A combined forecasting and packing model for air cargo loading: A risk-averse framework

Tseremoglou, I.; Bombelli, A.; Santos, Bruno F.

DOI
10.1016/j.tre.2021.102579

Publication date
2022

Document Version
Final published version

Published in
Transportation Research. Part E: Logistics and Transportation Review

Transportation Research Part E 158 (2022) 102579

Contents lists available at ScienceDirect

Transportation Research Part E

journal homepage: www.elsevier.com/locate/tre

A combined forecasting and packing model for air cargo loading: A risk-averse framework

Iordanis Tseremoglou *, Alessandro Bombelli, Bruno F. Santos

Air Transport and Operations, Faculty of Aerospace Engineering, Delft University of Technology, Building 62 Kluyverweg 1 2629 HS Delft, Netherlands

ARTICLE INFO

Keywords:
Air cargo loading
Knapsack problem
Unit load devices
Prediction uncertainty
Deep neural networks
Extreme points

ABSTRACT

In this paper, we present a combined forecasting and optimization decision-support tool to assist air cargo revenue management departments in the acceptance/rejection process of incoming cargo bookings. We consider the case of a combination airline and focus on the passenger aircraft belly capacity. The process is dynamic (bookings are received in a discrete fashion during the booking horizon) and uncertain (for some bookings the three dimensions are not provided, while the actual belly space available for cargo is only revealed a few hours before departure). Hence, analysts base decisions on historical data or human experience, which might yield sub-optimal or infeasible solutions due to the aforementioned uncertainties. We tackle them by proposing data-driven algorithms to predict available cargo space and shipment dimensions. A packing problem is solved sequentially once a new booking request is received, predicting shipment dimensions, if necessary, and considering the uncertainty of such prediction. The booking is accepted if it results in a feasible loading configuration where no previously accepted booking is offloaded. When applied in a deterministic context, our packing method outperformed the one used by the partner airline, increasing the loaded volume up to 20%. The framework was also tested assuming unknown shipment dimensions, comparing a risk-prone and a risk-averse strategy, with the latter accounting for uncertainty in dimension predictions and the former using mean values. While the average loaded volume decreases in the risk-averse case, the number of unplanned offloadings due to under-predicted dimensions decreases from 54% to 12% of the simulated cases, hence yielding a more robust acceptance strategy.

1. Introduction

During the last decades, globalization has significantly fostered the growth of international air trade. Although goods transported by air correspond to only 1% of the overall transported volume, the percentage spikes to 35% if value is used as a parameter. This percentage translates to a value of US $5.5 trillion and annual revenues of US $50 billions for the IATA members (IATA, 2019). Notwithstanding this relevance, air cargo for combination airlines is generally considered as of secondary importance with respect to the passenger counterpart, although the COVID-19 pandemic revamped cargo operations for many combination carriers. This disparity is also reflected into the imbalance of academic works, especially the ones focused on big data and machine learning, addressing the passenger (Chung et al., 2020) and the cargo side. This disparity is also enhanced by the intrinsic differences in terms of network planning and supply–demand balance (Bombelli et al., 2020) between passenger and cargo operations.

* Corresponding author.
E-mail address: I.Tseremoglou@tudelft.nl (I. Tseremoglou).

https://doi.org/10.1016/j.tre.2021.102579
Received 8 June 2021; Received in revised form 28 October 2021; Accepted 11 December 2021
Available online 20 January 2022

Air cargo can be transported either in full freighter aircraft or in the belly space of passenger aircraft, through special pallets or containers called Unit Load Devices (ULDs). ULDs have specific shapes and sizes in order to fit into designated positions in the lower deck or belly space (passenger aircraft) or main deck and lower deck (full freighters). Full freighters are dedicated to cargo transportation and have a fixed and large capacity. On the other hand, in passenger aircraft, passenger baggage have priority over cargo, meaning that cargo capacity is more limited and subject to unexpected changes. In both aircraft types, cargo capacity is split into capacity reserved for long-term contracts (allotments) and capacity available for sale during the booking period (short-term capacity). Short-term capacity management in passenger aircraft generally poses the biggest challenges from a planning perspective: this is due to the high uncertainty both on the supply side (belly cargo space available) and the demand side (offloading, no-shows). Quite often, there is also a mismatch between the booked (expected) shipment dimensions and the shipment dimensions that are actually received by the airline. Our work addresses this problem.

The utilization of cargo capacity in a way that maximizes profit constitutes the overall objective of the air cargo load planning problem faced by airlines, as defined by Brandt and Nickel (2018). In general, the air cargo load planning problem involves many stakeholders such as the Revenue Management (RM) and Handling & Operations Department. In this paper we adopt the perspective of the RM department while considering operational requirements.

To achieve the aforementioned objective, a RM analyst has to make an informed decision on whether to accept or reject an incoming booking request. An incoming booking request may consist of one or multiple shipments, which may differ in size or not. In the context of our partner airline and related literature, we identify the following four questions that the RM analyst must answer during the whole booking process:

1. What is the **capacity** of the aircraft?
2. What are the exact **dimensions** of the shipments?
3. What is the **number** and **type** of ULDs that can maximize transported volume?
4. How to **pack** the shipments inside the ULDs to maximize the amount of shipments transported, minimizing the risk of having to offload accepted shipments for unplanned lack of space?

This decision-making process is further complicated by the following challenges. When considering passenger aircraft, the available cargo capacity is dynamic and fluctuates until the last day before departure, as it depends on several factors such as payload, belly space and the expected passenger baggage for this specific flight. Regarding the booking acceptance/rejection policy, some airlines, as our partner airline, do not even require shippers to provide shipment dimensions when submitting a request, but only volume, weight, and other pieces of information such as booking origin. Hence, the true shipment dimensions are not always known when operational planning is performed, but their true values are revealed shortly before the shipment is loaded onto the aircraft. Finally, the palletization strategy depends mostly on the skills of palletization workers rather than on automated systems (Wada et al., 2017), but this specific aspect entails an extra level of uncertainty that is beyond the scope of this work.

The above complexities can be summarized in the form of two nested problems that the RM analyst must solve when receiving a new booking request. The outer problem consists of **forecasting** the shipment dimensions (if not provided) and the available aircraft capacity, providing answers to questions 1 and 2. The inner is a **packing** problem, i.e., based on the results of the forecasting phase, the optimal ULD configuration in combination with the optimal palletization of shipments into ULDs should be determined, hence addressing decision types 3 and 4.

Our goal is to provide a decision-support tool for the air cargo booking process. We focus on tackling the two distinct problems with a combined forecasting and packing model. Although the model is shown here in the context of an air cargo application, we believe its applicability extends to other fields, such as the maritime or the rail industry. More specifically, the main contributions of this paper can be summarized as follows

- First, to the best of our knowledge, a novel modeling framework addressing all the phases of the air cargo booking process is developed, while explicitly considering uncertainty in both aircraft capacity and shipment dimensions.
- Second, a modified heuristic for the Three-Dimensional Multiple Heterogeneous Knapsack Problem is introduced. The introduced heuristic is based on the work of Crainic et al. (2008) and includes several enhancements to deal with the specific characteristics of the air cargo loading problem. The chosen heuristic proved to perform better than the original version and than another variant based on a different sorting strategy. Although not guaranteeing optimal solution due to its heuristic nature, the selected algorithm performs well according to some specific Key Performance Indicators (KPIs) and, more importantly, is fast enough for a real-time implementation.
- Lastly, we performed an extensive evaluation of the model using real booking data from five flights provided by our partner airline. The evaluation includes an assessment of the efficiency of the heuristic, a sensitivity analysis of the impact of the confidence level to the volume of the accepted shipments, and a comparison between the effects of a risk-prone and a risk-averse strategy in terms of differences between the accepted volume by RM and the loaded volume by operations.

The rest of the paper is organized as follows. In Section 2 the relevant academic literature and how our model relates to the existing research is presented. In Section 3, the problem definition, the methodology, and the relevant academic literature are presented. Section 4 describes into detail the developed forecasting models and their performance. Section 5 provides the implementation details of the packing models. The combined forecasting and packing model is presented in Section 6, along with case studies which shed light into the efficiency of the proposed model. Finally, Section 7 provides conclusions and recommendations for future work.

2. Literature review

While our work entails forecasting techniques and packing modeling, our major contributions reside in the latter research area. As such, this literature review will revolve around packing formulations with a focus on logistics and air cargo loading applications, while references to forecasting models and techniques are provided throughout the paper when necessary.

When it comes to packing problems, we identified two characteristics that strongly influence the modeling effort. The first characteristic pertains the temporal domain of the problem. Items can all be known in advance, or might be revealed throughout the planning horizon. The two situations refer, respectively, to a *static* or to a *dynamic* problem. The second characteristic pertains the confidence the decision-maker has with respect to the dimensions of the items and/or the number and type of bins to use. Dimensions can be *known* or *unknown*. Given the combination of the two characteristics, we can hence identify four research areas.

Most packing models address static problems where dimensions are known with certainty. Focusing on three-dimensional problems, Crainic et al. (2008) address the three-dimensional packing problem for a set of identical bins through an Extreme Point (EP) Heuristic. The EPs represent the possible locations that can accommodate new items inside a bin. Crainic et al. (2011) generalize the bin packing problem by considering variations in the size and cost of items and bins and introduce the Variable Cost and Size Bin Packing Problem (VCSBP). They proposed a methodology based on lower and upper bounds and managed to solve instances of up to 1000 items. In Paquay et al. (2014), a mixed integer programming for solving a three-dimensional Bin Packing Problem (3DBPP) for air cargo is proposed, introducing constrains for fragility, weight distribution, and high vertical stability and orientation. However, this model managed to solve instances of only 8 or 12 shipments. The same model was extended in Paquay et al. (2017) by testing multiple heuristics such as Relax-and-Fix (R&F), Insert-and-Fix (I&F), and Fractional Relax-and-Fix (FRF) and showed that two of those heuristics (I&F and FRF) produced promising results for instances up to 80 shipments, with computational time up to 35 min. Most recently, in Paquay et al. (2018), a tailored two-phase constructive EP heuristic of the three-dimensional Multiple Bin Size Bin Packing Problem (MBSBPP) was developed that solved instances of up to 100 boxes, with the computational time not exceeding 12 s. Readers are referred to Brandt and Nickel (2018) for an extensive review of packing models with heuristics that have been developed for the air cargo load planning problem during the recent years.

Focusing on static problems with uncertain dimensions, Crainic et al. (2014) proposed a stochastic extension of the VCSBP considering variations in the renting cost of bins as well as uncertainty related to shipments in terms of volume and presence. This model was extended by Crainic et al. (2016) to account for uncertainty on the availability of bins in terms of existence and number. They use a progressive hedge meta-heuristic to solve loading problems of up to 500 items and 10 bin types. Both of the aforementioned problems are formulated as two-stage integer stochastic programs with recourse, where the appropriate number and type of bins are selected the first stage, while at the second stage, when demand information are revealed, the assignment of items to bins and the acquisition of extra bins is examined.

To the best of our knowledge, the literature pertaining dynamic packing applications for air cargo is extremely scarce. If we relax the application domain, on the other hand, dynamic bin packing problems have received quite some attention in the literature. We refer readers to Ivković and Lloyd (2009) and Berndt et al. (2020), where the fully dynamic bin packing problem is considered. In such a problem, items arrive and depart in an online fashion and repacking of previously packed items is allowed. Several approximation algorithms are presented, as well as a thorough analysis of the computational complexity of the different algorithms. In Gupta et al. (2017), an extension of the fully dynamic bin packing model is presented where repacking is minimized. The model is applied to the problem of data backup to minimize the number of disks used, as well as communication incurred in moving file backups between disks.

Finally, we move to works addressing unknown items dimensions in a dynamic setting: this is the framework where this paper is positioned. The only contribution we are aware of is Rizzo et al. (2019). In the paper, a RM system for air cargo is proposed that combines machine learning predictions with mathematical optimization methods. The model, called AI-CARGO, addresses a well-known issue of the air cargo supply chain, i.e., the general inconsistency between specified (by the shipper at the time of booking) and actual (received by the airline) volumes to be loaded. Machine learning is used to predict the actual volume, while dynamic programming is used to determine if the incoming booking request should be accepted or rejected.

In Table 1 we provide an overview of the aforementioned papers divided by addressed problem scope. To conclude the literature review, we also point out the similarities and differences between our approach and some of the referenced works. Our work differs from Crainic et al. (2016) in that, instead of taking recourse actions and make adjustments to the loading plan once the true information are revealed, we choose to protect from the uncertainty in number of bins and shipment dimensions beforehand, by establishing loading plans where uncertainty is already considered. We solve the Three-Dimensional bin packing model using the EP heuristic proposed by Crainic et al. (2008). However, our version of the heuristic is enhanced since (i) we consider shipment rotations in the algorithm and (ii) we modify their merit function to account for uncertainty in shipment dimensions. The modeling framework and the constraints are inspired by the work of Paquay et al. (2018). Our work differs in that we use probabilistic instead of deterministic constraints for load stability and capacity restrictions. Finally, our overarching goal is comparable with the one of Rizzo et al. (2019), as testified by the fact that the two works fall under the same category in Table 1. A major difference is that we use a three-dimensional and probabilistic approach when assigning items to bins, rather than a one-dimensional and deterministic approach. We believe our approach better reflects the dynamics and uncertainties of the air cargo acceptance/rejection booking process.

Table 1
Summary of research into packing problems divided into static/dynamic approaches and known/unknown items dimensions.

	Static problem	Dynamic problem
Known dimensions	Crainic et al. (2008) Crainic et al. (2011) Paquay et al. (2014) Paquay et al. (2017) Paquay et al. (2018)	Ivković and Lloyd (2009) Berndt et al. (2020) Gupta et al. (2017)
Unknown dimensions	Crainic et al. (2014) Crainic et al. (2016)	Rizzo et al. (2019) **This contribution**

3. Problem statement and methodology

The problem we are addressing can be summarized as follows. Let us consider a specific (OD airport pair, aircraft type, departure date) triplet operated by a combination airline with a passenger aircraft, with unknown available belly space for cargo, a set of shipments and a set of ULDs. Shipments are assumed to be cuboid boxes, each with unknown length, width, height, and known weight. The overall objective is to maximize the number of loaded shipments compatibly with a user-defined confidence level related to the uncertainty of belly space and their true dimensions at the moment of booking.

3.1. Forecasting problem

The forecasting block addresses prediction of (i) aircraft cargo capacity and (ii) shipment dimensions. Predicting cargo capacity can be treated as a time series forecasting problem, since historical capacity values define a one-dimensional sequence (where time is the variable) that can be used to predict future values. On the other hand, predicting shipment dimensions falls under the broader category of feature prediction problems and, more specifically, it can be categorized as a regression problem. As such, both problems can be addressed using a variety of techniques, ranging from statistical models to machine learning models.

In recent years, machine learning methods have become increasingly popular for predictive data analytics models that address both the aforementioned problems. In this work, we focused on Deep Neural Networks (DNNs), i.e., neural networks with more than one hidden layer. According to Le Roux and Bengio (2010), DNNs can provide better approximations to nonlinear functions than the models with a single hidden layer. The quintessential example of DNNs is the Multi-Layer Perceptron (MLP), which is a multi-layer feedforward neural network with more than one hidden layers. The fact that a large amount of high-dimensional booking data is used in this research to predict shipment dimensions, makes MLP an ideal tool for our needs.

For cargo space prediction we used Long Short-Term Memory networks (LSTMs) developed by Hochreiter and Schmidhuber (1997). LSTMs proved to perform exceptionally well for a long-time horizon forecasting problem and to capture the long-term dependencies of non-linear data (Sagheer and Kotb, 2019), as the historical flight data that are used in this work are. The detailed implementation and structure of both forecasting models is provided in Section 4.

However, how much trust can be attributed to a forecast has a deep impact on any decision-making process. Towards this end, a probabilistic approach should be adopted that provides probability distributions, instead of point estimates, as inputs to the packing model, in order to protect the RM analyst from the uncertainty included in the forecast. For example, a shipment whose predicted length is 100 cm, can still turn out to be 90 or 110 cm long: we want to assess the (un)likelihood of these deviations from the predicted value. Delgado et al. (2019) use a scenario tree to represent aircraft capacity uncertainty. On the other hand, our methodology is partially inspired by the work of Zhu and Laptev (2017) since it is based on a similar prediction framework. The goal is to quantify the uncertainty of the model prediction $\hat{y}^* = f^{\hat{W}}(x^*)$, where $f^{\hat{W}}(\cdot)$ is the neural network, $\hat{W}$ are the fitted parameters and x^* is the sample data. We aim to construct a probability distribution $\sim N(\bar{y}^*, \sigma^2)$ where $\bar{y}^*$ is the mean predicted value and σ^2 is the variance. The associated prediction interval is

$$[\bar{y}^* - z_{\alpha/2}\sigma,\ \bar{y}^* + z_{\alpha/2}\sigma] \tag{1}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard Normal distribution. This prediction interval can further assist analysts in the choice of the confidence level that is deemed appropriate for the airline. In addition, note that we assumed that predictions follow Normal distribution. This is a point we will further touch and expand later in the paper, by validating the normality assumption in the data provided by the partner airline.

Bayesian Neural Networks (BNNs), provide us with the required framework to construct the probability distribution of the prediction and the corresponding prediction interval. In BNNs, a prior distribution, such as $W \sim N(0,1)$ is used to define the network weight. Given a set of N inputs $X = \{x_1, \ldots, x_N\}$ and corresponding outputs $Y = \{y_1, \ldots, y_N\}$, the model aims to find the optimal posterior distribution $P(W|X,Y)$, also known as Bayesian inference. In this work, we approximate Bayesian inference using the recently developed Monte Carlo (MC) dropout technique proposed by Gal (2016). Dropout involves dropping a random subset of neurons after each hidden layer with probability p. Hence, the model output can be viewed as a randomly generated sample from the posterior predictive distribution. For a fitted neural network $f^{\hat{W}}(\cdot)$, a specific input x^*, and for M iterations, hidden units of the

neural network are dropped with probability p. A set of $\{\hat{y}_1^*, \dots, \hat{y}_M^*\}$ predictions is obtained, which define the empirical distribution of the model output for the specific input x^*. The predictive mean $\bar{y}^*$ is

$$\bar{y}^* = \frac{1}{M}\sum_{i=1}^{M} \hat{y}_i^* \tag{2}$$

while uncertainty is captured by the model variance. Since this model is designed to steer towards a risk-averse behavior, a more conservative approach is used. Specifically, the uncertainty of the model is not defined solely by the variance $\sigma^2_{\hat{y}^*}$ of the empirical distribution of $\hat{y}^*$. In addition, we incorporate an additional source of uncertainty which accounts for the model inherent noise, namely the residual standard error σ^2_{res} observed in an independent validation set $X = \{x'_1, x'_2, \dots, x'_N\}$, $Y = \{y'_1, y'_2, \dots, y'_N\}$. The residual standard error measures how well every predicted value $\{\hat{y}_1, \dots, \hat{y}_N\}$ approximates the corresponding real value $\{y'_1, \dots, y'_N\}$. Namely, a model might be consistent prediction-wise across different runs without predicting the true value (low $\sigma^2_{\hat{y}^*}$ and high σ^2_{res}), or might be inconsistent prediction-wise across different runs but with an average value that well represents the true value (high $\sigma^2_{\hat{y}^*}$ and low σ^2_{res}). With our approach, we try to ensure that the corresponding prediction interval will contain, for the majority of cases, the real values of the aircraft capacity and shipment dimensions. Hence, the total variance in our models is

$$\sigma^2 = \sigma^2_{\hat{y}^*} + \sigma^2_{res} \tag{3}$$

The final algorithm along with the mathematical formulation is presented in Algorithm 1.

Algorithm 1 Calculation of model's uncertainty and prediction interval

Input Sample data x^*, neural network $f^{\hat{W}}(\cdot)$, dropout probability p, number of observations N, number of iterations M
for $i = 1$ to M **do**
  $\boldsymbol{y_i^*}$ = **MC Dropout**$(f(x^*), p)$
end for
// Predicted mean
$\boldsymbol{\bar{y}^*} = \frac{1}{M}\sum_{i=1}^{M} y_i^*$
// Variance of empirical distribution y^*
$\boldsymbol{\sigma^2_{\hat{y}^*}} = \frac{1}{M}\sum_{i=1}^{M}(\hat{y}_i^* - \bar{y}^*)^2$
// Residual standard error
$\boldsymbol{\sigma^2_{res}} = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{y}_i - y'_i)^2$
// Total variance
$\boldsymbol{\sigma^2} = \sigma^2_{\hat{y}^*} + \sigma^2_{res}$
return $\bar{y}^*, \sigma$
Output Prediction interval

3.2. Packing problem

Our packing model serves a double purpose. First, it determines the optimal number and type of ULDs such that the utilization of the predicted aircraft capacity is maximized. This problem is the classic One-Dimensional Knapsack Problem (1D-KP), according to the typology by Wascher et al. (2007). Secondly, it determines the optimal assignment and loading scheme of shipments to ULDs such that the maximum number of shipments is loaded. This problem is a Three-Dimensional Multiple Heterogeneous Knapsack Problem (3D-MHKP). Following again the typology defined by Wascher et al. (2007), we define our problem a KP since it deals with the maximization of the overall number of items loaded in a set of bins, in contrast to bin packing problems where the goal is the minimization of the number of bins. Notwithstanding an academic literature that is not always consistent in this sense, and our use of the term *packing* or *bin packing* throughout the paper for sake of readability, our problem falls under the KP category.

The majority of the literature addressing KPs in the air cargo context, assumes that aircraft capacity or shipment dimensions are deterministic (Brandt and Nickel, 2018; Paquay et al., 2014, 2017, 2018). While we argue that some airlines might have access to precise information on shipment dimensions, in this paper we focus on the worst case scenario that considers uncertainty in both cargo capacity and shipment dimensions. We pair probabilistic (chance) constraints with a confidence level to ensure that the probability that cargo capacity or a shipment dimension does not exceed a certain threshold is greater or equal to the confidence level. From a RM perspective, this confidence level defines a trade-off between potentially higher load factors ("high-risk high-reward" approach) and ensuring the loading strategy is robust to uncertainty in the input parameters (conservative approach) that each airline can exploit according to its needs.

A generic formulation of a probabilistic KP that is applicable to both the 1D-KP and the 3D-MHKP is: given a knapsack with volume capacity V and set of $N = \{1, \dots, n\}$ items, each of them having profit p_i and a volume distribution with mean $\tilde{v}_i$, the objective is to determine the subset of items to place in the knapsack that maximizes the profit subject to volume capacity restrictions. The corresponding mathematical notation is as follows

$$\max \sum_{i=1}^{n} p_i x_i \tag{4a}$$

s.t.

$$\mathbb{P}(\sum_{i=1}^{n} \tilde{v}_i x_i \leq V) \geq \alpha \tag{4b}$$

where x_i is a binary variable, taking unitary value when item i is loaded in the knapsack. Note that, in case of load factor maximization, we have $p_i = \tilde{v}_i$. Furthermore, in case $\tilde{v}_i \sim N(\bar{v}_i, \sigma^2_{v_i})$, constraint (4b) can be reformulated as follows:

$$\sum_{i=1}^{n} \bar{v}_i x_i + \Phi^{-1}(\alpha) \cdot \sqrt{\sum_{i=1}^{n} \sigma_i^2 x_i} \leq V \tag{5}$$

where $\Phi^{-1}(\cdot)$ is the inverse Cumulative Distribution Function (CDF) of the Normal distribution $N(0, 1)$ It is noted that $x_i^2 = x_i$ since x_i is binary. While a detailed formulation of the chance constrained 1D-KP and 3D-MHKP is part of Section 5, a preliminary formulation is provided in the following.

As stated previously, ULDs come into different shapes and sizes and they are loaded into specific positions in the aircraft belly and blocked with latches. Readers are referred to Brandt and Nickel (2018) for an overview of the most popular ULD types. The floor latches can be rearranged, offering the airlines a range of different ULD configurations corresponding to the specific cargo capacity. Generally speaking, an airline should select the ULD configuration that is expected to accommodate the highest amount of shipments for a specific flight. For example, an airline might decide to use one large ULD instead of two small ULDs to accommodate large shipments that would not fit in the latter.

In this work, we address the procedure as follows. Based on the predicted cargo capacity distribution (using historical data to come up with the prediction), we determine the possible alternative ULD configurations with the probabilistic 1D-KP model. Then, the ULD configuration that accommodates the highest amount of shipments is selected. Towards this end, having as inputs the predicted shipment dimensions distributions, the probabilistic 3D-MHKP is solved for every ULD configuration defined by the 1D-KP model and the configuration that maximizes the number of shipments loaded is selected.

In addition to the basic geometry constraints requiring the shipment to lie entirely within the ULD and not to overlap with other shipments, some additional constraints are considered. More specifically, (i) shipments may be rotated up to six different orientations, (ii) they must be stably supported, and (iii) the weight capacity of the container must be satisfied. For the purposes of this work, all the aforementioned constraints, with the exception of weight capacity, are modified to chance constraints, to account for the uncertainty in the predicted shipment dimensions. We did not impose chance constraints on weight since, according to our dataset, it seemed to be a very reliable piece of information. Moreover, the following assumptions are introduced:

- Only single flight legs are considered, meaning that requirements that dictate the positioning of specific shipments inside the ULD to facilitate the offloading and loading procedures at transshipment airports are not addressed.
- ULD weight and balance restrictions to comply with Center of Gravity (CG) stability of the aircraft are not considered.

In our model, each shipment is characterized by length $l_i \sim N(\bar{l}_i, \sigma^2_{l_i})$, width $w_i \sim N(\bar{w}_i, \sigma^2_{w_i})$ and height $h_i \sim N(\bar{h}_i, \sigma^2_{h_i})$, that are computed by the forecasting block. In addition, each shipment can be placed inside a ULD using up to six possible orientations. This number is lower for shipments containing fragile or non-turnable items, as described in Junqueira et al. (2012) and Paquay et al. (2014, 2017).

Load stability ensures that shipments are loaded in a stable manner, and that they will not fall onto the container floor during loading or transport, resulting in cargo damage or handling personnel injuries. This can be achieved by requiring either full shipment base support, as implemented in Goncalves and Resende (2012) and Paquay et al. (2018), or partial base support, as in Techanitisawad and Tangwiwatwong (2004) and Fuellerer et al. (2010). In the context of this work, load stability will be addressed by enforcing the four vertices of each shipment base to be fully supported by the floor of the container, the top face of other shipments, or a cut in a ULD.

Each ULD j has length L_j, width W_j and height H_j and weight capacity WC_j. The ULD can be represented as a 3D coordinate system, where the origin (0, 0, 0) lies on the back-bottom-left vertex of the ULD. The length, width and height of the ULD lie on the x, y and z-axis respectively. ULDs are parallelepipeds that may have one or several cuts on their corners in order to fit in the aircraft fuselage. For the purposes of this work, the considered ULDs are either Lower Deck Pallets (LDPs) or LD-3 containers, which have a cut in their lower left corner (see Fig. 1). This specific cut can be modeled, in a similar manner as in Paquay et al. (2018), by the linear equation $z = -\frac{b}{a}x + b$, where $a = 40$ cm and $b = 51$ cm according to LD-3 specifications. The considered ULDs are the same ULDs that our partner airline is using. It is noted that more cuts can be added accordingly to account for different ULD types, or this feature can be dropped completely if this model is used for other transportation modes.

The weight of the shipments loaded in a ULD may not exceed the weight capacity WC_j of the specific ULD j. This type of constraint is known as weight capacity constraint, and readers are referred to Dereli and Das (2010) and Paquay et al. (2014, 2017) for indicative examples.

The fact that the Knapsack Problem is NP-Hard (Garey and Johnson, 1979), in combination with the requirement that the model should serve as a decision support tool in a quasi real-time environment, introduces the need for a heuristic. The goal of the heuristic is to calculate a good feasible packing solution within a very short computational time. In fact, our goal is to solve an optimization model anytime a new booking is received, to assess if that booking should be accepted or rejected. As such, the computational time should be consistent with the waiting time a customer is supposed to accept before receiving a decision. We deemed a few minutes (~5 min) as a reasonable computational time. In the context of this work, we use the Extreme Point (EP) heuristic, designed by Crainic et al. (2008). Our variation of the EP heuristic is further explained in Section 5.2. The conceptual design of our forecast and optimization model is summarized in Fig. 2.

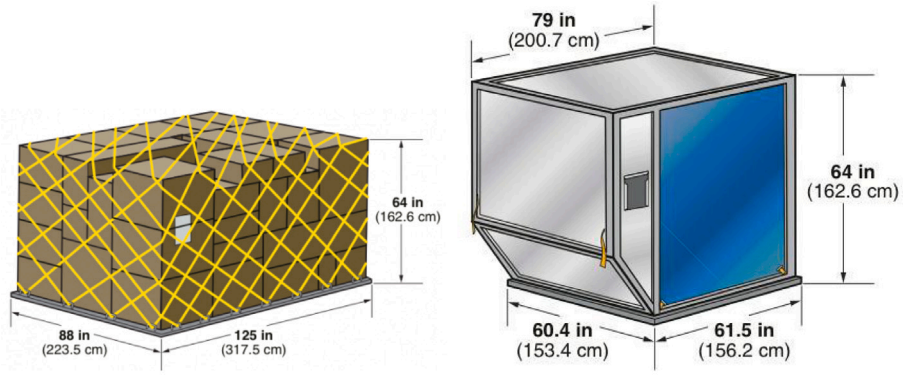


Fig. 1. Considered ULDs.
Source: From www.searates.com.

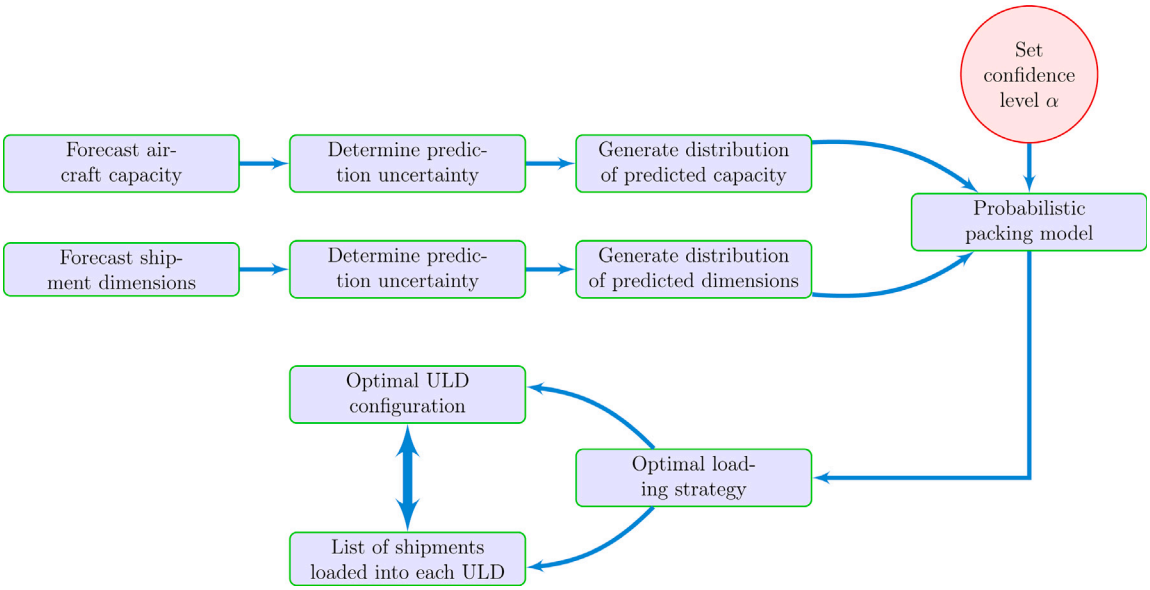


Fig. 2. Conceptual design of our modeling framework.

4. Forecasting models

4.1. Cargo capacity forecasting model

An LSTM network is used to predict the cargo capacity. The model input is based on cargo capacity data for the period 2010–2019, provided by the partner airline. The followed approach is presented in Fig. 3.

We focus on a specific (OD airport pair, flight prefix, flight number, aircraft type) tuple. The reason behind this choice is the large variability across each of these inputs when not considered altogether due, for example, to the different commodity types that are generally flown across different regions. As example, trying to aggregate all input data per aircraft type would conceal the route-specific (and commodity-specific) nature of cargo transport: the same aircraft type might carry a considerable different amount and kind of cargo if operating across two distinct routes.

From each specific data entry we extracted the corresponding month, day of the month, and weekday. Together with the palletized cargo capacity, this is the input for the LSTM network. It is highlighted that the palletized cargo capacity refers to the capacity corresponding solely to the volume of the loaded ULDs and not the total belly capacity of the aircraft. In general, each ULD cannot be placed in any position in the aircraft compartment: there are specific positions equipped with latches and rollers. Using the palletized cargo capacity, instead of the generic belly capacity, allows us to apply the 1D-KP model and determine the

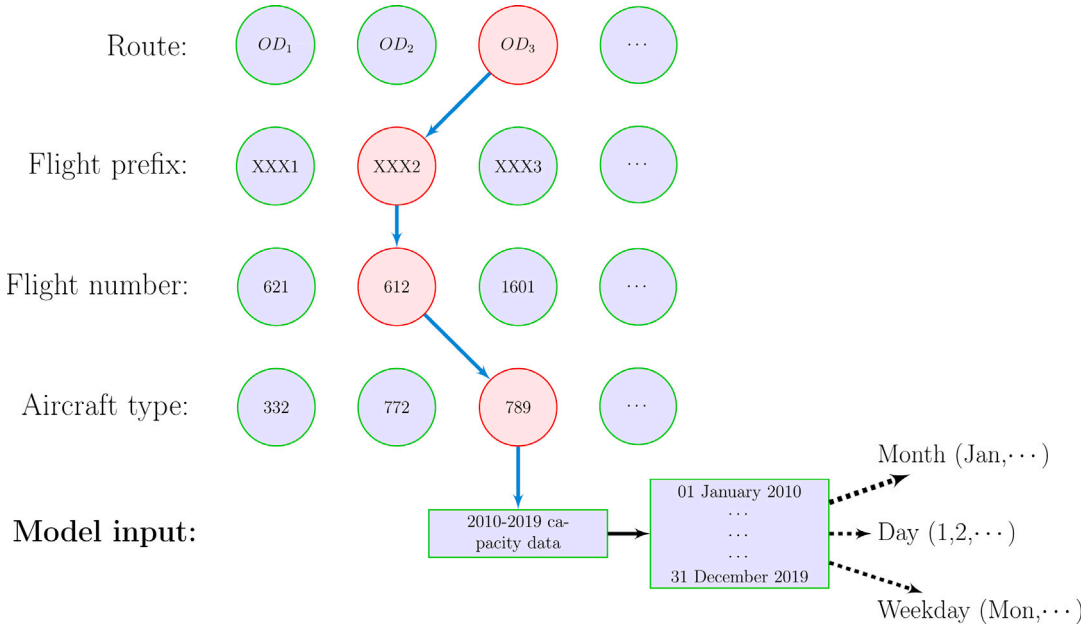


Fig. 3. Input selection approach for the LSTM network.

Table 2
Optimal hyperparameters of LSTM network.

Hyperparameters	Values
Number of hidden layers	3
Number of neurons per layer	50
Optimizer	Adam
Dropout probability	0.2
Number of epochs	50
Batch size	20
Loss function	MAE

optimal ULD configuration(s) directly on the predicted values, without having to compensate for the spacing between the ULDs due to position requirements. The palletized cargo capacity will be referred simply to as cargo capacity hereinafter.

We empirically determined the time-span for each prediction to be 60 days prior to the prediction date. This value allows to account for some seasonality in the time-series while keeping the computational effort reasonable.

4.1.1. LSTM architecture and performance metrics

In the training phase of the LSTM, we used grid search to determine the optimal hyperparameters. The resulting LSTM architecture, as defined by the optimal hyperparameters, is presented in Table 2.

To assess the performance of the LSTM network, we used Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) defined as follows

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

where N is the number of observations, y_i is the true value for the cargo capacity and $\hat{y}_i$ is the predicted value. In addition, since a probability distribution is also computed, we introduce the Prediction Interval Coverage Probability (PICP) metric. PICP identifies the number of occurrences where the prediction interval contains its true value and can be written as

$$PICP_{l(x_i),u(x_i)} = \frac{1}{N}\sum_{i=1}^{N} h_i \qquad \text{where} \quad h_i = \begin{cases} 1 & \text{if } l(x_i) \le y_i \le u(x_i) \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $l(x_i)$ and $u(x_i)$ are respectively the upper and lower bound of the prediction interval for observation i.

Table 3
Validation results of the LSTM forecasting model.

Flight #	Aircraft type	MAPE	MAE $[m^3]$	Training time [min]	Norm. distributed predictions (%)	PICP
Flight #1	A330-300	9.14%	6.23	8	96.8%	99.5%
Flight #2	B777-200	7.59%	5.72	9	95.9%	99.2%
Flight #3	B767-300	14.41%	5.56	11	94.9%	94.1%
Flight #4	A330-200	8.63%	7.82	24	96.5%	91.1%
Flight #5	B747-400	6.62%	6.18	2	98.1%	94.4%
Flight #6	B787-9	9.19%	7.81	4	98.6%	96.3%
Flight #7	B787-9	7.40%	5.50	5	98.7%	95.3%
Flight #8	B777-200	14.18%	9.80	8	96.1%	95.4%
Flight #9	A330-200	11.8%	8.66	8	97.1%	96.9%
Flight #10	B747-400	12.58%	9.88	4	97.2%	95.3%
Flight #11	A330-300	13.26%	8.90	4	96.4%	95.9%
Flight #12	B747-400	14.15%	9.44	3	97.6%	97.6%
Flight #13	B777-200	11.26%	7.32	2	96.7%	90.3%
Flight #14	A330-200	5.58%	4.97	4	90.5%	96.7%
Flight #15	B767-300	10.53%	7.14	2	95.3%	95.7%
Flight #16	B777-200	12.19%	10.01	5	100.0%	95.8%
Flight #17	A330-200	7.18%	6.12	11	93.1%	96.5%
Flight #18	A330-300	11.65%	8.53	10	97.1%	97.5%
Flight #19	A330-200	6.69%	5.44	10	95.3%	97.1%
Flight #20	B767-300	10.24%	10.47	4	100.0%	96.7%

4.1.2. Model validation

Training and testing of the LSTM network was performed on a laptop computer with Intel i5 9300, 16 GB RAM and NVIDIA 1660Ti GPU using the Python 3.7 programming language. The performance of the model was checked against the cargo capacity data of 20 flights and the results are summarized in the left-hand size of Table 3. The exact flight details are omitted for confidentiality reasons. The corresponding dataset is split into training and testing set, with a proportion of 75% and 25% respectively. The last 25% of the training set is withheld for validation.

In order to extract the uncertainty of the predictions, the fitted LSTM network of each flight was run for 1000 iterations using the Monte Carlo (MC) dropout technique, as described in Section 3.1. In each iteration, hidden units were dropped with probability $p = 0.2$. The mean and the total variance of the predicted daily capacity per flight is derived using Algorithm 1, while the corresponding distributions of the predicted values are tested for normality through the Shapiro–Wilk test. The results of the Shapiro–Wilk test for each flight over the probability distributions are presented in the second-to-last column of Table 3. It can be concluded that the normality assumption is verified for the vast majority of the generated distributions. This supports our assumption of normally distributed predictions for cargo capacity and our simplified version of chance constraints.

Using the normality assumption, for each predicted value we construct a 95% prediction interval, in accordance with Eq. (1). Based on the constructed prediction interval, we calculate the PICP for the predicted values per flight, which is presented in the corresponding column in Table 3. The achieved high values of the PICP indicate that the true value of the daily cargo capacity is included, for the majority of cases, in the generated distribution of the predicted cargo capacity.

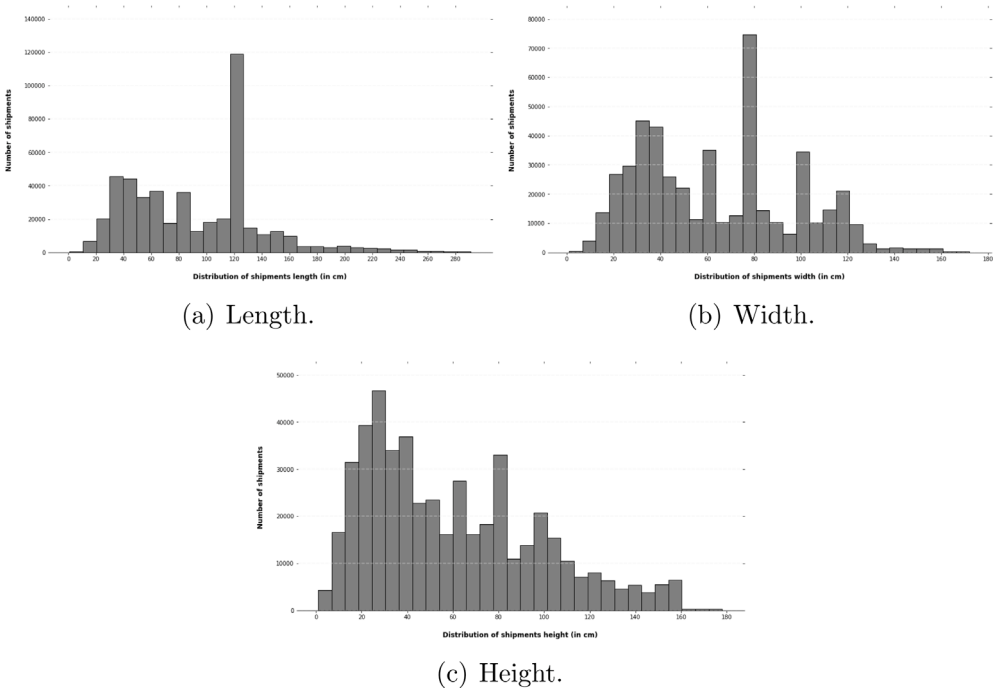
4.2. Shipment dimensions forecasting model

The distribution of the predicted shipment dimensions is computed using an MLP network. The model inputs are based on booking information provided by customers of the partner airline. These booking information is contained in a dataset spanning from January 2019 until December 2019. Each row of the dataset corresponds to a specific shipment and includes relevant information such as the booking origin, the booking destination, the reference number of the customer, the volume of the shipment and the type of the shipment, etc. Note that different shipments can be associated to the same booking (e.g., a group of 50 identical smartphones). While shipments belonging to the same booking can be split across different ULDs, we enforce that a booking can be accepted only if all its shipments can be accepted. For the purposes of this work, and driven partly by the study of Koch (2019), the input nodes of the neural network consist of (i) booking origin, (ii) booking destination, (iii) product code, (iv) commodity code, and (v) shipment volume.

The model produces three outputs, namely the length, width, and height of each shipment. It is noted that the dimensions of the shipments were not always included in the dataset, since it is not a mandatory piece of information to be provided by the customers of our partner airline. Since all three dimensions are needed to test and train the model, only bookings having dimensions information were considered to train and test the model.

4.2.1. Data preparation and pre-processing

In the considered dataset, booking entries that had missing values for at least one of the aforementioned fields where excluded. This resulted in a total of almost 1,150,000 entries. In order for the model to be efficient in terms of computational resources and



(a) Length. (b) Width. (c) Height.

Fig. 4. Distribution of shipment dimensions for the considered dataset.

Table 4
Normally distributed predictions for shipment dimensions [%].

Dimensions	Normally distributed predictions (%)
Length	75.2%
Width	70.1%
Height	69.9%

time, a representative sample of 500,000 shipment entries was chosen, by ensuring that every shipment feature was represented with the same percentage as in the original dataset.

For shipments where rotations were allowed along all three axis, dimensions were re-arranged considering length as the longest dimension, followed by width and height. Differently from other fields, where length or width have an unambiguous definition, we decided to follow the common convention that defines length as the longest dimension of a box. For shipments that could not be rotated in the height dimension, only length and width were re-arranged using the same logic. Additionally, the z-score was used to assess and remove outliers in either three dimensions. The resulting empirical distribution of the length, width and height for the final list of shipments are presented in Fig. 4. These distributions are very similar to the ones derived by Brandt and Nickel (2018), a fact which verifies the representativeness of the chosen sample.

Consistently with the cargo capacity analysis, the dataset is split into training and test set, with a proportion of 75% training and 25% testing. The last 25% of the training set is withheld for validation.

4.2.2. MLP architecture and performance metrics

Similarly to Section 4.1.1, we used grid search to compute the optimal hyperparameters. The resulting MLP architecture along with the optimal hyperparameters are presented in Fig. 5.

4.2.3. Model validation

Training and testing of the neural network was carried using the same laptop described above. Model training took 6 h and 32 min. Fig. 6 presents the model performance in terms of MAE and MAPE per dimension, together with the corresponding boxplots for MAE. For 75% of the predictions, the MAE is below 13.5, 9 and 11 cm for length, width and height respectively. The largest errors occur for length predictions. This is consistent with the fact that length is the largest dimension per shipment in our dataset and is, generally speaking, the one subject to the highest variability.

In order to generate the probability distributions for the predicted values and the associated prediction intervals, we run 1000 iterations of the fitted network using the MC dropout technique. To determine whether the resulting distributions can be approximated with a Normal distribution, we applied once more the Shapiro–Wilk tests. Results are shown in Table 4.

Hyperparameters	Values
Number of hidden layers	2
Number of neurons per layer	600
Optimizer	Adam
Dropout probability	0.2
Number of epochs	100
Batch size	20
Loss function	MAE

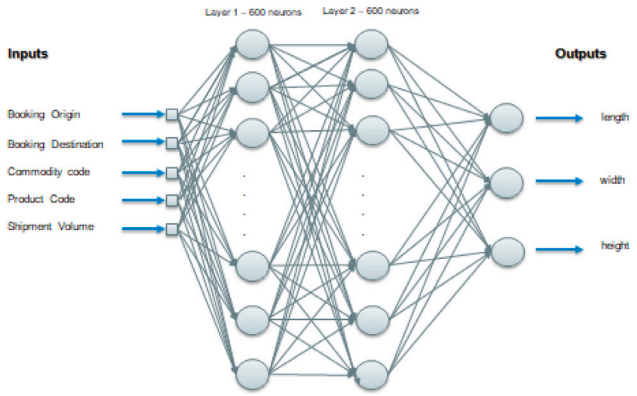


Fig. 5. MLP architecture.

Dimension	MAPE	MAE (cm)
Length	13.9%	12.9
Width	13.1%	6.5
Height	19.8%	8.8

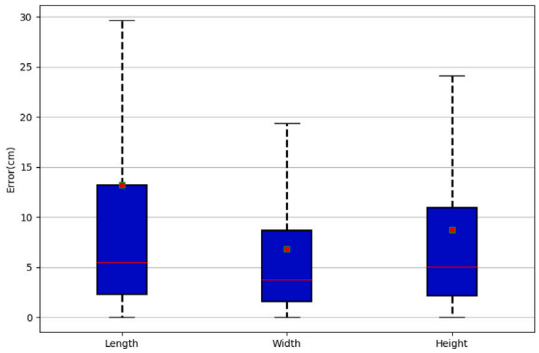


Fig. 6. MLP performance metrics.

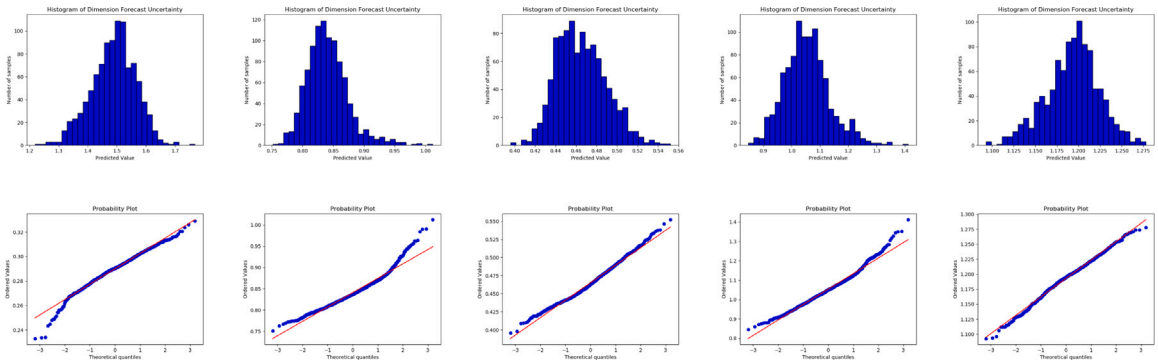


Fig. 7. Histograms and QQ plots for a subset of predicted dimensions distributions that did not pass the Shapiro–Wilk test.

The results demonstrate that the normality assumption is reasonable for the obtained distributions. To further fortify this assumption, QQ plots were created for the distributions that did not pass the Shapiro–Wilk test. Some characteristic examples are provided in Fig. 7, where the first row shows the histogram and the second row the associated QQ plot. As it is evident, no major deviations from the Normal distribution are observed.

For each probability distribution, a 95% prediction interval is established according to the algorithm presented in Section 3. Based on the constructed prediction intervals, we calculate the PICP for each dimension. The results are presented in Table 5. Similarly to cargo capacity predictions, high values of PICP were achieved.

Table 5
PICP for shipment dimensions prediction.

Dimensions	PICP
Length	95.1%
Width	98.6%
Height	95.2%

5. Packing models

5.1. One-Dimensional Knapsack Problem with chance constraints (1D-KP)

Given the predicted aircraft cargo capacity $V_{cc} \sim N(\bar{V_{cc}}, \sigma^2_{V_{cc}})$ and an unlimited set of n LD-3 containers and k LDPs with volume $V_{LD-3} = 4$ m^3 and $V_{LDP} = 10$ m^3 respectively, the probabilistic 1D-KP can be formulated as follows

$$\max \quad nV_{LD-3} + kV_{LDP} \tag{7a}$$

s.t.

$$\mathbb{P}(nV_{LD-3} + kV_{LDP} \leq \tilde{V}_{cc}) \geq \alpha \tag{7b}$$

$$n, k \quad \in \mathbb{Z}^+ \tag{7c}$$

If $\tilde{V_{cc}}$ is Gaussian, as proven in Section 4, Eq. (7b) becomes

$$\bar{V}_{cc} + \boldsymbol{\Phi}^{-1}(1-\alpha)\sigma_{V_{cc}} - nV_{LD-3} - kV_{LDP} \leq 0 \tag{8}$$

The optimal solution of this problem can be achieved with different combinations of decision variables n and k. Each combination of the decision variables corresponds to a different ULD configuration. Our ultimate goal is to maximize the number of shipments loaded. As such we argue that, although a larger available volume is generally advisable, the optimal solution deriving from this 1D problem might not be the most optimal when the actual shipments are considered. To this avail, we consider the optimal and the first sub-optimal solution as candidate ULD configurations for our packing model. The reason behind this choice becomes evident if we consider the following example. Let us assume, for sake of simplicity, that the predicted cargo capacity is $V_{cc} = 12$ m^3 and we have a shipment with dimensions $3 \times 2 \times 1.5$ m. The optimal solution of the 1D-KP corresponds to three LD-3 containers (i.e., $3 \cdot 4 = 12$ m^3). However, this specific shipment exceeds the dimensions of an LD-3 container and hence, it is offloaded. On the other hand, the first sub-optimal solution of the 1D-KP is 10 m^3, which corresponds to one LDP. The LDP can accommodate this shipment instead. It might also be argued that considering all possible ULD configurations should lead to a better final outcome, but this might increase the computational time unnecessarily, since a loading pattern for each configuration needs to be computed as shown in Section 6.2. In addition, we experimentally found out that going beyond the first sub-optimal solution did not lead to any tangible gain in the final objective.

5.2. Three-Dimensional Multiple Heterogeneous Knapsack Problem with chance constraints (3D-MHKP)

Given our need to compute a packing solution in a limited time, we resort to the EP heuristic algorithm developed by Crainic et al. (2008) to solve the 3D-MHKP. The EP algorithm of Crainic et al. (2008) addressed the Single Size Bin Packing Problem for items with known dimensions and that could not be rotated or overlap. The algorithm is further extended to account for the probability distributions of the shipments dimensions as well as additional constraints considered in the 3D-MHKP.

The input for the packing algorithm is, apart from the predetermined ULD configurations, a list of shipments $\mathbf{P} = (p_{1m}, p_{2m}, \dots, p_{nm})$ corresponding to a specific flight. Each shipment p_{im} may be composed of m multiple individual part shipments (this justifies the two-index formulation), that must be delivered together but can be split across different ULDs. Every part of a shipment p_{im} is characterized by length $l_{im} \sim N(\bar{l}_{im}, \sigma^2_{l_i})$, width $w_{im} \sim N(\bar{w}_{im}, \sigma^2_{w_{im}})$, height $h_{im} \sim N(\bar{h}_{im}, \sigma^2_{h_{im}})$ (as determined by the forecasting model), volume V_{im} and weight W_{im}. Unless differently specified, each shipment can be positioned inside a ULD according to one of the six rotations shown in Fig. 8. The allowable rotations are modeled binary parameters l^+_{im}, w^+_{im} and h^+_{im}. If a parameter is unitary, the associated dimension can be aligned with the vertical side of the ULD.

The potential location of a new shipment inside a ULD is selected based on a global list containing all the EPs in all the ULDs. After a shipment is placed with its back-bottom-left vertex on the selected EP (x_i, y_i, z_i), a new set of EPs is generated that correspond to potential positions for future packings, and this set is appended to the global list. The EPs generation process is summarized in Fig. 9. From the three vertices of the shipment highlighted with red dots, projections are taken towards the back-, bottom-, and left-most side of the ULD with respect to the reference frame. The six resulting new EP points are highlighted with green triangles. Note that the other vertices of the shipment, as well as the third potential projection from every red dot, are not considered because they would not be consistent with a packing strategy that develops from the origin of the reference frame. In addition, an EP is projected only in case it can be supported either by an underlying shipment, the ULD base, or the ULD cut as depicted in Fig. 9(b).

When the considered ULD is empty, the shipment is placed in position $(0, 0, 0)$ for LDPs or $(0, a, 0)$ for LD-3 containers. In all other cases, a subset of candidate EPs from the global list are evaluated as potential locations of shipment p_{im}, following an approach

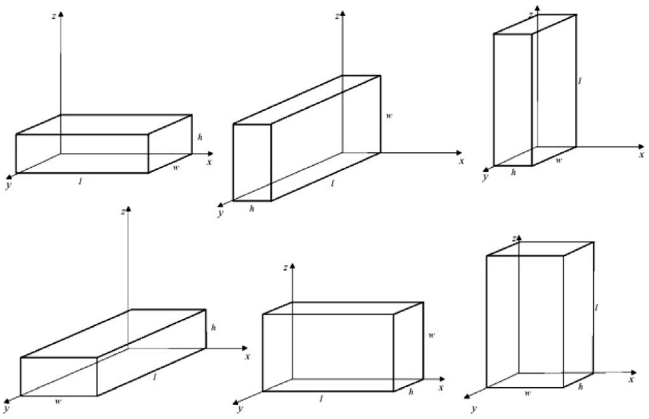
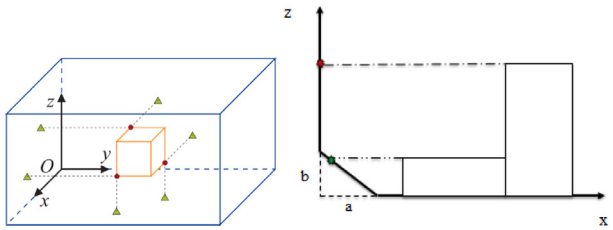


Fig. 8. Possible orientations of a shipment.
Source: From Jozefowska et al. (2018).



(a) EPs generation example. From Baldi et al. (2012)

(b) Valid and non valid EPs in a LD-3 container: only the green projection is usable.

Fig. 9. EPs generation process. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

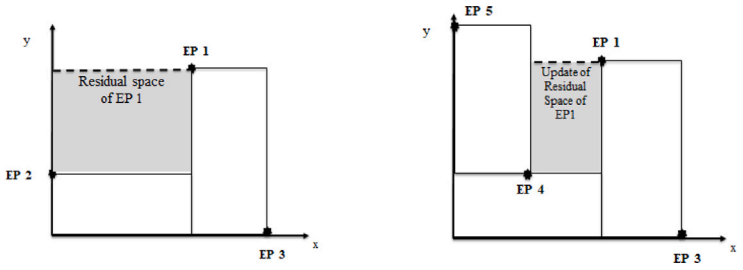


Fig. 10. Example of residual space update.

inspired by Crainic et al. (2008). This approach is based on a merit function that maximizes the utilization of the EP Residual Space (RS). The RS quantifies the free volume around an EP projected along the three dimensions, and consists of three components, RS_x, RS_y, RS_z. Each component captures the distance between the EP and the wall of the container or the nearest shipment along the corresponding axis. Every time a shipment is placed into the container, the RS of all EPs is updated (see Fig. 10).

The MF developed in this work is further extended to take into account the allowable rotations of the shipment as well as the probability distribution of the predicted dimensions. Our MF places the shipment in the EP and with the allowed orientation that minimizes the summation of the differences between the RS component and the corresponding shipment dimension along the relevant axis. In other words, given a specific shipment i, it is chosen the (ULD, EP, orientation) tuple that better occupies the available space around the EP. Our MF is defined as

$$\begin{aligned} MF = &(RS_x - l_{im}r_{i11} - w_{im}r_{i21} - h_{im}r_{i31}) \\ &+ (RS_y - l_{im}r_{i12} - w_{im}r_{i22} - h_{im}r_{i32}) \\ &+ (RS_z - l_{im}r_{i13} - w_{im}r_{i23} - h_{im}r_{i33}) \end{aligned} \tag{9}$$

where r_{iab} is a rotation binary variable introduced to model the six possible orientations of the shipment p_{im}, taking the value of 1 if side a of the shipment is aligned with side b of the ULD (Paquay et al., 2014). Since each side of a shipment should be aligned with a side of the ULD and vice versa, it holds that $\sum_{a=1}^{3} r_{iab} = 1$ and $\sum_{b=1}^{3} r_{iab} = 1$. Moreover, since l_{im}, w_{im}, h_{im} are not deterministic but follow a Normal distribution, the merit function can be modified as follows

$$\begin{aligned} MF = {} & RS_x + RS_y + RS_z \\ & - \sum_{b=1}^{3} \bar{l}_{im} r_{i1b} - \sum_{b=1}^{3} \bar{w}_{im} r_{i2b} - \sum_{b=1}^{3} \bar{h}_{im} r_{i3b} \\ & - \Phi^{-1}(\alpha) \sqrt{\sum_{b=1}^{3} \sigma_{l_{im}}^2 r_{i1b}} - \Phi^{-1}(\alpha) \sqrt{\sum_{b=1}^{3} \sigma_{w_{im}}^2 r_{i2b}} - \Phi^{-1}(\alpha) \sqrt{\sum_{b=1}^{3} \sigma_{h_{im}}^2 r_{i3b}} \end{aligned} \tag{10}$$

where $\Phi^{-1}(\cdot)$ is the inverse CDF of $N(0,1)$ and α is the desired confidence level. Considering the aforementioned constraints on r_{iab}, Eq. (10) can be simplified and re-arranged per RS component as

$$\begin{aligned} MF = {} & \left(RS_x - \left(\bar{l}_{im} + \Phi^{-1}(\alpha)\sigma_{l_{im}}\right) r_{i11} - \left(\bar{w}_{im} + \Phi^{-1}(\alpha)\sigma_{w_{im}}\right) r_{i21} - \left(\bar{h}_{im} + \Phi^{-1}(\alpha)\sigma_{h_{im}}\right) r_{i31} \right) \\ & + \left(RS_y - \left(\bar{l}_{im} + \Phi^{-1}(\alpha)\sigma_{l_{im}}\right) r_{i12} - \left(\bar{w}_{im} + \Phi^{-1}(\alpha)\sigma_{w_{im}}\right) r_{i22} - \left(\bar{h}_{im} + \Phi^{-1}(\alpha)\sigma_{h_{im}}\right) r_{i32} \right) \\ & + \left(RS_z - \left(\bar{l}_{im} + \Phi^{-1}(\alpha)\sigma_{l_{im}}\right) r_{i13} - \left(\bar{w}_{im} + \Phi^{-1}(\alpha)\sigma_{w_{im}}\right) r_{i23} - \left(\bar{h}_{im} + \Phi^{-1}(\alpha)\sigma_{h_{im}}\right) r_{i33} \right) \end{aligned} \tag{11}$$

After the MF for all the EPs is calculated, the EPs with any negative RS component are removed to ensure that the placed shipment does not overlap with any of the previously packed shipments. The remaining EPs are sorted according to increased values of the MF (the smallest the value of MF given the original definition in Eq. (9), the more residual space is used). These EPs are evaluated as potential positions of the shipment p_{im} based on the following constraints

1. the corresponding ULD dimensions must be satisfied, i.e., a shipment must lie within the corresponding ULD boundaries
2. the corresponding ULD weight capacity must be satisfied
3. load stability must be ensured, i.e., all four vertices of the shipment must be supported either by the ULD base, by other shipments, or by a combination of a cut and shipments
4. in case of part shipments, when a part of the shipment is not loaded, then all the parts should be offloaded.

and the EP that minimizes the MF is chosen. To account for uncertainty in the prediction of the dimensions, constraints 1. and 3. are modified to chance constraints, as they incorporate probability distributions. When assessing the feasibility of a shipment insertion into a ULD, two conditions must be met. First, if the intended shipment is loaded on top of another one, stability must be ensured (constraint 3). Fig. 11 depicts an example where the "augmented" (according to confidence level α) horizontal occupancy of shipment 2 is compared to the mean horizontal occupancy of shipment 1. If the first one is no larger than the second one, we consider shipment 2 feasibly supported by shipment 1 along the L-direction. Note that, although not shown in Fig. 11 due to its 2D nature, the same policy applies to the third dimension (W-direction). If the stability condition is met, the algorithm checks if the overall "augmented" dimensions do not exceed the ULD bounds. This procedure is shown in Fig. 12. On the left side, we show how to compute the cumulative probability distribution per dimension. Along the L-direction, only shipment 1 is considered. The contribution of shipment 2 is omitted because the stability constraint forces its horizontal occupancy to be smaller than the one of shipment 1. Along the H-direction, we consider the combined effect of shipment 1 and shipment 2 and compute the combined distribution N_h which, due to the normality assumption, takes the form shown in Fig. 12. On the right side, we show how it is checked if the shipment loading strategy is feasible in terms of ULD bounds. The color of the probability distribution for the H-dimension is different than black to highlight it refers to the summation of the two normal distributions shown on the left side. Again, the W-dimension is not shown for sake of readability, but the same concept applies there.

The pseudo-code describing the algorithm of the proposed packing heuristic applied to a single ULD and a list of shipments with no part shipments ($m = 1$ for all shipments), for sake of simplicity, can be found in Appendix.

6. Computational experiments

6.1. Combined model design

The combined model, where the forecast and the optimization blocks are used in sequence, is presented in this section. To pursue this goal, we adopt the perspective of the cargo RM department of a combination airline. Given a passenger flight scheduled to depart on day t, the planning horizon to book cargo space on that flight is defined as $[t - L_{ph}, t]$, where L_{ph} is the number of days prior to departure (i.e., the start of our planning horizon) when RM starts accepting bookings for such flight (for our partner airline, $L_{ph} = 14$). Note that we consider day t at the upper bound of the planning horizon. Although this might not be common practice, some last-minute bookings can still be accepted on the day of departure. For every new booking request that is received, the following steps must be carried out in sequence: (i) if not provided, compute dimensions and uncertainty distributions, (ii) add the booking to the list of accepted bookings and compute the optimal loading strategy, and (iii) accept the incoming booking request if a set of conditions is met.

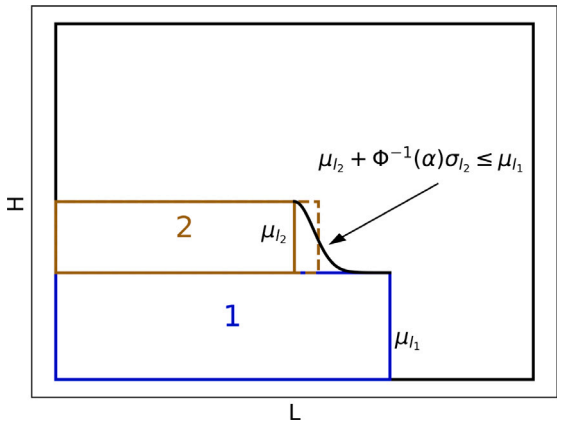


Fig. 11. Procedure to determine if a shipment can be feasibly supported by another one accounting for uncertainty in the dimensions.

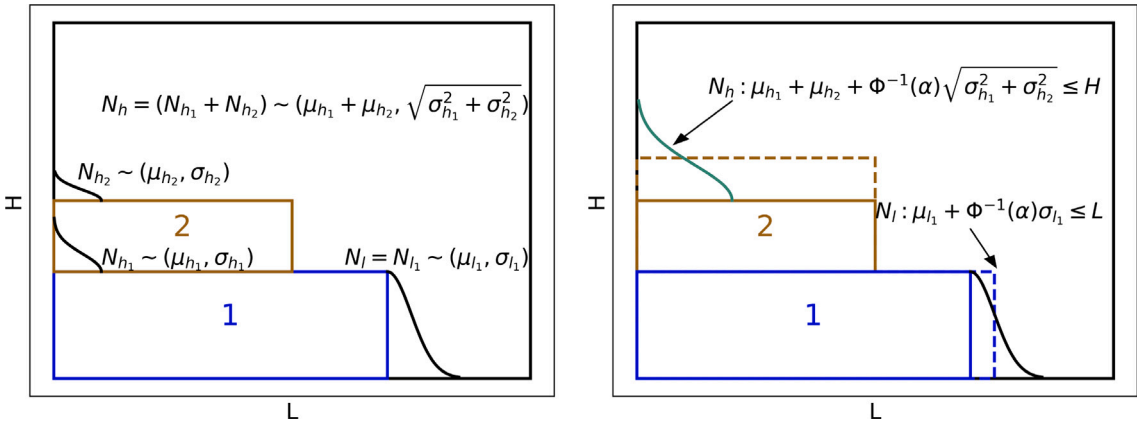


Fig. 12. Procedure to determine if a shipment loading configuration is feasible with respect to the ULD bounds. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This decision-making process is more straight-forward at the beginning of the planning horizon, where supply (available cargo space in ULDs) generally outmatches demand (already accepted bookings, plus incoming new booking). As the number of accepted bookings increases, the process becomes more challenging both from a practical perspective (decrease in available supply, increase in demand), and from a computational perspective, since the increase in bookings to consider increases the solution space of the optimization problem. The decision-making process that encompasses both the forecast and the packing model is shown in Fig. 13.

Initially, we generate the distribution of the predicted cargo capacity for the specific flight, $\tilde{V}_{cc} \sim (\bar{V}_{cc}, \sigma^2_{V_{cc}})$ with the LSTM network developed in Section 4.1. Then, the 1D-KP described in Section 5.1 is run, within a specific confidence level, in order to determine the available optimal ULD configuration(s). Every time a booking request with unspecified dimensions is received, the features of the booking described in Section 4.2 are passed to the MLP network, and the distribution of the predicted shipment dimensions $l_{im} \sim N(\bar{l}_{im}, \sigma^2_{l_i})$, $w_{im} \sim N(\bar{w}_{im}, \sigma^2_{w_{im}})$, $h_{im} \sim N(\bar{h}_{im}, \sigma^2_{h_{im}})$ is generated.

Afterwards, the 3D-MHKP model is solved for all the identified ULD configurations. The block diagram describing more into details the 3D-MHKP can be found in Fig. 14. For every candidate ULD configuration, the 3D-MHKP is solved by using the **First-Fit Decreasing** (FFD) heuristic, which proved to be the most effective among a set of other sorting heuristics. The candidate shipment, along with the already accepted shipments, are sorted by decreasing values of their volume and fed to the 3D-MHKP model developed in Section 5.2. In general, placing large shipments is difficult, especially when this happens at the later stages of the loading procedure. We address this issue by considering them first in the loading block. Then, the EP heuristic is applied to the first ULD until no more shipments can be accommodated and the offloaded shipments (if any) are considered for placement in the next ULD. The same procedure continues until there are no offloaded shipments or ULDs left. Finally, the ULD configuration that maximizes the number of shipments loaded is selected. As it concern ULDs, they were sorted selecting LD-3s first and LPDs afterwards. This choice stemmed from the preference of the partner airline for LD-3s, if possible, when transporting cargo in the belly space. Finally, it is noted that all the computational experiments are implemented in Python 3.7 using the same laptop described above.

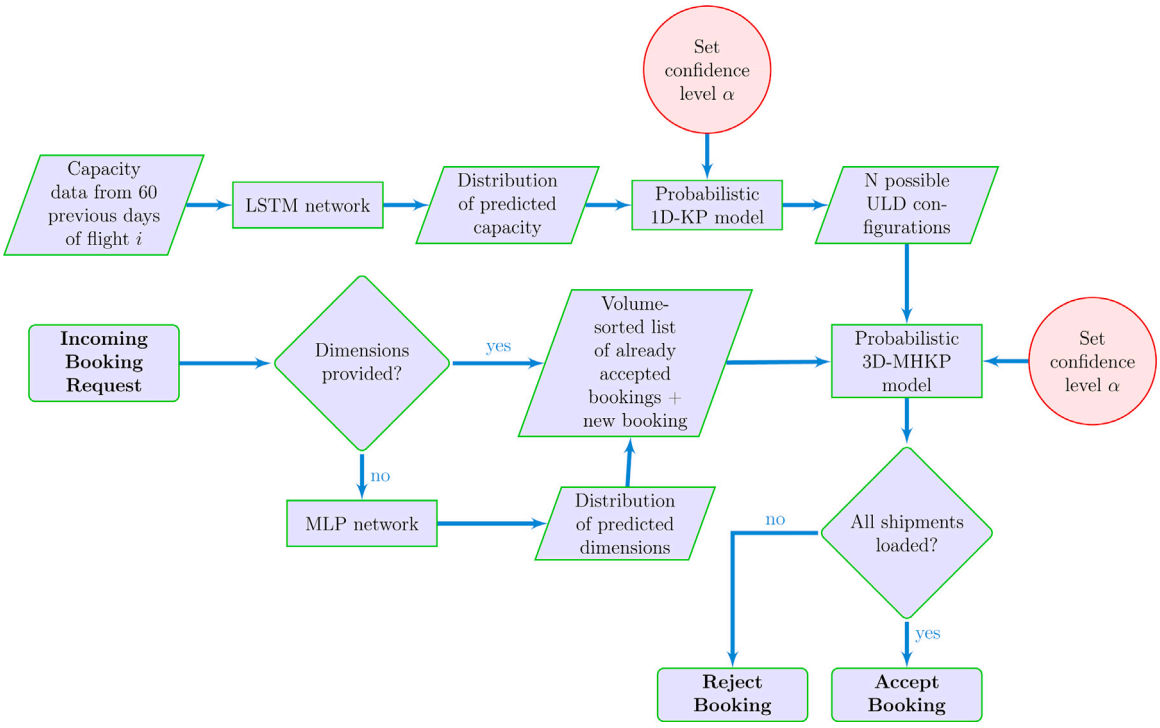


Fig. 13. Flowchart of the decision-making process defining the booking acceptance/rejection strategy.

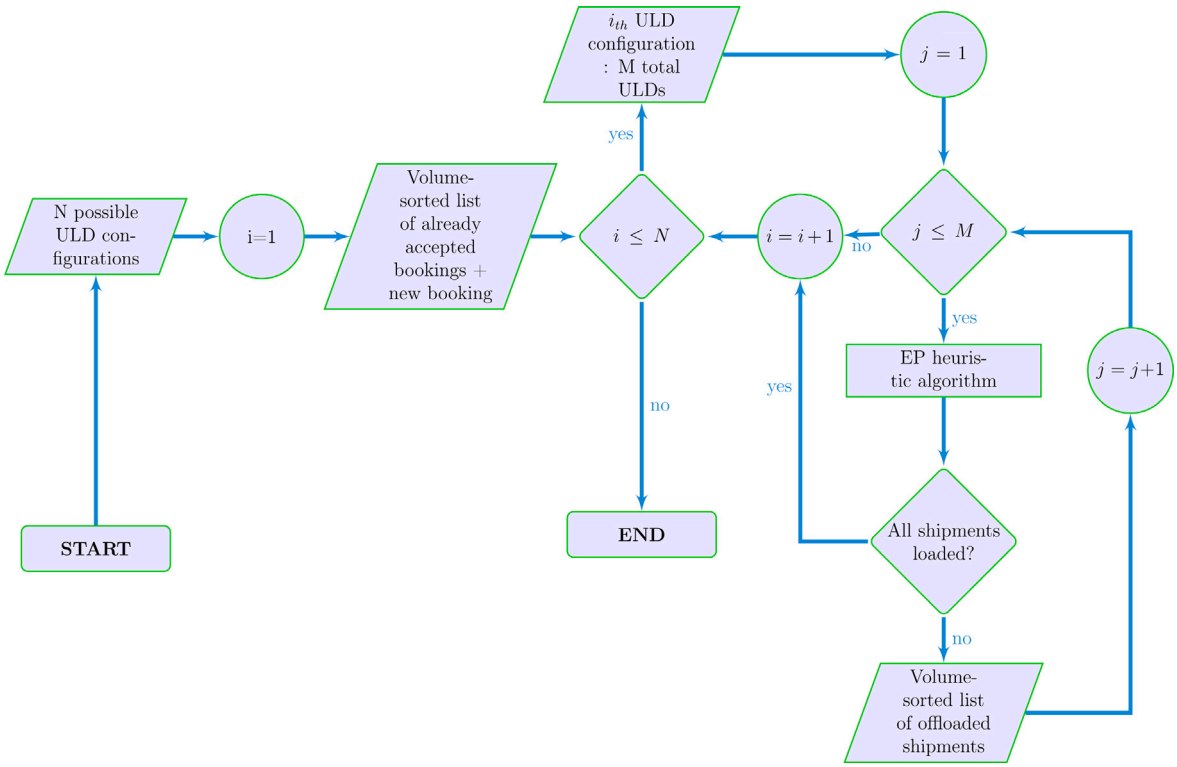


Fig. 14. 3D-MHKP model.

Table 6
Flight and booking information.

Flight #	Cargo capacity [m^3]	ULD configuration	Total shipments	Volume of booking requests [m^3]
Flight #1	74	7 LDPs - 1 LD-3	97	71.23
Flight #2	77	6 LDPs - 3 LD-3	91	72.65
Flight #3	96	8 LDPs - 4 LD-3	122	89.49
Flight #4	78	7 LDPs - 2 LD-3	194	70.48
Flight #5	76	6 LDPs - 4 LD-3	78	83.92

6.2. Case studies

In this section, our model will be applied to five real booking sequences referring to five flights performed with passenger aircraft, with data provided by the partner airline. This is a major European combination airline with two cargo hubs, a fleet of roughly 450 passenger aircraft and 6 full freighters, and that serves approximately 300 long haul destination across 110 countries. In one year, the airline generally handles 1.1 million tonnes of cargo. In our context, a booking may consist of multiple individual shipments with the same or different size. Each set of booking requests is sorted by chronological order in order to simulate the RM rationale of accepting/rejecting an incoming booking request. For the five flights, all bookings were received with specified dimensions which were determined to be accurate by operations (i.e., there was an excellent match between dimensions specified by shippers and real dimensions). As such, in this section and in Section 6.2.3 we will show results that consider shipment dimensions as known. On the other hand, in Sections 6.2.4 and 6.2.5 we will consider the case where dimensions are not given, to provide the full spectrum of options a RM analyst can face.

In our model, a booking is accepted when the computed loading strategy (ULD configuration + shipments loaded) does not offload any previously accepted bookings. This is consistent with our RM perspective that focuses on new bookings as they appear in the system, and considers the already accepted bookings not to be offloaded to avoid penalty fees (although they can be shuffled between ULDs after every new iteration). Moreover, in case of bookings consisting of multiple shipments, as explained in Section 5.2, if a part of a booking is offloaded, then the whole booking is rejected. If the booking is rejected, the last optimal loading strategy is stored, until a new booking request is received. Using this ordered sequence of bookings, our developed model is run every time a new booking request is received and, in case the aforementioned conditions are met, the booking is accepted. A good indicator of the solution quality of our model, apart from the volume of the shipments loaded, is the Acceptance Factor (AF). The AF is defined as the ratio between the volume of accepted requests and the volume of all incoming booking requests. As example, in case we have booking requests with total volume 10 m^3 and 9 m^3 are accepted, AF = 90%. Note that, while a high AF can be intuitively associated with high efficiency, a low AF does not necessarily mean the opposite. For example, if many more bookings than the aircraft capacity are received, the load factor might be close to 100%, but the AF will be low.

Since booking data are extremely sensitive, any indication of the considered flight or the original shipper is omitted for confidentiality reasons from our results. The corresponding flight and booking characteristics are summarized in Table 6, while in Fig. 15, a visual representation of the volume distribution of the bookings per flight is depicted. On the vertical axis, numbers refer to the booking number.

Given that for the set of five flights all booking dimensions were provided, we initially tested our EP packing heuristic without the forecast block. In particular, we compared it against the decision-making policy of the partner airline and different heuristics based on the EP paradigm.

6.2.1. Comparison with the airline decision-making policy

The decision-making policy of the airline addresses unknown dimensions by, given a booking volume without specified dimensions, translating it into a set of $40 \times 40 \times 40$ cm cubes until the value closest to the original volume is reached. The cubes are assembled to form a rectangular cuboid. Since multiple cube arrangements that result in the same volume can be formed, the chosen arrangement depends on some other specifications of the booking (e.g., origin, weight, etc.). The resulting cuboid is then placed in the first available ULD and in the first available location (generally using a bottom-back-left assignment policy). In addition, once a ULD reaches a pre-defined load factor, the ULD is considered complete and cannot be modified for the remaining portion of the booking window. The comparison between the two packing strategies is summarized in Table 7. Row **Airline** indicates the decision taken by the airline, whereas row **Model** depicts the decision output of the packing model. Used ULD configuration corresponds to the actual ULD configuration used by the airline, whereas the optimal ULD configuration is the ULD configuration determined by the packing model.

Our proposed packing model provides better results in terms of loaded volume and AF. The optimal ULD configuration determined by the model is the same as the ULD configuration used by the airline for flights #1, #4 and #5, while it changes for flights #2 and #3, where our model prefers the option of using one extra LDP instead of three LD-3 containers that the airline used. This choice becomes evident if we examine the bookings that are additionally rejected by our model when using the ULD configuration of the airline, namely booking 018 for flight #2 and booking 028 for flight #3. Due to the shipment large dimensions (booking 018) or the high number of part shipments having relatively big dimensions (booking 028), a configuration with one extra LDP replacing three LD-3 containers, even though this corresponds to less total cargo capacity (74 m^3 vs. 76 m^3 for flight #2 and

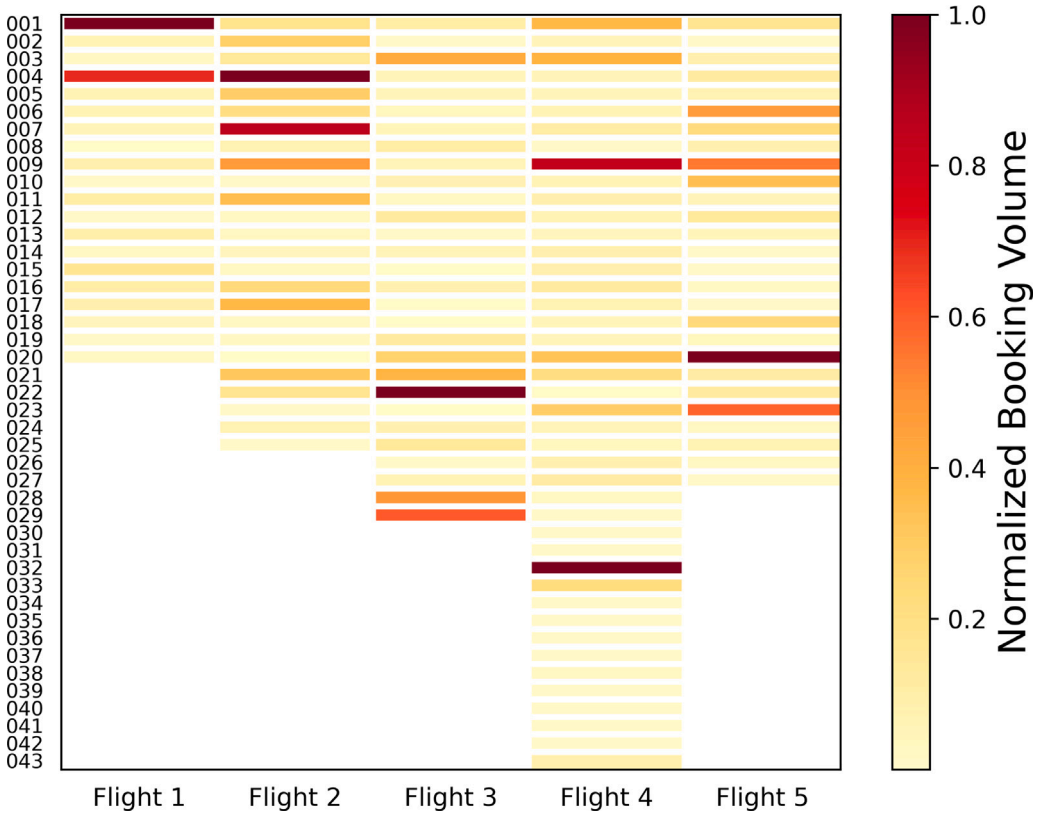


Fig. 15. Normalized booking volume distribution. A unitary value corresponds to the booking with the greatest volume for the specific flight (i.e., columns are independent color-wise). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

94 m^3 vs. 96 m^3 for flight #3), proves to be more suitable for accommodating those shipments. This observed result further justifies our design choice from Section 5.1, i.e., to take into account both the optimal and the first sub-optimal solution of the 1D-KP model when solving the 3D-MHKP. It is also interesting to observe a common pattern in the decision-making process of the airline. While acceptance/rejection decisions are relatively similar for the first 60/70% of bookings, the airline rejects a much larger amount of late bookings received. This is due to the aforementioned packing heuristic that does not allow reshuffling between ULDs. Hence, the empty space in a ULD flagged as complete is lost. On the other hand, our EP heuristic allows an even complete reshuffling of shipments anytime a new booking is received, hence allowing for more flexibility in the packing process.

The efficiency of the model in terms of computational time is also highlighted, being this a necessary requirement for RM. It is noted that the upper bound of the computational time corresponds to the time needed for the model to solve the complete set of bookings (e.g., it corresponds to the packing solution for 20 bookings for flight #1, 25 bookings for flight #2, etc.). For flights #1, #2, #3 and #5 the computational time remains below 7 s whereas for flight #4 it increases to 89 s. This sharp increase is attributed to the fact that a total of 194 individual shipments are considered for flight #4, in comparison with 97, 91, 122 and 78 individual shipments for flights #1, #2, #3 and #5 respectively.

An indicative visualization of the loading strategy implemented for flight #1 with the complete set of accepted bookings is presented in Fig. 16. In the first plot, the cut of the LD-3 container is highlighted in black.

It is highlighted that in the examined sets of booking requests, the rejection decisions in our model are associated solely with lack of capacity and not other reasons such as delayed delivery or qualitative rejection. Moreover, when interpreting the results, we have to remind that no weight and balance restrictions are considered and that the model is designed for a single flight. Consequently, any lack of capacity caused in the actual flights by requirements of specific shipments being in the same ULD to facilitate the offloading/loading process in multiple flight legs or by weight and balance restrictions, is not addressed in the current work. The developed packing model utilizes the residual space subject only to the constraints referred in Section 5.2.

6.2.2. Heuristic validation

To solve the bin packing problem at hand, a different heuristic paradigm such as adapted variants of tabu search or genetic algorithm could have been a viable choice. Notwithstanding, we decided to use our variant of the EP heuristic since it has been recognized as a state-of-the-art algorithm in air cargo bin packing models. As such, we provide some insights into validation of our strategy (using filling rates as KPI) when compared to a different variant of the EP heuristic. In fact, similarly to our work, Paquay

Table 7
Loading decision and performance metrics for considered flights.

Booking	Loading decision									
	Flight #1		Flight #2		Flight #3		Flight #4		Flight #5	
	Airline	Model	Airline	Model	Airline	Model	Airline	Model	Airline	Model
001	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
002	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
003	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
004	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
005	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
006	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
007	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
008	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
009	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
010	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
011	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
012	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
013	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
014	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
015	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
016	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
017	✗	✓	✗	✓	✓	✓	✗	✓	✓	✓
018	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
019	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
020	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗
021			✗	✓	✓	✓	✓	✓	✓	✓
022			✗	✓	✓	✓	✓	✓	✓	✓
023			✓	✓	✓	✓	✓	✓	✓	✓
024			✗	✗	✗	✓	✓	✓	✗	✗
025			✓	✓	✗	✓	✗	✓	✗	✓
026					✓	✓	✓	✓	✓	✓
027					✗	✓	✓	✓	✗	✓
028					✗	✓	✗	✓		
029					✗	✗	✓	✓		
030							✓	✓		
031							✗	✓		
032							✗	✓		
033							✗	✓		
034							✓	✓		
035							✓	✓		
036							✓	✓		
037							✓	✓		
038							✓	✓		
039							✓	✓		
040							✓	✓		
041							✓	✓		
042							✓	✓		
043							✗	✗		
Performance metrics										
Comp. time [s]	N/A	≤ 5	N/A	≤ 5	N/A	≤ 7	N/A	≤ 89	N/A	≤ 5
ULD Conf. (LDPs - LD-3s)	7-1	7-1	6-4	7-1	8-4	9-1	7-2	7-2	6-4	6-4
Volume $[m^3]$	62.22	68.76	60.86	71.73	61.53	76.53	60.58	69.32	45.46	49.28
AF	83.2%	96.51%	83.7%	98.7%	68.7%	85.5%	85.9%	98.3%	54.10%	58.72%

et al. (2018) have also developed a fast two-phase constructive heuristic based on such approach. However, they additionally account for fragility of the shipments and weight distribution inside the ULDs. Their approach is verified on instances of up to 100 items from a box dataset stemming from a real world case, through the evaluation of the ULDs filling rate. Specifically, they achieve for half the ULDs a filling rate of at least 50%. The computational time is less than 12 s. Since the shipments considered for this case study share approximately the same features with the boxes from the dataset used by Paquay et al. (2014), we compare the performance of our model in terms of ULD filling rates. The boxplots representing the filling rate for the ULDs used in each flight are presented in Fig. 17.

In the considered flights, half of the ULDs have a filling rate of at least 62%, whereas for four out of five flights this percentage increases to 75%. Moreover, a filling rate as high as 98% is achieved for a specific ULD in Flight #2. Overall, our model provides better results and, with the exception of flight #4, in a shorter computational time. However, we acknowledge that accounting for fragility and weight distribution might reduce the performance of our heuristic.

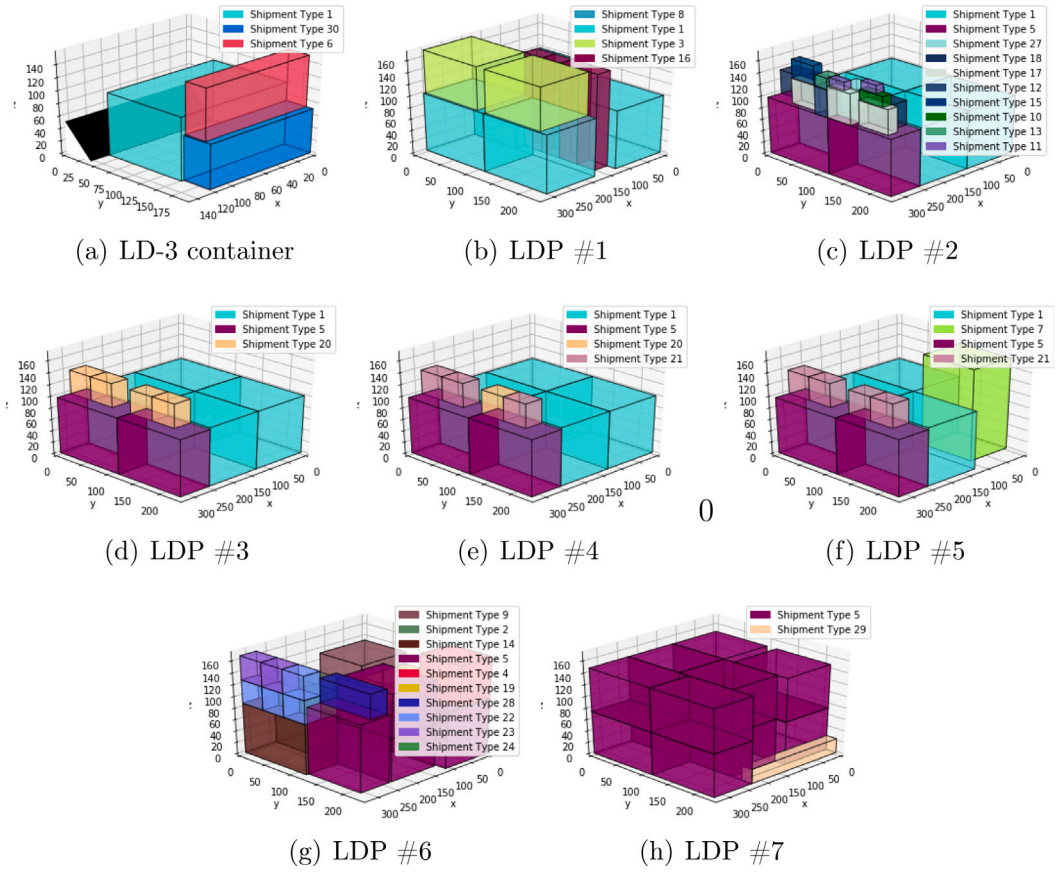


Fig. 16. Visualization of loading strategy (flight #1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

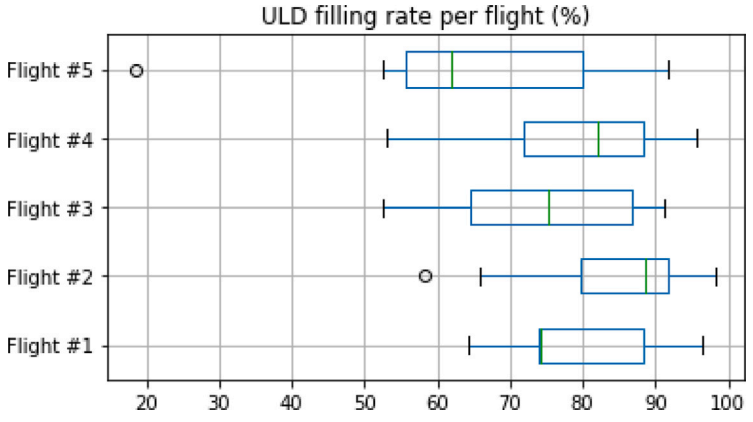


Fig. 17. Filling rates per flight [%].

6.2.3. Sensitivity analysis for cargo capacity forecast

In this section, we analyze the trade-off between the probability α of not violating the predicted aircraft capacity and the overall volume of the shipments loaded. Note that $\alpha = 1$ corresponds to the worst-case scenario that covers all possible realizations of the predicted aircraft capacity V_{CC}.

First, we generated the probability distribution of the predicted cargo capacity as explained in Section 4.1. The predicted capacity distributions were:

- Flight #1 : $V_{cc} \sim N(80.23, 7.4)$ m^3

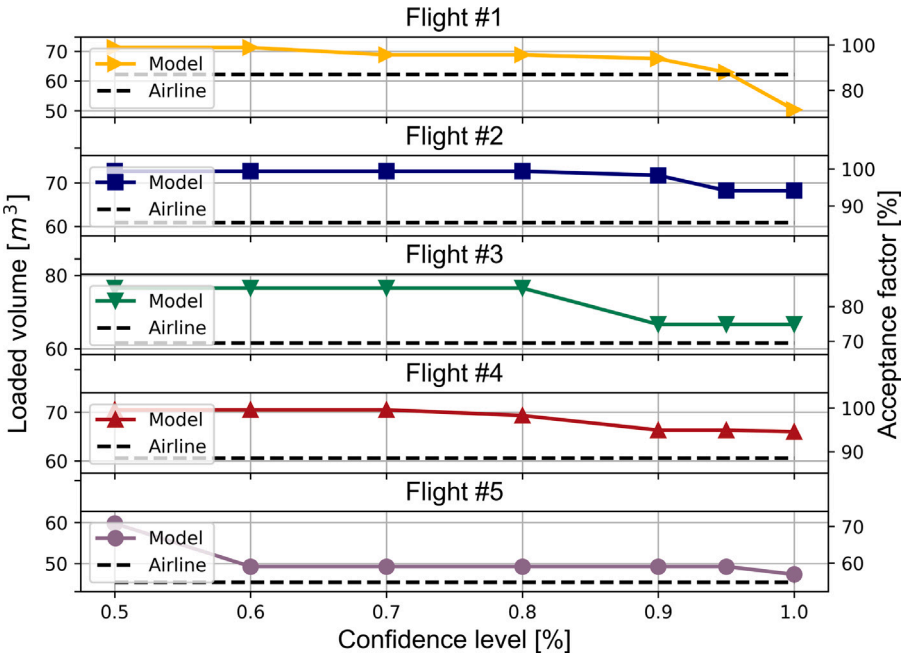


Fig. 18. Effect of confidence level in loaded volume and AF when varying α for overall cargo capacity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- Flight #2 : $V_{cc} \sim N(85.01, 7.1)$ m^3
- Flight #3 : $V_{cc} \sim N(100.51, 6.3)$ m^3
- Flight #4 : $V_{cc} \sim N(84.32, 5.2)$ m^3
- Flight #5 : $V_{cc} \sim N(79.30, 5.5)$ m^3

Then, the probabilistic 1D packing model was run for different confidence levels using as inputs the booking dimensions (still deterministic) as provided by our partner airline. The results in terms of loaded volume and the corresponding AF for each flight are presented in Fig. 18.

It is observed that, for all five flights, choosing a higher confidence level for the capacity forecast results in lower volume of shipments loaded. This is a direct consequence and validation of the intended functionality of the combined model, which is to protect the RM department from any uncertainty included in the forecast. This being said, with the exception of flight #1 for confidence levels $\alpha = 0.95$ and $\alpha = 0.99$, our combined model results in a higher volume of shipments loaded in comparison to our partner airline. This is due to the fact that our forecasting model, even for the most conservative approach (confidence level $\alpha = 0.99$) outputs an aircraft capacity that is still higher than the volume of the shipments loaded by the airline. This is in combination with the efficiency of the model that, even when the available cargo space is reduced, manages to find feasible loading patterns. Additionally, the sharp decrease that is observed in flight #1 at confidence level $\alpha = 0.99$ is due to rejected booking 004 (Fig. 15) that is a reasonably bulky shipment. As such, the rejection of this single booking is causing the overall loaded volume to sharply decrease.

6.2.4. *Sensitivity analysis for dimensions forecast*

We also performed a sensitivity analysis by changing the confidence level for shipment dimensions prediction. Hence, from now on we do not consider dimensions as known (otherwise the confidence level would play no role). For this scenario, to have a benchmark reference, we considered the aircraft capacity, which can be found in Table 6, as known. However, the optimal ULD configurations are still determined by the packing model. On the other hand, we assume that customers did not provide any shipment dimensions, which is corresponds to the worst-case real life scenario, as usually at least some customers do provide shipment dimensions when forwarding a request. Hence, we used the MLP model developed in Section 4.2 to generate the distribution of the predicted shipment dimensions. The results of the forecasting model for the five flights together with the corresponding prediction intervals are summarized in Fig. 19. It can be noted that the prediction is very accurate for shipments with smaller dimensions, mostly due to their higher frequency in our dataset. Prediction accuracy decreases for larger shipments and mostly in the length dimension.

The probabilistic packing model is run for different confidence levels. The results for every flight in terms of loaded volume along with the associated AF can be found in Fig. 20.

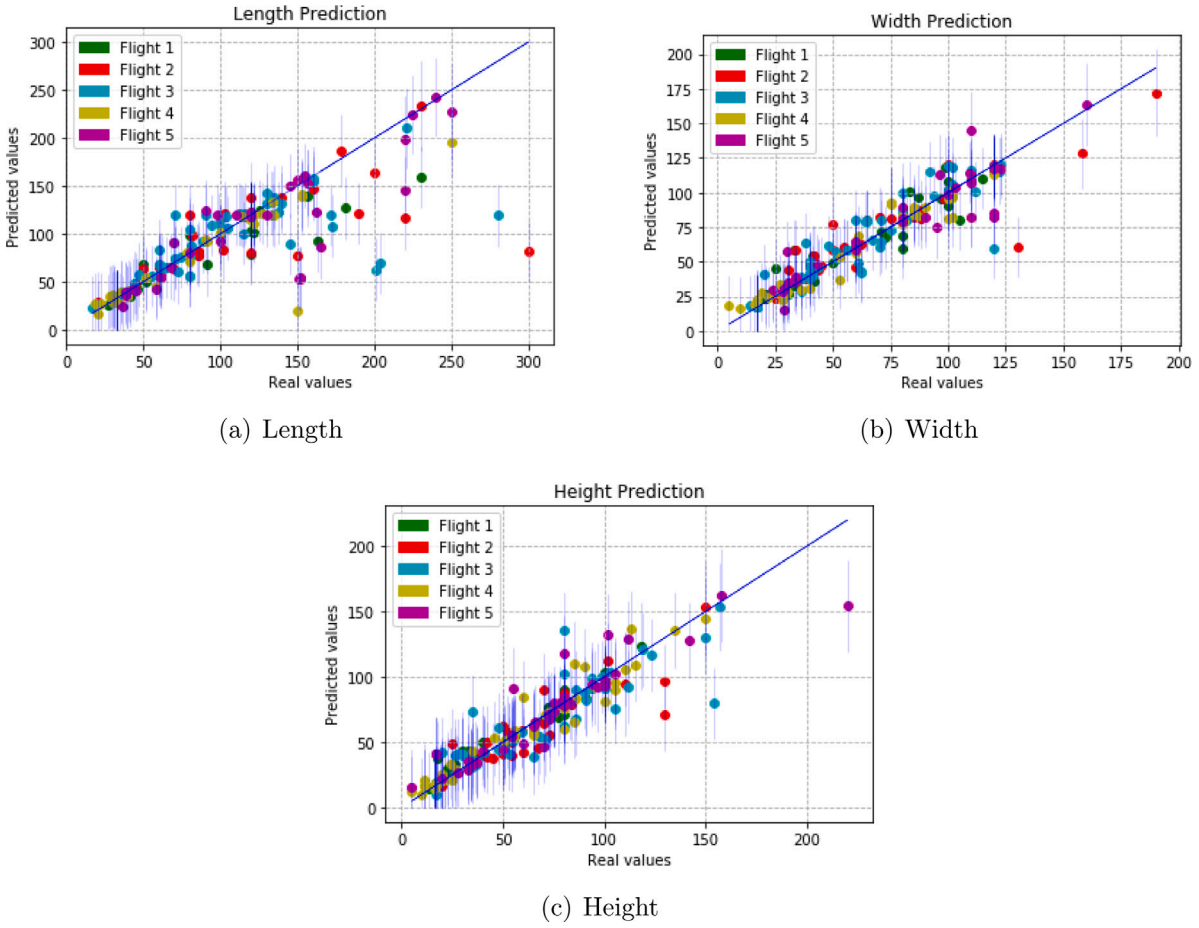


Fig. 19. Results of shipment dimensions forecasting model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

As expected, when increasing the confidence level, the volume of loaded shipments decreases. Differently than the previous case, the decrease is more evident and sharp. In comparison with our partner airline, our model performs better in terms of loaded volume only for confidence level $\alpha = 0.5$ for all flights, as well as for confidence level $\alpha = 0.6$ and 0.7 for flight #3 and #5. This can be partly explained by the conservative approach we adopted when estimating the total standard deviation of our predictions in exchange for a higher PICP (Section 3.1), yielding, in general, high standard deviation estimates for length, width and height. However this contributes towards the risk-averse behavior of the model, as will be proven in the next section. Moreover, applying the same confidence level for all shipments results in being needlessly conservative for shipments where we achieve high prediction accuracy, like shipments with smaller dimensions. In a real-life scenario, shipments of different origins and types might be clustered according to the reliability of historical data provided by the shippers and the accuracy of the predictions using this data. Hence, rather than applying the same confidence level for all shipments, cluster-specific confidence levels should be defined. However, this clustering analysis is outside the scope of this paper and might need and even larger training dataset. Hence, it is left as part of our future work.

6.2.5. Comparison between a risk-prone and a risk-averse behavior of the combined model

One of the novelties of our model resides in combining the perspective of RM (booking phase) and operations (loading phase). As such, we wanted to compare how a risk-prone and a risk-averse approach by RM would impact operations. It should also be noted that, if all bookings were known in a deterministic fashion, no difference between the approaches would arise, as shown in Section 6.2. When booking dimensions are not known, and must be predicted, a risk-prone approach by RM implies a lower confidence level, increasing the accepted volume but, simultaneously, the chances of overbooking and offloading due to under-predicted dimensions. While maximizing loaded volume is of paramount importance for every cargo operator, so is enabling a seamless supply chain where booked shipments fly as scheduled. In fact, consistently offloading shipments will result in a snowball effect that creates bottlenecks and missed due dates, which may generate large re-booking and penalty costs (Liu et al., 2019).

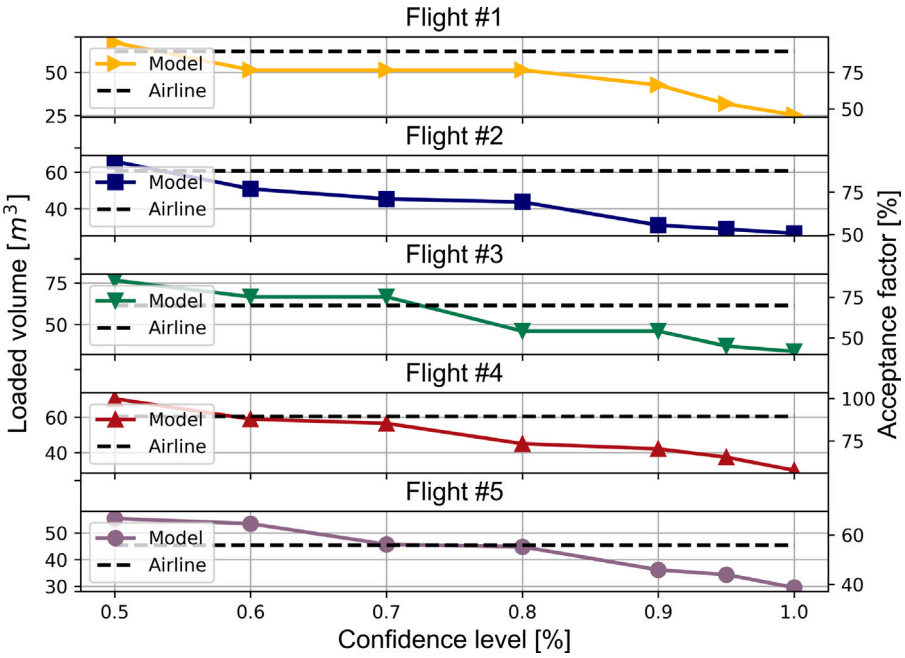


Fig. 20. Effect of confidence level in loaded volume and AF when varying α for shipment dimensions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Hence, in this section we are mostly interested in the number of offloadings and the discrepancy between the accepted (RM) and loaded (operations) volume associated to a risk-prone and risk-averse strategy.

To achieve this goal, we assumed that customers did not provide any shipment dimension, which corresponds to the worst-case scenario for RM, similarly to Section 6.2.4. Then, for the risk-prone approach we used $\alpha = 0.5$ when deciding whether to accept or not a booking (this corresponds to considering mean values for a Normal distribution), whereas in the risk-averse approach we used $\alpha = 0.6$. It is noted that we did not perform a similar analysis for cargo capacity since, as shown in Sections 6.2.3 and 6.2.4, the model is less sensitive to variations in the predicted cargo capacity.

To model the loading process, where shipments might appear with dimensions that do not match the predicted ones, we simulated a total of 1000 different scenarios. For each scenario, the dimension values of the shipments that were accepted by RM were randomly drawn from the corresponding distributions defined by the forecasting model and passed as inputs to the packing model. The available cargo capacity corresponds to the real cargo capacity of the flight instead. In principle, each scenario corresponds to a realization of the loading process.

We performed this analysis for each of the five flights that were already described, and show results in terms of offloadings and discrepancy between accepted and loaded volume in Figs. 21 and 22, respectively. We also want to point out that computational performances were slightly worse when compared to the fully deterministic cases described before. Always using the worst-case scenario (end of the planning horizon with the highest number of bookings considered) as reference, while for flights #1, #2 and #5 we remained below our 5 min threshold, for flights #4 and #5 the maximum computational time exceeded 10 min. Notwithstanding that this value does not jeopardize our solution approach, algorithmic efficiency for the full probabilistic framework is a point we want to improve as part of our future work.

Results confirm the risk-averse behavior of the model, as in all flights applying a confidence level of $\alpha = 0.6$ during the booking process results in having significantly fewer cases where shipments that were accepted during the booking phase are offloaded in the loading process. On the other hand, it is observed that using a confidence level of $\alpha = 0.5$ during the booking phase generates a sharp increase in the number of offloadings during the loading process. Apart from flight #5, as evident in Fig. 21, the ratio between the offloadings for $\alpha = 0.5$ and $\alpha = 0.6$ ranges between 3 and 7.5. If we take the average of the five flights, we have offloaded shipments in 532 cases if $\alpha = 0.5$ and in 153 cases if $\alpha = 0.6$. To get more insights into how this difference affects operations, Fig. 22 reports the accepted volumes for the five flights and the two confidence levels, and the average across the 1000 scenarios of the loaded volumes. When $\alpha = 0.5$, the average offloaded volume is 5.6 m^3, which corresponds to more than half an LPD or one LD-3 and a half. Note that the average is reduced by flight #5. For example, the offloaded volume for flight #1 exceeds 10 m^3. On the contrary, the average offloaded volume when $\alpha = 0.6$ is 1.2 m^3 and should be much easier to compensate and absorb. The relatively smaller difference in number of offloadings observed for flight #5 between the two confidence levels can be explained by the fact that approximately the same volume of shipments is accepted when adopting either a confidence level of $\alpha = 0.5$ or $\alpha = 0.6$ (see Fig. 20).

Clearly, this strategic advantage offered by the reduction of offloadings does come with a price, since the average loaded volume when $\alpha = 0.5$, as expected given the risk-prone nature of the approach, is higher. For flights 1# and #2, the difference is around

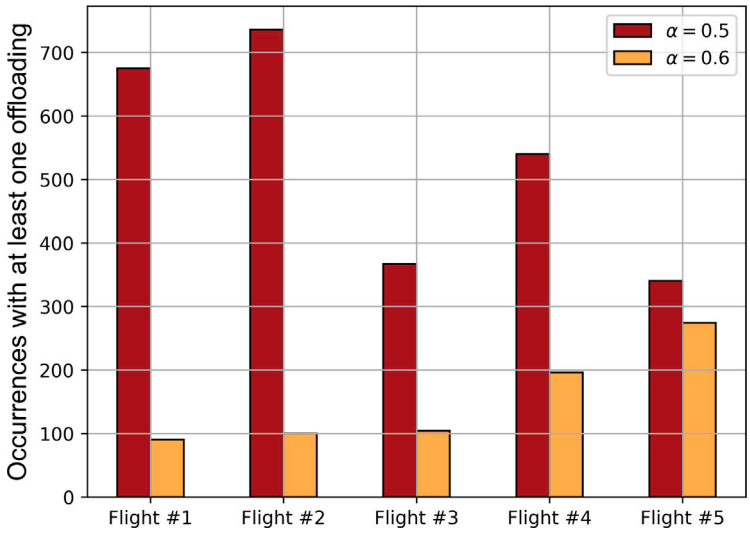


Fig. 21. Total number of scenarios per flight and per confidence level where at least one offloading occurred. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

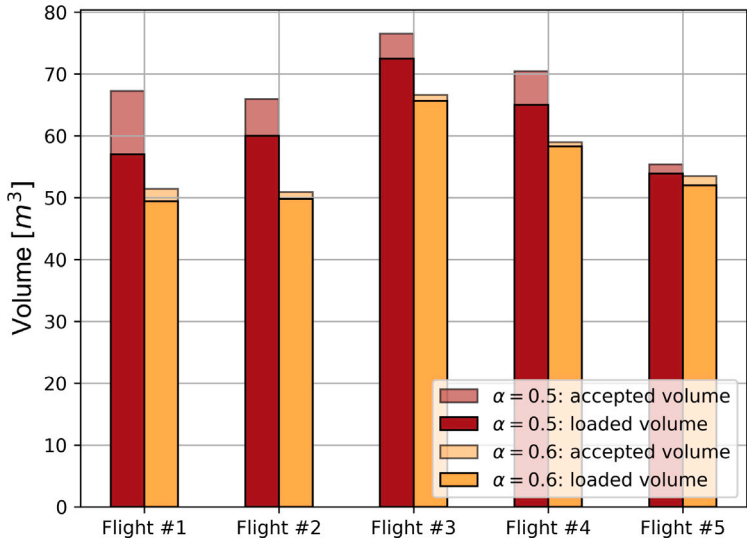


Fig. 22. Accepted and average loaded volume for the 1000 scenarios per confidence level and per flight. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

10 m^3 (the volume of an LDP). The difference is even larger if we consider the accepted volume, due to the slightly myopic acceptance criterion of the risk-prone approach. While we argue that the risk-averse approach still reaches good load factors and simultaneously limits offloadings, choosing the most appropriate confidence level is up to decision-makers of the airline. In principle, we are dealing with two conflicting objectives: maximizing expected loaded volume versus minimizing expected offloadings. For OD airport pairs connected multiple times a day, a lower α for the early flights might be compensated with a higher α for the later flights, so that they could absorb the offloadings generated earlier during the day. For OD airport pairs with scarce connections and early due dates, a high α is recommended to ensure on-time deliveries.

7. Conclusions

In this paper, we presented a novel combined forecasting and packing model, developed adopting the RM perspective of the cargo department of a combination airline. The overarching goal is to facilitate the decision-making process of accepting or rejecting a booking request. Due to its complexity and specificity, this problem has not been addressed deeply in the academic literature and, to our knowledge, there is no work that integrates forecasting and packing features under uncertainty into a single model.

The forecasting functionality of the model outputs a mean value and probability distributions of the predicted values for cargo capacity and shipment dimensions, and is implemented through an LSTM and MLP network. Both models provided very accurate predictions, while the produced distributions established prediction intervals with high PICP. The generated distributions are passed as inputs to the probabilistic packing model, which has a dual functionality: determine the optimal ULD configuration and the optimal palletization strategy under a specific confidence level. The former is achieved through a 1D-KP model while the latter through a 3D-MHKP model. The real-time environment requirement introduces the need for a good feasible solution in a short computational time. On that grounds, the 3D-MHKP model was designed using a state-of-the-art heuristic. Moreover, to comply with real operational needs, the heuristic was enriched with additional constraints referring to shipment possible orientations, stability, weight capacity of the ULD and complete shipments.

Experimentations were performed on booking datasets from five real flights provided by our partner airline. First of all, our findings highlight the computational efficiency of the deterministic packing model, as in any case, the computational time did not exceed the 89 s and it remained below 7 s for four out of five flights. Secondly, when using as inputs the shipment dimensions and cargo capacity as provided by our partner airline, hence in a deterministic context, in the five cases we achieved an improvement of 9.6%, 17.8%, 24.3%, 14.4%, and 8.4% in terms of volume of shipments loaded. As such, our bin packing heuristic outperformed the current heuristic adopted by the partner airline while being computationally efficient for a quasi real-time use.

The model was found to be more sensitive when varying the confidence level of the predicted shipment dimensions distribution, which is attributed to our design choice to assume a higher total standard deviation for a more risk-averse behavior. To further verify the risk-averse behavior, i.e., preventing discrepancies between the booking and the loading process, we simulated these two processes for two different confidence levels. While computational time was still reasonably low, for two flights out of five the computational time exceeded 600 s when considering the largest number of bookings. Our findings highlight that adopting a higher confidence level guarantees that, in approximately 90% of cases for flights #1–#3 and above 80% and 70% of cases for flights #4 and #5 respectively, all shipments that were accepted during the booking phase are eventually loaded. This is of course achieved giving up some potential loaded volume, which the risk-prone strategy is able to capture instead. We argue that, while we believe the risk-averse strategy to be more beneficial in the long term, our model provides a decision-making tool, where airlines can select the most appropriate confidence level. Factors such as due dates, most common nature of the commodity per flight leg, frequencies of the OD airport pair considered can all play a role in the selection of the most suitable confidence level. We argue, for example, that for OD airport pairs with high frequencies a low confidence interval can be chosen for the early flight legs, so that eventual offloadings can be re-captured in later flights that were planned with a higher confidence level.

We believe this model to be the first of its kind in the context of air cargo. As such, it can be improved in many ways. As it concerns the forecasting block, future work should address clustering shipments depending on their origin or the commodity type, to define cluster-specific confidence levels to evaluate the acceptance/rejection of an incoming booking request. For example, some commodity types might display a much more homogeneous range of dimensions, hence increasing the confidence of the prediction. This approach should help flight analysts lower the confidence level for shipments that are labeled as very reliable and activate a more risk-averse approach for shipments labeled as more unreliable. On a similar note, in this work we considered dimensions provided by customers to represent actual dimensions, i.e., to represent the ground truth. This is in line with the general feeling of our partner airline that customers are generally reliable when providing such piece of information. Notwithstanding, an improvement to our current model would entail measured and verified in-situ dimensions to be the ground truth, and dimensions provided by customers three additional model features. In such a way, reliability of customers (i.e., whether the dimensions they specify match the dimensions of the received shipments) can be taken into account and customer-specific confidence levels can be designed.

For the packing model, in this work we enforce that a shipment can be supported if all the four vertices defining its lower side are supported. In reality, this constraint can be relaxed by, for example, defining a minimum supported area even if one or two vertices are not supported (i.e., overhang shipments). Another future research direction is to include weight and balance restrictions and to devise the model so that it can account for multiple flight legs, being transshipment of paramount importance in the cargo business. This is a necessary step towards a model capable of addressing all the needs of the cargo department of a combination airline. Finally, and having the full probabilistic framework in mind, a wiser selection of a subset of the extreme points to consider in the packing heuristic or more tailored algorithmic improvements should help reduce even further the computational complexity for the most challenging scenarios. This might entail adopting ad-hoc variants of a tabu search or a genetic algorithm, potentially based on the EP paradigm, to decrease even further the gap with an exact formulation.

CRediT authorship contribution statement

Iordanis Tseremoglou: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Alessandro Bombelli:** Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Bruno F. Santos:** Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank the editor and three anonymous reviewers for their comments that improved the quality of the paper. In addition, we would like to show gratitude to the associated partner airline for providing their booking and flight capacity data needed for this work.

Appendix. Packing heuristic algorithm

Algorithm 2 Heuristic Packing Method.

Input P: list of i shipments to be loaded with dimensions $l_i \sim N(\bar{l}_i, \sigma^2_{l_i})$, width $w_i \sim N(\bar{w}_i, \sigma^2_{w_i})$, height $h_i \sim N(\bar{h}_i, \sigma^2_{h_i})$ and weight W_i.

$WeightCapacitySatisfied(W_i)$: function returning **true** if weight of shipments loaded do not exceed weight capacity of the ULD.

$CanBeSupported(l_i, w_i, h_i, \alpha)$: function returning **true** if all four vertices of shipment p_i (within the specified confidence level α) are properly supported.

$ULDBoundsSatisfied(l_i, w_i, h_i, \alpha)$: function returning **true** if the packed shipment p_i lies within the boundaries of the ULD within the specified confidence level α.

$EPProjectionSupported(l_i, w_i, h_i, \alpha)$: function returning **true** if the projected EP (within the specified confidence level α) is properly supported.

Set List of loaded shipment $\mathbf{L_p} \leftarrow \emptyset$
Set List of Extreme Points $\mathbf{EP_L} \leftarrow \emptyset$
Set Weight of shipments loaded $\mathbf{W_L} = 0$
Set List of Residual Space values $\mathbf{RS} \leftarrow \emptyset$
for all $\mathbf{p_i} \in \mathbf{P}$ **do**
  Set List of Candidate Extreme Points $\mathbf{EP_{Cand}} \leftarrow \emptyset$
  for all $RSvalues \in \mathbf{RS}$ **do**
    Compute $RSvalues$ components for every allowable rotation of shipment p_i
  end for
  Remove RS instances with negative $RSvalues$ components
  Compute corresponding Merit Functions
  Order Merit Functions by ascending order along with the associated **EPs** and shipment orientations
  Set $\mathbf{EP_{Cand}} \leftarrow$ **EPs**
  for $\mathbf{EP} \in \mathbf{EP_{Cand}}$ **do**
    if $WeightCapacitySatisfied$ **and** $ULDBoundsSatisfied$ **and** $CanBeSupported$ **then**
      Set $\mathbf{L_p} \leftarrow \mathbf{L_p} \cup \mathbf{p_i}$
      Compute $newEP$ projections as in Fig. 9(a)
      if corresponding $EPProjectionSupported$ **then**
        Set $\mathbf{EP_L}' \leftarrow \mathbf{EP_L}' \cup newEP$
      end if
      Set $\mathbf{EP_L} \leftarrow \mathbf{EP_L} \cup \mathbf{EP_L}'$
      Update Residual Space List $\mathbf{RS}'$ as in Fig. 10
      Set $\mathbf{RS} \leftarrow \mathbf{RS}'$
      Set $\mathbf{W_L} = \mathbf{W_L} + \mathbf{W_i}$
      break
    end if
  end for
end for
Output List of loaded shipments $\mathbf{L_p}$

References

Baldi, M., Perboli, G., Tadei, R., 2012. The three dimensional knapsack problem with balancing constraints. Appl. Math. Comput. 218, 9802–9818. http://dx.doi.org/10.1016/j.amc.2012.03.052.
Berndt, S., Jansen, K., Klein, K.M., 2020. Fully dynamic bin packing revisited. Math. Program. 179, 109–155.

Bombelli, A., Santos, B.F., Tavasszy, L., 2020. Analysis of the air cargo transport network using a complex network theory perspective. Transp. Res. E 138, 101959.
Brandt, F., Nickel, S., 2018. The air cargo load planning problem - a consolidated problem definition and literature review on related problems. European J. Oper. Res. 275, 399–410.
Chung, S.H., Ma, H.L., Hansen, M., Choi, T.M., 2020. Data science and analytics in aviation. Transp. Res. E 134, e101837.
Crainic, T., Gobato, L., Perboli, G., Rei, W., 2016. Logistics capacity planning: A stochastic bin packing formulation and a progressive hedging meta-heuristic. European J. Oper. Res. 253, 404–417.
Crainic, T., Gobato, L., Perboli, G., Rei, W., Watson, J., Woodruff, D.L., 2014. Bin packing problems with uncertainty on item characteristics: an application to capacity planning in logistics. Procedia-Soc. Behav. Sci. 111, 654–662.
Crainic, T., Perboli, G., Rei, W., Tadei, R., 2011. Efficient lower bound heuristics for the variable cost and size bin packing problem. Comput. Oper. Res. 38, 1474–1482.
Crainic, T., Preboli, G., Tadei, R., 2008. Extreme point-based heuristics for the three-dimensional bin packing. Inform J. Comput. 20, 368–384.
Delgado, F., Trincado, R., Pagnoncelli, B., 2019. A multistage stochastic programming model for the network air cargo allocation problem under capacity uncertainty. Transp. Res. E 131, 292–307.
Dereli, Z., Das, G., 2010. A hybrid simulated annealing algorithm for solving multi-objective container-loading problems. Appl. Artif. Intell. 23, 463–486.
Fuellerer, G., Doerner, K., Hartl, R., Iori, M., 2010. Metaheuristics for vehicle routing problems with three-dimensional loading constraints. European J. Oper. Res. 201, 751–759.
Gal, Y., 2016. Uncertainty in Deep Learning (Ph.D. thesis). University of Cambridge.
Garey, M., Johnson, D., 1979. Computers and Intractability: A Guide to the History of NP-Completeness.
Goncalves, J., Resende, M., 2012. A parallel multi-population biased random- key genetic algorithm for a container loading problem. Comput. Oper. Res. 39, 179–190.
Gupta, A., Guruganesh, G., Kumar, A., Wajc, D., 2017. Fully-dynamic bin packing with limited repacking. arXiv preprint arXiv:1711.02078.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780. http://dx.doi.org/10.1162/neco.1997.9.8.1735.
IATA, 2019. Cargo Stategy Report. www.iata.org.
Ivković, Z., Lloyd, E.L., 2009. Fully dynamic bin packing. In: Fundamental Problems in Computing. Springer, pp. 407–434.
Jozefowska, J., Pawlak, G., Pesch, E., Morze, M., Kowalski, D., 2018. Fast truck-packing of 3d boxes. Eng. Manage. Prod. Serv. 10, 29–40. http://dx.doi.org/10.2478/emj-2018-0009.
Junqueira, L., Morabito, R., Yamashita, D., 2012. Three-dimensional container loading models with cargo stability and load bearing constraints. Comput. Oper. Res. 39, 74–85.
Koch, M., 2019. An Optimisation and Forecast Framework for ULD Packing in the Air Cargo Supply Chain (Master's thesis). TU Delft, Netherlands.
Le Roux, N., Bengio, Y., 2010. Deep belief networks are compact universal approximators. Neural Comput. 8, 2192–2207.
Liu, Y., Yin, M., Hansen, M., 2019. Economic costs of air cargo flight delays related to late package deliveries. Transp. Res. E 125, 388–401.
Paquay, C., Schyns, M., Limburg, S., 2014. A mixed integer programming formulation for the three-dimensional bin packing problem deriving from an air cargo application. Int. Trans. Oper. Res. 23, 187–213. http://dx.doi.org/10.1111/itor.12111.
Paquay, C., Schyns, M., Limburg, S., 2018. A tailored two-phase constructive heuristic for the three-dimensional multiple bin size bin packing problem with transportation constraints. European J. Oper. Res. 267, 1–12. http://dx.doi.org/10.1080/00207543.2017.1355577.
Paquay, C., Schyns, M., Limburg, S., Oliveira, J., 2017. MIP-based constructive heuristics for the three-dimensional bin packing problem with transportation constraints. Int. J. Prod. Res. 1–12. http://dx.doi.org/10.1080/00207543.2017.1355577.
Rizzo, S.G., Lucas, J., Kaoudi, Z., Quiané-Ruiz, J., Chawla, S., 2019. AI-CARGO: A data-driven air-cargo revenue management system. CoRR abs/1905.09130. URL: http://arxiv.org/abs/1905.09130, arXiv:1905.09130.
Sagheer, A., Kotb, M., 2019. Time series forecasting of petroleum production using deep LSTM recurrent networks. Neurocomputing 323, 203–213. http://dx.doi.org/10.1016/j.neucom.2018.09.082.
Techanitisawad, A., Tangwiwatwong, P., 2004. A GA-based heuristic for the interrelated container selection loading problems. Ind. Eng. Manage. Syst. 3, 22–37.
Wada, M., Delgado, F., Pagnoncelli, B., 2017. A risk averse approach to the capacity allocation problem in the airline cargo industry. J. Oper. Res. Soc. 68, 643–651. http://dx.doi.org/10.1057/s41274-016-0135-x.
Wascher, G., Haußner, H., Schumann, H., 2007. An improved typology of cutting and packing problem. European J. Oper. Res. 183, 1109–1130.
Zhu, L., Laptev, N., 2017. Deep and Confident prediction for time series at Uber. In: Proc. IEEE Int. Conf. Data Min. Workshops. pp. 103–110.