handled as they are packaged into containers at the origin. In this study, variable length trains both before and after the hub are studied. The containers are moved from one train to another at the hub without handling of the goods within the container. This is regarded as a mode change, (Meng & Wang, 2011).

The simulation model has, however, been built so that rail or road journeys may occur in any segment before or after the hub. This may be the subject of further study as the existing data set does not have any road data.

Chapter Eight

Development of Simulation Model

8.1 Introduction

A mathematical formulation for a hub and spoke network was developed in Section 4.1. An MS Excel spreadsheet was used to execute this model in Section 4.2. Furthermore, the capability to solve the mathematical model was developed in MATLAB® using a genetic algorithm. To allow increased capability such as the introduction of delays, a simulation model was also developed to model the transport of containers in a hub and spoke logistic environment[2]. The simulation implements the constraints described in Section 4.1 for such an environment. Similarly, the fitness of the results is evaluated using the objective function of Section 4.1. The simulation allows us to introduce real world aspects to the model (such as maintenance delays, breakdowns or floods). The effect on the operation can then be observed, leading to further conclusions on what conditions are suited or not suited to optimising train lengths.

8.2 Working of simulation

The simulation model allows for the flow of entities, defined for this purpose as locomotives, wagons, crew and freight. By modelling the movement of these entities through a hub and spoke environment, a simulation model is created which can be used to find optimal train lengths. These entities can be combined in a predefined manner, meaning that all elements of a train in the required quantities would be required before a train can depart and travel in the simulation.

[2] This environment was described in Section 3.1

The model shall investigate the capacity of the system to handle freight in the form of containers, by assessing the following criteria:

i. Capacity to handle containers
ii. Capacity to handle trains
iii. Capacity to handle wagons
iv. Time required
v. Capacity to dispatch trains
vi. Capacity of main line to accept trains

The model is complex and has multiple parameters, interacting with the environment and each other, making decisions and satisfying contradicting goals. The critical parameters (defined in Section 4.1 and adapted to the simulation in Section 7.2) which are required for further analysis and to provide input to the research objectives are:

Inputs

- Production rate (multiple origins, O1Rate, O2Rate)
- Distances (O1HDist (Origin1 to hub), O2HDistl (Origin2 to hub), MainlineD (hub to port))

Outputs

- Arrival train length (at the port)
- Wagons arriving (at the port)
- Container freight arrivals (at the port)

Variables

- Wagons
- Trains
- Containers
- Crew
- Distance
- Time

Decision variables

- Train length departing from nodes
- Number of trains departing from nodes
- Train departure (Should this train run or not)

Each component of a multimodal system has a capacity associated with it. The basic components considered here are the rail and road networks from the customer site (origin) to the hub, the hub itself and the rail network serving the hub. These are shown in Figure 4.1. The associated capacities include road capacity, vehicle sizes, operating parameters and a similar set of associated capacities for the rail as well as handling and storage capacities for the hub. As utilisation of each component is varied, the capacity of the whole network is affected. The optimum capacity of the system was studied.

To develop a model that considers the environment in which it operates, the thinking must include the type of loading and the effect of variable length trains on the operation and capacity of the system.

8.3 Logistic components

The elements (all infrastructure) are built in the model, entities (all moving components in the model, viz. wagons, containers, crew, locomotives and trains) are processed by these elements, thus simulating the logistic flow through the system. The system consists of blocks, which generate and process entities. Each block has an input and output which can be configured as required, allowing the flow of processed entities between the blocks to achieve some objective. The blocks themselves process the entities. Combinations of blocks are set up and designed to generate modules having the following characteristics (these modules represent the logistic system elements described in Section 1.4.1):

i. Origins

The origin module presents a conceptual view of a factory and is designed to model the logistic process, not any manufacturing process or product or similar intricacies that may be peculiar to any specific factory. Instead, the factory or origin is viewed from the perspective of a logistic service provider, such as a railway.

Each origin can produce goods that are output as containers. The origins have the capability to order trains when required. The components of a train are sourced and arrive at the origin if they are available. However, before departure from this origin the complete train made up of wagons, locomotives and crew must be present. Trains and containers may be stored in limited quantities (in the stockpile and the yard at the origin). The circulation and balance of trains in the system determine whether components are available to make up a train when required. The departure of trains depends on these conditions being satisfied. Loading and train compilation take place while a train waits for an available slot to depart.

The origin module operates as follows: Goods are produced, packaged and loaded into containers. When sufficient containers have been produced a train is called. The train consists of locomotives, wagons and crew that are drawn from a finite resource pool. The resource pool will be replenished as trains return from the port, thus requiring a balance in the system and sufficient resource levels for the various conditions. Time is allocated for compiling the train and loading the containers before a train departure check. All components of a train must be present to allow a departure.

The concept described above is presented in Figure 8.1, illustrating how trains are compiled and depart from an origin. This concept of the origin module is executed in the Simulink model, depicted in Figure 8.2.

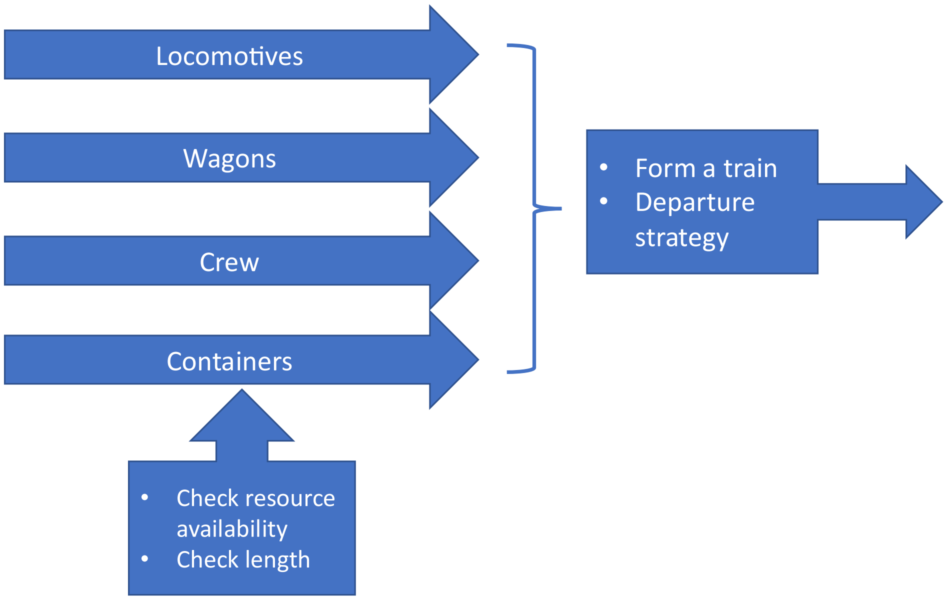


Figure 8.1 Origin module concept

Delays and inefficiencies in the process are accounted for as well as the availability of each component after a preceding journey. These may result in components of a train not being available and no train running in a slot.

Trains are released following one of the train departure strategies (described fully in Section 4.2):

a) The train can wait to make up the pre-determined full length.

b) When the slot is open the train can be released irrespective of whether it is full length or shorter.

To control the resources and balance the train between the origin and the port, a technique was developed to ‘generate’ and ‘destroy’ resources (a parameter in Simulink, separate from an entity like a locomotive), while reserving them for use after they have arrived at the origin.

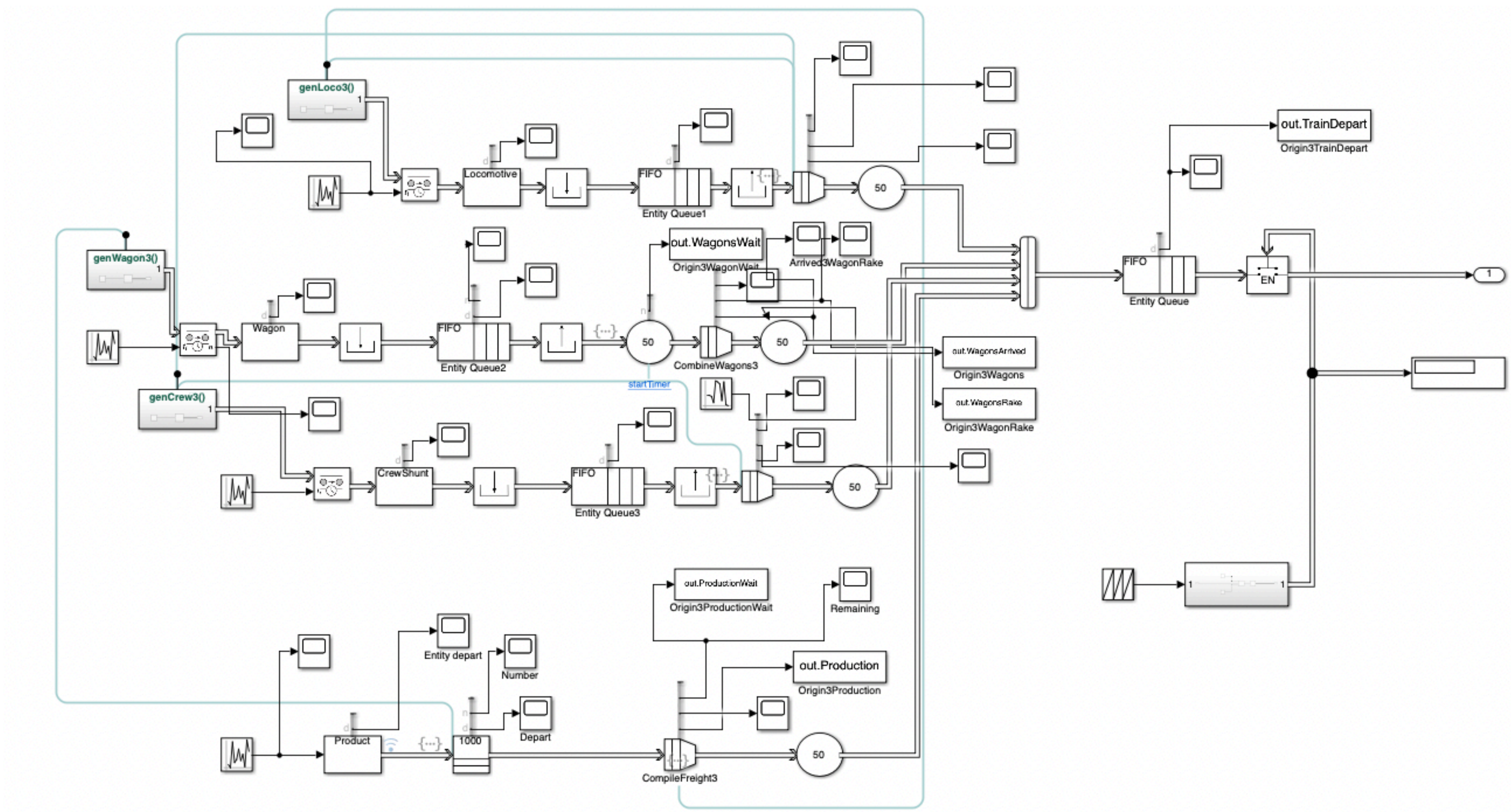


Figure 8.2 Origin module in Simulink

In Figure 8.2 the Origin module is presented. The blocks generate entities that represent the containers. These may be stored in a stockpile at the origin for a short time, the storage capacity can be set for each origin as required. The blocks also generate locomotives, wagons and crew. These are from a finite pool that must be replenished by the resources returning from the port. Finally, if all conditions have been satisfied, the train may depart when a slot becomes available.

ii. Transport modules

Transport modules may represent a rail or a road leg of the transport chain, shown conceptually in Figure 8.3 and as a Simulink model in Figure 8.4. They have a distance and a capacity attached to them. This module is also able to introduce nonlinear interferences in the form of routine delays of random time or more severe blockages of the path as may be the case for a derailment or other such event. Parallel paths are introduced in the model to allow for this functionality.

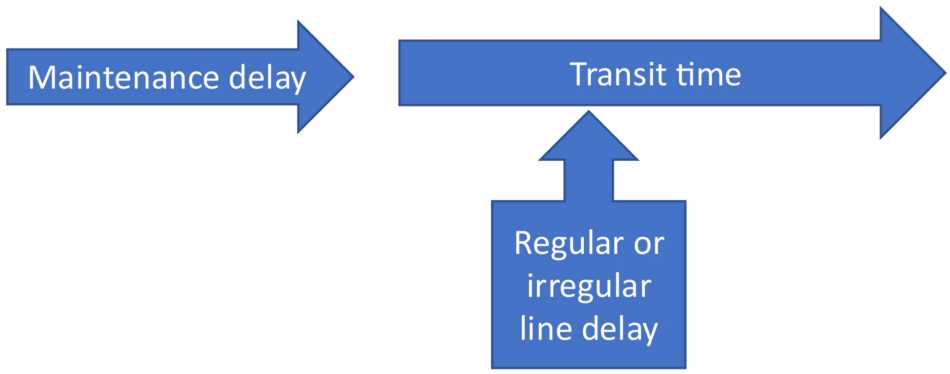


Figure 8.3 Transport module concept

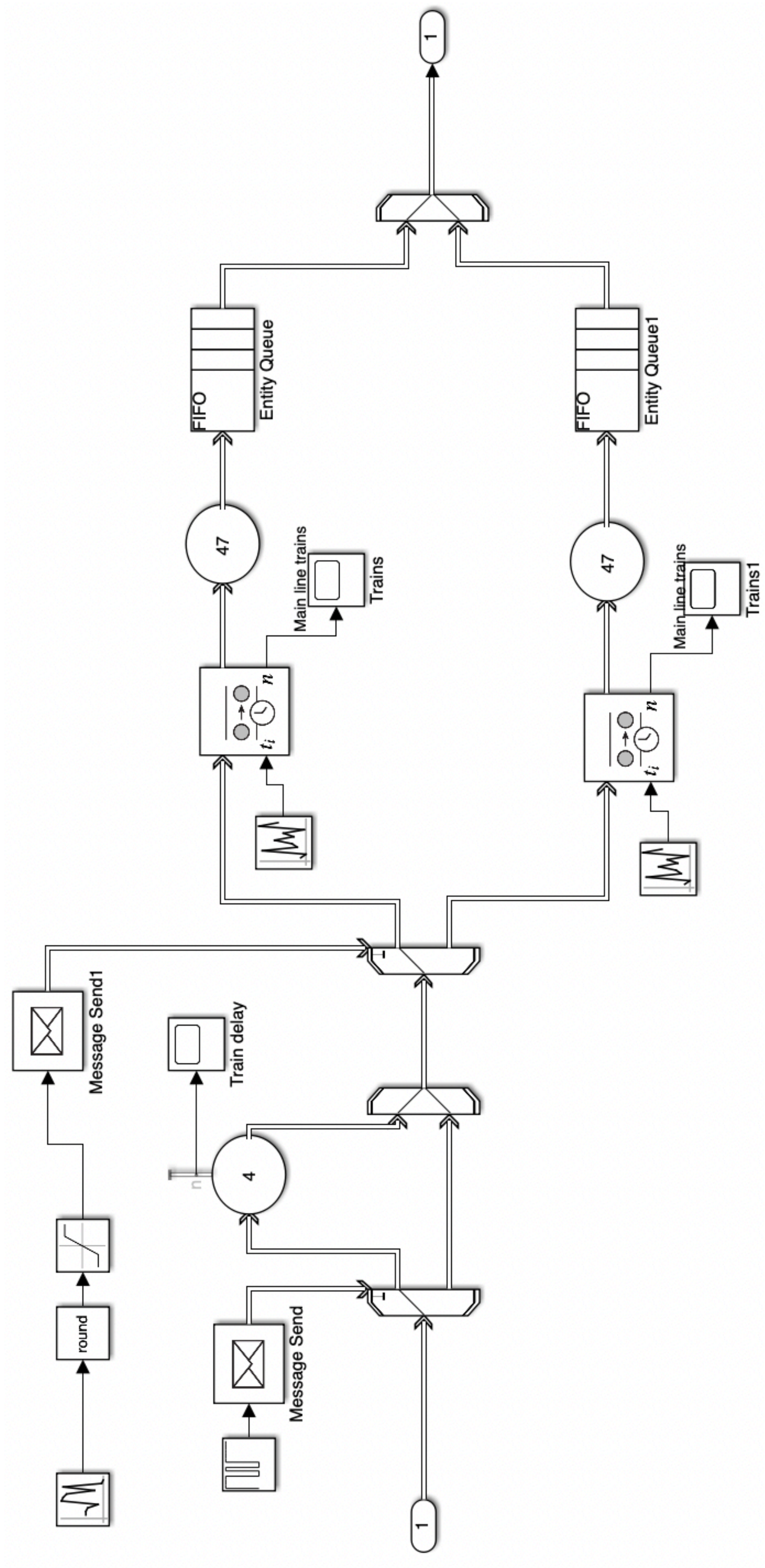


Figure 8.4 Transport module

This module facilitates the movement of trains or trucks along the route. The transit time is based on speed and distance parameters. The delay functionality allows for routine maintenance delays on individual vehicles.

iii. Hub

In the Hub, road and rail paths converge. Trains are compiled and will depart from the Hub on a longer rail journey to a destination, modelled as a Port. The logic depicted in Table 8.1 is applied.

Table 8.1 Hub logic defining intermodal behaviour

Action	**Description**
Combine	Combine various trains into one consolidated longer train
Intermode	A mode transfer occurs, build a train with freight arriving at the node via road trucks
Multiple	Freight from trucks and trains are combined into a longer train
Pass	Trains arrive and depart at the node without handling of the goods, the trains leave in the same form in which they arrived

Hubs therefore also have train compilation capabilities like the Origin elements and are shown conceptually in Figure 8.5 and as a Simulink model in Figure 8.6. Similar to the Origin module, a decision is made on whether a train can depart depending on all the conditions being satisfied.

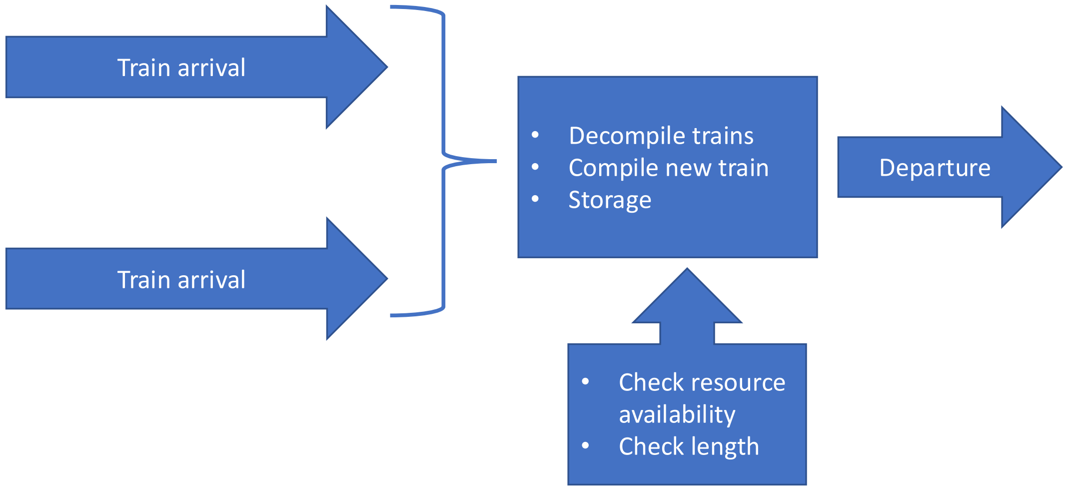


Figure 8.5 Hub module concept

In Figure 8.5, the concept for a hub is depicted. Multiple trains may arrive from various origins. At the hub they may be decompiled and compiled into new trains based on the requirements at the destination and on train length requirements.

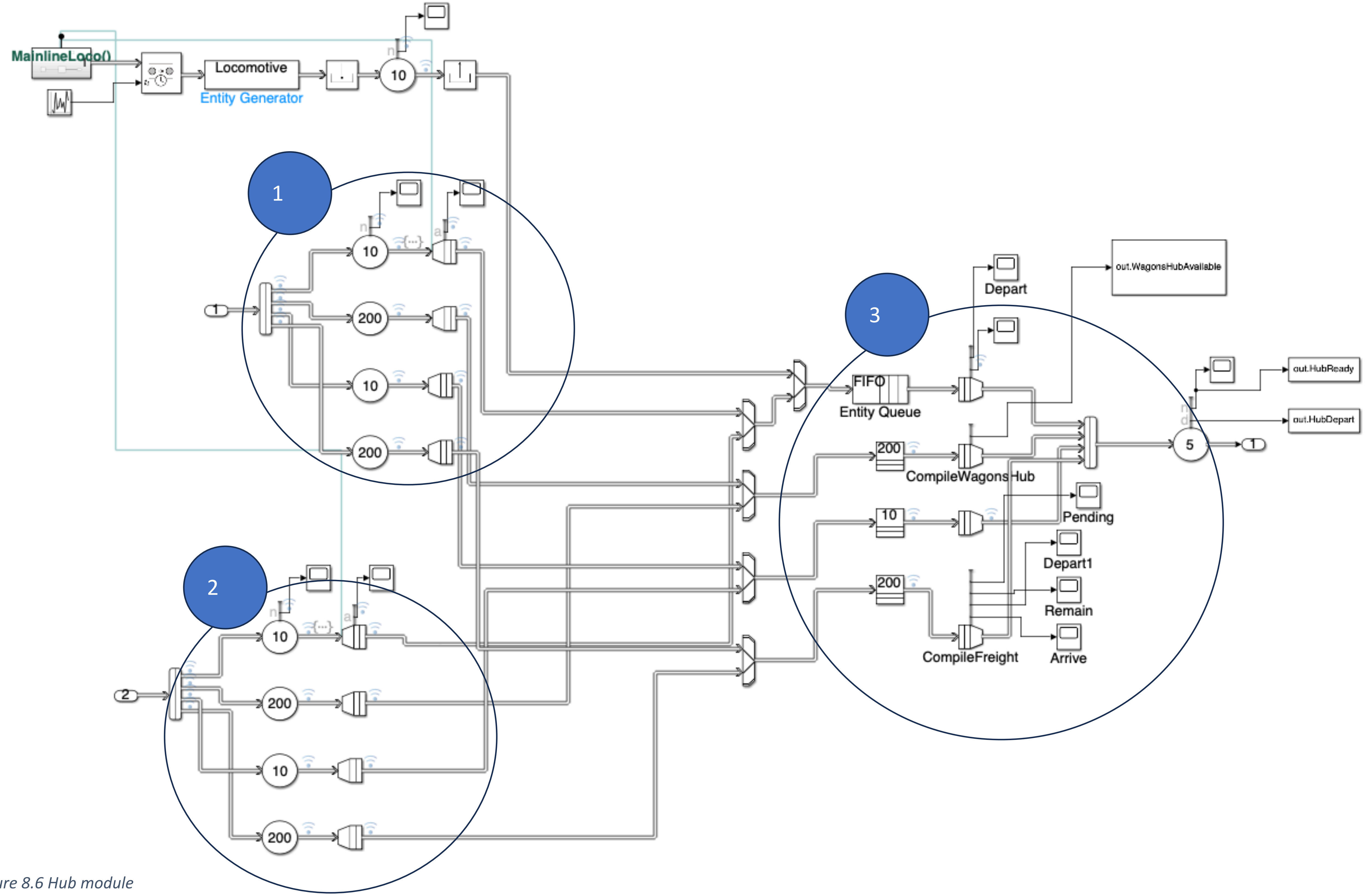


Figure 8.6 Hub module

In Figure 8.6 the Simulink model for the hub is depicted, based on the concept described in Figure 8.5. Region 1 in Figure 8.6 allows for train arrivals from an origin and these trains are decompiled. Region 2 allows for train arrivals from another unrelated origin and these trains are also decompiled. Region 3 allows for the train components to be compiled in a new train combination. Additional locomotives may also be added if required for train handling considerations.

Trains arriving at the Hub are broken up and compiled into a train of a new length determined by a Genetic Algorithm from the MATLAB Global Optimisation Toolbox in its search for the optimal train length. If trucks are used, the model does not consider how they are distributed after they have delivered their cargo as they are not the focus of the study.

iv. Port

Ports represent the destination and are the boundary of the model. At the port data is gathered on trains and the volume of goods arriving. The port also facilitates the handling of the train and crew, starting a return journey and is shown conceptually in Figure 8.7 and as a Simulink model in Figure 8.8.

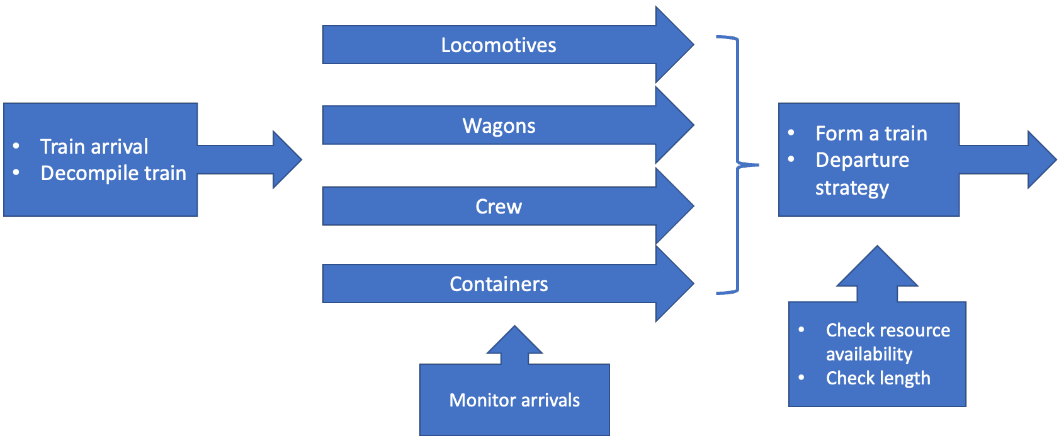


Figure 8.7 Port module concept

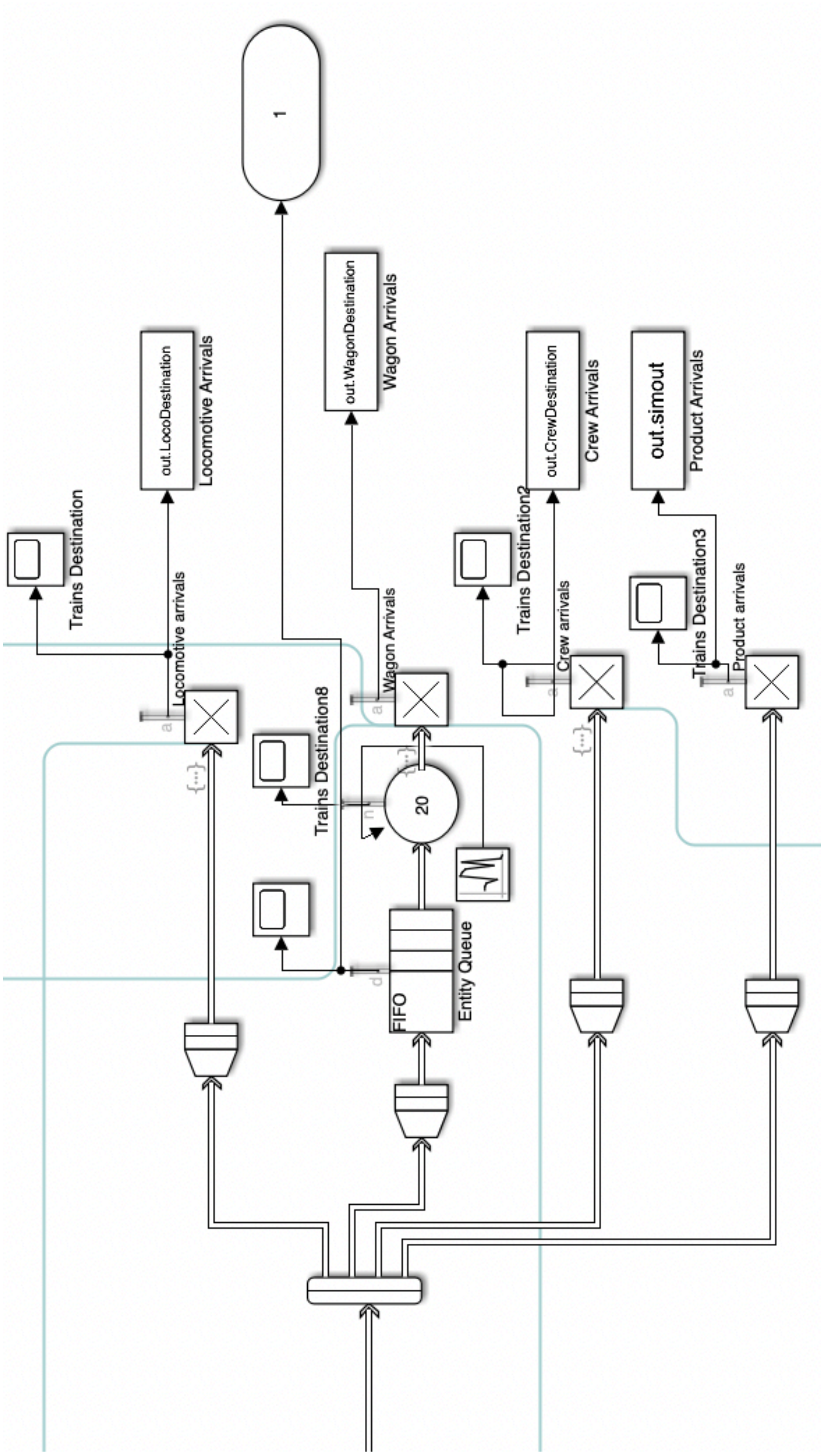


Figure 8.8 Port module

The port starts a return journey for the train, returning all locomotives, wagons and crew to the resource pool, making them available when required at an origin.

The trains are returned carrying containers from the port, they are offloaded and decompiled before being available for use at an origin again.

The Port module allows for the arrival, decompilation and offloading of trains. Arrivals and train lengths are monitored as they are critical for the optimisation. Trains are compiled and depart to the origin side of the model, allowing for the balancing of trains in the system.

8.4 Simulation operation

At each node, there are functions which must be performed:

i. Check the flow requirement
ii. Check the capacity of the node
iii. Decision on excess containers if any
iv. Flow the containers that can be accommodated
v. Build statistics on how these flows occur

Calibration

The Simulink model presented in the preceding sections represents a concept of a complex system, it was designed to investigate the research question. This model was tested with real world data obtained from Transnet in which the performance was tested and aligned with actual data. Permission was obtained to use the data. The model behaves in a similar way to a container logistic system.

The model was tested using data from a South African container corridor, shown in Table 8.2.

To confirm the validity and accuracy of the model, results were compared to an actual train plan (obtained from Transnet with permission) for a container flow.

Table 8.2 Comparison of actual train plan with model

	Plan		**Model**		**Correlation [%]**
	Month	Year	Month	Year	
Trains [no.]	12.04		11.8		98.01
Wagons [no.]	N/A	7456	590	7080	94.95
Containers [teu]	N/A	14913	1200	14400	96.56

In Table 8.2, the comparison of an actual train plan with the model is presented. The trains required per month and the wagons and containers required per year for the actual train plan are presented under **Plan**. These were the requirements for a particular flow of containers over a given route. These conditions were recreated in the model and the results obtained are presented under **Model**. The correlation between the Plan and the Model is high as listed in Table 8.2.

Chapter Nine
Simulation Model Results

The simulation model has been built and run as described in the previous chapters. The pertinent simulation results are presented here. Further simulation results are presented in Appendices E to F for completeness. The model was run on the container corridor, it was also run on a range of parameters designed to test a wide range of applications, demonstrating the scale of application of the simulation.

9.1 Simulation using a South African Container Corridor

The mathematical model developed in Chapter 4 was solved using a genetic algorithm, this was applied to a container corridor, described in Chapter 5. The simulation developed in Chapter 8 was also applied to the same container corridor.

Table 5.1 forms the basic input to the simulation model. The same layout as Figure 5.1 is created in the simulation. Trains are run at a speed of 26 km/h from the origins to the hub, H_1, and at 50 km/h between the hubs, H_1 and H_2, as these are typical main line speeds. In addition, the following delays are applied (these represent the typical delays experienced in practice):

a) Maintenance delay (random time between 0 and 2 hours)
b) Severe event disruption (random time between 2 and 8 hours)

Each train is assigned either the maintenance delay (which could be 0 hours) or the severe event disruption by a normally distributed random signal, and the simulation period was one week.

In the simulation, a volume throughput of 190 teu's was achieved. The train lengths were determined as:

a) Trainlength(1) y_{11}: 36
b) Trainlength(2) y_{21}: 29
c) Trainlength(3) y_{12}: 48

The cost was determined as R 1 820 400.

9.2 Simulation framework testing

The simulation was also tested on a range of conditions, described in Section 7.2 and these results are presented here.

9.2.1 Simulation parameters

The simulation was set up with the parameters for each run, varying the distance, speed, production rate and stack density in turn to generate a set of data for each scenario. The values of each parameter were chosen to represent different environments, traversing a broad range so the results could be applied to various circumstances. This enables an understanding of the impact of the parameter choices in these circumstances. This could be used to develop the theory that could then be applied to any real-world operation. The parameters are listed in Table 9.1.

Table 9.1 Environmental parameters

Run	Dist OH1 [km]	Dist OH2 [km]	DistMain line [km]	Speed OH1 [km/h]	Speed OH2 [km/h]	Speed Mainline [km/h]	O1RateP [gen.time]	O2RateP [gen.time]	Density [stack height]
1	40	30	500	18	18	24	0.8	0.8	1
2	40	30	500	26	26	50	0.8	0.8	1
3	80	60	1200	18	18	24	0.8	0.8	1
4	80	60	1200	26	26	50	0.8	0.8	1
5	40	30	500	18	18	24	0.02	0.02	1
6	40	30	500	26	26	50	0.02	0.02	1
7	80	60	1200	18	18	24	0.02	0.02	1
8	80	60	1200	26	26	50	0.02	0.02	1
9	40	30	500	18	18	24	0.8	0.8	2
10	40	30	500	26	26	50	0.8	0.8	2
11	80	60	1200	18	18	24	0.8	0.8	2
12	80	60	1200	26	26	50	0.8	0.8	2
13	40	30	500	18	18	24	0.02	0.02	2
14	40	30	500	26	26	50	0.02	0.02	2
15	80	60	1200	18	18	24	0.02	0.02	2
16	80	60	1200	26	26	50	0.02	0.02	2

The run plan was given in Table 7.1. Multiple simulations were generated for each run to optimise the train length. A genetic algorithm was used in MATLAB® to find the optimal train length. The final simulation run in each scenario uses the optimal train length and finds the volume throughput for the system in the given period. Table 9.1 presents the parameters chosen to represent the various scenarios required by the run plan of Table 7.1, representing the full combination of parameters that are possible and to be tested in the simulation. For the sections from the origins to the hub and the main line section, distances (short and long), speed (high and low) are chosen. Production rates at the origins and the stack density are also chosen.

9.2.2 Summarised results

The simulation tests were performed for the setup described in Chapter 8. High level results are presented in Table 9.2, where the pertinent values are the train lengths, container volume throughput, efficiency and wagon fleet sizes. The results are included for the optimised and standard length runs. Table 9.2 presents the optimised train lengths and corresponding fleet size and efficiency for each run, in addition to the freight production and arrival volumes. Similar data is presented for the standard train length runs.

Calculations were performed post simulation to find the required fleet size for a specific configuration as this is an important output informing a railway about the capital requirements should they operate in this manner. Data from the simulation was used to calculate fleet sizes using Eqn (30), derived to find the fleet size from the work rate:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}} \tag{30}$$

Results are presented in Figure 9.1 and Figure 9.2 showing the relationship between train length, volume throughput, efficiency and fleet size. Figure 9.1 presents the data from Result Set 5, being the optimisation data set. In this data set the train lengths are varied and the final train length in each run is the optimal length determined by the optimisation algorithm.

The hub and spoke layout considered here has two distinct operations, a first mile journey from the origins to the Hub, and a main line journey from the Hub to the Port. These sections have different characteristics and must be treated separately.

Table 9.2 High level results

Run	Parameter	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Optimise length																	
TrainLength	1	35	37	38	28	15	15	15	15	40	38	37	36	33	33	37	35
TrainLength	2	34	26	33	34	15	23	16	15	37	38	23	39	34	39	31	37
TrainLength	3	44	44	43	44	40	40	40	40	42	40	44	42	41	40	41	43
TrainLength	4	47	49	50	43	40	41	50	40	43	41	46	47	42	45	41	40
Fleetsize	Quantity	6.07	6.06	5.69	6.33	6.29	5.62	7.24	6.38	1.50	1.41	1.39	1.49	9.99	8.83	11.10	10.87
Origin1Prod	teu	269	269	269	269	210	230	192	210	269	269	269	269	408	468	496	444
Origin2Prod	teu	559	559	559	559	1950	1980	1950	1950	559	559	559	559	2075	2075	2075	2075
Freightarrivals	teu	352	352	344	352	560	560	560	560	84	80	88	84	492	480	492	516
Wagon fleet	Quantity	268.00	267.00	245.00	279.00	252.00	225.00	290.00	256.00	63.00	57.00	62.00	63.00	410.00	354.00	456.00	468.00
Efficiency	Percentage	42.51	42.51	41.55	42.51	25.93	25.34	26.14	25.93	10.14	9.66	10.63	10.14	19.81	18.88	19.14	20.48
Standard length																	
TrainLength	3	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50
Fleetsize	Quantity	8.45	10.06	9.33	8.57	24.20	24.12	24.23	24.48	24.48	5.24	2.75	5.27	13.48	12.81	14.52	13.64
Origin1Prod	teu	560	560	560	560	560	560	560	560	560	560	560	560	800	640	800	800
Origin2Prod	teu	559	559	559	559	2075	2075	2075	2075	2075	559	559	559	2075	2075	2075	2075
Freightarrivals	teu	800	800	700	700	1700	1700	1600	1700	1700	800	400	600	1600	1800	1600	1600
Wagon fleet	Quantity	422.6	503.0	466.4	428.4	1209.9	1206.1	1211.7	1223.9	1223.9	262.2	137.3	263.4	674.2	640.5	725.9	682.2
Efficiency	Percentage	71.4924	71.4924	62.5559	62.5559	64.5161	64.5161	60.7211	64.5161	64.5161	71.4924	35.7462	53.6193	55.6522	66.2983	55.6522	55.6522
Wagon fleet	Quantity	154.575	235.985	221.370	149.445	957.910	981.140	921.700	967.885	1160.885	205.200	75.315	200.435	264.160	286.540	269.880	214.185
Freightarrivals	Quantity	448	448	356	348	1140	1140	1040	1140	1616	720	312	516	1108	1320	1108	1084

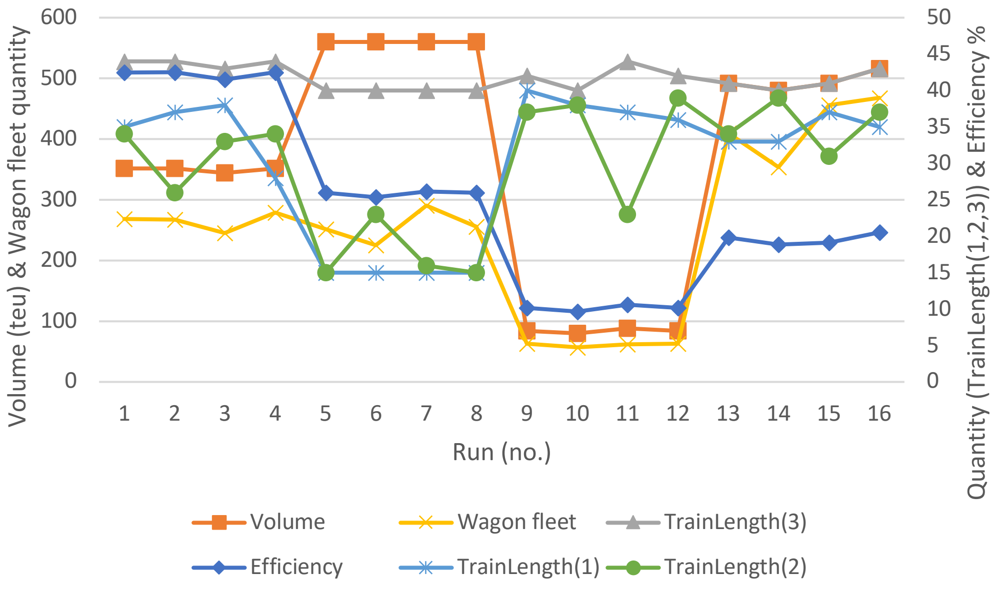


Figure 9.1 Volume relative to fleet size – optimal length

It should be noted that Runs 1-8 are single stacked, while Runs 9-16 are double stacked. Therefore, the volume throughput between them is not directly comparable. The maximum train length of 50 wagons on the main line (between the Hub and the Port) was not chosen during the optimisation as the efficiency target could be satisfied with lower train lengths.

On the journey from the Origins to the Hub (parameters TrainLength(1) & (2)), train lengths below the maximum length of 40 wagons were chosen by the algorithm. Note that for the low production and single stacked runs (5-8) the train lengths on this first mile journey were consistently low, generally 15 wagons long (the minimum), and with the highest value of 23 wagons. While higher than the other train lengths in this section, this is also low relative to the maximum constraint value of 40 wagons.

Considering the single stacked runs main line journey, it was found that shorter than maximum train lengths were chosen in all the runs, with Runs 1-4 being higher than Runs 5-8. Runs 1-4 all have high production rates.

Efficiency, ε_2, is defined as the fraction of total production which reaches the destination in each period, here the fraction of containers produced which reach the destination are considered as a measure of efficiency, presented in Eqn (31):

$$\sum_k \rho x_{12}(t_k) y_{12}(t_k) \geq \varepsilon_2 . P_{total}. \tag{31}$$

Considering the impact of the production rate on efficiencies in Table 9.3 shows that Runs 5-8 achieved a median efficiency of 25.93 % whereas efficiencies for Runs 1-4 are 42.51 %. This gives rise to the possibility that the high production rate, single stacked runs are more suited to train length optimisation.

Table 9.3 Median efficiency

Run	**1-4**	**5-8**	**9-12**	**13-16**
Efficiency median [%]	42.51	25.93	10.14	19.48

While the algorithm converged on shorter train lengths for Runs 5-8, their wagon fleet sizes were similar to Runs 1-4, evident in Figure 9.1 as the lower production rate in Runs 5-8 resulted in longer cycle times.

The efficiencies achieved during optimisation are significantly lower than the standard runs where fixed length trains are run. Note that the standard length runs are not driven by the efficiency or cost consideration, they run all trains at the maximum length if they can get all the components of the train.

The double stacked trains did not respond well to the train length optimisation with lower efficiencies and longer train lengths than the single stacked trains. If the container volume requirements are not sufficient to justify the transition to double stacked trains, it will not be cost effective. Double stacking is a capacity increase lever and will not be implemented before volumes increase above a certain threshold.

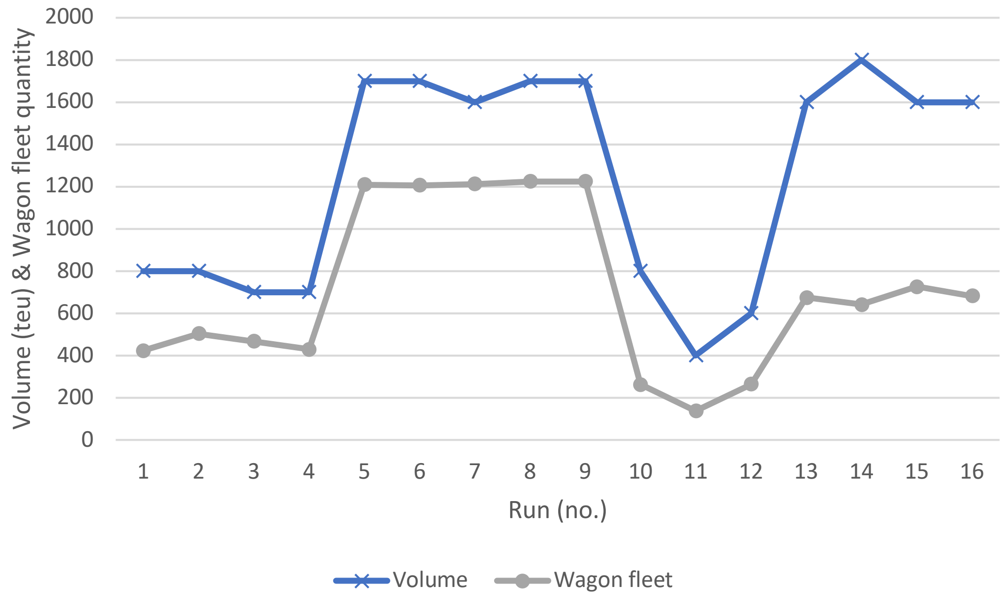


Figure 9.2 Volume relative to fleet size - Fixed length

A set of runs was conducted where no optimisation was performed, i.e. maximum train lengths were used. The results are presented in Figure 9.2 showing the container volume and wagon fleet size results over all the fixed train length runs. The trains ran with predetermined length of:

- TrainLength(1): 40 wagons
- TrainLength(2): 40 wagons
- TrainLength(3): 50 wagons
- TrainLength(4): 50 wagons

This set of runs corresponds to the *maximum length train scenario*, meaning that the railway compiles a train to the predetermined length and releases it. If the train is not ready, it will wait until the train is complete and the next slot is available before departure. Runs 1-8 are single stacked and Runs 9-16 are double stacked. Runs 1-4 and 9-12 have origins with low production rates, whereas in Runs 5-8 and 13-16 the production rates are high.

The data shows that during the high production runs the volume throughput is higher and the required wagon fleet size is higher, as is expected. To gain insight, this set of data must be compared with the optimised runs and these comparisons are presented in Figure 9.3.

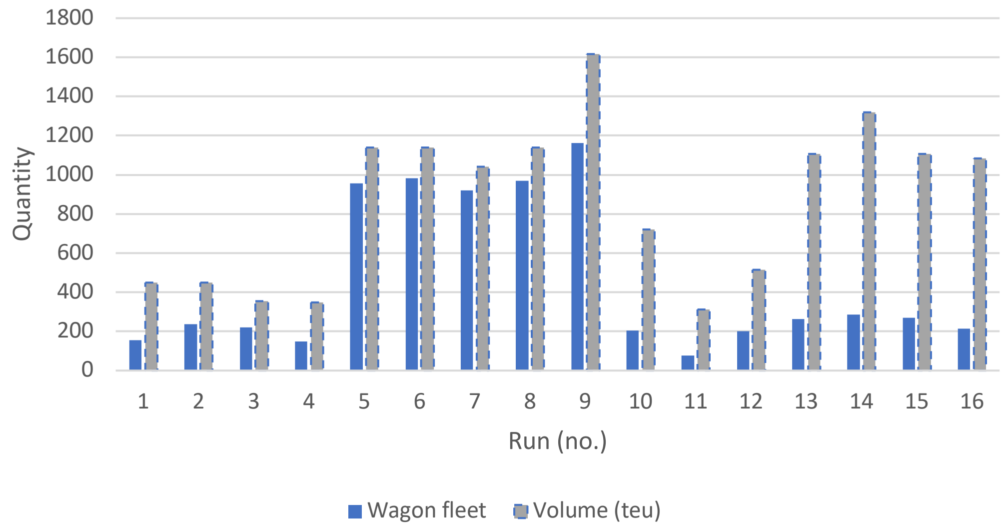


Figure 9.3 Comparison: Fixed vs Optimised length

Figure 9.3 shows the difference when the wagon fleet size and volume throughput data from the fixed length runs were defined as the reference and the data from the optimal length runs were subtracted from this reference value. Therefore, a positive result indicates that the fixed length run resulted in a higher wagon fleet size or volume throughput. For volume a positive result means that the fixed length run delivered a greater volume, which is the case for all the runs. While the optimised train lengths delivered less volumes, they were also at lower cost. Whether this is desirable will depend on the business requirements, which is outside the scope of this study.

Similarly, for the wagon fleet sizes a positive number indicates that the fixed train length fleets are larger than those required in the optimal length runs, this is the case for all the runs, although the magnitude of the difference varies. This means that there is a larger fleet to own, operate and maintain at greater cost to the business.

Interestingly, the double stacked, low production rate Runs (13-16) exhibit a large volume differential with a relatively small wagon fleet impact. This suggests that these conditions are less suitable to train length optimisation.

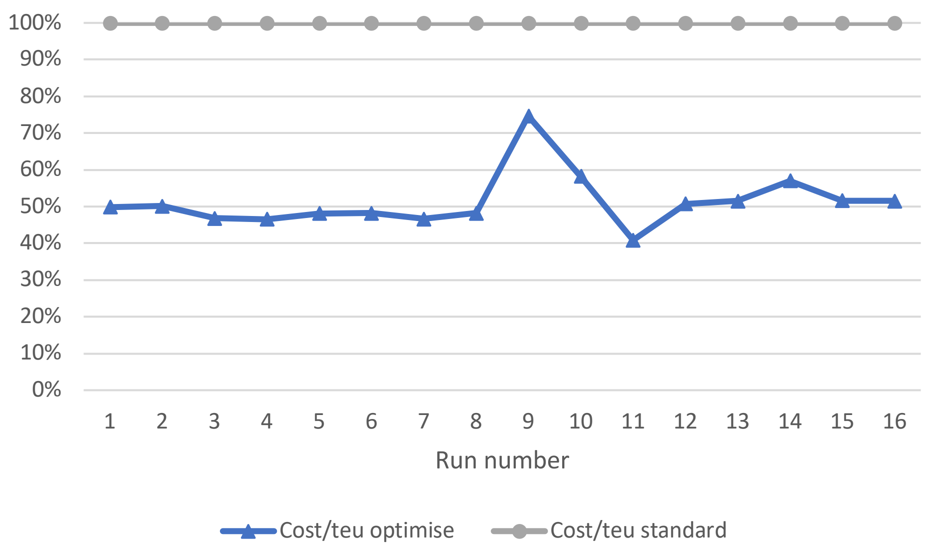


Figure 9.4 Relative cost per container[3]

In Figure 9.4, the relative cost difference is presented between the standard run and the optimised run. This is calculated using the objective function during the simulation. In the optimised run, the cost is consistently lower. This is as expected as the cost was minimised.

[3] A cost index is used to protect confidential information.

9.3 Comments on simulation results

In the simulation of the container corridor, less containers were transported than in the mathematical model and consequently the cost was lower. However, the simulation dealt with real world issues which the mathematical model does not. These include:

a) Availability of locomotives
b) Availability of wagons
c) Availability of crew
d) Train failures
e) Other delays due to external factors

These results highlight the different focus and application of the simulation and the mathematical model. While the mathematical model can be used to determine a theoretical throughput, it may be prudent to test this with the simulation to find how realistic it will be to achieve.

In performing tests with the model, it is evident in Figure 9.1 that shorter train lengths result in volume and cost benefits. During the simulation testing, it is evident in Figure 9.3 that the fixed length runs delivered higher volumes, but also that they required a larger wagon fleet. The size of a wagon fleet is determined by the turnaround time for a trip and indirectly by the volume demand. The cost to operate the fixed train length system was also higher, indicated in Figure 9.4.

The volume throughput is markedly higher in the standard runs with high production rates and single stacking. In all other cases the fixed length advantage has less impact, implying that, if necessary, a shorter train may be run, and this will not have a significant adverse effect on the volume throughput of the system. Specifically, it may be stated that shorter than maximum train lengths may be considered, when necessary, with less substantial volume throughput losses in a low production rate environment.

The most significant finding was the optimal train lengths found from the origin to the hub (the first mile) for the low production rate cases, being significantly shorter than the allowed maximum. This means that the railway could operate these sections at a cost saving and this corresponds with the findings of the mathematical model optimisation.

In the simulations, cost savings were achieved by running shorter length trains, conditions were identified where these had minimal impact, described above. It was also identified that double stacked, low production rate environments may negate any benefits of train length optimisation. In this case this is attributable to a volume threshold not being reached yet to justify double stacked trains.

The results have shown that compiling optimal length trains at hubs in a high production rate environment and more so at the origins does have a 30 to 50 % impact on the volume throughput of the system, with a corresponding cost benefit to the operation. Further benefits will be realised in a business where less rolling stock ownership is necessary.

These results shall be discussed in the next chapter, where the contribution to knowledge is considered.

Chapter Ten
Discussion

Much has been accomplished in the field of intermodal transportation over the last few decades, spanning multiple disciplines. In the literature, it is evident that this is a mature field and that there are various areas of study where it would be necessary to develop new approaches to advance the state of knowledge.

While existing research areas were concentrated around hub location problems, this study was concerned with developing a new approach to the emerging area of optimal train lengths. This involved developing a mathematical formulation to minimise the transport costs in an intermodal system. Current models were extended and developed in a different direction as this development focussed on train lengths and informed the development of a mathematical model and a logistic simulation. This approach then progressed to find optimal train lengths at minimum cost based on the environment, a previously unchartered area of study. The combination of techniques allowed the development of this problem from an engineering perspective.

Here, the development of capacity considering the train length, without infrastructure investment, was the motivation. The focus was on finding optimal train lengths to minimise the cost of transport. This was achieved by constructing a new mathematical model, conducting an optimisation, and furthermore, constructing a simulation of a hub and spoke environment. The primary contribution is the mathematical model and the optimisation and stands on its own merit. The simulation is a further contribution where the logistic environment is modelled, and the cost objective function (of the mathematical model) is used to make train length decisions. In essence, the optimisation and the simulation are separate and the decision on which a user would use depends on the application and the required output.

The mathematical model was developed based on a hub and spoke model. A bespoke optimisation solution was developed using a Genetic Algorithm. This was solved for a simple problem and for a corridor level optimisation problem. In both these cases the train lengths and departure decisions were optimised for a minimum cost objective.

The results indicate that the objective to achieve the minimum cost with constrained resources has been achieved in a limited sense. Optimal train lengths that are shorter than the maximum allowable train lengths have been found to have minimal negative impact on volume in a low production rate environment. However, this does not apply to a high production rate environment.

Furthermore, on the first mile of the journey it was possible to significantly reduce train lengths, specifically in a low production rate, single stacked environment, refer to Chapters 6 and 9. This did not apply to the double stacked cases, where longer train lengths were necessary in the first mile, refer to Chapter 9.

These insights represent a new area of thinking that is not found in the literature in this form. For the logistic industry, the direct benefits would be the ability to move volumes with less rolling stock and lower cost. It is a counterintuitive benefit as the industry would not normally consider reducing fleet sizes to maintain volumes due to varying train lengths.

To grow intermodal market share against road, one should consider the parameters of transportation cost, transit time, delay percentage, frequency and free time. Migration to intermodal transport is influenced by characteristics such as transportation distance, shipment size, accessibility, packaging, freight type, value, perishability and fragility.

These parameters are partly considered in this study where they are relevant, therefore transportation distance and shipment size were of significance. Increasing the train frequency and reducing transit time could attract volumes to intermodal transport. In the work presented here, train lengths were optimised and thus there would be more frequent use of slots, therefore the train would wait for less time to depart than if it were to wait to make up a predefined maximum length.

Therefore, running optimal length trains could contribute to growing intermodal market share as an additional benefit. This could be the subject of further study.

In running optimal length trains the results indicated that cycle times benefited positively, hence smaller rolling stock fleets are required. As trains may depart more frequently, railway companies have a lower risk of missing available slots and the resultant underutilisation of capital infrastructure.

The results have shown that compiling optimal length trains at hubs in a high production rate environment and more so at the origins does have a 30 to 50 % impact on the volume throughput of the system, with a corresponding cost benefit to the operation. Further benefits will be realised in a business where less rolling stock ownership is necessary.

Conclusions shall be drawn in the next chapter.

Chapter Eleven

Conclusion

In this study, the transportation network was examined based on the location and volumes of freight to be transported, in the context of an intermodal hub and spoke network. The intent was to minimise the cost of operation by varying the train lengths in the system and using existing infrastructure.

Based on available infrastructure and demand for goods in a particular scenario one could find an ideal logistic chain. The movements in that chain should be configured for the optimal volume throughput while considering the interaction with the other parts of the network.

The requirement for transportation arises as goods are produced and are transported in an evolving economy with new pressures arising from e-commerce and globalisation. Whereas rail is suited to large, established and consistent volumes, road is more suited to the just in time nature of the developing environment. While rail requires large, consistent loads to cover fixed costs, road is far more flexible. However, rail is less susceptible to externality costs and much more economical, in the order of 11 times per ton-km, (Kuo & Miller-Hooks, 2012). Therefore, combining rail and road to exploit their benefits and reduce their drawbacks would be beneficial. It is for this reason that the intermodal hub and spoke system has evolved worldwide.

To this end this thesis proposed and built a mathematical model representing a conceptual intermodal hub and spoke layout. The novel contribution here was the development of the model to handle train lengths and departure decisions. The constraints for the system were developed. The model was optimised mathematically through the development of a novel and bespoke genetic algorithm. A simulation model was developed separately to model operational complexities, such as delays, capacities and the balancing of resources, using the objective function of the mathematical model. The movement of freight through such a system has been modelled. Optimisations were performed to find the best train length for minimal cost.

While the mathematical model may be used to develop a theoretical view of the optimal or best solution, the simulation could assist is determining whether this solution can be implemented practically.

Upon conducting tests, optimisations and simulations in the operating environment, the results support the following conclusions:

- Optimisation did result in lower operating costs to achieve marginally lower volumes while meeting demand and other constraints, refer to Chapter 6.

- Testing through the simulation did result in shorter train lengths than the predefined maximum train length based on a cost driven objective function, however there was a 30 to 50 % resultant loss in volume, due to the real-world factors included, as was shown in Chapter 9.

- In running less than maximum length trains, the results indicated that cycle times benefited positively, hence smaller rolling stock fleets are required, resulting in a cost saving and significant capital benefit to a railway company, refer to Chapter 9.

- As trains may now depart more frequently, railways have a lower risk of missing slots and the resultant underutilisation of fixed infrastructure.

- Very short (at the minimum allowed) trains may be run in the first mile of the journey for single stacked trains in a low production rate environment, as was shown in Chapters 6 and 9.

- Running shorter length trains could contribute to growing intermodal market share as an additional benefit. This could be the subject of further study.

These conclusions have been made in the context that a logistic environment is a complex system with many variables, interfaces, and variations in how it can be set up, and one may develop many strategies for how it can be used to benefit the railway company and the economy.

During the development of this work, it was decided that due to the wide-ranging possibilities of operation, that a train make-up strategy was necessary to make decisions about when to release a train. Two basic scenarios were considered:

- Full train length - The train can wait to make up the full defined length.
- Slot priority - When the slot is open, the train can be released, irrespective of whether it is full length or shorter.

In analysing the results, the Slot priority scenario is chosen in the context tested. This set a framework for running optimised train lengths and defining different volume targets rather than running the predetermined maximum train lengths.

Of greater significance was the determination of the highly optimised train lengths on the first mile, implying that operations could be adjusted to make more efficient use of existing resources.

It was also found that as shorter turnaround times were achieved a railway could plan to use a smaller wagon fleet for a required service. They could plan this service using optimised train lengths with the knowledge that minimal losses in volume would result. This would have significant benefits in terms of capital and maintenance requirements.

Specific recommendations associated with the study:

i. Develop train plans considering the trade-off between running shorter length trains with reduced volumes and the savings related to a smaller rolling stock fleet.
ii. Adjust the geographic concentration of rolling stock based on where larger rolling stock fleets are required, for example if the circumstances require shorter trains on the first mile section, then smaller fleets may be required in that area.

This work is positioned at a tactical level i.e., it is not intended to change train lengths daily in an operational environment, nor is it intended to inform strategic planning to support infrastructure expansion. Instead, the primary benefit of this work would be to use operational modifications to extend the life of existing infrastructure by finding capacity that was previously inaccessible due to a fixed operating regime. A railway operator may use this work to determine the optimal train lengths to run in a specific environment.

Further work could involve using more sophisticated methods to determine the penalty parameters for the mathematical model.

Areas of potential future research include:

1. The energy saved by running shorter trains.

2. Do the benefits of standard gauge track become more worthwhile in the context of optimising train lengths and what conditions are required for this to work?
3. Investigate other optimisation algorithms to improve solutions.
4. Balancing of resources in different production rate environments.
5. Extend the study to foreign locations with much greater distances, expect a different interaction between the first mile and main line journeys.
6. Determine how to implement weather complications in the simulation.
7. Method to determine penalty parameters.
8. How does running higher frequency trains benefit intermodal market share?

Future research studies will be based on some of the above considerations.

Bibliography

Agamez-Arias, A. & Moyano-Fuentes, J., 2017. Intermodal transport in freight distribution: a literature review. Transport Reviews, 37(6), pp. 782-807.

Ahuja, R., Jha, K. & Liu, J., 2007. Solving Real Life Railroad Blocking Problems. Interfaces, 37(5), pp. 404-419.

Alumer, S. & Kara, B., 2008. Network hub location problems: The state of the art. European Journal of Operational Research, Volume 190, pp. 1-21.

Alvarez-SanJaime, O., Cantos-Sanchez, P., Moner-Colonques, R. & Sempere-Monerris, J., 2015. The impact on port competition of the integration of port and inland transport services. Transportation Research Part B, Volume 80, pp. 291-302.

Andersen, J., Crainic, T. & Christiansen, M., 2009. Service network design with asset management: Formulations and comparative analyses. Transportation Research Part C, pp. 197-207.

Anon., 2003. Organization for Economic Co-operation and Development. [Online]

Arkhipov, D., Wu, D., Wu, T. & Regan, A., 2020. A parallel genetic algorithm framework for transportation planning and logistic management. IEEE Access.

Arnold, P., Peeters, D. & Thomas, I., 2004. Modelling a rail/road intermodal transportation system. Transportation Research Part E, Volume 40, pp. 255-270.

Assad, A., 1980. Modelling of rail networks: Toward a routing/makeup model. Transportation Research B, pp. 101 - 114.

Bodin, L., Golden, B. & Schuster, A., 1980. A model for the blocking of trains. Transportation Research, Volume 14B, pp. 115-120.

Bontekoning, Y., Macharis, C. & Trip, J., 2004. Is a new applied transportation research field emerging? - A review of intermodal rail-truck freight transport literature. Transportation Research Part A, pp. 1-34.

Boyce, D. & Janson, B., 1980. A Discrete Transportation Network Design Problem with Combined Trip Distribution and Assignment. Transportation Research, Volume 14B, pp. 147-154.

Bruns, F., Goerigk, M., Knust, S. & Schobel, A., 2014. Robust load planning of trains in intermodal transportation. OR Spectrum, Volume 36, pp. 631-668.

Campbell, J. F., 1996. Hub location and the p-hub median problem. Operations Research, 44(6), pp. 923-935.

Campbell, J., 1994. Integer programming formulations of discrete hub location problems. European Journal of Operational Research, pp. 387-405.

Chiang, Y. & Roberts, P., 1980. A note on transit time and reliability for regular route trucking. Transportation Research, Volume 14B, pp. 59-65.

Crainic, T. & Kim, K., 2007. Intermodal Transportation. In: Handbook in OR & MS. s.l.:Elsevier, pp. 467-537.

Department of Transport, 2019. National Rail Policy Draft White Paper, s.l.: Department of Transport.

Eiselt, H. & Sandblom, C., 2012. Operations Research: A Model Based Approach. 2 ed. s.l.:Springer.

Farahani, R., Hekmatfar, M., Arabani, A. & Nikbakhsh, E., 2013. Hub location problems: A review of models, classification, solution techniques, and applications. Computers & Industrial Engineering, Volume 64, pp. 1096-1109.

Ferguson, T., n.d. Linear Programming: A Concise Introduction. s.l.:s.n.

Florian, M. & Gaudry, M., 1980. A conceptual Framework for the Supply Side in Transportation Systems. Transportation Research, Volume 14B, pp. 1-8.

Gaudry, M., 1980. A study of aggregate bi-modal urban travel supply, demand and network behavior using simultaneous equations with autoregressive residuals. Transportation Research, Volume 14B, pp. 29-58.

Graham, D., Cassady, C., Bowden, R. & LeMay, S., n.d. Modeling intermodal transportation systems: Establishing a common language, s.l.: s.n.

Guo, W., Atasoy, B., Beelaerts van Blokland, W. & Negenborn, R., 2020. Dynamic and Stochastic Shipment Matching Problem in Multimodal Transportation. Journal of the Transportation Research Board.

Hamblin, S., 2013. On the practical usage of genetic algorithms in ecology and evolution. Methods in Ecology and Evolution, Volume 3, pp. 184-194.

Hanczar, P., 2014. Solving inventory routing problems using location based heuristics. Operations Research and Decisions.

Harada, T. & Alba, E., 2020. Parallel Genetic Algorithms: A Useful Survey. ACM Computer Survey, August, 53(4), pp. 1-39.

Hayuth, Y., 1994. The overweight container problem and international intermodal transportation. Transportation Journal, 34(2), pp. 18-28.

Holguin-Veras, J. & Thorson, E., 2003. Modeling commercial vehicle empty trips with a first order trip chain model. Transportation Research Part B, Volume 37, pp. 129-148.

Ishfaq, R., 2010. Hub network design models for intermodal logistic: A dissertation, s.l.: The University of Alabama.

Jafarian-Moghaddam, A. R., 2021. Elastic train scheduling model. Applied Soft Computing, Volume 110.

Jonker, T. et al., 2019. Coordinated optimization of equipment operations in a container terminal. Flexible Services and Manufacturing Journal.

Kaneko, M. & Liu, S., 2015. Elimination of dominated strategies and inessential players. Operations Research and Decisions.

Karamouz, M., Zahraie, B., Kerachian, R. & Eslami, A., 2010. Crop pattern and conjunctive use management: A case study. Irrigation and Drainage, Volume 59, pp. 161-173.

Keaton, M., 1991. Service cost tradeoffs for carload freight traffic in the US rail industry. Transportation Research-A, Volume 25A, pp. 363-374.

Khan, S. et al., 2008. Design and implementation of an optimal fuzzy logic controller using genetic algorithm. Journal of Computer Science, 4(10), pp. 799-806.

Klatter, D., Luthi, H. & Schmedders, K., 2011. Operations Research Proceedings 2011. Zurich, Springer.

Kumar, A. & Anbanandam, R., 2020. Evaluating the interrelationships among inhibitors to intermodal railroad freight transport in emerging economies: A multi-stakeholder perspective. Transportation Research Part A, pp. 559-581.

Kuo, A. & Miller-Hooks, A., 2012. Developing Responsive Rail Services through collaboration. Transportation Research Part B, pp. 424-439.

Kurtulus, E. & Cetin, I., 2020. Analysis of modal shift potential towards intermodal transportation in short-distance inland container transport. Transport Policy, 89(2020), pp. 24-37.

Li, L., Negenborn, R. & De Schutter, B., 2013. A general framework for modeling intermodal transport networks. 2013 IEEE International Conference on Networking, Sensing and Control.

Li, Z., Lam, W. & Wong, S., 2012. Modeling intermodal equilibrium for bimodal transportation system design problems in a linear monocentric city. Transportation Research Part B, Volume 46, pp. 30-49.

Lin, Y. & Chen, X., 2014. Simulated annealing algorithm improved BP learning algorithm. Applied Mechanics and Materials, Volume 513, pp. 734-737.

Liu, P., Mu, D. & Gong, D., 2017. Eliminating overload trucking via a modal shift to achieve intercity freight sustainability: A system dynamics approach. Sustainability, Volume 9.

Martowibowo, S. & Damanik, B., 2021. Optimization of material removal rate and surface roughness of AISI 316L under dry turning process using genetic algorithm. Manufacturing Technology, 21(3).

Medeiros, A., Weijters, A. & van der Aalst, W., 2007. Genetic process mining: an experimental evaluation. Data Min Knowl Disc, Volume 14, pp. 245-304.

Meers, D. & Macharis, C., 2015. Prioritization in modal shift: determining a region's most suitable freight flows. European Transportation Research Review, Volume 7.

Meng, Q. & Wang, X., 2011. Intermodal hub-and-spoke network design: Incorporating multiple stakeholders and multi-type containers. Transportation Research Part B, pp. 724-742.

Monios, J. & Bergqvist, 2016. Drivers for vertical integration in the rail sector - using wagons as "relationship specific assets". The International Journal of Logistic Management, Volume 27, pp. 533-551.

Morlok, E., 1980. Types of transportation supply functions and their applications. Transportation Research-B, Volume 14B, pp. 9-27.
Mostert, M. & Limbourg, S., n.d. Intermodal network design for freight transportation in Belgium - location of rail/road and road/inland waterway terminals, s.l.: s.n.
Nicholas, S. & Buckley, T., 2019. South African Coal Exports Outlook, s.l.: Institute for Energy Economics and Financial Analysis.
Oviedo-Carranza, J. & Dominguez-Navarro, J., 2022. Optimal operation of a distributed generation microgrid based on the multi-objective genetic algorithms. Vigo (Spain), 20th International Conference on Renewable Energies and Power Quality.
Pittman, R. et al., 2020. The effectiveness of EC policies to move freight from road to rail: Evidence from CEE grain markets. Research in Transportation Business & Management.
Pourmohammad-Zia, N., Schulte, F., Souravlias, D. & Negenborn, R., 2020. Platooning of Automated Ground Vehicles to Connect Port and Hinterland: A Multi-objective Optimization Approach. Springer, pp. 428-442.
Racunica, I. & Wynter, L., 2005. Optimal location of intermodal freight hubs. Transportation Research Part B, pp. 453-477.
Saunders, M., Lewis, P. & Thornhill, A., 2007. Research Methods for Business Students Fourth Edition. s.l.:FT Prentice Hall.
Serper, E. & Alumur, S., 2016. The design of capacitated intermodal hub networks with different vehicle types. Transportation Research Part B, pp. 51-65.
Seymer, N., 1976. Intermodal comparisons of energy intensiveness in long distance transport. Transportation Research, Volume 10, pp. 275-279.
Sharma, J. & Singhal, R., 2014. Genetic algorithm and hybrid genetic algorithm for space allocation problems - A review. International Journal of Computer Applications, 95(4).
Sivanandam, S. & Deepa, S., 2008. Introduction to Genetic Algorithms. s.l.:Springer.

Skorin-Kapov, D., Skorin-Kapov, J. & O'Kelly, M., 1996. Tight linear programming relaxations of uncapacitated p-hub median problems. European Journal of Operational Research, Volume 94, pp. 582-593.

Song, D. & Dong, J., 2012. Cargo routing and empty container repositioning in multiple shipping service routes. Transportation Research Part B, pp. 1556-1575.

Sun, Y., 2020. Green and reliable freight routing problem in the road-rail intermodal transportation network with uncertain parameters: A fuzzy goal programming approach. Journal of Advanced Transportation, Volume 2020.

Talley, W. & Ng, M., 2020. Rail intermodal management at marine container terminals: Loading double stack trains. Transportation Research Part C, pp. 252-259.

Talvitie, A. & Dehghani, Y., 1980. Models for transportation level of service. Transportation Research-B, Volume 14B, pp. 87-99.

Talvitie, A. & Hasan, I., 1980. An equilibrium model system for transportation corridors and its application. Transportation Research, Volume 14B, pp. 165-174.

Tawfik, C. & Limbourg, 2019. Scenario-based analysis for intermodal transport in the context of service network design models. Transportation Research Interdisciplinary Perpectives.

Vale, C. & Ribeiro, I., 2014. Optimized intermodal roundtrips in transport networks. Paris, ResearchGate.

Vanderbei, R., 2008. Linear Programming: Foundations and Extensions. 3 ed. s.l.:Springer.

Wadhwa, S., Farahmand, K. & Kimberly, V., 2019. A deterministic mathematical model to support future investment decisions for developing inland container terminals. Research in Transport Economics, Volume 77.

Weigel, M., 1994. A railroad intermodal capacity model. Omaha, USA, Proceedings of the 1994 Winter Simulation Conference.

Wikipedia, 2016. Wikipedia. [Online]

Wilson, W., 1980. Performance models of flexibly routed transportation services. Transportation Research-B, Volume 14B, pp. 67-78.

Woroniuk, C. & Marinov, M., 2013. Simulation modelling to analyse the current level of utilisation of sections along a rail route. Journal of Transport Literature, 7(2), pp. 235-252.

Yan, S., Bernstein, D. & Sheffi, Y., 1995. Intermodal pricing using network flow techniques. Transportation Research-B, Volume 29B, pp. 171-180.

Zhang, Y., Atasoy, B. & Negenborn, R., 2021. Preference-Based Multi-Objective Optimization for Synchromodal Transport Using Adaptive Large Neighborhood Search. Journal of the Transportation Research Board.

Zunder, T., 2016. Recent developments with single wagon load services, policy and practice in Europe. Proceedings of the Federated conference on computer science and information systems, pp. 1097-1104.

Train length cost optimisation

```
% Using genetic algorithm process


% Constants
R1 = 150;  % Value for R1
R2 = 150;  % Value for R2
R3 = 100;  % Value for R3
R4 = 100;  % Value for R4
R5 = 100;  % Value for R5
R6 = 100;  % Value for R6
m = 4;     % Value for m
mH1 = 10; % Value for mH1
mO = 30000; % Overhead value
d21 = 170; % Value for d21
d31 = 158; % Value for d31
d12 = 630; % Value for d12
Dtotal1 = 200;
Dtotal2 = 400; % Value for Dtotal
rho = 2; % Value for rho
w2 = 120;  % Value for w2
w3 = 120;  % Value for w3
wH1 = 100; % Value for wH1
ORateP2 = 12; % Container production rate
ORateP3 = 15;


% Define the bounds
lb = [0;0;0;15;15;40];
ub = [1;1;1;40;40;50];


% define population size and number of variables
NPop = 50;
NVar = 6;


% Generate X chromosome

X  =  [randi([0,1],  63,  NPop);  randi([15,  40],  42,  NPop);
randi([40,50], 21, NPop)];


% Define the maximum number of iterations
maxIterations = 100;
```

```
% Initialize a variable to track the current iteration
iteration = 0;


% Initialize a variable to store the best solution found so far
bestSolution = [];
bestCost = Inf; % Initialize with a high value


% Start the loop
while iteration < maxIterations
    % Increment the iteration counter
    iteration = iteration + 1;
```

Container stack

```
% Generate w (container stack) matrix
w=zeros(63,NPop);
% Assign w2 to t0
w(1, :) = w2;


% Assign w3 to t0
w(22, :) = w3;


% Assign wH1 to t0
w(43, :) = wH1;
```

Calculate container stacks

```
for c = 1:NPop
    for r = 2:21
        if w(r-1,c)>rho*X(r+62,c)
        w(r,c) = w(r-1,c)+ORateP2-(rho*(X(r-1,c)*X(r+62,c)));
        else
            w(r,c) = w(r-1,c)+ORateP2;
        end

    end
```

```
    for r2 = 23:42
        if w(r2-1,c)>rho*X(r2+62,c)
        w(r2,c)          =         w(r2-1,c)+ORateP3-(rho*(X(r2-
1,c)*X(r2+62,c)));
        else
            w(r2,c) = w(r2-1,c)+ORateP3;
        end
    end


    for r3 = 44:63
        if w(r3-1,c)>rho*X(r3+62,c)
        w(r3,c) = w(r3-1,c)+rho*((X(r3-43,c)*X(r3+20,c))...
            +(X(r3-22,c)*X(r3+41,c))-(X(r3-1,c)*X(r3+62,c)));
        else
            w(r3,c) = w(r3-1,c)+rho*((X(r3-43,c)*X(r3+20,c))...
            +(X(r3-22,c)*X(r3+41,c)));
        end
    end


end
```

Calculate the cost objective function

```
CostX = zeros(21,NPop);
Cost_Sum = zeros(1,NPop);


for c4 = 1:NPop
cost = m * X(1:21,c4) .* X(64:84,c4) * d21 + ...
       m * X(22:42,c4) .* X(85:105,c4) * d31 + ...
       mH1  *  X(43:63,c4)  .*  X(106:126,c4)  *  d12  +  m0*(-
X(1:21,c4)+1) + ...
       m0*(-X(22:42,c4)+1) + m0*(-X(43:63,c4)+1) + ...
       R1 * max(0, X(64:84,c4) .* (1 - X(1:21,c4))) + ...
       R1 * max(0, X(85:105,c4) .* (1 - X(22:42,c4))) + ...
       R2 * max(0, X(106:126,c4) .* (1 - X(43:63,c4))) + ...
       R3 * max(0, Dtotal2 - (rho * sum(X(106:126,c4)))) + ...
       R4   *   max(0,   Dtotal1   -   (rho   *   sum(X(64:84,c4)   +
X(85:105,c4)))) + ...
       R5 * max(0, rho * X(1:21,c4).*(X(64:84,c4) - w(1:21,c4)))
+ ...
       R5   *   max(0,   rho   *   X(22:42,c4).*(X(85:105,c4)   -
w(22:42,c4))) + ...
```

```
       R6   *   max(0,   rho   *   X(43:63,c4).*(X(106:126,c4)   -
w(43:63,c4)));
CostX(:,c4) = cost;
Cost_Sum(1,c4) = sum(cost);
end
```

Genetics

```
% Sort chromosomes based on their costs
[sorted_costs, sorted_indices] = sort(Cost_Sum);


% Calculate the number of top chromosomes to keep (20%)
top_percentage = 0.2;
top_count = round(NPop * top_percentage);


% Select the best chromosomes and keep them
best_chromosomes = X(:, sorted_indices(1:top_count));


% Perform crossover on the remaining chromosomes (80%)
remaining_chromosomes = X(:, sorted_indices(top_count+1:end));


% Perform crossover operations on the remaining chromosomes
% Initialize offspring matrix
offspring = zeros(size(remaining_chromosomes));


% Perform crossover for pairs of parents
for i = 1:2:size(remaining_chromosomes, 2)
    % Select parent chromosomes
    parent1 = remaining_chromosomes(:, i);
    parent2 = remaining_chromosomes(:, i+1);

    % Perform double point crossover
    [child1, child2] = doublePointCrossover(parent1, parent2);

    % Place offspring into offspring matrix
    offspring(:, i) = child1;
    offspring(:, i+1) = child2;
end


% Generate random value P between 0 and 1
P = rand;
```

```
% Combine best chromosomes with offspring
new_population = [best_chromosomes, offspring];


% Update population size (if needed)
NPop = size(new_population, 2);
```

RESULTS

```
Costend = zeros(21,NPop);
Vol_end = zeros(21,NPop);


for c5 = 1:NPop
coste = m * new_population(1:21,c5) .* new_population(64:84,c5)
* d21 + ...
       m         *         new_population(22:42,c5)         .*
new_population(85:105,c5) * d31 + ...
       mH1         *         new_population(43:63,c5)         .*
new_population(106:126,c5) * d12 + ...
       m0*(-X(1:21,c4)+1)   +   m0*(-X(22:42,c4)+1)   +   m0*(-
X(43:63,c4)+1);
Costend(:,c5) = coste;
Vol_delivered      =      rho*(new_population(43:63,c5)      .*
new_population(106:126,c5));
Vol_end(:,c5) = Vol_delivered;
end


FinalCost = sum(Costend,1);
TotalVolume = sum(Vol_end,1);


% Find indices of chromosomes with the lowest FinalCost
[~, min_indices] = min(FinalCost);


% Select chromosomes with the lowest FinalCost
select_chromosomes    =    new_population(:,    FinalCost    ==
min(FinalCost));
Vol_select_chromosome = rho*sum(select_chromosomes(43:63,1) .*
...
    select_chromosomes(106:126,1));
Cost_Result = min(FinalCost);
% Check if a new best solution is found
    if min(FinalCost) < bestCost
        % Update the best solution
```

```
        bestSolution      =      select_chromosomes;      %
'select_chromosomes' contains the best solution
        bestCost = min(FinalCost);
    end

% Constraints
Wagon_constraint = R1 * max(0, select_chromosomes(64:84,1) .*
(1 - select_chromosomes(1:21,1))) + ...
       R1   *   max(0,   select_chromosomes(85:105,1)   .*   (1   -
select_chromosomes(22:42,1))) + ...
       R2   *   max(0,   select_chromosomes(106:126,1)   .*   (1   -
select_chromosomes(43:63,1)));


Demand_constraint   =    R3   *   max(0,   Dtotal2   -   (rho   *
sum(select_chromosomes(106:126,1)))) + ...
       R4      *      max(0,      Dtotal1      -      (rho      *
sum(select_chromosomes(64:84,1)                                 +
select_chromosomes(85:105,1))));


Container_constraint     =     R5     *     max(0,     rho     *
select_chromosomes(1:21,1).*...
    (select_chromosomes(64:84,1) - w(1:21,min_indices))) + ...
       R5           *           max(0,           rho           *
select_chromosomes(22:42,1).*(select_chromosomes(85:105,1)...
       -   w(22:42,min_indices)))   +   R6   *   max(0,   rho   *
select_chromosomes(43:63,1).*...
       (select_chromosomes(106:126,1) - w(43:63,min_indices)));

% Display iteration information or progress

    % Check termination condition (if necessary)
    % Terminate the loop if the termination condition is met
end

% Display or return the best solution found
disp('Best solution:');
disp(bestSolution);
```

```
disp(['Best cost: ', num2str(bestCost)]);
```

Output data

```
% Load the existing table
loadedTable = load('results_table2.mat');
resultsTable2 = loadedTable.resultsTable2;
```

```
% Define the names of your variables or columns
variableNames                                                              =
{'ORateP2','ORateP3','Dtotal','w2','w3','wH1','R1','R2','R3','
R4','R5','R6',...
    'Cost_Result',
'Vol_select_chromosome','select_chromosomes'};


% Create a table with results

newTable2                                                                  =
table(ORateP2',ORateP3',Dtotal',w2',w3',wH1',R1',R2',R3',R4',R
5',R6',...

Cost_Result',Vol_select_chromosome',select_chromosomes','Varia
bleNames', variableNames);
resultsTable2 = [resultsTable2; newTable2];


% Save the table to a .mat file for subsequent runs
save('results_table2.mat', 'resultsTable2');
```

```
function [offspring1, offspring2] =
doublePointCrossover(parent1, parent2)
    % Get the length of the chromosome
    chromosome_length = numel(parent1);
    chromosome_length1 = chromosome_length/2;


 % Randomly select two crossover points for the first half of
the chromosome
    crossover_point1 = randi(chromosome_length1 - 1);
    crossover_point2 = crossover_point1 +
randi(chromosome_length1 - crossover_point1);


    % Randomly select two crossover points for the second half
of the chromosome
    crossover_point3 = randi(chromosome_length -
chromosome_length1 - 1) + chromosome_length1;
    crossover_point4 = crossover_point3 +
randi(chromosome_length - crossover_point3);


    % Initialize offspring
    offspring1 = zeros(size(parent1));
    offspring2 = zeros(size(parent2));
```

```
% Perform crossover for the first half
    offspring1(1:crossover_point1) =
parent1(1:crossover_point1);
    offspring1(crossover_point1 + 1:crossover_point2) =
parent2(crossover_point1 + 1:crossover_point2);
    offspring1(crossover_point2 + 1:chromosome_length1) =
parent1(crossover_point2 + 1:chromosome_length1);


    offspring2(1:crossover_point1) =
parent2(1:crossover_point1);
    offspring2(crossover_point1 + 1:crossover_point2) =
parent1(crossover_point1 + 1:crossover_point2);
    offspring2(crossover_point2 + 1:chromosome_length1) =
parent2(crossover_point2 + 1:chromosome_length1);


    % Perform crossover for the second half
    offspring1(chromosome_length1 + 1:crossover_point3) =
parent1(chromosome_length1 + 1:crossover_point3);
    offspring1(crossover_point3 + 1:crossover_point4) =
parent2(crossover_point3 + 1:crossover_point4);
    offspring1(crossover_point4 + 1:end) =
parent1(crossover_point4 + 1:end);


    offspring2(chromosome_length1 + 1:crossover_point3) =
parent2(chromosome_length1 + 1:crossover_point3);
    offspring2(crossover_point3 + 1:crossover_point4) =
parent1(crossover_point3 + 1:crossover_point4);
    offspring2(crossover_point4 + 1:end) =
parent2(crossover_point4 + 1:end);




% Define the ranges for each group of rows
lowerBound_row64to105 = 15;
upperBound_row64to105 = 40;
lowerBound_row106to126 = 40;
upperBound_row106to126 = 50;


    % Generate random value P between 0 and 1
P = rand()*2-0.5;


% Modify offspring by multiplying one by P and the other by
(1-P)
```

```
 offspring1((chromosome_length1+1):end) =
round(parent2((chromosome_length1+1):end)...
    * P + parent1((chromosome_length1+1):end) * (1 - P));
 offspring2((chromosome_length1+1):end) =
round(parent1((chromosome_length1+1):end)...
   * (1 - P) + parent2((chromosome_length1+1):end) * P);


% Restrict values within the specified ranges
offspring1(64:105) = max(lowerBound_row64to105,
min(offspring1(64:105), upperBound_row64to105));
offspring1(106:126) = max(lowerBound_row106to126,
min(offspring1(106:126), upperBound_row106to126));


offspring2(64:105) = max(lowerBound_row64to105,
min(offspring2(64:105), upperBound_row64to105));
offspring2(106:126) = max(lowerBound_row106to126,
min(offspring2(106:126), upperBound_row106to126));
end
```

Appendix B – Simulation optimised runs

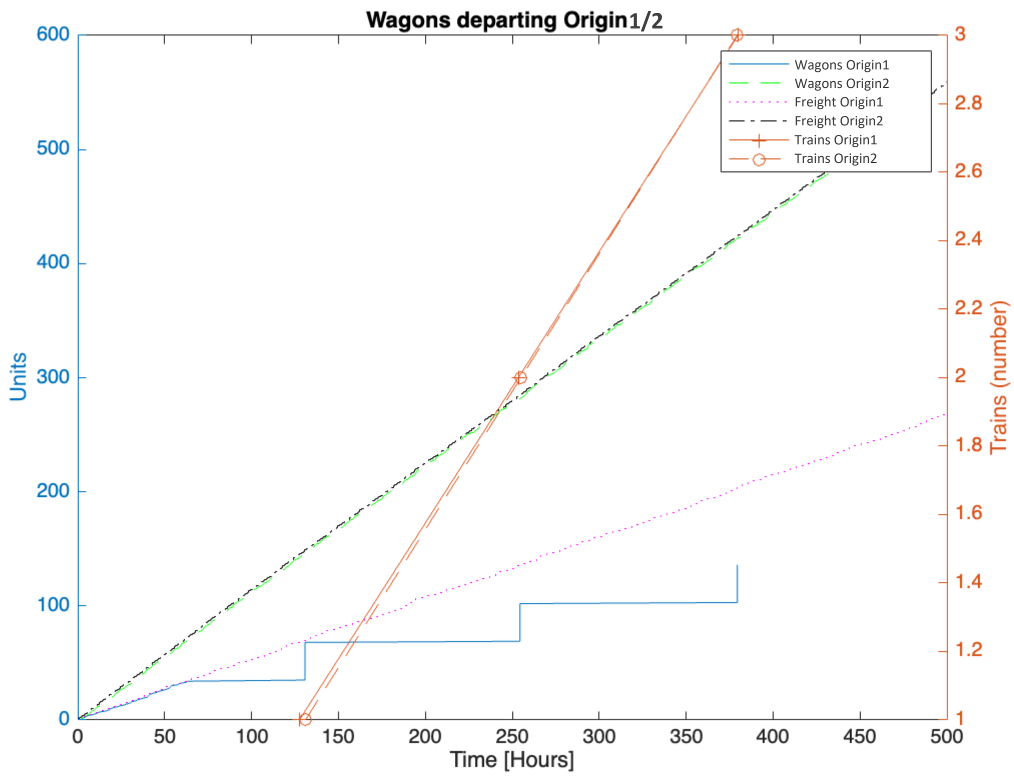
A large volume of data is generated during every simulation. This appendix contains an extract of the data which shows the performance relative to this study. These simulation runs are presented in Section 9.2.

Data analysis - Run 1

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

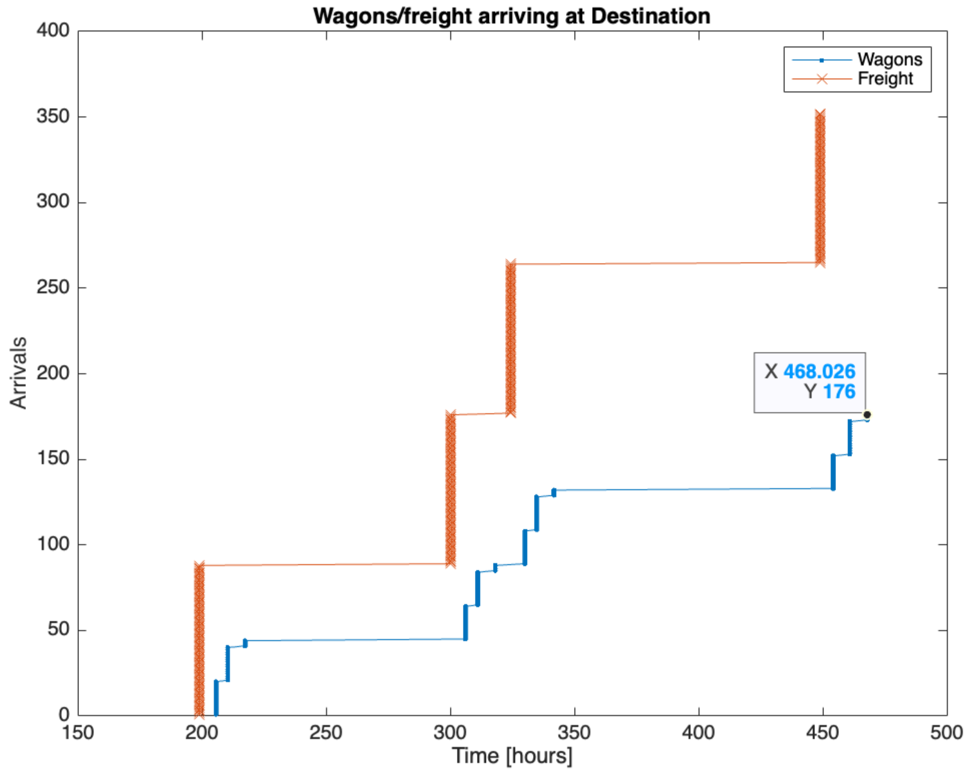
Origin 1 and 2 are production points.




```
Origin1Prod = 269
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
    35    34    44    47
```

The maximum wagon turnaround time is:

```
ans = 759.0637
WagonsDest = 176
TrainsDest = 4
TrainsDestDay = 0.1920
Freightarrivals = 352
Efficiency = 42.5121
CostTotal = 1.8423e+06
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

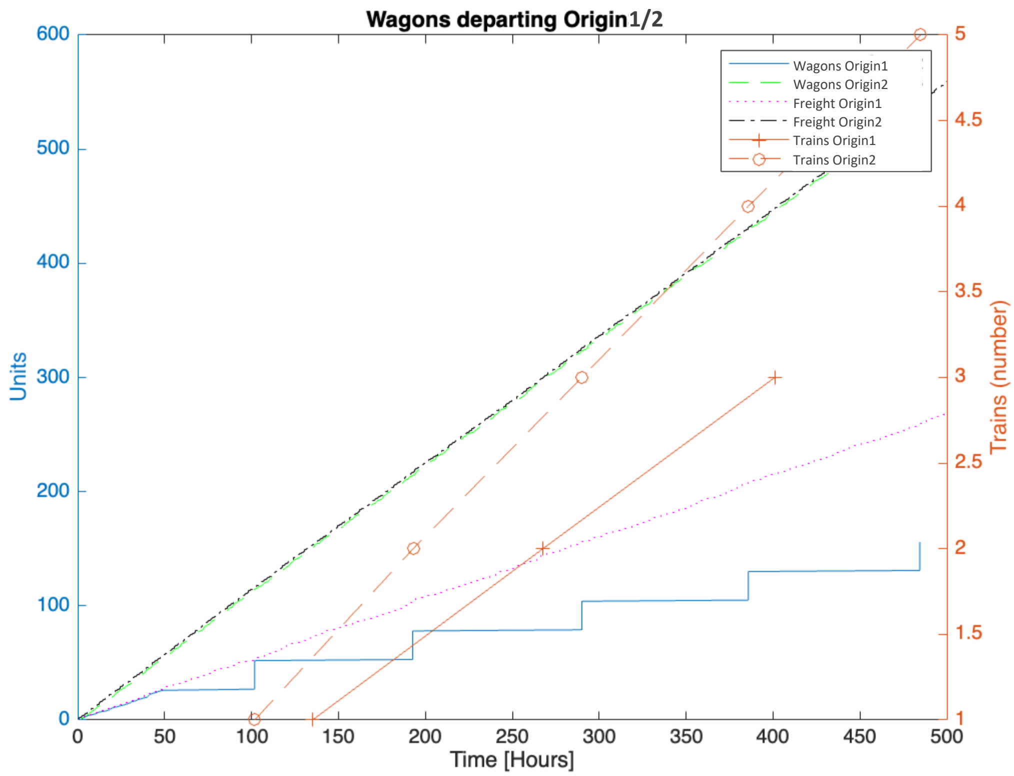
```
Fleetsize = 6.0725
```

Data analysis - Run 2

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

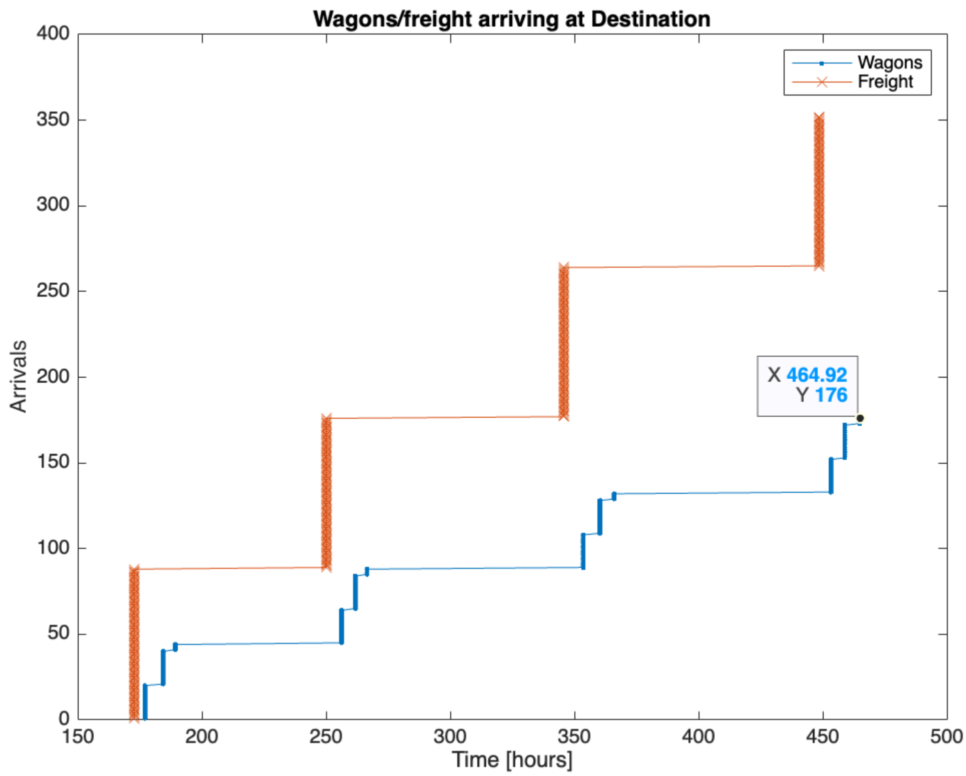
Origin 1 and 2 are production points.




```
Origin1Prod = 269
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
    37    26    44    49
```

The maximum wagon turnaround time is:

```
ans = 757.6841
WagonsDest = 176
TrainsDest = 4
TrainsDestDay = 0.1920
Freightarrivals = 352
Efficiency = 42.5121
CostTotal = 1.8632e+06
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

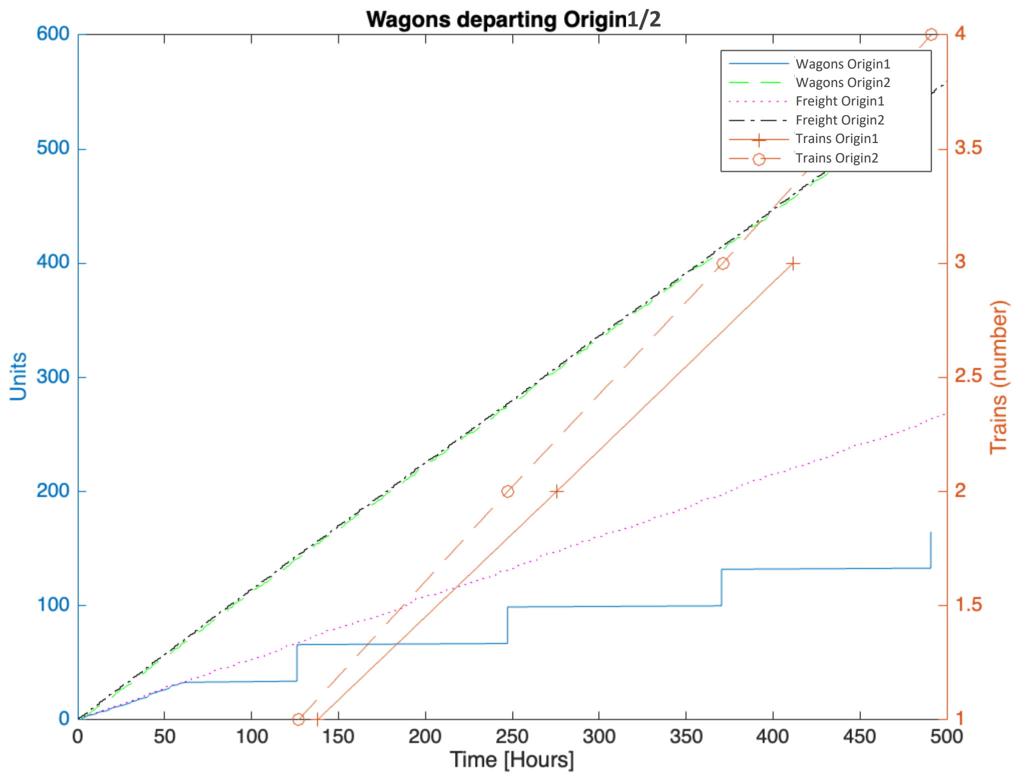
```
Fleetsize = 6.0615
```

Data analysis - Run 3

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

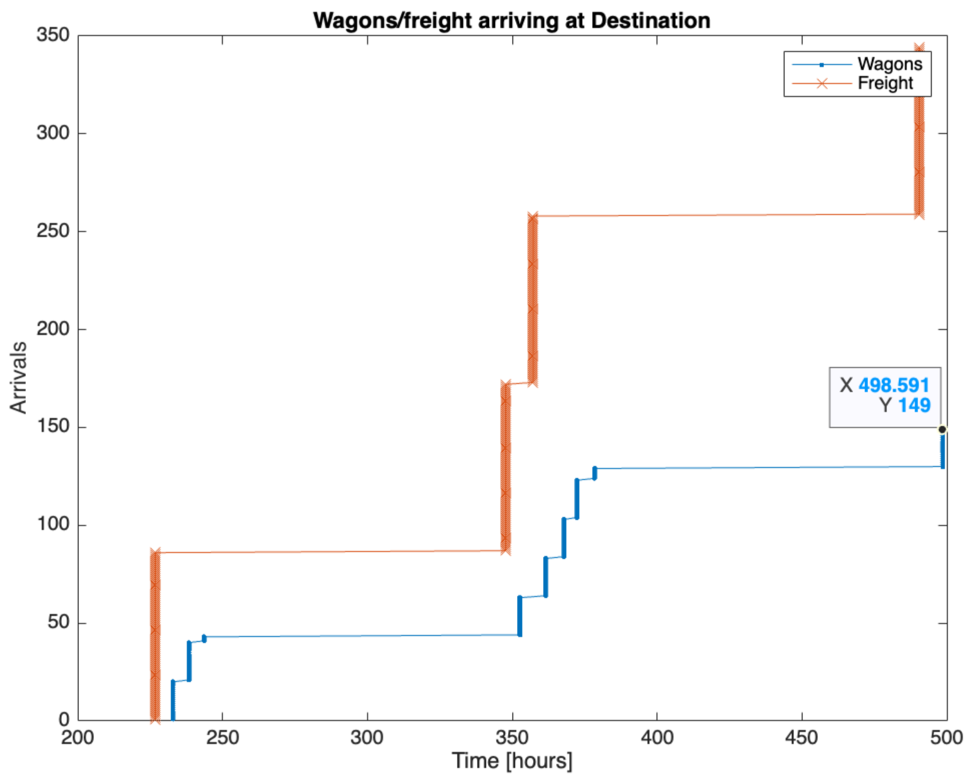
Origin 1 and 2 are production points.




```
Origin1Prod = 269
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
    38     33     43     50
```

The maximum wagon turnaround time is:

```
ans = 723.3819
WagonsDest = 169
TrainsDest = 3.9302
TrainsDestDay = 0.1887
Freightarrivals = 344
Efficiency = 41.5459
CostTotal = 4.3213e+06
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

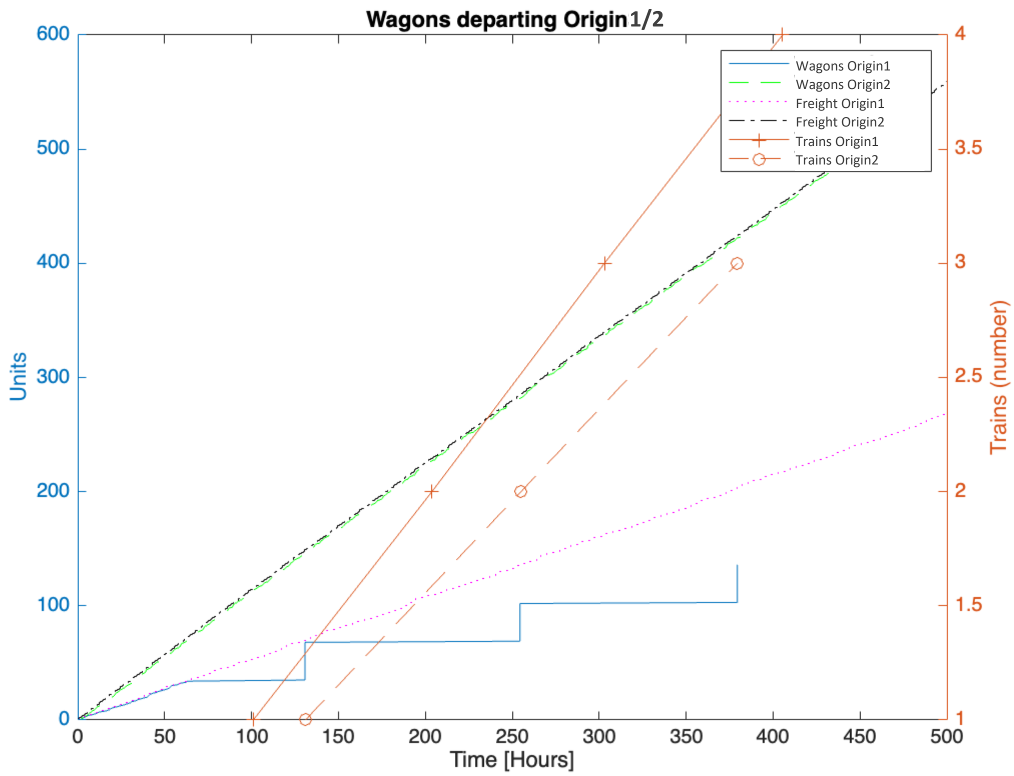
```
Fleetsize = 5.6861
```

Data analysis - Run 4

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

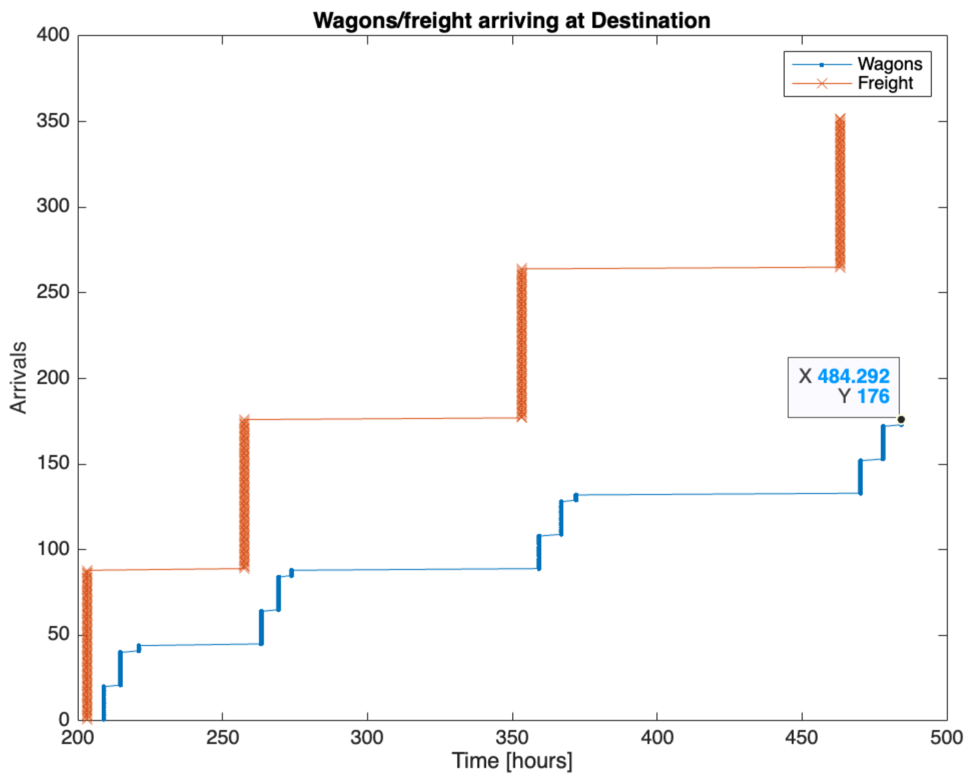
Origin 1 and 2 are production points.




```
Origin1Prod = 269
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
    28     34     44     43
```

The maximum wagon turnaround time is:

```
ans = 791.0342
WagonsDest = 176
TrainsDest = 4
TrainsDestDay = 0.1920
Freightarrivals = 352
Efficiency = 42.5121
CostTotal = 4.3763e+06
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

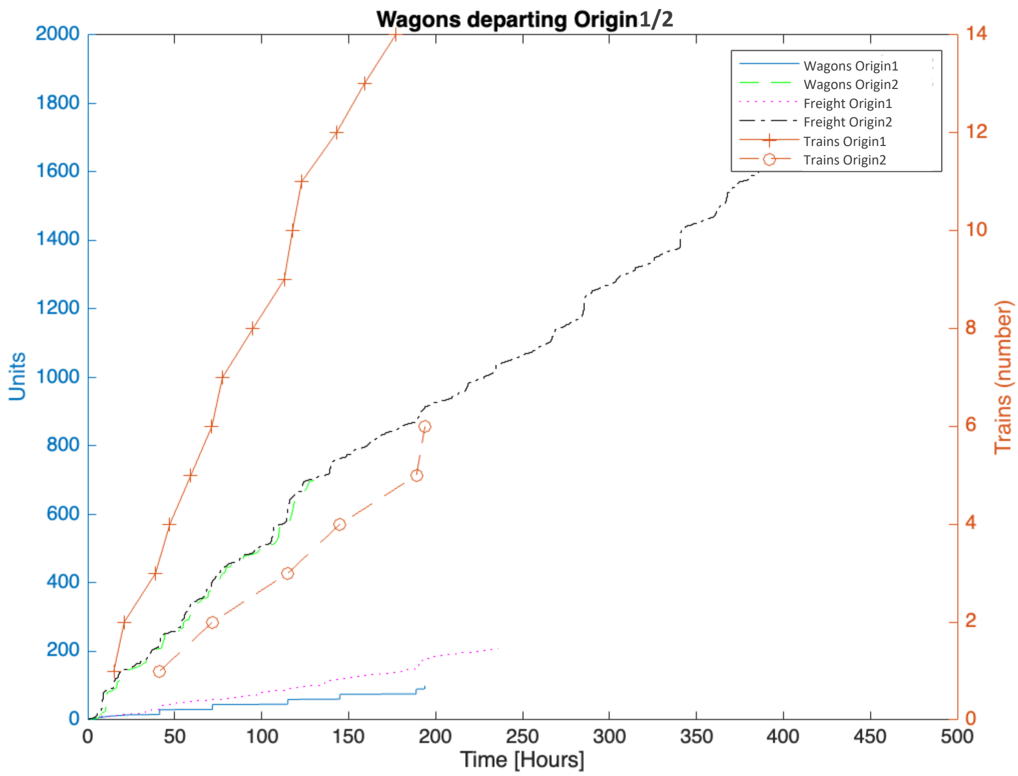
```
Fleetsize = 6.3283
```

Data analysis - Run 5

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

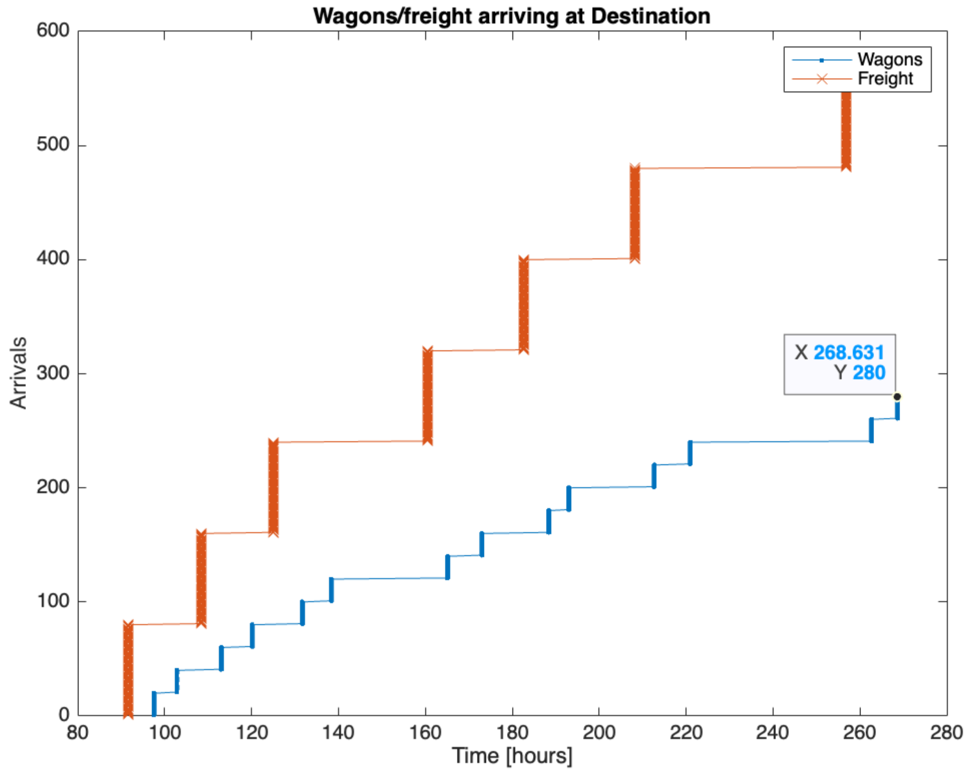
Origin 1 and 2 are production points.




```
Origin1Prod = 210
Origin2Prod = 1950
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
    15    15    40    40
```

The maximum wagon turnaround time is:

```
ans = 449.0904
WagonsDest = 280
TrainsDest = 7
TrainsDestDay = 0.3360
Freightarrivals = 560
Efficiency = 25.9259
CostTotal = 2922274
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

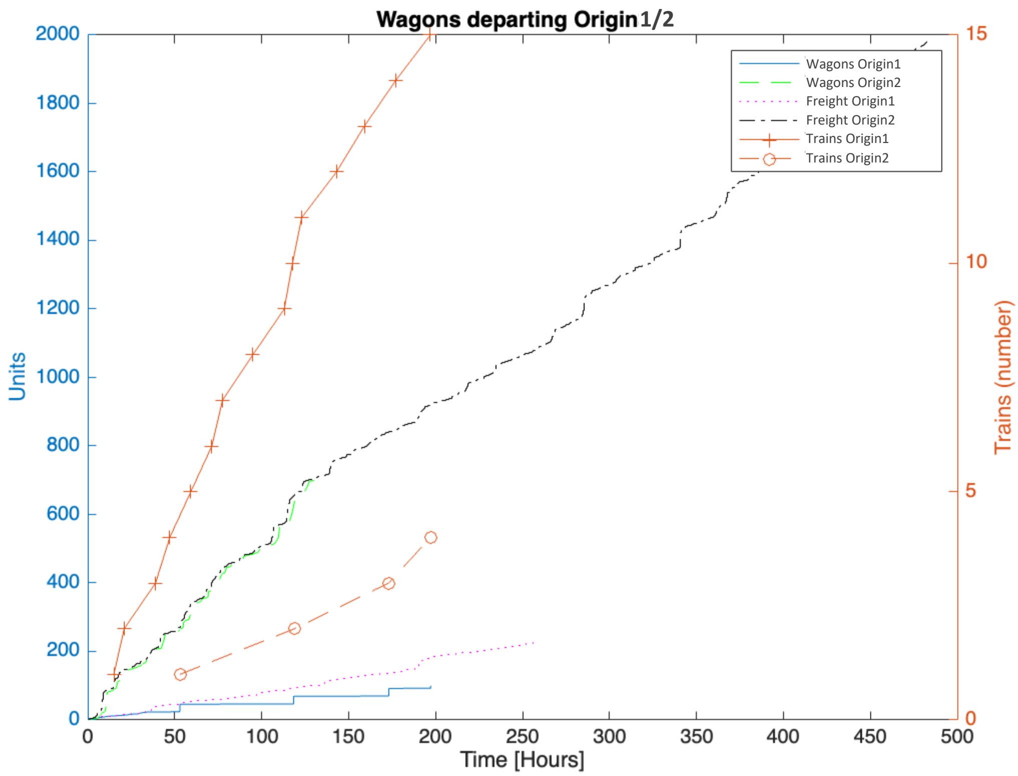
```
Fleetsize = 6.2873
```

Data analysis - Run 6

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

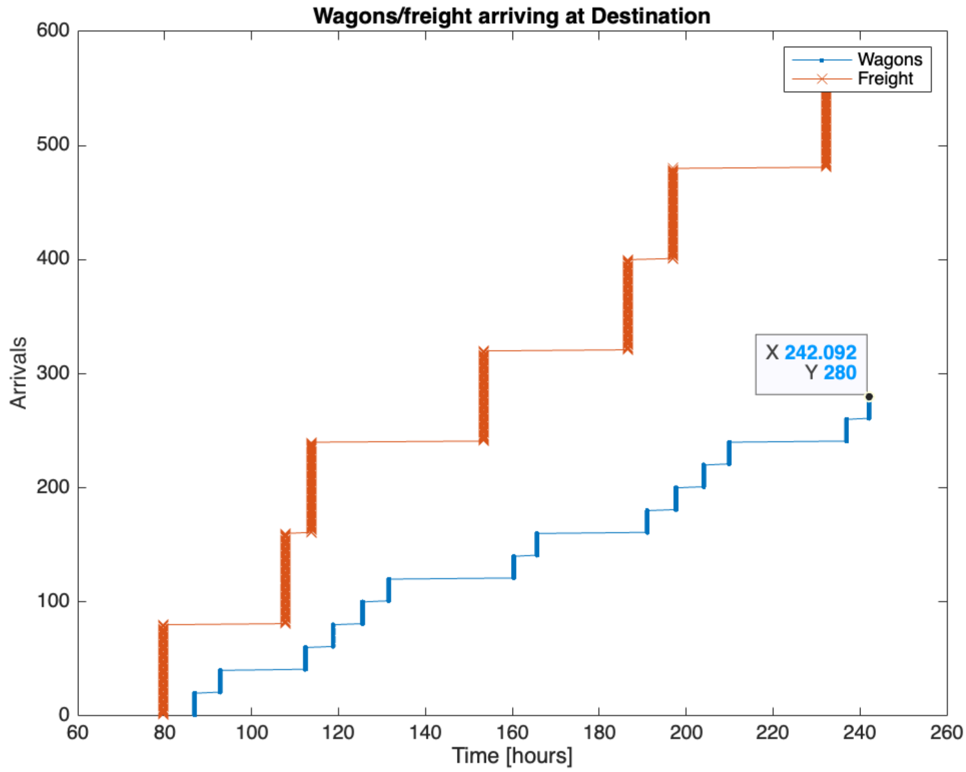
Origin 1 and 2 are production points.




```
Origin1Prod = 230
Origin2Prod = 1980
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
    15    23    40    41
```

The maximum wagon turnaround time is:

```
ans = 401.4878
WagonsDest = 280
TrainsDest = 7
TrainsDestDay = 0.3360
Freightarrivals = 560
Efficiency = 25.3394
CostTotal = 2.9350e+06
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

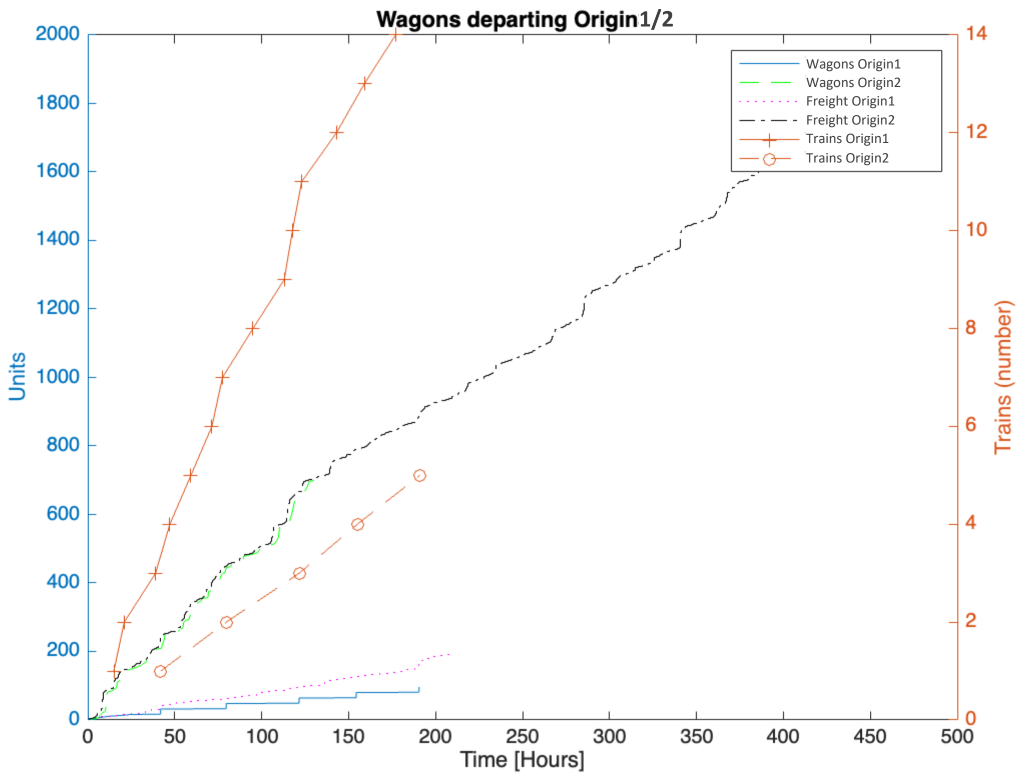
```
Fleetsize = 5.6208
```

Data analysis - Run 7

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

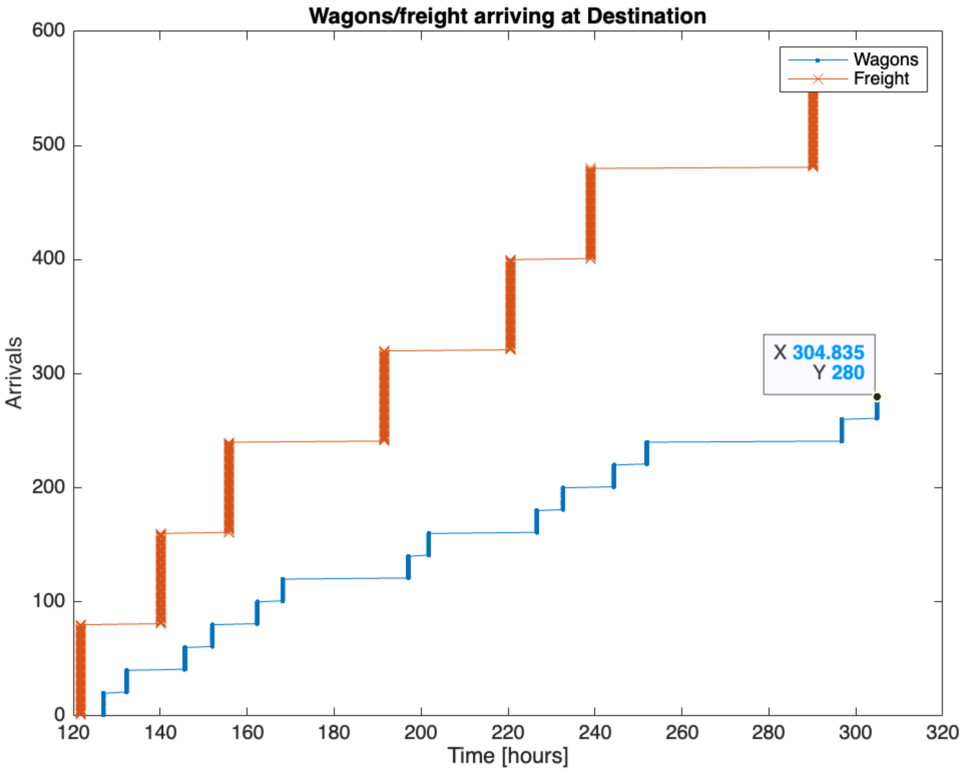
Origin 1 and 2 are production points.




```
Origin1Prod = 192
Origin2Prod = 1950
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
     15      16      40      50
```

The maximum wagon turnaround time is:

```
ans = 517.4489
WagonsDest = 280
TrainsDest = 7
TrainsDestDay = 0.3360
Freightarrivals = 560
Efficiency = 26.1438
CostTotal = 6915984
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

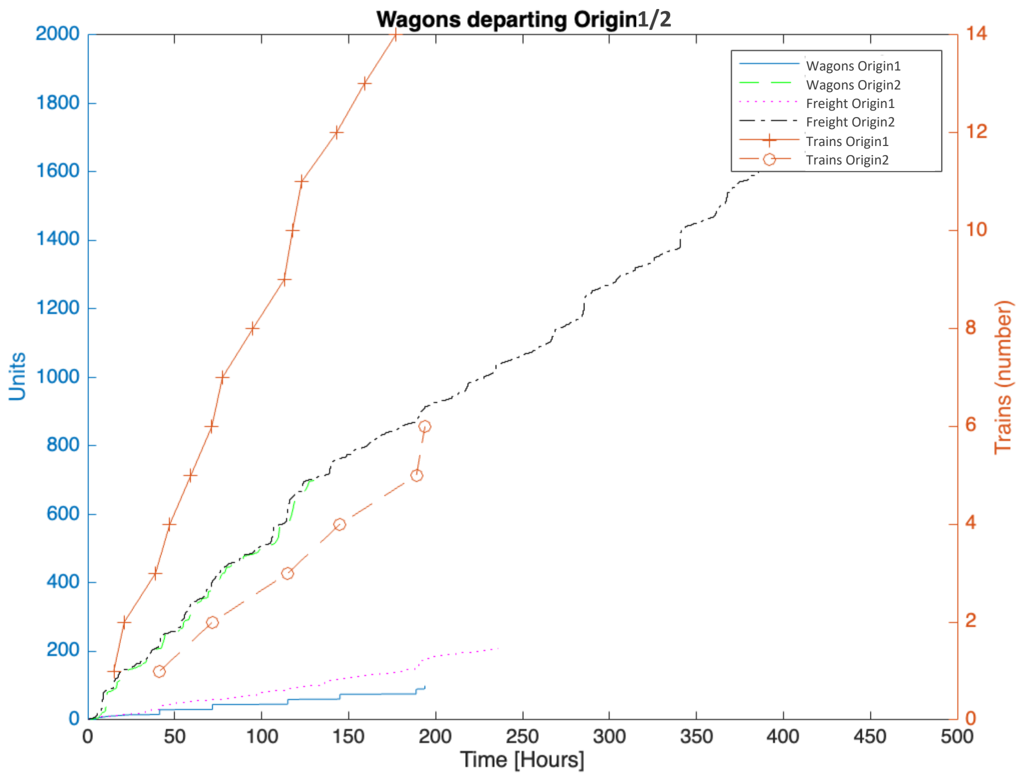
```
Fleetsize = 7.2443
```

Data analysis - Run 8

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

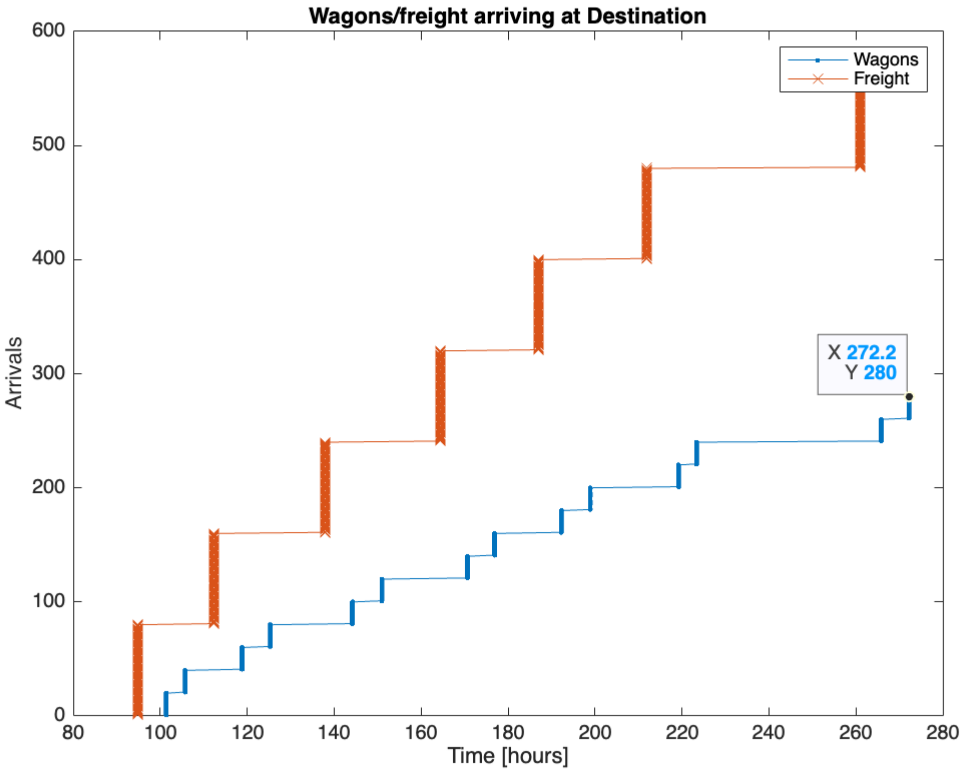
Origin 1 and 2 are production points.




```
Origin1Prod = 210
Origin2Prod = 1950
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
    15    15    40    40
```

The maximum wagon turnaround time is:

```
ans = 455.4371
WagonsDest = 280
TrainsDest = 7
TrainsDestDay = 0.3360
Freightarrivals = 560
Efficiency = 25.9259
CostTotal = 6927588
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

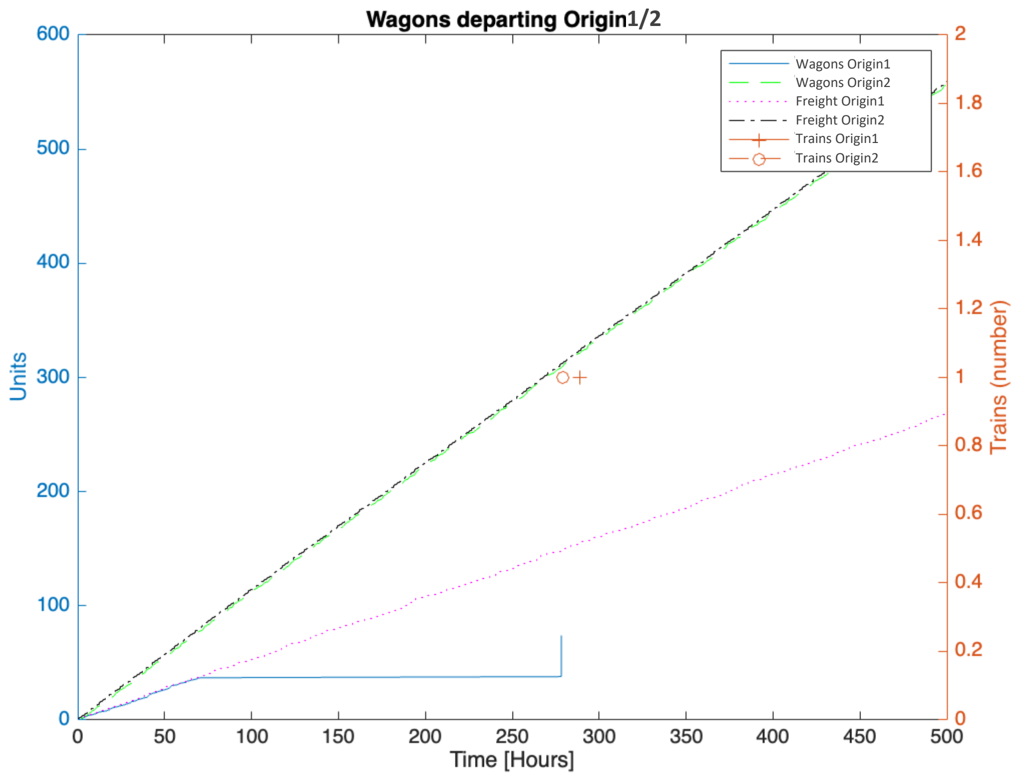
```
Fleetsize = 6.3761
```

Data analysis - Run 9

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

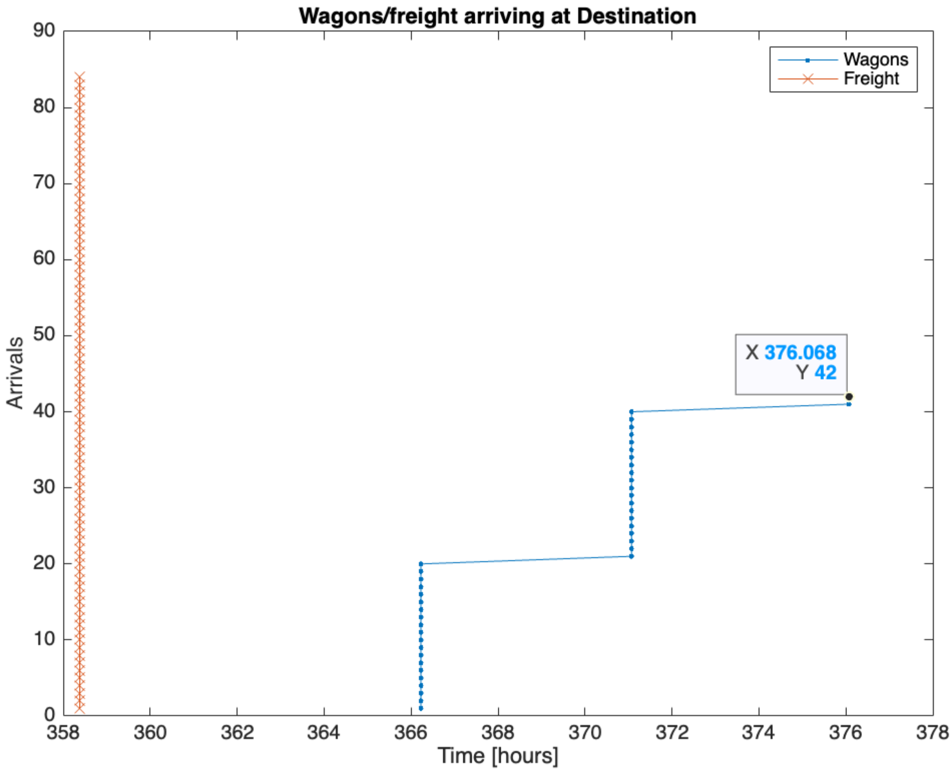
Origin 1 and 2 are production points.




```
Origin1Prod = 269
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
     40     37     42     43
```

The maximum wagon turnaround time is:

```
ans = 747.7369
WagonsDest = 42
TrainsDest = 1
TrainsDestDay = 0.0480
Freightarrivals = 84
Efficiency = 10.1449
CostTotal = 4.5855e+05
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

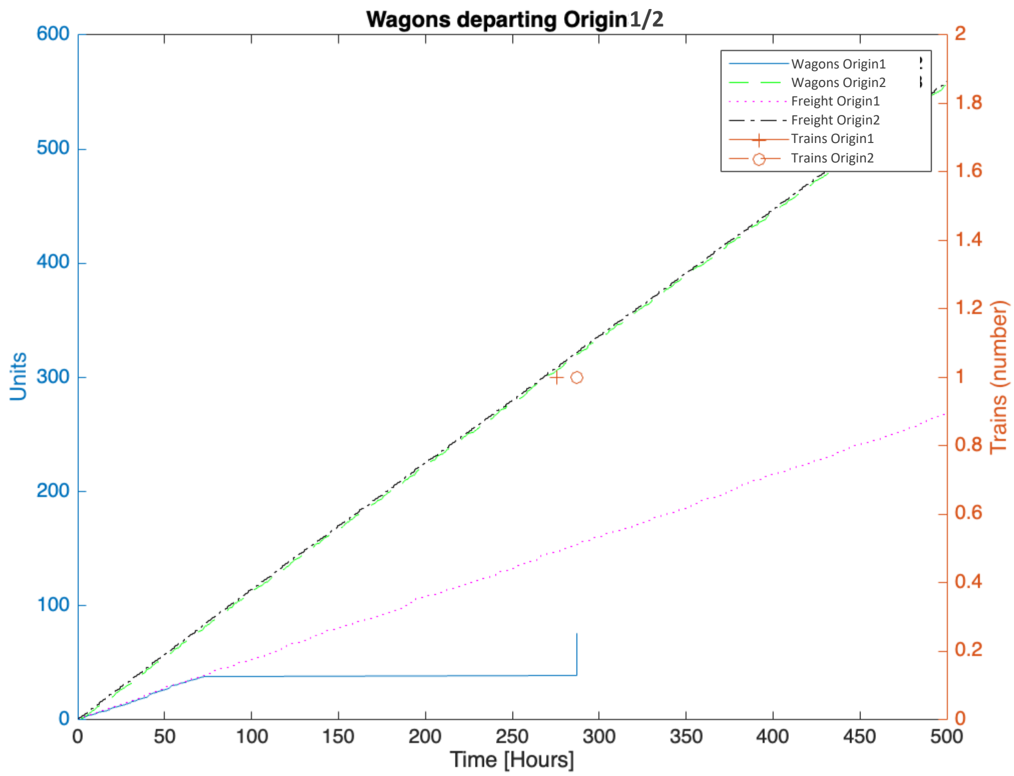
```
Fleetsize = 1.4955
```

Data analysis - Run 10

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

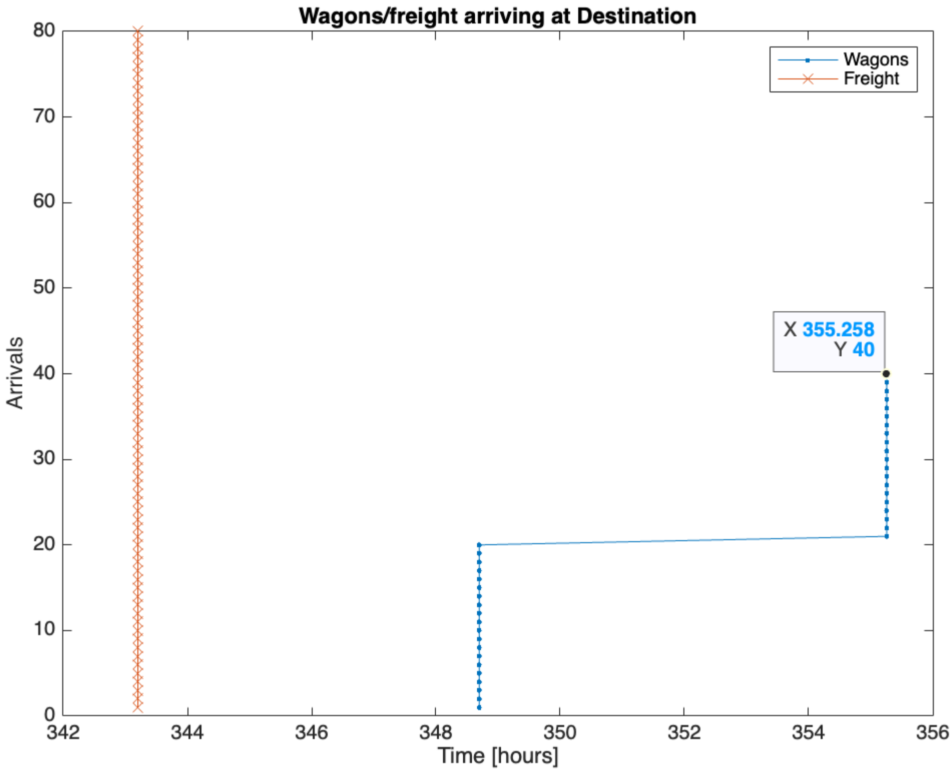
Origin 1 and 2 are production points.




```
Origin1Prod = 269
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
     38     38     40     41
```

The maximum wagon turnaround time is:

```
ans = 706.1161
WagonsDest = 40
TrainsDest = 1
TrainsDestDay = 0.0480
Freightarrivals = 80
Efficiency = 9.6618
CostTotal = 4.3824e+05
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

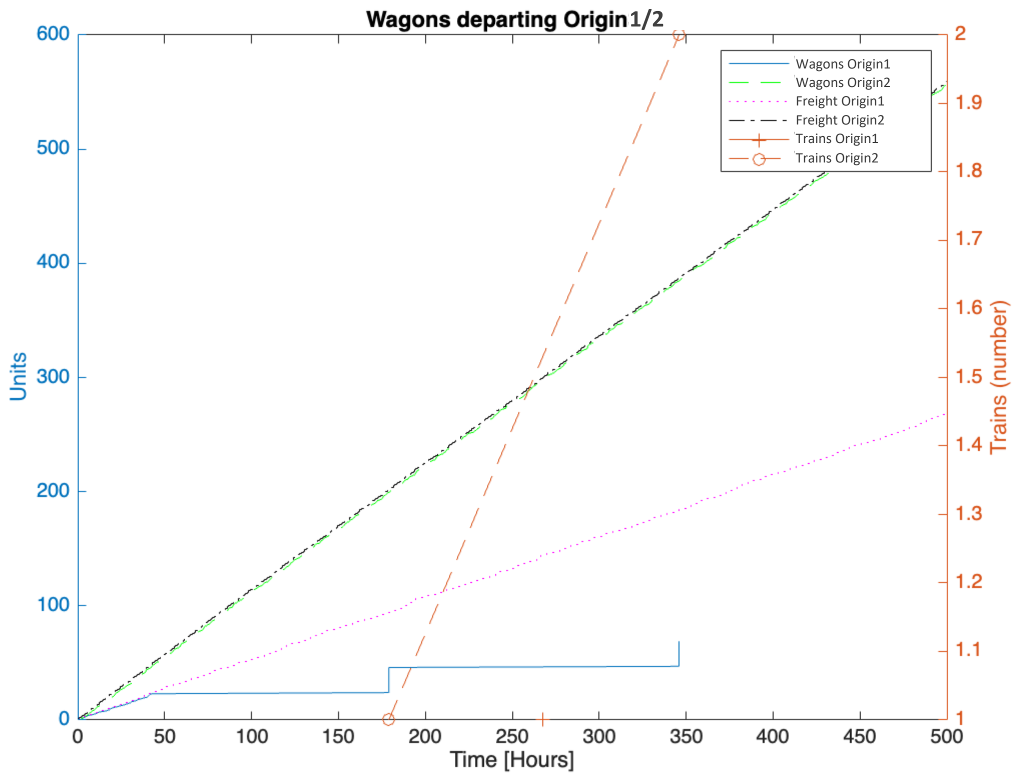
```
Fleetsize = 1.4122
```

Data analysis - Run 11

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

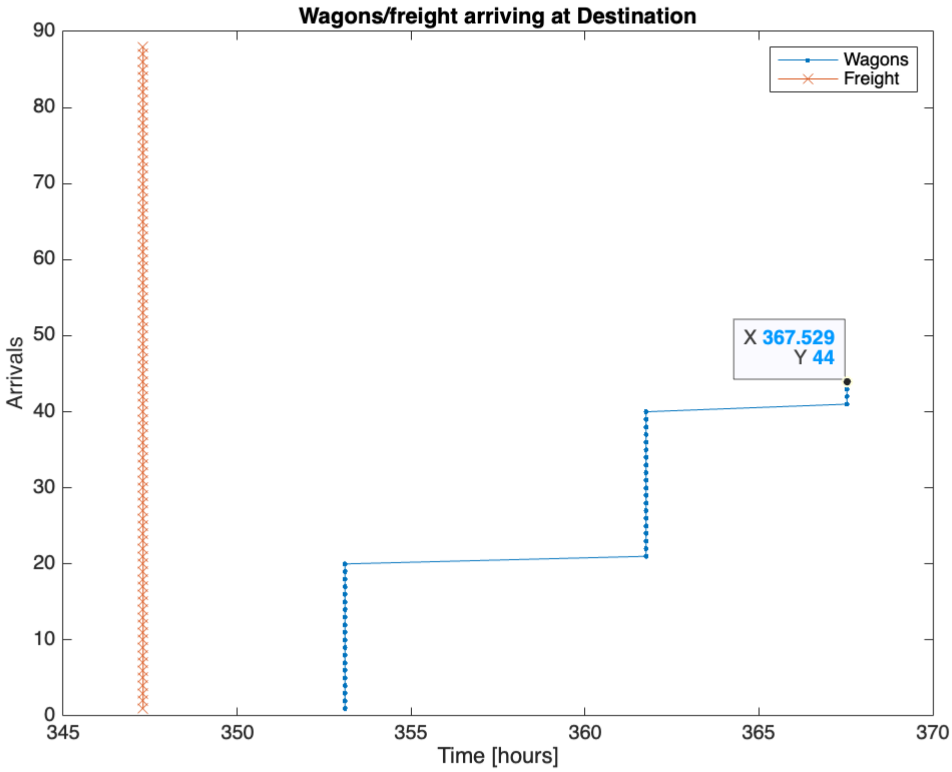
Origin 1 and 2 are production points.




```
Origin1Prod = 269
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
     37      23      44      46
```

The maximum wagon turnaround time is:

```
ans = 695.4589
WagonsDest = 44
TrainsDest = 1
TrainsDestDay = 0.0480
Freightarrivals = 88
Efficiency = 10.6280
CostTotal = 1.1318e+06
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

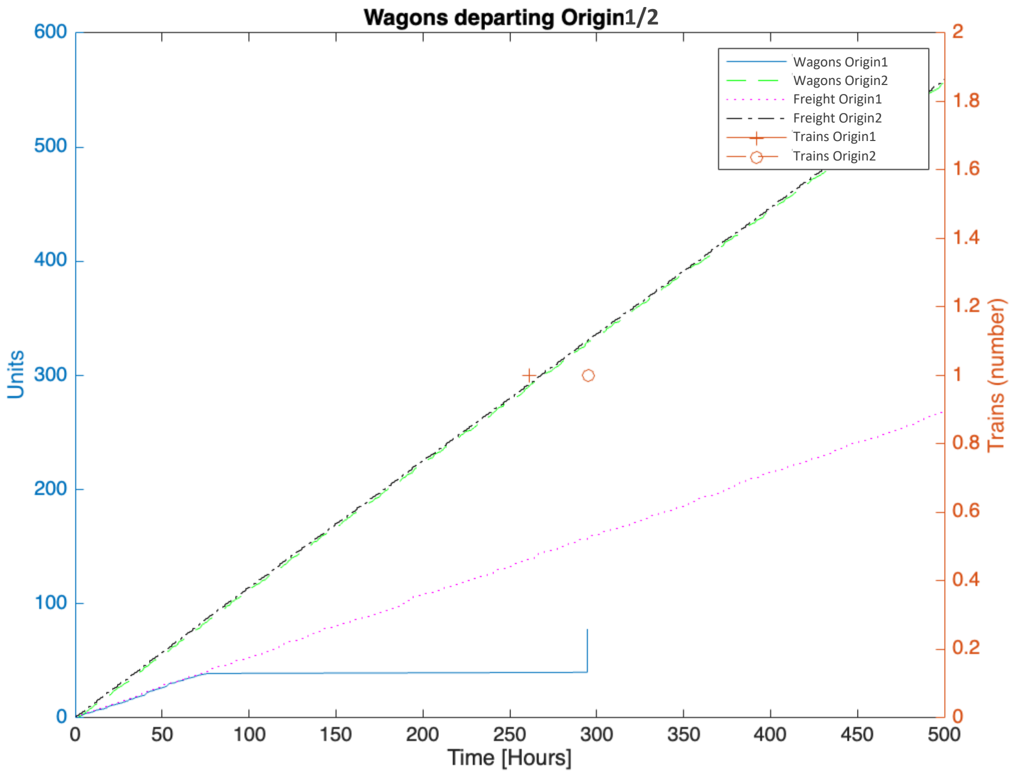
```
Fleetsize = 1.3909
```

Data analysis - Run 12

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

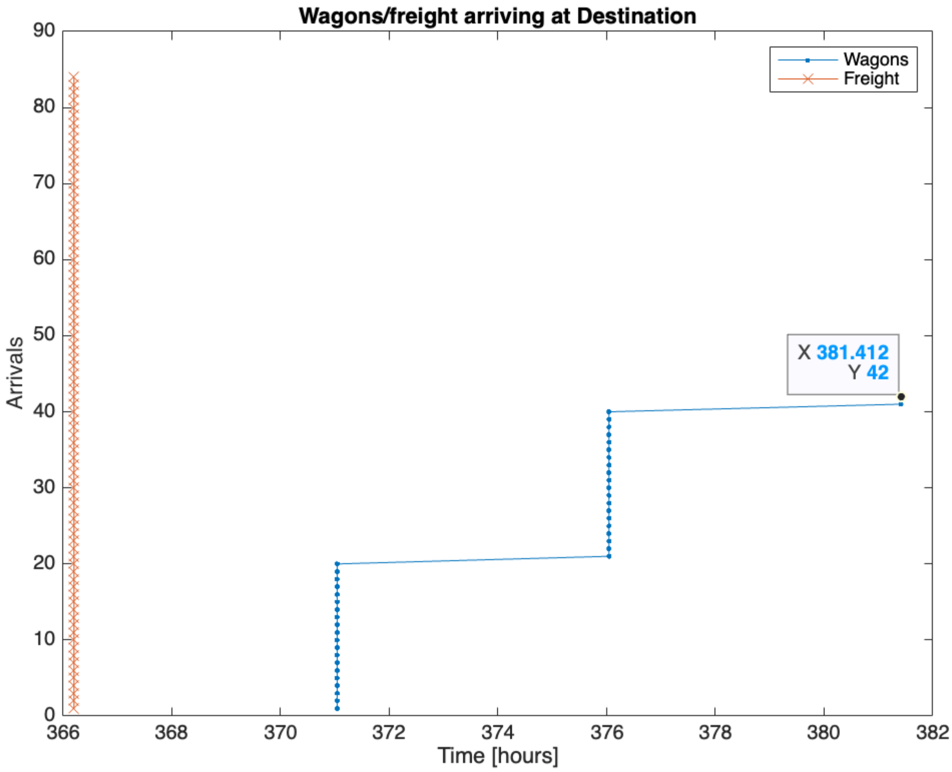
Origin 1 and 2 are production points.




```
Origin1Prod = 269
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
    36    39    42    47
```

The maximum wagon turnaround time is:

```
ans = 742.6231
WagonsDest = 42
TrainsDest = 1
TrainsDestDay = 0.0480
Freightarrivals = 84
Efficiency = 10.1449
CostTotal = 1.0757e+06
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

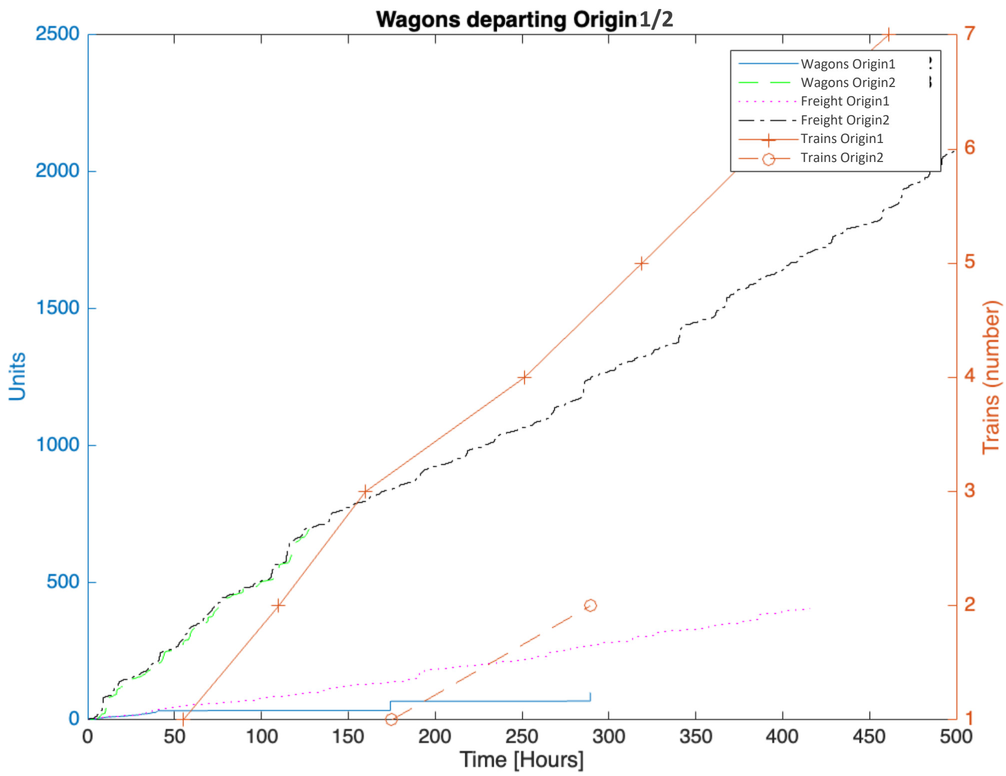
```
Fleetsize = 1.4852
```

Data analysis - Run 13

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

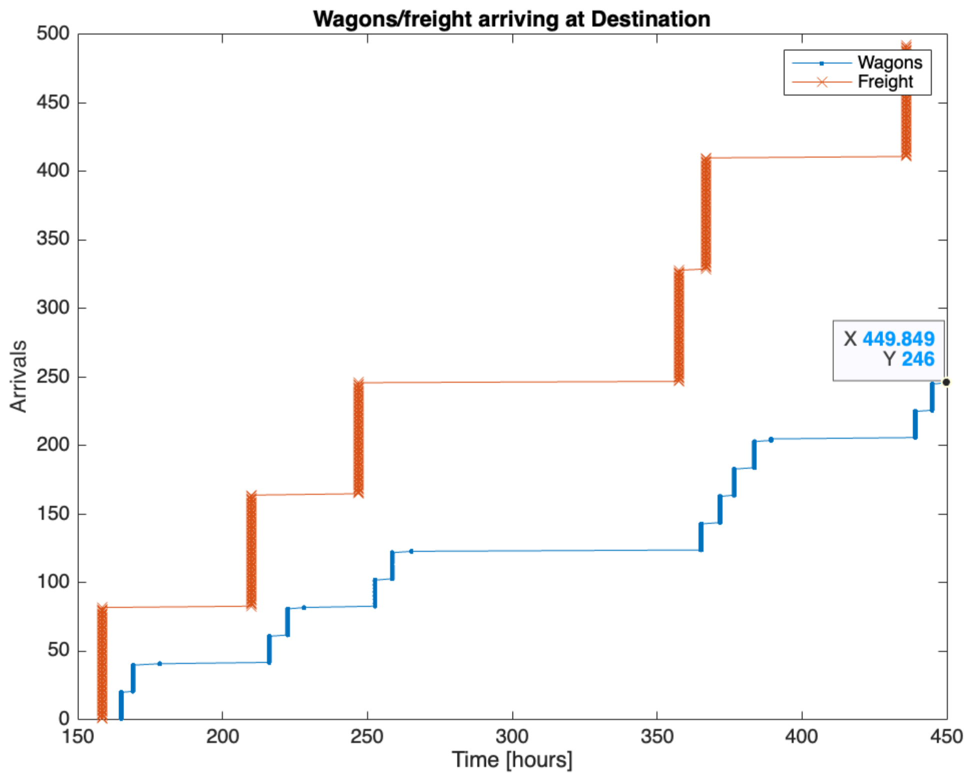
Origin 1 and 2 are production points.




```
Origin1Prod = 408
Origin2Prod = 2075
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
    33     34     41     42
```

The maximum wagon turnaround time is:

```
ans = 832.1129
WagonsDest = 246
TrainsDest = 6
TrainsDestDay = 0.2880
Freightarrivals = 492
Efficiency = 19.8147
CostTotal = 2.5970e+06
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

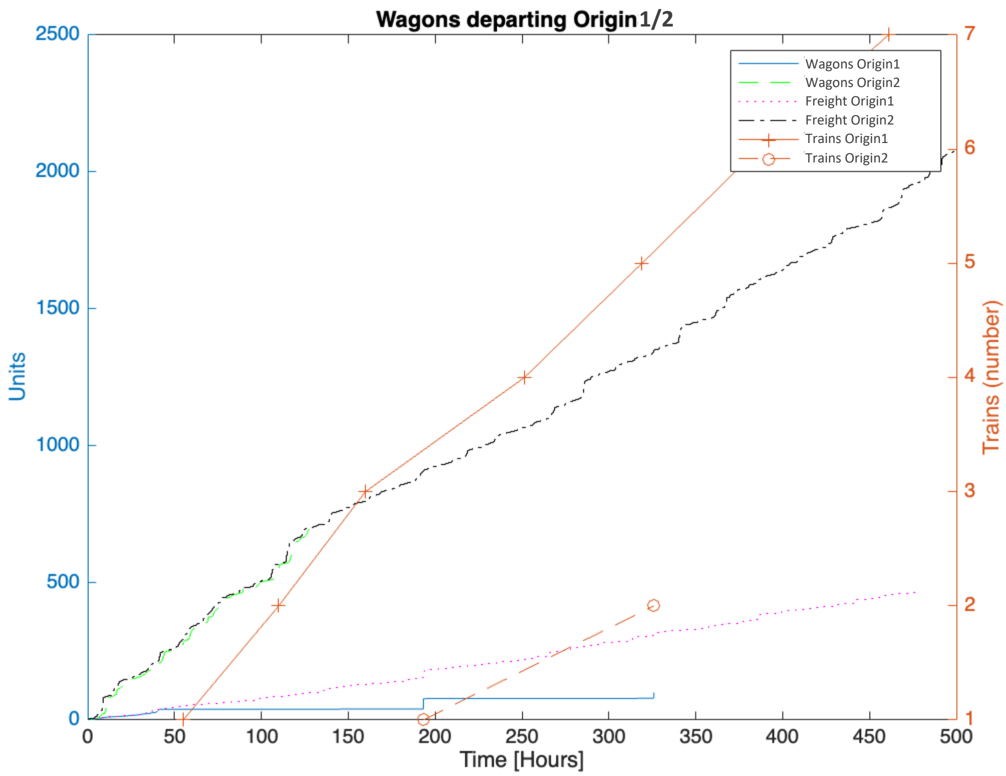
```
Fleetsize = 9.9854
```

Data analysis - Run 14

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

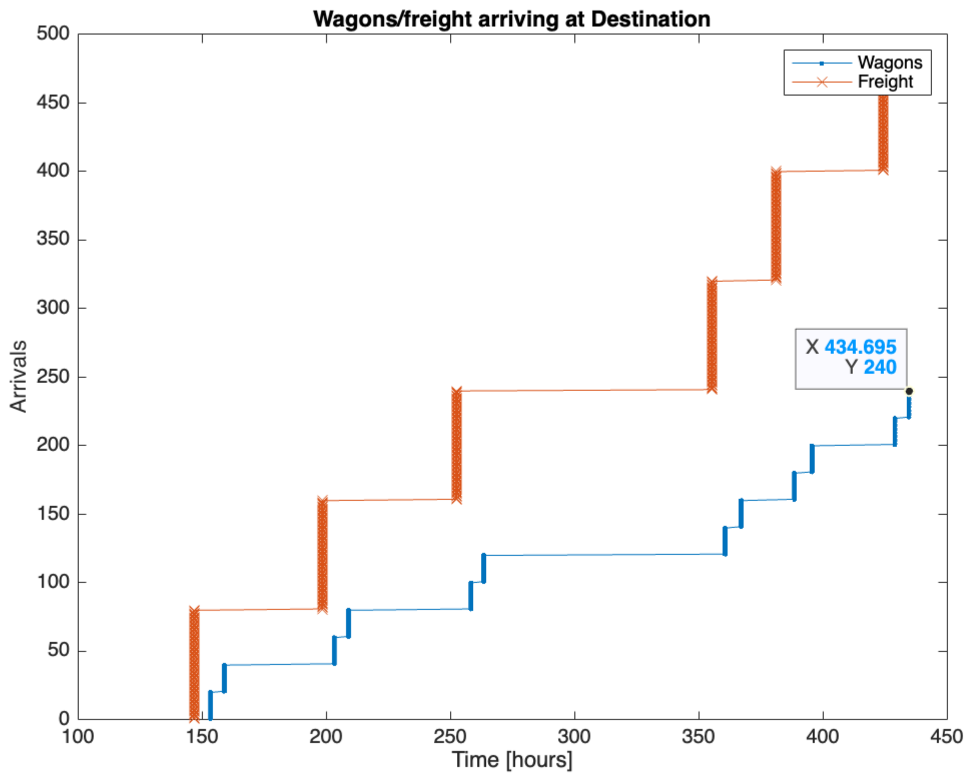
Origin 1 and 2 are production points.




```
Origin1Prod = 468
Origin2Prod = 2075
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
    33    39    40    45
```

The maximum wagon turnaround time is:

```
ans = 735.8800
WagonsDest = 240
TrainsDest = 6
TrainsDestDay = 0.2880
Freightarrivals = 480
Efficiency = 18.8753
CostTotal = 2.5448e+06
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

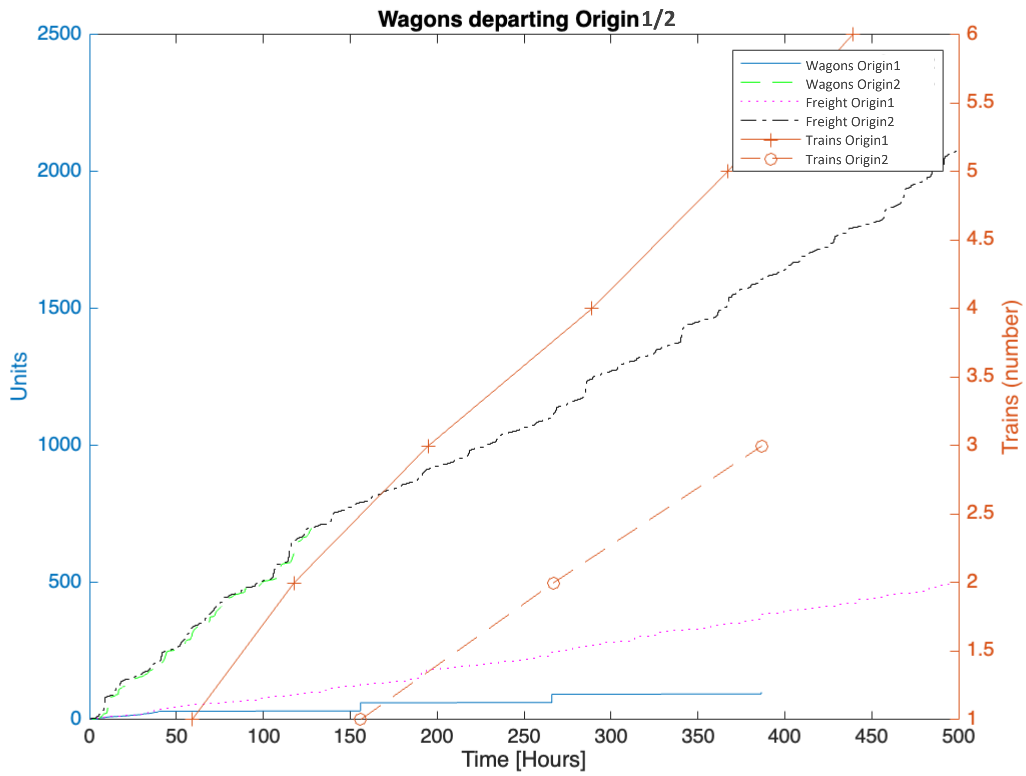
```
Fleetsize = 8.8306
```

Data analysis - Run 15

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

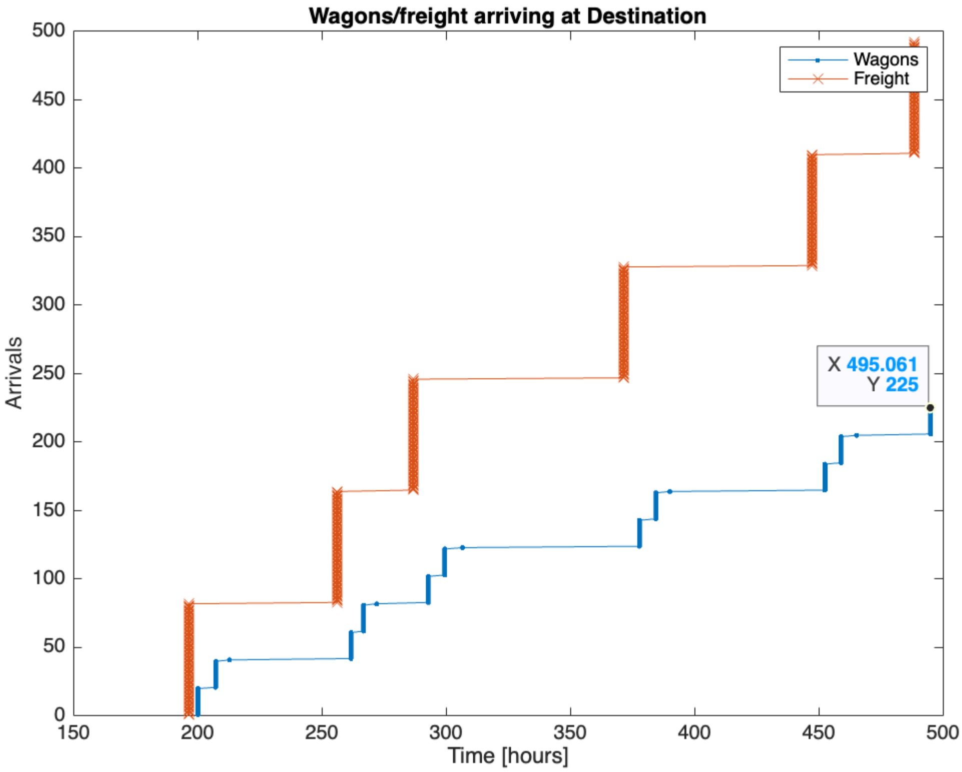
Origin 1 and 2 are production points.




```
Origin1Prod = 496
Origin2Prod = 2075
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
    37    31    41    41
```

The maximum wagon turnaround time is:

```
ans = 929.1230
WagonsDest = 245
TrainsDest = 5.9756
TrainsDestDay = 0.2868
Freightarrivals = 492
Efficiency = 19.1365
CostTotal = 6.1606e+06
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

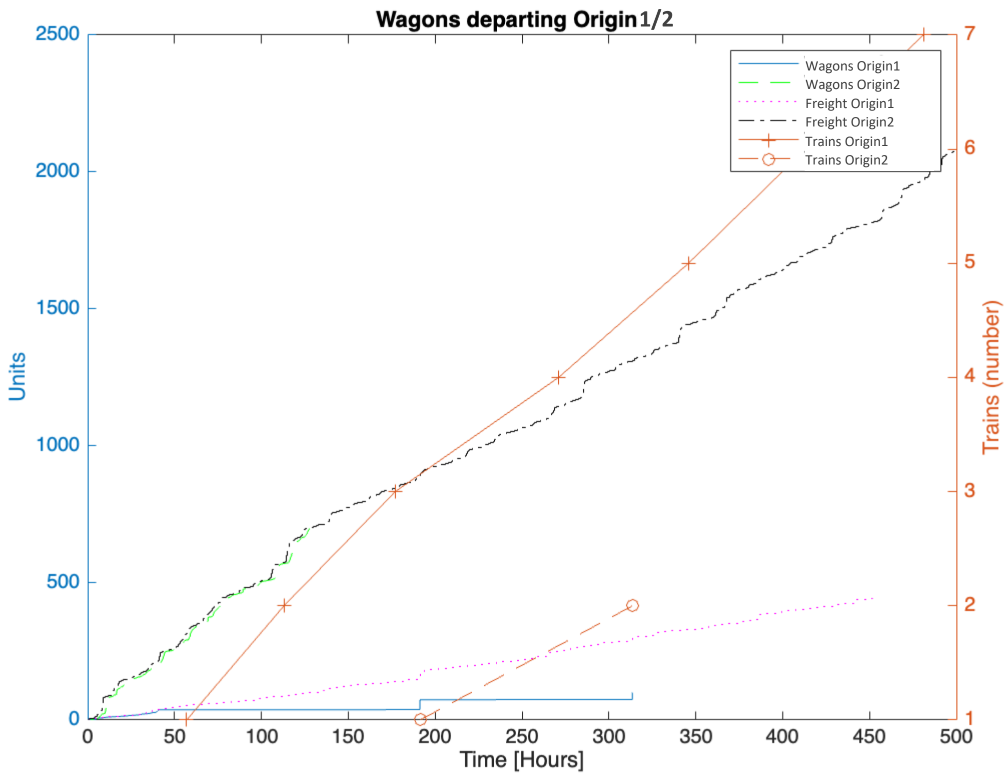
```
Fleetsize = 11.1042
```

Data analysis - Run 16

Data from the simulations runs is presented, an optimisation routine was running using a Genetic Algorithm to find the optimum train length.

Origin

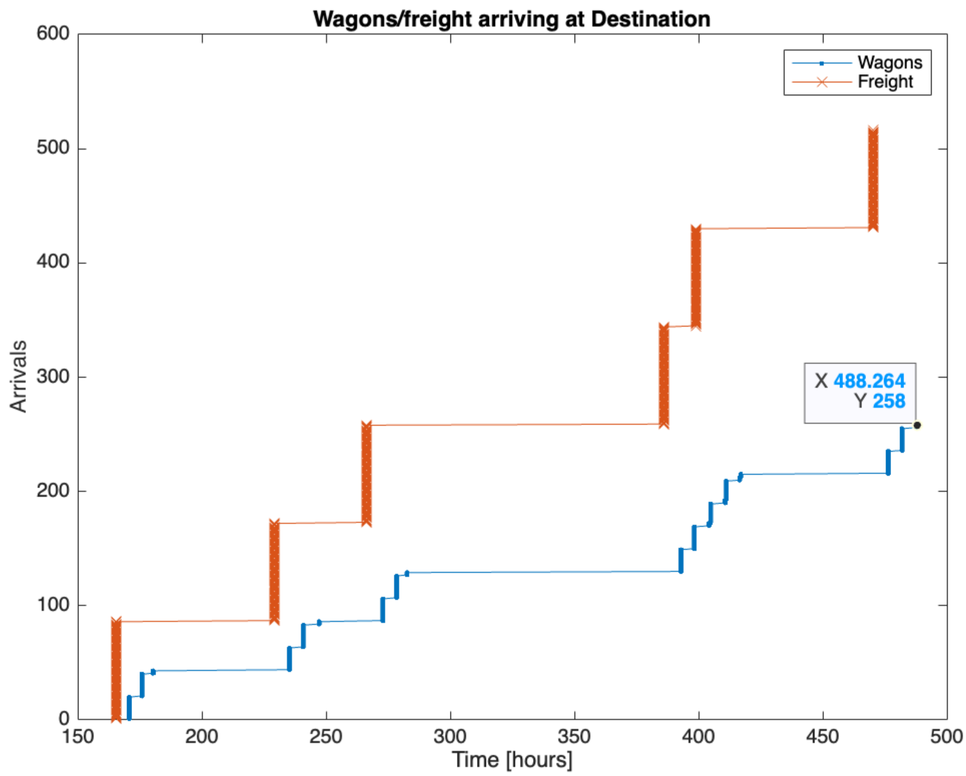
Origin 1 and 2 are production points.




```
Origin1Prod = 444
Origin2Prod = 2075
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been optimised for the highest volume throughput of the system, this affects turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The train lengths after optimisation are:

```
TrainLength = 1×4
    35      37      43      40
```

The maximum wagon turnaround time is:

```
ans = 905.9288
WagonsDest = 258
TrainsDest = 6
TrainsDestDay = 0.2880
Freightarrivals = 516
Efficiency = 20.4843
CostTotal = 6.4526e+06
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

```
Fleetsize = 10.8711
```

Appendix C – Simulation standard runs

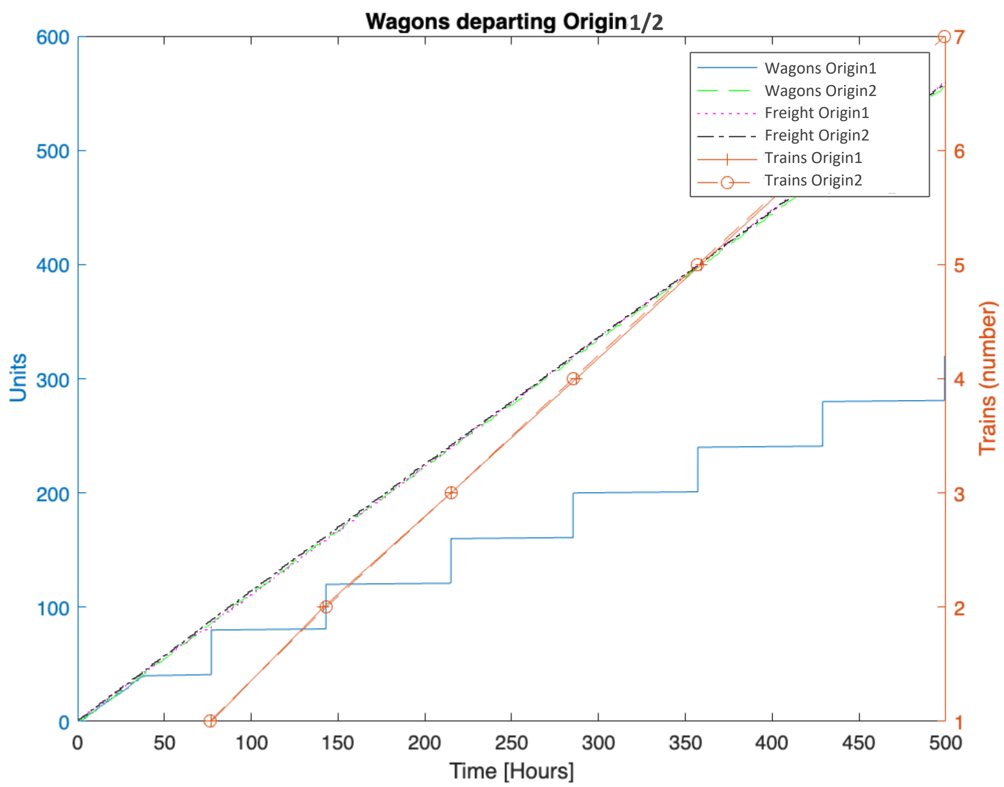
A large volume of data is generated during every simulation. This appendix contains an extract of the data which shows the performance relative to this study. These simulation runs are presented in Section 9.2.

Data analysis - Standard Run 1

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

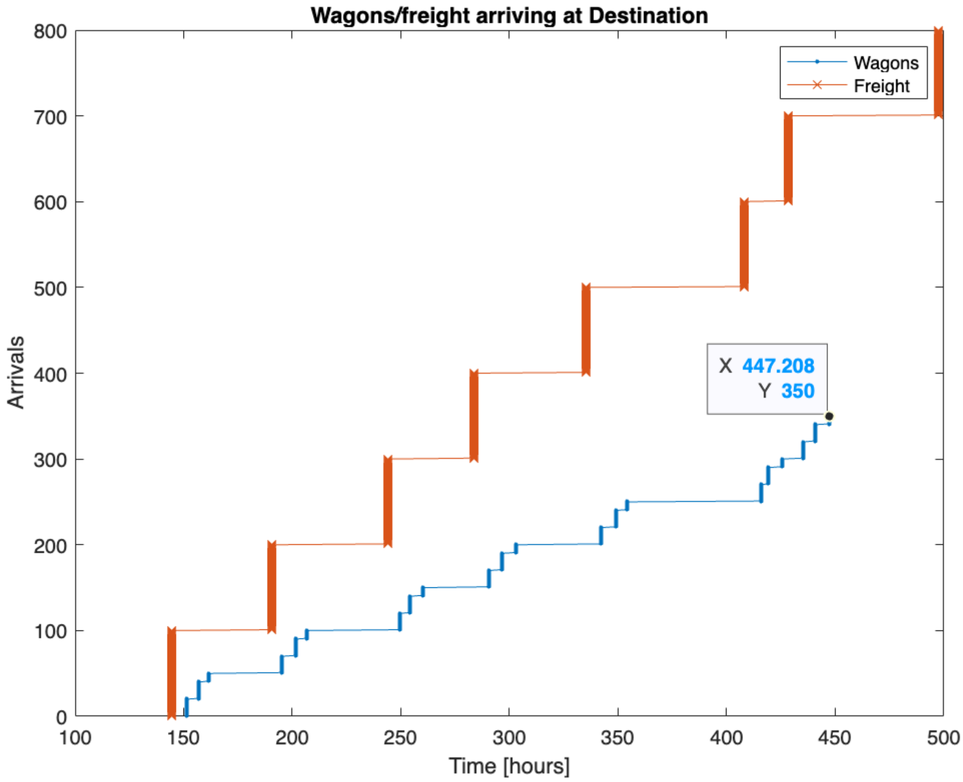
Origin 1 and 2 are production points.




```
Origin1Prod = 560
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 603.6778
WagonsDest = 350
TrainsDest = 7
TrainsDestDay = 0.3360
Freightarrivals = 800
Efficiency = 71.4924
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

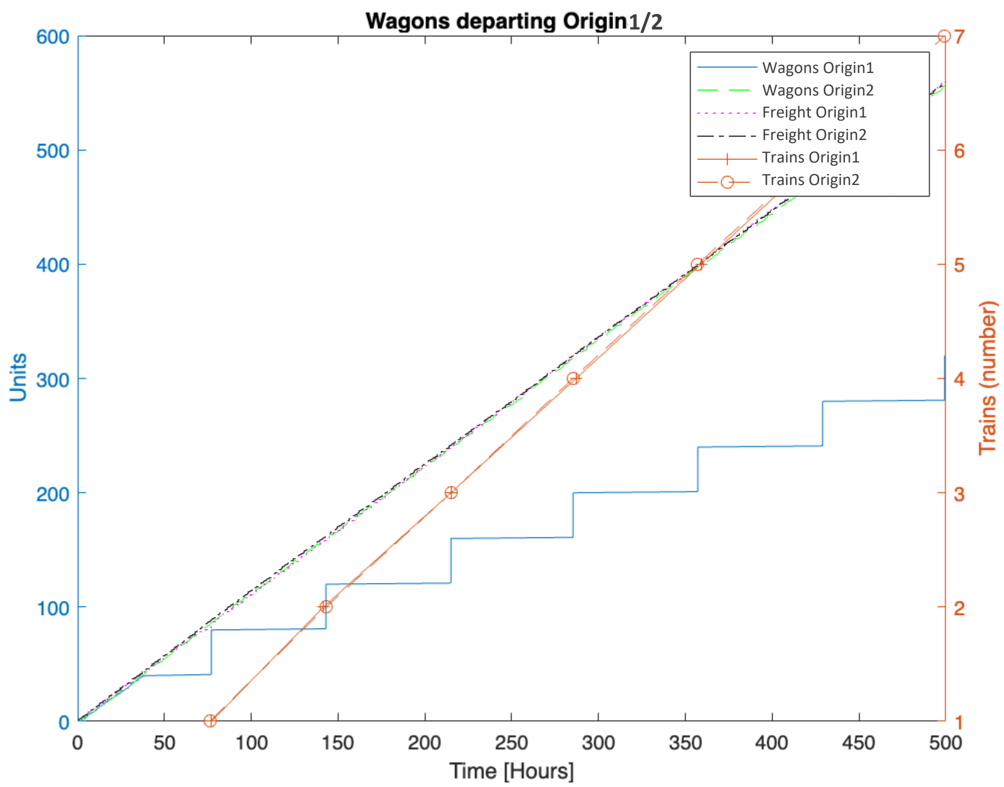
```
Fleetsize = 8.4515
```

Data analysis - Standard Run 2

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

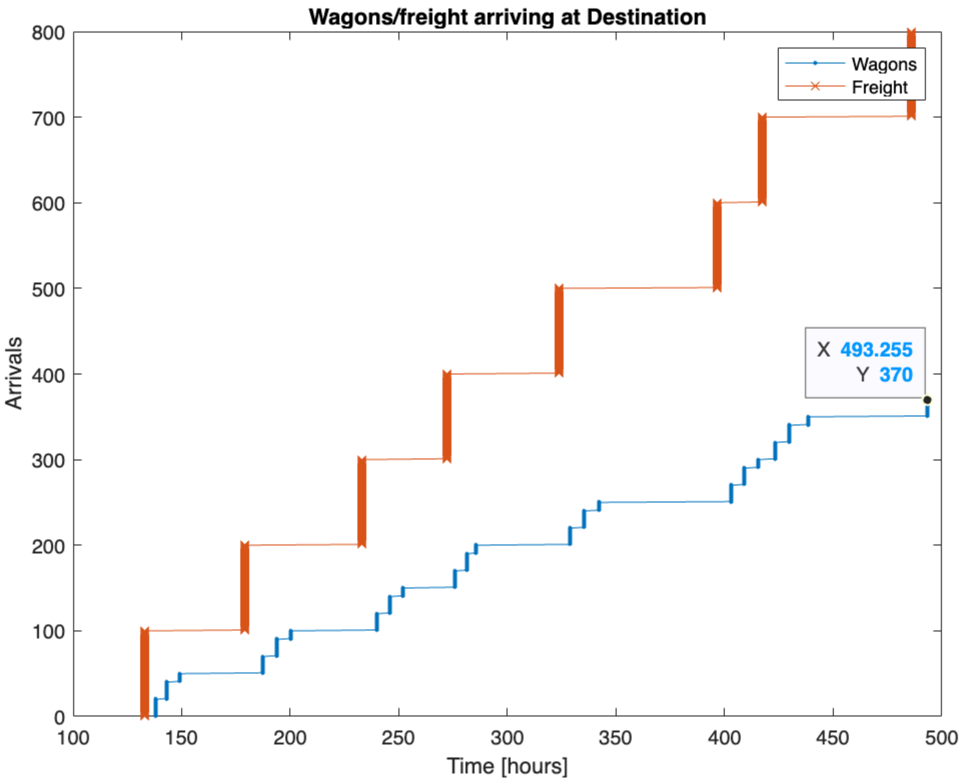
Origin 1 and 2 are production points.




```
Origin1Prod = 560
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 679.7104
WagonsDest = 370
TrainsDest = 7.4000
TrainsDestDay = 0.3552
Freightarrivals = 800
Efficiency = 71.4924
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

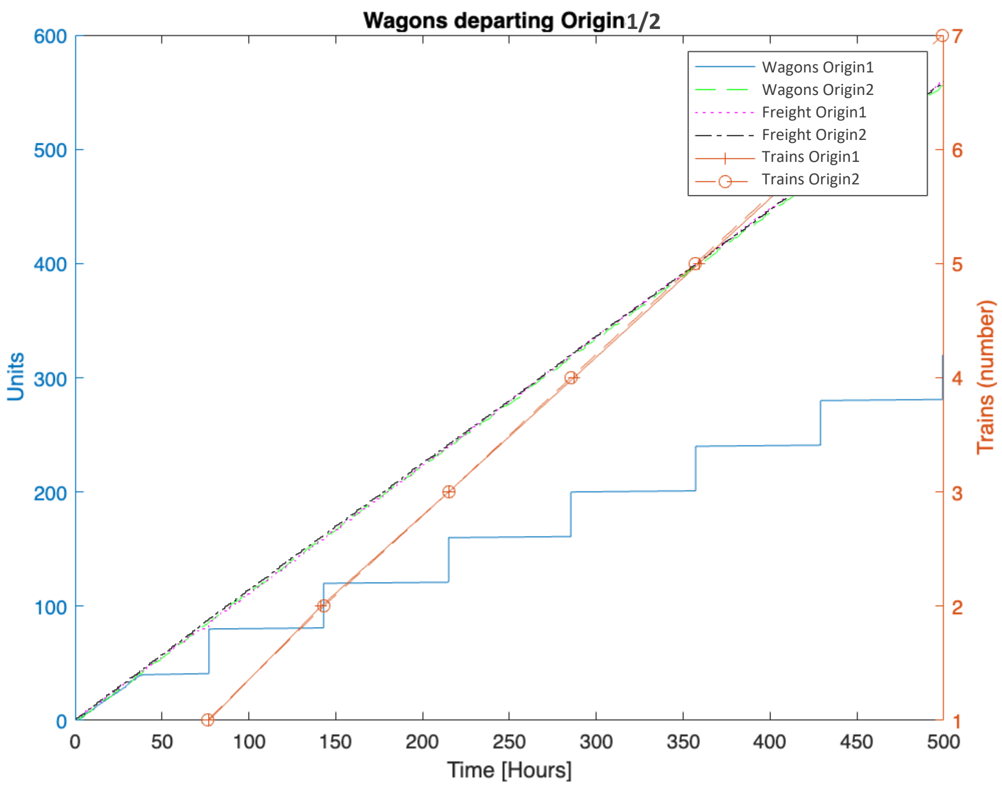
```
Fleetsize = 10.0597
```

Data analysis - Standard Run 3

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

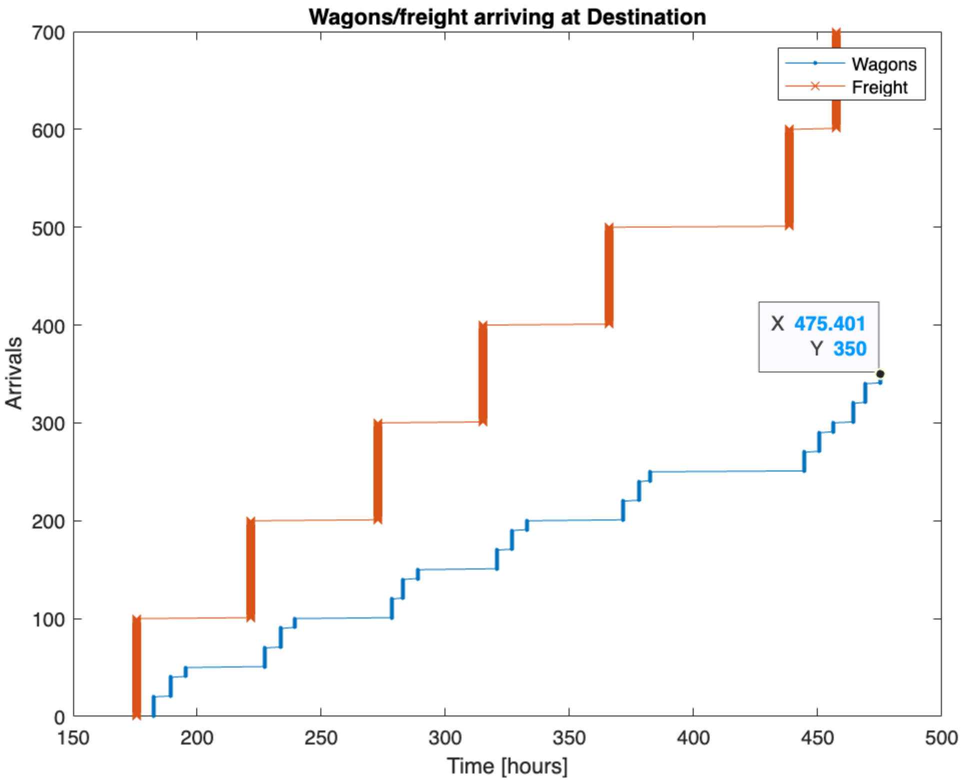
Origin 1 and 2 are production points.




```
Origin1Prod = 560
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 666.2453
WagonsDest = 350
TrainsDest = 7
TrainsDestDay = 0.3360
Freightarrivals = 700
Efficiency = 62.5559
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

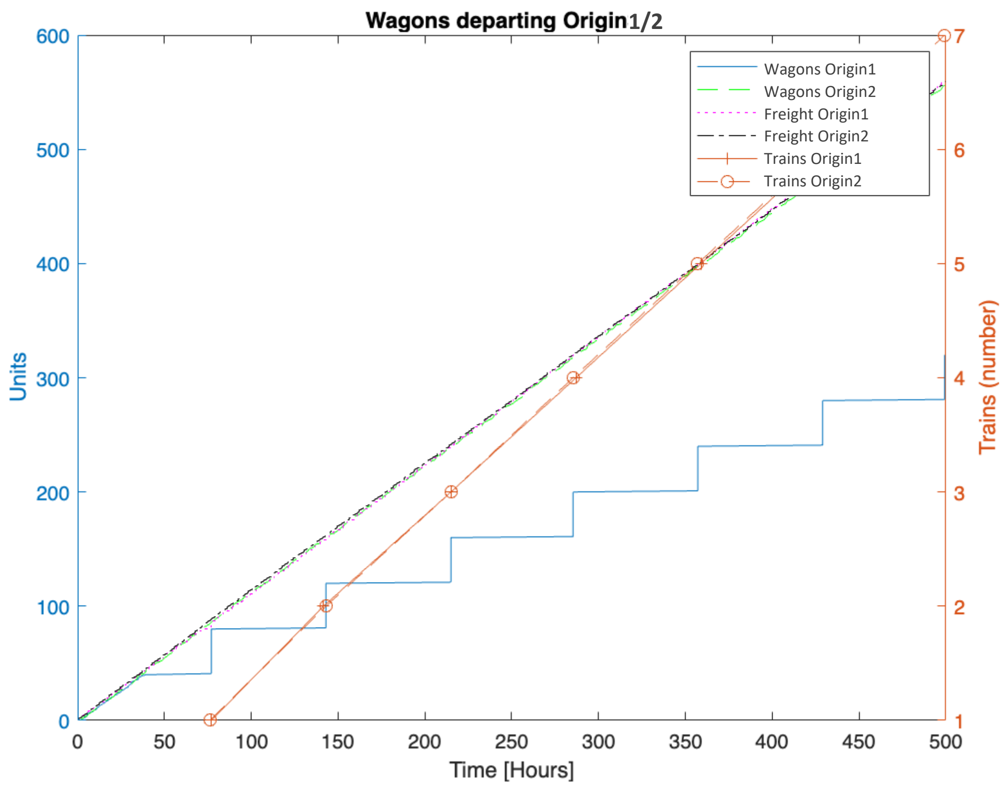
```
Fleetsize = 9.3274
```

Data analysis - Standard Run 4

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

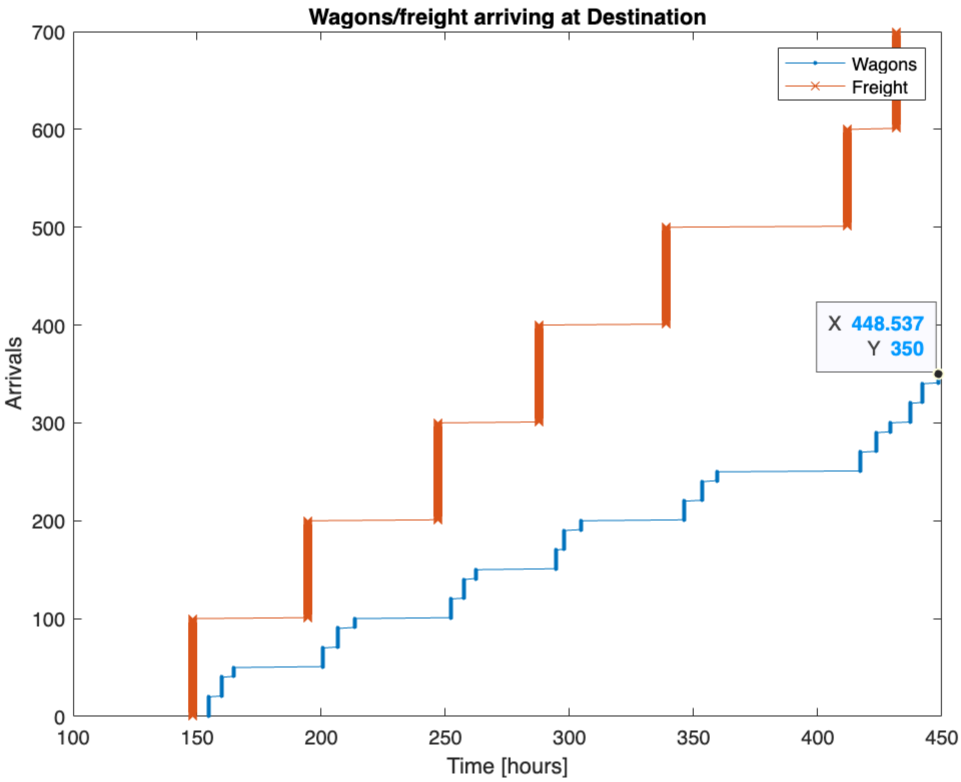
Origin 1 and 2 are production points.




```
Origin1Prod = 560
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 612.0624
WagonsDest = 350
TrainsDest = 7
TrainsDestDay = 0.3360
Freightarrivals = 700
Efficiency = 62.5559
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

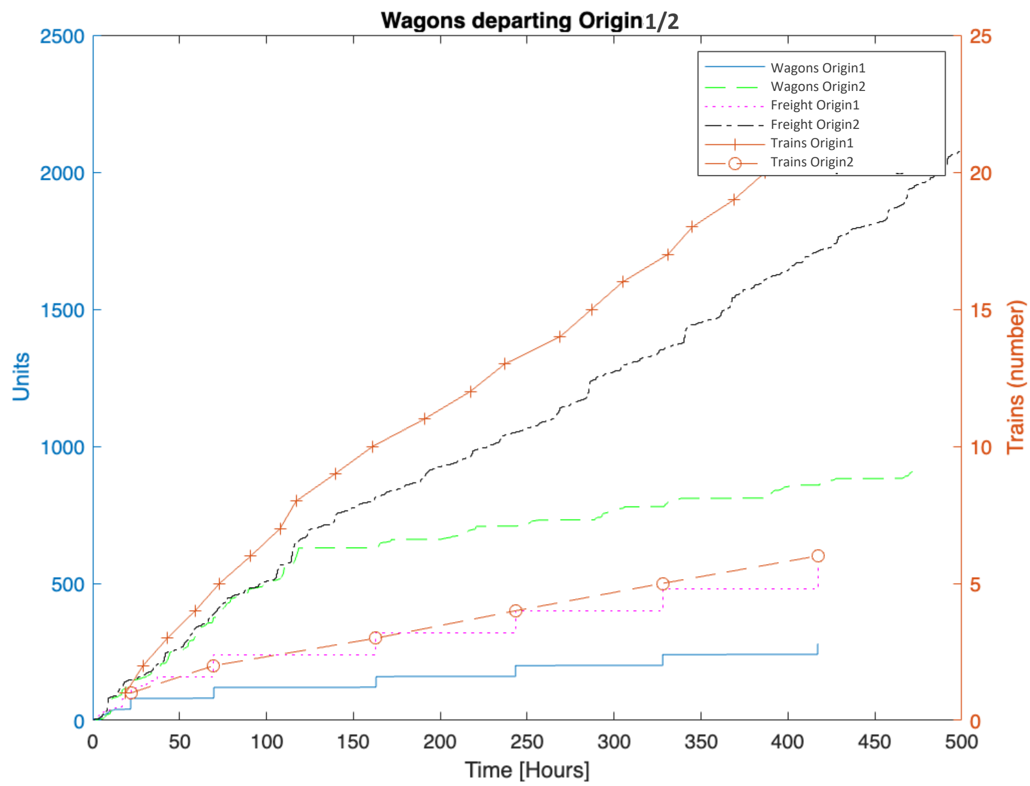
```
Fleetsize = 8.5689
```

Data analysis - Standard Run 5

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

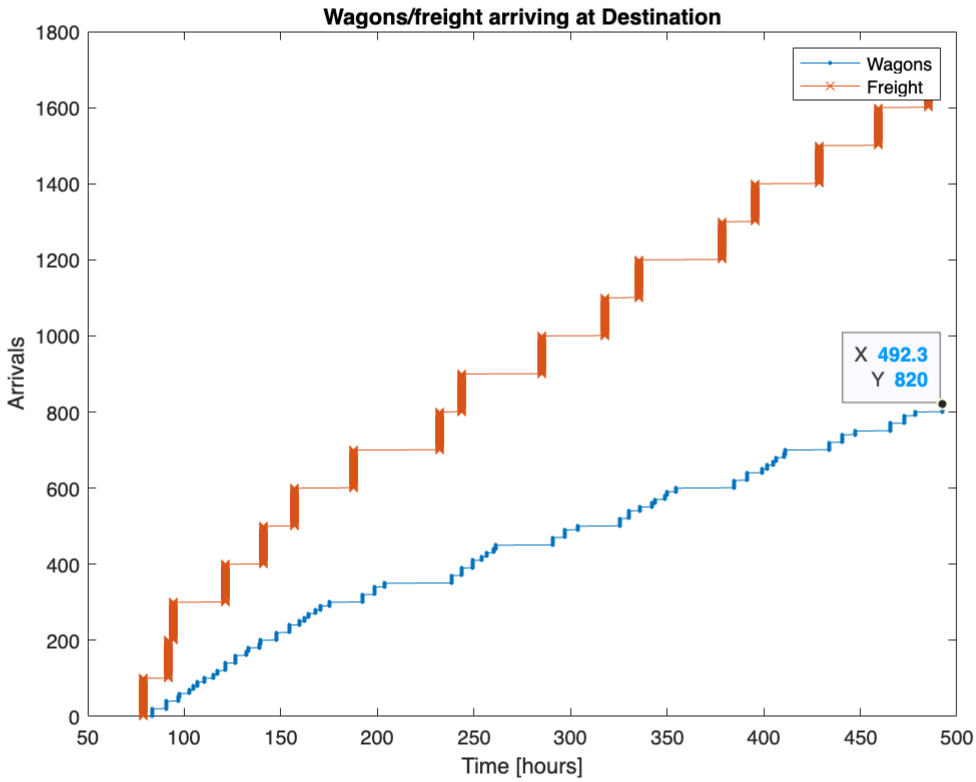
Origin 1 and 2 are production points.




```
Origin1Prod = 560
Origin2Prod = 2075
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 737.7510
WagonsDest = 820
TrainsDest = 16.4000
TrainsDestDay = 0.7872
Freightarrivals = 1700
Efficiency = 64.5161
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

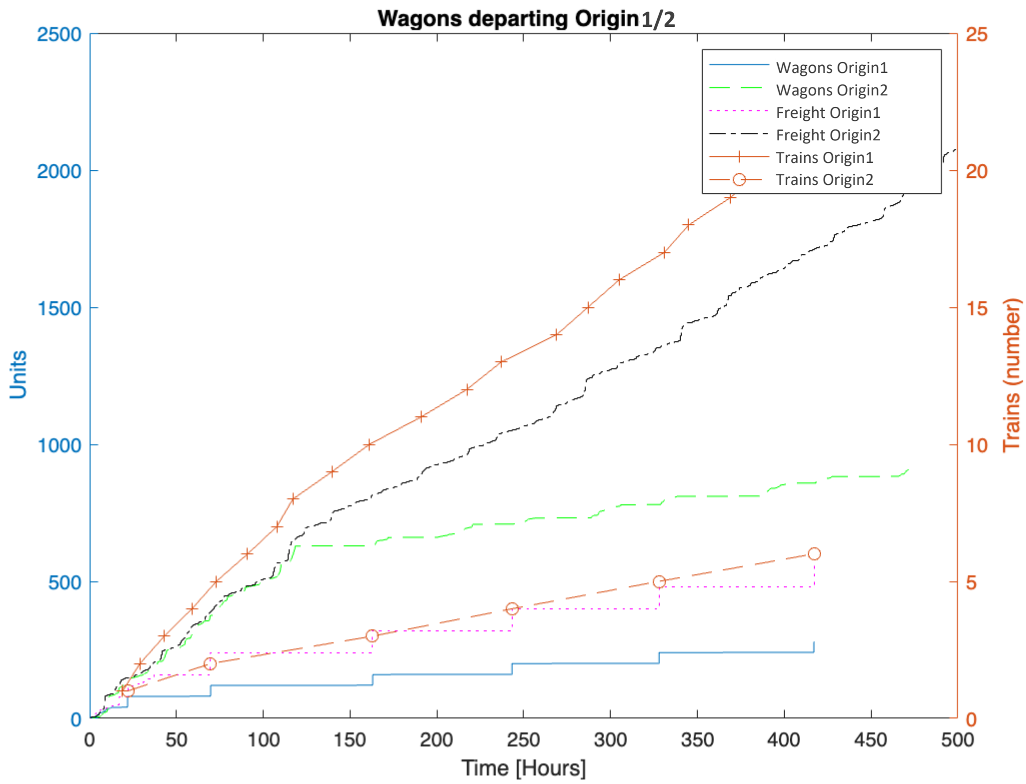
```
Fleetsize = 24.1982
```

Data analysis - Standard Run 6

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

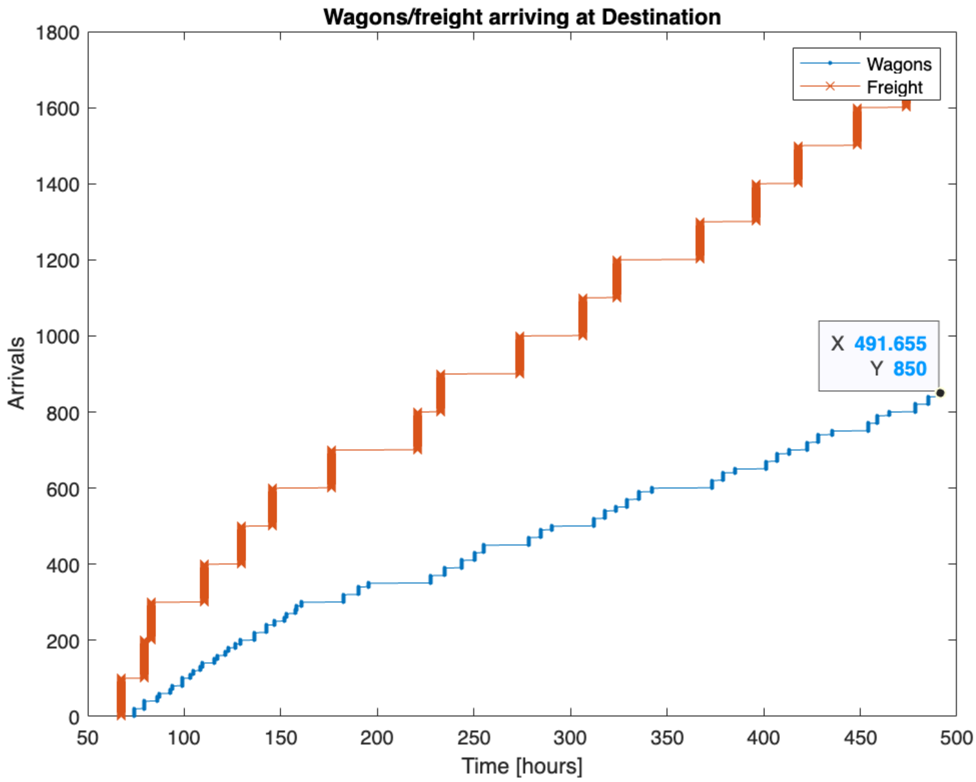
Origin 1 and 2 are production points.




```
Origin1Prod = 560
Origin2Prod = 2075
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 709.4951
WagonsDest = 850
TrainsDest = 17
TrainsDestDay = 0.8160
Freightarrivals = 1700
Efficiency = 64.5161
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

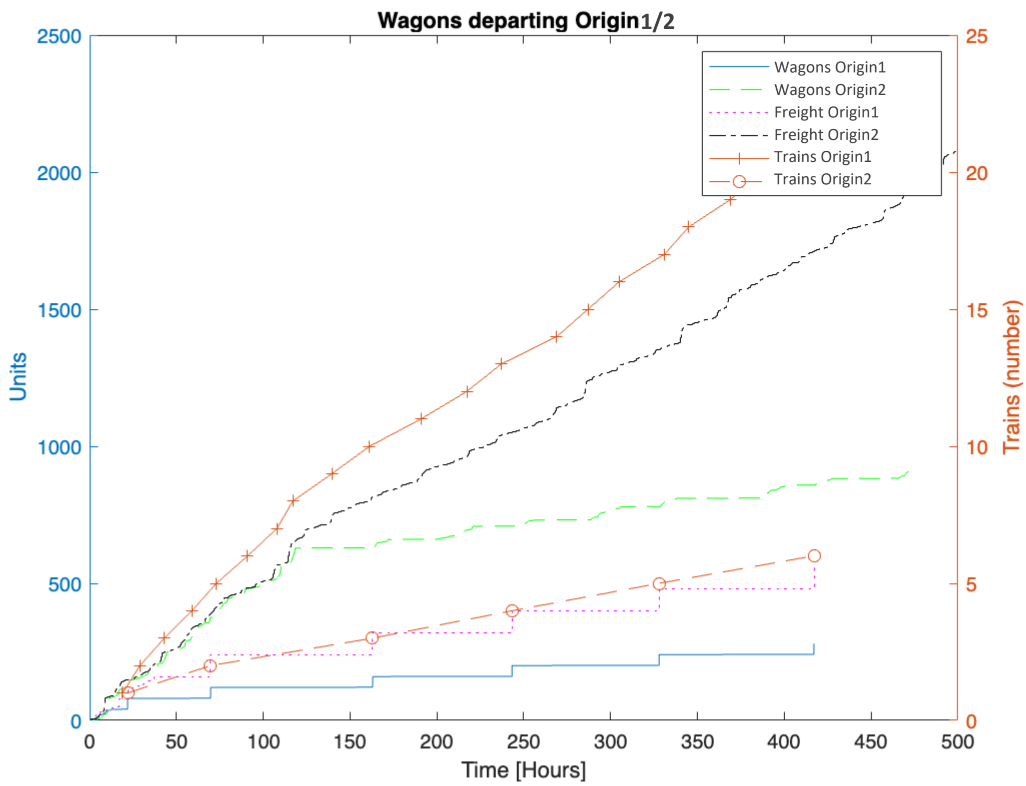
```
Fleetsize = 24.1228
```

Data analysis - Standard Run 7

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

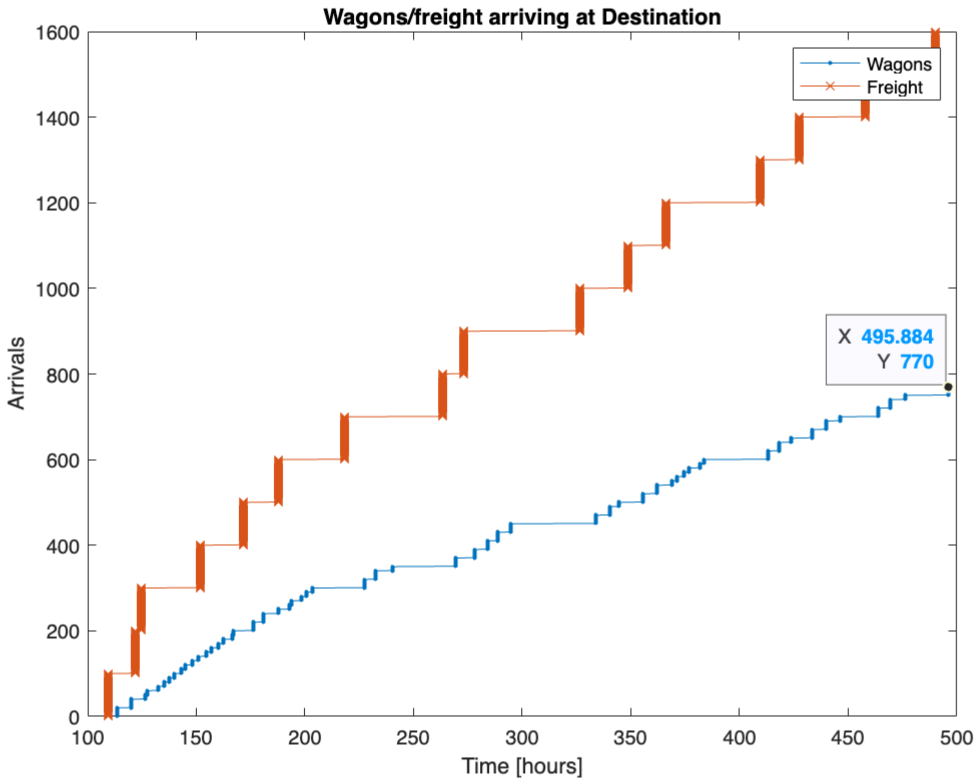
Origin 1 and 2 are production points.




```
Origin1Prod = 560
Origin2Prod = 2075
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 786.8170
WagonsDest = 770
TrainsDest = 15.4000
TrainsDestDay = 0.7392
Freightarrivals = 1600
Efficiency = 60.7211
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

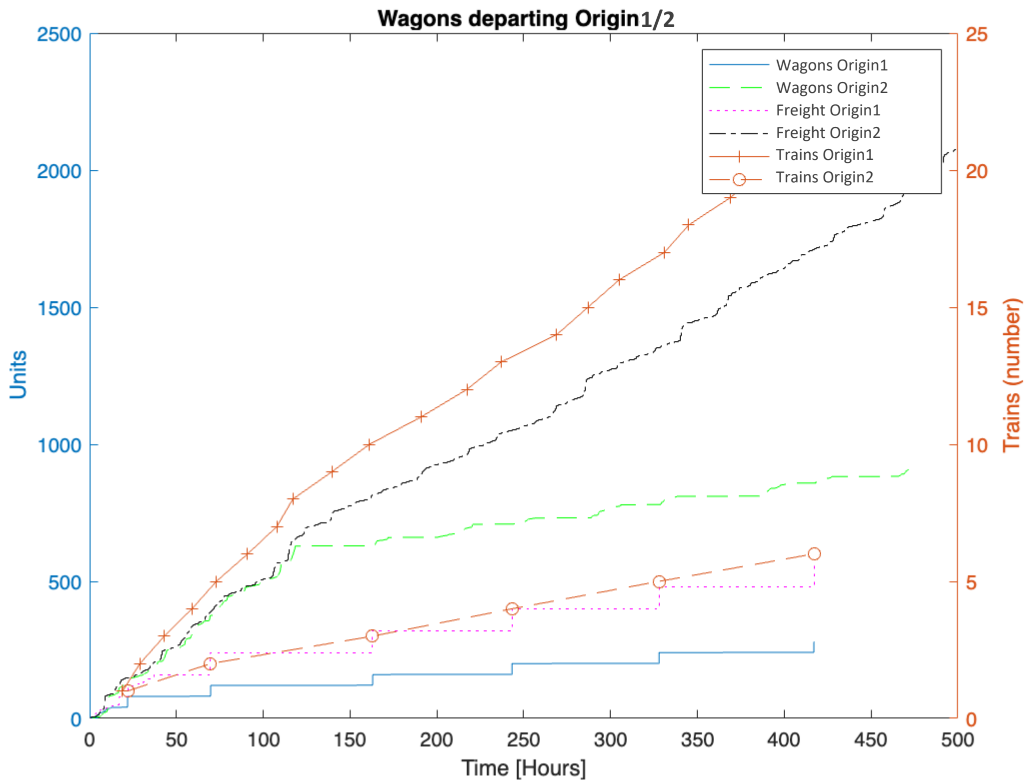
```
Fleetsize = 24.2340
```

Data analysis - Standard Run 8

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

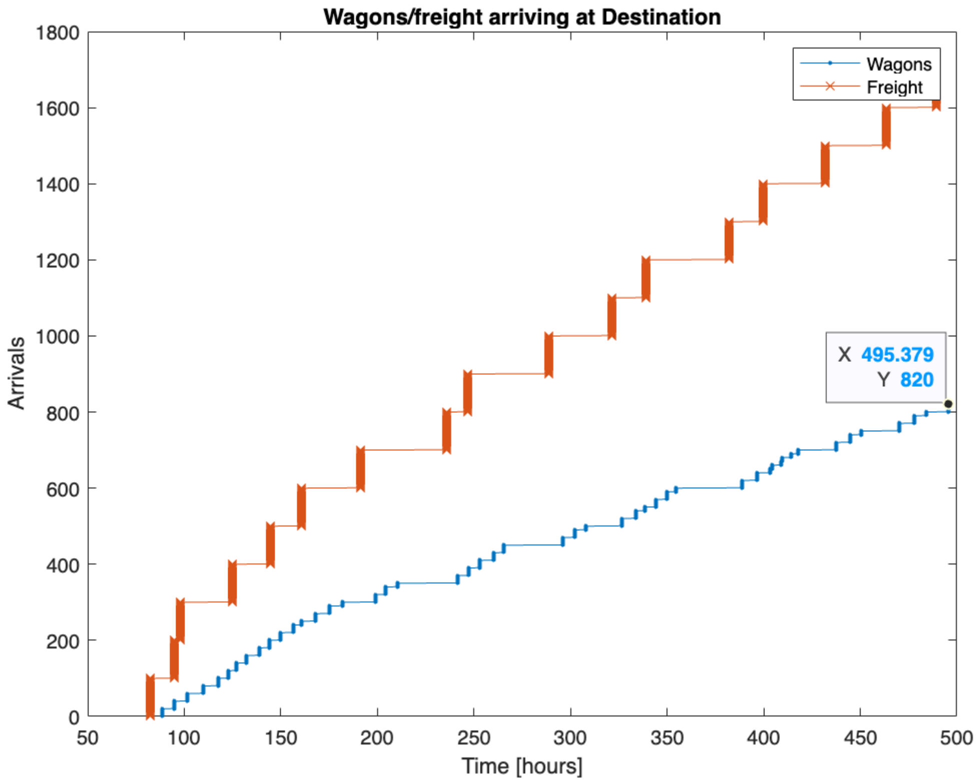
Origin 1 and 2 are production points.




```
Origin1Prod = 560
Origin2Prod = 2075
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 746.2717
WagonsDest = 820
TrainsDest = 16.4000
TrainsDestDay = 0.7872
Freightarrivals = 1700
Efficiency = 64.5161
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

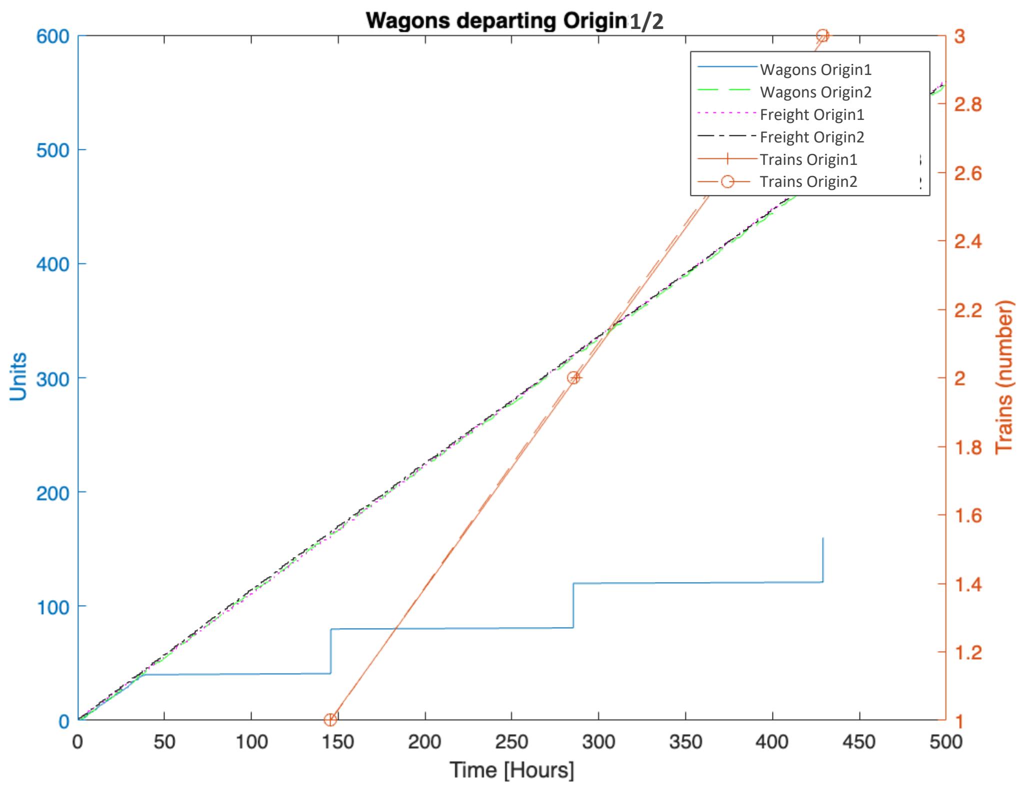
```
Fleetsize = 24.4777
```

Data analysis - Standard Run 9

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

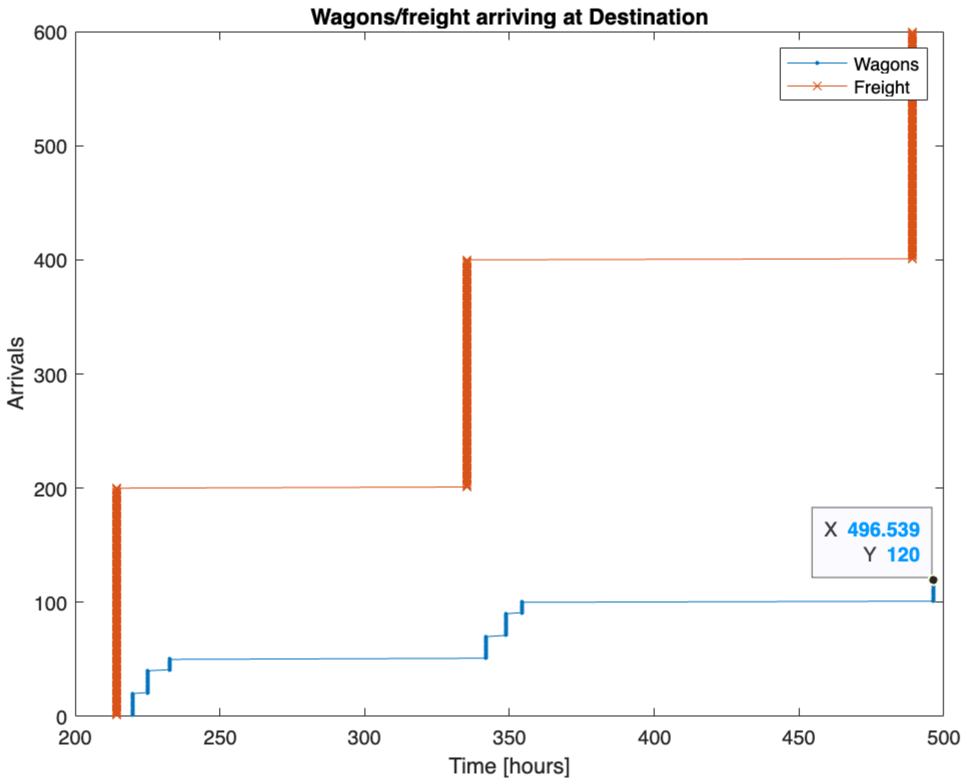
Origin 1 and 2 are production points.




```
Origin1Prod = 560
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 850.0777
WagonsDest = 120
TrainsDest = 2.4000
TrainsDestDay = 0.1152
Freightarrivals = 600
Efficiency = 53.6193
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

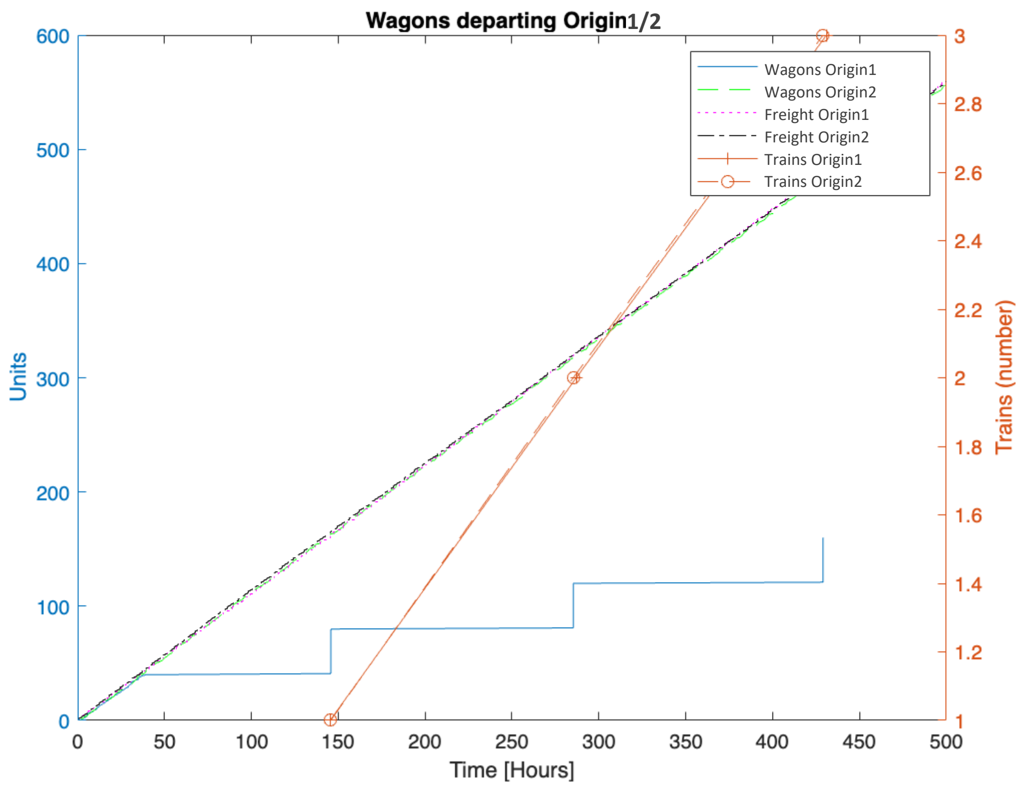
```
Fleetsize = 4.0804
```

Data analysis - Standard Run 10

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

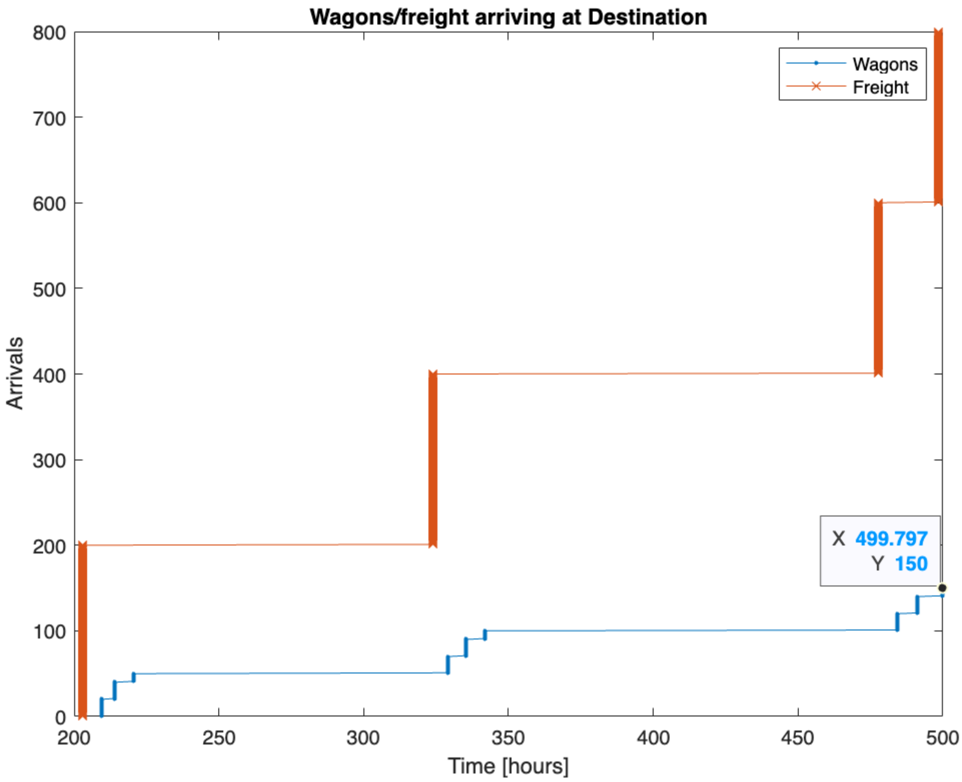
Origin 1 and 2 are production points.




```
Origin1Prod = 560
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 873.9948
WagonsDest = 150
TrainsDest = 3
TrainsDestDay = 0.1440
Freightarrivals = 800
Efficiency = 71.4924
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

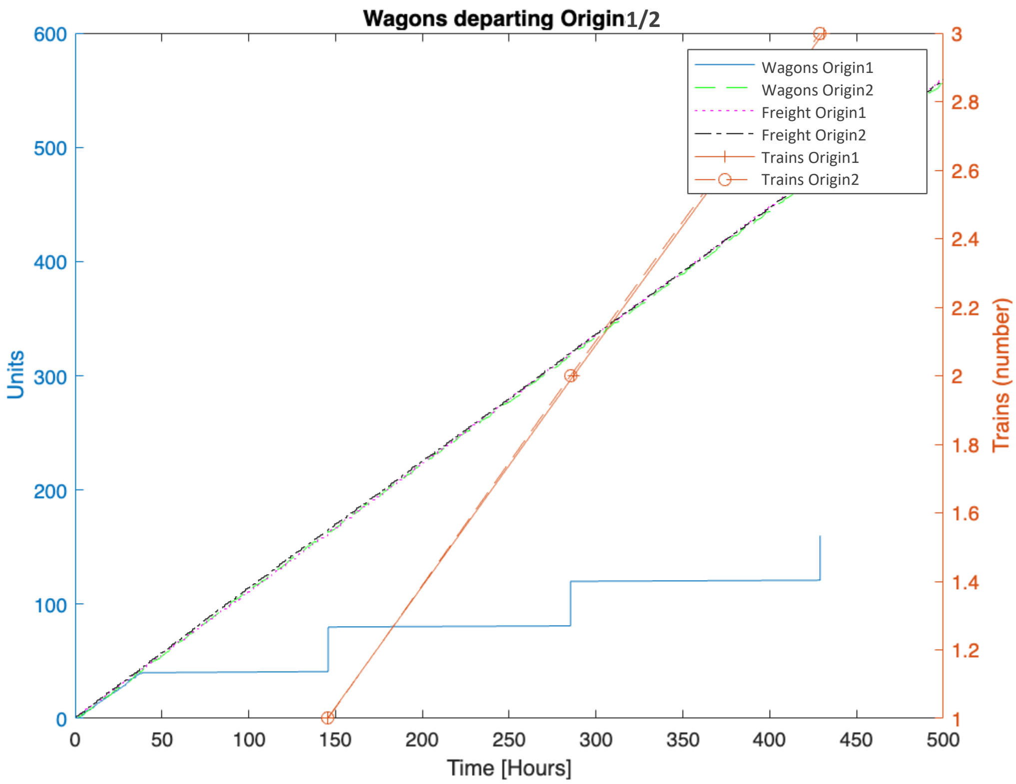
```
Fleetsize = 5.2440
```

Data analysis - Standard Run 11

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

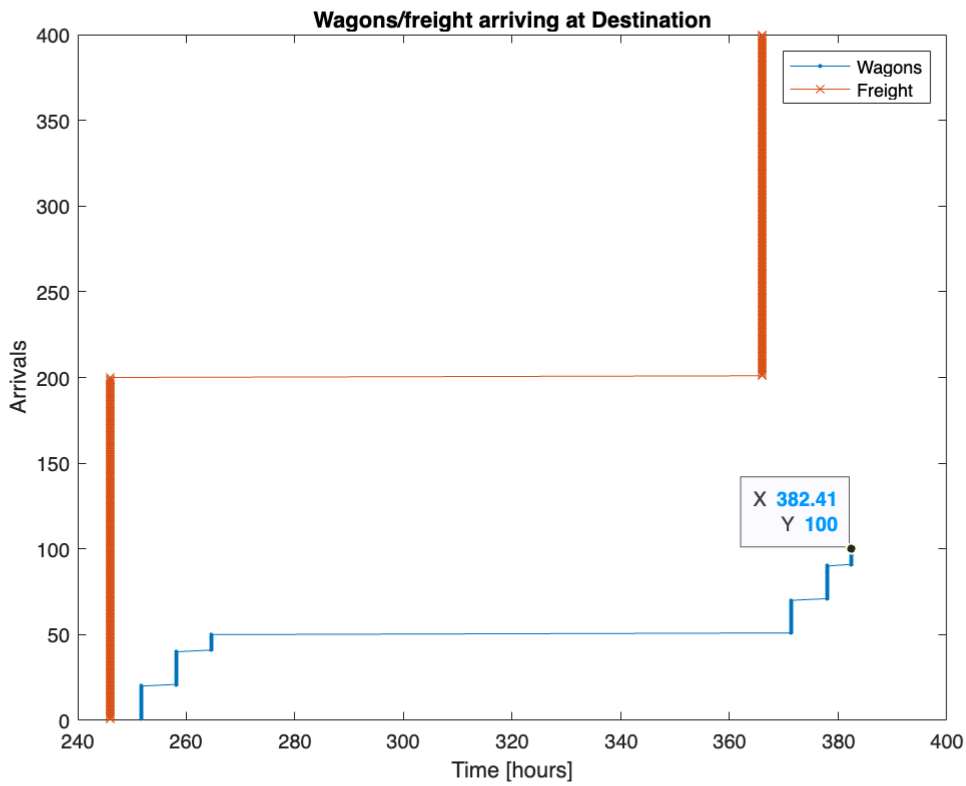
Origin 1 and 2 are production points.




```
Origin1Prod = 560
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 686.5681
WagonsDest = 100
TrainsDest = 2
TrainsDestDay = 0.0960
Freightarrivals = 400
Efficiency = 35.7462
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

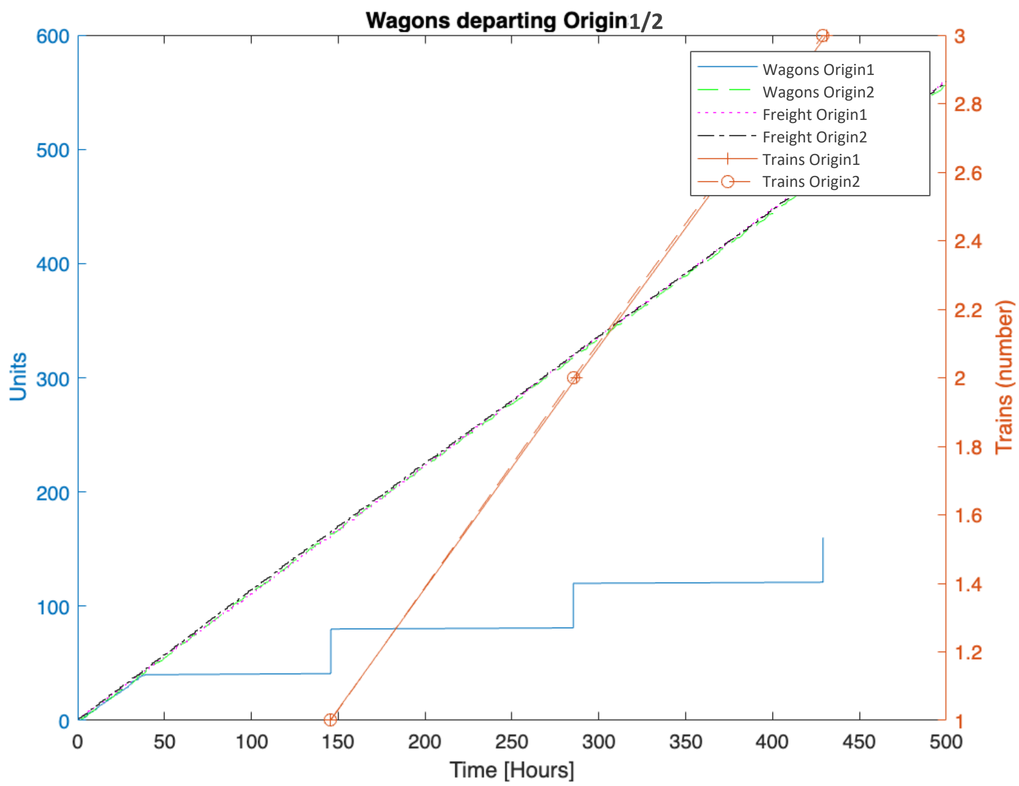
```
Fleetsize = 2.7463
```

Data analysis - Standard Run 12

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

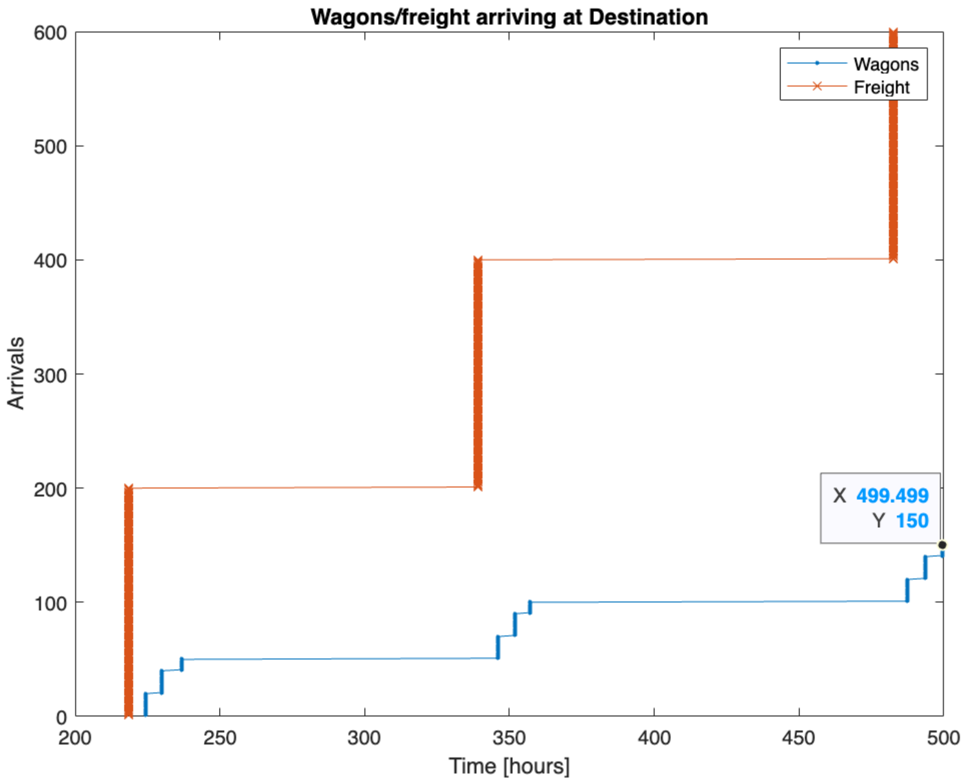
Origin 1 and 2 are production points.




```
Origin1Prod = 560
Origin2Prod = 559
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 878.1100
WagonsDest = 150
TrainsDest = 3
TrainsDestDay = 0.1440
Freightarrivals = 600
Efficiency = 53.6193
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

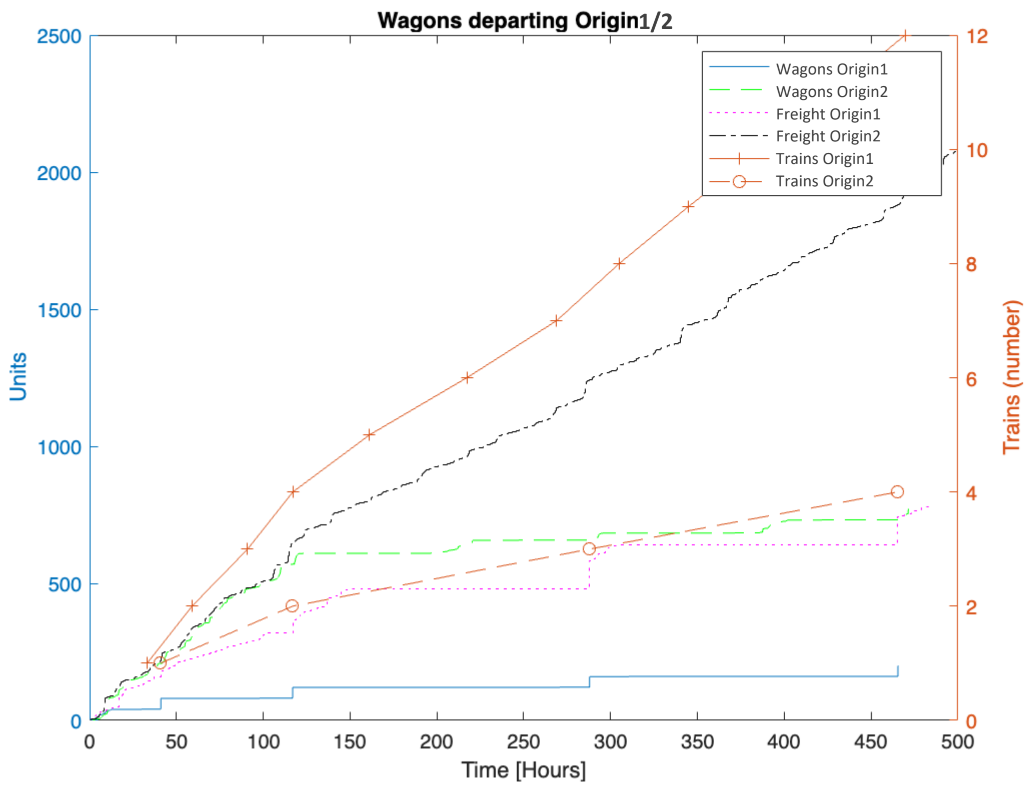
```
Fleetsize = 5.2687
```

Data analysis - Standard Run 13

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

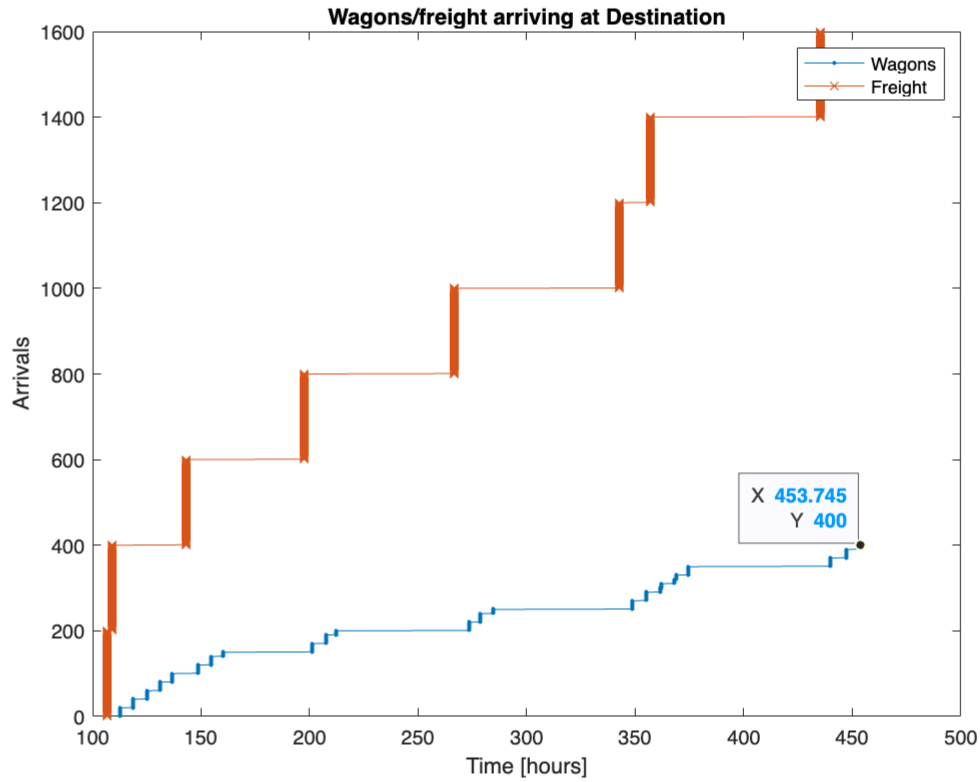
Origin 1 and 2 are production points.




```
Origin1Prod = 800
Origin2Prod = 2075
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 842.7024
WagonsDest = 400
TrainsDest = 8
TrainsDestDay = 0.3840
Freightarrivals = 1600
Efficiency = 55.6522
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

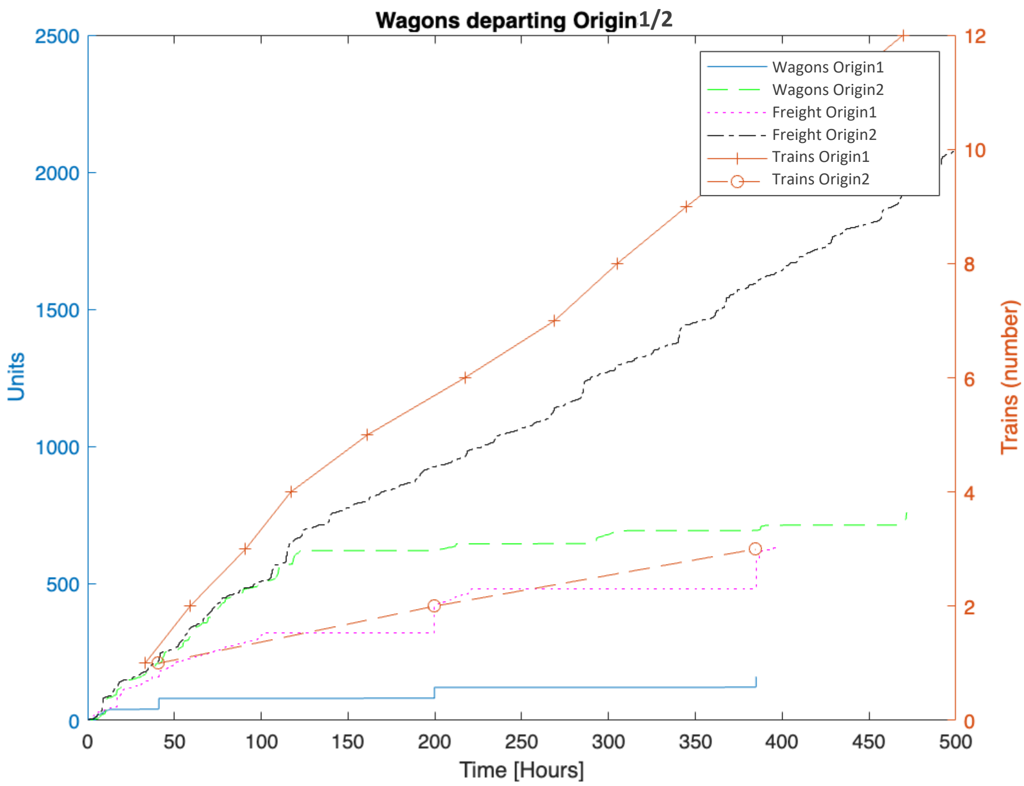
```
Fleetsize = 13.4832
```

Data analysis - Standard Run 14

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

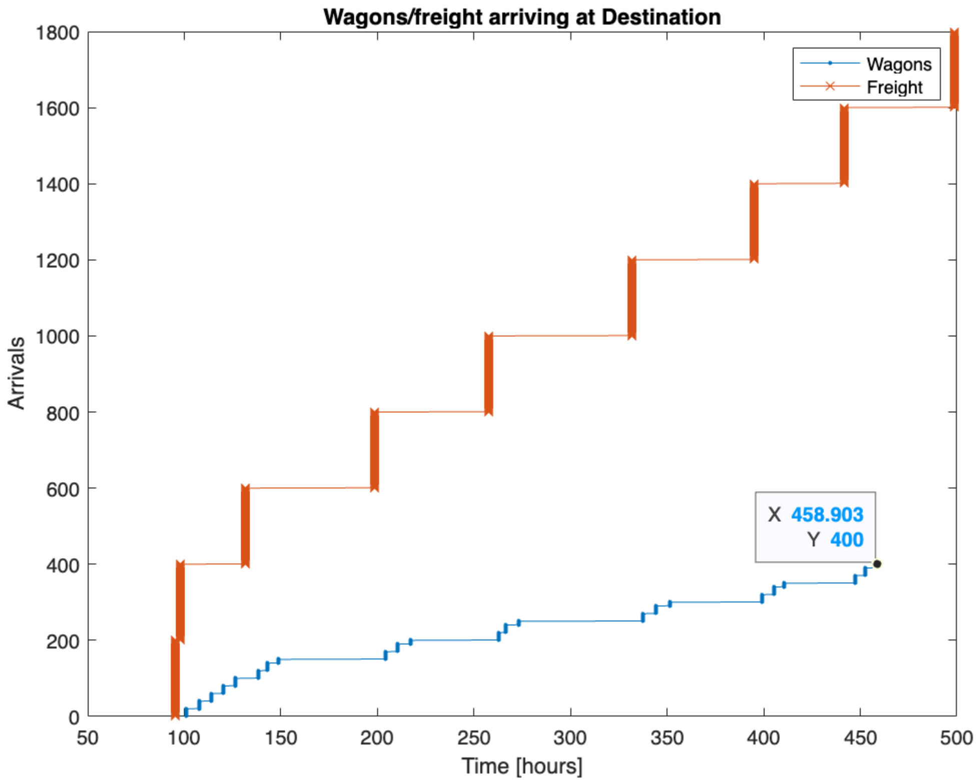
Origin 1 and 2 are production points.




```
Origin1Prod = 640
Origin2Prod = 2075
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 800.6727
WagonsDest = 400
TrainsDest = 8
TrainsDestDay = 0.3840
Freightarrivals = 1800
Efficiency = 66.2983
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

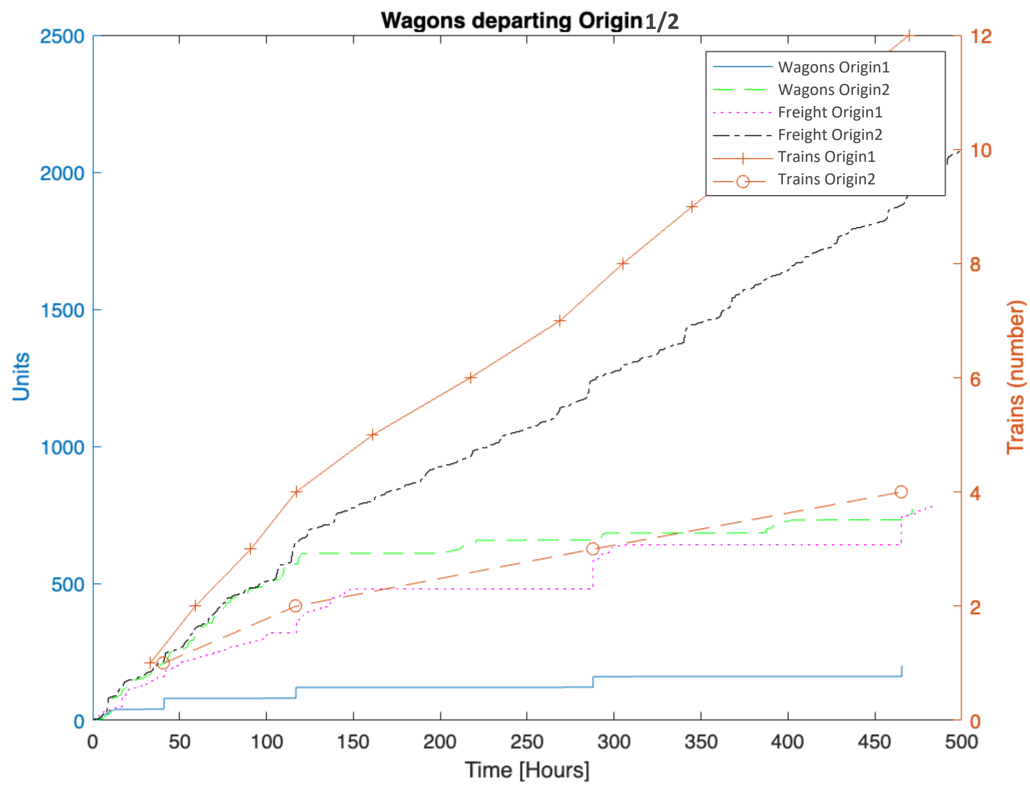
```
Fleetsize = 12.8108
```

Data analysis - Standard Run 15

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

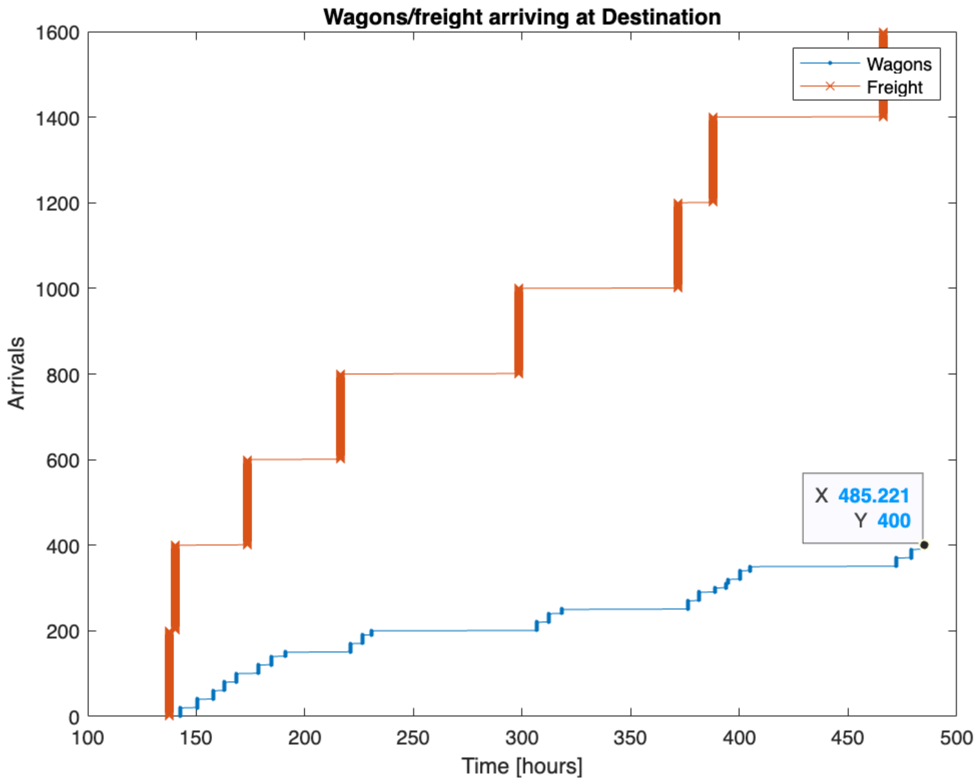
Origin 1 and 2 are production points.




```
Origin1Prod = 800
Origin2Prod = 2075
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 907.3483
WagonsDest = 400
TrainsDest = 8
TrainsDestDay = 0.3840
Freightarrivals = 1600
Efficiency = 55.6522
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

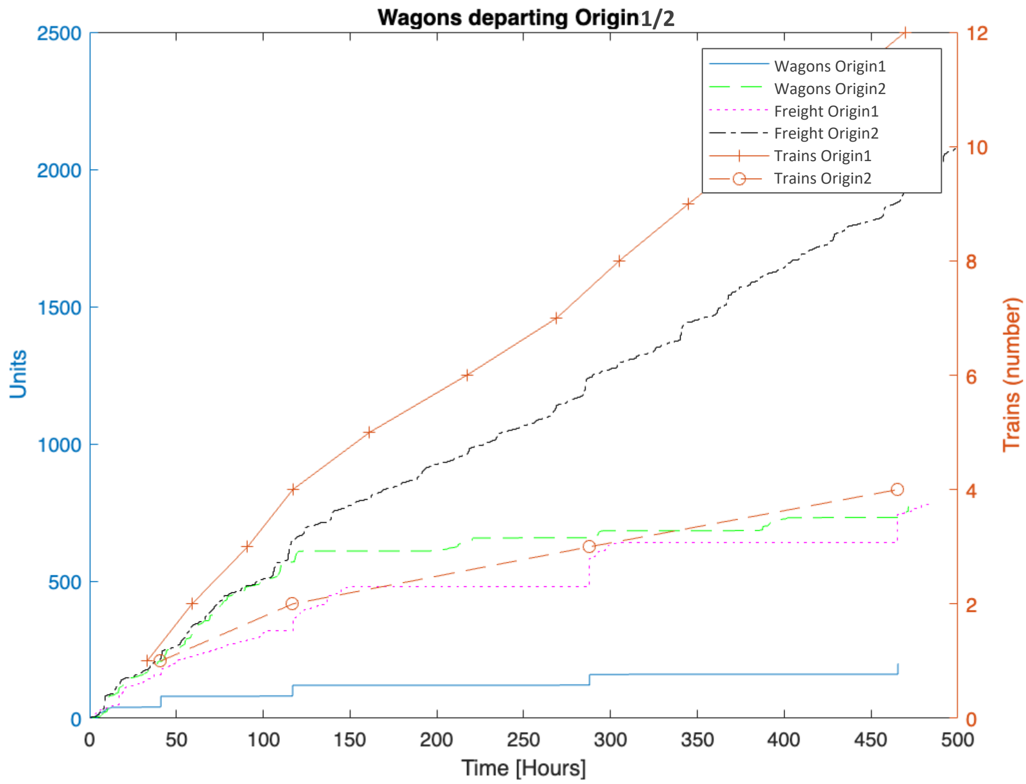
```
Fleetsize = 14.5176
```

Data analysis - Standard Run 16

Data from the simulation run is presented, no optimisation routine is performed on this Run, the train waits to make up the defined length and departs when the slot is open.

Origin

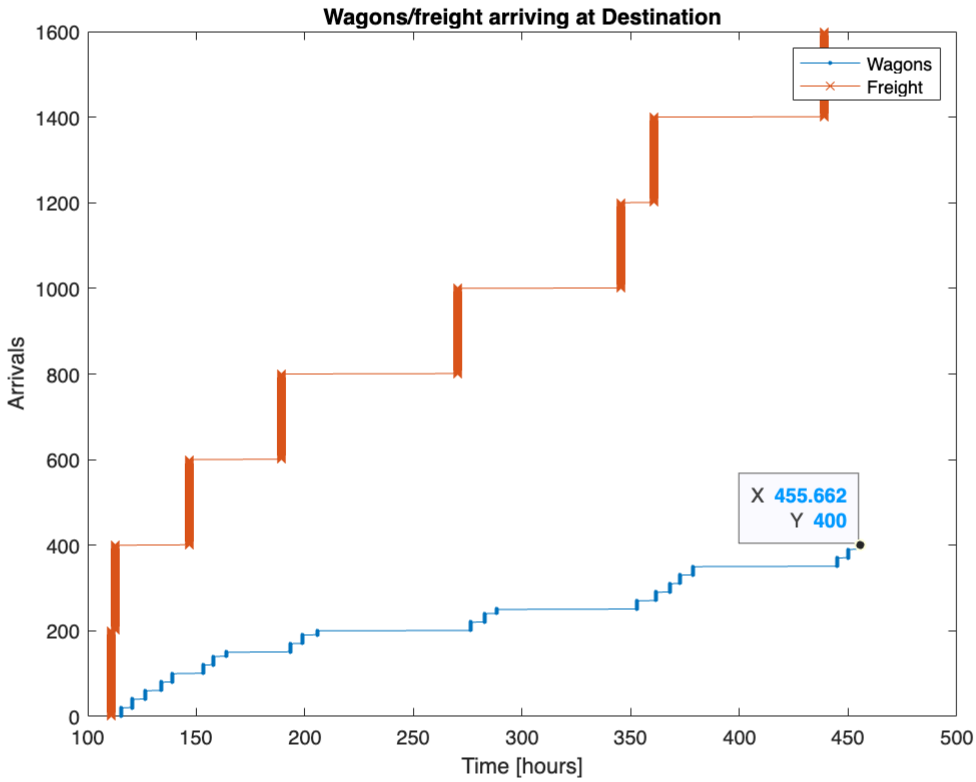
Origin 1 and 2 are production points.




```
Origin1Prod = 800
Origin2Prod = 2075
```

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.



Train length factors

The train length has been set at the maximum for the line section, this affects volume throughput and turnaround time, which has an effect on the fleet size. Therefore, a relationship has developed between the train length and the fleet size.

The maximum wagon turnaround time is:

```
ans = 852.7326
WagonsDest = 400
TrainsDest = 8
TrainsDestDay = 0.3840
Freightarrivals = 1600
Efficiency = 55.6522
```

Fleet sizes are calculated by:

$$\text{Fleet size} = \frac{\text{cycle time}}{24} * \frac{\text{loaded trains}}{\text{day}}$$

```
Fleetsize = 13.6437
```

Appendix D – Simulation results

Data analysis

Several sets of data were compiled. The model is complex and had to be debugged to ensure correct execution. Of the formal data sets (where complete runs were performed), complete data sets chosen for analysis are categorised as Result Sets 4 and 5, comprising of the following data:

- Result Set 5 – Optimised run. Optimisation is performed using a genetic algorithm to find the optimal train length, available in Appendix F.
- Result Set 4 – Standard run, no train length optimisation performed. Train lengths are set at predetermined maximum values, available in Appendix G.

In all cases that follow in this chapter the optimised run is presented first and the standard run second. The analysis shall focus on how the simulated train lengths are affected by various scenarios and what impact this has on the container volume throughput of the system.

It was observed that the data had similar characteristics (but were not the same) when the production rate and stack density of the containers were the same as shown in the Appendices where they are presented. Therefore, the results have been grouped into Runs 1-4, 5-8, 9-12 and 13-16. The results are presented with a representative Run from each of these groups.

Some graphs were observed to have a series that stops before the end of the time, while other series continue. This is due to the mapping algorithm in the MATLAB® script. Upon investigation it was found that this apparent stoppage of a series occurred in some cases where the value would remain constant from that point to the end (for example if no further wagons arrived at the port). While the data existed as a flat line, it was not drawn and had no effect on the analysis. An obvious case of this phenomena can be observed in Figure 0.3 and it occurs intermittently in the results. These simulation runs are presented in Section 9.2.

Data analysis - Run 1-4

Run 1 to 4 are low production rate, single stacked runs. These will be scenarios where low transportation capacity is required. Therefore, higher speeds may not have a benefit over lower speeds. Running shorter length trains may be possible without disadvantages.

Origin

Origins 1 and 2 are production points, they are factories producing goods that are packaged into containers. These containers may be transported to consumption points by any means. In this case rail is considered both before and after the hub.

Figure 0.1 presents departures from the origins for the optimised case, while Figure 0.2 presents a similar view for the standard case. These figures depict wagons, freight and trains departing from Origin1 and Origin2 over time.

The simulation has been run over a 500-hour period, chosen as this is sufficient for the system to stabilise. In the graphs the primary y-axis indicates quantity of wagons and of containers. The secondary y-axis indicates quantities of trains, as these are significantly smaller than the wagons and the containers. The x-axis indicates the time in hours.

The lines Wagons Origin1 and Wagons Origin2 indicate the wagons which are departing from these respective origins on the primary y-axis.

The lines Freight Origin1 and Freight Origin2 indicate the containers which have been loaded on the wagons and are departing from these respective origins on the primary y-axis.

The lines Trains Origin1 and Trains Origin2 indicate the trains which are departing from these respective origins on the secondary y-axis.

The logistic system is driven by the production rate in the origin (the factory). The production rate for each run is listed in Table 9-1. Containers are produced at the indicated rate and the origin module will call for all the components of a train as required. Provided the module can get the components, it will follow a predefined processing time to make up a train (as in the real world). Delays in obtaining any of the required components will have an impact on the departure rate of trains and therefore on the volume throughput of the system.

Table 9-1 also defines which runs are single stacked (2 containers per wagon) and double stacked (4 containers per wagon, note that this is not possible on a narrow gauge line as is used in South Africa).

In the case where an optimisation is performed, MATLAB® will attempt to minimise the cost by changing the train lengths and meeting the efficiency parameter.

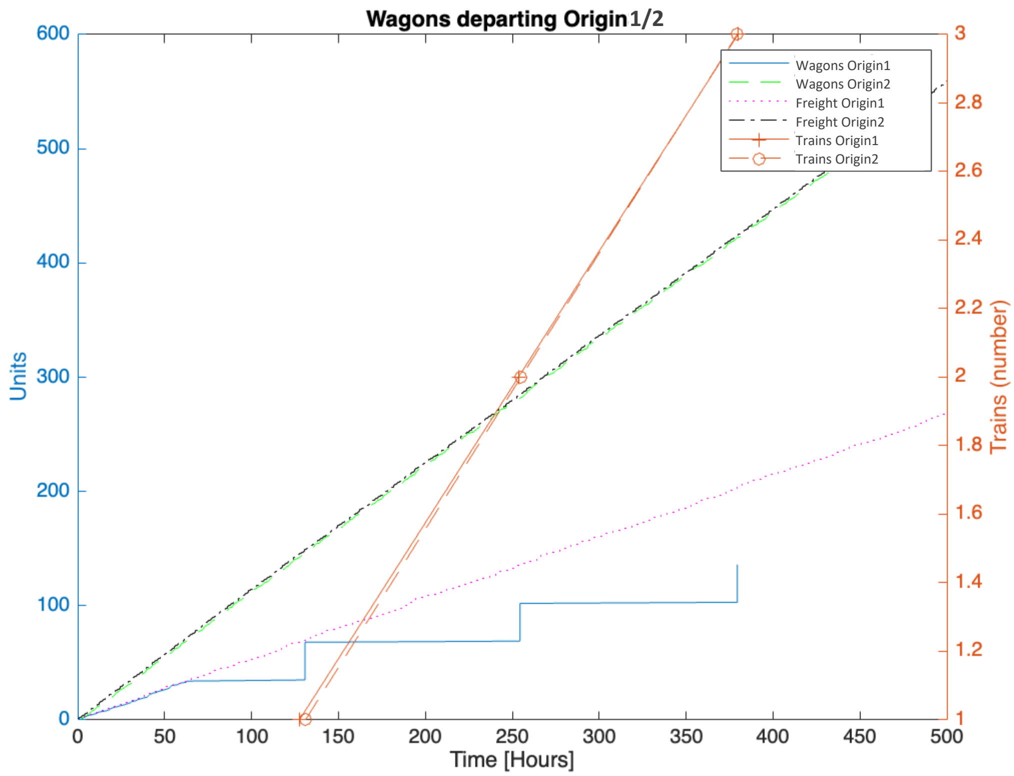


Figure 0.1 Optimised departures (Runs 1-4; short distance, slow speed, low production rate, single stack)

Figure 0.1 and Figure 0.2 show similar ramp up of train departures for the standard length case and optimum case.

Note that Origin 1 and Origin 2 have been designed with different timing mechanisms to control the slot times and train departures, to observe the simulation behaviour. Hence, the behaviour appears different at the point of measurement but does have the same effect for the train departure and make up. Essentially Origin 1 has a gate controlling access to the slot placed after the train is compiled and ready to depart while Origin 2 has gates placed at the point before the containers and wagons are combined.

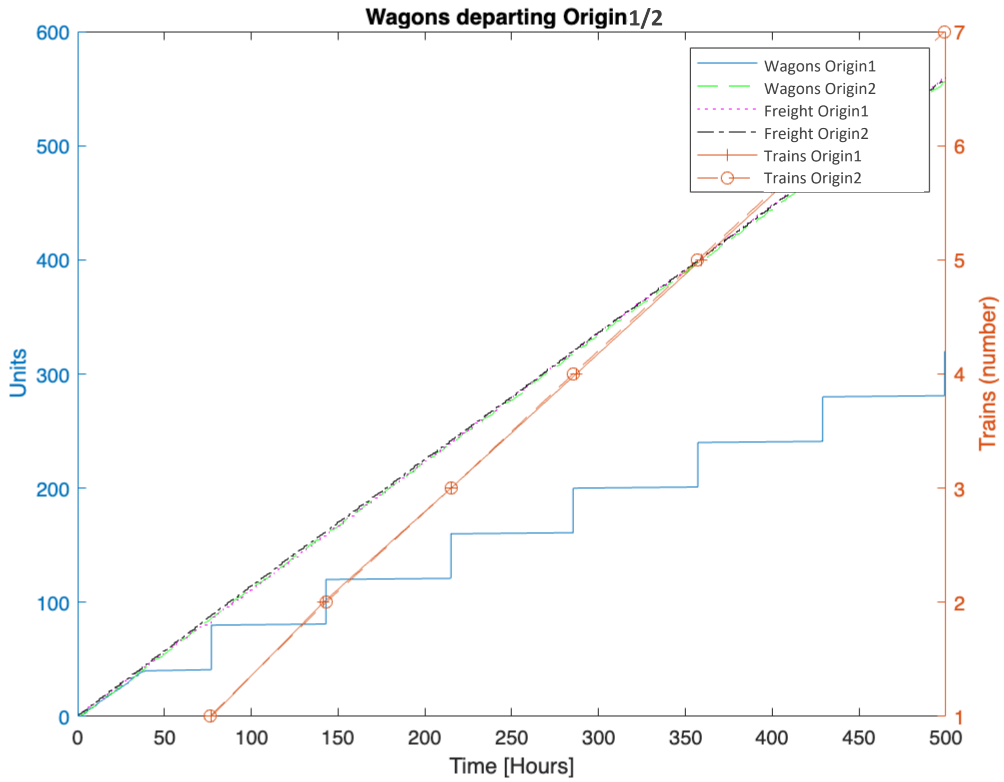


Figure 0.2 Standard departures (Runs 1-4; short distance, slow speed, low production rate, single stack)

Figure 0.2 was generated using an automated script in MATLAB®, as are all the result graphs presented here.

In the simulation, trains arrive at the Destination and the containers are offloaded, wagon and freight arrivals are shown in Figure 0.3. The freight wagons carry 2 teu's each in Runs 1-8. Figure 0.3 presents the arrivals at the destination for the optimised case, similarly, Figure 0.4 presents the standard case. The data is presented as wagons and freight arrivals over time.

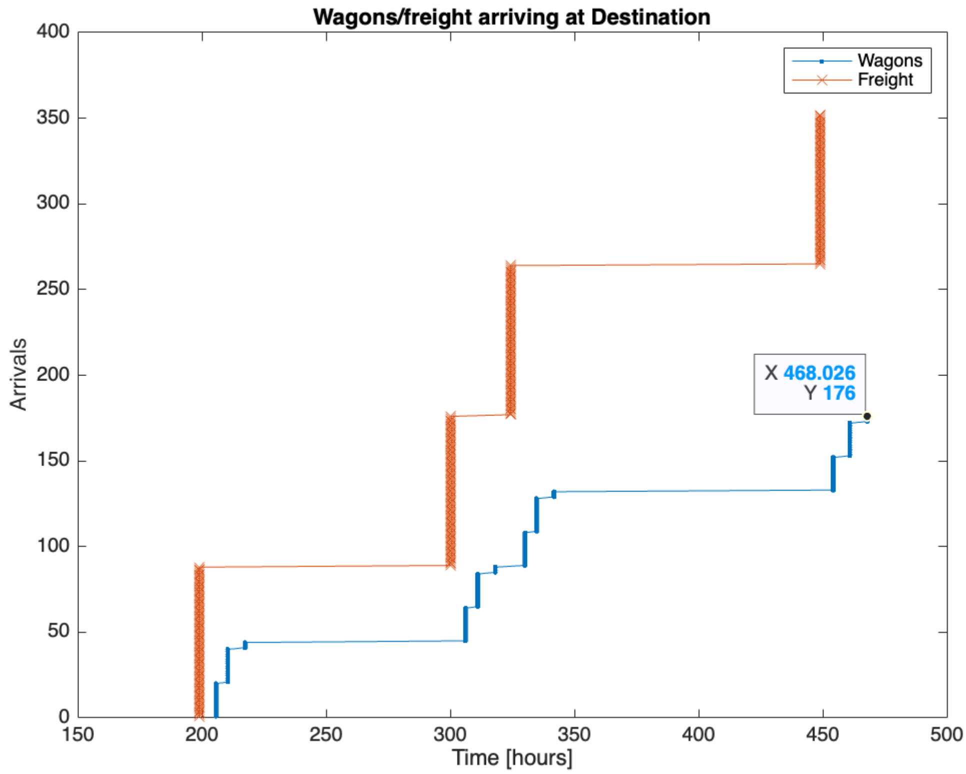


Figure 0.3 Run 1 optimised destination (short distance, slow speed, low production rate, single stack)

The destination is located at the termination of the long hub to hub rail journey. Therefore, at this point only one type of train, compiled at the hub arrives (not various trains from multiple origins). As the boundary of the study, the volumes arriving at the destination represent the throughput of the system.

Wagon arrivals have some variability as the wagons are processed (offloaded and moved to the yard and examined). Freight arrivals are measured in batches when a train arrives as the containers are not considered in this study after that point.

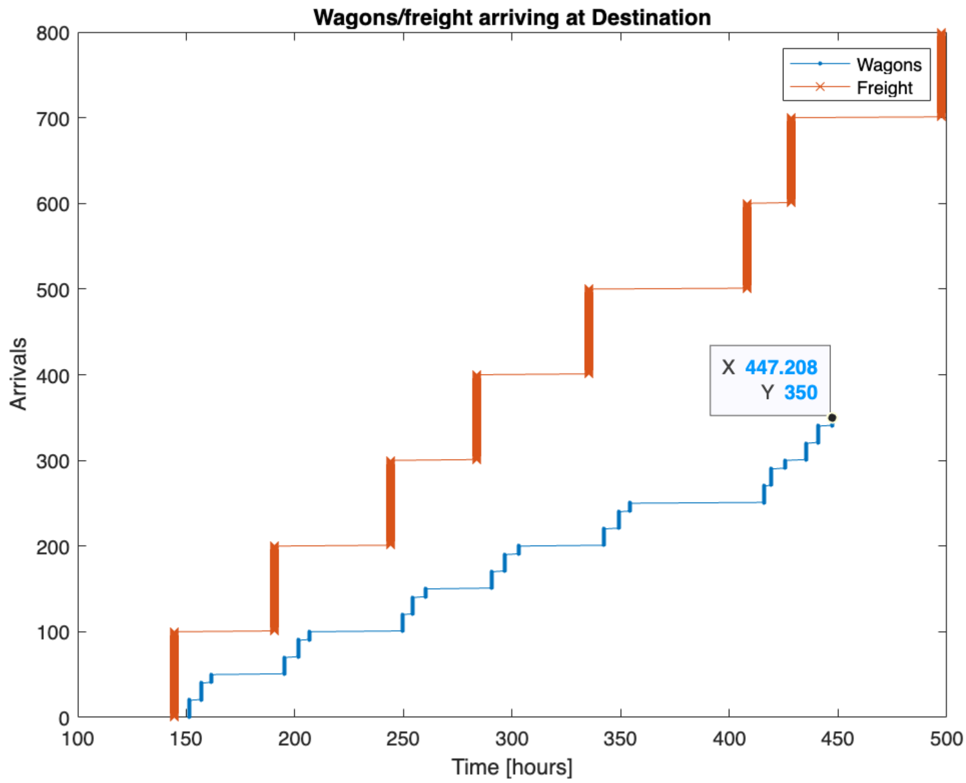


Figure 0.4 Run 1 standard destination (short distance, slow speed, low production rate, single stack)

Figure 0.3 and Figure 0.4 show higher arrivals at the destination for the standard case at 800 teu's compared 350 teu's for the optimised case.

Runs 5-8 are high production rate and low density runs. In this environment, transport capacity would be constrained.

Origin

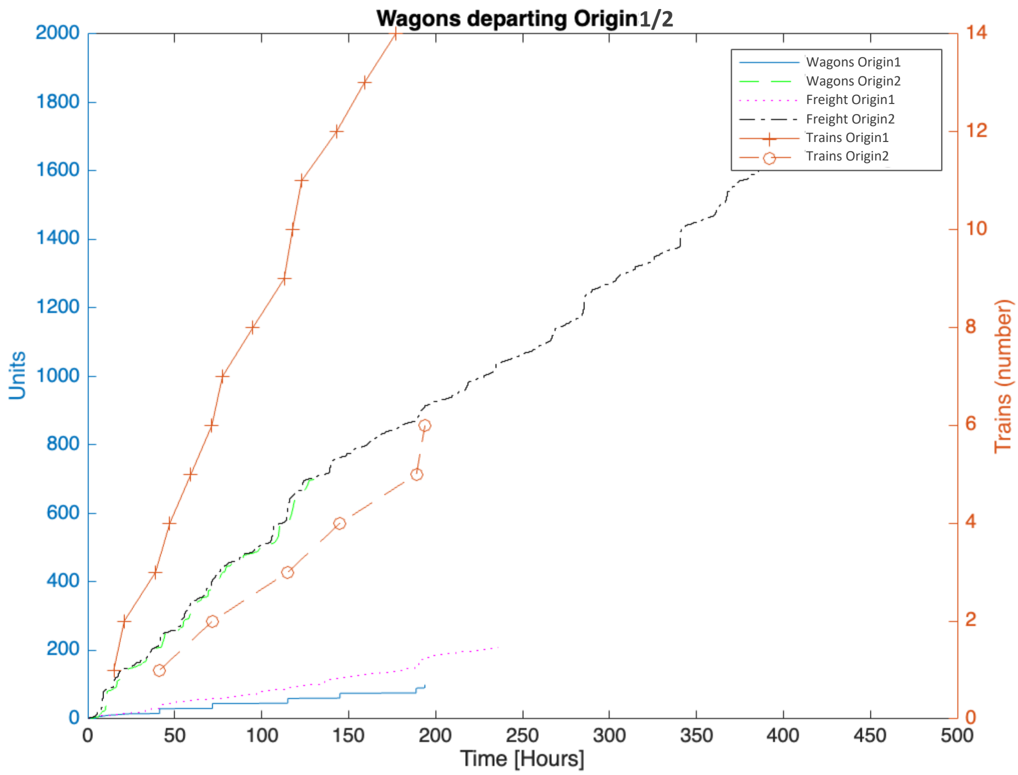


Figure 0.5 Optimised departures (Run 5; short distance, slow speed, high production rate, single stack)

It is observed that the optimised departures begin at a similar rate in Figure 0.5 and Figure 0.6, compared to the standard departures. However, once the resources become depleted, the train departure rate slows down in the standard case. This is a result of the system being unbalanced in terms of the rolling stock. This is not true for the optimised run though, as train departures continue at a steady pace in Figure 0.5.

This indicates that with the frequent, maximum length train departures in the standard case, the system is unbalanced and cannot cope with the volumes when operated in the fixed length type of operation.

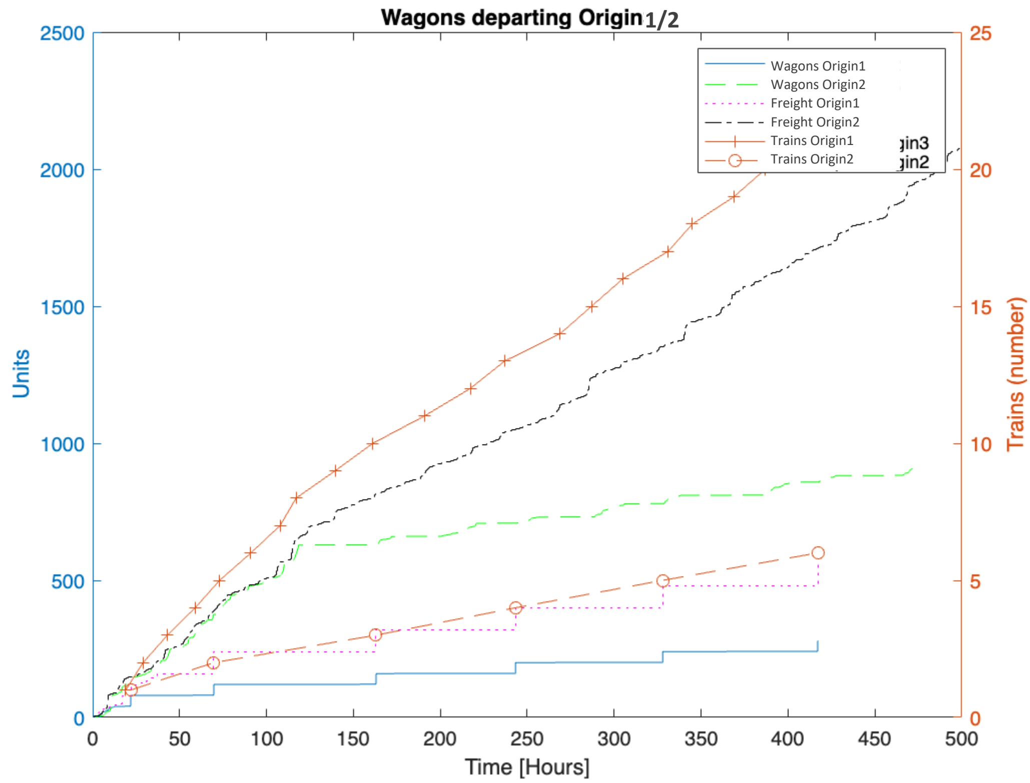


Figure 0.6 Standard departures (Run 5; short distance, slow speed, high production rate, single stack)

Destination

Trains arrive at the Destination, showing wagons and freight arrivals.

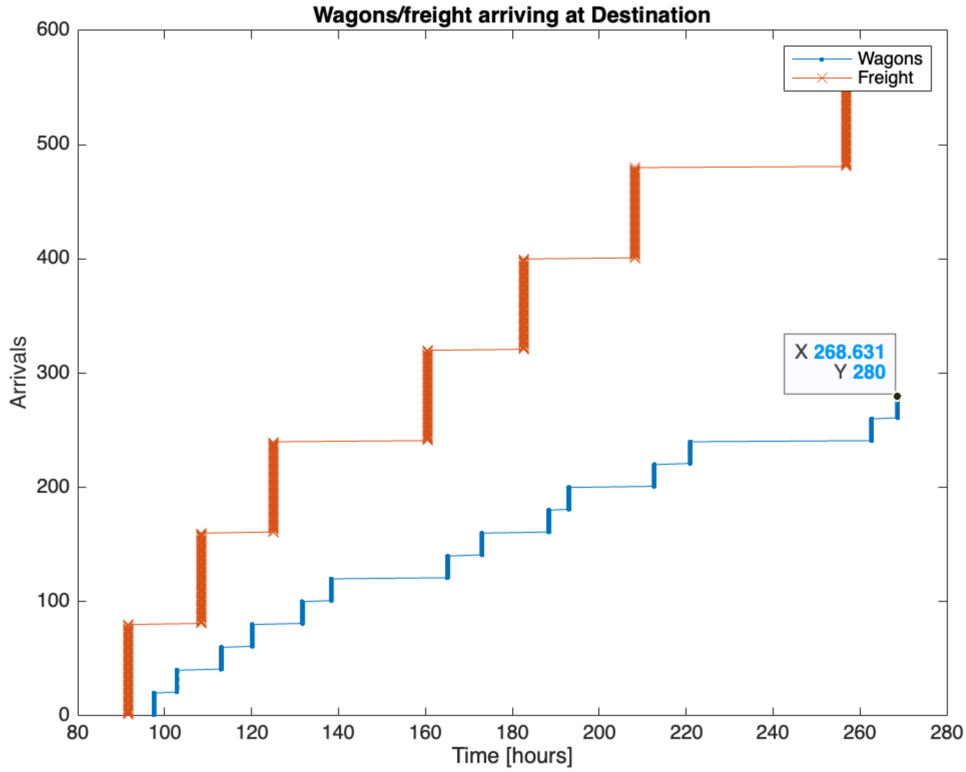


Figure 0.7 Optimised arrivals (Run 5; short distance, slow speed, high production rate, single stack)

In Figure 0.7, the arrivals at the port begin at a higher rate and slow down as the system becomes unbalanced, with insufficient resources to sustain the initial departure rates. This is also attributable to longer delays on the main line for some trains.

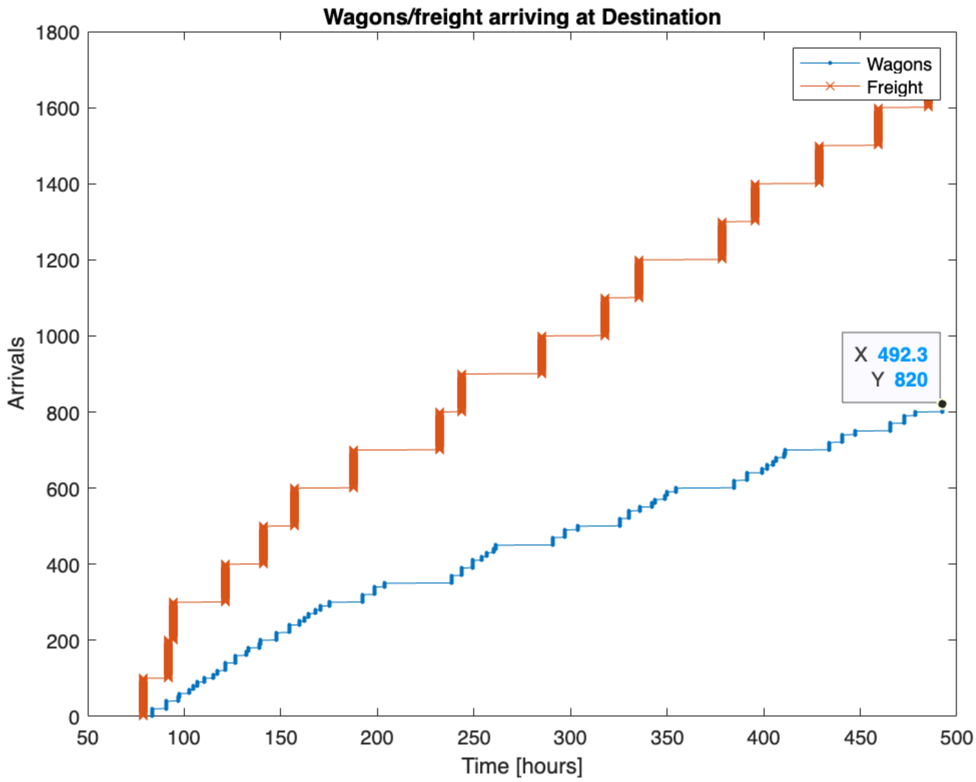


Figure 0.8 Standard arrivals (Run 5; short distance, slow speed, high production rate, single stack)

Figure 0.7 and Figure 0.8 show arrivals at the destination in much higher quantities for the standard case. In this case, standard length trains can depart on time and achieve significantly more volume throughput, even as they are affected by resource imbalances and main line delays.

Run 9-12 are low production rate and high density runs, with Runs 9-16 also being double stacked trains. Run 11 has been chosen as the representative run.

Origin

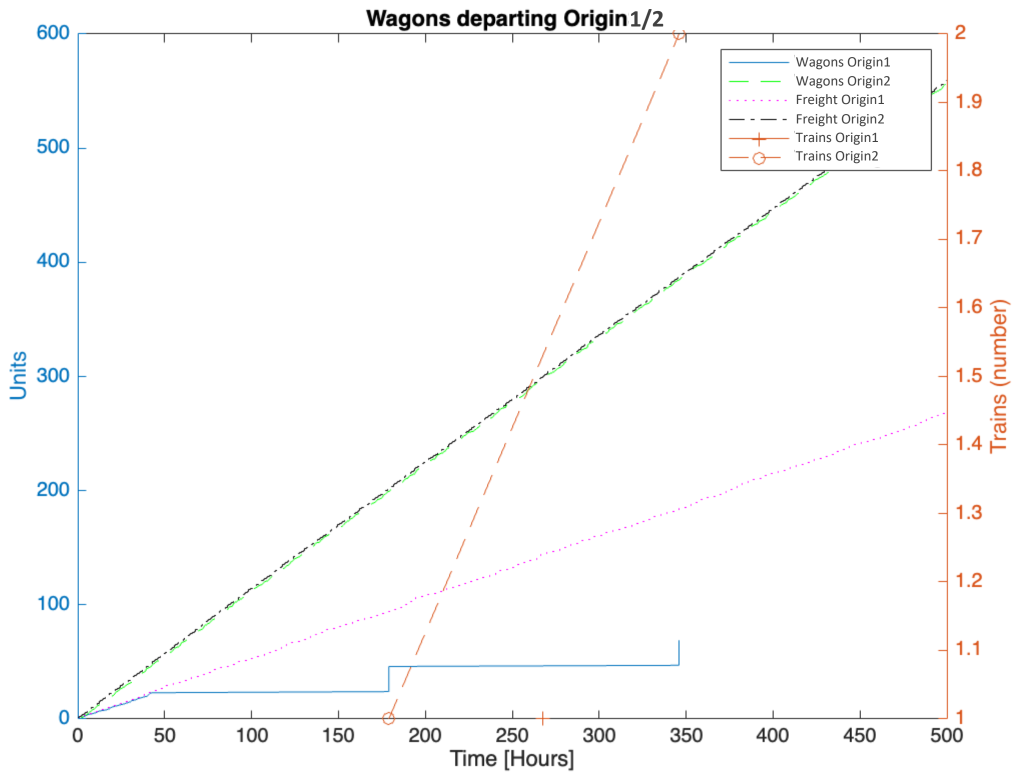


Figure 0.9 Optimised departures (Run 11; long distance, slow speed, low production rate, double stack)

It is observed that the freight from Origin 2 experiences a bottleneck and stops while waiting for a train in Figure 0.9. This can be attributed to Run 11 having long distances and slow speeds. Therefore, resources are constrained. Interestingly, for all cases in this partition (Run 9-12), trains from Origin2 could not depart due to the conditions not being met.

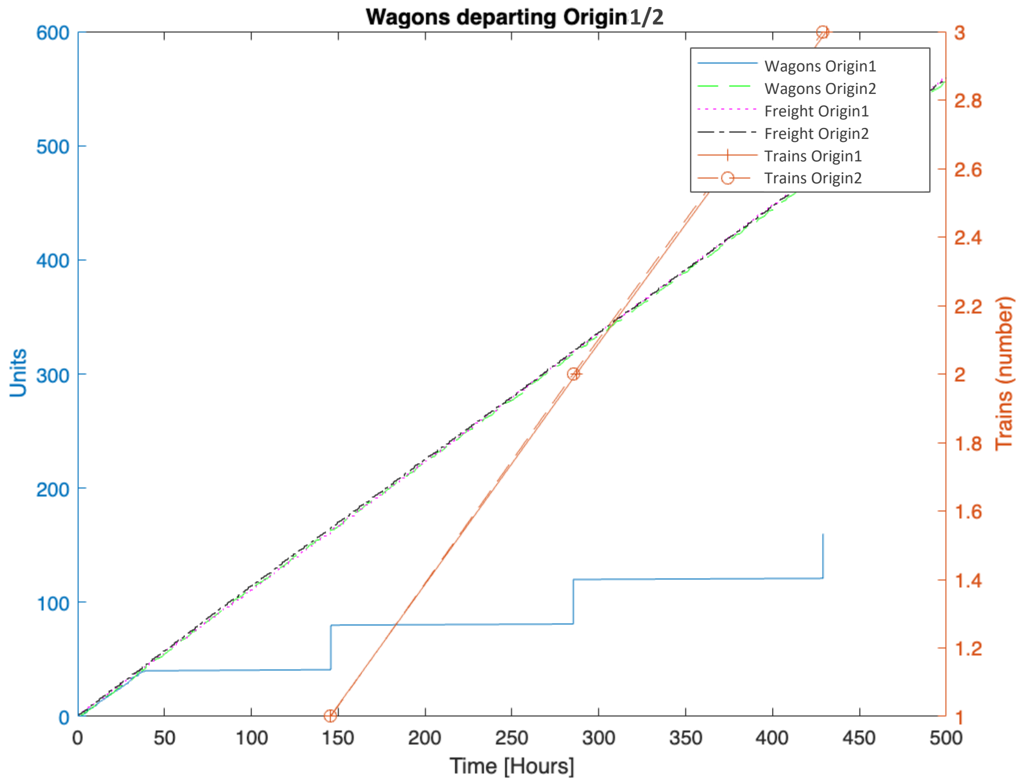


Figure 0.10 Standard departures (Run 11; long distance, slow speed, low production rate, double stack)

During the standard run in Figure 0.10, train departures occur simultaneously from Origin1 and Origin2. The standard run allows for a balanced use of the resources under these low productivity conditions.

Destination

Trains arrive at the Destination, showing wagons and freight arrivals in Figure 0.11.

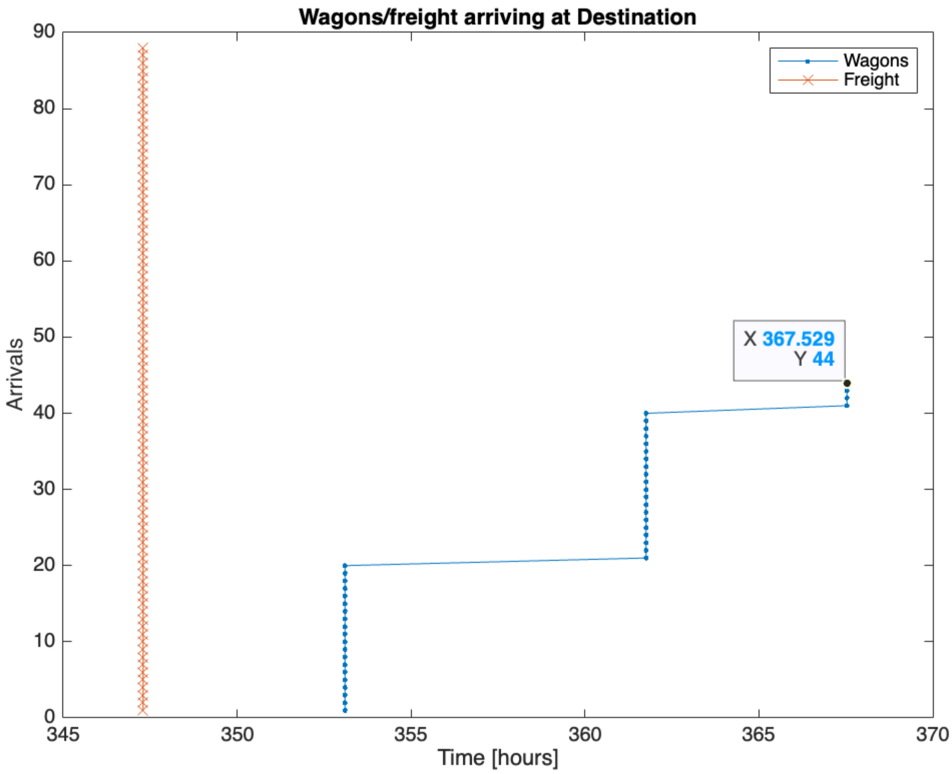


Figure 0.11 Optimised arrivals (Run 11; long distance, slow speed, low production rate, double stack)

Figure 0.11 and Figure 0.12 show similar arrival profiles for both optimised and standard runs, although the standard run arrivals are higher at 400 teu's compared to approximately 90 teu's. The data for Runs 9-12 exhibits the same peculiarities for all graphs with freight arrivals seemingly disconnected from the wagon arrivals

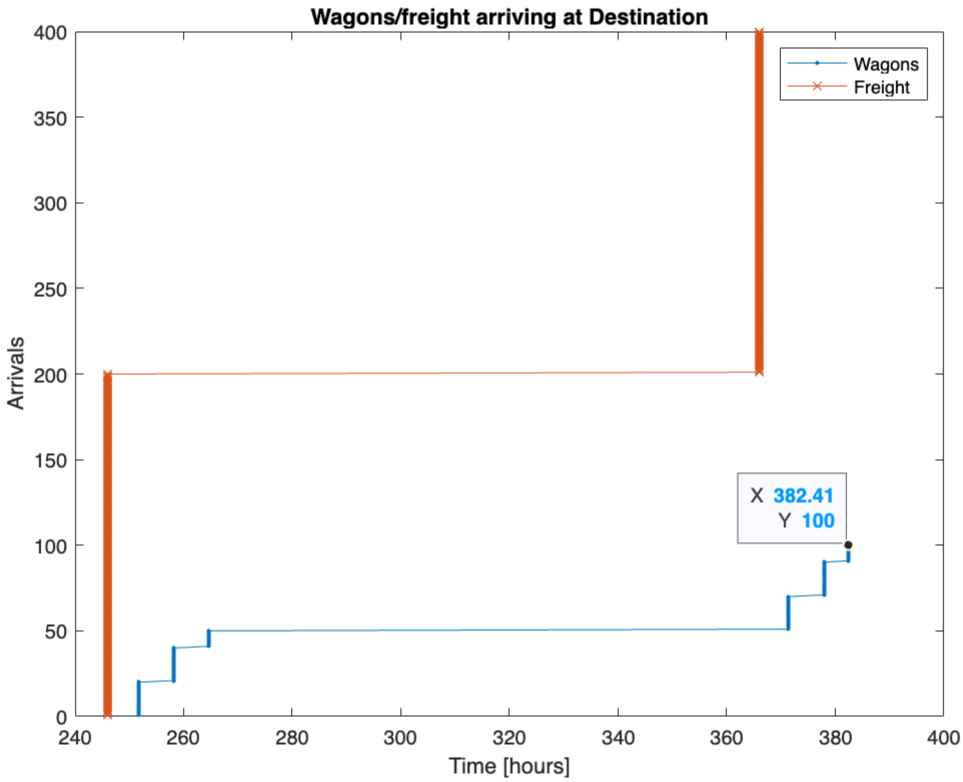


Figure 0.12 Standard arrivals (Run 11; long distance, slow speed, low production rate, double stack)

Run 11 is resource constrained and has very limited train arrivals, depicted in Figure 0.12.

Runs 13-16 are high production rate and high density. While they are similar to Runs 5-8 in terms of the production rate, the trains in these runs are double stacked, therefore they have double the capacity.

Origin

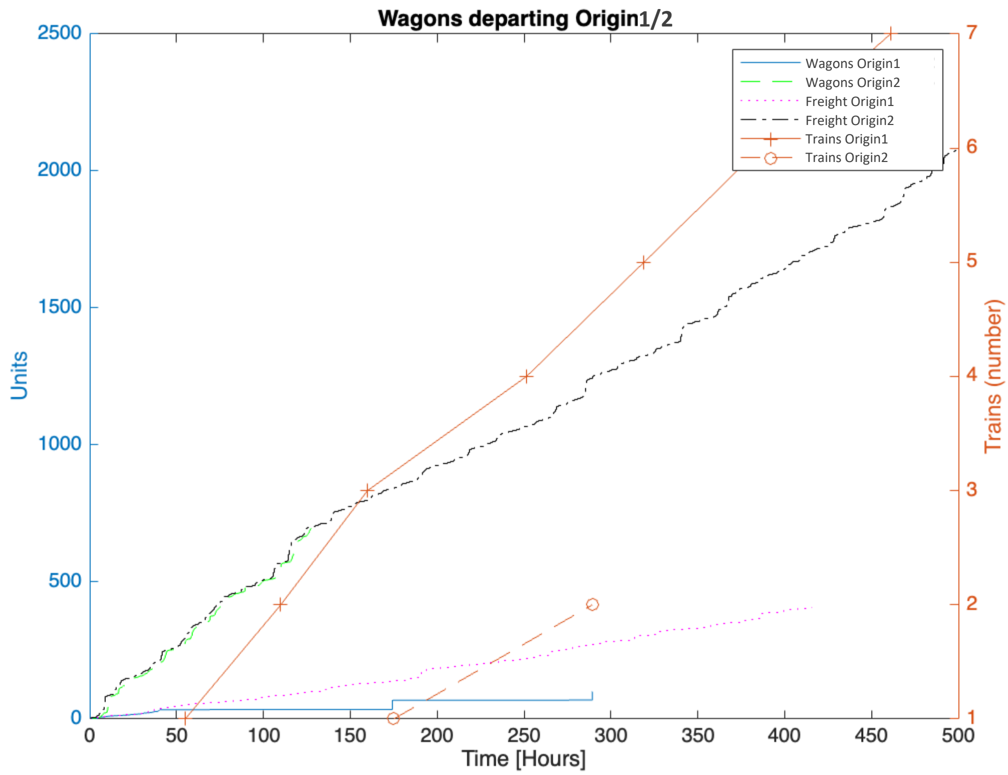


Figure 0.13 Optimised departures (Run 13; short distance, slow speed, high production rate, double stack)

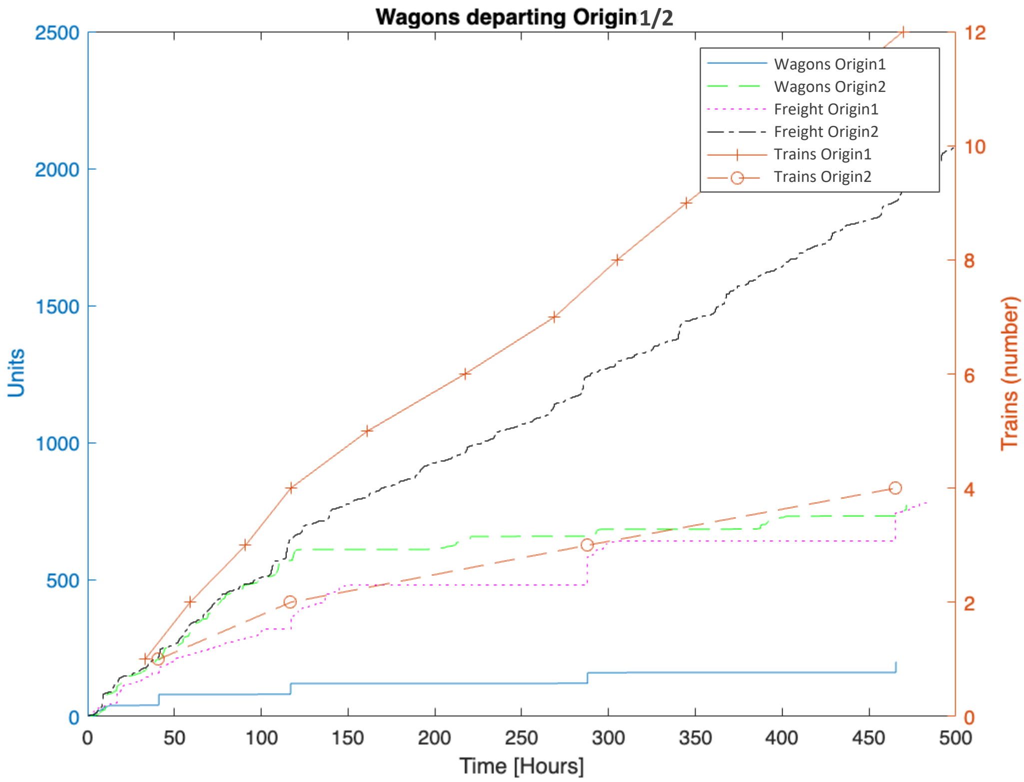


Figure 0.14 Standard departures (Run 13; short distance, slow speed, high production rate, double stack)

Figure 0.13 and Figure 0.14 show that both optimised and standard train departures begin more rapidly and slow to a lower rate due to resource constraints in the high production rate environment.

Trains arrive at the Destination, showing wagons and freight arrivals.

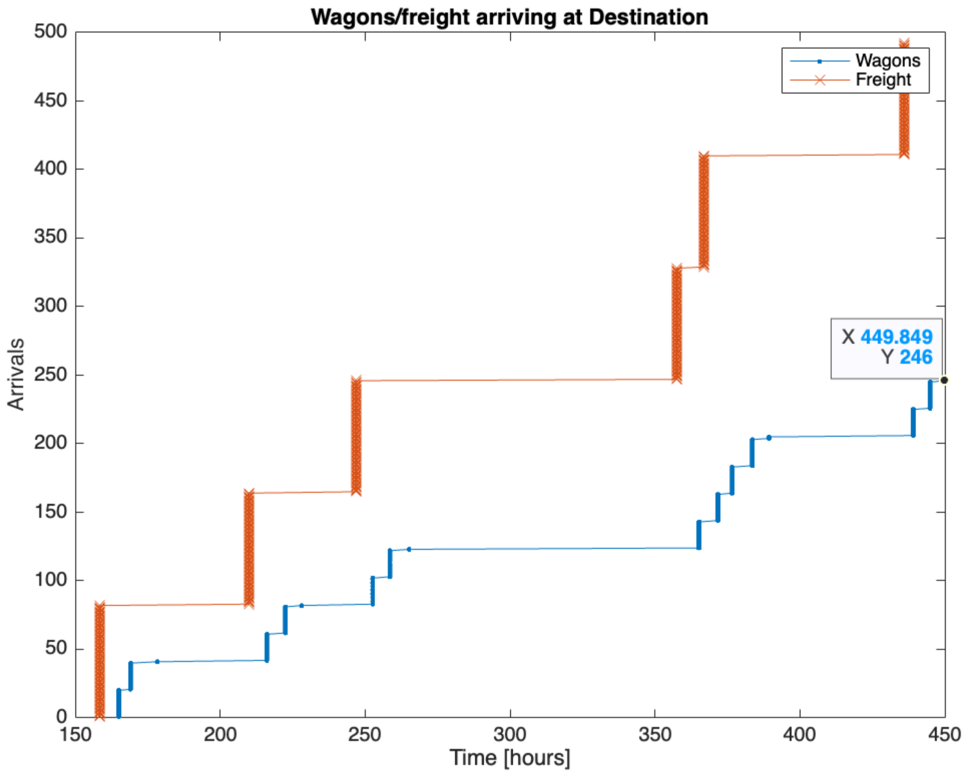


Figure 0.15 Optimised arrivals (Run 13; short distance, slow speed, high production rate, double stack)

In Run 13, depicted in Figure 0.15, high volume throughput is achieved for both the standard and optimised runs, however the optimised system becomes unbalanced later in the run, therefore as resources are constrained, the volume throughput is reduced.

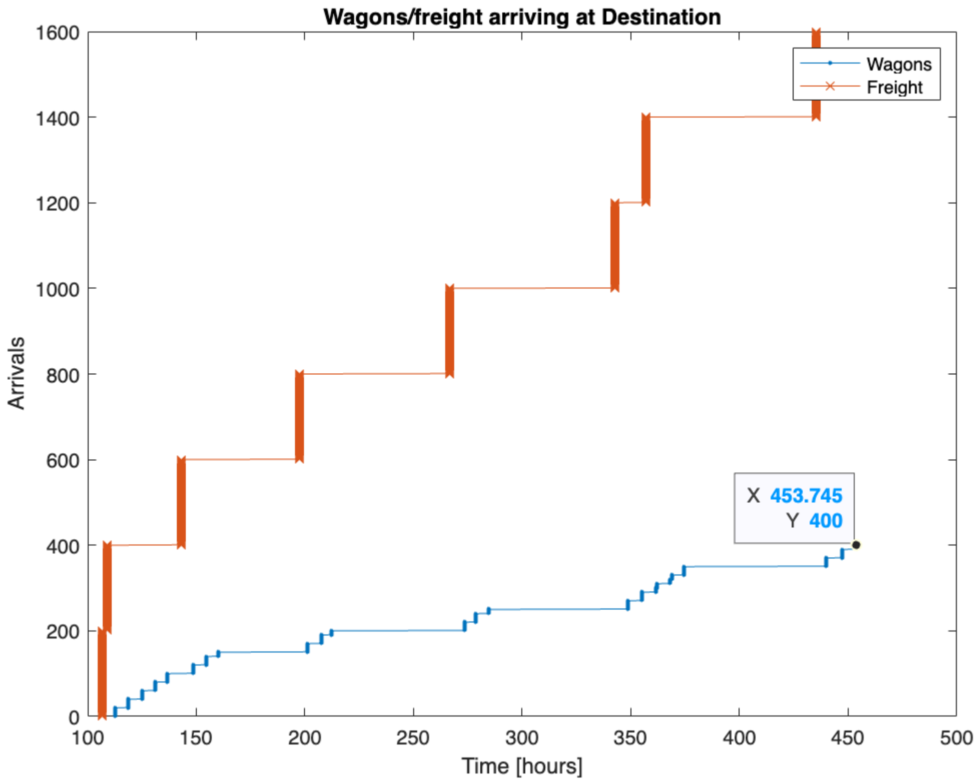
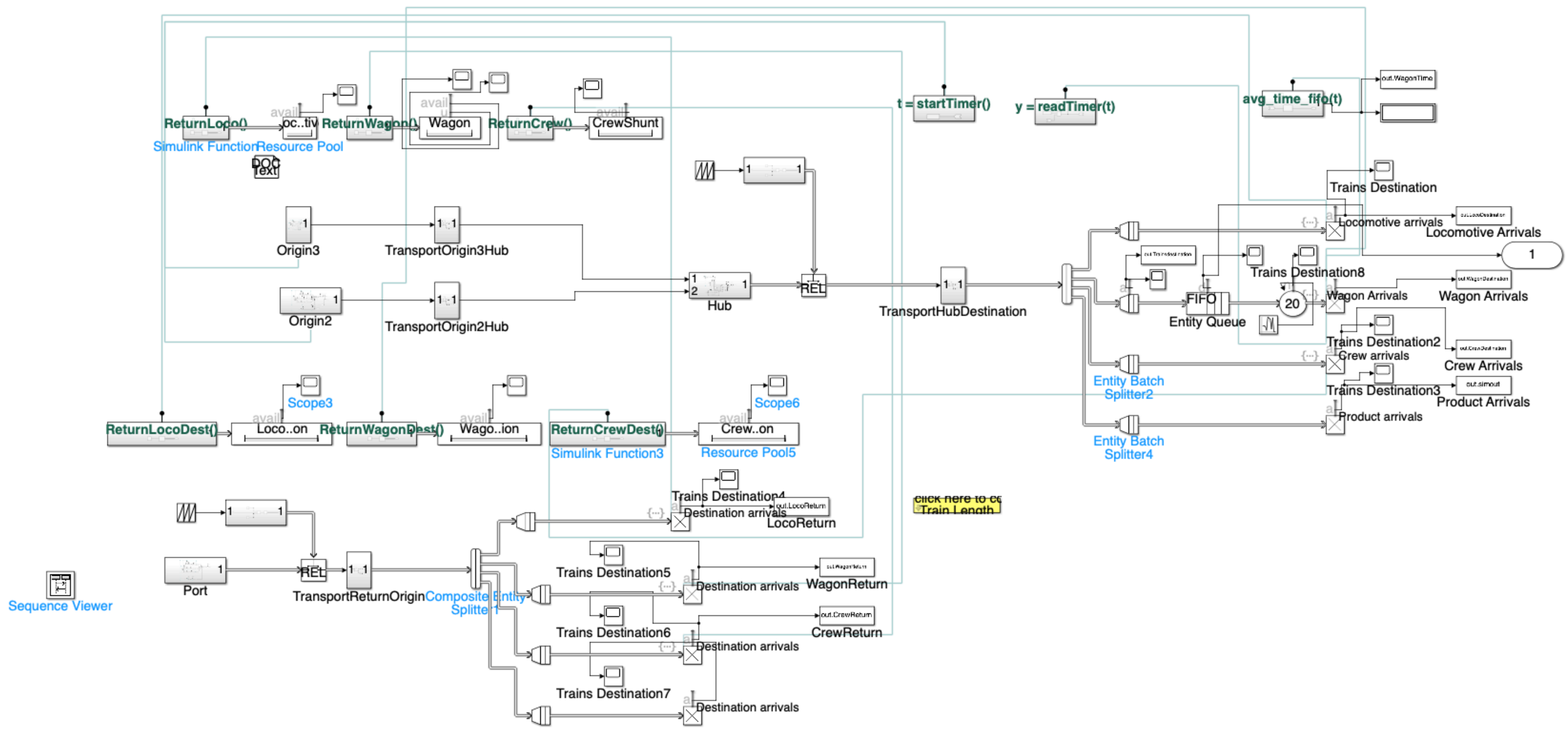


Figure 0.16 Standard arrivals (Run 13; short distance, slow speed, high production rate, double stack)

Figure 0.15 and Figure 0.16 show higher arrivals for the standard case, which appears less affected by a resource imbalance.

Appendix E – Simulation model – High level view



Detailed information about building models in Simulink using the SimEvents toolbox (for discrete event simulation) can be found here:

https://www.mathworks.com/help/simevents/index.html?s_tid=CRUX_lftnav

Appendix F – Simulation model setup

```
% HubSpokeSystem

modelname = 'HubSpokemodelbasic';
open_system(modelname);
scopes = find_system(modelname, 'LookUnderMasks', 'on', 'BlockType', 'Scope');
cellfun(@(x)close_system(x), scopes);
set_param(modelname, 'SimulationCommand', 'update');

sim(modelname);

%Optimisation
cellfun(@(x)close_system(x), scopes);

% Set model parameters Run8
O2HDist = Simulink.Parameter(3.08);
O3HDistl = Simulink.Parameter(2.31);
O2Rate = Simulink.Parameter(0.02);
O3Rate = Simulink.Parameter(0.02);
MainlineD = Simulink.Parameter(24);
%OHDistH = Simulink.Parameter(28);
disp('TrainLength before optimisation =');
disp(TrainLength);
%close_system([modelname '/Data Analysis/Order Backlog']);
TrainLength = RunTrainLengthOptimisation();
disp('TrainLength after optimisation =');
disp(TrainLength);

sim(modelname);
```

Appendix G – Simulation optimisation script

```
function finalResult = RunTrainLengthOptimisation()
% Find optimal resource capacities of an intermodal hub and spoke
system.
%  Apply Genetic Algorithm solver of MATLAB Global Optimisation
Toolbox and
%  SimEvents Discrete-Event Simulation to find optimal resource
capacties.

% The genetic algorithm solves optimisation problems by
repeatedly
% modifying a population of individual points. Due to its random
nature,
% the genetic algorithm improves chances of finding a global
solution.
% It does not require the functions to be differentiable or
continuous.

% Decision variables are the
%   # TrainLength(1-4)

% These are set by the genetic algorithm as it runs multiple
% simulations of the Hubspokemodelbasic model via the variable
% TrainLength(1-4).

% Open a parallel pool
pool = parpool;

% Setting plot and parallelization options
%opts = gaoptimset(...
%    'UseParallel', 'always',...
%    PlotFcns=@gaplotbestf, ...
%    Generations=25, ...
%    StallGenLimit=10);
opts = optimoptions('ga',...
    'PlotFcn','gaplotbestf',...
    'MaxGenerations',50,...
    'MaxStallTime',50);

% Loading the model to run on all parallel workers
pctRunOnAll('load_system(''HubSpokemodelbasic'')');

% Lower bound of decision variables
lb = [15 15 40 40];

% Upper bound of decision variables
ub = [40 40 50 50];

% Integer constraints: If the third was not an integer, it would
be [1, 2,
% 4]
IntCon = [1 2 3 4];

% Track time spent for optimisation
tic;
```

```
% Execute genetic algorithm solver
[finalResult, ~, ~] = ga(@volumeThroughput, 4, [], [], [], [],
...
    lb, ub, [], IntCon, opts);
toc;

% Shut down the parallel pool
delete(pool);

end
%%

% Cost function that assign different values to the decision
variables in
% the model
function obj = volumeThroughput(TrainLength)
% Assigns costs to the values of ResourceCapacity, which
correspond
% to [Origin2, Origin1, Hub, Port]
%VarC = [1000 1300 700 400] * TrainLength';

% Assigns variables to the base workspace for simulation
assignin('base', 'TrainLength', TrainLength);

% Simulation of the model and assigns output to the variable z
if isempty(find_system('type', 'block_diagram', ...
        'Name', 'HubSpokemodelbasic'))
    load_system('HubSpokemodelbasic.slx');
end

modelname = 'HubSpokemodelbasic';


set_param('HubSpokemodelbasic/ConfigResource','Origin2TrainLength
', ...
    num2str(TrainLength(1)),'Origin1TrainLength',...
    num2str(TrainLength(2)),'HubTrainLength',...
    num2str(TrainLength(3)),'PortTrainLength',...
    num2str(TrainLength(4)));

%[~, ~, z] = sim('HubSpokemodelbasic.slx');
z = sim(modelname, 'ReturnWorkspaceOutputs', 'on', 'SaveFormat',
'Dataset');
assignin('base','z',z);

%Fixed and variable costs
Cost31 = [cost];
DistOH2=80;
DistOH3=60;
DistMainline=1200;

% Calculates the objective function, based on TEUkm costs
%obj = Origin2Cost + Origin1Cost + MainlineCost
obj =
(2*TrainLength(1)*z.TrainDepart.signals.values(end)*DistOH2*Cost3
1) +...
```

```
(2*TrainLength(2)*z.TrainDepart2.signals.values(end)*DistOH3*Cost
31)+...

(2*TrainLength(3)*z.HubDepart.signals.values(end)*DistMainline*Co
st31)+...
    R1*max(epsi*(z.Production.Data(end)+z.Production2.Data(end))-
z.simout.signals.values(end),0);
end
```

Appendix H – Ethics waiver letter

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

HUMAN RESEARCH ETHICS COMMITTEE (NON-MEDICAL)

Registration number: REC-101114-044

20 January 2023

Re: Mr. Denesh Reddy (1589916)

Waiver letter number: HRECNMW23/02/01

To whom it may concern,

Mr. Reddy is a registered PhD student at the School of Mechanical, Industrial and Aeronautical Engineering at the University of the Witwatersrand, Johannesburg. This letter is to confirm that, at the time of writing, Mr. Reddy does not need ethical clearance for his PhD study entitled *'Optimizing the throughput of a complex multimodal freight network'*. This decision has been reached based upon a description of the project supplied by Mr. Reddy to the University Human Research Ethics Committee (Non-Medical), which has been evaluated by the Chairs and Deputy Chairs. If, however, if Mr. Reddy changes the methods of data collection and analysis for this study, this decision may no longer be valid. If such changes take place, this should be communicated to the University Human Research Ethics Committee (Non-Medical) as soon as possible. This waiver letter is valid until 19 January 2026.

Please feel free to contact me should you require any further information.

Thank you.

Yours sincerely,
S Schoeman

Shaun Schoeman (Administrative Officer)
Solomon Mahlangu House, 10th Floor, Room 10004, Jorissen Street, Braamfontein, Johannesburg
Private Bag 3, Wits 2050
T + 27(0)11 717 1408 | **E** Shaun.Schoeman@wits.ac.za | hrec-medical.researchoffice@wits.ac.za
www.wits.ac.za/research/about-our-research/ethics-and-research-integrity/