Solving the Air Express Transportation Problem using a combined Route Construction Algorithm and a Set-Partitioning Model

MSc Thesis Air Transport Operations

Floris Hermans

TU Delft

Solving the Air Express Transportation Problem using a combined Route Construction Algorithm and a Set-Partitioning Model

MSc Thesis Air Transport Operations

by

Floris Hermans

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on 16/01/2023.

Student number:	4596617	
Project duration:	09/2021 – 12/2022	
Thesis committee:	Dr. A. Bombelli	TU Delft, supervisor
	Dr. B.F. Santos	TU Delft, chair
	Dr. J. Sun	TU Delft, examiner

Acknowledgements

This thesis is the final step in finishing the master's program Air Transport Operations at the Delft University of Technology. Finding a thesis subject in line with my personal preferences was important to me considering the effort I needed and wanted to put into the project. The courses I attended during the first year of the master's program that I enjoyed most were mainly on the subject of operations optimization. Using different techniques and approaches to construct models that solve operational problems in order to find more efficient solutions has been of great interest to me since the start of the master's program. That is why my search for a thesis subject was pointing towards this area of research, preferably in combination with using real-life data since this would bring even more relevance to the solution. Alessandro Bombelli told me about an interesting thesis subject on bin packing of containers for aircraft. Whilst this subject is in line with my interests, another subject mentioned by Alessandro caught my attention even more: optimizing global shipment routing for an air express courier. Firstly, the idea of working on an optimization problem considering a global network really interested me because of the problem size and corresponding challenges. Secondly, this subject included real-life data from an air express courier which was in line with my mentioned interest in finding a solution with real-life relevance. This brought me to the interesting thesis subject of solving an air express transportation problem.

Upon finalizing this thesis, I would like to thank Alessandro for proposing this interesting thesis subject to me and even more for the conversations and guidance throughout the project. Your expertise and knowledge helped me out with some new or slightly different views on the problem during the more difficult parts of the research. Thanks for this guidance and of course the good coffee from the Italian coffeemakers at the Fellowship!

Since this research was conducted in combination with an air express courier, much-appreciated guidance and industrial insights have been given by my daily supervisor at the company, Francesco. I would like to thank you for the weekly meetings we had in which we discussed not only the potential theoretical solutions to the problem that I had in mind but also the implementation on the programming level. Your expertise first of all really helped me out in solving the problem at hand, but secondly added a business view to the problem which helped me to find a real-life relevant solution that could actually be useful for an air express courier. Even though our meetings have unfortunately mainly been online, your guidance has been extremely helpful for me so thank you for that.

Floris Hermans
Delft, December 2022

Contents

List of Figures

List of Abbreviations

3PL	Third-party logistic
AETP	Air Express Transportation Problem
AF	Available Factor
AM	The Americas
ATFFSP	Air Transportation Freight Forwarding Service Problem
dst	Destination Facility
edd	Expected Delivery Date
EU	Europe
FFTP	Freight Forwarding Transportation Problem
KPI	Key Performance Indicator
LF	Load Factor
LP	Linear Programming
LSFP	Liner Ship Fleet Deployment Problem
MILP	Mixed-integer Linear Programming
OD	Origin - Destination
org	Origin Facility
pu	Pickup Time
RCA	Route Construction Algorithm
RMSE	Root Mean Squared Error
SNDP	Service Network Design Problem
SPM	Set-Partitioning Model
TP	Transshipment Problem
ULD	Unit Load Device
UTC	Universal Time Coordinated
VRP	Vehicle Routing Problem
wgt	Weight

Introduction

Many global events such as the current energy crisis, the war in Ukraine, the COVID-19 pandemic, and much more influence the world trade of goods. Despite these events, the forecast by the IMF World Economic Outlook [1] shows that the annual growth of trade in goods will be higher in 2023 than it has been for the past decade. This calls for transportation companies to optimize their global processes and increase their efficiency more than ever before. One of the aspects to optimize is the construction and choice of routes for each of the shipments transported by freight forwarders and air express couriers. The current implementation of international transportation of goods by freight forwarders and express companies is based on a sequential selection of the next movement between facilities along a route from origin to destination. Movements in this case are both ground rides and air flights with a specific origin, destination, departure time, and arrival time. This approach works well in practice but lacks a view of steps further ahead in the network. A preview of the entire network can provide viable information in the selection of movements for each of the shipments. A more cost-effective solution might be to not select the next available movement at a specific facility but rather a movement with a later departure time.

The goal of this research is to formulate a model that finds a more cost-effective solution to this so-called Air Express Transportation Problem (AETP) than the historical implementation. A secondary goal is to construct adapted formulations to this model in which real-life strategies by the air express courier are implemented. In this research, both shipment and movement data are provided by the air express courier. The shipment data includes for example the pick-up and expected delivery time, the origin and destination facility, and the weight and volume of each shipment. A real-life relevant cost-effective solution must be found by the formulated models using this data, which not only improves the profitability of the courier but also improves the efficiency of the global trade of goods.

This thesis report is divided into three parts. Part I presents the scientific paper in which the problem is stated in further detail and the methodology of the proposed implementations is described. The results are discussed for each of the proposed implementations and both conclusions and recommendations for future work are drawn from these results. Part II of this thesis report elaborates on the available literature related to the subject of this thesis and presents potential research gaps. Additional information and results on subjects discussed in the scientific paper are given in Part III of this thesis report.

I

Scientific Paper

Solving the Air Express Transportation Problem using a combined Route Construction Algorithm and a Set-Partitioning Model

Floris Hermans,*

Delft University of Technology, Delft, The Netherlands

Abstract

The demand in the global trade of goods is thriving in the recent post-COVID times and is forecast to increase to the highest levels of the past decade. This calls for global transportation companies to implement increasingly efficient and highly optimized transportation processes to fulfill both demand and their need for economically and environmentally sustainable solutions. This paper presents a combined route construction algorithm (RCA) and a set-partitioning model (SPM) formulation to solve the air express transportation problem (AETP). An optimal set of routes is selected to transport a set of shipments from origin to destination whilst minimizing the total transportation time for all shipments included. The set of shipments and a set of movements is gathered from a global air express courier. Movements in this case are both ground rides and air flights with a specific origin, destination, departure time, and arrival time. The combined model finds an improvement in the historical implementation by the courier for 24.9% of the shipments for which at least one feasible route is constructed. Four alternative approaches that match the real-life strategies of an air express company are implemented. The number of shipments for which the model finds a feasible solution is increased by 9.9% by allowing late delivery of a shipment. The preference of air express companies to select historically implemented paths for specific origin-destination pairs is also implemented. Compared to the original approach, the number of shipments following a historically implemented path increases by 12.3% using this adapted formulation. A bi-objective formulation of the SPM allows air express companies to select a trade-off between the most efficient set of routes and the set of routes with the highest historical occurrence. Finally, a chance-constrained formulation is included to assume the upper bound of the movement capacity based on the strategy of an air express company in terms of the level of risk.

1 Introduction

The current energy crisis, the war in Ukraine, the COVID-19 pandemic, and many more current events influence the world trade of goods. This has a great impact on the decisions made and solutions found by transportation companies. The solutions sought by these companies are not only based on the pure profitability of their businesses but also consider both economic and environmental sustainability. Data published in the CBP World Trade Monitor of May 2022 [CPB, 2022] shows that the pandemic caused the biggest decrease in international trade on record in the first half of 2020. However, the second half of that year showed the steepest increase in world trade of goods, resulting in a higher amount of goods being traded across national borders than ever before in early 2021. Forecast by the IMF World Economic Outlook [IMF, 2022] shows that the annual growth of trade in goods will be higher in 2022 and 2023 than it has been for the past decade, whilst taking into account the influence of the war in Ukraine that has a huge impact on global trade. This increase in demand for the international trade of goods combined with the need for economic and environmental sustainability calls for transportation companies to be increasingly efficient and optimize their processes even more than before.

The current implementation of international transportation of goods by freight forwarders and air express companies is based on a sequential selection of the next movement between facilities along a route from origin to destination. Movements in this case are both ground rides and air flights with a specific origin, destination, departure time, and arrival time. This approach works well in practice but lacks a view of steps further ahead in the network. A preview of the entire network can provide viable information in the selection of movements for each of the shipments. A more cost-effective solution might be to not select the next available movement at a specific facility but rather a movement with a later departure time. [Archetti and Peirano, 2020] introduce this optimization problem as the air transportation freight forwarder service problem (ATFFSP) and develop a mixed-integer linear programming (MILP) formulation. More recent research by [Angelelli et al., 2020] uses matheuristics to find a solution to this problem using less computational time. In this work, we introduce an adapted formulation as the air express transportation problem (AETP) in which the objective is total

*Msc Student, Air Transport and Operations, Faculty of Aerospace Engineering, Delft University of Technology

transportation time minimization for a set of shipments. The problem is solved using a route construction algorithm (RCA) and a set-partitioning model (SPM) sequentially. The SPM requires an input set of routes for each shipment, which is constructed in two phases by the RCA. The first phase finds feasible routes for each origin-destination pair (od-pair) and the second phase selects a subset of these routes for each shipment in the data set. This paper proposes different linear programming (LP) formulations to solve the problem for five different real-life implementations. These implementations include a general on-time formulation, a formulation that allows penalized late delivery, a historic route prioritizing formulation, a movement capacity forecasting formulation, and a bi-objective formulation. Note that each implementation addresses a different nuance of the actual AETP as previously mentioned.

The rest of the paper is organized as follows. Section 2 provides an overview of relevant available literature. A description of the problem and assumptions is given in Section 3. Section 4 explains the methodology for both the RCA and the SPM. This section also includes an explanation of the initialization of relevant data. The results for different real-life approaches are shown and discussed in Section 5. Finally, Section 6 provides a conclusion to the work proposed in this paper and states recommendations for future work.

2 Literature Review

The main focus of available literature relevant to the AETP is freight transportation planning. The area of freight transportation planning considers vehicle routing problems (VRPs), transshipment problems (TPs), and service network design problems (SNDPs). The TP, introduced by [Orden, 1956], is considered in this paper and is an extension to the transportation problem in terms of the usage of intermediate facilities as so-called transshipment centers. According to [Andersen et al., 2009], an SNDP not only considers the definition of the service network in terms of selecting the routes for the specific service and the attributes of each service, but it also considers the determination of flows of shipments over the arcs and through the nodes of a service network. This latter decision in SNDPs is of particular interest to this research. In this paper, it is assumed that the set of routes covered by a fleet of vehicles through a network is given since it is part of the tactical planning phase for transportation companies. Therefore, determining the optimal set of these routes for a specific fleet, as is done by solving a VRP, is not considered. The ATFFSP, which is used as a foundation and adapted in the research conducted in this paper, includes elements from TPs, SNDPs, and, according to [Archetti and Peirano, 2020], freight forwarding transportation problems (FFTPs).

[Lim et al., 2005] convert the classical TP to consider many-to-many transshipment problems on a pure network and using multi-facility. According to the definition in the literature review by [Guastaroba et al., 2016], a pure network excludes direct deliveries from origin to destination and thus forces the use of at least one intermediate facility. This is not strictly necessary in the research conducted in this paper, but the authors also introduce a time window for pickup and delivery at the origin and destination respectively, which is of particular interest to this research. The authors distinguish fixed from flexible schedules which can both be provided by transportation companies. In the network, fixed schedules are present as departure and arrival times for flights for example. These fixed schedules rather than flexible schedules are considered in the research proposed in this paper. The authors distinguish in cases of multiple shipments for each origin and destination, as would be the case for clustered pickup of shipments before transport to a facility for example, and single shipment for each origin and destination in the network, as would be the case for a network with nodes for each individual customer. The focus of the research proposed in this paper is the latter case of individual single shipment - single delivery. This case is discussed by [Miao et al., 2012], in which hard and soft time windows are combined with penalties. In this case, a soft time window refers to a preferred service time interval in which the shipment is preferred to be delivered. When not delivered within the soft time interval but still in the hard time interval, a penalty value is added to the objective function. When the hard time window can also not be met and delivery is not possible within this window, a larger penalty will be added to the objective function. In this formulation, the objective is to minimize the total cost including the penalties as described. The research proposed in this paper includes hard time windows for the delivery of shipments. Late delivery, and the corresponding penalty, are included in one of the alternative approaches. [Chen et al., 2006] also consider hard time windows at both the origin and destination. The authors add capacity constraints and storage costs at the intermediate facilities to the proposed model.

Research by [Barnhart and Schneur, 1996] investigates express shipment delivery by carriers within specific time intervals. The authors propose a column generation algorithm to solve the SNDP to quasi-optimality and propose two approaches where one considers a fixed aircraft fleet and the other considers an unspecified fleet composition that is optimized as part of the problem. The first case of a fixed aircraft fleet is of particular interest in the scope of the research conducted in this paper since transportation companies will decide on this as part of their strategic planning, while the problem tackled in this research is more tactical in nature. [Barnhart and Schneur, 1996] implement the fact that not every spoke can be reached by aircraft, so also ground movement is considered, just as in the real-life case covered in this paper. The problem is formulated as binary linear programming and the column generation is implemented as a two-step heuristic. [Daeki et al., 1999] aim to model and find a solution to a large-scale SNDP with given time intervals. The authors use problem-reduction methods including column and row generation techniques to significantly reduce the problem size in terms of the number of constraints and variables. Research by [Daeki et al., 1999] shows that the implemented techniques are very effective for large-scale SNDPs which makes it interesting for the scope of the research conducted in this paper. The proposed heuristic is further improved by [Barnhart et al., 2002].

Looking into the third subproblem which has important elements that connect to the ATFFSP, research on freight forwarding transportation considers for example the decision on what available movement to select for each section in a network. Available research distinguishes in the handling of shipments by company-owned vehicles or outsourcing to so-called third-party logistics (3PL) companies. The research proposed in this paper considers a movement schedule of company-owned vehicles only since this is the main focus of large air transportation companies. According to [Archetti and Peirano, 2020], there are more problems that share characteristics with the ATFFSP such as the liner shipping problem, widely known in maritime transportation. In these problems, line-based service in terms of container shipping is offered with a fixed schedule. These problems include the operational planning problem of allocating shipments to lines or even to containers on specific lines. [Archetti and Peirano, 2020] state that, on a tactical planning level, the main problem can be formulated as a liner ship fleet deployment problem (LSFD) which covers the assignment of vehicles to specific routes. The available research on these problems tends to model this problem from SND approach.

The research proposed in this paper adapts the ATFFSP formulation by [Archetti and Peirano, 2020] to an AETP in which the objective is to minimize total transportation time for a total set of shipments whilst meeting capacity constraints on company-owned movements.

3 Problem Description

The research proposed in this paper aims to find an optimal set of routes to transport all shipments from the origin to the destination facility while minimizing the overall transportation time. A route is referred to as a sequential set of movements used to transport a shipment between facilities from the origin to the destination. The path that is selected for a shipment refers to the sequential set of facilities entered by the shipment. Further elaboration on and an example of the definition of routes and paths is given in Section 4.2. Besides origin and destination facility, shipment characteristics include both a pickup time and expected delivery date and calculated weight. Movement schedule characteristics include movement origin and destination facility, scheduled time of departure and arrival, and vehicle type. Mentioned data is first used in the route RCA to construct a set of feasible routes for each shipment. The SPM then finds the optimal set of these routes to transport all shipments. A capacity constraint is set for each movement in the SPM, in which the capacity is assumed based on historical load data of movements with an equal origin, destination, and vehicle type. Processing times at specific facilities are also considered in the RCA. These processing times are assumed based on historical data on the arrival and departure times of shipments at specific facilities.

Prior to the actual construction of routes for each shipment in the RCA, historic and alternative paths are built. Historically implemented paths for the transportation of shipments are stored for each unique origin-destination pair (od-pair). In addition to these historic paths, alternative paths are constructed based on the movements in the movement schedules. Presetting the degree of complexity to these routes, i.e. the number of facilities to include in a path from origin to destination, new paths are constructed and stored for each unique od-pair. Both the historic and alternative paths are used in the RCA to select sequential sets of movements that follow these paths as feasible routes.

Air express companies operate on a global scale and therefore generate huge shipment data sets and movement schedules. The research in this paper considers a subset of this shipment data and investigates the use case of shipments with an origin facility in Europe (EU) and a destination facility in the Americas (AM). A selection of available od-pairs within this use-case is made to include the most relevant and most occurring pairs. This selection significantly reduces the number of od-pairs to consider in the RCA, and therefore reduces the total computational time, whilst including and finding routes for a large number of shipments in the data set.

4 Methodology

The RCA is divided into two phases and is part of pre-processing for the SPM as shown in Figure 1. A number of initialization steps are implemented to construct usable input from given data sets for both the RCA and SPM. These steps are explained in Section 4.1 and include algorithms for the construction and determination of processing times, movement capacity, and paths based on historical data. As mentioned in Section 3, movement schedules provide information to extend the number of feasible paths for od-pairs. The actual implementation of

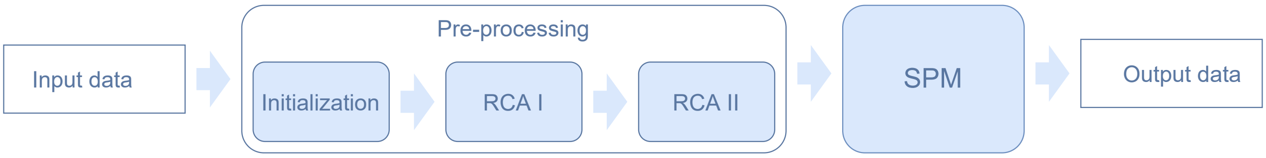


Figure 1: Overview methodology.

both phases of the RCA and the SPM are explained in Sections 4.2 and 4.3, respectively. Different approaches for the SPM that comply with different real-life strategies for air express companies are also discussed.

4.1 Initialization

This section explains the use of historical data in the initialization phase of the model. Section 4.1.1 explains the use of historical shipment data to find processing times for future shipment arrivals at specific facilities. The assumption on movement capacity of each movement in the movement schedule is done following the method explained in Section 4.1.2. Both historical and alternative paths can be used as a basis for the route construction part of the model. The alternative path-finding algorithm and the set level of complexity based on historical data are explained in Section 4.1.3.

4.1.1 Processing Times

From historical data on arrival and departure times of shipments at specific facilities, a data set is constructed of corresponding processing times that are determined as the difference between said departure and arrival time. The RCA includes movements that are loaded at the origin and unloaded at the destination. Therefore, data of shipments that have historically been transshipped at a facility is not included in this initialization of processing times. The determined processing times are stored based on the specific facility and time of arrival. The processing times used in the RCA are sampled from this data. Five different sampling techniques are compared on their root mean squared error (RMSE) for the determined processing times for each of the shipments in a test data set arriving at a facility at a specific arrival time. These techniques are based on the determination of a weighted average of the historical processing times at the facility of arrival. The weights are determined as the inverse of the difference in arrival time of the shipment compared to historical arrival times at the specific facility. Using the inverse of this value results in higher weights for historic processing times for arrival times closer to the arrival time of the shipment.

The method of rounding the arrival times impacts the performance of selected techniques. Included techniques round historic and new arrival times to *hour of the day* (A), *hour of the week* (B), *day of the week* (C), *working or off hour* (D), *week or weekend day* (E). For the *hour of the day* technique, historic processing times are stored for the arrival time rounded to the hour of the day, e.g. Wednesday 6 a.m. is stored as 6. The same arrival time for the *hour of the week* technique would be 54 ($2 * 24 + 6$). In the *day of the week* technique, this arrival time would be 3 since Wednesday is the third day of the week. The remaining techniques use a similar approach to the *day of the week* and *hour of the day* techniques, but add an extra factor to the weights to distinguish in *week or weekend day*, and in *working or off hour*, respectively. A higher factor is given to historical processing times with arrival time that has an equal classification to the shipment.

Figure 2 shows the calculation of the difference in arrival time for shipment x and historic arrival times for shipments a, b, and c using the *hour of the week* technique. The historic arrival times are assumed equal for each week that is considered. In some cases, a specific historic arrival time in the previous or next week is closer to the current arrival time than the historic arrival time in the current week. Therefore, to compute the difference in arrival time, not only times in the *current week* but also times in the *previous week*, noted by *, and *next week*, noted by **, must be considered. For each of the historic processing times, the smallest difference in arrival time is selected, e.g. $min(|x-a|, |x-a^*|, |x-a^{**}|)$. In the example shown in Figure 2, the smallest differences are between x and a (36 hours), between x and b (30 hours), and between x and c^* (72 hours). So in this case, the arrival time in the previous week is considered for historic shipment c.

The RMSE is determined for each of the techniques as the square root of the mean of the square of all of the error, with an error being the difference between a historic processing time and a sampled processing time

$$RMSE = \sqrt{\frac{\sum_{k=1}^{n}(pt_k - \bar{pt_k})^2}{n}} \quad (1)$$

where pt_k is the historic processing time at the facility for shipment k and $\bar{pt_k}$ is the sampled processing time for a shipment arriving at the facility at the same time as shipment k. The total number of historic processing times at the facility is denoted as n. Figure 3 shows the RMSE values for each of the techniques and shows the

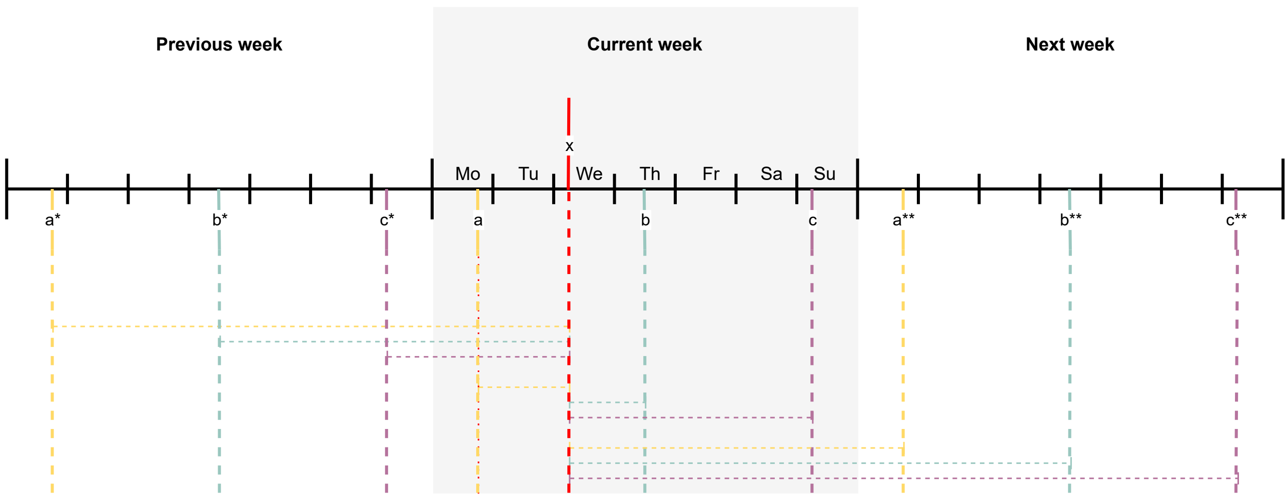


Figure 2: Arrival time difference interpretation using the *hour of the week* technique.

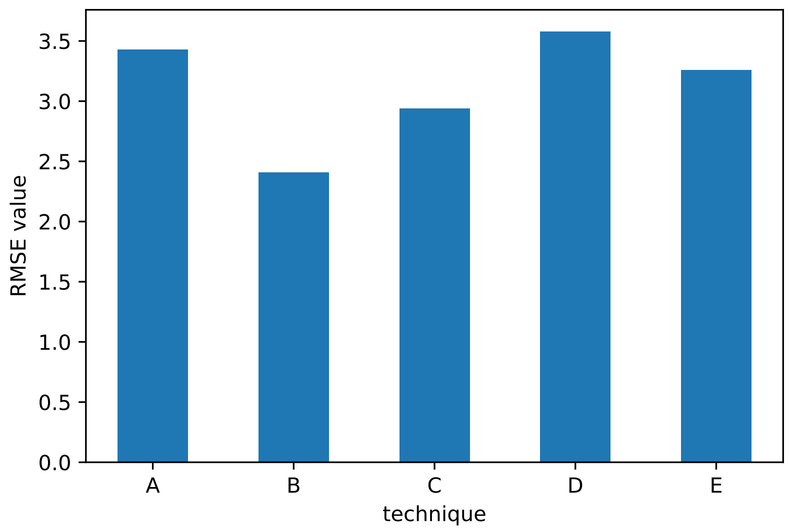


Figure 3: RMSE values for processing time sampling techniques: A) *hour of the day*, B) *hour of the week*, C) *day of the week*, D) *working or off hour*, E) *week or weekend day*.

lowest RMSE value for the *hour of the week* technique (B). This technique thus performs best and is therefore used in the initialization of the input data of the RCA.

4.1.2 Movement Capacity

The capacity of each movement is constrained in the SPM. A limit is set for the combined weight of shipments selected for each movement. This is done to prevent the model from assigning an amount of weight for a movement that exceeds the known capacity of that movement since this would result in the infeasibility of a solution in a real-life implementation. This capacity limit is based on historical data that is categorized in movement origin, movement destination, and vehicle type. An estimation of the capacity for each movement in the schedule is made based on these three characteristics. Using real-life data sets from an air express company comes with the possibility of certain data points being incomplete, inconsistent, or not present. This is taken into account in the estimation of the capacity limit for each movement. If one or more characteristics of a specific movement are unknown, an estimation is made based on the remaining characteristics. Air express companies might sell a part of the available capacity to third parties. The historical factor of the capacity that is actually used by the company is therefore considered in the estimation of capacity for movements in the schedule.

The RCA constructs routes based on the assumption that a shipment is loaded onto a movement at the origin facility of that movement and unloaded from the movement at the destination facility of that movement. Historical capacity data is based on the weight of shipments added to unit load devices (ULDs) or containers

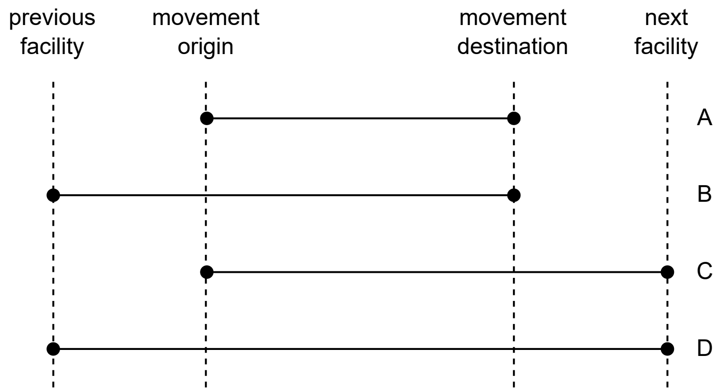


Figure 4: ULD unloading or transshipment options for a movement with specific origin and destination.

before being loaded onto a movement. These ULDs and therefore shipments can be both unloaded or transshipped at a facility. Figure 4 shows four options for a ULD when considering a specific section of a movement from an origin facility to a destination facility. Option A refers to the ULD being loaded onto the movement at the origin and unloaded at the destination. Option B shows that the ULD is loaded at a previous facility and is transshipped at the origin of the movement. Similar to option A, the ULD is unloaded at the destination of the movement. The opposite of option B is referred to by option C, where a ULD is loaded at the origin of the movement and transshipped to the next facility at the destination of the movement. The fourth and final option is transshipment at both the origin and the destination of a movement. This is shown by option D, in which a ULD is loaded at a previous facility and unloaded at the next facility. As the RCA part of the model assumes loading at the origin facility and unloading at the destination facility of a movement, the relevant option to include is option A. Based on historical data, an *available factor* (AF) can be computed for each historical movement as

$$AF = \frac{wgt_A}{wgt_{ABCD}} \tag{2}$$

where the added weight (wgt_A) of shipments with option A, i.e. relevant to the RCA, is divided by the added weight of all shipments (so including each of the options previously mentioned) on the historical movement (wgt_{ABCD}) to find the available factor. Note that wgt_{ABCD} is less than or equal to the total capacity of the historical movement. Based on the computed available factors for each historical movement, an assumption is made on the capacity of movements in the movement schedule. This assumption is based on the origin, destination, vehicle type, and scheduled departure time of a movement. The assumption is made considering the distribution of historical values with equal characteristics. An air express company can implement strategies for this assumption varying from a risk-prone strategy to a risk-averse strategy. These different strategies are considered and implemented using a chance-constrained approach discussed in Section 4.3.

4.1.3 Historic and Alternative Paths

The RCA considers historic and alternative paths for specific od-pairs and finds feasible combinations of movements from the movement schedule that follow these paths. From an air express company perspective, historically implemented paths are preferred over alternative paths. Historic paths are therefore considered prior to the alternative paths in the RCA. The historic paths are sorted in descending order of historic occurrence to ensure that the RCA considers these paths in a decreasingly relevant order.

The alternative paths are constructed using Algorithm 1. The algorithm uses the movement schedule to find feasible next movements at a facility and constructs sets of sequential facilities connected by these movements. The starting facility is the origin that is given as input to the algorithm. Sequential sets are stored as an alternative path once the latest added facility is equal to the destination given as input to the algorithm. The maximum number of facilities between the origin and destination in an alternative path is restricted. This is referred to as *level of complexity* and the actual level value is based on the complexity of historic routes. Figure 5 shows that over 90% of all historic routes are fulfilled using paths of up to level 6. Therefore, the maximum level of complexity is preset to 6 for the alternative path-finding algorithm.

4.2 Route Construction Algorithm

The RCA constructs sets of feasible routes for shipments by connecting movements in the movement schedule in sequential sets according to the previously mentioned historical and alternative paths. This is done in two

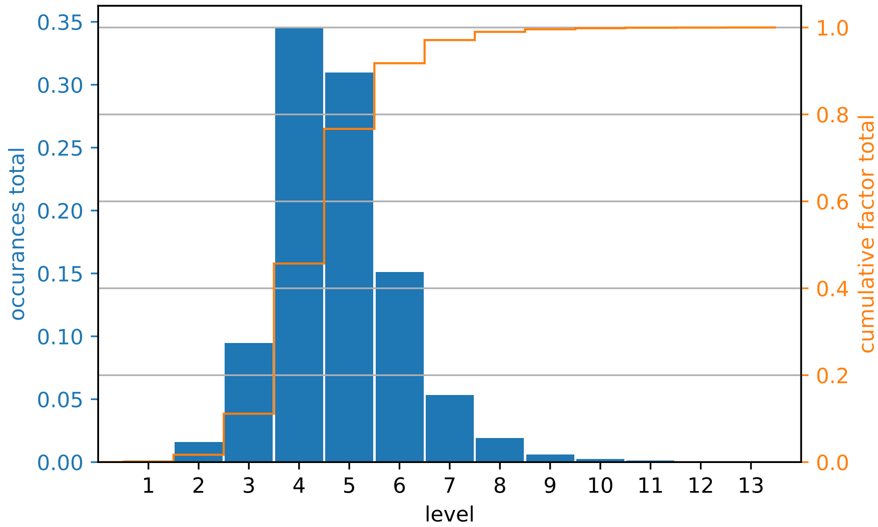


Figure 5: Histogram level of complexity for historic paths including the cumulative factor of all paths considered.

Algorithm 1 Alternative paths for specific od-pair.

Input origin: org, destination: dst, level of complexity: $level$, destinations of movements departing at specific facility: $mvmts\_fac$
Initialize $paths$ = empty list for paths for od-pairs
for $l = 0$ to $level$ **do**
  for $section = 0$ to $range(l)$ **do**
    $mvmts\_org = mvmts\_fac[org]$
    for fac_0 in $mvmts\_org$ **do**
      $mvmts\_section = mvmts\_fac[fac\_0]$
      $\ldots$
      for fac_l in $mvmts\_section$ **do**
        if $fac\_l = dst$ **then**
          $path = [org, fac\_0, \ldots, fac\_l, dst]$
          add $path$ to $paths$
        end if
      end for
    end for
  end for
end for
return paths
Output paths for specific od-pair

phases. Algorithm 2 shows the working principle of phase I of the RCA. This first phase is the construction of actual routes for unique od-pairs present in the shipment data. The total number of routes constructed is limited to reduce computational time for both the RCA and the SPM. The second phase of the RCA is shown in Algorithm 3 and selects a subset of the constructed routes for each shipment to comply with the pickup time (pu) and expected delivery date (edd) of the specific shipment.

Algorithm 2 constructs routes for specific od-pairs, so part of the input is an origin (org) and a destination (dst). The corresponding historic and alternative paths for the od-pair are considered in the input variable $paths\_od\_pair$ and initialized as $paths\_od$. To find feasible combinations of movements to construct routes, variable $mvmts\_od\_pairs$ is used in the algorithm that returns a list of all movements in the movement schedule between a specific facility pair. The processing times at specific facilities at specific arrival times are also considered. A predetermined variable $next\_mvmts$ returns a list of feasible movements departing from a specific facility after a given arrival time at that facility and said processing time. The algorithm investigates feasible combinations of movements to actually transport a shipment from origin to destination using the sequential set of facilities given by the historic and alternative paths. Figure 6 shows an example of the RCA for od-pair LHR-YVR (London Heathrow Airport to Vancouver International Airport). From historical and alternative paths, four paths with different levels of complexity are considered and used as input to the algorithm to find corresponding sets of movements. A direct path from origin to destination (LHR-YVR), two paths that

Algorithm 2 Route construction algorithm (I) for specific od-pair.

Input origin: org, destination: dst, historic and alternative paths: $paths\_od\_pair$, movements between specific facilities: $mvmts\_od\_pairs$, feasible next movements for a specific movement: $next\_mvmts$
Initialize $routes$ = empty list for routes for od-pair, $costs$ = empty list for costs of routes for od-pair
$paths\_od = paths\_od\_pair$ for specific od-pair
m = maximum length of all paths in $paths\_od$
for $n = 0$ to m **do**
  $paths\_n =$ paths in $paths\_od$ with length n
  for $path$ in $paths\_n$ **do**
    j = number of sections in $path$
    for $section$ 0 to $section$ j **do**
      movements section 0 = $mvmts\_od\_pairs$[origin section 0, destination section 0]
      for mvmt section 0 in movements section 0 **do**
        movements section 1 = $next\_mvmts$[mvmt section 0, destination section 1]
        ...
        movements section j = $next\_mvmts$[mvmt section j-1, destination section j]
        for mvmt section j in movements section j **do**
          $route$ = [mvmt section 0, ..., mvmt section j]
          $cost$ = arrival time mvmt section j - departure time mvmt section 0
          add $route$ to list of $routes$ for od-pair
          add $cost$ to list of $costs$ for od-pair
        end for
      end for
    end for
  end for
end for
return $routes$, $costs$
Output routes and corresponding costs for specific od-pair

transport through one facility between origin and destination (LHR-YHM-YVR and LHR-CVG-YVR), and a path that transports through two facilities between origin and destination (LHR-EMA-CVG-YVR). The RCA constructs five feasible combinations of movements from the movement schedule, routes I, II, III, IV, and V. Each route has a specific departure time at the origin and a specific arrival time at the destination. Algorithm 3 selects a feasible subset of options for a shipment based on the characteristics of each route.

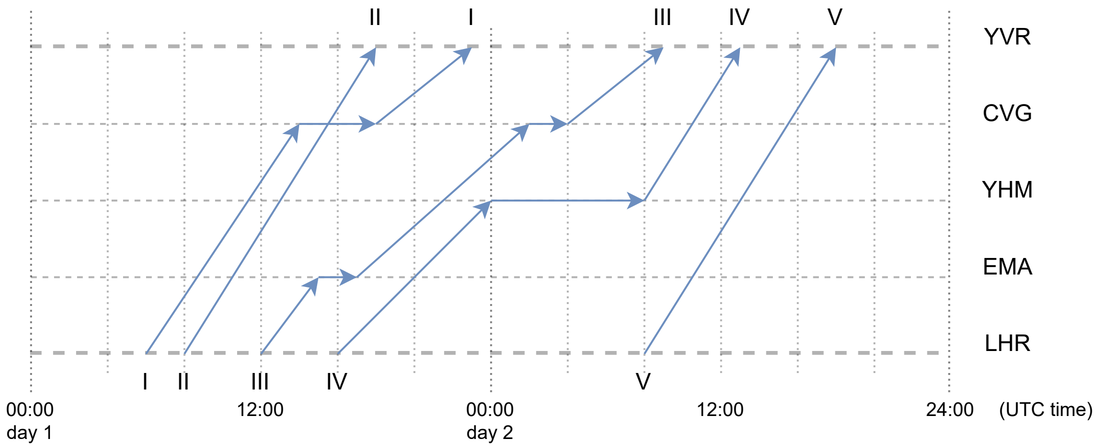


Figure 6: Constructed routes for od-pair LHR-YVR based on historical and alternative paths.

As mentioned, Algorithm 3 extends the RCA to select a feasible subset of routes for a specific shipment from the total set of routes for the specific od-pair of the shipment. The selected routes must meet two requirements. The departure time of the first movement must be later than or at the pu of the shipment and the arrival time of the final movement at the destination must be prior to or at the edd. These limits are visualized by the red lines in figure 7. Based on these requirements, the RCA selects routes II, III, and IV as feasible options for the shipment. The departure time of the first movement of route I is prior to the pu so this route is excluded. The arrival time of the final movement of route V is after the edd so this route is also excluded.

Algorithm 3 Route construction algorithm (II) for specific shipment.

Input shipment origin: *org*, shipment destination: *dst*, shipment pickup time: *pu*, shipment expected time of delivery: *edd*, routes for od-pairs: *routes* (output algorithm 1), costs for routes for od-pairs: *costs* (output algorithm 1)
Initialize shp_routes_mvmts = empty list for routes for shipment, shp_routes_costs = empty list for costs of routes for shipment
$routes\_od = routes[org, dst]$
$costs\_od = costs[org, dst]$
for *route* in $routes\_od$ **do**
  cost = corresponding cost for *route* in $costs\_od$
  mvmt 0 = first movement in *route*
  mvmt j = final movement in *route*
  if (departure time mvmt 0 $> pu$) & (arrival time mvmt $j < edd$) **then**
    add *route* to shp_routes_mvmts
    add *cost* to shp_routes_costs
  end if
end for
return shp_routes_mvmts, shp_routes_costs
Output routes and corresponding costs for specific shipment

A rule-of-thumb heuristic is implemented in this part of the RCA to reduce the solution space and thereby decrease the computational time for the SPM. The number of routes constructed by the RCA and input to the SPM is limited by implementing the *buffer method*. This method compares the cost, which is in this case the total transportation time from *pu* to the arrival time at the destination facility, of newly constructed routes to previously added routes for a specific shipment. A new route is added if and only if the cost is below or within a preset buffer time over the lowest cost of all previously added routes. Using a buffer time of 48 hours ensures a significant decrease in the size of the solution space input to the SPM whilst still being able to find the optimal set of routes to transport all shipments.

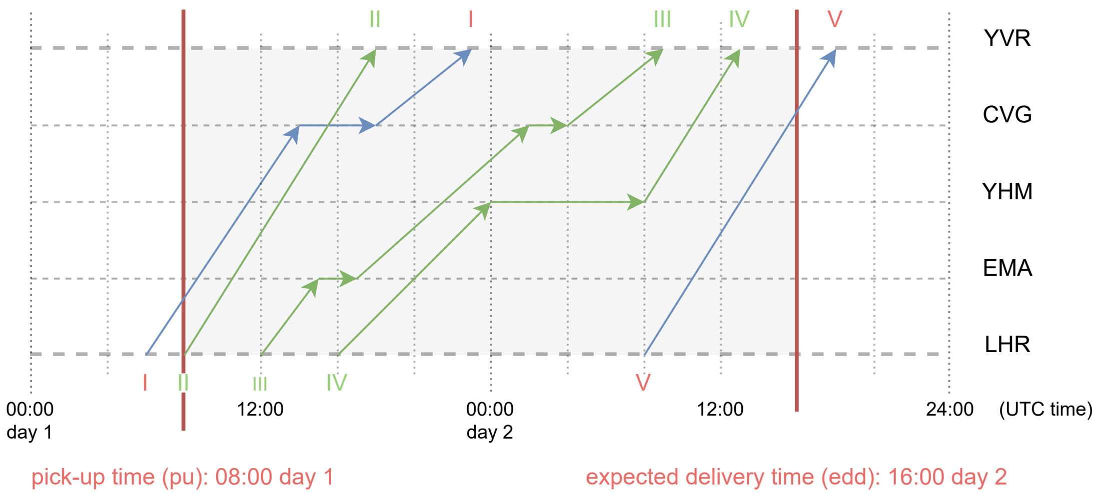


Figure 7: Selection of constructed routes for a shipment with od-pair LHR-YVR and given pick-up time and expected delivery time

4.3 Set-Partitioning Model

The goal of the SPM is to select one and only one feasible route in the solution set for each shipment, constructed using the RCA, such that the total cost (time) to transport all shipments is minimized, whilst satisfying capacity constraints for each of the movements in the used movement schedule. The output of the RCA discussed in Section 4.2 is input to the SPM which is formulated as a linear programming (LP) model. As previously mentioned, different approaches that correspond to real-life implementations are included in the model. These variants include an approach to include late delivery, an approach to prioritize historically implemented routes, a bi-objective model that allows for the combined implementation of the latter and original approach, and an approach to use different strategies in determining the available capacity of movements in the movement schedule. The original formulation of the SPM is stated and explained below. The used sets, indices, parameters, and decision variables are summarized in Table 1. As said, the objective is to minimize the total cost for all shipments.

Sets	
S	set of shipments, indexed by s
M	set of movements, indexed by m
R_s	set of routes for shipment s, indexed by r
R_{sm}	set of routes for shipment s including movement m, indexed by r
Parameters	
C_{sr}	cost of route r for shipment s
W_s	weight of shipment s
CAP_m	capacity of movement m
Decision variables	
x_{sr}	if route r is selected for shipment s

Table 1: Sets, indices, parameters, and decision variables for the original formulation of the SPM.

$$\min \sum_{s \in S} \sum_{r \in R_s} C_{sr} \cdot x_{sr} \tag{3}$$

Subject to:

$$\sum_{r \in R_s} x_{sr} = 1 \quad \forall s \in S \tag{4}$$

$$\sum_{s \in S} \sum_{r \in R_{sm}} W_s \cdot x_{sr} \leq CAP_m \quad \forall m \in M \tag{5}$$

$$x_{sr} \in \{0, 1\} \quad \forall s \in S, r \in R_s \tag{6}$$

The objective function 3 shows cost minimization as the main goal of the model. Cost C_{sr} is defined as the total transportation time for a shipment s for route r. The total transportation time is the time in hours between *pu* and arrival at the destination facility. Selected routes for each shipment get decision variable x_{sr} value 1 and the corresponding cost or time for these specific routes is added. The objective is to minimize this sum. Constraint 4 sums the decision variable x_{sr} over all feasible routes r for a shipment s. The constraint ensures that one and only one route is selected for each shipment. Constraint 5 ensures for each movement that the capacity of this movement is not exceeded. The constraint sums over all shipments $s \in S$ and for each of these shipments over all available routes $r \in R_{sm}$, the routes for the shipment that include movement m, and adds all weights W_s of the selected shipments on this specific movement m. Finally, Constraint 6 determines the domains of decision variable x_{sr}.

4.3.1 Late Delivery

The on-time formulation of both the RCA and the SPM does not allow for construction and therefore selection of routes that deliver a shipment after the expected delivery date. This might result in a situation where the first phase of the RCA does find feasible routes for a specific od-pair, but the second phase does not find a selection of these routes that is feasible for a specific shipment based on the *pu* and *edd* of this shipment. Obviously, air express companies prefer selecting a costly, late delivery of a shipment over no delivery of a shipment. To implement this real-life case in the SPM, the *late delivery* approach allows an adapted formulation of Algorithm 3 to also select routes that have an arrival time at the destination after the *edd* of a shipment. These routes must be penalized in the objective function to force the SPM to prioritize on-time delivery options if available. This penalty value is initialized using the output of the RCA. The penalty is set equal to the highest cost for all available routes of all shipments as determined by the RCA. The value is added to the cost for selected late delivery options using parameter CP_{sr}, which is the penalty cost of a route r for a shipment s. Again, the cost is defined as the transportation time between *pu* and arrival at the destination facility. This cost is multiplied by a cost multiplication factor equal to the highest known cost in the RCA part of the model to construct the penalty cost CP_{sr}. The objective function in the SPM is adapted to

$$\min \sum_{s \in S} \sum_{r \in R_s} C_{sr} \cdot x_{sr} + CP_{sr} \cdot x_{sr} \tag{7}$$

where parameter CP_{sr} has a cost penalty value if a route r for a shipment s has an arrival time after *edd* and therefore delivers the shipment late. Value 0 is given if route r delivers the shipment early. This results in the addition of the penalty value to the objective function if and only if a route is selected that delivers a shipment late.

4.3.2 Likeliness

An air express company prefers the selection of historically implemented paths rather than newly constructed, alternative paths. As mentioned in Section 4.1.3, historical paths are considered prior to alternative paths in the RCA. A maximum is set in the RCA to the total number of routes constructed for a unique od-pair as a rule-of-thumb heuristic to reduce computational time for both the RCA and the SPM. Considering routes from historic paths prior to routes from alternative paths in the RCA ensures the inclusion of these preferred routes in the solution space input to the SPM. To further fulfill the preference of air express companies to select routes based on historic paths, a *likeliness* approach is introduced. This approach adapts the cost of each route to force the model to prioritize routes with the highest historical count, meaning that routes that are constructed based on historically important and more frequently implemented paths are prioritized in the model. The adaption of the cost for each route is therefore based on the historical occurrence or count of the corresponding path. First, the count of each route is determined and stored as the factor of the total added count of all routes in the model, so for route i

$$factor_i = \frac{count\ route_i}{\sum_{i=1}^{n} \ count\ route_i} \tag{8}$$

A cost multiplication factor is determined for each route by normalizing and inverting these factors. Taking the inverse of each value ensures that a route r with a higher historical occurrence will get a lower multiplication factor and will therefore be prioritized in the SPM. The adapted cost values are used by introducing cost parameter CF_{sr} for a shipment s for route r, which is equal to cost C_{sr} multiplied by the mentioned cost multiplication factor. The objective function of the SPM is adapted in this approach to

$$\min \sum_{s \in S} \sum_{s \in R_s} CF_{sr} \cdot x_{sr} \tag{9}$$

The definition of the discussed likeliness approach is extended by including late delivery into a secondary *likeliness late delivery* approach. This approach allows for penalized late delivery whilst also including the mentioned multiplication factor. This is done by introducing adapted cost variable CFP_{sr}, which both penalizes late delivery and prioritizes higher historical occurrence. The objective function of the SPM is adapted in this approach to

$$\min \sum_{s \in S} \sum_{r \in R_s} CF_{sr} \cdot x_{sr} + CFP_{sr} \cdot x_{sr} \tag{10}$$

4.3.3 Bi-objective

The likeliness approach discussed in Section 4.3.2 adapts the original SPM objective to prioritize the most likely routes, i.e. routes constructed from paths with the historically highest occurrence. This is preferred by air express companies to fulfill the demand for shipment transportation by using known routes, but this objective also imposes a disadvantage compared to the original formulation. In the original formulation, the SPM is more likely to select newly found routes constructed from alternative paths. These routes can potentially be more efficient compared to routes constructed from historical paths. This potential of transporting shipments from origin to destination in less amount of time and therefore lower total cost can be of significant importance to an air express company. A *bi-objective* approach is introduced to investigate this potential and to allow for different strategies in a trade-off between the selection of the most efficient set of routes and the most likely set of routes. This is done by introducing a weight factor γ that takes values between zero and one. The objective function of the SPM is adapted to a combination of the objective functions in Equations 3 and 9, including mentioned weight factor γ

$$\min \sum_{s \in S} \sum_{r \in R_s} \gamma \cdot C_{sr} \cdot x_{sr} + (1 - \gamma) \cdot CF_{sr} \cdot x_{sr} \tag{11}$$

where the γ-value defines the weight set to either part of the objective function. Higher gamma values correspond to a higher preference for routes constructed using the original formulation and lower values correspond to a preference for likely routes. Solving the likeliness approach formulation of the SPM for a variety of γ-values between zero and one, allows an air express company to select the preferred strategy in terms of the trade-off between newly found alternative paths and paths with a historically high occurrence.

As previously done for the likeliness approach, a combination of the bi-objective approach and the late delivery approach is included in the model. This combination allows an air express company to select a preferred strategy in terms of path selection trade-off, whilst including late delivery options in the solution space. The objective function in Equation 11 is adapted in this approach to a combination of the objectives functions in Equations 7 and 10, including weight factor γ

$$\min \sum_{s \in S} \sum_{r \in R_s} \gamma \cdot (C_{sr} \cdot x_{sr} + CP_{sr} \cdot x_{sr}) + (1 - \gamma) \cdot (CF_{sr} \cdot x_{sr} + CFP_{sr} \cdot x_{sr}) \tag{12}$$

4.3.4 Chance Constraints

The assumed capacity of each movement in the movement schedule is an important input variable to the SPM. The original formulation of the SPM assumes the total available capacity of a movement as a limit for the added weight of all selected shipments on that movement. As mentioned in Section 4.1.2, the model assumes the loading of a shipment on the movement at the movement origin and unloading at the movement destination. An actually available factor, AF, for each historical movement can be determined by dividing the corresponding capacity with these characteristics by the total capacity of the movement, as shown in Equation 2. Based on historical data of this AF of movements with equal characteristics on the origin, destination, vehicle type, and scheduled departure time, an assumption can be made for the AF and therefore available capacity of movements in the movement schedule. This assumption influences the feasibility of real-life implementation by air express companies. E.g. if the assumed available capacity exceeds the actual available capacity of a movement, the added weights of all shipments selected on that movement by the SPM can exceed the actual available capacity, therefore resulting in an infeasible solution in real life.

An air express company can implement different strategies on the assumption of the AF, varying from a risk-prone to a risk-averse strategy. More risk-averse strategies assume lower AFs on average. This results in a higher chance of actual feasible solutions but also decreases the average load factor of the included movements. This trade-off is implemented in the model by introducing *chance constraints* for the actual capacity of each movement. These constraints allow the model to assume the limits set to the capacity of each movement based on a preset level of risk. The general formulation of a chance constraint, described by [Prekopa, 2003], is

$$\mathbb{P}(h(x, \xi) \geq 0) \geq p \tag{13}$$

where $h(x, \xi)$ states the constraints as a system of inequalities and p is the set probability that takes a value between zero and one. This general formulation is adapted and used as chance constraints for the capacity constraints in the SPM. The actual capacity of movements is assumed to follow a normal distribution. For each combination of mentioned characteristics on movement origin, destination, vehicle type, and scheduled departure time, a mean value (μ) and a standard deviation (σ) of the available capacity can be computed. The formulation of the chance constraints uses the percent point function (ppf) formulation $\Phi^{-1}(\alpha)$, which is the inverse of the cumulative distribution function $\Phi(\alpha)$, to find a corresponding normal distribution constant. In this formulation, α is the input variable that states the probability of guarantee and corresponds to the level of risk taken by an air express company. The capacity constraints in the original formulation of the SPM are adapted to

$$\sum_{s \in S} \sum_{r \in R_{sm}} W_s \cdot x_{sr} \leq CAP_m^{\alpha} \quad \forall m \in M \tag{14}$$

where the assumed available capacity of movement m for risk level α is given by the adapted capacity parameter CAP_m^{α}. This parameter is calculated as $CAP_m^{\alpha} = \mu_m + \Phi^{-1}(\alpha) \cdot \sigma_m$. In this formulation, μ_m and σ_m are respectively the mean value and standard deviation of historically available capacities for movements with equal characteristics to movement m. A lower value for α corresponds to lower risk and thus provides the model with a lower available capacity in given constraints. Higher α values correspond to higher risk and therefore higher available capacity. Since the capacity constraint defines a maximum value or upper limit, $\Phi^{-1}(\alpha) \cdot \sigma_m$ is added to μ_m. Important to notice is that $\Phi^{-1}(\alpha)$ takes negative values for $\alpha < 0.5$. A factor, defined by value $\Phi^{-1}(\alpha) \cdot \sigma_m$, is subtracted from the mean value for these values of α, resulting in assumed available capacity values below the mean value for a movement.

As previously done by combining the likeliness or bi-objective approach and the late delivery approach, mentioned in Section 4.3.2, the formulation of the chance-constrained approach can also be adapted to allow for penalized late delivery options. The objective function for this approach is in accordance with the objective function for the original late delivery approach in Equation 7. The capacity constraints for this combined approach are in accordance with the chance constraints shown in Equation 14.

5 Results

The research in this paper uses historical input data that is gathered from an air express company. The RCA and the SPM described in Section 4 and the mentioned different approaches are tested using this input data. Different approaches are compared and results are compared to the actual historical implementation by the air express company. As mentioned in Section 3, air express companies operate on a global scale and generate huge shipment and movement schedule data sets. A subset of this data is selected as the use case for the research in this paper based on the origin, destination, and time of pickup of the shipments. As mentioned, shipments with an origin in Europe (EU) and a destination in the Americas (AM) are the subjects of the used data sets. Also, a time interval of two weeks has been selected for the pickup time of included shipments. Data on shipments and movements in the first week has been used to initialize specific input variables for the RCA and SPM, such as historic paths and processing times. Shipment data for the second week has been used to select relevant od-pairs for the RCA and actual shipments for the SPM. Movement data from the second week has been used for finding alternative paths and for the actual construction of routes in the RCA. Splitting the data set into two separate weeks as explained, is done to touch upon the real-life implementation of the combined approach. Air express companies can use historic data prior to the current moment in time for the construction of historic paths and historic processing times. Future movement schedules and scheduled shipments with specific origins and destinations allow an air express company to construct feasible routes using the RCA and solve for an optimal set of routes to transport all shipments using the SPM. The research in this paper also uses known arrival times of shipments in the data set of the second week to compare to the arrival times in the solution of the model and thereby test the performance of the combined model.

The specific sizes and characteristics of the input data are elaborated on in Section 5.1. The models are tested on a MacBook Pro (Big Sur 11.2.3) which equips with a 2.3 GHz Dual-Core Intel Core i5 processor and 8GB 2133 MHz LPDDR3 memory. The models are implemented in Python 3.8.1 with pulp 2.6.0. Results of the original implementation of the combined RCA and SPM are discussed in Section 5.2. Section 5.3 shows the results of and comparisons between the different approaches discussed in Section 4.3.

5.1 Data Set

As mentioned, shipment data with origin in Europe and destination in the Americas is considered in this research. With a time interval for pick-up times of the shipments set to two weeks, a large data set of unique shipments is identified. Mentioned characteristics included are the origin, destination, time of pick-up, expected delivery date (edd), and the calculated weight of the shipment (wgt) in kilograms. The data set is gathered from real-life data from an air express courier and can therefore be incomplete and lacking the mentioned relevant attributes. The relevant shipments are filtered in the initialization part of the model. To test the performance of the combined RCA and SPM approach, the results have to be compared to the actual historical implementation by the air express company. A comparison can be made between the arrival times at the destination of a shipment for both the model and historical implementation. An important aspect of the available historical data is therefore that the last section of a route of a shipment must be known. If the final movement and therefore arrival time at the destination is known, a comparison can be made between the model solution and the historical implementation. Selecting this subset of mentioned relevant data shows that around 50% of the gathered data set of unique shipments are input to the model. The mentioned incompleteness of the data set influences the performance of the model. Efficient historical routes might not be available in the RCA part of the model which causes the model to select alternative and potentially more costly options.

The construction of routes for each available od-pair is done in the first phase of the RCA, according to algorithm 2. The od-pairs chosen as input to this algorithm are the od-pairs present in the mentioned reduced relevant shipment data. Based on the historical demand for an air express company, the importance or relevance of specific od-pairs can be derived. This relevance is shown in historical shipment data by the corresponding historical occurrence for od-pairs. Higher historical occurrence indicates higher relevance. A subset of the shipment data can be chosen in which the most relevant od-pairs are considered. The goal is to reduce the total amount of od-pairs input to the first part of the RCA whilst considering a significant fraction of the relevant shipment data set. Reducing the total amount of od-pairs results in reducing computational time for the RCA and therefore the combined RCA and SPM. Sorting od-pairs on historical occurrence in descending order, the blue line in Figure 8 shows the fraction of all relevant shipments included for the number of unique od-pairs included. This figure clearly shows that a relatively small subset of od-pairs can be selected whilst including, and thus allowing the model to find feasible routes for, a significant subset of the shipments included in the relevant data. A threshold value can be set for the minimal factor of shipments that has to be added to the subset when adding a new od-pair. This threshold value is shown by the orange line in Figure 8.

Considering a set of threshold values between $1.0 \cdot 10^{-8}$ and $5.0 \cdot 10^{-2}$, Figure 9 shows the corresponding computational times for the different phases of the model. The average computational time per shipment decreases for higher threshold values and thus fewer od-pairs and shipments are included for both the initialization and

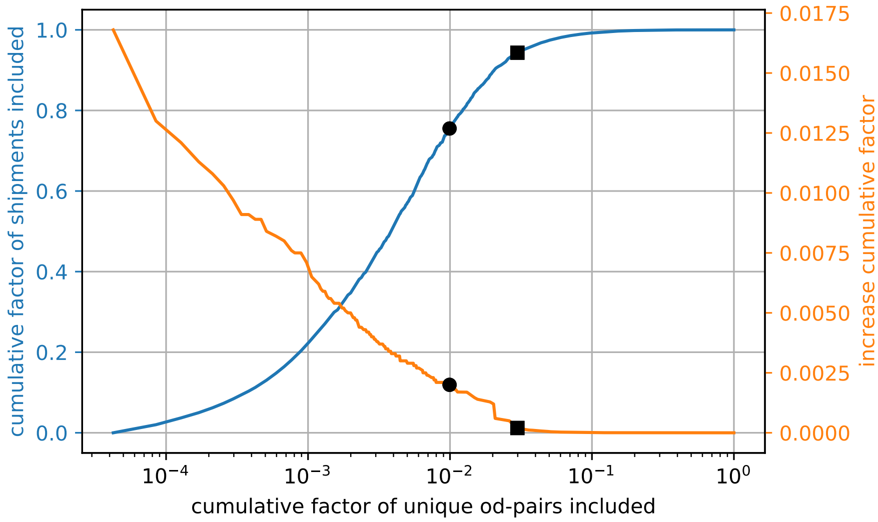


Figure 8: Cumulative factor of shipments included for the cumulative factor of unique od-pairs included. Note that od-pairs are considered in decreasing order of the number of shipments per od-pair. The black square and circle represent the relevant threshold values discussed in Section 5.1.

RCA I phase of the model. The average computational time per shipment increases for the RCA II and SPM phases of the model. The red line representing the added average computational time of each phase shows two threshold values for which a relatively higher factor of od-pairs and thus shipments results in a lower average computational time compared to a lower factor of od-pairs (or higher threshold values). These two threshold values are 0.0002, represented as a black square, and 0.002, represented as a black circle. Figure 8 shows the corresponding cumulative factor of od-pairs and factor of shipments for these threshold values. The figure shows that a threshold value of 0.0002 (black square) results in a subset that includes the top 3% most relevant od-pairs which accounts for just over 94% of all shipments. A threshold value of 0.002 (black circle) results in a subset that includes the top 1% most relevant od-pairs which accounts for just over 75% of all shipments. The latter threshold value is set for selecting the relevant subset since this value corresponds to the lowest average computational time of both threshold values. The relevant subset of shipments is used as input for the combined RCA and SPM approach. The results are discussed in Sections 5.2 and 5.3.

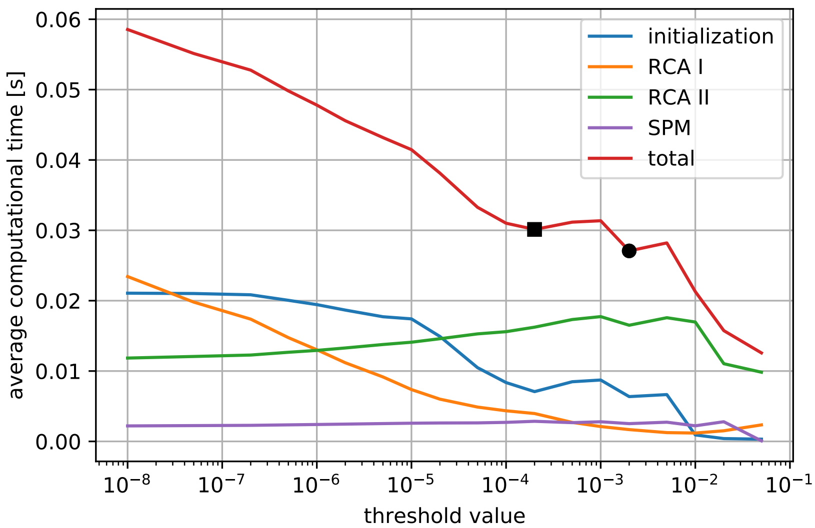


Figure 9: Average computational time per shipment for different phases of the model. The black square and circle represent two relevant threshold values elaborated on in Section 5.1.

5.2 Original Implementation

This section discusses the results of the original implementation of the combined RCA and SPM approach. Results of the original on-time delivery implementation are shown and discussed in Section 5.2.1. Section 5.2.2 includes the results for the additional late delivery implementation, mentioned in Section 4.3.1.

5.2.1 On-Time Delivery

Table 2 shows the results of the original implementation of the combined RCA and SPM approach. The table shows the values of the used key performance indicators (KPIs) that reflect the results of the model. An elaboration is given below on the meaning and calculation of each KPI and the given values are discussed.

	KPI	**Value**	**Description**
1	feasible shipments	50.3%	of input set of shipments
2	infeasibility paths	80.8%	of 49.7% infeasible shipments
3	infeasibility routes	0%	of 19.2% feasible paths
4	infeasibility time interval	100%	of 19.2% feasible paths
5	average computational time (s)	0.093	per shipment
6	average cost (h)	60.506	per shipment
7	average historic cost (h)	43.601	per shipment
8	average cost subset x (h)	59.817	per shipment
9	average historic cost subset x (h)	43.069	per shipment
10	cost decrease	24.9%	of feasible shipments
11	equal cost	3.5%	of feasible shipments
12	cost increase	71.6%	of feasible shipments
13	unavailability path	84.6%	of 71.6% shipments with cost increase
14	unavailability movement	6.0%	of 15.4% shipments with available path
15	unavailability time interval	0%	of 94.0% shipments with available movement
16	unavailability processing times	100%	of 94.0% shipments with available movement

Table 2: Results original combined implementation of the RCA and SPM.

1. Percentage of shipments for which at least one feasible route is constructed in the RCA. KPIs 2-4 map the reasons for infeasibility. Further notation: feasible shipments.
2. Percentage of remaining shipments for which no feasible path is available for the corresponding od-pair.
3. Percentage of remaining shipments for which no feasible routes are constructed based on available paths for corresponding od-pair.
4. Percentage of remaining shipments for which no feasible route is available with both departure after the pick-up time and arrival before the expected delivery time.
5. Average computational time for all shipments in the relevant input data set, including both feasible and infeasible shipments.
6. Average transportation time (in the solution) for all feasible shipments.
7. Average historical transportation time for all shipments for which the historical arrival time at the destination is known.
8. Average transportation time (in the solution) for all shipments in subset x, where subset x is the subset of feasible shipments for which the historical arrival time at the destination is known.
9. Average historical transportation time for all shipments in subset x.
10. Percentage of feasible shipments in the relevant input data set with lower transportation time compared to the historical implementation.
11. Percentage of feasible shipments in the relevant input data set with equal transportation time compared to the historical implementation.
12. Percentage of feasible shipments in the relevant input data set with larger transportation time compared to the historical implementation. KPIs 13-16 map the reasons for the selection of a more costly route.
13. Percentage of shipments for which the historical implementation is not chosen by the model due to the unavailability of the historical path.

14. Percentage of shipments for which the historical implementation is not chosen by the model due to the unavailability of movements to construct a route that follows the historical path.
15. Percentage of shipments for which the historical implementation is not chosen by the model due to the unavailability of a route that follows the historical path within the given time interval.
16. Percentage of shipments for which a historical path following but more costly implementation is chosen by the model due to larger assumed processing times and therefore selection of movements with departure time after the departure time of historical movements.

The *feasibility* KPIs 1 to 4 are summarized in Figure 10. The figure shows that the model finds a feasible solution for 50.3% of the shipments (KPI 1). For 40.2% ($49.7 \cdot 0.808$) of the shipments, no feasible route is found because of the absence of feasible paths for the specific od-pair (KPI 2). Due to no available routes in the pick-up and delivery time window, the model is unable to find a feasible route for the remaining 9.5% of shipments (KPI 4). Much performance and an increase in the number of feasible shipments can be achieved by using a more complete data set. More complete data on historical shipments will improve the construction of historic paths and more complete data on movement schedules will improve both the construction of alternative paths and the construction of routes in the RCA.

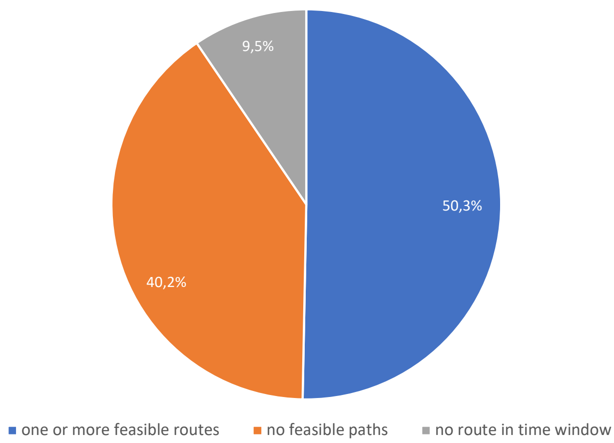


Figure 10: Summary results original approach, based on *feasibility* KPIs 1 to 4 in Table 2.

The average computational time per shipment is 0.093 seconds for the original implementation (KPI 5). KPIs 6 to 9 consider the average cost, which is the transportation time between pick-up and delivery, for specific subsets of shipments. The average cost is 60 hours and 30 minutes in the solution (KPI 6) and 43 hours and 36 minutes in the historical implementation (KPI 7). Considering subset x, shipments with both a feasible solution and a known historical transportation time, the average cost is 59 hours and 49 minutes in the solution (KPI 8) and 43 hours and 4 minutes in the historical implementation (KPI 9). The higher average cost for the solution is mainly caused by the objective of the model to select a feasible route independent of the availability or feasibility of the historically implemented route. The causes are considered in more detail in KPIs 10 to 16.

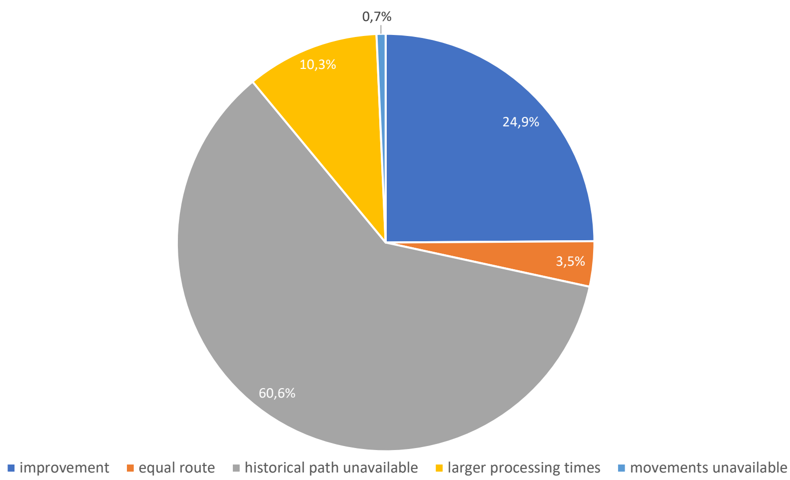


Figure 11: Summary results original approach, based on *improvement availability* KPIs 10 to 16 in Table 2.

The *improvement availability* KPIs 10 to 16 are summarized in Figure 11. The figure shows that the model finds a solution with lower transportation time and thus an improvement on the historical implementation by the air express company for 24.9% of the shipments in the relevant subset (KPI 10). Equal route and thus equal transportation time is selected for 3.5% of the shipments (KPI 11). For 60.6% ($71.6 \cdot 0.846$) of the shipments, the historical implementation is not selected due to the unavailability of the historical path for the first part of the RCA (KPI 13). Of the remaining 11.0% of the shipments, the historic implementation is not selected for 0.7% ($11.0 \cdot 0.06$) due to the unavailability of movements to construct a route that follows the historic path (KPI 14). The solution of the final remaining 10.3% of the shipments does include routes that follow the historical path, but follow a route with larger transportation time than the historical implementation due to longer processing times at facilities (KPI 16). This is caused by the risk-averse strategy explained in Section 4.1.1. The overview in Figure 11 shows that data incompleteness in terms of availability is the main contributor to the selection of more costly routes. Therefore, much performance can be achieved by improving data completeness.

Figure 12 shows the difference in transportation times for shipments between the found solution and the historical implementation. A $\Delta-$value ($\Delta = C_{solution} - C_{historical}$, where C refers to the cost or transportation time of a shipment) is determined for each shipment in the previously mentioned subset x for which both a historical implementation and a solution is known. The orange and blue areas in Figure 12 represent respectively the cost decrease and cost increase of the model solution compared to the historical implementation. An important notice is the difference in distribution between the negative values for cost decrease in Figure 13(a) and positive values for the cost increase in Figure 13(b). The majority of the values of cost decrease is roughly between -24 and 0 hours, meaning that the improvement on the historical implementation is mostly within one day. Note that the outliers show significant improvement for a small number of shipments. The increase in cost in Figure 13(b) shows peaks for $\Delta-$values around 24, 36, and 48 hours. For the corresponding shipments, the model has selected a final movement with an arrival time of respectively one, one and a half, or two days after the arrival time of the historically used movement.

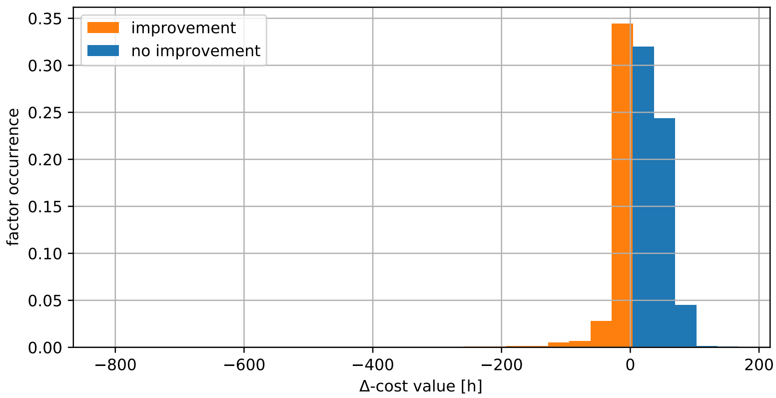
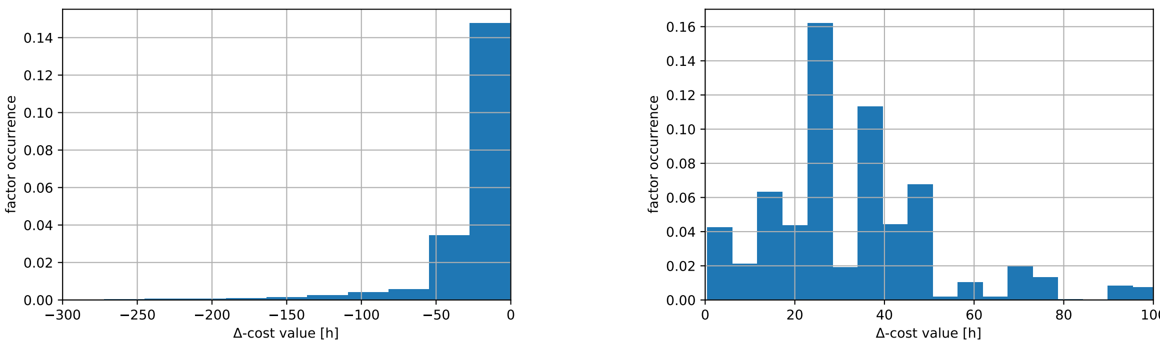


Figure 12: Difference in transportation time between found solution and historical implementation.



((a)) Negative $\Delta-$values.

((b)) Positive $\Delta-$values.

Figure 13: Extension of Figure 12, a represents cost decrease and b represents cost increase.

To summarize the mentioned results in this section, the original on-time delivery implementation of the combined model can be used by an air express company to improve delivery time for around a quarter of the shipments for which a feasible route is constructed. A feasible route is constructed for around half of the shipments in the input data subset. Whilst finding improvement for this amount of shipments for an air express company operating on a global scale is an achievement, much performance can be achieved by both improving on the completeness of the used input data and improving the assumption of processing times at facilities. Improving the accuracy of the algorithm that assumes the processing times at facilities results in a less risk-averse strategy in this assumption and therefore results in lower processing times on average. Both improvements will not only increase the number of shipments for which a feasible route is constructed but also provide potentially more efficient routes and thereby increasing the number of shipments for which improvement on the historical implementation is found.

5.2.2 Late Delivery

Similar KPIs as used and explained in the previous section for the results of the on-time delivery implementation are used to show the results of the late delivery implementation that is explained in Section 4.3.1. These KPIs and corresponding values are shown in Table 3 and the results are discussed below.

	KPI	**Value**	**Description**
1	feasible shipments	55,3%	of input set of shipments
2	infeasibility paths	80.5%	of 44.7% infeasible shipments
3	infeasibility routes	0%	of 19.5% feasible paths
4	infeasibility pickup time	100%	of 19.5% feasible paths
5	average computational time (s)	0.199	per shipment
6	average cost (h)	66.778	per shipment
7	average historic cost (h)	43.601	per shipment
8	average cost subset x (h)	66.163	per shipment
9	average historic cost subset x (h)	43.541	per shipment
10	cost decrease	24.6%	of feasible shipments
11	equal cost	3.2%	of feasible shipments
12	cost increase	72.2%	of feasible shipments
13	unavailability path	85.1%	of 72.2% shipments with cost increase
14	unavailability movement	5.8%	of 14.9% shipments with available path
15	unavailability pickup time	0%	of 94.2% shipments with available movement
16	unavailability processing times	100%	of 94.2% shipments with available movement

Table 3: Results late delivery included combined implementation of the RCA and SPM.

Summarizing the feasibility KPIs, the model finds a solution for 55.3% of the shipments (KPI 1), which is a 9.9% increase compared to the original on-time implementation. For 36.0% ($44.7 \cdot 0.805$) of the shipments, no feasible route is found because of the absence of feasible paths for the specific od-pair (KPI 2). Caused by no available routes departing after the specified pick-up time, the model is unable to find a feasible route for the remaining 8.7% of shipments (KPI 4).

	KPI	**Value**[1]	**Value LD**[2]	**Difference**
1	feasible shipments	50.3%	55.3%	*+9.9%*
5	average computational time (s)	0.093	0.199	*+114.0%*
6	average cost (h)	60.506	66.778	*+10.4%*
8	average cost subset x (h)	59.817	66.163	*+10.6%*

Table 4: Comparison on-time[1] and the late delivery[2] implementation for selected KPIs from Tables 2 and 3.

The average computational time per shipment is 0.199 seconds for the late delivery implementation (KPI 5), an increase of 114.0% compared to the on-time implementation. The average cost is 66 hours and 47 minutes in the solution (KPI 6), which is an increase of 10.4% compared to the original implementation, and 43 hours and 36 minutes in the historical implementation (KPI 7), which is equal to the original implementation since the same set of shipments is considered. Considering subset x, the average cost is 66 hours and 10 minutes in the solution (KPI 8) and 43 hours and 32 minutes in the historical implementation (KPI 9), an increase of respectively 10.6% and 1.1% compared to the original implementation.

Summarizing the improvement availability KPIs 10 to 16, the combined RCA and SPM model that allows penalized late delivery of a shipment finds a solution with lower transportation time and thus an improvement on the historical implementation for 24.6% of the shipments in the relevant subset (KPI 10). Equal route and thus equal transportation time is selected for 3.2% of the shipments (KPI 11). For 61.4% ($72.2 \cdot 0.851$) of the shipments, the historical implementation is not selected due to the unavailability of the historical path for the first part of the RCA (KPI 13). For the remaining 10.8% of the shipments, the historic implementation is not selected for 0.6% ($10.8 \cdot 0.058$) due to the unavailability of movements to construct a route that follows the historic path (KPI 14). The solution of the final remaining 10.2% of shipments does include at least one route that follows the historical path, but those routes have a larger total transportation time than the historical implementation caused by higher processing times at facilities (KPI 16).

To summarize the mentioned results in this section, the late delivery implementation of the combined model can be used by an air express company to improve delivery time for around a quarter of the shipments for which a feasible route is constructed. This is a similar result to the on-time delivery implementation discussed in section 5.2.1. A feasible route is constructed for over half of the shipments in the input data subset, around a 10% increase compared to the on-time implementation. Note that this increase in the number of shipments for which a feasible route is constructed goes with an increase in average transportation time of also around 10%. Yet, considering the real-life implementation of the model by an air express company, the preference is to find a feasible route for the highest amount of shipments, independent of the corresponding increased average cost. An air express company will therefore select the late delivery rather than the on-time delivery implementation of the model in case of real-life implementation. Again, finding improvement for around a quarter of the feasible shipments for an air express company operating on a global scale is an achievement, but much performance can be achieved by improving both the completeness of the used data and the assumption of processing times at facilities. Both improvements will not only increase the number of shipments for which a feasible route is constructed but also provide potentially more efficient routes and thereby increasing the number of shipments for which improvement on the historical implementation is found.

5.3 Approaches

Section 4.3 explains different approaches to the combined RCA and SPM that comply with the real-life strategies of an air express company. The results of these approaches are discussed in this section. The results of the *likeliness* approach, the *bi-objective* approach, and the *chance-constrained* approach are discussed in Sections 5.3.1, 5.3.2, and 5.3.3 respectively. For each of these approaches, both an on-time and a late delivery implementation are included and discussed. A comparison is made between these results and the results for the original approaches discussed in the previous section.

5.3.1 Likeliness Approach

In the likeliness approach, a cost multiplication factor is included in the objective function to force the model to select routes based on historical paths rather than alternative paths. The results of the model in which this likeliness formulation is implemented are shown in Table 5. Note that results for both the on-time delivery and the late delivery implementation are provided. An explanation of each KPI is given below.

	KPI	**Value**[1]	**Value LD**[2]	**Description**	**Difference**
1	feasible shipments	50.3%	55.3%	of input set of shipments	*+9.9%*
2	average computational time (s)	0.108	0.231	per shipment	*+113.9%*
3	average cost (h)	61.084	67.311	per shipment	*+10.2%*
4	cost decrease	23.4%	23.2%	of feasible shipments	*-0.9%*
5	likeliness factor original[3]	0.851	0.872	of feasible shipments	*+2.5%*
6	likeliness factor	0.956	0.958	of feasible shipments	*+0.2%*

Table 5: Results likeliness approach implementation of the combined RCA and SPM. [1] value for the on-time delivery likeliness approach. [2] value for the late delivery likeliness approach. [3] KPI for the likeliness factor of the original implementation of the model, discussed in Section 5.2.

1. Percentage of shipments for which at least one feasible route is constructed in the RCA. Further notation: feasible shipments.
2. Average computational time for all shipments in the relevant input data set, so both feasible and infeasible shipments.
3. Average transportation time for all feasible shipments.
4. Percentage of feasible shipments in the relevant input data set with lower transportation time compared to the historical implementation.

5. The factor of feasible shipments for which the model has selected the route with the highest historical occurrence in the original implementation.
6. The factor of feasible shipments for which the model has selected the route with the highest historical occurrence in the likeliness implementation.

The percentages of feasible shipments (KPI 1) are in line with the values given in Sections 5.2.1 and 5.2.2 since this is determined by the RCA which is equal for both the original and the likeliness implementation. The average computational time (KPI 2) is 0.108 seconds and 0.231 seconds for the on-time and late delivery implementation of the likeliness approach, respectively. The average cost (KPI 3) is 61 hours and 5 minutes for the on-time delivery implementation of the likeliness approach, which is an increase of 1.0% compared to the original approach (KPI 6, Table 2), and 67 hours and 19 minutes for the late delivery implementation of the likeliness approach, which is an increase of 0.8% compared to the original approach (KPI 6, Table 3). These increases in cost are caused by the adapted objective of selecting routes with higher historical occurrence rather than routes with the lowest cost. An improvement is found (KPI 4) for 23.4% of the feasible shipments in the on-time delivery implementation of the likeliness approach, which is a decrease of 6.0% compared to the original approach (KPI 10, Table 2), and for 23.3% of the feasible shipments in the late delivery implementation of the likeliness approach, which is a decrease of 5.7% compared to the original approach (KPI 10, Table 3). The decreased level of improvement is also caused by the adapted objective of selecting routes with higher historical occurrence rather than routes with the lowest cost.

The likeliness factor for the original approach (KPI 5) of the model is 0.851 and 0.872 for the on-time and late delivery implementation, respectively, indicating the selection of the highest occurring historical path for 85.1% and 87.2% of the feasible shipments. These relatively high factors are the result of the underlying nature of the prioritization of routes constructed from historical paths by the RCA. These routes, therefore, have a higher chance of being included in the solution space that is input to the SPM compared to the routes constructed from alternative paths. The likeliness factor increases by 12.3% to 0.956 for the on-time delivery implementation of the likeliness approach (KPI 6). This is shown in combination with the increase of 1.0% in average cost by the solid blue line in Figure 14. The dot and the cross represent the original and likeliness approach, respectively. For the late delivery implementation, the likeliness factor increases by 9.9% to 0.958. This is shown in combination with the increase of 0.8% in average cost by the dotted blue line in the figure.

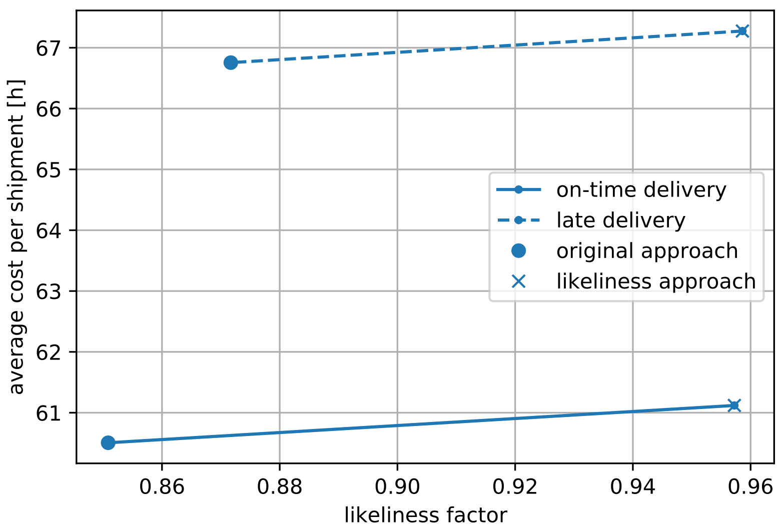


Figure 14: Average cost per shipment in the likeliness approach compared to the original approach for both the on-time and the late delivery implementation.

An air express company can use mentioned results to select a strategy of cost minimization or historical route preference. Note that such selection of strategy is based on the KPI values shown in table 5. Based on the mentioned values for these KPIs in this section, a strategy can be selected based on an increase of the likeliness factor of circa 10% and an increase in average transportation time of circa 1% for both approaches. Given the potential decrease in actual costs (so not the current cost definition of transportation time) that the selection of historical routes rather than alternative routes offers, a conclusion can be drawn that an air express company should implement the likeliness approach rather than the original approach. The potential decrease in actual cost is the result of for example more frequent usage of fewer facilities rather than less frequent usage of more facilities and therefore needing both fewer and cheaper contracts due to a stronger position in negotiations. However, since the cost in the formulation in this research is given by transportation time rather than actual costs associated with the transportation, no direct comparison can be made yet between the mentioned increase in cost and the potential decrease in actual cost by selecting historical routes. Future work can improve the strategy selection and provide a more economically substantiated choice by including the actual costs of transportation.

5.3.2 Bi-objective Model

The bi-objective formulation of the SPM part of the combined model is shown and explained in Section 4.3.3. The bi-objective formulation allows an air express company to select a strategy in the trade-off between selecting the most efficient set of routes (weight γ) or the most likely set of routes (weight $1-\gamma$), rather than a binary choice between either of the two approaches. The results of the bi-objective implementation are shown in Table 6, where the definition of each KPI is equal to the definition given in the previous section. Two values are given for each implementation for KPIs 3 to 5. This shows the extreme values for each KPI and represents the KPI values for γ-values zero and one, respectively.

	KPI	**Value**[1]		**Value LD**[2]		**Description**	**Difference**	
	γ-value	*0.0*	*1.0*	*0.0*	*1.0*		*0.0*	*1.0*
1	feasible shipments	50.3%		55.3%		of input set of shps[4]	*+9.9%*	
2	avg. comp. time[3] (s)	0.113		0.373		per shipment	*+230.1%*	
3	average cost (h)	61.652	60.506	67.827	66.755	per shipment	*+10.0%*	*+10.3%*
4	cost decrease	23.4%	24.9%	23.2%	24.6%	of feasible shipments	*-0.9%*	*-1.2%*
5	likeliness factor	0.956	0.851	0.958	0.872	of feasible shipments	*+0.2%*	*+2.5%*

Table 6: Results bi-objective approach implementation of the combined RCA and SPM. [1] value for the on-time delivery bi-objective approach. [2] value for the late delivery bi-objective approach. Abbreviations avg. comp. time[3] for average computational time per shipment and shps[4] for shipments are used for visualization purposes.

The percentages of feasible shipments (KPI 1) are in line with the values given in Section 5.3.1 since this percentage is determined in the RCA, which is prior to the SPM and equal for each previously discussed and the bi-objective approach. The average computational time (KPI 2) is 0.113 seconds and 0.373 seconds for the on-time and late delivery implementation of the likeliness approach, respectively. The average cost (KPI 3) is lower for higher γ-values due to the shift in objective to prioritize more efficient routes rather than more likely routes. The average cost reduces by 1.9% and 1.6% between $\gamma = 0.0$ and $\gamma = 1.0$ for the on-time and the late delivery implementation, respectively. As expected, the percentages of shipments for which improvement is found (KPI 4) are in line with the values for both the original implementation in Tables 2 and 3 for $\gamma = 1.0$ and the likeliness implementation in Table 5 for $\gamma = 1.0$. And, as expected, the values for the likeliness factors are in line with the values for both the original implementation for $\gamma = 1.0$ and the likeliness implementation for $\gamma = 0.0$ in Table 5 as well.

Figures 15 and 16 elaborate on the relation between the average cost and the likeliness factor for a range of γ-values between 0.0 and 1.0 for both the on-time delivery implementation (Figure 16(a)) and the late delivery implementation (Figure 16(b)). The figures show the expected lower average costs and corresponding lower likeliness factors for higher values of γ. Similar behavior is visible for both the on-time and late delivery implementation. A significant decrease in average cost is visible from value 0.0 to 0.1, caused by the change in objective from purely selecting historical routes for $\gamma = 0.0$ to the objective of also allowing the selection of alternative and less costly routes for $\gamma = 0.1$. A significant decrease in the likeliness factor is visible from value 0.9 to 1.0, which is caused by the absence of the likeliness cost multiplication factor for the latter $\gamma-$value. The absence of this factor results in no specific prioritization of historically implemented routes and thus a lower likeliness factor.

Just as for the likeliness approach discussed in Section 5.3.1, an air express company can use mentioned results to select a strategy of cost minimization or of historic route preference. By using a bi-objective formulation of the model, a trade-off between mentioned strategies can be made rather than a binary choice of either strategy. Where the results in Section 5.3.1 show a preference for implementing the likeliness approach rather than the original approach, Figures 15 and 16 can be used by an air express company to make this choice in trade-off based on the likeliness factor and the average cost per shipment. The large decrease in cost together with a minimal decrease in likeliness factor between $\gamma = 0.0$ and $\gamma = 0.1$ shows that an air express company can further improve on the mentioned implementation of the likeliness approach by implementing the bi-objective approach with $\gamma = 0.1$. Whilst these figures are useful to an air express company in the current implementation, the figures also show that performance can be achieved by providing a smoother trade-off. The current implementation

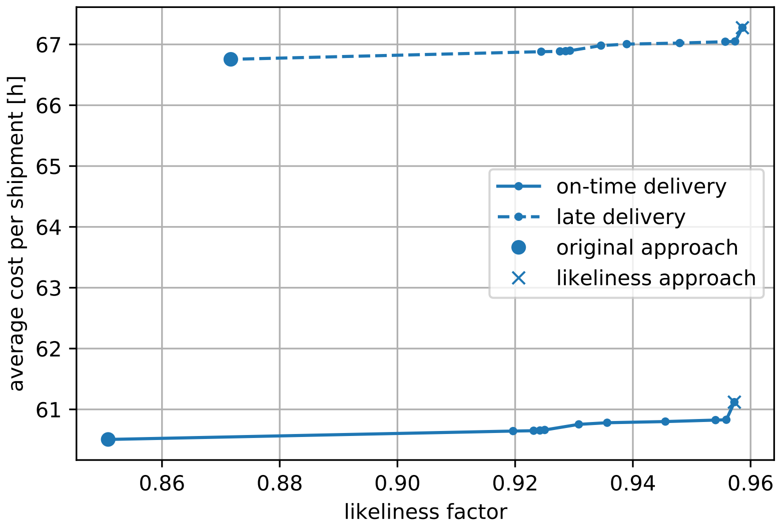
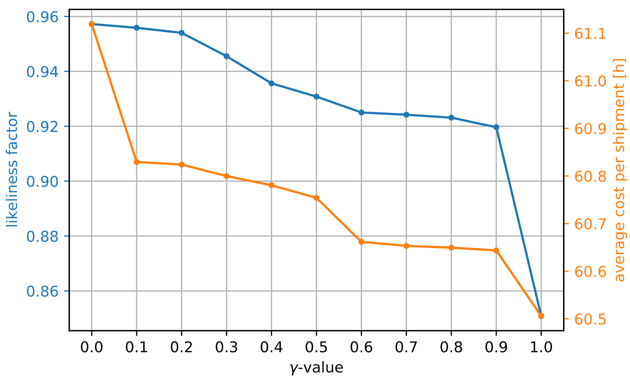
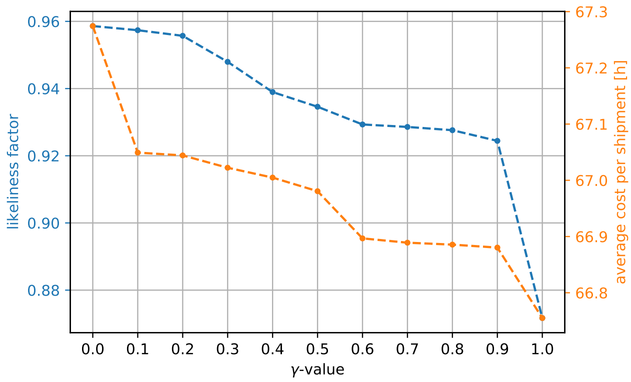


Figure 15: Average cost per shipment versus likeliness factor for different values of γ in the bi-objective approach.



((a)) Delivery time constrained implementation.



((b)) Late delivery implementation.

Figure 16: The likeliness factor and average cost per shipment for given $\gamma-$values, where a represents the delivery time constrained implementation and b represents the late delivery implementation.

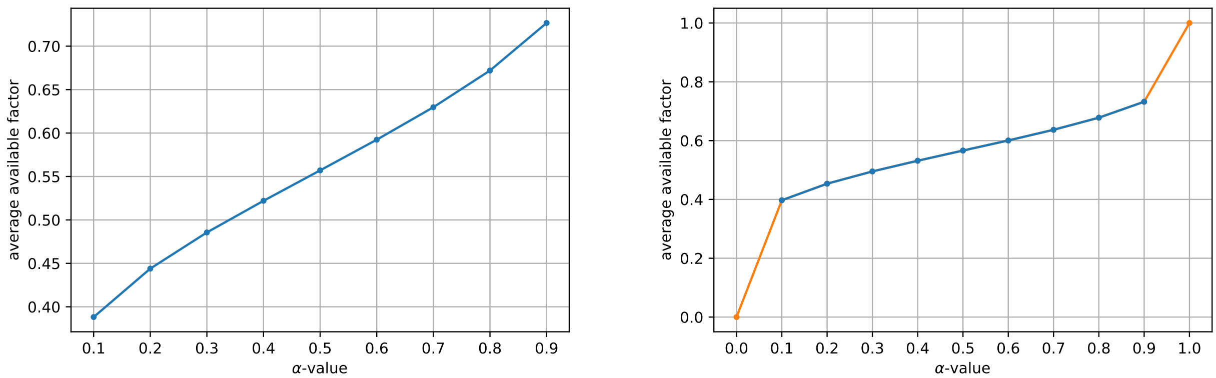
shows a large step in average cost between γ-values 0.0 and 0.1 and a large step in likeliness factor between γ-values 0.9 and 1.0. Future research can improve on the combination of the cost multiplication factor and the weight factor to provide a smoother trade-off and therefore a more precise choice for an air express company. As previously mentioned, since the cost in the formulation in this research is given by actual transportation time rather than actual costs associated with the transportation, future work can improve the strategy selection and provide a more economically substantiated choice by including the actual costs of transportation.

5.3.3 Chance Constraints

The chance-constrained formulation of the SPM part of the combined model is shown and explained in Section 4.3.4. Based on the computed available factor (AF) of historical movements with characteristics on origin, destination, vehicle type, and scheduled departure time, an assumption can be made for the AF, and therefore available capacity, of each movement in the movement schedule. The feasibility of the real-life implementation of the solution of the model by air express companies depends on this assumption. If the assumed available capacity exceeds the actual available capacity of a movement, the model is able to assign more weight to that movement than the actual capacity, thus resulting in potentially infeasible solutions.

The chance-constrained formulation of the SPM includes the implementation of different strategies, varying in levels of risk in the assumption of the AF, of an air express company. A more risk-averse strategy assumes lower available capacities on average and therefore ensures a higher chance of the solution being feasible in real-life implementation. The assumed capacity for each movement s is given by the right hand side of Equation 14: $CAP_m^\alpha = \mu_m + \Phi^{-1}(\alpha) \cdot \sigma_m$. Variable α takes values between 0 and 1, where $\alpha \to 0$ results in $\Phi^{-1}(\alpha) \to -\infty$,

$\alpha = 0.5$ results in $\Phi^{-1}(\alpha) = 0$, and $\alpha \to 1$ results in $\Phi^{-1}(\alpha) \to +\infty$. Figure 17(a) shows the average factor of the capacity that is available for different values of α in the chance-constrained formulation. The figure clearly shows that higher values for α and thus higher risk strategies account for higher assumed available capacity, which is in line with the expectation. Note that α-values 0.0 and 1.0 are not taken into account, because the normal distribution constant and therefore the assumed AF takes values towards respectively minus infinity and plus infinity for each movement in the movement schedule for these α-values. Since the available factor is a value between zero and one, a limit is set for the assumed factor to be within these values. Figure 17(b) shows the assumed available factors if α-values 0.0 and 1.0 are taken into account, where the assumed factors would be zero and one for each of the movements in the movement schedule for these α-values respectively.



((a)) AF for α-values 0.1 to 0.9.

((b)) AF for α-values 0.0 to 1.0.

Figure 17: Average available factor (AF) for movements in the movement schedule for different α-values, where b includes $\alpha = 0.0$ and $\alpha = 1.0$.

Both the on-time and late delivery implementation of the chance-constrained approach of the combined RCA and SPM have been tested for different α-values. The different α-values result in different available capacities for the movements in the movement schedule and thus result in different upper bound values for the capacity constraints in the SPM model. This influences the routes selected by the model and thereby the number of shipments with a feasible solution and the average cost for the transportation of these shipments. These results are summarized using four KPIs in Table 7, where the definition of each KPI is equal to the definition given in Section 5.3.1. Values are given for the outer bound α-values of 0.1 and 0.9, where the latter value accounts for an assumed available capacity for each movement which is roughly equal to the assumed capacity of each movement in the previously discussed approaches. Therefore, the values for each KPI for $\alpha = 0.9$ are roughly equal to the values for the original approach in Tables 2 and 3.

	KPI	**Value**[1]		**Value LD**[2]		**Description**	**Difference**	
	α-value[3]	*0.1*	*0.9*	*0.1*	*0.9*		*0.1*	*0.9*
1	feasible shipments	0.3%	50.3%	0.6%	55.3%	of input set of shps[5]	*+100.0%*	*+9.9%*
2	avg. comp. time[4] (s)	0.153		0.286		per shipment	*+86.9%*	
3	average cost (h)	63.826	60.458	73.751	66.755	per shipment	*+15.6%*	*+10.4%*
4	cost decrease	31.0%	24.9%	26.7%	24.5%	of feasible shipments	*-13.9%*	*-1.6%*

Table 7: Results chance constrained approach implementation of the combined RCA and SPM. [1] value for the delivery time constrained chance constrained approach. [2] value for the late delivery including chance constrained approach. Note that α-value[3] is only applicable for KPIs 2, 3, and 4. Abbreviations avg. comp. time[4] for average computational time per shipment and shps[5] for shipments are used for visualization purposes.

Lower α-values result in lower available capacities for each movement in the movement schedule and therefore less feasible routes can be selected. This results in a lower bound value for the number of feasible shipments (KPI 1) of just 0.3% and 0.6% for $\alpha = 0.1$ for respectively the on-time and late delivery implementation of the chance-constrained approach. The average computational time (KPI 2) is 0.153 seconds and 0.286 seconds for the on-time and late delivery implementation, respectively. Since lower α-values result in smaller solution spaces of available routes for each shipment due to lower average available capacities on each movement, the model is potentially forced to select more costly, but feasible routes for the shipments for these lower α-values. This is shown by the average cost (KPI 3) which is 5.6% and 10.5% higher for the lower bound α-value 0.1 compared to the upper bound α-value 0.9 for respectively the on-time and late delivery implementation. Despite the significantly lower amount of feasible shipments and the larger average cost for the lower bound α-value 0.1, the model does find an improvement on the historical implementation for 31.0% and 26.7% of the feasible shipments in the on-time and late delivery implementation, respectively.

As shown in Figure 17(a), different α-values in the chance-constrained approach result in different assumed AFs and therefore available capacities of the movements in the movement schedule. The influence of different available capacities on the selection of routes by the model results in different load factors of the included movements for different α-values. The load factor (LF) of a movement is the factor of used capacity divided by the total capacity (cap_t) of that movement

$$LF = \frac{cap_t - cap_a + wgt}{cap_t} \tag{15}$$

where used capacity is determined as the sum of the unavailable capacity (the total capacity minus the available capacity (cap_α)) and the added weight (wgt) for a movement. The available capacity and therefore the LF depends on the selected α-value. The average LF of the included movements is shown in Figure 18 for both the on-time and late delivery implementation of the chance-constrained model. Higher values of α account for a lower average LF. This shows that the LF is mainly influenced by the set AF for corresponding α-value since the AF increases for higher α-values and the unavailable capacity therefore decreases. This decreasing LF shows that the added available capacity of newly included movements exceeds the added weight of new shipments with a feasible solution for higher α-values. Note that the average LF does not decrease between $\alpha-$values 0.5 and 0.6. A turning point in available capacity for these values causes a significant increase in both the number of feasible shipments and the number of selected movements. In this case, however, the additional weight of the added shipments balances the additional available capacity of the added movements. Therefore, the LF between these values does not decrease with a similar slope to the other steps in $\alpha-$values. The values of the average LF differ slightly for the on-time and late delivery implementation. This slight difference is caused by the additional shipments for which the RCA does select feasible routes if late delivery is allowed.

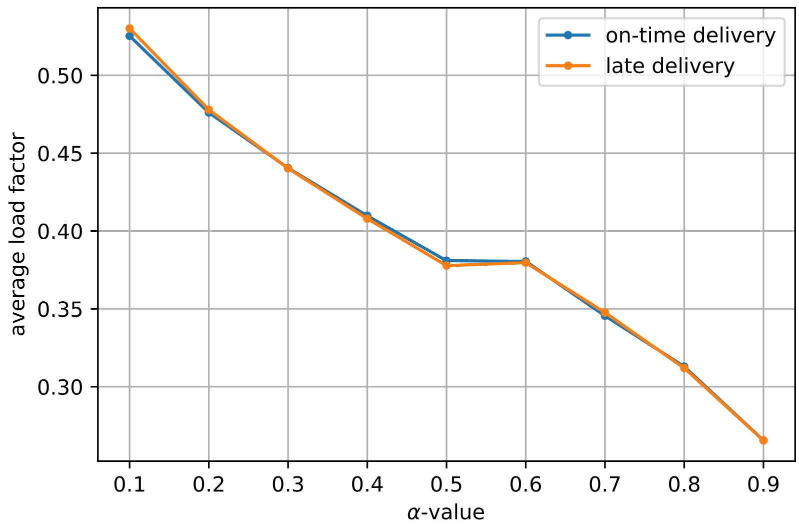


Figure 18: Average load factor (LF) for movements in the movement schedule for different α-values.

An air express company can select an α-value based on their strategy in risk regarding the actual feasibility of found solutions. Future work must investigate the actual feasibility of the real-life implementation of the found solutions for set levels of risk α to allow an air express company to define a strategy in terms of the level of risk based on the actual feasibility of these solutions for different α-values.

6 Conclusion and Recommendations

The objective of the research proposed in this paper is the formulation and implementation of a combined route construction algorithm (RCA) and set-partitioning model (SPM) to solve the Air Express Transportation Problem (AETP) for a real-life data set obtained from a global air express courier. Using a relevant subset of both shipment and movement data, the model is solved using different formulations that are in line with real-life air express courier preferences and decision strategies. In the original formulation, the model finds an improvement for around a quarter of the feasible shipments. The number of feasible shipments is around half of the shipments in the input set of shipments for the on-time delivery implementation. The number is improved by including penalized late delivery of a shipment in the late delivery implementation of the original formulation. This implementation increases the number of feasible shipments by around 10% which is accompanied by an increase in average cost per shipment of around the same percentage, 10%. Despite this increase in average transportation time, the late delivery implementation is preferred over the on-time delivery implementation by an air express courier because of the goal of the courier to rather deliver a shipment late than not deliver the shipment. More performance of the model in terms of feasible shipments and improvement in transportation time can be achieved by improving both input data completeness and the algorithm that finds an assumed processing time at a specific facility. Improving the data completeness not only results in a larger number of available historical and alternative paths but also in a larger number of constructed routes. This increases the solution space for each shipment and therefore both the chance of at least one feasible route for a shipment and the chance of a route that improves on the historical implementation. Currently, the assumption of processing times at facilities is done using a risk-averse strategy. Assumed processing times can be larger than the actual processing time which can result in not selecting an actually feasible next movement at a facility. Improving the assumption of processing times results in lower processing times on average and therefore construction of routes with less transportation time due to the selection of earlier movements at a facility. Future work must improve the algorithm to implement a more accurate and less risk-averse formulation which results in a lower average cost in the model.

Reduction of the actual average cost can be achieved by forcing the model to select more likely routes, which are defined as routes that follow a path with higher historical occurrence. This reduces the number of unique paths and therefore the number of visited facilities in the solution. This proposes a potential reduction in the actual average costs due to for example less costly contracts with fewer facilities. The adapted formulation of this likeliness approach improves the likeliness factor of the solution for both the on-time and late delivery implementation by around 10%, whilst the total transportation time increases only around 1%. This suggests that the air express company must select the likeliness formulation rather than the original formulation given the potential reduction in actual cost proposed by the increased likeliness factor. However, future work must investigate the actual costs of the solution rather than the transportation time to be able to compare the advantage of the increased number of selected routes that follow a historical path to the downside of the increased average transportation time.

The trade-off between selecting the most efficient set of routes and the most likely set of routes is implemented in the bi-objective formulation, where factor γ defines the weight set to either objective. A γ-value of 0.1 results in a nearly equal likeliness factor whilst significantly reducing the average transportation time compared to the likeliness approach for both the on-time and late delivery implementation. This formulation of the bi-objective model using a given γ-value is therefore preferred over the formulation of mentioned likeliness approach. An important note is that the current implementation does not provide a smooth trade-off between both objectives, in particular between γ-values $0.0 - 0.1$ and $0.9 - 1.0$. Future work can improve the smoothness, and therefore select a γ-value more accurately, by combining the weight factor and the construction of the cost multiplication factor for each route rather than the current implementation in sequential order.

On top of the mentioned approaches, an air express courier can implement a chance-constrained formulation of the model in which the available capacity of each movement is assumed based on a set level of risk α. This level of risk defines the real-life feasibility of a solution in terms of the potential exceeding of the capacity of a movement by the added weight of selected movements. Future work can compare the actually available capacities of each selected movement to the assumed available capacity for each α-value and determine the actual feasibility of the found solutions. This allows an air express company to select more risky solutions, which accounts for lower average cost and a higher number of feasible shipments, whilst still ensuring the actual feasibility of the solution.

References

[CPB, 2022] (2022). Cpb world trade monitor. `https://www.cpb.nl/en/world-trade-monitor-may-2022`.

[IMF, 2022] (2022). Imf world economic outlook. `https://www.imf.org/en/Publications/WEO/Issues/2022/10/11/world-economic-outlook-october-2022`.

[Andersen et al., 2009] Andersen, J., Crainic, T. G., and Christiansen, M. (2009). Service network design with asset management: Formulations and comparative analyses. *Transportation Research Part C: Emerging Technologies*, 17:197–207.

[Angelelli et al., 2020] Angelelli, E., Archetti, C., and Peirano, L. (2020). A matheuristic for the air transportation freight forwarder service problem. *Computers and Operations Research*, 123.

[Archetti and Peirano, 2020] Archetti, C. and Peirano, L. (2020). Air intermodal freight transportation: The freight forwarder service problem. *Omega (United Kingdom)*, 94.

[Barnhart et al., 2002] Barnhart, C., Kim, D., and Ware, K. (2002). Network design for express shipment delivery.

[Barnhart and Schneur, 1996] Barnhart, C. and Schneur, R. R. (1996). Air network design for express shipment service. *Operations Research*, 44:852–863.

[Chen et al., 2006] Chen, P., Guo, Y., Lim, A., and Rodrigues, B. (2006). Multiple crossdocks with inventory and time windows. *Computers and Operations Research*, 33:43–63.

[Daeki et al., 1999] Daeki, K., Barnhart, C., Ware, K., and Reinhardt, G. (1999). Multimodal express package delivery: A service network design application. *Transportation Science*, 33:391–407.

[Guastaroba et al., 2016] Guastaroba, G., Speranza, M. G., and Vigo, D. (2016). Intermediate facilities in freight transportation planning: A survey.

[Lim et al., 2005] Lim, A., Miao, Z., Rodrigues, B., and Xu, Z. (2005). Transshipment through crossdocks with inventory and time windows. *Naval Research Logistics*, 52:724–733.

[Miao et al., 2012] Miao, Z., Fu, K., and Yang, F. (2012). A hybrid genetic algorithm for the multiple crossdocks problem. *Mathematical Problems in Engineering*, 2012.

[Orden, 1956] Orden, A. (1956). The transhipment problem. *Management Science*, 2:276–285.

[Prekopa, 2003] Prekopa, A. (2003). Probabilistic programming.

II

Literature Study
previously graded under AE4020

Development of an optimisation tool to improve shipment assignment in air transportation freight forwarder service problems

Literature review

Floris Hermans
4596617

TU Delft

Development of an optimisation tool to improve shipment assignment in air transportation freight forwarder service problems

Literature review

by

Floris Hermans
4596617

This literature study is submitted in fulfillment for the course AE4020

Student number: 4596617
Supervisor: Dr. A. Bombelli, TU Delft

TU Delft

Contents

Acronyms

3PL Third-Party Logistics. 8

AoD Airport of Departure. 1, 3, 5, 7, 14, 15, 19, 28

AoL Airport of Leaving. 1, 3, 5–7, 14–17, 19, 21, 28

ATFFSP Air Transportation Freight Forwarder Service Problem. 1–3, 8, 9, 11, 12, 26, 28, 29

BP Branch & Price. 11

CG Column Generation. 9–13, 22, 25, 26, 28

ESPPRC Elementary Shortest Path Problem with Resource Constraints. 11

GA Genetic Algorithm. 6

HSN Hub and Spoke Network. 7

LSFD Liner Ship Fleet Deployment. 8

MCNF Multicommodity Network Flow Problem. 11

MDVRP Multidepot Vehicle Routing Problem. 11

MDVRPI Multidepot Vehicle Routing Problem with Interdepot routes. 11

MP Master Problem. 11, 22, 23, 25, 26

PDCGM Primal-Dual Column Generation Method. 11

RMP Restricted Master Problem. 11, 23, 25, 26

SA Simulated Annealing. 6

SNDP Service Network Design Problem. 5–7

TP Transshipment Problem. 5, 6

TS Tabu Search. 6

VRP Vehicle Routing Problem. 5, 11

VRPTW Vehicle Routing Problem with Time Windows. 11, 23–26, 28

List of Figures

1
Introduction

This report focuses on the international package shipment from origin to destination, including the presence of multiple substations, so-called intermediate facilities, for the shipments. From the origin, the shipment can be brought to a warehouse from where a truck brings a set of shipments to the airport of leaving (AoL). When flown to the airport of destination (AoD), the shipments are customs cleared and brought further to their destination. Each location in this network can be called a node and each route between nodes can be called an arc. Freight forwarding companies offer transportation of the shipment over these arcs which is called a service. A particular shipment will be subject to a set of different services between its origin and destination. The cost of each service in the network depends on multiple factors such as the duration or length of the service, the dimensions and weight of the shipment, and potential cost reduction when combining multiple shipments. It is important to minimize the total cost for all services used to meet demand and deliver all shipments from origin to destination because of both economic and environmental sustainability.

Besides the fact that companies are already in need of more sustainable solutions from a profitability point of view, the demand for the world trade of goods is increasing again which also increases the effect of efficient solutions for the freight forwarding companies and thus increases the need for these solutions even more. The World Trade Statistical Review 2021 shows that the COVID-19 pandemic resulted in a 23% decrease in the world trade of goods in the second quarter of 2020. However, in the last quarter of 2020, the trade in goods was already up to 1% compared to the pre-COVID numbers of the last quarter of 2019. This is said to be due to the adjustment to new working conditions and the working of COVID-19 vaccines.

This increase in demand for the world trade of goods combined with the need for economic and environmental sustainability calls for freight forwarding companies to be increasingly efficient and optimize their processes more than ever before. The goal of the research proposed in this report is therefore to formulate a model for the selection of an optimal set of services in which the total cost is minimized whilst meeting all demand and considering certain real-life constraints such as the capacity of certain vehicles or intermediate facilities.

Archetti and Peirano [2020] introduce this optimization problem as the Air Transportation Freight Forwarder Service Problem (ATFFSP) and give a mixed-integer linear programming (MILP) formulation. By using real-life data from a company, the authors test the model and compare the solutions to the solutions historically implemented by this company. In more recent research, Angelelli et al. [2020] use matheuristics to find a solution for this problem. This usage of matheuristics significantly decreases the computational time for larger instances and gives good solutions. When considering the real-life case, one could also focus on a slightly different version of the ATFFSP in which the freight forwarding company prioritises time minimizing instead of cost-minimizing for delivery. This changes both the objective and a selection of constraints compared to the models above. More differences between the ATFFSP and the real-life implementation of the problem will be discussed in chapter 4 and in more detail in chapter 5. Despite the differences, the research by the authors described above does give solid foundation to model an adjusted algorithm that tackles the changed objective. Therefore, this report not only aims to adapt the formulation by both authors but also further improve the work by Archetti and Peirano [2020] and Angelelli et al. [2020] on ATFFSP in terms of computational efficiency. Note that this report will consider a real-life case with data from an industrial partner and thus aims to improve the operations for a real-life instance of the adapted ATFFSP. It will consider multiple constraints found in the real-life application of shipment deliveries by freight forwarder companies.

A more detailed description of the problem that is addressed in this report is given in chapter 2. This chapter also elaborates on the general planning during a master thesis research and the specific planning of the research that is proposed in this report. Before being able to improve the real-life instance of the problem, literature research is done as is shown in chapter 3. This chapter focuses on research in the available literature that can be relevant to formulate an accurate model. The solution techniques and objectives in this available research are discussed and noted. Chapter 4 elaborates on the most relevant available research to the scope of this report. It gives a theoretical basis for the research that is to be conducted. This chapter also states the differences to currently available research and the difficulties to overcome to formulate a model for the adapted ATFFSP. These differences and a proposed method are elaborated on in chapter 5.

2 Research Outline

This chapter summarizes the outline of the research discussed in this report. The problem and objectives of the research are discussed.

As stated in the introduction, the demand for world trade of goods is increasing again and the need for both economic and environmental sustainability rises every day. The research in this report focuses on the goal of freight forwarding companies to find even more efficient solutions in their field of process optimization. These companies deliver shipments from origin to destination by using different sets of services. These services can be for example the pickup at an origin location by a truck, the delivery at an intermediate facility, the transport from an AoL to an AoD by an aircraft, or the delivery at the destination. Knowing the schedules for different vehicles, the companies make choices on what services to use for which shipments. On the one hand, these choices directly influence the profitability of the company. It can be more cost-efficient for example to store a set of shipments at a certain location and transport it after consolidating with other shipments. On the other hand, these choices indirectly influence profitability. The customer service of the company is influenced by making choices to minimize the time of delivery for the shipments. It can be more time-efficient for example to not store a shipment to be waiting for direct transport but using an indirect route instead. For example not a direct flight from AoL to AoD, but with an intermediate airport as transshipment stop. This may be less cost-minimizing, but it does increase customer service.

These objectives can be obtained by formulating accurate optimization models. The main objective of the research proposed in this report is in line with the first objective of cost minimization. As will be further explained in chapter 5, this research aims to develop an optimization tool to improve the cost minimization for freight forwarding companies in their service of transporting shipments from origin to the destination including intermediate facilities. The models for solving the ATFFSP proposed by Archetti and Peirano [2020] and Angelelli et al. [2020] are used as a theoretical basis. The aim of the research in this report is to further improve on these models in terms of computational time and optimality. More importantly and also stated in the introduction, the real-life case differs from the ATFFSP. The proposed research aims to adapt the model to this real-life case to not only find practical relevance but also to compare the results to the historical implementation by the industrial partner.

3
Solution techniques and objectives

This chapter gives an overview of the research available in literature that is relevant to the research proposed in this report. Since the problem discussed in this report includes parts of multiple relevant sub-problems and since the proposed improvement of current research includes parts of other solution techniques, this chapter reviews a variety of available literature. Section 3.1 mainly focuses on transportation planning problems. The subsections further elaborate on some relevant problems in this field. Section 3.2 reviews the available literature on bin packing problems. This is a field where the consolidation of shipments can be further optimized. Finally, section 3.3 discusses available literature on so-called column generation. After the relevant research on these topics is given, chapter 4 elaborates on the most important works to give a theoretical basis for the research in this report.

3.1. Transportation Planning

Stating the general focus of the research in this report as transportation planning, much of the available literature can be found in the literature review of Guastaroba et al. [2016], which focuses on freight transportation planning. This survey focuses mainly on merging shipments at so-called intermediate facilities between origin and destination to minimize cost. The three classes of problems considered in this literature review are vehicle routing (VRPs), transshipment (TPs), and service network design (SNDPs). Before elaborating on the latter two problems which are particularly relevant for the research proposed in this report, four features should be stated in line with the scope of this report. According to the definition by Guastaroba et al. [2016], a pure network is considered which excludes direct deliveries from origin to destination and thus forces the use of at least one intermediate facility. From the perspective of a freight forwarder company, these intermediate facilities would be the AoL and AoD for example. A second feature, the usage of multiple available intermediate facilities, is called multifacility. The demand for multiple destinations is supplied from multiple origins which makes it a many-to-many structure. However, when considering specific shipments from each a specific origin to a specific destination, the structure is one-to-one. Finally, this report considers the delivery of multiple types of shipments which makes it multicommodity.

In this report, the set of routes that are covered by fleets of trucks and aircraft throughout the network is assumed to be given. Therefore, determining the optimal set of these routes for a specific fleet, as is done by solving a VRP, is not considered. The transshipment problem, introduced by Orden [1956], is considered in this report and is an extension to the transportation problem in terms of the usage of intermediate facilities as so-called transshipment centres. According to Crainic et al. [2009], a SNDP not only considers the definition of the service network in terms of selecting the routes for the specific service and the attributes of each service, but it also considers the determination of flows of shipments over the arcs and through the nodes of a service network. This latter decision in SNDPs is of particular interest for this report and will be elaborated further on in this chapter.

The specific air transportation freight forwarder service problem as stated in the introduction was introduced by Archetti and Peirano [2020] as a combination of the latter two problems. Before this, the problem of selecting the best set of services for cost minimisation of the operations of a freight forwarding company was not yet subject to much research. However, since this problem includes elements of different problems on which significantly more research is conducted, a detailed basis of relevant previous literature can be given. According to Archetti and Peirano [2020], these elements are problems on transshipment problems (TPs), service network design (SNDPs), and, in addition to the problems considered by Guastaroba et al. [2016], freight forwarding transportation problems.

3.1.1. Transshipment Problem

Lim et al. [2005] convert the classical TP to consider many-to-many transshipment problems on a pure network and using multifacility. The authors also introduce a time window for pickup and delivery at origin and destination respectively and consider capacity constraints at the transshipment centres. The authors distinguish fixed from flexible schedules which are both provided by transportation companies. In the network, fixed schedules are present as departure and arrival times for flights for example. Flexible schedules do not have fixed departure and arrival times such as for truck transportation. Both these schedules will be considered in the research proposed in this report. The authors further distinguish in cases of multiple shipments for each origin and destination, as would be the case for clustered pickup of shipments before transport to a warehouse or AoL for example, and single shipment for each origin and destination in the network, as would be the case for a network with individual nodes for each customer. The focus of the research proposed in this report is the latter case of individual single shipment - single delivery.

This case is discussed by Miao et al. [2012b], in which hard and soft time windows are combined with penalties. In this case, a soft time window refers to a preferred service time interval in which the shipment is preferred to be delivered. When not delivered within the soft time interval but still in the hard time interval, a penalty value is added to the objective function. When the hard time window can also not be met and delivery is not possible within this window, a larger penalty will be added to the objective function. In this formulation of Miao et al. [2012b], the objective is to minimize the total cost including the penalties as described. It is found that two proposed heuristics, an adaptive tabu search (TS) and an adaptive genetic algorithm (GA), both outperform a computational time-constrained exact approach using CPLEX. Depending on the size of the problem, each of both heuristics has pros and cons as discussed in the paper. Research by Miao et al. [2012a] proposes a hybrid GA to find feasible solutions. This research uses a different definition for the penalty than previously stated. In this case, shipments that are available at the origin but not able to be shipped within the hard time window get a penalty. Again, the objective is to minimize cost including these penalties.

Just as in the research conducted by Miao et al. [2012b], Chen et al. [2006] consider hard time windows for both origin and destination. This research adds capacity constraints and storage costs for the intermediate facilities. The paper proposes a TS heuristic, a simulated annealing (SA) heuristic, and a hybrid combination of the latter two. Chen et al. [2006] find that the TS outperforms both the SA and hybrid algorithms. The formulation given by the authors is extended by Marjani et al. [2012] where it is formulated as a bi-objective integer-programming model which includes shipments between intermediate facilities. As a dual objective, the algorithm minimizes the total cost of shipment transportation and storage as well as the tardiness, meaning the degree to which a certain time interval is not met. The problem is solved using different meta-heuristics and the outcomes are compared. Marjani et al. [2012] find that the so-called variable neighbourhood search (VNS) outperforms TS and SA heuristics for this problem.

In addition to the algorithms proposed by the previously cited authors, Lapierre et al. [2004] present a new model in which the number and location of the transshipment centres are determined and is solved using a metaheuristic based on a combination of VNS and TS. To find a solution, real cost structures are used based on data from a U.S. carrier, but demand was randomised. These solutions are compared to the optimal solution found using an exact method. The hybrid algorithm is found to outperform the separate algorithms in terms of efficiency.

3.1.2. Service Network Design Problem

As stated before, a SNDP does not only consider the definition of the service network in terms of selecting the routes for the specific service and the attributes of each service, but it also considers the determination of flows of shipments over the arcs and through the nodes of a service network. In other words, one aspect is to define the service network itself. This includes determining and selecting the routes or arcs to be considered, the intermediate facilities to be used, the vehicles to be used to transport the shipments, or the frequency of this service. Once the network is defined, the second aspect is to decide how to move shipments from origin to destination using the network of services, i.e. what arcs to follow and what nodes to pass for each shipment to minimize the total cost of all services used in the network for example.

Guastaroba et al. [2016] state that the origin-destination structure is one-to-one, meaning that each shipment has a specific origin and specific destination. Thereby, the definition of a shipment consists of an origin-destination pair, the size and weight of the shipment, and often also the level of service (LoS). The authors also state that in real life, all shipments from a specific area are picked up and merged at a warehouse or AoL. In the scope of this report, the pickup of these shipments will also be considered. Note that intermediate facilities can also be origins or destinations. This would be the case if a shipment is brought to the facility by the sender or picked up at the facility by the receiver for example. Guastaroba et al. [2016] considers SNDPs for less-than-truckload carriers, time-definite freight common carriers, and express delivery carriers. This report focuses on the latter two carriers, which both provide pickup and delivery service for small shipments. The industrial partner contributing to the research conducted from this report is a combination of the latter two carriers. The company promises the customers pickup and delivery of the shipments within a specific time interval. These carriers often consolidate shipments at a specific origin centre, for example, a warehouse or the AoL, and use different types of vehicles to deliver shipments to the following intermediate facility or destination. For short distances, trucks can be used to deliver the shipments to the following intermediate facility. For larger distances, aircraft can be used to bring the shipments from an AoL to an AoD.

Much available literature on the time-definite freight common carriers makes use of the hub-and-spoke network (HSN) model. An extensive explanation of this model is given by Lin et al. [2003]. Where the authors identify three types of HSN, it is sufficient for this report in the scope of air transportation to note that in general, a HSN consists of a central 'hub' which is connected to terminals or 'spokes' and all shipments go through the hub. Models in available literature differ in the number of hubs, the number of constraints included such as capacity constraints, and the inclusion of fleet management (Lin and Chen [2004] and Lin and Chen [2008]). Research conducted by Lin and Chen [2004] also includes uncertainty in demand data. The authors determine the size and weight of each shipment by using random variables. This is of particular interest for the research conducted from this report since the size and weight of shipments are often not known beforehand in real-life freight forwarding.

Research by Barnhart and Schneur [1996] investigates express shipment delivery by carriers within specific time intervals. The authors propose a column generation algorithm to solve the SNDP near-optimal and can distinguish in a fixed aircraft fleet or an unspecified size and make-up of the aircraft fleet. In the scope of the research conducted from this report, the first case of a fixed aircraft fleet is of particular interest since the industrial partner will decide on this as part of their strategic planning. Barnhart and Schneur [1996] implement the fact that not every spoke can be reached by aircraft, so also ground movement is considered, just as in the real-life case covered in this report. The problem is formulated as binary linear programming and the column generation is implemented as a two-step heuristic.

Daeki et al. [1999] aim to model and find a solution to a large scale SNDP with given time intervals. The authors use problem reduction methods including column and row generation techniques to significantly reduce the problem size in terms of the number of constraints and variables. The research by Daeki et al. [1999] shows that the implemented techniques are very effective for large scale SNDPs which makes it particularly interesting for the scope of the research conducted from this report. This proposed heuristic is further improved by Barnhart et al. [2002].

3.1.3. Freight Forwarding Transportation

Looking into the third subproblem which has important elements that connect to the ATFFSP, research on freight forwarding transportation considers for example the decision on what available service to use for each arc in the network, the effect of merging shipments when applying different policies, routing choices, and planning on multiple levels. Available research distinguishes in the handling of shipments by company-owned vehicles or outsourcing to so-called third-party logistics (3PL) companies.

Krajewska and Kopfer [2009] find a solution to the integrated operational transportation planning problem by presenting a tabu search heuristic algorithm. This extension to the traditional vehicle routing and scheduling problem includes the possibility to outsource certain shipments to a set of potential 3PL companies. Apart from finding that a combination of own vehicles and outsourcing to third parties is in fact cost reduction effective, the authors also show that their algorithm can approximate the optimal number of vehicles in the company's fleet when considering the total fulfilment costs. This has an impact on real-world freight forwarding companies since they can adapt the size of their fleet to match the long-term planning of shipments. The importance of integrating with 3PL companies is further shown by Tyan et al. [2003], where the authors conduct research on different consolidation policies in combination with global 3PL companies. The optimal solution to the mathematical programming model given by the authors gives an insight into the quantities of shipments that should be assigned to different flights to still meet service requirements and minimize cost. An overview of more available literature on 3PL problems is given by Aguezzoul [2014], where the performance of various research is measured and distinguished based on five categories including multi-attribute decision-making techniques, statistical approaches, artificial intelligence, mathematical programming, and hybrid methods. The authors also state the most common criteria for selecting 3PL company to be cost, relationship, services, quality, information, flexibility, and delivery. Work by Jung et al. [2008] shows the importance of collaboration between, in their case, the manufacturer and the 3PL company to find an optimal supply chain plan. Since both parties want to minimize the transparency of information in real life, the decentralized framework presented in the research is based on minimal information sharing between the manufacturer and 3PL.

When considering more efficient and sustainable freight forwarding transportation, the field of research on multimodal transportation, in other words, transportation of shipments with at least two different modes of transportation, is important to investigate. Steadieseifi et al. [2014] conduct a literature review on this platform whilst considering multiple levels of planning. The scope of research conducted from this report focuses mainly on operational planning. In this field of planning, the authors distinguish in terms of problem type, the presence of transshipment, additional scheduling issues, usage of decentralized decision making and difference in additional objective components. The available literature is also sorted on the used solution methodologies, including heuristics, metaheuristics, hybrid heuristics, and simulation. Steadieseifi et al. [2014] conclude that future research on this operational planning should improve by better approximation, decomposition of the problem, and tighter formulation. The authors also acknowledge that parallel computation could mean significant improvement for higher-dimensional problems.

The available literature on the three subproblems given above provides a solid foundation for building a model that solves the air transportation freight forwarder service problem. According to Archetti and Peirano [2020], more problems share characteristics with the ATFFSP such as the liner shipping problem, widely known in maritime transportation. In these problems, line-based service in terms of container shipping is offered with a fixed schedule. These problems include the operational planning problem of allocating shipments to lines or even to containers on specific lines. Archetti and Peirano [2020] state that, on a tactical planning level, the main problem can be formulated as a liner ship fleet deployment problem (LSFD) which covers the assignment of vehicles to specific routes. The available research on these problems tends to model this problem from SND approach.

3.1.4. Air Transportation Freight Forwarder Service Problem

Based on the literature described above, Archetti and Peirano [2020] propose a MILP formulation for the air transportation freight forwarder service problem. The solutions found by using available data gathered from a real-life freight forwarding company are tested in terms of computational time to find the optimal solution and compared to the actual solutions implemented by the company. Further research by Angelelli et al. [2020] also focuses on the ATFFSP and implements a matheuristic algorithm that is based on the solution of a set-partitioning formulation and a set of feasible routes between origins and destinations. The authors show that the proposed matheuristic algorithm is effective in finding good solutions for larger size problems within less computational time than previous research available in the literature. The method proposed in this report will be based on the models by both authors and will be given in chapter 5.

3.2. Bin Packing

As stated in the introduction of this report and despite being the focus of this report, the ATFFSP is not the only important problem to be optimised to minimize cost for freight forwarding companies. This report also considers available literature on the so-called bin packing problem (BPP). Solutions to this problem find optimal loading of shipments or items into containers or unit load devices (ULDs). This is of particular importance to freight forwarding companies since minimizing the unused space when packing items into ULDs accounts for the reduced total amount of ULDs needed and thus for the reduced total amount of flights or services needed.

As there is a great variety in characteristics of different types of bin packing problems, it is important to distinguish between these types. A widely known typology for cutting and packing problems is given by Dyckhoff [1990]. The author uses four criteria to categorise the problems in this area. Firstly, Dyckhoff states that there is a difference in dimensionality for each type of problem. The geometric layouts can be one-, two-, three-, or N-dimensional. Where N in the latter case has a value larger than three. Dyckhoff further differentiates in the kind of assignment, where the available two options are to assign all objects (bins) and a selection of items or to assign all items and a selection of objects. The first option can be rephrased as including as many items as possible in a fixed number of bins. The second option would be to find the minimum number of bins to fit all items. When considering the objects, a further differentiation in Dyckhoff's typology is using one object, multiple identical objects, or multiple different objects. Finally, for the assortment of small items, Dyckhoff distinguishes between a few items of different figures, meaning different sizes and shapes, many items of many different figures, many items of relatively few different figures, and items of a similar figure. This concludes the four categorisation criteria by Dyckhoff. Wäscher et al. [2007] improve the typology of Dyckhoff by using new categorisation criteria. The authors remove the n-dimensional type of problem since these can be viewed as a variant rather than a completely different type. In the assortment of small items, Wäscher et al. [2007] identify only three types including identical small items, weakly heterogeneous items, or strongly heterogeneous items. As for the assortment of large objects, the authors distinguish in one or several large objects. In the case of one large object, the dimensions can be both fixed and variable. In the case of several large objects, all dimensions are fixed, and the objects can be identical, weakly heterogeneous, or strongly heterogeneous. When considering the real-life case type of ULD packing for freight forwarding companies, the bins, or large objects or ULDs can be identified as weakly heterogeneous assortment. According to the definition by Wäscher et al. [2007], this identifies the problem as a multiple bin size bin packing problem (MBSBPP). Note that small items or shipments, in this case, can be defined as strongly heterogeneous and the assignment is an input minimization.

3.2.1. One-Dimensional BPP

As identified by both Dyckhoff [1990] and Wäscher et al. [2007], problems on bin packing can be distinguished in terms of dimensionality. Research on one-dimensional bin packing problems often uses an exact approach or heuristics to find a solution for the adapted cutting stock problem. Basic heuristics for these problems are First-Fit, Next-Fit, Best-Fit, and nature-inspired algorithms. These algorithms are reviewed by Yesodha [2012].

However, as stated by Gilmore and Gomory [1963], such linear programming problems can involve ten to 100 million columns when considering real-life implementation. Since this would account for huge computational times, Gilmore and Gomory [1963] introduce a column generation (CG) procedure that enables the use of only a set of the columns instead of the entire length of columns. In such a procedure, the most negative reduced prices of an optimization problem determine what columns are added to the solution space. This procedure can be combined with other approaches like branch-and-bound and used as a basis for an exact procedure like the procedure introduced by Vance et al. [1994]. The authors allow CG at any node of the branch-and-bound tree for efficient computation. Carvalho [1999] extends the procedure by Vance et al. [1994] by investigating an arc flow formulation with side constraints. The algorithm proposed by Carvalho [1999] generates a set of columns at each iteration of the branch-and-bound procedure which corresponds to a feasible packing of the bin. Flow conservation constraints ensure the appropriate number of items to be packed.

3.2.2. Multi-Dimensional BPP

Apart from the literature on one-dimensional bin packing, much research has been conducted in the literature on two- and three-dimensional versions of the problem. Where two-dimensional problems are often solved in the literature using exact approaches or heuristics, only a few papers on the three-dimensional bin packing problem propose an exact algorithm. Westerlund et al. [2005] propose a MILP approach for optimal heterogeneous figured item packing in large unit objects and provide a mathematical formulation. The authors include volume constraints, non-overlapping constraints, orientation constraints, and some additional constraints. The authors note that their formulation can easily be extended to cover N-dimensional problems. This is of practical importance since logistical aspects such as time constraints can be considered.

Most of the papers in which the MBSBPP is considered as a three-dimensional problem, however, do not present an exact approach but rather use heuristics to find a feasible solution. Available literature differs in constraints considered, such as constraints on the weight limit, weight distribution, orientation, fragility, stability, and shape. Paquay et al. [2016] briefly review publications on these heuristics and use the constraints relevant for the real-life implementation of the problem in their MILP formulation. The authors include basic geometric constraints to ensure that each item is in exactly one ULD and does not overlap other items. The authors include more specific constraints on orientation, the special shape of the ULDs, stability, fragility, and weight distribution.

As stated and important to the scope of this report, the formulation by Paquay et al. [2016] includes the specific case of ULDs as objects or bins. In their research, these ULDs are classified as truncated parallelepipeds which have this shape to fit in the fuselage of the aircraft. Different ULD shapes are considered for different positions in the aircraft. Four possible cuts for a container are considered and described by the linear equation $z = ax+b$. The authors investigate three sets of possibilities where the ULDs are identical and parallelepipeds, identical and not parallelepipeds, or not necessarily identical nor parallelepipeds. The authors find that increase in the number of items or ULDs would significantly increase computational time and therefore refer to the combined use of heuristics and exact algorithms.

Recently, researchers have implemented reinforcement learning (RL) to solve the three-dimensional BPP to cope with the limitations of heuristics in terms of generality because of the situation-specific design of those algorithms. Focusing on the travelling salesman problem (TSP), Bello et al. [2016] present a framework that uses neural networks and reinforcement learning to find the solution to combinatorial optimization problems. The authors find that without much heuristic designing, close to optimal results can be achieved. Based on the potential of deep reinforcement learning (DRL) to solve these problems, as shown by Bello et al. [2016], Hu et al. [2017] apply a DRL-based method to solve three-dimensional BPP and use real-life data to demonstrate the effectiveness of the method. Where most existing literature on these three-dimensional BPPs assume that information on all items is known in advance, Zhao et al. [2021] propose a DRL method to solve the so-called online formulation of the BPP where items must be packed immediately after arrival with no information on following items. The agent in this algorithm predicts feasible placements for the item and modulates the action probabilities output by the actor during training. The authors show that their algorithm significantly outperforms previously available methods and state that the method can easily be extended to implementations such as handling lookahead items, packing multiple bins simultaneously, and re-orienting items.

3.3. Column Generation

As stated in the introduction and in chapter 2 of this report, the main focus is to improve and adapt the models currently available on the ATFFSP. The two models provided in subsection 3.1.4 are the theoretical basis of the work to be undertaken and will be elaborated upon in chapter 4. A new heuristic algorithm that improves these two models in terms of computational time will be constructed. To do this, a column generation algorithm, as first suggested by Ford and Fulkerson [1958] and later introduced as a fundamental idea by Dantzig and Wolfe [1960], will be implemented. This algorithm reduces the solution space by only considering a viable subset and transforms a master problem into a restricted master problem which has a lower dimensionality and thus accounts for less computational time needed. Relevant recent research will be discussed to use this fundamental idea for the specific problem discussed in this report. Chapter 4 elaborates on this literature by stating the basic principles of CG and how to adapt a model to be solved by a CG algorithm.

Lübbecke and Desrosiers [2005] give an overview of the available literature on topics considering integer programming column generation, which is useful in creating a column generation based model for the ATFFSP. The authors explain the classical decomposition principle used in linear programming. They state the idea of using a small subset of columns to solve a restricted master problem (RMP) instead of using the entire set of columns to solve the master problem (MP). The authors explain how to use dual variables and consequently the reduced cost to find an answer to the so-called pricing problem. The process is repeated until no more negative reduced cost coefficients are available. If so, the set of columns is believed to solve the RM optimally as well as the RMP. The authors extend this basic explanation and look into different types of problems and different types of decompositions and relaxations. The available literature is stated for each type. The authors group available literature on their specific application.

This overview of the available literature is relevant to get a view on what applications are yet able to be solved using a column generation algorithm. In the scope of this report, it is also important to know how to adapt or decompose a formulation to be subject to a column generation technique. Feillet [2010] give a tutorial on how to implement CG and branch-and-price (BP) methods to solve VRP problems with time windows (VRPTW). Even though models for VRPs are not directly relevant to the research in this report, the authors do give a good view on how to adapt a model formulation to be able to be solved by CG (or BP). In their paper, the authors first give the objective function and constraints of the VRPTW model formulation. They state that this model has a weak linear relaxation. A Dantzig-Wolfe decomposition is first needed to get a better set covering model. The authors show the decomposition and the improved model. The linear relaxation of this model is used for the formulation of a MP and a RMP.

Muter et al. [2014] propose two different CG algorithms for the so-called multidepot vehicle routing problem with interdepot routes (MDVRPI). This problem is an extension to the multidepot vehicle routing problem (MDVRP) and is modelled as a set covering problem in which the variables are feasible routes. These routes are generated by using two different pricing sub-problems. The authors distinguish in these two formulations by using one or two levels. The one-level formulation directly finds new feasible routes by solving an elementary shortest path problem with resource constraints (ESPPRC). The two-level formulation involves some more difficulties but the authors show how to overcome these and state that this second formulation performs better, especially for larger instances. Gondzio et al. [2016] also show how to use a CG technique for larger instances. The authors use a primal-dual column generation method (PDCGM) for large-scale optimization. The authors use the so-called interior point method which gives suboptimal and well-centred dual solutions. As said, the authors extend the research yet available by considering larger instances. They also give the application of their algorithm in some relevant real-life problems. In the scope of the research in this report, the numerical experiments performed by the authors on transportation networks, specifically the multicommodity network flow problem (MCNF), are important. The decomposition and CG formulation for this problem is shown and experimented on. It is shown that the PDCGM is indeed a promising method for large-scale optimization problems in terms of the number of iterations and computational time.

To conclude this chapter, in addition to the research done in the current available literature and focusing on the ATFFSP rather than the BPP, this report proposes research that further improves the work done by Archetti and Peirano [2020] and Angelelli et al. [2020] on the air transportation freight forwarding service problem in terms of computational efficiency. This is done by proposing a CG algorithm and using relevant decomposition techniques as briefly discussed in this chapter. In addition, this report will consider a real-life case with data from an industrial partner and thus aims to improve the operations for a real-life instance of the ATFFSP. The theoretical basis for this research is given in chapter 4.

4

Methodology

As discussed previously, the research proposed in this report aims to develop an optimization tool to improve cost minimization for the services provided by freight forwarder companies. The aim is to improve the formulations given in previous work by Archetti and Peirano [2020] and Angelelli et al. [2020] in terms of computational efficiency and optimality. The theoretical basis of the work to be undertaken consists therefore of the MILP model as formulated by Archetti and Peirano [2020] and the matheuristic approach as proposed by Angelelli et al. [2020]. The general formulation for a CG algorithm given by Lübbecke and Desrosiers [2005] and the specific implementation by Feillet [2010] both add to this theoretical basis to find a computationally more efficient model. Using real-life data from the actual network of an industrial partner, solutions for latter models are found and compared to the actual solutions as historically implemented by the industrial partner. Finally, solutions from the improved model proposed in the research conducted from this report are compared to both the actual historical implementations and the models previously cited and given in this chapter. The hypothesis to be tested is whether the new model is an actual improvement compared to the previous models in terms of computational efficiency and an improvement to the actual implementation in terms of cost minimization.

4.1. ATFFSP formulations

To include a theoretical basis of the works of both Archetti and Peirano [2020] and Angelelli et al. [2020], it is important to notice the similarities and differences between the two proposed models. Many parameters, decision variables, and sets show similarities but often differ slightly in their notation. Therefore, before stating the mathematical formulation of both methods, the parameter and variable notations are compared in table 4.1. Once these notations are stated, the time-space notation of Archetti and Peirano [2020] is given and the route construction algorithm used by Angelelli et al. [2020] is summarized. With these formulations known, the decision variables of both models are compared in table 4.1 and the MILP model of Archetti and Peirano [2020] and the set-partitioning formulation for the ATFFSP of Angelelli et al. [2020] are stated respectively. Both formulations include dominance rules which are compared. Once this is done, both models have been given and can be used as the theoretical basis for the research conducted from this report.

As stated, table 4.1 shows the notations used for parameters in both pieces of research. This includes parameters for shipments such as the origin, destination, pickup and delivery time, and weight for example. This also includes parameters for the available services such as the start and end location, start time, transit time, and service cost. Notations for the different types of locations and different types of services are also given. Finally, the parameters (and their calculation if necessary) relevant for cost determination of services or stocking are given. Note that empty cells in this table show that the authors do not make use of the parameter in their formulation. Also note that the set used by Archetti and Peirano [2020] for the shipments stocking costs will be explained below, as part of the explanation of the time-space representation.

	Archetti and Peirano [2020]	Angelelli et al. [2020]
time periods	$H = (1, \ldots, T)$	T time periods of 12 hrs
shipment parameters:	$k \in C$	$k \in K$
- origin	p^k	o_k
- destination	d^k	d_k
- earliest time pickup	α^k	e_k
- latest time delivery	β^k	l_k
- early delivery	θ_k^-	θ_k^-
- late delivery	θ_k^+	θ_k^+
- weight	wgt^k	w_k
- volume	vol^k	
service parameters:	$w \in S$	$s \in S$
- start location	p^w	o_s
- end location	d^w	d_s
- start time	α^w	e_s
- transit time	τ^w	d_s
- service cost	γ^w	c_s
locations:	L	
- origins	$OR = \cup_{k \in C} p^k$	O
- warehouses	WH	H
- AoLs	AL	AL
- AoDs	AD	AD
- destinations	$DEST = \cup_{k \in C} d^k$	D
transport services:	S	S
- dedicated trucks	TR_{ded}	TR_{ded}^{O-AL} $o_s \in O$, $d_s \in AL$ TR_{ded}^{O-H} $o_s \in O$, $d_s \in H$ TR_{ded}^{H-AL} $o_s \in H$, $d_s \in AL$
- groupage trucks	TR_{grou}	TR_{gr}^{H-AL} $o_s \in H$, $d_s \in AL$
- air transport	AC	AC $o_s \in AL$, $d_s \in AD$
- foreign agents	AG	FA $o_s \in AD$, $d_s \in D$
total service cost:	γ^w	c_s
- dedicated trucks	unitary cost for distance	unitary cost for distance
- foreign agents	fare by foreign agent	fare by foreign agent
- groupage trucks:		
- weight ranges	$R^w = \{1, 2, \ldots, \lvert R^w \rvert\}$	$R = 1, 2, \ldots, \lvert R \rvert$
- lower & upper bound	(l_r, u_r)	(l_r, u_r)
- volumetric weight	$Volwgt_{TR}^k = vol^k \cdot 300$	
- chargeable weight	$CW_{TR}^k = max(\frac{wgt^k}{100}; \frac{Volwgt_{TR}^k}{100})$	
- unitary cost	FR^{wr} for $l_r \leq CW_{TR}^k \leq u_r$	f_{sr} where: $r = 2, \ldots, \lvert R \rvert$
- service cost	$\gamma^w = FR^{wr} \cdot CW_{TR}^k$	
- air companies	equal to groupage trucks, but:	equal to groupage trucks
- volumetric weight	$Volwgt_{Air}^k = vol^k \cdot 167$	
- chargeable weight	$CW_{Air}^k = max(wgt^k; Volwgt_{Air}^k)$	
stock holding	$st_{i,j} \lvert (i,j)$ $(i,j) \in A_6 \cup A_7 \cup A_8$	st_i $i \in H \cup AL \cup AD$

Table 4.1: Comparison of parameter notation used by the authors considered. Note that an empty cell shows that the certain parameter is not used by the specific authors and is therefore left empty.

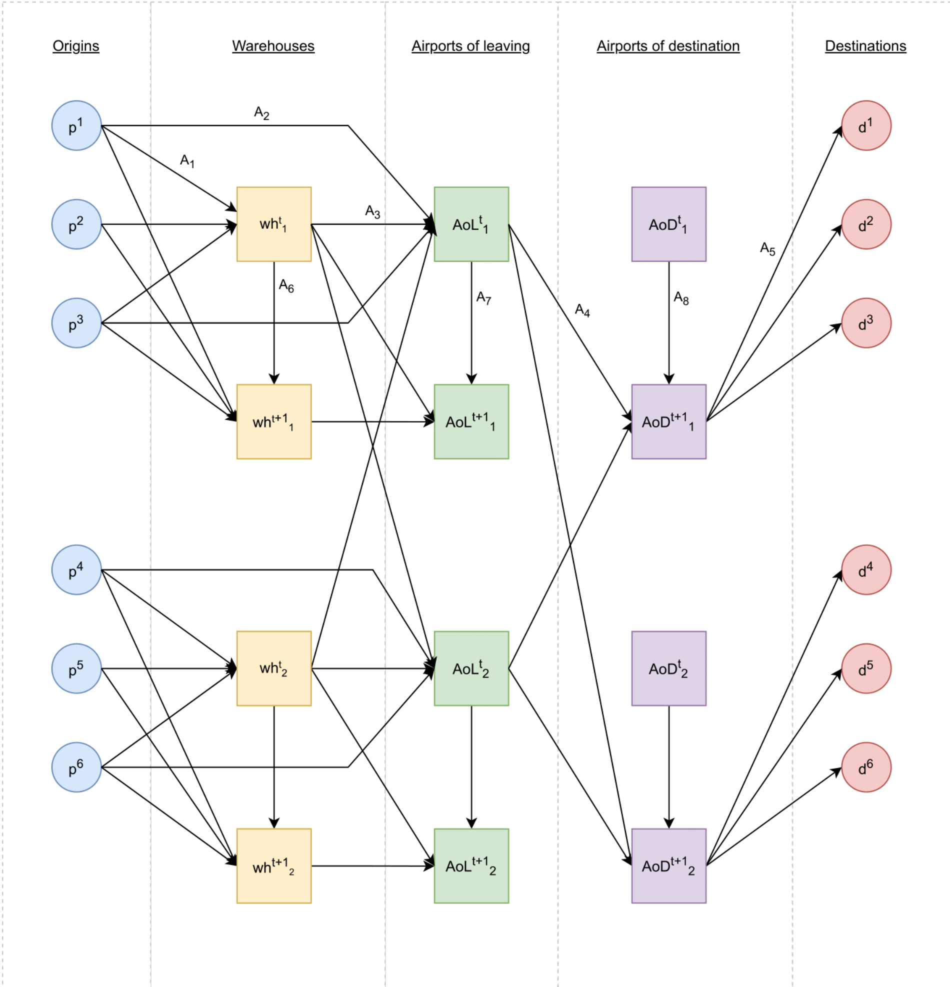


Figure 4.1: State-space representation of the network as implemented by Archetti and Peirano [2020]. Source: Adapted from Archetti and Peirano [2020].

4.1.1. Time-Space Network

In general, a network used in an air transportation freight forwarder service problem involves a shipment k starting at its origin o_k and moving through the network to its destination d_k. To transport the shipment from origin to destination, transportation service is needed between each node of the network. This node can be the origin, a warehouse, an airport of leaving (AoL), an airport of destination (AoD), or the destination. Archetti and Peirano [2020] distinguish four different segments in this network. First, shipments are transported from their origin to either a warehouse or directly to the AoL. At both intermediate facilities stated, shipments can be grouped or consolidated to reduce further shipment costs. Moving shipments from a warehouse to an AoL is the second type of segment considered by the authors. From an AoL to an AoD, air transportation service is used where consolidation of shipments with the same AoD may reduce costs for the freight forwarding company. Finally, the shipments are transported from the AoD to their destination once customs cleared. The segments and nodes explained can be visualised in terms of a time-space network. Figure 4.1 shows this visualisation where an arrow from an intermediate facility at time t to the same intermediate facility at time $t+1$ accounts for stocking of the shipment. The time-space network as visualised in this figure is given by $G = (V, A)$.

The set of nodes (V) in the time-space network used by Archetti and Peirano [2020] is given as: $V = V_{OR} \cup V_{WH} \cup V_{AoL} \cup V_{AoD} \cup V_{DEST}$, where the elements are formulated as:

$$
\begin{aligned}
V_{OR} &= OR \\
V_{WH} &= \{wh_j^t | t \in H, j \in WH\} \\
V_{AoL} &= \{AoL_j^t | t \in H, j \in AL\} \\
V_{AoD} &= \{AoD_j^t | t \in H, j \in AD\} \\
V_{DEST} &= DEST
\end{aligned}
$$

The entire set of arcs (A) in the time-space network is given as: $A = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6 \cup A_7 \cup A_8$. The authors distinguish in a set of transportation services $\tilde{A} = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5$ and a set of stocking services $\bar{A} = A_6 \cup A_7 \cup A_8$. Formulation of both sets and their elements are given as:

$$
\begin{aligned}
&A_1 = (i,j) \quad \forall w \in TR_{ded} | i = p^w \wedge i \in V_{OR}, \alpha^w \geq \alpha^k; j = d^w \wedge j \in V_{WH} | t = \alpha^w + \tau^w \\
&A_2 = (i,j) \quad \forall w \in TR_{ded} | i = p^w \wedge i \in V_{OR}, \alpha^w \geq \alpha^k; j = d^w \wedge j \in V_{AoL} | t = \alpha^w + \tau^w \\
&A_3 = (i,j) \quad \forall w \in TR_{ded} \cup TR_{grou} | i = p^w \wedge i \in V_{WH} | t = \alpha^w; j = d^w \wedge j \in V_{AoL} | t = \alpha^w + \tau^w \\
&A_4 = (i,j) \quad \forall w \in AC | i = p^w \wedge i \in V_{AoL} | t = \alpha^w; j = d^w \wedge j \in V_{AoD} | t = \alpha^w + \tau^w \\
&A_5 = (i,j) \quad \forall w \in AG | i = p^w \wedge i \in V_{AoD} | t = \alpha^w; j = d^w \wedge j \in V_{DEST} \\
&A_6 = (wh_j^t, wh_j^{t+1}) \forall t \in H | 1 \leq T-1; wh_j^t \in V_{WH} \\
&A_7 = (AoL_j^t, AoL_j^{t+1}) \forall t \in H | 1 \leq T-1; AoL_j^t \in V_{AoL} \\
&A_8 = (AoD_j^t, AoD_j^{t+1}) \forall t \in H | 1 \leq T-1; AoD_j^t \in V_{AoD}
\end{aligned}
$$

4.1.2. Route Construction Algorithm

Angelelli et al. [2020] use a different approach to model the air transportation freight forwarder service problem. The authors propose a route construction algorithm to generate all feasible routes for a shipment from origin to destination. After reducing this set of feasible routes using acceleration techniques and a heuristic method, a set-partitioning formulation is proposed to choose the optimal set of routes to minimize the company's costs. The route construction algorithm will first be explained. Note that this explanation does include the determination of cost for each route but the explanation of the acceleration techniques and the proposed heuristic method will be given after the statement of the MILP model by Archetti and Peirano [2020] and the set-partitioning method by Angelelli et al. [2020]. This is done to be able to compare the dominance rules considered by both authors.

Angelelli et al. [2020] formulate the set of feasible routes ω for each shipment k from its origin to its destination as Ω_k. An important definition is that a path is a partial route, so not the entire route from origin to destination. The concept of consecutive services is also important for the construction of feasible routes. A service, s', is in a set of consecutive services of a previous service, s, denoted as $s' \in CS(s)$, if the origin of s' is equal to the destination of the previous service ($d_s = o_{s'}$) and if $e_s + t_s \leq e_{s'}$. The other way around, a service s' is in a set of previous services of service s, $PS(s)$ if $s \in CS(s')$. The authors give an example to clarify this formulation of consecutive services and continue by identifying two types of routes that can be followed in this network. The first type does not include a visit to a warehouse by the shipment k but rather direct transportation from origin to AoL. A set of routes of this type is denoted as PD_k. The second type does include warehouses and thus assumes that the shipment k is first transported from origin to warehouse before being transported from warehouse to AoL. A set of routes of this second type is denoted as PWH_k. The entire set of feasible routes Ω_k is the union of the sets of these two types. A more detailed definition of both types is given by:

$$
\begin{aligned}
PD_k = \{(s, s', s'') \in TR_{ded}^{O-AL} \times AC \times FA \mid o_s = o_k, e_k \leq e_s, s' \in CS(s), s'' \in CS(s'), \\
d_{s''} = d_k, e_{s'} + t_{s'} \leq l_k + \eta, e_{s''} + t_{s''} \leq l_k + \beta\}
\end{aligned}
\tag{4.1}
$$

$$
\begin{aligned}
PHW_k = \{(s, s', s'', s''') \in TR_{ded}^{O-H} \times TR_{gr} \cup TR_{ded}^{H-AL} \times AC \times FA \mid o_s = o_k; e_k \leq e_s, s' \in CS(s), \\
s'' \in CS(s'), s''' \in CS(s''), d_{s'''} = d_k, e_{s''} + t_{s''} \leq l_k + \eta, e_{s'''} + t_{s'''} \leq l_k + \beta\}
\end{aligned}
\tag{4.2}
$$

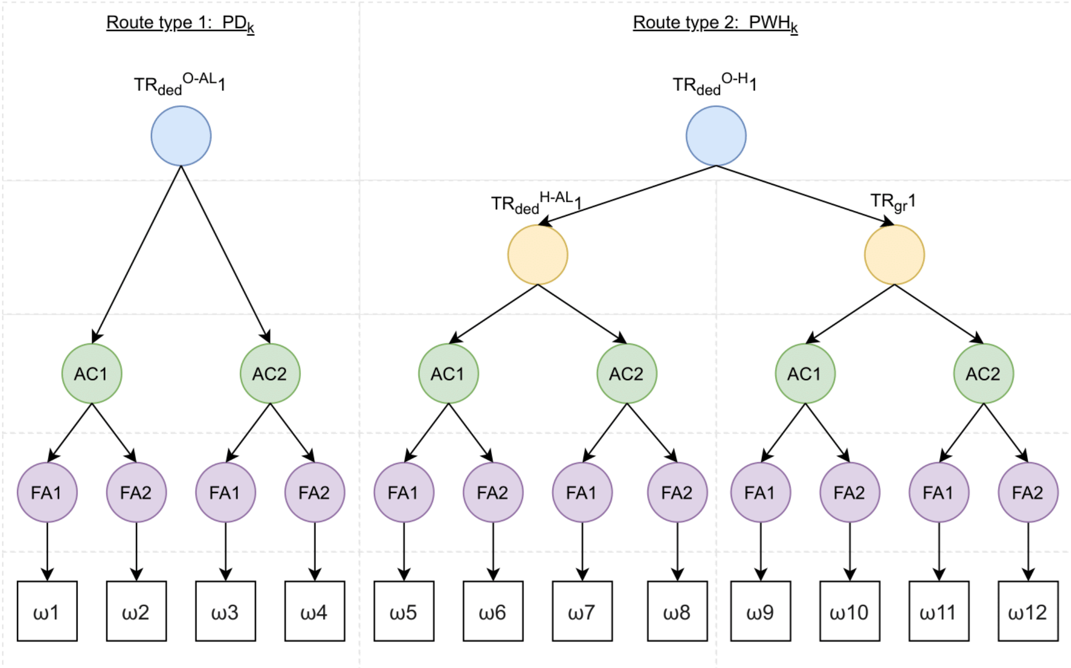


Figure 4.2: Route construction procedure as implemented by Angelelli et al. [2020]. Source: Adapted from Angelelli et al. [2020].

In these definitions, $e_k \leq e_s$ ensures that shipment k is available for pickup before or at the time that service s is available. A limit is imposed in the starting times and is calculated by adding a time of η days for air services and β days for foreign delivery services to the requested delivery time.

Important to note is that shipments from the origin to either a warehouse or an AoL are assumed to be transported by a dedicated truck. Consolidating or grouping is assumed possible for shipments from a warehouse to an AoL. Therefore, when considering routes of type 2 or PWH_k, two different legs are added for the shipment between warehouse and AoL. This is shown in figure 4.2, where the services $s \in TR_{ded}^{O-H}$ are considered as the first leg and the services $s' \in CS(s) \cup (TR_{ded}^{H-AL} \cup TR_{gr})$ as the second leg.

As touched upon in table 4.1, Angelelli et al. [2020] use a definition for the cost of a route, c_ω, which is a sum of the transportation cost, holding cost, and penalties or gains for a late or early delivery respectively. The cost for consolidation or airline services is not included in this definition. Important is the notation of services on the same route, where s^- stands for services in $PS(s)$ and s^+ stands for services in $CS(s)$. The cost of services provided by dedicated trucks or foreign agents is denoted as c_s. Since Archetti and Peirano [2020] already use a definition for $\bar{A}$ and $\tilde{A}$, the definitions by Angelelli et al. [2020] using these formulations are presented as $\bar{A_b}$ and $\tilde{A_b}$.

$$\tilde{A_b} = TR_{ded}^{O-AL} \cup TR_{ded}^{O-H} \cup TR_{ded}^{H-AL} \cup FA$$
$$\bar{A_b} = TR_{gr} \cup AC$$

The cost can now be given by:

$$c_\omega = \sum_{s \in \omega | s \in \tilde{A_b}} c_s + \sum_{s \in \omega} \sum_{o_s \in H \cup AL \cup AD} st_{o_s} \cdot w_k \cdot (e_{s^+} - (e_s + t_s)) + \Delta TT^{sk}$$

Where:

$$\Delta TT^{sk} = \begin{cases} ((e_s + t_s) - l_k) \cdot \theta^+ \text{when } ((e_s + t_s) - l_k) \geq 0, s \in FA \mid d_s = d_k \\ ((e_s + t_s) - l_k) \cdot \theta^- \text{when } ((e_s + t_s) - l_k) < 0, s \in FA \mid d_s = d_k \end{cases}$$

	Archetti and Peirano [2020]	Angelelli et al. [2020]
binary variables:		
- 1 if arc* is used for shipment	$x_{(i,j)}^{kw} = \{0,1\} \quad \forall k \in C; (i,j) \in \tilde{A}$ $f_{(i,j)}^{k} = \{0,1\} \quad \forall k \in C; (i,j) \in \bar{A}$	$x_{k\omega} \in \{0,1\} \forall k \in K, \omega \in \Omega_k$
- 1 if weight range is applied	$y^{wr} = \{0,1\} \forall r \in R^w$ $w \in TR_{grou} \cup AC$	$y_{sr} \in \{0,1\} \forall r \in R, s \in \bar{A_b}$
- 1 is service in range is used	$z^{wrk} = y^{wr} x_{(i,j)}^{wk}$	$z_{srk} \in \{0,1\} \forall k \in K, s \in \bar{A_b}, r \in R$
delay of shipment	$TT_k^+ \in \mathbb{N}$	
anticipation of shipment	$TT_k^- \in \mathbb{N}$	

Table 4.2: Comparison of decision variable notation used by authors considered. Note that an empty cell shows that the decision variable is not used by the specific author and is therefore left empty. *arc is the correct notation for the implementation by Archetti and Peirano [2020], note that this variable refers to the entire route rather than one arc in the implementation by Angelelli et al. [2020].

Angelelli et al. [2020] use dominance rules, mid-cost frontier, and a heuristic rule to reduce the solution space and select only the subset of the most promising routes. Before explaining these implementations, the models used by both Archetti and Peirano [2020] and Angelelli et al. [2020] will be stated and the constraints will be discussed. To do this, an overview of the decision variables used in the models and their notations by each author is given in table 4.2. Three used binary decision variables take value 1 if an arc or route is used for a shipment, if a weight range is applied for a service, and if a service with a specific weight range is used for a shipment. Note that all variables take value zero otherwise. First, the MILP model as implemented by Archetti and Peirano [2020] will be given and the constraints will be discussed. After this, the set-partitioning formulation by Angelelli et al. [2020] will be given and again, the constraints will be discussed.

4.1.3. Mathematical Formulations

Archetti and Peirano [2020] linearize their formulation of the objective function using the variables z^{wrk}, given in table 4.2. Their objective is to minimize the total cost. This includes transportation cost, stocking cost, and penalties or gains. The authors introduce the set of nodes $\tilde{\tilde{A}} = A_1 \cup A_2 \cup A_5 \cup A_3 | w \in TR_{ded}$. The objective function is given by:

$$\begin{aligned} \min \sum_{k;(i,j)\in\tilde{\tilde{A}}} \gamma^w x_{(i,j)}^{wk} + \sum_{k\in C;(i,j)\in\{A_3|w\in TR_{\text{grou}}\}\cup A_4;r\in R^w} FR^{wr} CW^k z^{wrk} + \\ \sum_k \theta_k^+ TT_k^+ - \sum_k \theta_k^- TT_k^- + \sum_{k\in K;(i,j)\in\bar{A}} st_{(i,j)} \cdot wgt^k \cdot f_{(i,j)}^k \end{aligned} \tag{4.3}$$

The constraints defined by the authors will be stated first and discussed below.

$$\sum_{(i,j)\in A_1\cup A_2|i=p^k} x_{(i,j)}^{wk} = 1 \quad k \in C \tag{a}$$

$$\sum_{(i,j)\in A_5|i=d^k} x_{(i,j)}^{wk} = 1 \quad k \in C \tag{b}$$

$$\begin{aligned} \sum_{(i=p^k,j=wh_s^t)\in A_1} x_{(i,j)}^{wk} + f_{(wh_s^{t-1},wh_s^t)}^k \\ -f_{(wh_s^t,wh_s^{t+1})}^k - \sum_{(i=wh_s^t,j)\in A_3} x_{(i,j)}^{wk} = 0 \quad k \in C, wh_s^t \in V_{WH} \end{aligned} \tag{c}$$

$$\sum_{(i,j=AoL_s^t)\in A_3} x_{(i,j)}^{wk} + \sum_{(i=p^k,j=AoL_s^t)\in A_2} x_{(i,j)}^{wk}$$
$$+f^k_{(AoL_s^{t-1},AoL_s^t)} - f^k_{(AoL_s^t,AoL_s^{t+1})} - \sum_{(i=AoL_s^t,j)\in A_4} x_{(i,j)}^{wk} = 0 \quad k \in C, AoL_s^t \in V_{AoL} \qquad (d)$$

$$\sum_{(i,j=AoD_s^t)\in A_4} x_{(i,j)}^{wk} + f^k_{(AoD_s^{t-1},AoD_s^t)}$$
$$-f^k_{(AoD_s^t,AoD_s^{t+1})} - \sum_{(i=AoD_s^t,j=d^k)\in A_5} x_{(i,j)}^{wk} = 0 \quad k \in C, AoD_s^t \in V_{AoD} \qquad (e)$$

$$\sum_{(i,j)\in A_5} (\alpha^w + \tau^w) x_{(i,j)}^{wk} - \beta^k \le -TT_k^- + TT_k^+ \quad k \in C \qquad (f)$$

$$x_{(i,j)}^{wk} \le M_j \quad k \in C, (i,j) \in A_2 \qquad (g)$$

$$\sum_{k\in C} CW^k x_{(i,j)}^{wk} \ge l^r y^{wr} \quad (i,j) \in A_4 \cup \{A_3 \mid w \in TR_{grou}\}, r \in R^w \qquad (h)$$

$$\sum_{r\in R^w} y^{wr} = 1 \quad w \in AC \cup TR_{grou} \qquad (i)$$

$$z^{wrk} \le x_{(i,j)}^{wk} \quad r \in R^w, w \in AC, k \in C, \quad (i,j) \in A_4 \cup \{A_3 \mid w \in TR_{grou}\} \qquad (j)$$

$$z^{wrk} \le y^{wr} \quad r \in R^w, w \in AC \cup TR_{grou}, k \in C \qquad (k)$$

$$z^{wrk} \ge x_{(i,j)}^{wk} + y^{wr} - 1 \quad r \in R^w, w \in AC, k \in C, \quad (i,j) \in A_4 \cup \{A_3 \mid w \in TR_{grou}\} \qquad (l)$$

$$x_{(i,j)}^{wk} \in \{0,1\} \quad k \in C, (i,j) \in \tilde{A} \qquad (m)$$

$$f_{(i,j)}^{k} \in \{0,1\} \quad k \in C, (i,j) \in \bar{A} \qquad (n)$$

$$z^{wrk} \in \{0,1\} \quad r \in R^w, k \in C, w \in AC \cup TR_{grou} \qquad (o)$$

$$y^{wr} \in \{0,1\} \quad r \in R^w, w \in AC \cup TR_{grou} \qquad (p)$$

$$TT_k^+, TT_k^- \in \mathbb{N} \quad k \in C \qquad (q)$$

Constraint a is added to make sure that each shipment is picked up at the origin and delivered at the warehouse or AoL, depending on using a service in set A_1 or set A_2. To ensure this, the constraint forces only one arc in both of these sets to be activated for a specific shipment k. A similar notation is used in constraint b to make sure that each shipment is picked up at the AoD and delivered at the destination. Only one arc in set A_5 can be activated for a specific shipment k.

Constraints c, d, and e are used to balance the nodes of warehouses, AoLs, and AoDs respectively. These constraints force the sum of incoming arcs minus the sum of outgoing arcs for each of these nodes to be zero for each shipment. This ensures that each shipment k that enters a specific node of one of these types also leaves.

Constraint f considers the final set of arcs, A_5, from AoD to the destination of each shipment. The possible delay or anticipation of shipments for services in this set of arcs are defined as variables TT_k^- and TT_k^+.

There is a possibility that transportation of a shipment from its origin directly to an AoL is not possible, mostly because the specific AoL is not equipped to handle customs operations. Variable M_j takes value 0 if this is the case and takes value 1 if the AoL can be used to handle customs operations. Constraint g forces the decision variable $x_{(i,j)}^k$ to be less than or equal to the value of M_j to exclude arcs from origin to AoL if customs handling must be done at a warehouse.

The binary variable y^{wr} takes value 1 if a certain weight range r is applied for service w, as denoted in Table 4.2. Constraint h defines the applicable range for a service by ensuring that the combined chargeable weight of the shipments in that service exceeds the lower bound of the range. In addition, constraint i ensures that only one range is chosen.

The binary decision variable z^{wrk} takes value 1 if certain arc or service, w, is used for a shipment, k, assuming that a range, r, is applied to this service. This includes air services in A_4 and groupage truck services in A_3. Table 4.2 shows how to calculate z^{wrk}. Constraints j, k, and l are used to implement this definition into the model. Constraints j and k force z^{wrk} to be zero if either or both of the variables $x^{wk}_{(i,j)}$ and y^{wr} have value zero. Constraint l forces z^{wrk} to take value one if both latter variables have value one.

The domains of each of the decision variables stated in table 4.2 are determined by constraints m, n, o, p, and q. This concludes the MILP formulation by Archetti and Peirano [2020]. For further theoretical basis, the set-partitioning model as implemented by Angelelli et al. [2020] will be discussed. The main idea, objective value, and used constraints are stated.

This set-partitioning problem uses the set of routes as constructed by the route construction algorithm and finds the optimal set to deliver all shipments from origin to destination as cost-effective as possible. It uses the definition of cost for each route, c_ω, as previously defined and discussed in this report. This cost does not include the cost for services with consolidated shipments. The costs of these services are determined according to the combined weight of the shipments and are added to the objective function in case these services are used. This objective function in the set-partitioning formulation minimizes the sum of costs of all used routes including these consolidated services and is given by Angelelli et al. [2020] as:

$$\min \sum_{k \in K} \sum_{\omega \in \Omega_k} c_\omega x_{k\omega} + \sum_{k \in K, s \in \bar{A}_b, r \in R} w_k f_{sr} z_{srk} \tag{4.4}$$

The constraints used by the authors are stated and discussed below:

$$\sum_{\omega \in \Omega_k} x_{k\omega} = 1 \quad \forall k \in K \tag{r}$$

$$\sum_{k \in K, \omega \in \Omega_k} w_k x_{k\omega} \geq l_r y_{sr} \quad \forall s \in \bar{A_b}, r \in R \tag{s}$$

$$\sum_{r \in R} y_{sr} = 1 \quad \forall s \in \bar{A_b} \tag{t}$$

$$z_{srk} \geq y_{sr} + \sum_{\omega \in \Omega_k | s \in \omega} x_{k\omega} - 1 \quad \forall k \in K, s \in \bar{A_b}, r \in R \tag{u}$$

$$x_{k\omega} \in \{0,1\} \quad \forall k \in K, \omega \in \Omega_k \tag{v}$$

$$y_{sr} \in \{0,1\} \quad \forall r \in R, s \in \bar{A_b} \tag{w}$$

$$z_{srk} \in \{0,1\} \quad \forall k \in K, s \in \bar{A_b}, r \in R \tag{x}$$

Constraint r ensures that for each shipment, one and only one route, or set of services, is selected and used. The selection of a weight range for services that include consolidation is done by constraint s. It adds the weights of all shipments in the service and ensures that the correct weight range is chosen according to the total weight of these shipments. Constraint t adds to this definition by ensuring that only one weight range is chosen. The definition of variable z_{srk} can be formulated as a binary variable that has value 1 if a service in a certain weight range is chosen and used. This definition is given by constraint u. Finally, the domains of each of the decision variables stated in table 4.2 are determined by constraints v, w, and x. This concludes the set-partitioning model as formulated by Angelelli et al. [2020].

4.1.4. Dominance Rules

The main models by both Archetti and Peirano [2020] and Angelelli et al. [2020] are now summarized in this report. As previously noted, both algorithms use dominance rules to reduce the solution space before solving the model because of computational efficiency. Two types of dominance rules are used by both authors. Both use a definition for cost dominance and fast delivery dominance. These rules are

$service\ 1 > service\ 2$	**Archetti2020**	**Angelelli2020**
cost dominance:	$\alpha^{w_1} = \alpha^{w_2}$	$e_{s1} = e_{s2}$
	$\alpha^{w_1} + \tau^{w_1} = \alpha^{w_2} + \tau^{w_2}$	$t_{s1} = t_{s2}$
		$o_{s1} = o_{s2} \wedge d_{s1} = d_{s2}$
- dedicated trucks & foreign agents	$\gamma^{w_1} < \gamma^{w_2}$	$c_{s1} < c_{s2}$
- groupage trucks & AC	$FR^{w_1 r} \leq FR^{w_2 r} \forall r \in R^{w_1}; R^{w_2}$	$f_{s_1 r} < f_{s_2 r} \forall r \in R$
	$\lvert R^{w_1} \rvert = \lvert R^{w_2} \rvert$	
	$l_{rw_1} \leq l_{rw_2} \lvert r \in R^{w_1}$	
fast delivery dominance:	$\alpha^{w_1} = \alpha^{w_2}$	$e_{s1} = e_{s2}$
	$\gamma^{w_1} \leq \gamma^{w_2}$	$c_{s1} \leq c_{s2}$
	$\tau^{w_1} < \tau^{w_2}$	$t_{s1} < t_{s2}$
		$o_{s1} = o_{s2} \wedge d_{s1} = d_{s2}$

Table 4.3: Notation for cost dominance and fast delivery dominance used by Archetti and Peirano [2020] and Angelelli et al. [2020].

compared in table 4.3. It is clear that both authors use similar rules but differ in their notation. For cost dominance, the service with the lowest unitary or service is prioritised. For fast delivery dominance, the service which uses the smallest amount of time is prioritised. Note that this service should have equal or lower unitary costs than the opposing service. Apart from these two dominance rules, Angelelli et al. [2020] introduce three more rules including waiting dominance, mid-leg dominance, and last-leg delivery dominance. These rules are explained below.
To define waiting dominance, path p is formulated as the partial route considered with service s as the last service for this path. Two paths, p_1 and p_2, are considered and contain the last services s_1 and s_2 respectively. The waiting dominance rule is applied to dedicated truck services between warehouse and AoL and services by foreign agents since these starting locations include stocking costs. Path p_1 dominates path p_2 if these stocking costs are lower than the cost difference between the two services s_1 and s_2. This is denoted as:

$$
\begin{gathered}
e_{s_1} > e_{s_2} \\
d_{s1} = d_{s_2} \\
o_{s1} = o_{s_2} \\
e_{s_1} + t_{s_1} = e_{s_2} + t_{s_2} \\
c_{s_1} + st_{o_{s1}} \cdot \left(e_{s_1} - e_s + t_s\right) \cdot w_k < c_{s_2} + st_{o_{s2}} \cdot \left(e_{s_2} - e_s + t_s\right) \cdot w_k
\end{gathered}
$$

The final two dominance rules by Angelelli et al. [2020] are applied to full routes rather than partial paths. Two routes considered, ω_1 and ω_2, differ in specific leg meaning that both use the same previous service s^- and following service s^+ for the specific leg. The different services used for ω_1 and ω_2 in that leg are s_1 and s_2 and have the same origin and destination. This can be written as: $o_{s_1} = o_{s_2} = d_{s^-}$ and $d_{s_1} = d_{s_2} = o_{s^+}$.

The so-called mid-leg dominance rule is applied to dedicated truck services between origin and warehouse, origin and AoL, and warehouse and AoL. Route ω_1 dominates route ω_2 if the sum of costs for the service used, cost of stocking at the current location, and cost of stocking at the following location is less than or equal for route ω_1 rather than for route ω_2. This is written as:

$$
\begin{gathered}
c_{s_1} + w_k \cdot st_{o_{s_1}} \left(e_{s_1} - (e_{s^-} + t_{s^-})\right) + w_k \cdot st_{o_{s^+}} \left(e_{s^+} - \left(e_{s_1} + t_{s_1}\right)\right) \\
\leq \\
c_{s_2} + w_k \cdot st_{o_{s_2}} \left(e_{s_2} - (e_{s^-} + t_{s^-})\right) + w_k \cdot st_{o_{s^+}} \left(e_{s^+} - \left(e_{s_2} + t_{s_2}\right)\right)
\end{gathered}
$$

The final dominance rule, called last-leg delivery rule, is implemented for services by a foreign agent. It considers penalties or gains for late or early deliveries respectively. In short, ω_1 dominates ω_2 if the total cost of using service s_1 by the foreign agent is less than the total cost of using service s_2 by the foreign agent. This rule is written as:

$$e_{s_1} \leq e_{s_2}$$
$$e_{s_1} + t_{s_1} \geq e_{s_2} + t_{s_2}$$
$$c_{s_1} + st_{o_{s_1}} \cdot w_k \cdot (e_{s_2} - e_{s_1})) + \Delta TT^{s_1} < c_{s_2} + \Delta TT^{s_2}$$

Apart from these dominance rules, Angelelli et al. [2020] include two more ways to reduce the solution space before finding a solution. The first of these is the so-called min-cost frontier. This is an implementation of cost dominance for both air services and grouped truck services. Since these services use weight ranges for service cost determination, as shown in table 4.1, different services can be the cheapest option for different ranges. It is unlikely that a specific service is the cheapest option for all available ranges. The min-cost frontier, therefore, finds the cheapest service for each range. The set of services considered is $\hat{S} \in TR_{gr} \cup AC$. This includes services with the same origin, destination, release time, and transit time. The authors create a virtual service, vs, if there is more than one service available in the set. For each range available, the unitary cost is determined as: $f_{vs,r} = \min\{f_{sr} | s \in \hat{S}\}$. Note that when services use different weight ranges, all possible ranges for a shipment should be considered. For example, service 1 uses weight ranges 0-15 and 15-30, and service 2 uses ranges 0-10, 10-20, and 20-30. To include all possible ranges, the min-cost frontier should be considered for ranges 0-10, 10-15, 15-20, and 20-30.

Finally, the authors include a heuristic rule to further reduce the solution space. This heuristic is based on the principle that air transportation services are characterized by the need of customers for fast shipment. Apart from this customer need, stocking costs at airports are relatively high compared to other intermodal options. Therefore, the authors assume that it is best for international shipments to choose a series of services where each shipment begins within a certain, restricted, time lapse after the previous service because of high stocking costs and customer needs. The heuristic that is proposed considers only a service s_2 as a successive service of s_1 if s_2 starts within a given time period after s_1 ends. This is if $e_{s_2} \in [e_{s_1} + t_{s_1}, e_{s_1} + t_{s_1} + \Delta]$. For values of Δ, the authors consider six options based on a fixed value or on the required transit time: Δ = 1 day, 2 days, 3 days, $((l_k - e_k)/2)$, $((l_k - e_k)/3)$, or $((l_k - e_k)/4)$. Once an option for Δ is chosen, the solution space can be reduced by the 'rule-of-thumb' as described above.

4.2. Column Generation

The main models that are used as the theoretical basis for the research in this report are extensively summarized in section 4.1. To be able to improve on these algorithms in terms of computational efficiency, a column generation technique will be used. This section summarizes the general formulation of CG as given by Lübbecke and Desrosiers [2005] and gives an example of how to decompose and adapt a model to be solved using a CG technique as been shown by Feillet [2010].

4.2.1. General formulation

Lübbecke and Desrosiers [2005] consider the following linear program as a standard formulation of a master problem (MP). The objective of the MP is:

$$\min \sum_{j \in J} c_j \lambda_j \tag{4.5}$$

This is subject to the following constraints:

$$\sum_{j \in J} \mathbf{a}_j \lambda_j \geq \mathbf{b}$$
$$\lambda_j \geq 0, \quad j \in J$$

Note that in this formulation, J is the entire set of columns. In each step of the algorithm, a vector of dual variables, $\mathbf{u} \geq 0$, is used to find a reduced cost coefficient given by:

$$\arg\min\{c_j - \mathbf{u}^T \mathbf{a}_j | j \in J\}$$

Since the entire set of columns J can be very large for larger instances, the technique uses a smaller subset of these columns to find a solution to the restricted master problem (RMP). This subset can be denoted as J'. The optimal solutions to the RMP are primal, $\bar{\lambda}$, and dual, $\bar{\mathbf{u}}$. Columns are added until there are no more negative reduced cost coefficients available. These coefficients are determined by:

$$\bar{c}^\star = \min\{c(\mathbf{a}) - \bar{\mathbf{u}}^T \mathbf{a} \mid \mathbf{a} \in \mathcal{A}\}$$

If there are no more negative reduced cost coefficients, and the primal and dual optimal solutions can thus be determined, and more importantly, if the dual solution $\bar{\mathbf{u}}$ is also feasible for the entire set of solutions, it is stated that this primal solution $\bar{\lambda}$ does not only solve the RMP, but also the MP.

The authors also state a general formulation of the decomposition principle by Dantzig and Wolfe [1960]. The linear program has the following objective function and constraints:

$$\min \mathbf{c}^T\mathbf{x} \tag{4.6}$$

Which is subject to:

$$\begin{aligned} A\mathbf{x} &\geq \mathbf{b} \\ X\mathbf{x} &\geq \mathbf{d} \\ \mathbf{x} &\geq 0 \end{aligned}$$

The authors further state that there is a difference between extreme points $\{\mathbf{p}_q\}_{q\in Q}$ and nonnegative combination of extreme rays $\{\mathbf{p}_r\}_{r\in R}$. Now, $\mathbf{x}$ can be written as a convex combination of the latter two:

$$\mathbf{x} = \sum_{q\in Q} \mathbf{p}_q\lambda_q + \sum_{r\in R} \mathbf{p}_r\lambda_r, \quad \sum_{q\in Q} \lambda_q = 1, \quad \lambda \in \mathbb{R}_+^{|Q|+|R|}$$

By using two linear transformations, $c_j = \mathbf{c}^T\mathbf{p}_j$ and $\mathbf{a}_j = A\mathbf{p}_j$ the linear program can be written as its so-called extensive formulation. The objective function becomes:

$$\min: \sum_{q\in Q} c_q\lambda_q + \sum_{r\in R} c_r\lambda_r \tag{4.7}$$

Which is subject to the following constraints:

$$\begin{aligned} \sum_{q\in Q} \mathbf{a}_q\lambda_q + \sum_{r\in R} \mathbf{a}_r\lambda_r &\geq \mathbf{b} \\ \sum_{q\in Q} \lambda_q &= 1 \\ \lambda &\geq \mathbf{0} \end{aligned}$$

Even though this formulation involves a large number of variables, it is said to have a substantially lower number of rows than the previous formulation. The coefficients for the reduced cost can now be determined by:

$$\bar{c}^\star := \min\left\{\left(\mathbf{c}^T - \bar{\mathbf{u}}^T A\right)\mathbf{x} - \bar{v} \mid D\mathbf{x} \geqslant \mathbf{d}, \mathbf{x} \geqslant \mathbf{0}\right\}$$

Again, if this expression gives no more negative reduced cost coefficients, the algorithm is terminated since there are no more columns that will significantly reduce the cost when present in the final solution.

4.2.2. Adapt formulation

Subsection 4.2.1 shows the general formulation used in literature to apply column generation to a model. Feillet [2010] shows how to decompose and adapt a model based on the vehicle routing problem with time windows (VRPTW). This subsection briefly explains the VRPTW and states the general model as also used by the authors. The authors show that there is a need for a decomposition of the model to apply a column generation technique. This decomposition is shown and the steps of the column generation technique are noted in this subsection.

In general, a VRPTW considers a network of customer nodes in which each customer must be visited in a certain time window by using a set of vehicles that deliver a certain demand of shipments from a depot node. It is assumed that early arrival and waiting at the customer node is allowed, but late delivery is not. The set of nodes in the network is $V = \{v_0, v_1, ..., v_n\}$, where v_0 is the depot node and the other nodes are the customers in this network. The demand of each customer is d_i and the time window in which delivery is needed is $[a_i, b_i]$. The service time needed for delivery is denoted as $serv_i$. Note that the cost and time for a vehicle to travel between two nodes are c_{ij} and t_{ij} respectively. Further, the amount of vehicles is U and the capacity of each vehicle is Q. The goal in a VRPTW is to visit each customer exactly once whilst minimizing the total cost of the movements of all used vehicles. This objective is written as:

$$\min \sum_{1 \leq u \leq U} \sum_{(v_i, v_j) \in A} c_{ij} x_{ij}^u \tag{4.8}$$

Where decision variable x_{ij}^u has value 1 if arc (v_i, v_j) is used by vehicle u. The constraints of the model are denoted and explained below:

$$\sum_{a \leq u \leq U} \sum_{\{v_j \in V | (v_i, v_j) \in A\}} x_{ij}^u \geq 1 \quad \forall v_i \in V \backslash \{v_0\} \tag{a'}$$

$$\sum_{\{v_j \in V | (v_i, v_j) \in A\}} x_{ij}^u - \sum_{\{v_j \in V | (v_j, v_i) \in A\}} x_{ji}^u = 0 \quad \forall\, (v_i \in V, 1 \leq u \leq U) \tag{b'}$$

$$\sum_{\{v_i \in V | (v_0, v_i) \in A\}} x_{0i}^u \leq 1 \quad \forall (1 \leq u \leq U) \tag{c'}$$

$$\sum_{(v_i, v_j) \in A} d_i x_{ij}^u \leq Q \quad \forall (1 \leq u \leq U) \tag{d'}$$

$$s_i^u + serv_i + c_{ij} - s_j^u + M x_{ij}^u \leq M \quad \forall \left((v_i, v_j) \in A, v_j \neq v_0, 1 \leq u \leq U \right) \tag{e'}$$

$$s_i^u + serv_i + c_{i0} - b_0 + M x_{i0}^u \leq M \quad \forall\, ((v_i, v_0) \in A, 1 \leq u \leq U) \tag{f'}$$

$$a_i \leq s_i^u \leq b_i \quad \forall\, (v_i \in V, 1 \leq u \leq U) \tag{g'}$$

$$x_{ij}^u \in \{0, 1\} \quad \forall \left((v_i, v_j) \in A, 1 \leq u \leq U \right) \tag{h'}$$

Note that s_i^u is the starting time of vehicle u coming from node i. In the model above, constraint a' ensures that each customer node is visited at least once by the used vehicles. Constraint b' is the balancing constraint that ensures for each node that each entering vehicle also leaves. Each vehicle must leave the depot node at most once which is ensured by constraint c'. The capacity of the vehicles must not be exceeded which is ensured by constraint d'. This constraint adds the demands of the customers met by a specific vehicle and makes sure this cumulative demand is less than or equal to the capacity of the vehicle. Constraints e' and f' ensure that the vehicles reach the next customer (or depot for constraint f') in time. Constraint g' adds the aspect of time window to this by ensuring that the starting time at a node i is within time window $[a_i, b_i]$. Finally, constraint h' defines the binary decision variable x_{ij}^u which is previously explained.

Given the general model for the VRPTW, Feillet [2010] state that its linear relaxation is very weak. The result is that using a branch-and-bound technique to find an optimal solution is only possible for small instances. Therefore, the authors state, it is preferred to find a new model with better relaxation and use branch-and-price methods. This new model can be found by using the so-called Dantzig-Wolfe decomposition of the given model. The authors further explain how to perform this decomposition before using the new model to apply a column generation technique.

Before rewriting the model and thus applying the Dantzig-Wolfe decomposition, a few notations are important to note. The set of feasible vehicle routes is Ω, where route r_k has cost c_k. Binary decision variable a_{ik} takes value 1 if route r_k visits customer v_i. Binary decision variable b_{ijk} takes value 1 if route r_k includes the arc from node v_i to node v_j. Further note that binary decision variable θ_k^u takes value 1 if route r_k is selected for vehicle u. The authors state that previously given constraints $b' - h'$ are Ω^U which results in the fact that the entire model can be rewritten to the following objective function and constraints:

$$\text{min: } \sum_{1 \leq u \leq U} \sum_{(v_i, v_j) \in A} c_{ij} \left(\sum_{r_k \in \Omega} b_{ijk} \theta_k^u \right) \tag{4.9}$$

Subject to:

$$\sum_{1 \leq u \leq U} \sum_{\{v_j \in V | (v_i, v_j) \in A\}} \left(\sum_{r_k \in \Omega} b_{ijk} \theta_k^u \right) \geq 1 \quad \forall \, (v_i \in V \setminus \{v_0\}) \tag{i'}$$

$$\sum_{r_k \in \Omega} \theta_k^u = 1 \quad \forall (1 \leq u \leq U) \tag{j'}$$

$$\theta_k^u \in \{0, 1\} \quad \forall \, (r_k \in \Omega, 1 \leq u \leq U) \tag{k'}$$

A further notation is given for: $c_k = \sum_{(v_i, v_j) \in A} b_{ijk} c_{ij}$, and $a_{ik} = \sum_{\{v_j \in V | (v_i, v_j) \in A\}} b_{ijk}$. Also, the notation $\theta_k = \sum_{1 \leq u \leq U} \theta_k^u$ is used to rewrite this model to the following objective function and constraints:

$$\text{min: } \sum_{r_k \in \Omega} c_k \theta_k \tag{4.10}$$

Subject to:

$$\sum_{r_k \in \Omega} a_{ik} \theta_k \geq 1 \quad \forall \, (v_i \in V \setminus \{v_0\}) \tag{l'}$$

$$\sum_{r_k \in \Omega} \theta_k \leq U \tag{m'}$$

$$\theta_k \in \mathbb{N} \quad \forall \, (r_k \in \Omega) \tag{n'}$$

In this formulation, constraint l' ensures that each customer node is on the selected routes. This constraint corresponds to constraint a' in the original VRPTW formulation. The total number of vehicles used in the final solution is limited by constraint m'. Note that the decision variable θ_k takes integer values, opposed to the binary values of θ_k^u, as is defined by constraint n'.

Since the set of routes Ω grows exponentially for larger instances, a column generation technique is used to solve this model which makes it a branch-and-price approach rather than a branch-and-bound approach. As explained in subsection 4.2.1, not the entire set of feasible routes or columns is used to perform this CG. The master problem (MP) is defined as the linear relaxation of the model given above in equation 4.10 and constraints $l' - n'$. The subset Ω_1 is used in the so-called restricted master problem (RMP). This RMP is formulated as:

$$\text{min: } \sum_{r_k \in \Omega_1} c_k \theta_k \tag{4.11}$$

Subject to:

$$\sum_{r_k \in \Omega_1} a_{ik} \theta_k \geq 1 \quad \forall \, (v_i \in V \setminus \{v_0\}) \tag{o'}$$

$$\sum_{r_k \in \Omega_1} \theta_k \leq U \tag{p'}$$

$$\theta_k \geq 0 \quad \forall \, (r_k \in \Omega_1) \tag{q'}$$

To extend the subset of feasible routes, a subproblem is formulated in which the reduced cost of a potential route is determined. Before being able to do this, a dual program of the RMP must be formulated with dual variables related to the constraints in the RMP. This results in a nonnegative dual variable, λ_i, which is related to constraint o' and a nonpositive dual variable, λ_0, which is related to constraint p'. The dual program, $D(\Omega_1)$ is given by the following objective function and constraints:

$$\text{max:} \sum_{v_i \in V \backslash \{v_0\}} \lambda_i + U\lambda_0 \tag{4.12}$$

Subject to:

$$\sum_{v_i \in V \backslash \{v_0\}} a_{ik}\lambda_i + \lambda_0 \leq c_k \quad \forall\,(r_k \in \Omega_1) \tag{r'}$$

$$\lambda_i \geq 0 \quad \forall\,(v_i \in V \backslash \{v_0\}) \tag{s'}$$

$$\lambda_0 \leq 0 \tag{t'}$$

The goal is to find an optimal solution to the RMP. When this solution is found, the corresponding values for the dual variables, λ^*, give the optimal solution for the dual program $D(\Omega_1)$. A very important step is noting that if this solution is also feasible for the dual program with the entire set of routes, $D(\Omega)$, then the solution is also assumed to solve the MP.

The next step in this CG approach is focused on adding more feasible and cost reductive routes to the subset Ω_1 for the subproblem. Note that these new routes $r_k \in \Omega \backslash \Omega_1$. The search for these routes is done according to the following condition:

$$c_k - \sum_{v_i \in V \backslash \{v_0\}} a_{ik}\lambda_i^* - \lambda_0^* < 0$$

Which can be written as:

$$\sum_{(v_i, v_j) \in A} b_{ijk}\left(c_{ij} - \lambda_i^*\right) < 0 \tag{4.13}$$

Feillet [2010] note this expression shows that the subproblem can be categorized as an elementary shortest path problem with resource constraints (ESSPPRC). New elementary paths must be found with negative costs to improve the solution. This must be done whilst still meeting the capacity and time window constraints. The authors briefly summarize a solution procedure based on dynamic programming. This procedure is an extended version of the classical Bellman's algorithm. The search for new routes continues until a sufficient number of cost reductive columns has been added to Ω or when there are no more columns to add with negative reduced cost. This final set of columns is used to find a feasible solution for the RMP and equivalently to the MP.

This concludes the tutorial by Feillet [2010] on how to adapt a model, in this case a model for the VRPTW, to be solved using a CG technique. The aim of the research in this report is to use a similar decomposition technique to adapt the formulation of the ATFFSP. In addition, the aim is to use a similar CG approach to find solutions to the problem in a computationally more efficient way than yet shown in the available literature. Also, this report focuses on a real-life implementation of the ATFFSP. This involves a number of changes to the definition and formulation of the ATFFSP. The specifics are stated in chapter 5. This chapter also defines the research questions and the proposed method to answer these questions in more detail.

5

Use Case

Chapter 3 has given an overview of the research in available literature which has relevance to the topic of this report. After reviewing this available literature, chapter 4 elaborated on the most important literature to be used as a theoretical basis. Before using this theoretical basis to formulate a method, this chapter focuses on the specific research questions to be answered in section 5.1 and proposes a method and more importantly states the differences between the real-life case of this report and current literature in section 5.2.

5.1. Research Questions

As stated previously in this report, the proposed research will aim to find a column generation based algorithm that solves the air transportation freight forwarder service problem. The aim is to further improve the work by Archetti and Peirano [2020] and Angelelli et al. [2020] done on this topic in terms of computational efficiency. The research has practical relevance since real-life data from an industrial partner will be used to improve a real-life network with this model. To conclude, this research aims to develop an optimization tool to improve the cost minimization for freight forwarders in their service of transporting shipments from origin to the destination including intermediate facilities such as warehouses or airports. The main research question can thereby be written as:

> "How to minimize the cost of services used by freight forwarding companies whilst fulfilling demand of transport of shipments from origin to destination with the presence of intermediate facilities?"

To be able to find an optimal solution to this problem from a practical point of view, the network must be modelled as accurately as possible to the real-life implementation by the industrial partner of this research. Data on available services must be gathered including for example capacity of the vehicles used, time windows of operation of the vehicles used, cost of operation of the vehicles used, cost reduction when consolidating shipments, and much more. A sub-question for the research considered in this report can be written as:

> "What data is needed to model the network as accurately as possible to the real-life implementation by an industrial partner to find a solution with practical relevance?"

With the relevant data available, the model should be formulated correctly to be in line with the actual service network used by the industrial partner. This is important since this provides the possibility to compare the solutions of the model to the solutions implemented by the company. This can be written as a sub-question to the main research question:

> "How to model the total service network used by the industrial partner using the relevant real-life data?"

Once an accurate model is formulated using this real-life data, the model should be optimized to minimize the cost for the freight forwarding company. It is important that all relevant constraints are identified and formulated. An efficient algorithm must be formulated to find a cost-minimizing solution within reasonable computational time. A sub-question can be written as:

> “How to formulate an efficient algorithm which minimizes the total cost for a freight forwarding company by optimizing the selection of available services in the network whilst meeting the total demand of shipments from origin to destination?”

To conclude, the main objective of the research considered in this report is to achieve the development of an optimization tool to improve the cost minimization for freight forwarders in their service of transporting shipments from origin to the destination including intermediate facilities. This will be achieved by implementing real-life data, gathered from an industrial partner, into the formulation of an accurate network that represents the actual network of this partner and finding an optimal solution that minimizes the costs of services used in this network, whilst meeting the total demand.

5.2. Proposed method

As shown in chapter 4, the models proposed by both Archetti and Peirano [2020] and Angelelli et al. [2020] give a different approach to solve the ATFFSP. The first approach is based on solving a MILP which takes significant computational time for larger instances. The second approach uses a subproblem to find feasible routes first and uses a second problem to find the optimal set of these routes to solve the ATFFSP. Since this approach uses just a subset of the solutions and is subject to a number of acceleration techniques, Angelelli et al. [2020] show that their algorithm needs less computational time than the MILP approach by Archetti and Peirano [2020].

As a basis for further improvement, this report proposes an approach that is mainly based on the two-level approach of first finding a set of feasible routes and then deciding on the optimal combination of these routes to deliver all shipments. This is chosen because of two main reasons. Firstly, this approach is shown to be the most effective of the two discussed in chapter 4 in terms of computational time. Secondly, the tutorial by Feillet [2010], as shown in chapter 4, gives a clear explanation on how to convert a model to be solved using a CG technique. The authors also assume a set of routes to be available and find cost reductive columns or routes to add to the subset which is used to solve the main problem. This can be very helpful in setting up a similar approach for the ATFFSP instead of the VRPTW. The route construction algorithm including the acceleration techniques as proposed by Angelelli et al. [2020] can be used in the improved model to find a set of feasible routes. The set-partitioning formulation by the authors can be adapted to involve a column generation technique to find the optimal set of routes more effectively in terms of computational time. Before this can be implemented, the formulation of the ATFFSP has to be adapted to accurately match the real-life implementation by the industrial partner.

As said, there are a number of differences between the formulation of the ATFFSP used in this report and the actual implementation of real-life companies. These differences include the number and types of (intermediate) facilities, the presence of processing time, the possible disruptions in a network, and the main objective. These differences will be further explained below.

In the network, there is a possibility of using one or more facilities between the origin facility and the destination facility. In the ATFFSP formulation, only a direct connection between AoL and AoD was allowed. The new situation calls for the addition of more intermediate steps/ facilities between these airports. This can be compared to the presence of warehouses in the ATFFSP formulation. These warehouses could be used, but not necessarily. A similar formulation can be written for the facilities in the total set of facilities excluding the origin and destination facility. By doing this, a shipment does not necessarily need to be transported directly from AoL to AoD, but a more cost or time effective can be chosen which might include an intermediate stop.

Secondly, the ATFFSP formulation does not take the processing time at a facility into account. In the real-life implementation considered in this report, there is in fact an amount of processing time between the inbound and outbound of a shipment at a facility. In the ATFFSP formulation, two options are direct outbound of a shipment or stocking the shipment at the facility at a certain cost. The latter option could be useful to implement processing time at facilities. One could describe stocking cost as zero as long as it is within the expected processing time. The costs are only charged after the processing time. For stocking time less than the minimal processing time, one could add a significantly high penalty to ensure this is not possible in the solution. This would make sure that shipments are at the facility for at least the minimal processing time.

An important element to include in the formulation is the presence of possible disruptions and therefore changes in possible routes and time windows. Disruptions may appear whilst shipments are transported over the network. These disruptions affect not only the specific vehicle movement (and included shipments) which is subject to delay for example but the entire network. Therefore, two things are important. First of all, the current locations of shipments/ active movements must be included in the formulation. Secondly, the time intervals must be changed for the affected shipments/ movements. The origin location of a shipment can be changed based on the current location and the network can be evaluated again. More importantly, the time of arrival at the next facility or at the destination is changed. By implementing these elements into the formulation, the network can be revised after a disruption and a more accurate solution can be found.

Finally, the main objective for freight forwarding companies and thus the objective in problems concerning the optimisation of their operations such as the ATFFSP is to minimise cost. In the actual problem considered by the industrial partner in this report, the main focus is in line with express companies: minimise the time used to deliver the shipments. Since time equals money for these companies, this objective can also be rephrased to minimise time. Another option is to formulate the model bi-objective and set to focus on minimising the time before minimising the cost. This adaption has to be made to the current formulation to find a time minimizing solution to deliver all shipments. This solution can then be compared to the solution which was historically implemented by the industrial partner.

After implementing these changes into the formulation, the routes can be constructed and a solution can be found using a column generation technique. Solutions can be compared to the actually implemented solutions by the partner and the computational time used can be compared to the computational time when the original set-partitioning formulation was used. These results will conclude the research that is proposed in this report.

Bibliography

A. Aguezzoul. Third-party logistics selection problem: A literature review on criteria and methods, 2014. ISSN 03050483.

E. Angelelli, C. Archetti, and L. Peirano. A matheuristic for the air transportation freight forwarder service problem. *Computers and Operations Research*, 123, 11 2020. ISSN 03050548. doi: 10.1016/j.cor.2020.105002.

C. Archetti and L. Peirano. Air intermodal freight transportation: The freight forwarder service problem. *Omega (United Kingdom)*, 94, 7 2020. ISSN 03050483. doi: 10.1016/j.omega.2019.02.009.

C. Barnhart and R. R. Schneur. Air network design for express shipment service. *Operations Research*, 44:852–863, 1996. ISSN 0030364X. doi: 10.1287/opre.44.6.852.

C. Barnhart, D. Kim, and K. Ware. Network design for express shipment delivery, 2002.

I. Bello, H. Pham, Q. V. Le, M. Norouzi, and S. Bengio. Neural combinatorial optimization with reinforcement learning. 11 2016. URL http://arxiv.org/abs/1611.09940.

J. M. V. D. Carvalho. Exact solution of bin-packing problems using column generation and branch-and-bound, 1999.

P. Chen, Y. Guo, A. Lim, and B. Rodrigues. Multiple crossdocks with inventory and time windows. *Computers and Operations Research*, 33:43–63, 1 2006. ISSN 03050548. doi: 10.1016/j.cor.2004.06.002.

T. G. Crainic, G. Perboli, and R. Tadei. Ts2pack: A two-level tabu search for the three-dimensional bin packing problem. *European Journal of Operational Research*, 195:744–760, 6 2009. ISSN 03772217. doi: 10.1016/j.ejor.2007.06.063.

K. Daeki, C. Barnhart, K. Ware, and G. Reinhardt. Multimodal express package delivery: A service network design application. *Transportation Science*, 33:391–407, 1999. ISSN 00411655. doi: 10.1287/trsc.33.4.391.

G. B. Dantzig and P. Wolfe. Decomposition principle for linear programs. *Operations Research*, 8: 101–111, 2 1960. ISSN 0030-364X. doi: 10.1287/opre.8.1.101.

H. Dyckhoff. A typology of cutting and packing problems, 1990.

D. Feillet. A tutorial on column generation and branch-and-price for vehicle routing problems. *4OR*, 8: 407–424, 2010. ISSN 16142411. doi: 10.1007/s10288-010-0130-z.

L. R. Ford and D. R. Fulkerson. A suggested computation for maximal multi-commodity network flows. *Management Science*, 5:97–101, 10 1958. ISSN 0025-1909. doi: 10.1287/mnsc.5.1.97.

P. C. Gilmore and R. E. Gomory. A linear programming approach to the cutting stock problem. 2, 1963.

J. Gondzio, P. González-Brevis, and P. Munari. Large-scale optimization with the primal-dual column generation method. *Mathematical Programming Computation*, 8:47–82, 3 2016. ISSN 18672957. doi: 10.1007/s12532-015-0090-6.

G. Guastaroba, M. G. Speranza, and D. Vigo. Intermediate facilities in freight transportation planning: A survey, 2016. ISSN 15265447.

H. Hu, X. Zhang, X. Yan, L. Wang, and Y. Xu. Solving a new 3d bin packing problem with deep reinforcement learning method. 8 2017. URL http://arxiv.org/abs/1708.05930.

H. Jung, F. F. Chen, and B. Jeong. Decentralized supply chain planning framework for third party logistics partnership. *Computers and Industrial Engineering*, 55:348–364, 9 2008. ISSN 03608352. doi: 10.1016/j.cie.2007.12.017.

M. A. Krajewska and H. Kopfer. Transportation planning in freight forwarding companies. tabu search algorithm for the integrated operational transportation planning problem. *European Journal of Operational Research*, 197:741–751, 9 2009. ISSN 03772217. doi: 10.1016/j.ejor.2008.06.042.

S. D. Lapierre, A. B. Ruiz, and P. Soriano. Designing distribution networks: Formulations and solutions heuristic. *Transportation Science*, 38:174–187, 2004. ISSN 00411655. doi: 10.1287/trsc.1030.0064.

A. Lim, Z. Miao, B. Rodrigues, and Z. Xu. Transshipment through crossdocks with inventory and time windows. *Naval Research Logistics*, 52:724–733, 12 2005. ISSN 0894069X. doi: 10.1002/nav.20113.

C. C. Lin and S. H. Chen. The hierarchical network design problem for time-definite express common carriers. *Transportation Research Part B: Methodological*, 38:271–283, 2004. ISSN 01912615. doi: 10.1016/S0191-2615(03)00013-4.

C. C. Lin and S. H. Chen. An integral constrained generalized hub-and-spoke network design problem. *Transportation Research Part E: Logistics and Transportation Review*, 44:986–1003, 2008. ISSN 13665545. doi: 10.1016/j.tre.2008.02.001.

C. C. Lin, Y. J. Lin, and D. Y. Lin. The economic effects of center-to-center directs on hub-and-spoke networks for air express common carriers. *Journal of Air Transport Management*, 9:255–265, 2003. ISSN 09696997. doi: 10.1016/S0969-6997(03)00019-X.

M. E. Lübbecke and J. Desrosiers. Selected topics in column generation. *Operations Research*, 53: 1007–1023, 11 2005. ISSN 0030364X. doi: 10.1287/opre.1050.0234.

M. R. Marjani, S. M. M. Husseini, and B. Karimi. Bi-objective heuristics for multi-item freights distribution planning problem in crossdocking networks. *International Journal of Advanced Manufacturing Technology*, 58:1201–1216, 2 2012. ISSN 02683768. doi: 10.1007/s00170-011-3444-x.

Z. Miao, K. Fu, and F. Yang. A hybrid genetic algorithm for the multiple crossdocks problem. *Mathematical Problems in Engineering*, 2012, 2012a. ISSN 1024123X. doi: 10.1155/2012/316908.

Z. Miao, F. Yang, K. Fu, and D. Xu. Transshipment service through crossdocks with both soft and hard time windows. *Annals of Operations Research*, 192:21–47, 1 2012b. ISSN 02545330. doi: 10.1007/s10479-010-0780-4.

I. Muter, J. F. Cordeau, and G. Laporte. A branch-and-price algorithm for the multidepot vehicle routing problem with interdepot routes. *Transportation Science*, 48:425–441, 8 2014. ISSN 15265447. doi: 10.1287/trsc.2013.0489.

A. Orden. The transhipment problem. *Management Science*, 2:276–285, 4 1956. ISSN 0025-1909. doi: 10.1287/mnsc.2.3.276.

C. Paquay, M. Schyns, and S. Limbourg. A mixed integer programming formulation for the three-dimensional bin packing problem deriving from an air cargo application. *International Transactions in Operational Research*, 23:187–213, 1 2016. ISSN 14753995. doi: 10.1111/itor.12111.

M. Steadieseifi, N. P. Dellaert, W. Nuijten, T. V. Woensel, and R. Raoufi. Multimodal freight transportation planning: A literature review. *European Journal of Operational Research*, 233:1–15, 2 2014. ISSN 03772217. doi: 10.1016/j.ejor.2013.06.055.

J. C. Tyan, F.-K. Wang, and T. C. Du. An evaluation of freight consolidation policies in global third party logistics, 2003. URL www.sciencedirect.comwww.elsevier.com/locate/dsw.

P. H. Vance, C. Barnhart, E. L. Johnson, and G. L. Nemhauser. Solving binary cutting stock problems by column generation and branch-and-bound*, 1994.

J. Westerlund, L. G. Papageorgiou, and T. Westerlimd. A problem formulation for optimal mixed-sized box packing, 2005.

G. Wäscher, H. Haußner, and H. Schumann. An improved typology of cutting and packing problems. *European Journal of Operational Research*, 183:1109–1130, 12 2007. ISSN 03772217. doi: 10.1016/j.ejor.2005.12.047.

R. Yesodha. A comparative study on heuristic procedures to solve bin packing problems. *International Journal in Foundations of Computer Science Technology*, 2:37–49, 11 2012. doi: 10.5121/ijfcst.2012.2603.

H. Zhao, Q. She, C. Zhu, Y. Yang, and K. Xu. Online 3d bin packing with constrained deep reinforcement learning, 2021. URL www.aaai.org.

III

Supporting work

1
Processing Times

The RCA part of the model selects feasible sets of sequential movements that follow historical or alternative paths. Movements in the movement schedule are considered for each section of the path. As a movement is selected for a section of the path, a next movement is feasible for the next section of the path if the scheduled departure time is after the arrival time of the previous movement including added processing time at the facility. The processing time at a facility is considered since a shipment is not available for departure at the exact moment the previous movement arrives, but is subject to the processing at the facility first. Historical data on processing times at specific facilities for specific arrival times is used to assume a processing time for a new shipment with a specific arrival time at the facility. The historical processing times are stored for each specific facility and arrival hour of the week. The hour of the week is formulated starting from zero at midnight on Sunday of the previous week to 168 $(7 * 24)$ at midnight on Sunday of the current week.

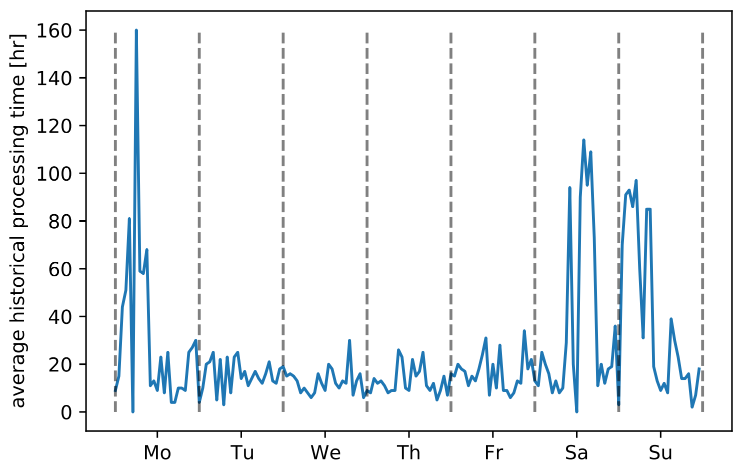


Figure 1.1: Average historical processing times for different hours of the week for facility A.

The dependence on both facility and hour of the week is shown in Figures 1.1 and 1.2. These figures show the average historical processing time for different hours of the week for two different facilities. A dependence on working days or weekend days is also clearly visible. Facility A in Figure 1.1 shows significantly higher average processing times for Saturday, Sunday, and early Monday compared to the working days of the week. Facility B in Figure 1.2 also shows dependence on the hour of the day, resulting in higher processing times on average around midnight (at the vertical dotted lines) than around noon (between the vertical dotted lines). Both dependencies are taken into account by assuming processing times based on the historical hour of the week. Algorithm 1 shows the implemented method to find an assumed processing time based on the stored historical data.

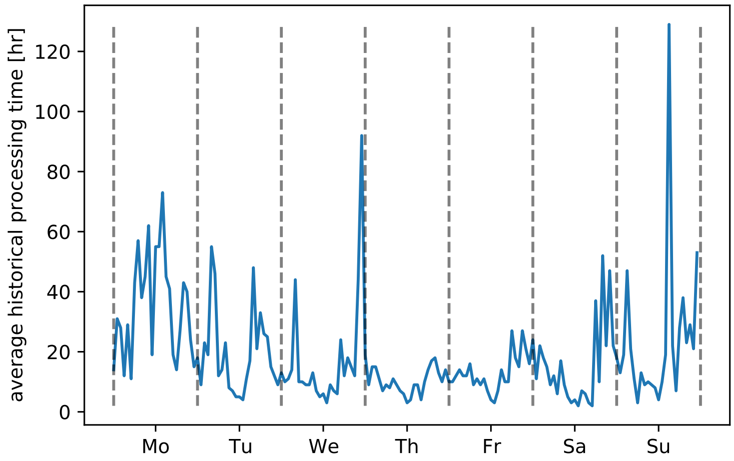


Figure 1.2: Average historical processing times for different hours of the week for facility B.

An assumed processing time is computed for each hour of the week ($hour$) for each facility. This is done to directly assign a processing time to each shipment arriving in each available facility at each possible hour of the week in the RCA. The algorithm assigns weights ($weights$) to the historic processing times at a specific facility (fac) and calculates the weighted average ($pt\_assumed$) of these processing times (pts). The algorithm assigns higher weights to historic processing times for which the arrival time (hr) is closer to the current arrival time ($hour$). This is done by computing the absolute difference between the historical arrival time and current arrival time ($|hour - hr|$) for each of the historical arrival times. The inverse of each value is taken to assign a higher weight to processing times for which hr is closer to $hour$ and therefore more relevant. The weighted average is computed for each hour of the week for each facility. The computed corresponding processing times can be added to the arrival times of movements in the movement schedule to find the earliest time of departure for a shipment arriving at a facility on a specific movement.

Algorithm 1 Processing time for specific arrival times at specific facilities

Input list of all facilities: $facs$, list of all hours of the week: hrs_wk, dictionary with historic processing times at specific facilities for specific arrival times: $pt\_facilities$
Initialize dictionary to store assumed processing times at specific facilities at specific arrival times: $pt\_facs\_hrs\_wk$
for fac in $facs$ **do**
  for $hour$ in hrs_wk **do**
    $diff\_hrs$ = empty list
    pts = empty list
    $pt\_fac = pt\_facilities[fac]$
    for hr,pt in pt_fac **do**
      add $|hour - hr|$ to $diff\_hrs$
      add pt to pts
    end for
    $weights = \frac{1}{diff\_hrs}$
    $pt\_assumed = \frac{sum(weights * pts)}{sum(weights)}$
    add $pt\_assumed$ to $pt\_facs\_hrs\_wk$ for $fac,hour$
  end for
end for
return $pt\_facs\_hrs\_wk$
Output dictionary with assumed processing times for every hour of the week at specific facilities

2
Available Factors

The available capacity of historical movements is used to make an assumption of the available capacity of future movements in the movement schedule. This available capacity is used in the chance-constrained approach discussed in the scientific paper. In the scope of the research proposed in the scientific paper, the available factor (AF) of a historic movement is determined as the factor of the known capacity that is loaded at the origin of the movement and unloaded at the destination of the movement. This is the relevant factor in the research since the formulation is based on loading a shipment on a movement at the origin and unloading at the destination. The weight of shipments on the historical movement that has been loaded prior to the origin and/ or unloaded after the destination is not part of the available capacity. The computed AFs are stored based on characteristics of the origin of the movement (org), the destination of the movement (dst), the vehicle type (vht), and the scheduled time of departure (std). This data is used to assume an AF for each of the movements in the movement schedule according to algorithm 2.

Input to algorithm 2 is the movement schedule ($mvmt\_data$) and the stored historical AFs based on mentioned movement characteristics (af_odvs). The algorithm considers each movement ($mvmt$) individually and gathers the characteristics for the org, dst, vht, and std of the $mvmt$. If an AF has previously been determined and stored in af_odvs for the specific combination of these characteristics, this AF is selected for the movement and added to the list of determined AFs for each movement in the movement schedule, af_mvmts. If the combination is unknown, a subset of the $mvmt\_data\_hist$ is selected with as many equal characteristics as available. Preferably a subset with equal org, dst, vht, and std. If no additional movements are available with these four characteristics, a subset with equal org, dst, and vht is selected. If no additional movements are available with these characteristics, a subset with equal org and destination dst is selected. Finally, if no additional movements are available for this od-pair, the total $mvmt\_data\_hist$ set is used as the subset. The selected subset is used to find an average AF which will be stored for the specific movement. The AFs for each of the movements in the schedule is used to determine the available capacity based on the strategy in the chance-constrained approach.

Algorithm 2 Determination of available factor (AF) for each movement in the movement schedule

```
Input movement schedule including characteristics on each movement: mvmt_data, historical movement data including capacity and (un)loading characteristics: mvmt_data_hist, dictionary with known AF for specific origin (org), destination (dst), vehicle type (vht), and departure time (std) combinations: af_odvs
Initialize af_mvmts = empty list
for mvmt in mvmt_data do
    odvs = org, dst, vht, std of mvmt
    odv = org, dst, vht of mvmt
    od = org, dst of mvmt
    if odvs in af_odvs then
        AF = af_odvs[odvs]
    else
        if movements in mvmt_data_hist with equal odvs then
            subset = subset of mvmt_data_hist with equal odvs
        else if movements in mvmt_data_hist with equal odv then
            subset = subset of mvmt_data_hist with equal odv
        else if movements in mvmt_data_hist with equal od then
            subset = subset of mvmt_data_hist with equal od
        else
            subset = mvmt_data_hist
        end if
        AF = available factor based on subset data
    end if
    add AF to af_mvmts for mvmt
end for
return af_mvmts
Output assumed available factor (AF) for each movement in the movement schedule
```

3

Dynamic Approach

The original and adapted formulations of the combined RCA and SPM in the scientific paper are based on a static implementation of the AETP. The model is input with shipment and movement data for a given time period. Historical and alternative paths are determined based on historical shipment data and given movement schedules, respectively. A set of routes is constructed for each shipment that is included in the input data in the RCA part of the model. The SPM then selects the optimal set of routes to transport all shipments from origin to destination. This process is executed once in sequential order to find the optimal solution for a given input set of shipments. A real-life implementation for air express couriers calls for a more dynamic implementation of the problem due to the dynamic input of new shipments based on customer demand. The courier selects a pickup time interval for new shipments and inputs the subset of shipments in this interval to the model. This dynamic formulation is implemented in the model and discussed below.

The main adaption to the static model is the implementation of the sequential running of both the RCA and SPM part of the model for each subset of shipments rather than running both parts once for the entire set of shipments. The variables built in the pre-processing part prior to the RCA can be extended using the new shipment data in each iteration. Important variables to update are the processing times (PTs) at facilities, the available factors (AFs) for specific movements, the historical paths, the alternative paths, and the capacity of specific movements. Including more data on both historical PTs and historical AFs ensures more accurate assumptions for future values used in the model. Adding the implemented paths for each iteration to the set of historical paths enlarges this set of paths. A larger amount of unique paths to follow in the RCA part of the model results in a larger amount of unique routes in the solution space of the shipments and therefore increases the potential improvement for future implementations. Also, adding the implemented paths to the historical set of paths influences the occurrence of each path, which is relevant in both the likeliness and bi-objective implementation of the model, discussed in the scientific paper. The alternative paths found for shipments with previously unknown od-pair must be added to the set of alternative paths. This prevents the computation of alternative paths again for future shipments with equal od-pair, thus reducing the computational time for future computation. Finally, the capacity of each movement in the movement schedule must be updated after each iteration to account for the added weight of shipments selected for specific movements in that iteration.

The requirement for this dynamic implementation is that the combined model is solved for each subset of shipments (batch) before the next time interval and therefore next input of shipments. Meeting this requirement allows the model to directly solve for a new input of shipments, which is preferred by the air express courier. Following this upper bound constraint for the computational time, different time intervals are tested for batches in a test data set. The computational time for each batch is stored and compared to the set batch time interval. Figure 3.1 shows the minimum, mean, and maximum computational time for each of the iterations in the test data set for different batch time intervals. Since the computational time is upper bound constraint following mentioned requirement, the maximum computational time is relevant and must not exceed the batch time interval. This is visualized in Figure 3.2, where the batch time is added as a computational time limit for reference. The figure clearly shows that the model solves the dynamic problem in less computational time compared to the set batch time interval for each interval time. Therefore, the requirement of the air express company is met for each set interval time.

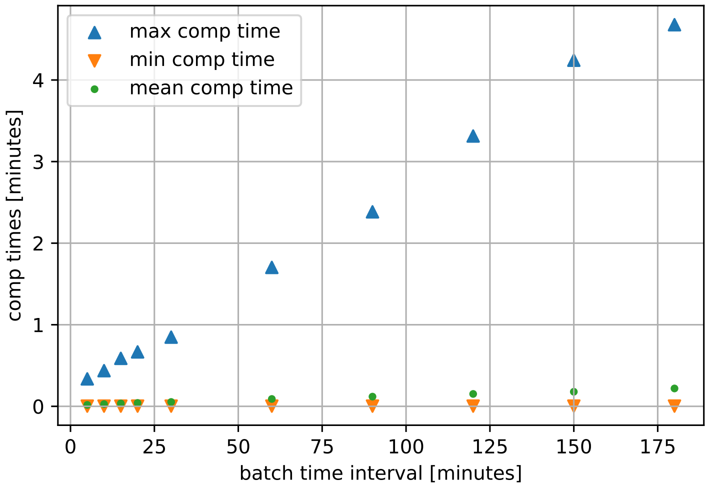


Figure 3.1: Characteristics computational time for different set batch time intervals on a test data set.

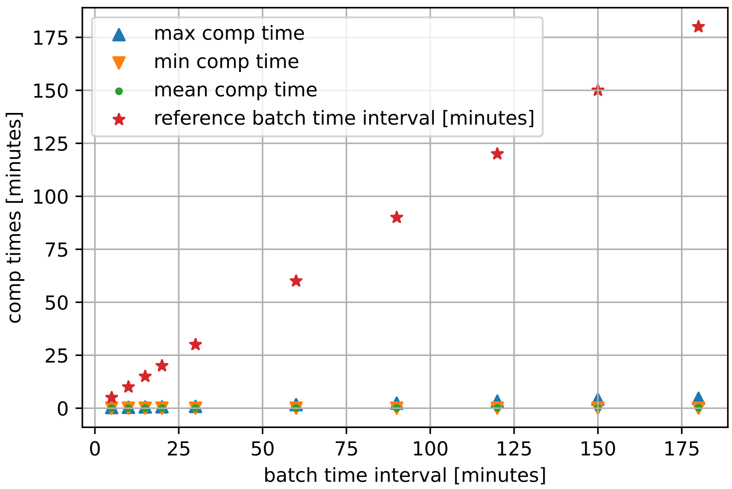


Figure 3.2: Characteristics computational time for different set batch time intervals on a test data set including reference batch time interval.

As the model performance meets the requirement of an air express company in terms of computational time for each batch time interval, an optimal value is selected for this batch time interval. This selection is done by finding the minimal total computational time needed to solve the model for all batches in a specific time period. Since the computational time of the model depends directly on the batch size, the average batch size is computed for different batch time intervals.

Figure 3.3 shows the computational time to solve the model for all batches for different batch sizes. The figure shows that the total computational time is minimal for batch sizes between 100 shipments and 1000 shipments. Batch sizes around 200 and 500 shipments show the lowest total computational time so are selected as preferred bounds for the average batch size for the dynamic model. As previously mentioned, air express companies find a solution for a set of shipments after a preset time interval. Therefore, a corresponding batch time interval has to be selected that meets the requirement of including an average number of

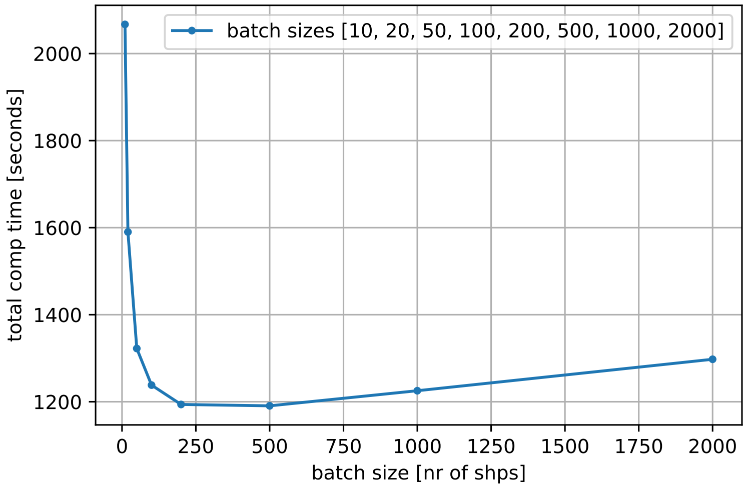


Figure 3.3: Ratio of computational time per batch size.

shipments between 200 and 500.

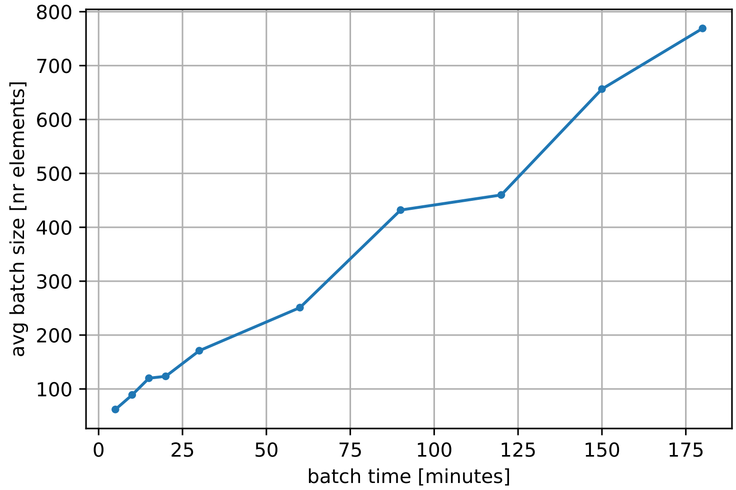


Figure 3.4: Average batch size in the number of shipments for each batch time interval.

Figure 3.4 shows the average batch size for different set batch time intervals. The figure shows that a batch time interval between 60 minutes and 120 minutes can be selected to fulfill the requirement by the air express company for the average batch size to be between 200 and 500 shipments. From a dynamic point of view, an air express company prefers smaller batch time intervals in order to quickly solve the model and select feasible routes for a set of shipments. Therefore, a batch time interval of 60 minutes is selected in this research.

4
Path Lengths

The different strategies implemented for the different approaches discussed in the scientific paper influence the solution found by the model for a specific input set of shipments. This is caused by differences in the objective or differences in the formulation or values of constraints. The maximum level of complexity for both historic and alternative paths included in the model is set to 6, based on the distribution of path lengths implemented in historical data. The distribution of path lengths of the routes selected for shipments for different approaches is shown in Figures 4.1, 4.2, 4.3, and 4.4. Both the on-time and the late delivery implementation are considered for each approach and compared to the actual historic implementation of the shipments.

Figure 4.1 shows the distribution of path lengths for the solution to the original formulation of the AETP. The historical implementation shows a normal distribution with significantly higher occurrence for level 3 paths and a similar occurrence for levels 2 and 4. Levels 1, 5, and 6 show a significantly lower occurrence. The figure shows that the air express courier selects a level 3 route, so with two intermediate facilities between origin and destination, for over 50% of the shipments. The solution of the on-time delivery formulation of the model shows similar behavior of selecting a level 3 route for the majority of the shipments. In this case, a route with two intermediate stops is selected for nearly two-thirds of the shipments. Following the objective of the AETP of selecting the routes with the lowest transportation time for a shipment, a level 1 route is selected for just over a quarter of the shipments. This shows that the model selects lower-level routes on average than the historical implementation. The distribution of path length in the solution for the late delivery formulation of the model shows similar behavior to the on-time delivery formulation. However, allowing late delivery in this formulation adds feasible routes to the solution space with a higher transportation time. This results in a higher occurrence of higher-level paths in the solution compared to the on-time delivery formulation. This is shown in the figure by a lower factor for levels 1, 2, and 3 whilst a higher factor is related to levels 4 and 5.

Figure 4.2 shows the distribution of path lengths for the solution to the model using the likeliness approach for both an on-time and a late delivery implementation. Again, these implementations are compared to the historical implementation by the air express courier. Similar behavior is shown in the distributions in Figure 4.1. The solutions show a higher occurrence of lower-level paths compared to the historical implementation. The solution of the late delivery implementation again shows slightly lower occurrence for lower-level paths but slightly higher occurrence for higher-level paths compared to the on-time delivery solution. Similar behavior is also shown for the chance-constrained approach in Figure 4.3. The preference of the model to select less costly, lower-level routes is shown by a higher occurrence of lower-level paths compared to the historical implementation. Also, the difference between the late delivery implementation and the on-time delivery implementation is shown by the lower occurrence of lower-level paths and higher occurrence of higher-level paths of the late delivery solution compared to the on-time delivery solution.

As mentioned in the scientific paper, the bi-objective approach implements a trade-off between the selection of the most efficient set of routes according to the original implementation and the selection of the most likely set of routes according to the likeliness implementation. Figure 4.4 shows the distribution of path lengths for the solution of the model using the bi-objective approach for both an on-time and a late delivery implementa-

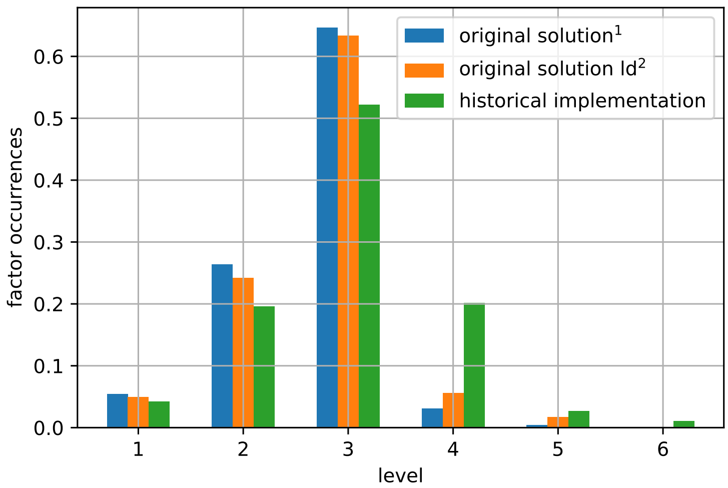


Figure 4.1: Path lengths for selected routes in the historical implementation and in the solution for both the on-time[1] and late delivery[2] formulation of the model using the original approach.

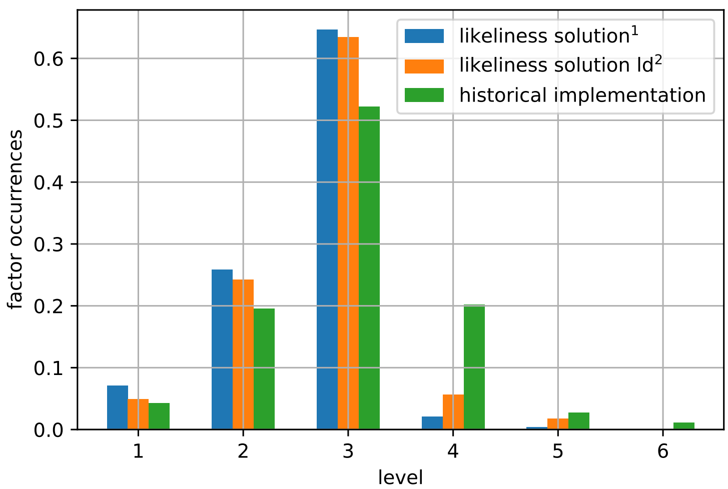


Figure 4.2: Path lengths for selected routes in the historical implementation and in the solution for both the on-time[1] and late delivery[2] formulation of the model using the likeliness approach.

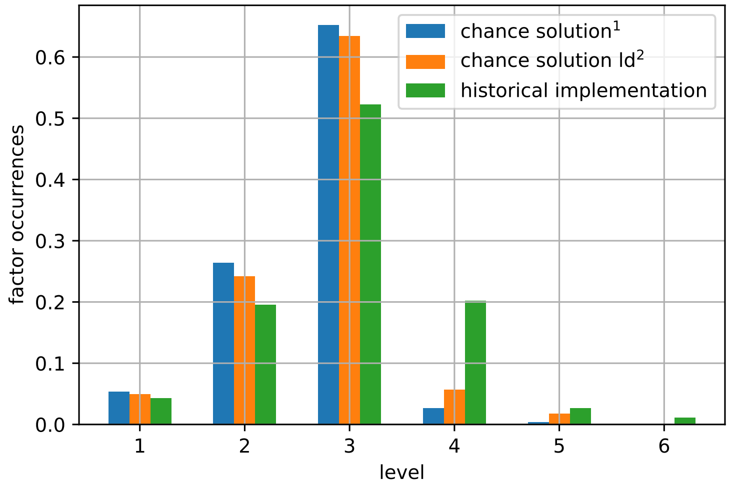


Figure 4.3: Path lengths for selected routes in the historical implementation and in the solution for both the on-time[1] and late delivery[2] formulation of the model using the chance-constrained approach.

tion. An average is taken for the different strategies implemented based on different γ-values. Figures 4.1 and 4.2 show the distribution for the original and likeliness approach, respectively. Since these two approaches represent the extreme values of the bi-objective approach, the distribution averages shown in Figure 4.4 have values between the values shown for the two separate approaches. As mentioned and shown by comparing Figures 4.1 and 4.2, the distribution of the two approaches is similar for both the on-time and late delivery implementation. This results in a similar distribution for the bi-objective approach. Again, the model shows a preference to select less costly, lower-level routes compared to the historical implementation as is shown by the higher occurrence of lower-level paths. The difference between the on-time and late delivery implementation is shown by the slightly lower occurrence for lower-level paths and higher occurrence for higher-level paths in the late delivery implementation compared to the on-time delivery implementation.

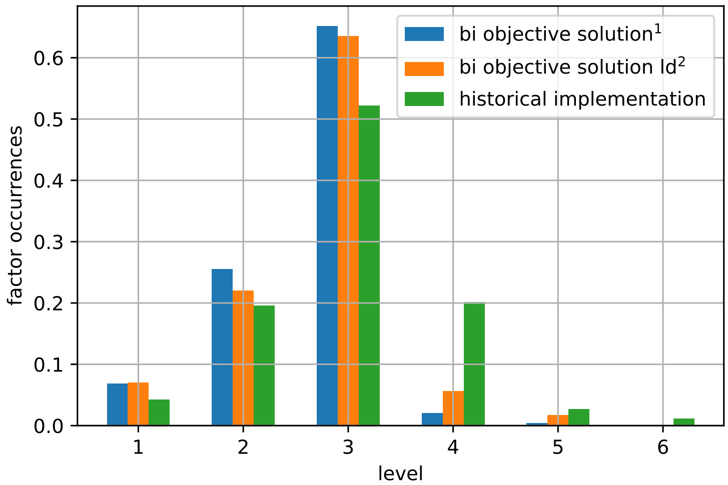


Figure 4.4: Path lengths for selected routes in the historical implementation and in the solution for both the on-time[1] and late delivery[2] formulation of the model using the bi-objective approach.

Bibliography

[1] Imf world economic outlook. https://www.imf.org/en/Publications/WEO/Issues/2022/10/11/world-economic-outlook-october-2022, July 2022.