Université de Montréal

Forecasting for Operational Planning of M1M Systems

par

Mohammad Amini

Department of Computer Science and Operations Research (DIRO)
Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Artificial Intelligence

December, 2021

Université de Montréal

Faculté des arts et des sciences

Ce mémoire intitulé

Forecasting for Operational Planning of M1M Systems

présenté par

Mohammad Amini

a été évalué par un jury composé des personnes suivantes :

Fabian Bastin
(président-rapporteur)

Emma Frejinger
(directeur de recherche)

Teodor Gabriel Crainic
(codirecteur)

Margarida Carvalho
(membre du jury)

Résumé

Cette thèse porte sur la prévision de la demande des expéditeurs et des offres de capacité des transporteurs pour la planification opérationnelle des systèmes 'Many-to-One-to-Many' (M1M). Un tel système agit comme un décideur intermédiaire entre les expéditeurs et les transporteurs en coordonnant le transport des marchandises des expéditeurs aux destinataires en utilisant les capacités offertes par les transporteurs. Le décideur prend ses décisions dans un horizon de planification opérationnelle, en optimisant ces décisions en tenant compte de l'incertitude sur les périodes futures. Pour accompagner les décisions du décideur, il est essentiel de prédire les nouvelles demandes des expéditeurs et les nouvelles capacités des transporteurs. Cela conduit à des problèmes de prévision de séries chronologiques à plusieurs variables et à plusieurs étapes. L'objectif de ce travail est d'analyser l'impact des erreurs de prévision sur la qualité de la solution pour un problème de planification opérationnelle M1M donné.

Cette thèse présente d'abord la structure du système M1M, la planification opérationnelle et les tâches de prévision associées. Nous décrivons la caractérisation des demandes des expéditeurs et des offres des transporteurs ainsi que comment la prévision peut soutenir les décisions en définissant les informations nécessaires pour le décideur sur l'horizon opérationnel.

Nous couvrons ensuite l'optimisation de l'affectation charge-transporteur en introduisant un modèle d'optimisation déterministe et les prévisions utilisées en entrée. En l'absence de données réelles, nous générons un ensemble de données synthétiques qui est ensuite utilisé pour estimer les modèles de prévision. L'objectif est de définir quelques modèles de prévision qui présentent des erreurs afin que nous puissions évaluer leur impact sur la qualité de la solution pour le problème de planification opérationnelle. Dans ce contexte, nous comparons plusieurs modèles ARIMA pour prédire les futures demandes et offres. Nous constatons que les modèles avec des erreurs de prévision relativement faibles peuvent conduire à des améliorations significatives de la qualité de la solution. Enfin, nous exposons quelques pistes de recherches futures.

Mots clés : Prévision de demande, séries chronologiques, M1M Systems, planification opérationnelle.

Abstract

This thesis is about forecasting new shipper-demand requests and carrier-capacity offers for operational planning of Many-to-One-to-Many (M1M) systems. Such a system acts as an intermediary decision-maker between shippers and carriers, coordinating the transportation of goods from shippers to consignees (shipment recipients) using capacity offered from carriers. The decision-maker makes the decisions within an operational planning horizon, optimizing these decisions accounting for uncertainty over future time periods. To support the decisions, forecasting new shipper-demands and carrier capacities is essential. This leads to multi-variate multi-step time series forecasting problems. The objective of this work is to analyze the impact of forecast errors on the solution quality for a given M1M operational planning optimisation method.

This thesis first presents the M1M system structure, operational planning, and related forecasting tasks. It explains the characterization of the shipper requests and carrier offers. We describe how forecasting can support the decisions by defining the needed information for the decision-maker over the operational horizon.

We then cover the optimization of load-to-carrier assignments introducing a deterministic formulation and the forecasts used as input. In the absence of real data, we generate a synthetic data set that is then used for estimating forecasting models. The aim is to define a few forecasting models that exhibit some errors so that we can assess their impact on the solution quality for the operational planning problem. In this context, we compare several ARIMA models to predict future requests and offers. We find that the models with relatively low forecast errors can lead to significant improvements in solution quality. Finally, we outline a few directions for future research.

Keywords: Freight Demand Forecasting, Time Series Models, M1M Systems, Operational Planning.

Contents

Résumé 5

Abstract 7

List of tables 11

List of figures 13

List of acronyms and abbreviations 15

Acknowledgments 17

Chapter 1. Introduction 19

Chapter 2. Operational Planning of M1M Systems 21

2.1. M1M System 21

2.2. Operational Planning 24

2.3. Forecasting Tasks 26

Chapter 3. Literature Review 29

3.1. Time Series Forecasting 29

3.1.1. Autoregressive and Moving Average Methods 30

3.1.2. Exponential Smoothing Methods 31

3.1.3. ARIMA Models 32

3.2. Demand Forecasting for Operational Planning 34

Chapter 4. Methodology 37

4.1. Load-to-carrier Assignment Optimization Model 37

4.2. Load-to-carrier Assignment Solution Method 40

4.3. Forecasting Models 42

4.4. Synthetic Data Generation Process 43

Chapter 5. Computational Results 47
5.1. Synthetic Data Generation 47
5.2. Forecasting Results 49
5.2.1. Stationary Distribution Verification 49
5.2.2. Model Selection 50
5.2.3. Forecast Performance Analysis 53
5.3. Impact of Forecast Errors on Solution Quality 56
Chapter 6. Conclusion and Future Work 59
References 61

List of tables

5.1 Initial parameter setting for shippers and carriers' datasets . . . 48
5.2 Augmented Dickey-Fuller test for a shipper and a carrier datasets. . . . 50
5.3 Estimated partial autocorrelation function for 10-lags of a shipper and a carrier time series data. . . . 51
5.4 Estimated autocorrelation function for 10-lags of a shipper and a carrier time series data. . . . 52
5.5 Model selection information criteria and estimated parameters of shippers' dataset. 53
5.6 Model selection information criteria and estimated parameters of carriers' dataset. 53
5.7 Forecasted new requested volumes and offered volumes for the next 10 time periods using ARIMA(2,0,0). . . . 54
5.8 Performance evaluation results for the successive 10 time periods of a shipper and a carrier time series data. . . . 55
5.9 Total assignment cost (ground truth G and using forecasts F) over ten time periods. . . . 56
5.10 Forecast errors produced by ARIMA(2,2,0) and ARIMA(2,0,0) for shippers' requested volumes over the operational horizon. . . . 57
5.11 Forecast errors produced by ARIMA(2,2,0) and ARIMA(2,0,0) for carriers' offers volumes over the operational horizon. . . . 57
5.12 Percentage error of objective function value due to forecasting. . . . 57

List of figures

2.1 M1M system structure, elements, interactions, and decisions (Guo et al., 2021; Crainic et al., 2021). 22
2.2 Operational Planning Horizon 24
4.1 Multi-period heuristic algorithms framework 41
5.1 Autocorrelation of request and offer values. 48
5.2 Requests and offers values partial autocorrelation correlogram. 51

List of acronyms and abbreviations

M1M	*Many to One to Many*
OPH	*Operational Planning Horizon*
3-PL	*Three-party logistics*
4-PL	*Four-party logistics*
IDSP	*Intelligent Decision Support Platform*
OD	*Origin Destination*
CNN	*Convolutional Neural Network*
LSTM	*Long Short-term Memory*
AR	*Autoregressive*
MA	*Moving Average*
ARIMA	*Autoregressive Integrated Moving Average*
ACF	*Autocorrelation Function*
PACF	*Partial Autocorrelation Function*
VAR	*Vector Autoregressive*
SARIMA	*Seasonal Autoregressive Integrated Moving Average*
ADF	*Augmented Dickey-Fuller*
AIC	*Akaike Information Criterion*
Min_Max	*Minimum Maximum*

MAPE	*Mean Absolute Percentage Error*
SMAPE	*Symmetric Mean Absolute Percentage Error*

Acknowledgments

First of all, I would like to thank Professor Emma Frejinger and Professor Teodor Gabriel Crainic, my research director and co-research director, for their support, advice, and financial support. They offered me an interesting project, introduced me to the world of freight transportation systems, and allowed me to have unique experiences.

I also extend a special thanks to Ms. Maria Ricciardelli, who patiently stood by me, supported me, and advised me during the past challenging year.

I also thank my parents for everything they have done for me.

Chapter 1

Introduction

Freight transportation is perhaps one of the most vital industries, mainly in advanced countries. The importance of transport systems has been particularly visible during the Covid-19 pandemic. For example, consumers and transport service providers use online interactions to a larger extent than before. In turn, the performance of freight transport systems affects all actors involved in sending and receiving shipments. These actors often have conflicting objectives. Shippers and shipment recipients want to pay lower transportation costs. Carriers aim to decrease operational costs, increase revenue and their service quality. Moreover, society as a whole can benefit from efficient systems with a lower environmental footprint.

Integrating shipper-demand requests and carrier-capacity offers in the decision-making process could optimize operations and achieve a balanced set of stakeholders' goals. In that matter, a multi-layer M1M decision-making structure has been introduced for logistics service providers (Guo et al., 2021; Crainic et al., 2021). The idea is an integrated decision-making process consisting of several highly coordinated sub-processes to satisfy the overall operational planning goals. The sub-processes consist of shippers' demand planning, carriers' capacity planning as well as consolidation and economic planning. In this work, we focus on the decision-making problem at the so-called Intelligent Decision Support Platform (IDSP) level. Henceforth we refer to IDSP as the decision-maker. In the introduced structure, shipper-demand requests and carrier-capacity offers occur in different locations and times. By consolidating shipments from different shippers into vehicles and by synchronizing operations, the decision-maker aims to optimize the selection of shipper-demand requests, carrier-capacity offers, load-to-carrier assignments, shipment itineraries, and carrier routes in time and space. Several aspects of this system are uncertain, and forecasting can be used to reduce this uncertainty.

The purpose of forecasting new shipments and capacities is to support decision-making. The decision-maker works within different locations and makes its decisions continuously over time. Forecasting for such a decision-maker is challenging because recorded data by the system varies over time and space. From the time perspective, we deal with a multivariate multi-step time series forecasting problem. From the shippers' side, they make requests for their shipments per period over the operational horizon. From the carriers' side, they offer capacity per period over the operational horizon. The loads and capacities are characterized by their physical, time, and economic attributes. We create the time series data for each shipper and carrier based on their past requests and offers.

In this thesis, we use statistical models to predict shipper-demand requests and carrier-capacity offers for the operational decision-making problem at hand. The aim is to assess the impact of forecast errors on the solution quality of the optimization problem. For this purpose, we generate synthetic data for shippers' demand and carriers' capacity offer using autoregressive models. We use the data to train different ARIMA models whose forecasts are used as input to a deterministic optimization model. The latter is a slight modification of the single-leg multiperiod model proposed in Crainic et al. (2021). We feed the provided forecast data to the optimization model by two forecasting models. Based on several problem instances, we draw the conclusion that the magnitude of the impact on solution quality is smaller than the magnitude of the forecast errors used as input.

This thesis is structured as follows. In Chapter 2, we introduce operational planning of M1M systems, and we describe the related forecasting problem. In Chapter 3, we review some classical time series forecasting models that can be used for our problem. In Chapter 4, we present the methodology. That is, a formulation of the optimization problem and forecasting models for shipper-demand requests and carrier-capacity offers over the operational horizon. We also introduce the data generation process to produce data for forecasting models. In Chapter 5, we discuss the detail of the evaluation and computational results of our proposed models. Finally, Chapter 6, presents the conclusion of this thesis and future work.

Chapter 2

Operational Planning of M1M Systems

In this chapter, we describe the M1M system that aims to fulfill efficiency and profitability stakeholders' interests in freight transportation. In Section 2.1, we introduce its elements and how they interact. In Section 2.2, we discuss operational planning: decisions, activities, and system optimization. We describe a data generation process at the operational level and forecasting problems related to the M1M system optimization in Section 2.3.

2.1. M1M System

M1M system structure and its interactions are depicted in Figure 2.1. The system of interest is composed of a business, communication, and decision-making structure (Guo et al., 2021; Crainic et al., 2021; Taherkhani et al., 2021), comprising three main elements: 1) shippers 2) carriers and 3) a decision-maker. Shippers, including manufacturers, wholesalers, distributors, and three and four logistics service providers (3-PLs and 4-PLs), make shipper-demand requests for transporting their product loads. Carriers, including 3-PLs and 4-PLs and transportation service providers of diverse modes and types (motor carriers, railroads, airlines, rivers and maritime carriers), offer capacities for transportation. Consignees, clients, or customers are the recipients of the products, which could be a distribution center or the customers who purchased the products. Each shipment starts at a shipper's location and must be delivered to a consignee's location. Large shippers and freight forwarders usually sign contracts with carriers or 3-PLs and 4-PL to transport their shipments for a long period. In this case, the shipper-demand requests are typically fixed. On the other hand, some shippers have relatively small shipper-demand requests, so they do not sign a contract but use shipping services in the spot market before shipping travel. Carriers give services to both long-term contract and spot market shippers.

Both shippers and carriers have their concerns; shippers are interested in more efficient transportation services, including their preferred delivery time and cost-effective services, which guarantees low investment and safety of their products. Carriers are interested in more

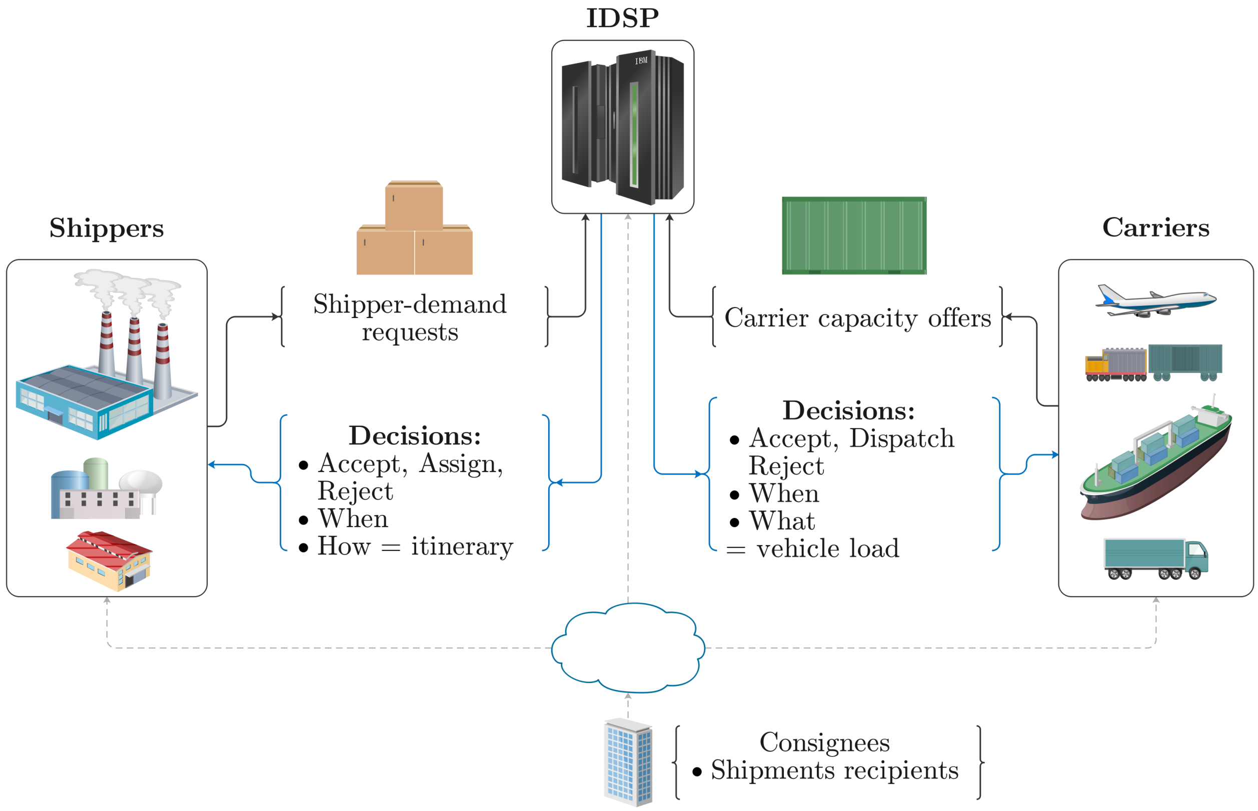

Fig. 2.1. M1M system structure, elements, interactions, and decisions (Guo et al., 2021; Crainic et al., 2021).

efficient transportation systems, mainly to increase their revenue. In that sense, a decision-maker, a platform-based intermediary called IDSP, is introduced for the M1M system (Guo et al., 2021; Crainic et al., 2021). IDSP provides automated planning and optimizes operations to satisfy the shippers, carriers, and platform's interests. The decision-maker aims to optimize load-to-carrier assignments, selection of routes and offers, and shipment itineraries in time and space.

Figure 2.1 also illustrates the interactions, including data and decision exchanges of the M1M system. In that regard, from the shippers' side, product loads in different sizes and weights should transport loads from many different shippers' locations to their related consignees' locations within their desired delivery time. From the carrier's side, vehicles with various capacities carry the loads through different routes.

The M1M system interacts with shippers and carriers to retrieve data needed for the decision-maker. The decisions are based on provided information from shipper-demand requests and carrier-capacity offers characterized by related operational, time-related, and economic attributes. Operational attributes are related to the physical aspects of requests and offers required for handling shipments. Time attributes of requests and offers are related to the timeliness of pick-up and delivery operations. Economic attributes are related to revenue and penalties for IDSP.

IDSP works on one and the same continuous (infinite) time stream. It receives both time-dependent shipper-demand requests and carrier-capacity offers and responses through a continuous stream of decisions. In that sense, IDSP receives both sides' information and makes its decisions at time steps. The decision-maker may suspend a request or an offer for the next decision time or make the decisions when receiving requests and offers. The IDSP decides to accept or reject a request or an offer. At the time of accepting a request, IDSP decides to assign the product loads to accepted offers' capacities, the departing vehicles. The dispatch time of the vehicle is based on the offered services, i.e., on the service departure and arrival times between a specific origin and destination (OD). If the schedule is not predefined, IDSP has a choice of dispatching time; it can decide at what decision time the vehicle should dispatch. If the schedule is predefined, then the decision-maker decides to dispatch the vehicle at the time of assigning the request to the capacity.

For example, a shipper wants to send product loads (e.g., boxes) with a specified volume between OD facilities respecting a preferred delivery time. Conversely, carriers offer capacities in different volumes to transport product loads. For the accepted requests and offers, IDSP looks at both sides' information to see at what time and what itinerary should decide to assign the product loads to which capacities based on the volume of both sides and when the vehicles should be dispatched. If the schedule is predefined when IDSP assigns the loads, it decides to dispatch the vehicle. Mass industrial producers with signed long-term contracts usually have fixed schedules from carriers' side to transport their product loads.

Based on the decisions IDSP takes, different activities occur at the operational level. To transport a shipment from an origin to a destination, three types of activities could occur: 1) consolidation operations, 2) transfer operations, and 3) delivery operations. In consolidation operations, assuming that a given request has been accepted, at first, the shipment should be picked up from the shipper's door (initial consolidation), then carried to the nearest terminal within the area of the shipper. At the terminal, there are different split and join handling operations (second consolidation); the shipment could transfer or be classified based on the product. If the shipment is a single product, a box, for instance, the transfer activities are unload, cross-dock handling then loads to the outgoing vehicle. For the multi-product shipment, the transfer activities are to reclassify, unload, sort, move, then load to the outgoing vehicle. There are three possible destinations regarding delivery operations for outgoing vehicles after departing the terminal: 1) the vehicle could go to the final destination, which is the consignee's door, 2) a terminal in the consignee's area, and 3) an intermediary terminal to transfer.

The M1M system may work within any network at the regional, national, continental, and geographical levels. The system operates between several service zones of a network. The zones can be urban areas or any geographically relevant area, which contain shippers and consignees' facilities. A related terminal may service each zone and be managed or used by IDSP to perform required activities like sub-contracting warehousing or cross-dock transfer. The network can be of different structures, and there are many related planning problems at the strategic, tactical, and operational levels (Crainic and Laporte, 1997). In the following section, we focus on operational aspects.

2.2. Operational Planning

IDSP receives many shipper-demand requests and carrier-capacity offers and aims to optimize its decisions regarding load-to-carrier assignments, shipment itineraries, and carrier routes. Synchronizing the operations related to handling such requests and offers requires operational planning, which reduces handling time and operational costs, also provides convenient transportation.

In this section, we discuss the three main components of operational planning in the M1M system (i) time, (ii) requests and offers information, and (iii) the decision-making process. Regarding the time component, the decision-maker receives time-dependent requests and offers, then makes its decisions at different points in time. IDSP can make the decisions when a new request or an offer appears, when a certain number of requests or offers are received, or make the decisions according to a predefined schedule (once or twice a day, for instance). Regardless of which form we choose, IDSP works based on discrete time steps at which decisions are taken at the operational level.

In this work, the time steps of decisions define the IDSP's schedule length at the operational level based on when a certain number of requests or offers are received. We work with a discrete-time sequence of observations, where a decision is taken at a time instance t, and the subsequent decision will be at time $t+1$. It continues with equal time intervals until the end of time horizon T. Each one of these intervals is called a time period. Hence, we define the IDSP's operational planning horizon (OPH) within periods of equal discrete length $t = 1, 2, 3, \ldots, T$. Figure 2.2 illustrates the operational planning horizon of the M1M system.

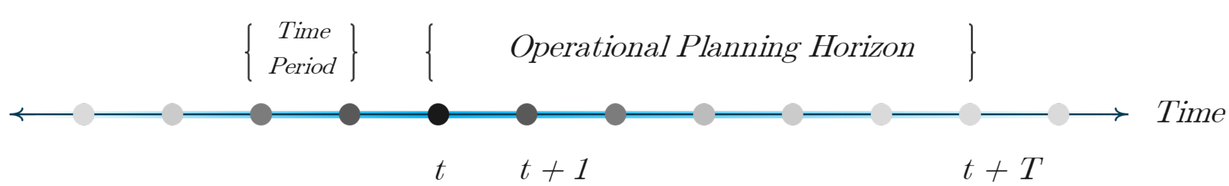

Fig. 2.2. Operational Planning Horizon

The time-related, operational, and economic attributes of requests and offers provide the information needed for the decision-maker. In that sense, operational planning is concerned with "when" and "what" issues (Crainic and Laporte, 1997). For example, when the IDSP decides to accept or reject a request, or what quantity or volume of loads should the decision-maker consider for what capacity.

The time-related attributes of the requests and offers related to transportation times are defined to treat "when" issues in operational planning. In this context, shipper-demand requests are characterized by:

a) their delivery times,
b) when IDSP should decide to accept or reject the requests, and
c) when the requests are ready to be picked up from shippers' facilities.

From carriers' side, the offers are characterized by:

(1) when IDSP should decide to accept or reject the offers,
(2) when vehicles arrive and depart all locations,
(3) when vehicles are ready to work,
(4) when vehicles should release at the destination terminals, and
(5) travel times between terminals.

Attributes of requests and offers related to revenue management of the M1M system consist of revenue and costs for IDSP. Shippers should pay IDSP for their shipments' handling, and if the load is delivered out of the preferred delivery time, IDSP has to pay the shippers based on the service level of the agreement contract. From the carriers' side, IDSP should pay the carriers for their fees corresponding to fixed and variable costs for handling and transportation.

Regarding the load-to-carrier assignment decision-making process, IDSP makes its decisions over the OPH based on information provided by both sides' operational, time, and attributes. Assuming the decisions are taken when receiving a certain number of requests and offers, IDSP decides to accept or reject the requests or offers at the current time t. For a rejected request, the shipper receives a notification for further negotiation. If the request is accepted, the decision-maker decides to assign the request to an accepted capacity offer regarding both sides' attributes and the itinerary. IDSP can dispatch the vehicle at present t or dispatch the vehicle at the following periods if it has the option of dispatching.

When the decisions are taken at each period based on the information at that period, ignoring future events, it is called a myopic decision-making process. Given the dynamic nature of the problem, this can lead to poor decisions. Consider, for example, at the current time t, an IDSP receives a request for a single product and will receive another request in the next period. If the decision-maker has the request information in the following periods at the current time t, it may hold the request until the coming period to assign both requests together. Whereas in a myopic way, at present t, IDSP decides to accept or reject the request, and if it still wants to hold the request, it does not know for how long. This scenario can happen for the dispatching time of the vehicle as well. In that sense, it is more likely that IDSP rejects a request in the myopic process because it does not have a suitable offer for it when making its decisions. Or may assign the request to a vehicle and dispatch the vehicle, not at the perfect time regarding using the vehicle's total capacity. Hence, the decision-maker can optimize its decisions if it has more knowledge of the coming periods.

To fulfill the goal, more information should be used to support decisions. Instead of a myopic model, we consider the future data that belong to the following time periods of the operational horizon. The information needed for IDSP to make the decisions at current time t is composed of present attributes' information of requests and offers, past time periods decisions' information, and expected requests and offers information from the following periods. To optimize a decision at the current time t, IDSP uses an integration of past, present, and forecasts of future information, which makes it a multi-period decision-making process.

The planning horizon is application specific. The required information at the current time t can be uncertain as it is by, e.g., weather, geographical region, spatial effects. Hence, to decrease the uncertainty, forecasting for new requests and offers is needed.

In the M1M system at the operational level, IDSP makes the decisions using an optimization model through the two decision-making processes over OPH. Forecasting for new requests and offers at each period in both decision-making processes are two inputs information for the optimization model. In the next section, we focus on forecasting tasks needed for the operational planning of the M1M system.

2.3. Forecasting Tasks

IDSP makes its decisions based on current and future information over the operational planning horizon. Current information consists of received requests and offers' information at current time t and past time periods decisions information. Future information consists of 1) known information from a signed contract or point forecast and 2) unknown or partially known information.

We define point forecast as the most likely predicted single value of each new shipper demand and carrier capacity over the operational horizon. At the current time t, we therefore forecast over the following time periods of the OPH. The difficulty of the forecasting task increases with the length of the horizon. How many successive time periods can be taken as known information for the decision-maker is related to the reliability of the forecasts and the planning requirements. Accordingly, if we have a deterministic optimization problem, we use point forecasts of new demand requests and capacity offers over the operational horizon.

If the formulation is stochastic, then we can leverage information about the uncertainty of the forecasts.

Regarding the forecasting process, the system operates continuously, and data can be captured as activities unfold. That means each request and offer is recorded with a timestamp and the associated activities. That leads to highly detailed data in both time and space dimensions, including all related attributes. For example, each shipment is associated with the shippers' facility's exact addresses and the related consignees' locations, as well as the volume, quantity, delivery time, and pick-up time. In that sense, the requests and offers' data are observed over time and space with varied timestamps and locations. Hence, large-scale unstructured data is captured which each request and offer can have its own time and space dimension.

Although data can be recorded in every detail, it needs to be structured based on the IDSP decision-making process. The decisions are taken over the OPH within different terminals associated with zones. Thus, the operational planning horizon defines the time dimension, and the terminals represent the space dimension of the M1M system. The observed requests and offers' data needs to be organized within the time steps of OPH and related terminals.

We consider the forecasting of new requests and offers by OD pair over the OPH. The data needed for the forecast comes from the history of past periods' recorded attributes of requests and offers. Hence, forecasting for multiple steps ahead using past period observations leads us to a multi-variate multi-step time-series forecasting problem. In the actual setting, depending on the recorded data, it is likely that the shippers-demand requests exceed the capacity offers. That means requests are constrained by offers, and the system may not observe the true requests, which is known as censored or constrained data (see, e.g., Fields et al., 2021, for recent study).

This work focuses on the mentioned multi-variate multi-step time-series forecasting tasks and their impacts on the operational decisions for a specific OD pair. We assume a deterministic formulation of the operational planning problem and do point forecasts for shippers-demand requests and carrier-capacity offers over the OPH. We measure the shipments and capacities by their observed volume attributes over one of their associated time attributes (availability time). Unfortunately, we never gained access to real data. Therefore, we generate synthetic data to train the forecasting models. Although we might need to deal with censored data in the actual setting of the system before the forecasting process, it is not the focus in this work. In the following, we review some classical time series models.

Chapter 3

Literature Review

Due to its multidimensional characteristics, freight demand can be more complex to forecast than passengers' transportation demand problems. Nevertheless, researchers have been more interested in understanding passengers' demand (Systematics, 1997). Freight forecasting is highly application-specific, and the studies that have been done in this field are related to specific applications. Even though machine learning has seen recent success, the vast majority of freight forecasting studies use statistical or econometric models. The latter tend to perform relatively well in small data regimes (Laage et al., 2021; Makridakis et al., 2018). The two fields of machine learning and statistics/econometrics are related, yet with distinct approaches to forecasting. In essence, machine learning is data-driven, while statistics/econometrics rely on models as representations of reality and related hypothesis testing. Unfortunately, we did not get access to real data for the work in this thesis. We therefore use statistical models for data generation and forecasting. The literature on time series modeling is vast. Here we focus on statistical time series models, including Autoregressive and Moving Average, Exponential Smoothing, and Autoregressive Integrated Moving Average (ARIMA) models. The reason for reviewing these models is because they perform well with small sizes data sizes and have the ability to describe time components and correlations within the data. Our review will continue with a discussion of demand forecasting for operational planning.

3.1. Time Series Forecasting

Box et al. (1970) introduces statistical techniques by using an iterative forecasting process to estimate parameters from data. The classical method for estimating models involves three steps: 1) model selection/definition, 2) parameter estimation and 3) model verification. These steps are common to the models we review in the following subsections.

3.1.1. Autoregressive and Moving Average Methods

We focus on the discrete time-domain category of quantitative methods. In particular, Autoregressive (AR) and Moving Average (MA) models introduced by Yule (1927) from a random time basis process. Yule (1927) introduced an AR model by modeling a harmonic movement of a pendulum. He developed the sinusoidal motion for three intervals then represented the movement y_t as a function of time which typically can be a linear second-order differential equation of the function as

$$r\frac{d^2y}{dt^2} + p\frac{dy}{dt} + qy = \varepsilon g_t \tag{3.1.1}$$

where r, p, and q are the constant parameters, g_t is the impulse force disturbance at time $t = 0$, and ε is the magnitude of the disturbance. He decomposed (3.1.1) for discrete time series by differencing the first and second intervals of the observations as

$$\psi_2\nabla^2\tilde{y}_t + \psi_1\nabla\tilde{y}_t + \psi_0\tilde{y}_t = \varepsilon_t \tag{3.1.2}$$

where, ψ_0, ψ_1,ψ_2 are parameters, ε_t is an error term at time t. The deviation from the mean is

$$\tilde{y}_t = y_t - \mu, \tag{3.1.3}$$

$$\nabla y_t = y_t - y_{t-1}, \tag{3.1.4}$$

$$\nabla^2 y_t = y_t - 2y_{t-1} + y_{t-2}, \tag{3.1.5}$$

and we obtain

$$\tilde{y}_t = \alpha_1\tilde{y}_{t-1} + \alpha_2\tilde{y}_{t-2} + \varepsilon_t, \tag{3.1.6}$$

where (3.1.6) is the second-order autoregressive model, and α_1,α_2 are parameters.

The generalized form of (3.1.6) for univariate AR of order p can be written as

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t, \tag{3.1.7}$$

where, α_0,...,α_p are unknown parameters of past p-lags that are estimated from data. The error terms ε_t are independent and identically distributed (i.i.d.) with zero mean and variance σ^2.

3.1.2. Exponential Smoothing Methods

Brown (1959) introduced exponential smoothing methods, and Winters (1960) used them for short-term forecasting to generate predictions to be used in optimization-model solving. With the help of an exponential system, Brown (1959) showed that it is possible to predict the next time period sales by weighted sales' average of the present time period, which can continue over the following periods. He used simple exponential smoothing, double exponential smoothing, and triple exponential smoothing methods to predict three different time series. The first model used a simple exponential smoothing model corresponding to a time series with no long-term trend and no clear seasonality pattern. In that sense, at the current time $t = 1$, each time period has a forecasting weight, and weighted averages compute the forecast for the next time period from the past periods at $t = 2$. The weights decrease exponentially with the lags. We obtain

$$y_{t+1} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots + \alpha(1-\alpha)^p y_{t-p}, \tag{3.1.8}$$

or equivalently,

$$y_{t+1} = \sum_{i=0}^{p} \alpha(1-\alpha)^i y_{t-i}, \tag{3.1.9}$$

where y_{t+1} is the forecast for time period $t+1$, and $0 \leq \alpha \leq 1$ is the smoothing parameter which controls the weights for past periods.

This model can not handle trends and seasonality. Holt (2004) enhanced the method for predicting with seasonality and trend. His introduced method consists of trend and level smoothing. A comparison between Holt and Brown exponential smoothing has been made by Xiao et al. (2014) for freight demand forecasting. He compared the two methods using 15 years of a real dataset. He showed that even though most studies adopt Brown exponential smoothing, the Holt method performed better in his experiments.

Standard exponential smoothing methods are simple to implement but can handle simple seasonal streams. A class of exponential smoothing methods for freight transportation at an operational level has been presented in Godfrey and Powell (2000). They introduced a Damped Trend Multi-Calendar (DTMC) exponential smoothing model that can forecast demands with characterized OD, volume, and type of shipper requests. Their model is based on Gardner and Everette (1985), a damped trend and multiplicative season applicable per season, which they changed the seasonal parameter to their calendar factors. Their forecasting equation for DTMC is

$$y_t(m) = \left(S_t + \sum_{i=1}^{m} \phi^i T_t \right) \prod_{j \in J} I_t^m(j), \tag{3.1.10}$$

where $y_t(m)$ is the forecast for $m-th$ period ahead of their calendar, and S_t is the exponential smoothing formulation of subsequent time period computed the same as (3.1.9), ϕ is damping (multiplicative) factor for trend, T_t is the smoothed trend at the end of period t, and $I_t^m(j)$ is the calendar factor over, all set of calendar attributes J.

They examined their model for a six-year real dataset aggregated from 25 cities and compared their model with a seasonal autoregressive MA as a benchmark. They used root-mean-squared error (RMSE) to evaluate the performance

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_t(m) - y_t(m))^2}{n}}, \tag{3.1.11}$$

and showed that the DTMC model performs better than the benchmark.

3.1.3. ARIMA Models

ARIMA models offer a different perspective on time series forecasting. The two most generally used techniques for time series forecasting are exponential smoothing and ARIMA models, which give complementary approaches to the problem. While exponential smoothing methods try to explain the data's trend and seasonality, ARIMA models strive to characterize the data's autocorrelations. ARIMA models allow both AR and MA components in modeling which makes it more flexible than the other statistical models. Box et al. (1970) introduced the ARIMA model as

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_{p+d} y_{t-p-d} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} + \varepsilon_t, \tag{3.1.12}$$

where (3.1.12) is the ARIMA(p,d,q) model consisting of AR of order p, the integration of difference of d past steps, and MA of order q. Parameters α_1,...,α_p, and θ_1,...,θ_q are unknown and most be estimated from data, and ε_1,...,ε_q are the i.i.d. error terms at past lags.

Judging which model is the best fit in terms of (p, d, q) orders is an empirical and iterative process. One way is to determine the initial orders of the model by estimating the autocorrelation function (ACF) of order p from the data of y as

$$\hat{\gamma}(p) = \frac{1}{N} \sum_{n=1}^{N-P} (y_t - \bar{y})(y_{t-P} - \bar{y}), \tag{3.1.13}$$

where $\hat{\gamma}(p)$ is the estimated autocovariance function for lag p, and $\bar{y}$ is the sampled mean. Then, we have

$$\hat{\rho}(p) = \frac{\hat{\gamma}(p)}{\hat{\gamma}(0)}, \tag{3.1.14}$$

where $\hat{\rho}(p)$ is the estimated ACF for a lag p. For partial autocorrelation function (PACF) the aim is to find the part of correlation between y_t and y_{t-p} that is not reflected by the observations $\{y_{t-p+1}, \ldots, y_{t-1}\}$. From that we have a conditional correlation of order p as

$$\hat{\phi}(p) = \frac{Cov(y_t, y_{t-p}|y_{t-1}, \ldots, y_{t-p+1})}{\sqrt{Var(y_t|y_{t-1}, \ldots, y_{t-p+1})Var(y_{t-p}|y_{t-1}, \ldots, y_{t-p+1})}}, \tag{3.1.15}$$

where $\hat{\phi}(p)$ is the estimated PACF for lag p. We note that in modern statistical learning, models are selected based on their out-of-sample fit (error on validation sets) rather than according to ACF and PACF. The latter can nevertheless provide interesting insights.

Guo et al. (2010) used an ARIMA model to forecast monthly railway freight volumes in China. Their observations consist of 96 samples of railway freight from 2002 to 2009. The ACF and PACF analyses of 12 months show that their observed data is non-stationary with trends and has a 12-month seasonal cycle. They defined the ARIMA(p, d, q) model as

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \cdots - \theta_q u_{t-q}, \tag{3.1.16}$$

$$z_t = \nabla^d y_t, \tag{3.1.17}$$

where z_t is defined based on the time series data y_t with d step difference the observed data from time t. ϕ and θ are the parameters of the AR and MA, respectively, and u_t is the random disturbance at time t (Guo et al., 2010). They compared their model with a Holt-Winter model and explain that although the Holt-Winter model performs well on the short-term forecast, the ARIMA model produces more accurate forecast value as the number of periods grows. They used four testing approaches to find the best fit: 1) stability test, 2) residual analysis, 3) model excessive set test, and 4) model low set test.

We can define ARIMA-based models for different types of time-series data. Schulze and Prinz (2009) introduced a seasonal ARIMA (SARIMA) model for container transportation within the origin Germany and three destinations. They analyzed the seasonal behavior of quarterly data from 1989 to 2006 and quarterly forecasts for 2007 and 2008. They build four forecasting models for each destination and the total output of the origin ports. To build the model by analyzing the ACF, PACF, and statistical test, they applied a double differencing linear transformation filter for nonseasonal and seasonal time steps formulated by

$$\nabla^1 \nabla^4 = (y_t - y_{t-1}) - (y_{t-4} - y_{t-5}). \tag{3.1.18}$$

The order of differencing is used to make the time-series stationary. By statistical test, they showed that the data could be assumed stationary with 1 step differencing in nonseasonal and quarterly seasonal. For forecasting, they built and initialed quarterly SARIMA(0,1,1)(0,1,1) which is

$$\hat{y}_{T+1} = y_T + y_{T-3} - y_{T-4} + \beta_1 \cdot e_T + \beta_4 \cdot e_{T-3} + \beta_1 \beta_4 \cdot e_{T-4}, \tag{3.1.19}$$

where T is the quarterly periods, β is the estimated parameter of MA model. Regarding the forecast interval for h-step ahead they showed

$$\hat{y}_{T+h} = \hat{y}_{T+h-1} + \hat{y}_{T+h-4} + \hat{y}_{T+h-5}, \tag{3.1.20}$$

which after 5 step-ahead, the errors decreased to zero. They compared their model with a Holt-Winters model, and for residual analysis, they used root mean squared errors and Theil's U (Schulze and Prinz, 2009) as follow

$$\mathrm{U} = \frac{\mathrm{RMSE}_{(\mathrm{forecast})}}{\mathrm{RMSE}_{(\mathrm{naive})}}, \tag{3.1.21}$$

where forecasting at y_{t-4} has been used as a naive model. Their comparison shows that the SARIMA model is slightly more accurate than the Holt-Winters model.

Freight demand is characterized by several attributes that we need to consider in prediction. Some variables may influence each other, which means there is cross-correlation. This leads to a multivariate time-series forecasting problem. One of the approaches to model the problem is using a vector autoregressive (VAR) models (Box et al., 2015). The complexity of such models depends on the number of variables involved in the modeling. Using the VAR model, we write time-series equations with the number of variables to forecast the next period. In each equation, we consider the other variables at each lag. If we have p lags and n variables, then we have $p \times n$ values for each forecaster equation. As long as we have n parameters needed to estimate, the total variables would be $p \times n^2$. Hence the parameters needed to estimate will increase in the order of $O(n^2)$. The equation for the VAR with the order of 1 is

$$\boldsymbol{Z}_t = \boldsymbol{\Phi} \boldsymbol{Z}_{t-1} + \boldsymbol{a}_t, \tag{3.1.22}$$

and for $n = 2$ we have

$$\boldsymbol{Z}_t = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \boldsymbol{Z}_{t-1} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}, \tag{3.1.23}$$

where $\boldsymbol{Z}_t = (z_{1t}, \ldots, z_{nt})'$ is the stationary variables vector, $\boldsymbol{\Phi}$ is the matrix form of the parameters needed to estimate, $\boldsymbol{a}_t$ is the error vector at time t.

3.2. Demand Forecasting for Operational Planning

The literature focused on demand forecasting specifically for operational planning of transport systems is relatively scarce. In this section we review literature related to forecasting for third-party logistic service providers. Bayraktar et al. (2008) analyzed the impact of demand forecasting on the information distortion in shipper demands for a 3PLs company, using an exponential smoothing model. They described that although the seasonality component decreases the forecast accuracy, it positively reduces the distortion parameter in customer demands between orders from carriers and sales to customers.

Ren et al. (2020) proposed an integrated approach composed of convolutional neural networks (CNN) and LSTM models for demand forecasting related a China's 3PLs company. Their model uses weekly time-space data, and a CNN model extracts the space dimension, removes the noise, and reduces the parameters as the first layer of their model. Then an LSTM model uses the output of the first layer with only time dimension to model the time series data and its complexities.

In the sense of unknown factors' effect on forecasting, Pradita et al. (2020) analyzed two events that significantly impact container demand forecasting for an Indonesian 3PLs company. They defined factors related to shippers, carriers, and customers in the supply chain as internal factors, and economic situation, government policies, oil price, and natural disasters as external effects. Then, with a preprocessing approach, they adjusted the data and used different statistical methods to find the best performing model.

In summary, there are studies in the literature on both time series models and machine learning algorithms for sequential data. If we had access to a real, large source of data, it would be relevant to train both types of models and compare their predictive performance measuring, e.g., mean squared error on validation sets. Unfortunately, we do not have access to real data. We therefore select ARIMA models for our multi-step multi-variate forecasting problem (Section 2.3). Moreover, we choose a second-order autoregressive model to generate synthetic data. The next chapter introduces the models for forecasting shipper-demand requests and carrier-capacity offers in the M1M system. We also describe the formulation of the operational planning problem (Fomeni et al., 2021) that we use.

Chapter 4

Methodology

The IDSP is designed to optimize load-to-carrier assignments, selection of routes and offers, and shipment itineraries in time and space. It makes its decisions over the operational planning horizon within terminals based on the decision-making process presented in Section 2.2. The forecasting process is defined based on the data type generated by the system. We have a multi-step multi-variate time-series forecasting problem that we model using approaches introduced in Chapter 3. Forecasts are inputs to the multi-period decision-making process.

In this chapter, we introduce a deterministic optimization model for the decision-making process in Section 4.1. We discuss a synthetic data generation process in Section 4.4 and determine the assumptions needed to consider for forecasting models. Finally, in Section 4.3, we introduce two time-series forecasting models aiming to produce point forecasts for future shipments and capacities over the operational planning horizon.

4.1. Load-to-carrier Assignment Optimization Model

For optimizing load-to-carrier assignments, Fomeni et al. (2021) introduce a multi-period optimization model, based on Bin Packing principles, for the single-segment case (i.e., two terminals linked by several transportation modes). In this section, we consider a simplification of the model for a specific OD pair. The formulation is the same as Fomeni et al. (2021) except a slight adaptation to integrate time series forecasts, as we further explain below.

Fomeni et al. (2021) introduce a Bin-Packing problem taxonomy $D/C/B/K/T[\cdot]$ where

- D stands for the D-dimensionality of physical attributes.
- C represents the vehicle fixed cost, which could be the same for all or variable for each vehicle.
- B is the vehicle size, which we could be all vehicles as the same or variable sizes.
- K stands for the shipment types, which could be single or multi-shipment types.
- T illustrates the single or multi-periods of the operational horizon.

- $[\cdot]$ represents additional attributes, which could be different load-to-carrier related costs and rules.

The multi-period model, 1/V/V/S/M[I2B], considers both physical and time attributes of shipments and capacities for the decisions. Here we consider $D = 1$ (1-dimension physical characteristic), $C = V$ is the fixed cost of capacities that can vary for each vehicle, $B = V$ is different volumes of vehicles, $K = S$ which means single commodity type, $T = M$ meaning the multi-period of the operational horizon, $[\cdot] = [I2B]$ is the load-to-carrier assignment.

In this context, the physical features of loads and capacities correspond to volume. The time attributes are defined by a) shipments' pickup time and latest delivery time, and b) vehicles' starting work time and release time. They define the following time windows:

- $\mathcal{T}_k = [\,\underline{t}_k, \bar{t}_k\,]$ is the availability time window of a shipment where $\underline{t}_k$ is the pickup time and $\bar{t}_k$ is the latest delivery time;
- $\Gamma_v = [\,\underline{\Gamma}_v, \overline{\Gamma}_v\,]$ is the availability time window of a capacity where $\underline{\Gamma}_v$ is when the vehicle is ready to work and $\overline{\Gamma}_v$ is when the load is released; and
- $\mathcal{T}_{kv}$ is the feasible load-to-carrier assignment interval considering $\mathcal{T}_k$ and Γ_v.

This work focuses on forecasting new requested volumes and offered volumes over the operational planning horizon. We define the OPH with schedule length $T \in \mathbb{Z}_+$, consisting of time periods $t = 1, 2, \ldots, T$. This process repeats over the same schedule length, represented by a deterministic formulation of the optimization model. Hence, we have a fixed horizon where, at each t, we produce point forecasts of new requested volumes and offered volumes based on their availability times over $t + 1, t + 2, \ldots, t + T$. At each decision time period, the information provided by forecasting is considered known to the decision-maker.

Fomeni et al. (2021) consider N_t shipper requests and M_t carrier offers per time period, each characterized by its volume and temporal attributes. In this work, we consider a set of N shippers, and a set of M carriers, each making requests and offers over time. We construct one time series for each shipper and each carrier and assume that they make at most one request/offer per time period. This is a restriction compared to Fomeni et al. (2021) because the set of shippers and carriers is assumed fixed and known. We make this assumption to ensure that there is historical data for each actor. In practice, if a new shipper or carrier enters the market, one would first need to use a forecast model trained on existing data and update it as data for this specific actor becomes available. The assumption of one request/offer per time period is not very restrictive as a time period can be relatively short in an operational planning problem. We now define this setting more formally.

From the shippers' side, let $\mathcal{K}_t$ be a set of shippers which have requested volumes to transport for a specific destination with availability time $\underline{t}_k$ that falls in time period t. Let $|\mathcal{K}_t| = N$ be the number of shippers at time period t. Each shipper makes at most one request per time period. We define the shipper-demand request y^s_{tk} to be the volume of each shipper's shipment $k \in \mathcal{K}_t$ at time period t as a function of its past values. The notation s, stands for a shipper. From that for N number of shippers in $\mathcal{K}_t$, we define the request vector $\boldsymbol{y}_t^s$ consisting of requests $\boldsymbol{y}_t^s = (y_{t,1}^s, y_{t,2}^s, \ldots, y_{t,N}^s)^T$, where $\boldsymbol{y}_t^s \in \mathbb{R}_+^N$.

We forecast $\hat{\boldsymbol{y}}_{t+1}^s, \ldots, \hat{\boldsymbol{y}}_{t+T}^s$, and define a matrix of shippers' requested volumes forecasts

$$\hat{\boldsymbol{Y}}^s = \begin{pmatrix} \hat{y}_{t+1,1}^s & \hat{y}_{t+2,1}^s & \cdots & \hat{y}_{t+T,1}^s \\ \hat{y}_{t+1,2}^s & \hat{y}_{t+2,2}^s & \cdots & \hat{y}_{t+T,2}^s \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{t+1,N}^s & \hat{y}_{t+2,N}^s & \cdots & \hat{y}_{t+T,N}^s \end{pmatrix}^T, \tag{4.1.1}$$

where $\hat{\boldsymbol{Y}}^s \in \mathbb{R}_+^{T\times N}$.

From the carriers' side, let $\mathcal{V}_t$ be a set of carriers offering capacity volumes for a specific destination with availability time $\underline{t}_v$ that falls in time period t. Let $|\mathcal{V}_t| = M$ be the number of carriers at time period t. Each carrier makes at most one offer per time period. We define carrier-capacity offer y_{tv}^c for volume of each carrier's vehicle $v \in \mathcal{V}_t$ at time period t as a function of its past values. The notation c, stands for a carrier. For M distinct number of carriers we define the offer vector $\boldsymbol{y}_t^c$ consists of offers $\boldsymbol{y}_t^c = (y_{t,1}^c, y_{t,2}^c, \ldots, y_{t,M}^c)^T$, where $\boldsymbol{y}_t^c \in \mathbb{R}_+^M$.

We forecast $\hat{\boldsymbol{y}}_{t+1}^c, \ldots, \hat{\boldsymbol{y}}_{t+T}^c$, and define the matrix of carriers' offered volumes forecasts

$$\hat{\boldsymbol{Y}}^c = \begin{pmatrix} \hat{y}_{t+1,1}^c & \hat{y}_{t+2,1}^c & \cdots & \hat{y}_{t+T,1}^c \\ \hat{y}_{t+1,2}^c & \hat{y}_{t+2,2}^c & \cdots & \hat{y}_{t+T,2}^c \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{t+1,M}^c & \hat{y}_{t+2,M}^c & \cdots & \hat{y}_{t+T,M}^c \end{pmatrix}^T, \tag{4.1.2}$$

where $\hat{\boldsymbol{Y}}^c \in \mathbb{R}_+^{T\times M}$. For the decision-maker at current time t, by concatenating $\boldsymbol{y}_t^s$ with $\hat{\boldsymbol{Y}}^s$, and $\boldsymbol{y}_t^c$ with $\hat{\boldsymbol{Y}}^c$, we define the matrix request $\boldsymbol{Y}^s$ and the matrix offer $\boldsymbol{Y}^c$ such a that $[\boldsymbol{Y}^s]_{tk} = y_{tk}^s$ and $[\boldsymbol{Y}^c]_{tv} = y_{tv}^c$.

The optimization problem consists in minimizing the total cost over the OPH. The decision variables are:

- $\xi_{t,v} = 1$ if vehicle $v \in \mathcal{V}_t$ at time period $t \in \Gamma_v$ is selected, otherwise 0,
- $\zeta_{t,kv} = 1$ if shipment $k \in \mathcal{K}_t$ with long-term contract is assigned to vehicle $v \in \mathcal{V}_t$ at time period $t \in \mathcal{T}_{kv}$, otherwise 0,
- $u_{t,k} = 1$ if shipment $k \in \mathcal{K}_t$ is assigned to a spot market vehicle at time period $t \in \mathcal{T}_k$, otherwise 0.

Using the forecasts we just introduced, the resulting multi-period formulation is

$$\min_{\xi,\zeta,u} \sum_{v\in\mathcal{V}_t} \sum_{t\in\Gamma_v} f_v \xi_{tv} + \sum_{k\in\mathcal{K}_t} \sum_{v\in\mathcal{V}_t} \sum_{t\in\mathcal{T}_{kv}} a_{tkv}\zeta_{tkv} + \sum_{k\in\mathcal{K}_t} \sum_{t\in\mathcal{T}_k} p_{tk} u_{tk} \tag{4.1.3}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}_t} \hat{y}_{tk}^s \zeta_{tkv} \le \hat{y}_{tv}^c \xi_{tv}, \qquad \forall v \in \mathcal{V}_t, t \in \Gamma_v, \tag{4.1.4}$$

$$\sum_{v\in\mathcal{V}_t}\sum_{t\in\mathcal{T}_{kv}}\zeta_{tkv}+\sum_{t\in\mathcal{T}_k}u_{tk}=1,\quad \forall k\in\mathcal{K}_t, \tag{4.1.5}$$

$$\sum_{t\in\Gamma_v}\xi_{tv}\le 1,\qquad \forall v\in\mathcal{V}_t, \tag{4.1.6}$$

$$\xi_{tv}\in\{0,1\},\qquad \forall v\in\mathcal{V}_t, t\in\Gamma_v, \tag{4.1.7}$$

$$\zeta_{tkv}\in\{0,1\},\quad \forall k\in\mathcal{K}_t, v\in\mathcal{V}_t, t\in\mathcal{T}_{kv}, \tag{4.1.8}$$

$$u_{tk}\in\{0,1\},\qquad \forall k\in\mathcal{K}_t, t\in\mathcal{T}_k, \tag{4.1.9}$$

where, f_v is the fixed cost of selecting and using a vehicle, a_{tkv} is the variable cost of assigning a shipment to the selected capacity at each time-period, p_{tk} is the shipping cost of assigning a shipment to a spot market vehicle at each time-period.

The objective function (4.1.3) is computed by minimizing the total fixed cost of using the accepted capacities, the total assignment cost of shipments to the accepted vehicles, and the total shipping cost of assigning the loads to the spot market vehicles over the OPH. Constraints (4.1.4) enforce the physical requirements of both sides, which means the total assigned shipments' volume cannot be more than the accepted capacity's volume. Constraints (4.1.5) ensure that all accepted shipments are assigned to at least one vehicle. Constraint (4.1.6) guarantees that the decision-maker does not accept a vehicle more than once. Binary constraints on the decision variables are imposed by constraints (4.1.7), (4.1.8), and (4.1.9).

Compared to the model in Fomeni et al. (2021), we use each set $\mathcal{K}_t$ and $\mathcal{V}_t$ at each time period instead of one set of shipments and capacities over the schedule length because the decision-maker makes its decisions per time period. We modified the constraints (4.1.4) where we inserted the forecasted requested volumes and offered volumes over the operational horizon because shipper-demand requests and carrier-capacity offers are frequent per time period.

4.2. Load-to-carrier Assignment Solution Method

With respect to solving the multi-period formulation explained in Section 4.1, Fomeni et al. (2021) proposed four constructive heuristic algorithms, HM1 - HM4. The main goal of the solution approaches proposed by the authors is to adapt the well-known First-Fit Decreasing and Best-Fit Decreasing algorithms where the shipments are sorted in descending order of sizes in both algorithms. Then by using the first algorithm, one assigns shipments to the first large carrier capacity. The second algorithm assigns shipments to the already loaded carrier capacity in which it fits (Martello and Toth, 1990; Dyckhoff, 1990).

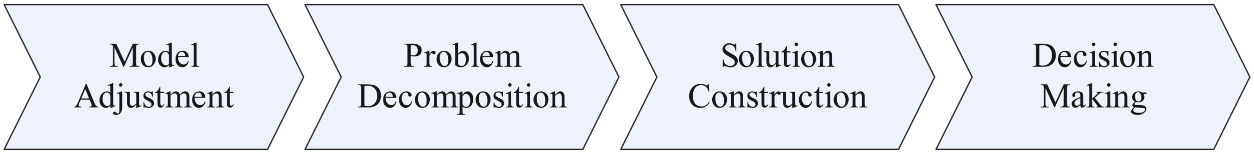

Fig. 4.1. Multi-period heuristic algorithms framework

HM1 - HM4 use the same framework to solve the multi-period problem. Figure 4.1 shows the structure of proposed heuristic algorithms. First, for each availability time in the operational horizon, one reproduces the selection variables of the capacity offer and shipper demand to adjust the model. Then, one breaks the multi-period problem into a series of single-period subproblems with respect to the availability times over the horizon and sequentially solves each of them by introduced single-period algorithms in the paper. To find the best possible solution, one determines the best time period for selecting and packing each capacity to avoid duplication of selection in the process. After obtaining the solution, related load-to-carrier assignment decisions will be made, including assigning the shipments to the most cost-efficient selected capacity and resorting unassigned loads and capacities. Finally, all shipments will be allocated to either selected or spot-market capacities by repeating the process.

Our choice of heuristic to solve the 1/V/V/S/M [I2B] model depends on assessment of the carrier capacities to identify their most efficient time period and which single-period algorithm is used to solve the problem. If capacities are rated by their cost contribution, we use HM4 to solve the model. If capacities are not ranked, HM1 - HM3 would be applicable.

The results in Fomeni et al. (2021) show that the proposed algorithms perform well in terms of finding a feasible solution to minimize the related cost. They evaluate the performance based on computational time and the difference between the best-known solution and the best feasible solution called the optimality gap. One set of instances – referred to as Set 1 Type 1 by the authors – where capacities' volumes are relatively larger than shipments' volumes, the HM1 returns the optimality gap between 1.06% and 16.24% and computational time between 0.01 s and 2.50 s. HM2 gives the optimality gap in the range of 5.53% to 15.71%, with computing time in the range of 0.02 s to 8.11 s. HM3 yields an optimality gap between 7.31% and 16.05%, and a computing time between 0.03 s and 15.84 s. Finally, HM4 returns an optimality gap in the range of 0.94% to 8.84% with the computational time between 0.05 s to 6.90 s.

Based on the results, HM4 outperforms the other algorithms on Set 1 and Type 1 for ranked capacities. In this research, vehicles are available in their orders and are assessed with no ranking for our specific problem. Hence, we use HM1 in our experiments to find a feasible solution. HM1 is based on the shipper demands, focusing on assigning all loads to the carrier capacities. The algorithm first assigns the selected shipper demands to empty sets of accepted carrier capacities. For the remaining unassigned shipments, the algorithm tries to sequentially allocate them to the best possible accepted vehicle based on related costs. If such a vehicle is not available, a new carrier capacity may be selected and the algorithm tries to assign remaining shipments to the new vehicle based on its costs' efficiency. Afterward, if the shipper's demands remain unassigned, the algorithm uses spot-market vehicles to complete its assignment decisions.

4.3. Forecasting Models

In this section, we consider two simple models from the literature to forecast new shipments and capacities: an autoregressive (AR) model and an ARIMA model. In Section 3.1 we described the general form of AR(p) on (3.1.7), which will be used to define the AR model for requests and offers as

$$y^s_{tk} = \alpha_{0k} + \alpha_{1k} y^s_{(t-1),k} + \alpha_2 y^s_{(t-2),k} + \cdots + \alpha_{p,k} y^s_{(t-p),k} + \varepsilon_{tk}, \tag{4.3.1}$$

$$y^c_{tv} = \alpha_{0v} + \alpha_{1v} y^c_{(t-1),v} + \alpha_2 y^c_{(t-2),v} + \cdots + \alpha_{p,v} y^c_{(t-p),v} + \varepsilon_{tv}, \tag{4.3.2}$$

where $\alpha_{p,k}$ and $\alpha_{p,v}$ are the unknown parameters of past p-lags apart needed to infer from the generated data. Random terms ε_{tk} and ε_{tv} are i.i.d. with zero mean and constant variance σ^2. How many lags we should consider for the forecasting model depends on the significance in partial autocorrelation function analysis of the generated data. PACF is defined to show other correlations within the captured data between y^s_{tk} and $y^s_{(t-p),k}$, which the AR process cutoff to zero after p-lag.

We can explain the forecasting formulation for both sides in terms of a) y^s_{tk} and y^c_{tv} with their past volumes' values, b) the generated error ε_{tk} and ε_{tk} at each time period, where the errors are uncorrelated, and each error term is a linear function of its preceding. From that, three ARIMA-based models presented in Section 3.1:

(1) y^s_{tk} and y^c_{tv} are a function of their past values, the error terms at past and current time periods, as well as a function of differencing;

(2) y^s_{tk} and y^c_{tv} are only a function of error terms at past and present time periods;

(3) y^s_{tk} and y^c_{tv} are a function of the sum of weighted past values and the error term at the current time period.

Regarding model selection with the help of auto correlation function, we select the first formulation to address the second forecasting model. Based on (3.1.12) we define

$$y^s_{tk} = \alpha_{0k} + \alpha_{1k} y^s_{(t-1),k} + \cdots + \alpha_{(p+d),k} y^s_{(t-p-d),k} - \theta_{1k}\varepsilon_{(t-1),k} - \cdots - \theta_{q,k}\varepsilon_{(t-q),k} + \varepsilon_{tk}, \tag{4.3.3}$$

$$y^c_{tv} = \alpha_{0v} + \alpha_{1v} y^c_{(t-1),v} + \cdots + \alpha_{(p+d),v} y^c_{(t-p-d),v} - \theta_{1v}\varepsilon_{(t-1),v} - \cdots - \theta_{q,v}\varepsilon_{(t-q),v} + \varepsilon_{tv}, \tag{4.3.4}$$

where (4.3.3) and (4.3.4), are the ARIMA(p,d,q) models by direct use of (3.1.17) in the model. Unknown parameters $\alpha_{p,k}$, $\alpha_{p,v}$, $\theta_{q,k}$, and $\theta_{q,v}$ need to be estimated. Random terms $\varepsilon_{q,k}$ and $\varepsilon_{q,v}$ are i.i.d. at past lags.

In practice, depending on the data under study, there are different approaches in the literature to make the time-series data stationary. For example, in our case, when we have a stochastic forecasting process with deterministic variation over mean to make the data stationary, Yue and Pilon (2003) compared differencing methods with detrending approaches. They showed that using differencing methods can distort the existing forecast process, while detrending can remove the deterministic trend without any change. Since we aim to generate synthetic data (details in the following section) and it is stationary, we can define (4.3.3) and (4.3.4) as follows

$$y^s_{tk} = \alpha_{0k} + \sum_{n=1}^{p} \alpha_{n,k} y^s_{(t-n),k} + \varepsilon_{tk} - \sum_{n=1}^{q} \theta_{n,k}\varepsilon_{(t-n),k}, \tag{4.3.5}$$

$$y^c_{tv} = \alpha_{0v} + \sum_{n=1}^{p} \alpha_{n,v} y^c_{(t-n),v} + \varepsilon_{tv} - \sum_{n=1}^{q} \theta_{n,v}\varepsilon_{(t-n),v}, \tag{4.3.6}$$

where (4.3.5) and (4.3.6) are the ARIMA($p, 0, q$) models or can interpret as ARMA(p,q) models, order p and q come from PACF and ACF analysis where AR process shows sharply cutoff to zero after p-lags. Parameters $\alpha_{n,k}$, $\alpha_{n,v}$, $\theta_{n,k}$, and $\theta_{n,v}$ need to be estimated.

Determining which model is the best fit is an empirical and iterative process in which we set the initial orders by analyzing ACF based on (3.1.13) and (3.1.14). Also, based on (3.1.15) we are aiming to find the part of correlation between y^s_{tk} and $y^s_{(t-p),k}$ that are not reflected by the values of $\{y_{((t-p),k)+1}, \ldots, y_{(t-1),k}\}$ by analyzing PACF. Then by applying the orders of AR and MA, we can estimate the unknown parameters by model fitting. Of course, in our context we know the ground truth model so this analysis is for illustrative purposes.

In the following section, we explain the synthetic data generation.

4.4. Synthetic Data Generation Process

The forecasting process is defined based on the shipper-demand requests and carrier-capacity offers' data types generated by the system. In Fomeni et al. (2021) they provide synthetic data for the optimization model. In this regard, in this research, we generate the data synthetically based on past values of requests and offers by considering both sides' physical and temporal attributes.

From physical attributes, each shipment and capacity is measured with mentioned characterizations in Section 2.2. Each shipper-demand request contains one shipment. We generate the data by reflecting the volume of freights and capacities. From temporal features, the decision-maker works over the operational horizon where shipments and capacities are available for decisions within their related time windows. We take the pickup time of loads and vehicles' starting work time as their availability time. Other time attributes are fixed in a rule-based manner, similar to Fomeni et al. (2021). From that, the availability time of both sides defines the time dimension, which is discretized in equal intervals and falls in operational time periods.

We generate the volume data for N number of shippers and M number of carriers with their corresponding requests and offers. We assume the loads and capacities are uncorrelated when generating the data for each side. In practice, the captured data may contain cross-correlation within different shipments concerning unknown effects.

The decision-maker works between two specific terminals, and requests and offers are frequent per period. In this case, we will have univariate time-series data per shipper and per carrier between two terminals. According to the autoregression model, we generate the data with respect to each request and offer as a function of their past values. In our results, we use a second-order autoregressive model. For shippers we use (4.3.1) with $p = 2$ lags apart. For carriers we use (4.3.2) with $p = 2$ lags apart. For both sides, we assume that the errors are i.i.d. Normal with $\mu = 9.5$ and $\sigma = 3.16$ for shipper demands, and $\mu = 100$ and $\sigma = 33.33$ for carriers capacity offers.

Since we are considering the deterministic components of time, we should add the level of time-series α_{0k}, and α_{0v}. That means the forecasting process should follow the trend pattern in the past, and we should consider this to make the series stationary. We assume the generated data follows the trend pattern of its past.

We define the following pseudo algorithm to generate the synthetic data records for the N number of shippers. For the carriers' side, we follow the same process.

Algorithm Data generation

Input: Parameters $\alpha_{0k}, \alpha_{1k}, \alpha_{2k}$,
parameters μ, and σ,
values of $y^s_{(t_0-1)}, y^s_{(t_0-2)}$, T,
m number of data points,
n number of shippers;
Output: $(m \times n)$ generated time-series DataFrame;
Initialize shippers' DataFrame;

```
1: for each number of shippers do
2:     for each number of data points do
3:         Call autoregressive function y_t^s (4.3.1);
4:         Truncate;
5:         Call discrete time-series data points generator;
6:         Append y_t^s to shippers' DataFrame.
7:     end for
8: end for
```

There are two truncated normal distributions for both sides' values, where volumes are randomly generated within their related bounds (line 4). It does so by generating a random distribution of ε_t with constant μ and σ. Since 97% of values fall in 3σ from the μ, to more satisfying the interval bounds, we calculate the parameters as half of the interval length for mean and 1/3 of this for standard deviation. The remaining 3% of values are set to the bounds. All values are greater than or equal to zero. We should note that for the synthetic data generation, we set the constant term mentioned in (4.3.1), and (4.3.2) to zero to satisfy the boundaries; in fact, the calculated μ at the start point y^s_{tk}, and y^c_{tv} can be taken as the level of the normally distributed data.

In a real setting, the captured data usually contains censored data, missing values, and outliers that we need to deal with before modeling. Furthermore, the data reliability during the data fusion section needs to be assessed. In this work, we assume that requests are not constrained by capacity, so the data is not censored or truncated. We do not simulate the effect of any other data issues either. In other words, we consider the data to be an accurate representation of demand and capacities, respectively. In the next chapter, we generate the data, describe the implemented forecasting models for both sides, analyze the results and compare the solutions obtained using the different forecasts.

Chapter 5

Computational Results

In this chapter, we describe the experimental phase of our research. We start by presenting the data produced by the synthetic time-series data generator introduced in Section 4.4. Recall that we use this data because we do not have access to real data.

We estimate the parameters of the forecasting models presented in Section 4.3. We then assess the ARIMA models in Section 5.2 by estimating orders and parameters, and by analyzing their ACF, PACF, and AIC. We compare the results of this performance evaluation to select the models to use in evaluating the impact of forecasting precision on the quality of the optimization results. We select the best-performing model, as well as a model with a significantly lower performance in terms of forecasting errors, to provide the data, requested volumes and offered volumes, to the optimization model (Section 4.1).

In Section 5.3, we finally assess the impact of forecasting on the solution quality of the optimization model, by solving it with the data provided by the two selected ARIMA models, and by comparing performances based on the total cost, percentage error, average error, and $10-90$ percentile error range.

5.1. Synthetic Data Generation

In Section 4.4, we introduced two second-order AR models to generate data for shippers and carriers. We set the initial parameters for (4.3.1) and (4.3.2) and generate the data. Table 5.1 shows the initial parameter setting for generating the data for the requests and offers' values. We set the initial values for parameters α_1 and α_2 based on Yule-Walker rule for stationary second order AR model where the parameters rely in the triangle $-1 \leq \alpha_2 \leq 1-|\alpha_1|$ (Eshel, 2003). We set the initial value of $y_{t_0}-1$ and $y_{t_0}-2$ in the requests and offers' values range the same as in Fomeni et al. (2021). We set the $T=10$ based on the operational horizon. We set the initial values of μ and σ to satisfy the shipments and capacities' volume range mentioned in 4.4.

Table 5.1. Initial parameter setting for shippers and carriers' datasets

	Requests	Offers
N. records	1010	1010
α_1	-0.25	-0.25
α_2	0.35	0.35
y_{t_0-1}	1	50
y_{t_0-2}	2	70
T	10	10
μ	9.5	100
σ	3.16	33.33

The total 1,010 data records were generated for each of the 285 shippers (requests) and 39 carriers (offers). The data is generated within $[1, 20]$ and $[50, 250]$ intervals for shippers and carriers, respectively. For each time series, we use 1,000 data points for training the models and 10 data points for the test set as ground truth. In other words, we do not perform cross-validation.

Figure 5.1 shows the data generated for the two time series and a subsample of 200 observations. We check the forecastability by plotting each time series correlogram and verifying the autocorrelation within the data. Figure 5.1 illustrates the autocorrelation of requests and offers of a shipper and a carrier, which is notable by two past lags significance from time $t = 0$ with 95% confidence interval. We use the synthetic generated data to feed the forecasting models in the following section.

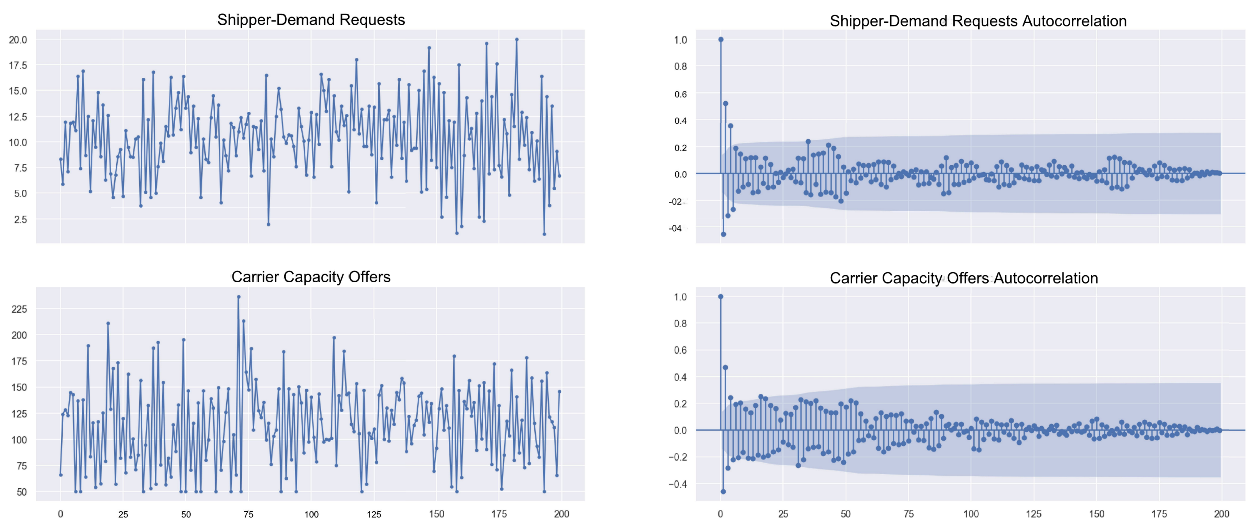

Fig. 5.1. Autocorrelation of request and offer values.

5.2. Forecasting Results

In this section, we are following the forecasting process explained in Section 4.3. Recall that we generate data according to an AR model. Using that data, we estimate ARIMA(p,d,q) models. Lacking access to real data, the purpose is to produce simple forecast models that exhibit some error and then evaluate the impact of those forecast errors on the solutions to the optimization problem. Given that we use simple statistical models, we can use classic hypothesis tests.

5.2.1. Stationary Distribution Verification

By assumption, the generated data based on Section 4.4 is stationary. Nevertheless, in this section, we validate the generation process by testing the stationary of the time series. In this context, y^s_{tk} and y^c_{tv} are stationary if the joint probability distribution of n values of y^s_{tk} and y^c_{tv} are on the same level with a set of p-shifted n observations. From that, we use the Augmented Dickey-Fuller (ADF) test, which is a statistical significance test. Based on Fuller (1976), the Dickey-Fuller (DF) test is a statistical test on a first-order autoregressive model where the null hypothesis is non-stationary existence within the data when the lag parameter α_{1k} is one. The alternative hypothesis is the stationary of the time series. For such a representation of

$$y^s_{tk} = \alpha_{0k} + \beta_k t + \alpha_{1k} y^s_{(t-1),k} + \varepsilon_{tk}, \tag{5.2.1}$$

we can describe DF as

$$\nabla y^s_{tk} = y^s_{tk} - y^s_{(t-1),k} = \alpha_{0k} + \beta_k t + \gamma y_{(t-1),k} + \varepsilon_{tk}, \tag{5.2.2}$$

where $\gamma = (\alpha_{1k} - 1)$, and y^s_{tk} is explained by a linear relationship of the previous lag with a trend time component $\beta_k t$ and a level α_{0k}.

The goal is to find out if we can remove the dependency of the value of y^s_{tk} on its previous lag $y^s_{(t-1),k}$ in (5.2.1) by differentiating them in (5.2.2). The null hypothesis is when we assume the existence of an explanatory variable which $\alpha_{1k} = 1$ and $\gamma = 0$, and then, the alternative hypothesis is when $-1 < 1 + \gamma < 1$. The null hypothesis can be rejected by a standard t-test where the p-value returns less than 0.05.

Using regression ∇y^s_{tk} as a first differentiating lag provides weak stationary because there may be still an autocorrelation within the error terms, and ADF is introduced to use p-lags past in the model $\nabla y_{(t-p),k}$ with the same statistical test as

$$\nabla y^s_{tk} = \alpha_{0k} + \beta_k t + \gamma y^s_{(t-1),k} + \sum_{n=1}^{p} \rho_{n-1} \nabla y_{(t-n+1),k} + \varepsilon_{tk}, \tag{5.2.3}$$

where ρ is the differentiating parameter at lag n. Several approaches were introduced in the literature for the ADF test. We use the Akaike information criterion (AIC) approach (Sakamoto et al., 1986).

Table 5.2 reports the ADF test results for a shipper and a carrier data requests and offers. The ADF test statistic using the AIC method returns the minimum value of information criteria -18.441 and -18.523 for the requests and offers' values, respectively. The prominence of p-value is noticeable where the value is much less than 0.05, and we can strongly reject the null hypothesis and verify the stationary of the time-series data. Choosing 1-past lag shows the order needed to apply for the regression to achieve the minimum value of the AIC method by using (5.2.2), for the 0.1, 0.01, 0.05 critical values. The fairly similar values of requests and offers in the table are due to the normal distribution assumed in both cases.

Table 5.2. Augmented Dickey-Fuller test for a shipper and a carrier datasets.

	Requests	Offers
ADF test statistic	-18.441	-18.523
p-value	$2.159e^{-30}$	$2.107e^{-30}$
p-lags	1	1
n-observations	1008	1008
0.01 critical values	-3.436	-3.436
0.05 critical values	-2.864	-2.864
0.1 critical values	-2.568	-2.568

By verifying the stationary of the time series we can use the ARIMA-based models without differencing order explained in the next section.

5.2.2. Model Selection

For model selection, we can analyze the ACF and PACF to determine the (p,d,q)-orders. We use two ARIMA models for forecasting requests and offers values over the following time periods. The first model is an ARIMA(p,d,q) model, and the second is an AR model based on ARIMA$(p,q,0)$ aiming to build a model with a higher forecasting error rate. Based on Section 4.3, we set the order of differencing d to zero to select the best model.

We estimate PACF by computing $\hat{\phi}(p)$ from equation (3.1.15) for each lag with its past lags without considering the lags in-between, to select the p-order of the AR model. That means the total number of lags of $((N/2) - 1)$ we can consider in computing. Table 5.3 demonstrates the PACF estimation for a sample of past 10-lags for both datasets. Since forecasting for new requested volumes and offered volumes follows the same process, we show y^s_{tk} and y^c_{tv} by y_t in forecasting results, for the sake of simplicity. We depict the results for a shipper-demand requests and a carrier-capacity offers data, then we forecast $\hat{\boldsymbol{y}}_{t+1}, \ldots, \hat{\boldsymbol{y}}_{t+10}$ for N number of shippers and M number of carriers, as explained in Section 4.1.

Table 5.3. Estimated partial autocorrelation function for 10-lags of a shipper and a carrier time series data.

Lags	Requests	Offers
y_{t-1}	$-4.509e^{-1}$	$-3.628e^{-1}$
y_{t-2}	$3.653e^{-1}$	$3.322e^{-1}$
y_{t-3}	$2.451e^{-2}$	$1.433e^{-2}$
y_{t-4}	$2.870e^{-3}$	$-3.021e^{-2}$
y_{t-5}	$1.201e^{-2}$	$1.022e^{-2}$
y_{t-6}	$-3.012e^{-2}$	$1.441e^{-2}$
y_{t-7}	$2.557e^{-2}$	$-5.403e^{-2}$
y_{t-8}	$5.875e^{-2}$	$6.646e^{-3}$
y_{t-9}	$-1.728e^{-2}$	$8.018e^{-3}$
y_{t-10}	$-3.807e^{-3}$	$-3.444e^{-2}$

Fig. 5.2. Requests and offers values partial autocorrelation correlogram.

We use the statistical significance test to select the p-order of the AR models by 95% confidence interval. We plot the PACF correlogram of both sides in Figure 5.2. For the sake of clarity, we plot 100 observations. The two past lags significance from $t = 0$ is clear, we can also see after 2-lags the process sharply cutoff to zero. That means we can estimate the AR model with order of 2 for both sides.

To select the q-order of the MA model, we estimate ACF by computing $\hat{\rho}(q)$ using equations (3.1.13) and (3.1.14), for each lag. Table 5.4 shows the ACF estimation for a sample of past 10-lags of both sides' data. In the estimated values, we aim to determine how many lags are needed to take into account to eliminate any autocorrelation within the generated error terms. In that sense, we use the statistical significance test to select the q-order of the MA models by 95% of a confidence interval. Figure 5.1 points out that the autocorrelation correlogram is symmetric plotting; we consider just the positive lags. The

Table 5.4. Estimated autocorrelation function for 10-lags of a shipper and a carrier time series data.

Lags	Requests	Offers
y_{t-1}	−0.450	−0.362
y_{t-2}	0.494	0.420
y_{t-3}	−0.289	−0.211
y_{t-4}	0.260	0.164
y_{t-5}	−0.164	−0.091
y_{t-6}	0.116	0.074
y_{t-7}	−0.062	−0.083
y_{t-8}	0.081	0.059
y_{t-9}	−0.053	−0.053
y_{t-10}	0.054	0.013

past two lags from the $t = 0$ time period are significant; hence we can estimate the order of 2 for the MA model.

In practice, finding orders of the model is an iterative process by fitting the model with the estimated orders and using statistical tests, which may lead us to overestimate the orders. To make our estimation more accurate, we use the AIC method for selecting the orders, the same as when we used the AIC method for the statistical significance test, instead of the standard t-distribution test in the stationary analysis. In general, AIC for estimating p and q orders are computed as

$$AIC = \frac{2(k - \ln(\hat{L}))}{N} \approx \frac{N \ln(\hat{\sigma}_a^2) + 2k}{N}, \tag{5.2.4}$$

where $\hat{L}$ is the maximum likelihood estimation of σ_a^2 denoted by $\hat{\sigma}_a^2$, N is the number of observations contributed in estimation, $k = p + q + 1$ is the number of estimated orders which may include a constant term (Box et al., 2015). We aim to minimize the residual variance $\hat{\sigma}_a^2$ and iteratively adding k to the model as a penalty term to find the optimal. Consequently, the best estimation of p and q orders is when the equation is minimized, and we can take it as criteria for the model selection. Tables 5.5 and 5.6 show the model selection information criteria using the AIC method with their related estimated parameters for the requests and offers data.

Even though all AR models in Tables 5.5 and 5.6 satisfy the standard t-test, we can see that each model's AIC criteria are different. We select the model for forecasting with the minimum AIC criteria, ARIMA(2,0,0) for both datasets. It can also be verified by comparing the truth models' orders highlighted in the tables. In terms of comparing parameters as another indicator, we set $\alpha_1 = -0.25$, and $\alpha_2 = 0.35$ to generate the ground truth for both shippers and carriers, depicted in Table 5.1. The selected model for shippers' data inferred the parameters better than the other models compared to the data generator's parameters.

Table 5.5. Model selection information criteria and estimated parameters of shippers' dataset.

	AIC	α_1	α_2	α_3	θ_1
ARIMA(1,0,0)	5235.1	-0.44	0	0	0
ARIMA(2,0,0)	5093.2	-0.28	0.36	0	0
ARIMA(3,0,0)	5094.5	-0.29	0.37	0.02	0
ARIMA(0,0,1)	7249.8	0	0	0	-0.26
ARIMA(2,0,1)	5094.5	-0.21	0.39	-0.07	0
ARIMA(1,0,1)	5141.2	-0.82	0	0	0.46
ARIMA(3,0,1)	5096.5	-0.29	0.37	0.02	-0.0014
ARIMA(2,2,0)	6119.0	-1.32	-0.51	0	0

Table 5.6. Model selection information criteria and estimated parameters of carriers' dataset.

	AIC	α_1	α_2	α_3	θ_1
ARIMA(1,0,0)	9925.9	-0.36	0	0	0
ARIMA(2,0,0)	9812.0	-0.24	0.33	0	0
ARIMA(3,0,0)	9813.7	-0.24	0.33	0.017	0
ARIMA(0,0,1)	11891.3	0	0	0	-0.21
ARIMA(2,0,1)	9813.9	-0.22	0.33	0	-0.018
ARIMA(1,0,1)	9852.6	-0.78	0	0	0.48
ARIMA(3,0,1)	9815.7	-0.24	0.33	0.017	0.0001
ARIMA(2,2,0)	10672.1	-1.33	-0.53	0	0

The estimated parameters are similar in the three models in the carriers' data, which indicates that those models may produce similar forecasts. We predict the successive 10 time periods on each time series data by fitting the selected model. Table 5.7 illustrates the forecasted new requested volumes and offered volumes, the ground truth, and the residuals for each time period.

In this section, we used several tests to find the best orders of ARIMA-based models and provide the best estimation of parameters by comparing the models to the ground truth values. In the next section, we analyze the performance of represented forecasting models with the help of two performance measurement metrics to select the two models.

5.2.3. Forecast Performance Analysis

Depending on the data under study, selecting which performance metric to use can be different. To analyze the residuals, we use regression metrics to assess the performance. We use two data parameter-free metrics to evaluate the performance because we have two different datasets for shippers and carriers with varying data parameters. We use the minimum-maximum (Min_Max) error metric for the first performance measurement and the mean absolute percentage error (MAPE) for the second metric. Since the generated and predicted

Table 5.7. Forecasted new requested volumes and offered volumes for the next 10 time periods using ARIMA(2,0,0).

	Requests			Offers		
Lags	Forecasted	Ground truth	Residuals	Forecasted	Ground truth	Residuals
y_{t+1}	13.8	13.3	-0.5	94.1	100.9	6.8
y_{t+2}	7.3	8.7	1.4	122.0	149.1	27.1
y_{t+3}	15.1	12.1	-3	103.5	125.8	22.3
y_{t+4}	1.0	9.5	8.5	117.2	134.2	17
y_{t+5}	11.9	11.5	-0.4	107.8	102.6	-5.2
y_{t+6}	7.4	9.9	2.5	114.6	66.3	-48.3
y_{t+7}	10.4	11.1	0.7	109.9	102.6	-7.3
y_{t+8}	10.6	10.2	-0.4	113.3	121.6	8.3
y_{t+9}	15.9	10.9	-5	110.9	72.5	-38.4
y_{t+10}	7.7	10.3	2.6	112.6	130.1	17.5

values vary over the mean, we aim to see how the models perform for very small and very large volumes of requests and offers through the Min_Max Error function. It does so by averaging the minimum value of actual requests or offers concerning the minimum value of forecasted demands or offers $\hat{y}_t$ over the maximum of both sides' actual and predicted values. The minimum-maximum error function over schedule length T is defined as

$$Min\_Max\,Error(y_{t|T}, \hat{y}_{t|T}) = 1 - \frac{1}{T}\sum_{n=1}^{T}\frac{\min(y_{t+n}, \hat{y}_{t+n})}{\max(y_{t+n}, \hat{y}_{t+n})}. \tag{5.2.5}$$

We use the MAPE function (5.2.6) as the second performance metric to show the error percentage of predicted requests and offers' values. The MAPE metric returns the proportional error of estimated values over the schedule length T,

$$MAPE(y_{t|T}, \hat{y}_{t|T}) = \frac{1}{T}\sum_{n=1}^{T}\frac{|y_{t+n} - \hat{y}_{t+n}|}{\max(\epsilon, |y_{t+n}|)}, \tag{5.2.6}$$

where ϵ is a small constant to avoid deviation by zero when the system observed zero values.

We forecast the new requested volumes and offered volumes for the successive 10 time periods by fitting the models in Table 5.6 and then computing the Min_Max Error and MAPE to verify the selected ARIMA(2,0,0) by our previous analysis. Table 5.8 reports the performance evaluation results for a shipper-demand requests and a carrier-capacity offers' volumes.

Since we analyze the forecasting residuals in Table 5.8, the lower the value, the better it is. The selected model, ARIMA(2,0,0), performs better in both evaluation metrics than the other models. From this, as for the accuracy of the chosen model, the Min_Max Error metric shows 77.6% accuracy in observing very small or very large shipments' volumes. On the other hand, the MAPE measures 0.6% sensitivity regarding the actual values with their corresponding forecasted values. This evaluation confirms our previous analysis regarding

Table 5.8. Performance evaluation results for the successive 10 time periods of a shipper and a carrier time series data.

	Requests		Offers	
	Min_Max Error	MAPE	Min_Max Error	MAPE
ARIMA(1,0,0)	0.263	0.156	0.171	0.220
ARIMA(0,0,1)	0.275	0.182	0.175	0.224
ARIMA(2,0,0)	0.224	0.006	0.163	0.213
ARIMA(3,0,0)	0.226	0.009	0.165	0.214
ARIMA(2,0,1)	0.226	0.009	0.164	0.214
ARIMA(1,0,1)	0.226	0.009	0.152	0.202
ARIMA(3,0,1)	0.232	0.013	0.165	0.214
ARIMA(2,2,0)	0.573	0.268	0.536	0.555

ACF, PACF, and AIC criteria for selecting the ARIMA(2,0,0) to provide new requested volumes' information for the decision-maker. Both metrics show very close errors for models with orders (3,0,0), (2,0,1), and (1,0,1), because the data is normally distributed.

The ARIMA(1,0,1) performs better for the carriers' dataset in both performance metrics regarding very small and very large capacities' values and the relative errors at all forecasted time periods. To find the best model we increased the number of data records to 10,000 and repeated the performance evaluation steps. In the case of ARIMA(2,0,0), the Min_Max Error metric returns 0.136, and shows 89.1% of accuracy that is 2.3% more accurate than the ARIMA(1,0,1) with 0.159 of error. The MAPE returns 18.6% of sensitivity in terms of actual values with their related forested, which performs better than the ARIMA(1,0,1). From that, we select the ARIMA(2,0,0) to provide new offered volumes' information for the decision-maker.

In evaluating the other models, ARIMA(0,0,1) tends to the mean after 1-step ahead forecasting over both shippers and carriers' datasets, which shows the property of MA models. We observe, however, that, with the exception of ARIMA(2,2,0), all the other models are not significantly worse than the best one. Therefore, in order to provide more insights into the impact of prediction errors on optimization, we select the worst model, ARIMA(2,2,0), which is expected to perform poorly. The model yields Min_Max Error metrics of 57.3% and 53.6% error, respectively. As to MAPE metric, the ARIMA(2,2,0) returns 26.8%, and 53.6% of sensitivity with respect to the ground truth with their corresponding forecasted values.

We use the two selected models to forecast new requested volumes and offered volumes over the operational horizon and create matrices $\boldsymbol{Y}^s$ and $\boldsymbol{Y}^c$ explained in Section 4.1 to provide information for the optimization modeling.

5.3. Impact of Forecast Errors on Solution Quality

In this section we report the main results. We evaluate the impact of forecasting results on the performance of the optimization model 1/V/V/S/N[I2B] described in Section 4.1. We solve the (4.1.3)-(4.1.9) by HM1, then analyze the impact of forecast errors on solution quality. We measure this impact for each instance based on percentage error δ that is computed for a given forecasting model in three steps:

1. Solve (4.1.3)-(4.1.9) using ground truth values, $\mathbf{Y}^s$, and $\mathbf{Y}^c$, giving solution value G.
2. Solve (4.1.3)-(4.1.9) using forecasts $\hat{\mathbf{Y}}^s$ and $\hat{\mathbf{Y}}^c$.
3. Evaluate the cost F according to (4.1.3)-(4.1.9) of the solution obtained in Step 2 using ground truth values $\mathbf{Y}^s$ and $\mathbf{Y}^c$. Note that in case of excess demand, it is covered by spot-market vehicles at a cost.

The resulting metric is then

$$\delta = \frac{F - G}{G} \cdot 100. \tag{5.3.1}$$

Table 5.9 reports the calculated total costs for Set 1 and Type 1 in Fomeni et al. (2021) instance over 10 successive time periods. We use forecasts generated by ARIMA(2,0,0) and ARIMA(2,2,0) to provide data for aforementioned three steps.

As for illustrating the performance of the forecasting models over different experiments we use average forecast error, mean absolute relative error, and $10th$ and $90th$ percentile forecast error metrics for shippers' requested volumes across the operational horizon on Table 5.10. We also use aforementioned performance metrics for carriers' offered volumes reported in Table 5.11.

Table 5.9. Total assignment cost (ground truth G and using forecasts F) over ten time periods.

	Total assignment cost		
(Requests, Offers)	ARIMA(2,2,0)	ARIMA(2,0,0)	Ground truth
(100, 97)	12641	10852	9984
(200, 51)	24087	23625	23600
(300, 84)	36169	33573	32868
(500, 11)	57308	56522	55207
(750, 74)	100224	95182	90594
(1000, 73)	142490	129150	121576

Table 5.12 shows the computed percentage error for the two models along with the forecast errors on the 10 days horizon in Table 5.10 and Table 5.11. As for analyzing the impact of forecasting errors on the solution quality, we use the same two forecasting models to produce data from Section 5.2.2. On the first experiment using $(100, 97)$ of requests and offers, ARIMA(2,0,0) shows 17.92% better performance in comparison to ARIMA(2,2,0) with respect to load-to-carrier selection and assignment decisions. On the other hand, by

Table 5.10. Forecast errors produced by ARIMA(2,2,0) and ARIMA(2,0,0) for shippers' requested volumes over the operational horizon.

	ARIMA(2,2,0)				ARIMA(2,0,0)			
Requests	Mean Absolute Error (MAE)	Mean Absolute Relative Error	10th Percentile Error MAE	90th Percentile Error MAE	Mean Absolute Error (MAE)	Mean Absolute Relative Error	10th Percentile Error MAE	90th Percentile Error MAE
100	6.41	0.78	1.0	14.5	2.92	0.58	0.55	6.60
200	6.48	0.76	1.0	14.0	2.91	0.41	0.40	6.45
300	7.53	0.85	1.0	17.0	3.03	0.42	0.50	6.10
500	8.07	0.94	1.0	17.0	2.83	0.39	0.50	5.85
750	8.04	0.95	1.0	18.0	2.81	0.41	0.40	5.70
1000	7.74	0.89	0.4	17.0	2.84	0.37	0.40	5.70

Table 5.11. Forecast errors produced by ARIMA(2,2,0) and ARIMA(2,0,0) for carriers' offers volumes over the operational horizon.

	ARIMA(2,2,0)				ARIMA(2,0,0)			
Offers	Mean Absolute Error (MAE)	Mean Absolute Relative Error	10th Percentile Error MAE	90th Percentile Error MAE	Mean Absolute Error (MAE)	Mean Absolute Relative Error	10th Percentile Error MEA	90th Percentile Error MAE
97	69.39	0.72	11.5	131.4	30.18	0.33	5.8	54.1
51	64.55	0.56	6.1	143.1	23.14	0.21	4.2	51.6
84	50.31	0.49	8.7	110.5	24.39	0.26	3.0	52.9
11	110.51	1.26	60.2	171.5	22.0	0.29	4.4	42.4
74	86.86	0.88	13.6	184.6	27.32	0.27	4.3	54.4
73	155.81	1.63	25.3	270.6	31.95	0.37	8.2	60.1

Table 5.12. Percentage error of objective function value due to forecasting.

	Percentage Error Objective Function	
(Requests, Offers)	ARIMA(2,2,0)	ARIMA(2,0,0)
(100, 97)	26.61	8.69
(200, 51)	2.06	0.10
(300, 84)	10.04	2.14
(500, 11)	3.80	2.38
(750, 74)	10.62	5.06
(1000, 73)	17.20	6.22

referring to Table 5.10 and Table 5.11, we can see that on the same experiment using 100 requests and 97 offers, the ARIMA(2,0,0) returns 3.94 lower average forecast error for requests and 39.18 lower average forecast for offers, as well as, 0.20 lower average relative forecast error for requests and 0.39 lower average relative error for offers. As for $10th$ and $90th$ percentile metrics, the results for the same experiment show 0.45 lower in $10th$ percentile forecast error, and 7.9 lower in $90th$ percentile error using 100 requests for ARIMA(2,0,0) compare to the ARIMA(2,2,0) model. In addition, for the carrier's side, the result for 97 offers shows 5.7 lower in $10th$ percentile forecast error and 77.3 lower in $90th$ percentile error for ARIMA(2,0,0) compare to the ARIMA(2,2,0) model. That means the selected model with lower forecast error performs better than ARIMA(2,2,0) regarding assignment decisions. To verify that, we compute the percentage errors and forecast errors by experimenting with different sets illustrated in Tables 5.10, 5.11 and 5.12. Going down the tables, the result shows lower percentage errors and forecast errors using the ARIMA(2,0,0) beating the ARIMA(2,2,0) by a handy margin. In other words, the selected model in all experiments outperforms the ARIMA(2,2,0) model.

The mean absolute relative errors of ARIMA(2,2,0) are in the range 49-163% and ARIMA(2,0,0) in the range 21-58% considering both shippers and carriers. The resulting deterioration in solution quality nevertheless range between 0.1-8.7% for ARIMA(2,0,0) and 2.1-26.6% for ARIMA(2,2,0). Optimizing assignment and consolidation decisions hence allows to reduce the magnitude of the impact of forecast errors.

Chapter 6

Conclusion and Future Work

The task of forecasting shippers' requested volumes and carriers' offered volumes for operational planning of M1M systems is inevitable and a necessity considering the uncertainty of new information involved. This work focused on assessing the impact of forecast errors on operational planning quality. Several autoregressive moving average approaches, where we forecast new requested volumes and offered volumes characterized by their physical and time attributes over the following time periods of the operational horizon. We generated the quantile over quantile random normal distributed time series data for both sides with respect to predefined boundaries. We systematically evaluated the performance of several ARIMA models following a forecasting process to select the best model as the forecaster. The forecaster provided new information for the decision-maker. We evaluated the impact of forecasting errors on the optimization model, where we computed, e.g., the total cost, percentage errors and average errors.

To summarize our contribution, we provide the data for the optimization model by using two ARIMA models from the forecasting process. We solve the optimization model using ground truth and forecast values and evaluate the cost by computing percentage error for the two ARIMA models. We analyzed the impact of forecast errors produced by these two models on the solution quality. We measured the impact based on the percentage error measurement metric. We also analyzed the forecast errors produced by the forecasting models for different sets for both sides' requests and offers' volumes over the operational horizon. Our analysis showed how the model with lower forecast errors could positively impact the decisions. While there is an increase in forecasting accuracy, we observed that the decision-maker could rely on the forecasting provided data as the known information to use in its decisions. We concluded that 1) even an imprecise forecast could be useful (the magnitude of the impact on solution quality is smaller than the magnitude of the forecast errors), at least in the context of our experiments, and 2) optimizing the assignment and consolidation decisions over several periods allows to reduce the magnitude of the impact of forecast errors.

Future work should continue to focus on the impact of forecasting errors on optimization modeling, in particular using data-driven approaches which handle time-series data, e.g., machine learning and neural networks. In real applications, when the actual model is unknown, estimating forecast models' parameters is more challenging than using the synthetic generated data in this research. Hence we should investigate the results according to different true and forecast models. Moreover, in general, future work should be dedicated to finding the best forecasting models for complex optimization-based decision-support methods to insure the reliability of forecasting provided data. Moreover, future work should focus on extending forecasting models to provide information for the decision-maker that works in general freight transportation networks consists of several origins and destinations which leads us to a time-space forecasting problem. We also extend our focus on learning the forecasting models jointly with the optimization modeling as the long perspective of future work.

References

Bayraktar, E., Koh, S. L., Gunasekaran, A., Sari, K., and Tatoglu, E. (2008). The role of forecasting on bullwhip effect for e-scm applications. *International Journal of Production Economics*, 113(1):193–204.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (1970). *Time series analysis: forecasting and control*. John Wiley & Sons.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Brown, R. G. (1959). *Statistical forecasting for inventory control*. McGraw/Hill.

Crainic, T. G., Gendreau, M., and Rei, W. (2021). *Optimization for Operational Planning of M1M Systems on Hyper-Corridors*. CIRRELT, Centre interuniversitaire de recherche sur les réseaux déntreprise, la logistique et le transport, Université de Montréal.

Crainic, T. G. and Laporte, G. (1997). Planning models for freight transportation. *European Journal of Operational Research*, 97(3):409–438.

Dyckhoff, H. (1990). A typology of cutting and packing problems. *European Journal of Operational Research*, 44(2):145–159.

Eshel, G. (2003). The yule walker equations for the ar coefficients. *Internet Resource*, 2:68–73.

Fields, E., Osorio, C., and Zhou, T. (2021). A data-driven method for reconstructing a distribution from a truncated sample with an application to inferring car-sharing demand. *Transportation Science*, 55(3):616–636.

Fomeni, F. D., Crainic, T. G., and Rei, W. (2021). Multi-period bin packing model and effective constructive heuristics for corridor-based logistics capacity planning. *Computers & Operations Research*, 132:105308.

Fuller, W. A. (1976). *Introduction to statistical time series*. John Wiley & Sons.

Gardner, J. R. and Everette, S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, 4(1):1–28.

Godfrey, G. A. and Powell, W. B. (2000). Adaptive estimation of daily demands with complex calendar effects for freight transportation. *Transportation Research Part B: Methodological*, 34(6):451–469.

Guo, W., Crainic, T. G., Gendreau, M., and Rei, W. (2021). *Operational Planning for Many-to-One-to-Many Freight Transportation*. CIRRELT, Centre interuniversitaire de recherche sur les réseaux déntreprise, la logistique et le transport, Université de Montréal.

Guo, Y. H., Shi, X. P., and Zhang, X. D. (2010). A study of short term forecasting of the railway freight volume in china using arima and holt-winters models. In *2010 8th International Conference on Supply Chain Management and Information*, pages 1–6. IEEE.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5–10.

Laage, G., Frejinger, E., and Savard, G. (2021). Periodic freight demand estimation for large-scale tactical planning. *ArXiv Preprint ArXiv:2105.09136*.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS One*, 13(3):e0194889.

Martello, S. and Toth, P. (1990). *Knapsack problems: algorithms and computer implementations*. John Wiley & Sons, Inc.

Pradita, S. P., Ongkunaruk, P., and Leingpibul, T. D. (2020). Utilizing an intervention forecasting approach to improve reefer container demand forecasting accuracy: A case study in indonesia. *International Journal of Technology*, 11(1):144–154.

Ren, S., Choi, T. M., Lee, K. M., and Lin, L. (2020). Intelligent service capacity allocation for cross-border-e-commerce related third-party-forwarding logistics operations: A deep learning approach. *Transportation Research Part E: Logistics and Transportation Review*, 134:101834.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986). Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel*, 81(10.5555):26853.

Schulze, P. M. and Prinz, A. (2009). Forecasting container transshipment in germany. *Applied Economics*, 41(22):2809–2815.

Systematics, C. (1997). *A guidebook for forecasting freight transportation demand*, volume 388. Transportation Research Board.

Taherkhani, G., Bilegan, I. C., Crainic, T. G., Gendreau, M., and Rei, W. (2021). Tactical capacity planning in an integrated multi-stakeholder freight transportation system. Accepted for Omega.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3):324–342.

Xiao, Y., Liu, J. J., Hu, Y., Wang, Y., Lai, K. K., and Wang, S. (2014). A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting. *Journal of Air Transport Management*, 39:1–11.

Yue, S. and Pilon, P. (2003). Interaction between deterministic trend and autoregressive process. *Water Resources Research*, 39(4).

Yule, G. U. (1927). Vii. on a method of investigating periodicities disturbed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 226(636-646):267–298.