



FORECASTING WORLDWIDE EMPTY CONTAINER AVAILABILITY WITH MACHINE LEARNING TECHNIQUES

Christoph Martius¹, Fraunhofer Center for Maritime Logistics and Services CML, Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany, christoph.martius@cml.fraunhofer.de

Lutz Kretschmann, Fraunhofer Center for Maritime Logistics and Services CML, Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany, lutz.kretschmann@cml.fraunhofer.de

Carlos Jahn, Fraunhofer Center for Maritime Logistics and Services CML, Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany, carlos.jahn@cml.fraunhofer.de

Ole John, Fraunhofer Center for Maritime Logistics and Services CML, Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany, ole.john@cml.fraunhofer.de

ABSTRACT	1
1. INTRODUCTION	2
2. DATA SET	4
2.1. Events in the Data Set	4
2.2. Descriptive Analysis of the Data	5
3. GENERAL PREDICTION APPROACHES	7
3.1. Neural Network Approach	8
3.2. Probabilistic Model	11
3.3. Evaluation Approach	11
4. RESULTS	12
5. CONCLUSIONS	16
ACKNOWLEDGMENTS	17
REFERENCES	17
AUTHOR(S) BIO NOTE(S)	19

ABSTRACT

Due to imbalances in the global transport of containerised goods, liner shipping companies go to great lengths to match the regional supply and demand for empty containers by transporting equipment from surplus to deficit regions. Making accurate forecasts of regional empty container availability could support liner companies and other involved actors by taking better relocation decisions and thus avoid unnecessary transportation costs of empty equipment. Against this background, this paper introduces two novel approaches based on machine learning and probabilistic techniques to predict the future availability of empty containers for more than 280 locations worldwide. The machine learning and probabilistic prediction models are built by analysing a unique data set of more than 100 million events from past container journeys. These events represent different stages during the transport process of a container. Both models use a two-step forecast logic: First, the expected future location of a container is predicted. Second, the expected timestamp for arriving at that location is estimated. The machine learning model uses artificial neural networks and mixture density networks to forecast the movements of containers. The models are quantitatively assessed and compared to the actual availability of containers and two more conventional forecasting approaches. The results indicate that the probabilistic prediction approach can keep up with conventional approaches while the neural network approach significantly outperforms the other approaches concerning every evaluation metric.

¹ Author for correspondence



Keywords: Maritime logistics forecasts, empty container relocation, machine learning, mixture density network

1. INTRODUCTION

The transportation of freight within supply chains and to consumers is a critical component of a globalised world. The timely availability of raw materials and finished goods plays an essential role for most companies. Accordingly, the efficient movement of cargo supports production, trade, and consumption (Crainic 2003, pp.1–3). Transport companies depend on the availability of an empty container with suitable characteristics in terms of equipment size, type, and condition to fulfil any transport order. In the ideal case – the balanced trade – containers would be stripped and filled with new cargo at the same location shortly after. Where this is impossible, empty equipment must either be stored in a depot until its subsequent use or transported to a region with an empty container deficit. Such relocations of containers from import-dominant to export-dominant areas take place locally (from one terminal to another), regionally (from one location to another) or globally (from one country to another) (Theofanis and Boile 2009; Di Francesco et al. 2009). High costs are associated with container repositioning, primarily due to over-land and marine transport and storage and handling of containers (Karmelić et al., 2012; Moon et al., 2013).

Empty containers make up a significant part of the container transports - up to 40% in over-land and 20% of marine transports (Konings and Thijs 2001; Karmelić et al. 2012, p.223). It is estimated that costs of empty container relocations amount to \$15 billion to \$20 billion annually, which represents around 5% to 8% of the carriers' total costs (Sanders et al., 2015). Hence, liner shipping companies go to great lengths to match the regional supply and demand for empty containers by optimising equipment repositioning to minimise costs. Additionally, the carrier reduces the risk of a competitor receiving an order and can realise the transports himself by satisfying the empty container demand at each port (Crainic 2003).

Liner shipping companies have various options to reduce the costs related to empty container logistics. Most common, lower transportation costs in import-dominant regions, sharing equipment between companies, or collaborating with competitors are used in practice. Another opportunity arises from recent advances in data-driven and machine learning approaches. Due to the digitalisation of transport processes in the context of the maritime industry 4.0, an increasing amount of transport-related data is collected daily. This enables the application of data-driven decision support models in various use cases, e.g. stowage planning (Shen et al. 2017), fuel management (Fagerholt et al. 2015), and container inspection (Klöver et al. 2020). In other logistic domains, such models, especially using machine learning techniques, are already applied to support business decisions, such as guiding self-driving vehicles (van Meldert and Boeck 2016; Jeon et al. 2011) or to optimise warehouse picking operations (Seward 2015). Where predictive models succeed in estimating supply and demand for empty containers, they possibly enable shipping companies to:

- realise container relocations in time if a surplus or deficit is foreseeable,
- adjust transport prices to meet expected market developments (dynamic pricing),
- plan transportation processes by considering future container availability to minimise transportation costs.

Overall, an accurate estimation of future demand and supply of empty containers contributes to the efficient utilisation of container fleets. Moreover, reducing regional and interregional equipment repositioning can also lower transport emissions (Crainic 2003; Schlingmeier 2017).

To this day, several scientific papers cover data-driven approaches to forecast container transport and port handling volumes. This also includes machine learning-based prediction models. For example, Shankar et al. (2020) use long-short-term-memory neural networks (LSTMs) to forecast the



container handling of the port of Singapore based on the historic quarterly throughput from 1995 to 2018 and compare the performance of their model to various time-series forecasting approaches (Shankar et al. 2020). Tsai and Huang (2017) use an artificial neural network approach to predict container movements between ten Asian ports. They train a regression model using features extracted from more than one million port-to-port container transports recorded in Taiwanese customs clearance documents. Additionally, they use financial data, like gross domestic product, interest rates and the industrial production index, as features (Tsai and Huang 2017). Another example is found in Xiao et al. (2012), who use a hybrid model which combines autoregressive integrated moving average (ARIMA) (cf. Box et al. 2015) with artificial neural networks to forecast univariate container throughput data from Tianjin Port (Xiao et al. 2012). Chan et al. (2019) compare various data-driven models to forecast container throughput of the port of Ningbo-Zhoushan. Their data basis is the yearly handling volumes between 2004 and 2015 (Chan et al. 2019). Starting from a similar database of quarterly container throughput in North Adriatic ports, Dragan et al. (2014) test different data-driven approaches, including ARIMA, to forecast regional container throughput on a port level (Dragan et al. 2014).

Even though several authors proposed and tested methods to forecast container throughput at ports, most are characterised by:

- relatively small data volumes, which prevents training robust models,
- no differentiation of container types,
- predictions that cover only small regions or even individual ports.

This paper tries to overcome these shortages. Two novel data-driven models are introduced – a deep learning model and a probabilistic model – to estimate the future availability of empty containers of different types for more than 280 locations worldwide. The machine learning model consists of feed-forward neural networks (cf. Bishop 2006) and mixture density networks (Bishop 1994). The prediction models are built by analysing a unique data set of more than 100 million events from past container movements. The events represent the different states of the transport process of a container, in particular, these are: “dispatch at container depot”, “loading on ship”, “discharge from ship” and “return to container depot”.

While related work mainly proposes to forecast the port throughput based on historical economic data of the ports (e.g. Dragan et al. 2014; Pang and Gebka 2017; Shankar et al. 2020), the models presented in this paper are built on journey data of individual containers. The subsequent events of each container are predicted for a large set of containers by forecasting the expected event location and realisation time. Based on the individual forecasts for a high number of containers, the supply and demand for various types of empty containers at worldwide locations is estimated. The proposed approach uses heterogeneous features for the prediction and enables the estimation of container availability for any port in the world.

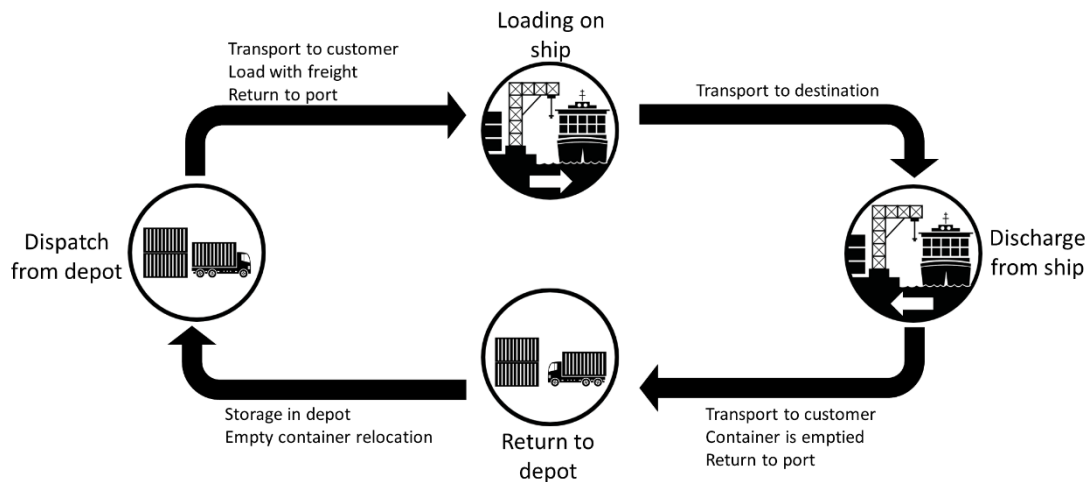
The remainder of the paper is structured as follows: First, the data set used to train and evaluate the prediction models is introduced in Section 2. Subsequently, Section 3 briefly describes the general prediction approach and methodologically introduces the two prediction models. The performance of the models is evaluated in Section 4. Finally, the paper concludes with a discussion of the performance and avenues for future research in Section 5.

2. DATA SET

2.1. Events in the Data Set

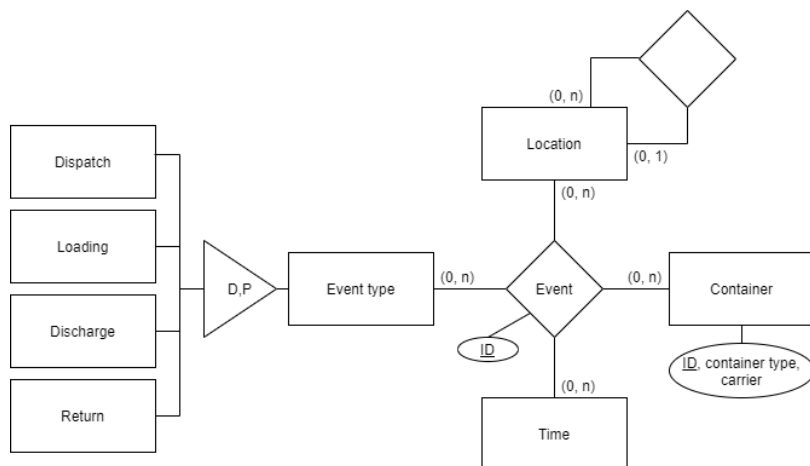
The machine learning and statistical prediction models are built by analysing a unique data set of more than 102 million events from past container journeys. These events represent different states during the transport process of a container, in particular the events “dispatch at container depot” (event name: dispatch), “loading on ship” (loading), “discharge from ship” (discharge) and “return to container depot” (return). Between these four events, different actions might take place. Usually, the container is taken out of a depot, transported to the customer, filled by the customer and subsequently returned to the port between a dispatch at container depot and the subsequent loading on ship. After loading, the container is transported from the origin to the destination before it is discharged from the ship. Subsequently, the container is transported to the customer, emptied and returned to the depot’s location. The container is stored in the depot between a return and a subsequent dispatch from the container depot. Eventually, an empty transport to another location can occur between these two events. Figure 1 depicts the events of a container journey.

Figure 1: Events during the transport process of containers



Source: own figure

Figure 2: Entity-relationship model of events



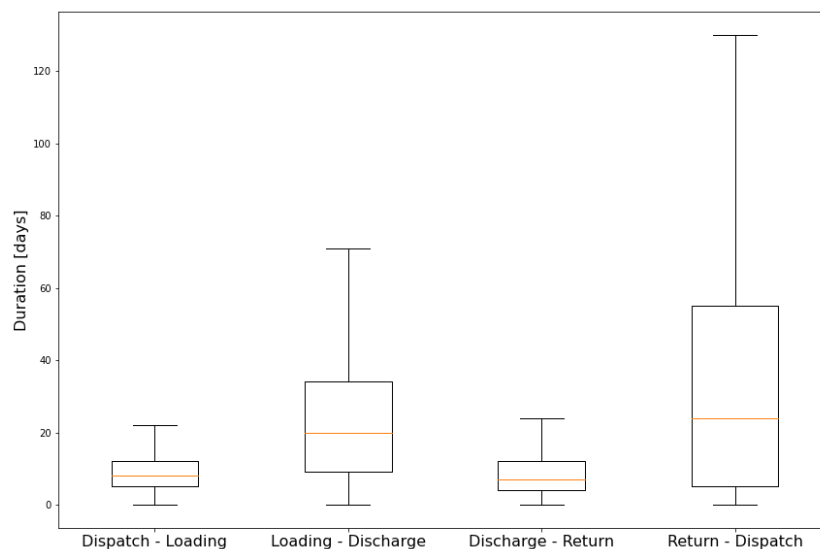
Source: own figure

Each event in the data set can be identified by a unique ID and belongs to a specific container. The type of the event (dispatch, loading, discharge, return) is provided and the location and time stamp. The events of the data set represent the movements of laden containers solely. Transports of empty containers are not included in the data set. A container ID can identify the containers in the data. Additionally, the type (e.g. forty-foot high cube) and the owner (carrier) of the container is provided. Figure 2 visualises the information related to an event with an entity-relationship model (cf. Chen 1976).

2.2. Descriptive Analysis of the Data

A container has, on average, an event every 19.7 days following the data. However, the duration between two subsequent events differs for each event type. For example, while the average time between a loading and a subsequent discharge (*transportation time*) corresponds to 23.5 days (lower quartile: 9 days, upper quartile: 34 days), the time between a return and a dispatch (*dwelling time*) averages to 35.2 days (5 days, 55 days). Figure 3 depicts the duration distribution between subsequent events with box and whiskers plots (cf. Tukey 1977).

Figure 3: Duration between subsequent events



Source: own figure

287 different locations are included in the data set. These locations either correspond to individual ports or countries. The data set comprises container events in 204 countries. Additionally, 83 locations are included in the data representing events at the most important ports like Shanghai, Ningbo, and Rotterdam. The events happening at these ports are treated separately; thus, they are excluded from the events of a country during models' training and evaluation.

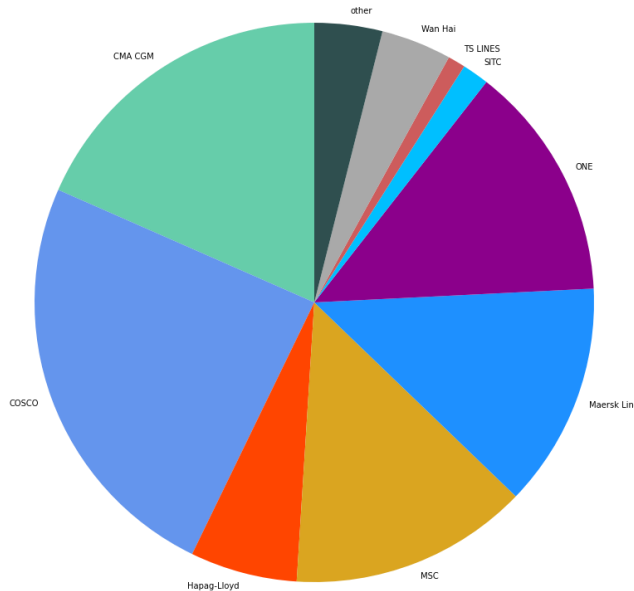
The containers included in the data set belong to one of 60 liner shipping companies. Therefore, container movements of each of the 14 largest liner shipping companies² are reflected in the data. However, the share of the events of a carrier does not correspond to its market share: For example, even though COSCO's market share corresponds to 12.2% as of July 2021³, 24.4% of the events

² According to <https://de.statista.com/statistik/daten/studie/28853/umfrage/kapazitaeten-der-containerschiffsflotte-der-groessten-reedereien/> (accessed on 09.09.2021)

³ According to <https://www.statista.com/statistics/198206/share-of-leading-container-ship-operators-on-the-world-liner-fleet/> (accessed on 12.08.2021)

relate to a COSCO container while just 12.9% of the events are of containers of the largest market player Maersk. Figure 4 visualises the share of the events of the liner shipping companies in the data.

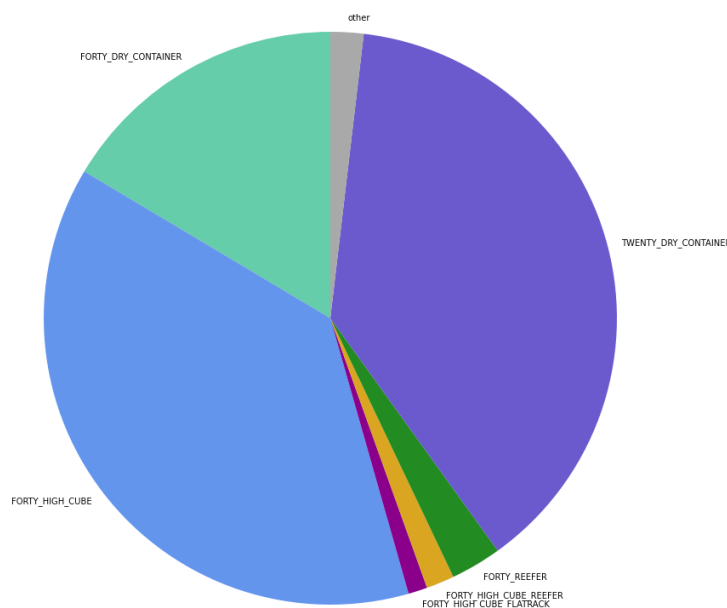
Figure 4: Share of liner shipping companies by number of movements



Source: own figure

Containers of 23 different types are included in the data. These container types include the most common variations of containers (dry container, reefer, flat racks, open tops, etc.) with a length of either twenty-, forty- or forty-five-foot. According to Che et al. (2011), twenty-foot, forty-foot dry and forty-foot high cube containers are among the most commonly used container types (Che et al. 2011, p. 506). These three container types account for 92.5% of the events in the data. Figure 5 visualises the share of the events of the different container types.

Figure 5: Share of container types

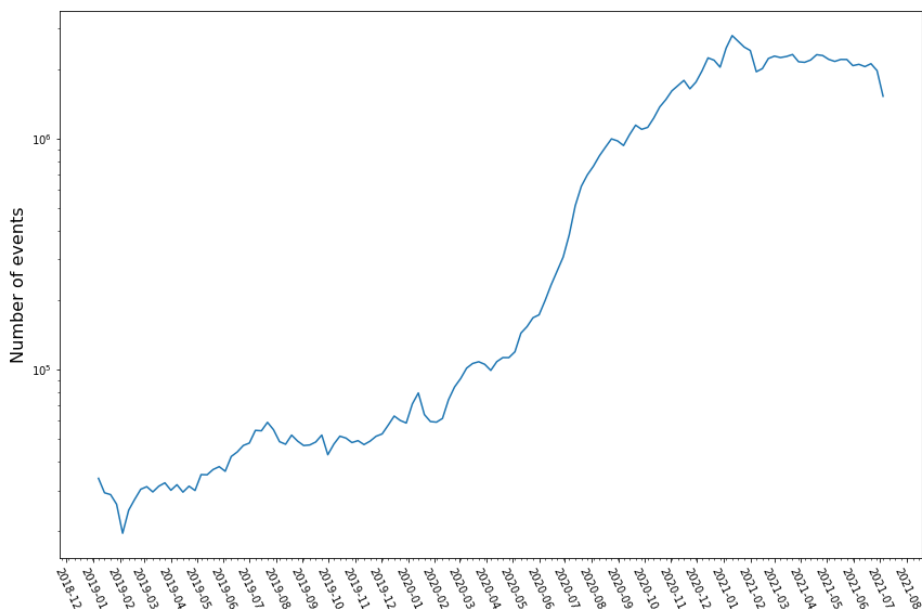


Source: own figure

The events in the data set are collected over more than 30 months: The first events date back to 1st January 2019 and the last to the 22nd July of 2021. 102,214,134 events of 12,893,347 containers are included in the data. Even though it covers a long time span and contains a large amount of total events, several challenges are associated with the available data set. The most important challenge is that the number of events per week increases considerably over the available period. While 219,653 events are tracked over the first two months of the covered time span, 19,108,155 events are tracked in May and June 2021, cf. Figure 6. Hence, most of the containers in the data are not tracked over the whole time span but only for some sporadic events over some weeks or months.

Furthermore, there are both missing events (or completely missing journeys) and inconsistencies in the data on a limited but not negligible level, even though the quality and completeness of the data increase continuously over the time span covered by the data set. All these aspects represent challenges to extracting information on the future demand and supply of empty containers and thus influence the quality of the models trained with the data.

Figure 6: Weekly number of events in the data set



Source: own figure

3. GENERAL PREDICTION APPROACHES

To support the determination of the regional empty container availability, the forecasting approaches aim to estimate the number of containers returned to container depot and dispatched at container depot at any location in the data set. The approach to estimating the number of returns and dispatches at the container depots presented in this paper is to forecast individual containers' movements (events). For each movement of a container, the location and date of the subsequent events has to be estimated. Then, based on the forecasts of the individual containers, the total number of returns and dispatches at the regional container depots can be derived, e.g. by a database query. The novelty of this prediction approach is that the future container throughput is forecasted by considering the current state of the global container fleet. Thereby, the model should anticipate the effects of the current situation. For example, suppose that a significant container shortage in an export-dominant region is observed at a certain point in time. In that case, the model can anticipate that this shortage causes the import to decline at the port, which mainly receives the freight from this region.



The prediction approach of a single container movement consists of five steps: First, features representing descriptive characteristics are derived from the last event, e.g. container type, location, or the time passed. Subsequently, a destination prediction model processes these features, producing a probability for each considered location to be the destination. Based on this probability distribution, one location is sampled as the destination. The defining features and the predicted destination are passed to a duration prediction model, which outputs another probability distribution representing the transition times of the container from the origin to the predicted destination. Again, one duration is sampled from the probability distribution. Finally, the date of the next container event is determined by adding the sampled duration to the date of the last event of the container. Subsequently, multiple upcoming container events can be estimated by repeating the prediction process with already forecasted events.

This paper introduces two different prediction models, which implement the described prediction approach. The first model uses two artificial neural networks to forecast the locations and dates of future events, while the second is a probabilistic model.

3.1. Neural Network Approach

The first approach combines two different types of neural networks to forecast the movements of an individual container as described by the general prediction approach. First, a feed-forward neural network (Bishop 2006, cf. Chapter 5 ff.) is used to predict the next event's location. Subsequently, a mixture density network (cf. Bishop 1994) is utilised to estimate the transition time based on the descriptive features of the last event and the predicted location.

Destination prediction:

To generate a probability distribution representing the probability p_l of each location $l_i \in L$ to be the location of the next container event, where L corresponds to the set of all locations in the data set, a feed-forward neural network is used. The model's input are descriptive features representing selected characteristics of the last event and the associated container. Table 1 lists a brief description of the features used to train the destination prediction model. An individual feed-forward neural network is trained for each of the four event types in the presented approach. Even though, it is expected that the forecasting approach can be successfully applied if the event type is included as a feature and hence, only one destination prediction model needs to be trained. Each categorical feature is passed as a one-hot encoded vector to the model.

Table 1: Features of the destination prediction model

Feature	Description	Type	Unique values
Origin	Location of the last event of the container. This can either be a specific port or the country where the event took place.	Categorical	287
Container type	String-based indicator of the type of the container, e.g. <i>FORTY-DRY-CONTAINER</i>	Categorical	23
Carrier	String-based indicator of the carrier of the container, e.g. <i>MSC</i>	Categorical	60
Days passed	Number of days passed between the last event of the container and the date the forecast is made.	Numeric	N.A.
Event type	String-based indicator of the type of the last event, either dispatch, loading, discharge, or return	Categorical	4

Source: authors

The destination prediction models are utilised to output a probability distribution given the descriptive features of an event as input. Each probability distribution is supposed to assign a

probability p_{l_i} , with $0 \leq p_{l_i} \leq 1$ for each location $l_i \in L$, in such a way that $\sum_i^{|L|} p_{l_i} = 1$. The output of the neural network is scaled with the softmax activation function (cf. Nwankpa et al. 2018) to satisfy this constraint.

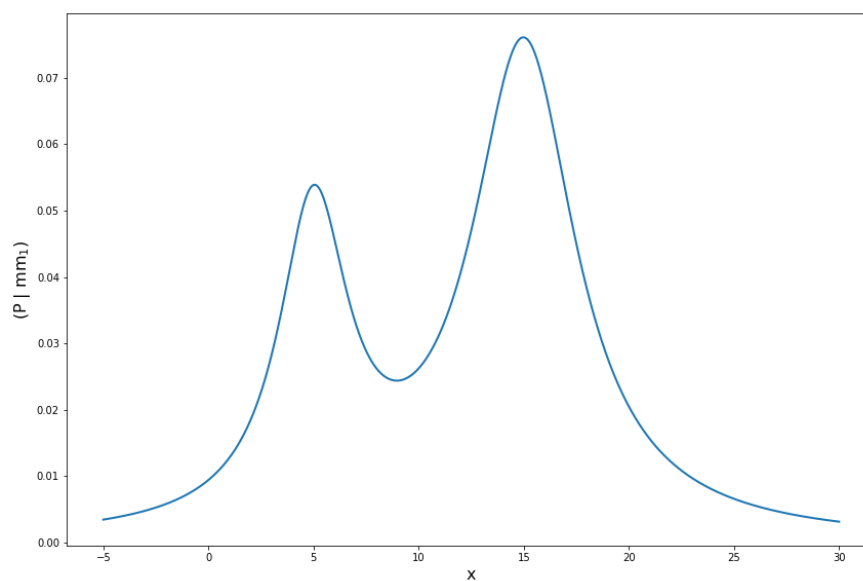
The features are processed by three fully-connected layers consisting of 700 neurons with Rectified Linear Unit (ReLU) (cf. Nair and Hinton 2010) as activation function before they are passed to the output layer. The models are trained with the historic events of the data set. Thereby, the single location of the next event serves as target. The proposed model can learn to assign a probability for each location even though only a single destination is passed as target by learning from multiple events described by the same or similar features. The destination prediction models are trained for 10 epochs with the adam optimiser with a batch size of 1028.

Duration prediction:

Mixture density networks are used to estimate the date of the next event. Mixture density networks are a special type of neural network designed to output a mixture model given some input features (cf. Bishop 1994). A mixture model can be thought of as a probabilistic model consisting of k components C_1, \dots, C_k (or subpopulations) which describe the characteristics of a collection of elements (population). Each component itself is a probabilistic model (e.g. probability density function) representing the characteristics of a subgroup of the overall population (Baxter 2010).

A component can be parametrised by different values depending on the type of probability distribution of the component (e.g. Gaussian distribution, exponential distribution). For example, if the component C_1 is a Gaussian distribution, the component can be represented by two values: mean (μ_1) and standard deviation (σ_1), hence $C_1 = \mathcal{N}(\mu_1, \sigma_1)$. Additionally, each component of a mixture model is weighted by a weighting factor $\alpha_1, \dots, \alpha_k$, with $\sum_i^k \alpha_i = 1$. Figure 7 visualises a mixture model consisting of two components, $mm_1 = P(x|C_1, C_2, \alpha_1, \alpha_2)$. The first component is a Gaussian distribution with a mean of 5 and standard deviation of 2 ($C_1 = \mathcal{N}(5, 2)$) and the second component a Gaussian distribution with mean 15 and standard deviation of 3 ($C_2 = \mathcal{N}(15, 3)$). The first component is weighted by $\alpha_1 = 0.3$ and the second component by $\alpha_2 = 0.7$ (cf. Baxter 2010).

Figure 7: Exemplary mixture model



Source: own figure

Mixture density networks estimate the parameters of a mixture model. Hence, the mixture density network needs to output for each component a weighting factor α and the parameters of the distribution, e.g. μ and σ . In case that all components can be parametrised by μ_1, \dots, μ_k and $\sigma_1, \dots, \sigma_k$ the output o of the mixture density network can be thought of as a vector $o = \{M, \Sigma, A\}$, where $M = [\mu_1, \dots, \mu_k]$, $\Sigma = [\sigma_1, \dots, \sigma_k]$ and $A = [\alpha_1, \dots, \alpha_k]$. To satisfy the constraint that $\sum_i^k \alpha_i = 1$, softmax is applied to A . A slightly modified exponential linear unit (ELU) activation (cf. Clevert et al. 2015) is applied to the remaining output values of the model ($\lambda_1(x), \lambda_2(x)$) to prevent that the values of M and Σ become smaller or equal to zero (cf. Borchers 2019):

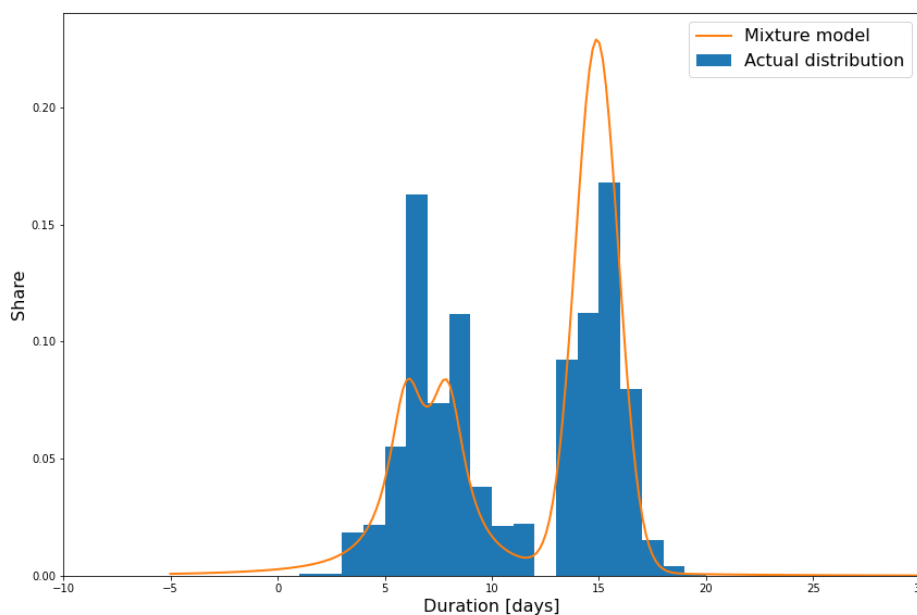
$$M(x) = \text{ELU}(\lambda_1(x)) + 1, \text{ respectively}$$

$$\Sigma(x) = \text{ELU}(\lambda_2(x)) + 1.$$

The model learns the values of M, Σ , and A by minimising the average negative log-likelihood of the ground truth values (transition times of the containers between the events) to be represented by the mixture model constructed by the mixture density network (cf. Borchers 2019). Thereby, the model is supposed to learn to construct the mixture model to reflect the distributions of the historical data.

The presented models are composed of three fully-connected layers with 500 neurons. The input to the model is a vector representing the one-hot encoded origin, destination, carrier, and container type. Thereby, the model is supposed to learn to output a mixture model for each input feature vector to reflect the distributions of the historical data. Our best-performing mixture density networks are combined from two Cauchy distributions and one Gaussian distribution. One mixture density network is trained for each event type as for the destination prediction. The models are trained for 40 epochs with the adam optimiser and a batch size of 128,000 to speed up the training process. During the inspection of the model's output, it was observed that the model can learn realistic distributions even for feature combinations that are not part of the historical data by generalising from historical data of similar feature combinations.

Figure 8: Predicted mixture model for transportation time from Shanghai to the United States



Source: own figure

Figure 8 depicts the mixture model for the transportation time learnt by the mixture density network and the distribution of the actual values of the following features: the port of Shanghai is

selected as the origin, the United States as the destination, Wan Hai as the carrier, and no specific container type is provided. 3,589 events with this feature combination are within the training data.

To determine the date of the next event of a container, a duration d_{sel} is sampled from the mixture model provided by the mixture density network - given the descriptive features of the last event of the container - and added to the date of the last event. The selected duration d_{sel} has to satisfy the constraint $d_{sel} > \delta_{event}$, where δ_{event} corresponds to the days that have passed since the last event. Thereby, it is ensured that the next event will happen in the future.

3.2. Probabilistic Model

The probabilistic model follows the same prediction procedure as the neural network approach. However, the learning process is simpler than the neural networks. The general idea is that the predicted location and duration is the same as one historical event described by the same or similar descriptive features. Thus, the probabilistic model used three features (origin, event type and days passed) to predict the location and four features (origin, destination, event type and days passed) to estimate the date of the next event.

The probabilistic model learns a probability distribution for each feature combination from the historic container movements. Thereby, the feature combination serves as input for the probabilistic model. The probabilistic model outputs a location duration based on the input feature combination. The probability distribution assigns a probability p_{l_i} , with $0 \leq p_{l_i} \leq 1$ for each location $l_i \in L$, in such a way that $\sum_i^{|L|} p_{l_i} = 1$. The probability distribution for each feature combination depends on the historic destinations, respectively duration of containers with feasible feature combination. A feature combination is considered feasible if the location-based features and the event type are the same as the descriptive features of the last event of the container and the duration is larger than the days passed since the last event of the container under consideration.

For example, 10 containers were shipped from Lisbon following the historical data. 3 of those were discharged from the ship 50 days later in Shanghai and 2 after 70 days in Shanghai. 3 arrived in New York after 28 days. The remaining two were sent to Hamburg, Germany and were discharged after 17 days. If the probabilistic model is used to predict the next location and duration of a container that was loaded in Lisbon 20 days ago, then the probability that Shanghai is selected as a destination corresponds to $\frac{5}{8} = 62.5\%$. In case Shanghai is sampled as the destination, the statistical model selects a shipment duration of 50 days with a probability of $\frac{3}{5} = 60\%$.

3.3. Evaluation Approach

The objective of the forecasts is to estimate a weekly container availability index of specific container types (either twenty-foot or forty-foot) at each location in the data set. The container availability index (CAX) is calculated following the approach used by xChange⁴, which is defined as the number of containers of a given type ct (e.g. twenty-foot standard container) returned at a specific location l during a certain week w ($returns_{l,ct,w}$) divided by the sum of returns $returns_{l,ct,w}$ and the number of dispatches of containers with the same type ct at location l within week w ($dispatches_{l,ct,w}$).

⁴ <https://container-xchange.com/features/cax/>



$$CAX_{l,ct,w} = \frac{(\text{returns}_{l,ct,w})}{(\text{returns}_{l,ct,w}) + (\text{dispatches}_{l,ct,w})}$$

The absolute error and squared error between the predicted and actual container availability for each evaluation feature combination (combination of location l , week w and container type ct) in the evaluation set is calculated to quantify the quality of the container availability index predicted by the models. The mean absolute error (MAE) resp. mean squared error (MSE) is computed by averaging the absolute errors resp. squared errors over all evaluation feature combinations of the evaluation set. A small MAE and MSE indicate that the predicted container availability indices are similar to the actual values of the data in the evaluation set. If the MAE of two different models are quite similar, but the MSE of one model is significantly larger than the MSE of the other, it indicates that the model with the higher MSE predicts for at least one evaluation feature combination a significantly different container availability index than reflected by the actual data.

The models are evaluated on rolling forecasting origins (cf. Hyndman and Athanasopoulos 2018): multiple evaluation data sets representing different time spans are used to quantify the performance of the two prediction models. The evaluation data is selected following the sample-out approach for each forecasting origin: events that took place before the forecasting origin are used for training the models, while the objective of the models is to predict the event after the forecasting origin. Each evaluation data set covers the events of a time span of 12 weeks. Thereby, the models have to forecast up to the next 20 events of each container (on average 4 to 5 events). The prediction models are trained and evaluated on 18 forecasting origins. The first forecasting origin is 01.11.2020 and the last 28.02.2021. The training data includes all the events dating back to January 2019.

In addition to the actual values, the quality of the two prediction models is compared to two benchmark functions, Benchmark Naïve and Benchmark SES. Benchmark Naïve predicts the same number of events for each evaluation feature as in the last week before the forecasting origin. Benchmark SES instead applies a simple exponential smoothing on the amounts for each feature combination over the 12 weeks before the forecasting origin (cf. Hyndman and Athanasopoulos 2018, Chapter 7.1).

The number of events in the data set increases from week to week, and the share of four event types is not evenly distributed due to missing data as described in Section 2. This leads to a bias regarding the number of events happening at the ports each week. Therefore, the evaluation data of each time span has to be cleaned to mitigate this issue. The basic idea of the cleaning process is that containers are only included in the evaluation data, which are (probably) constantly tracked over a time span and thereby show realistic movements. Four criteria are defined that must be satisfied by a container to be included in the evaluation data set:

1. The container is tracked at least 150 days before the forecasting origin.
2. The container has at least one event during the 150 days after the last day of the evaluation time span.
3. The events of the container happen in the expected order (dispatch → loading → discharge → return → dispatch).
4. The time between two consecutive events of a container is never larger than 150 days.

4. RESULTS

In this section, the forecast results of our two prediction models over the 18 forecasting origins are presented, compared to the predictions of the benchmarking approaches, and some insights are

highlighted. Table 2 shows the MAE and MSE averaged over the 18 forecasting origins for the two prediction approaches and the two benchmarks for locations where at least 50 containers are returned or dispatched from the depot.

Table 2: MAE and MSE over the evaluation period

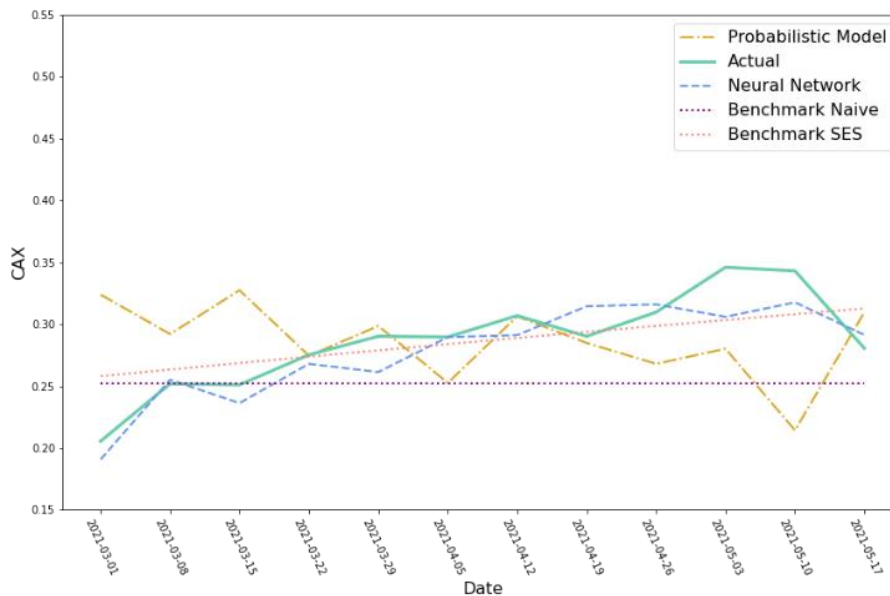
Approach	Mean Absolute Error		Mean Squared Error	
	Twenty-foot	Forty-foot	Twenty-foot	Forty-foot
Neural Network	0.0753	0.0711	0.0093	0.0086
Probabilistic Model	0.1048	0.0870	0.0170	0.0122
Benchmark Naive	0.1086	0.0925	0.0194	0.0147
Benchmark SES	0.1018	0.0864	0.0176	0.0131

Source: authors

The forecasts of the neural network approach significantly outperform the other three approaches with a MAE of 0.0753 for the twenty-foot container (MSE: 0.0093) and 0.0711 for the forty-foot container (MSE: 0.0086). While the prediction performance of the neural network is promising, the accuracy of the probabilistic approach is similar to the accuracy of the two benchmark models. Even though the observation that the probabilistic model performs similarly to the two benchmarks is still promising: it indicates that the applied prediction approach, which derives the container availability based on forecasts of individual containers, can compete with predictions on port level.

As shown in Table 2, the prediction quality of each model is better for forty-foot containers than twenty-foot containers. Larger availability of training and evaluation data for forty-foot containers might explain this observation, cf. Figure 5. Generally, it was noticed during evaluation that the performance of the neural network approach seems to be positively influenced by higher availability of training and evaluation data.

Figure 9: Predicted container availability for the port of Qingdao (forecasting origin: 01.03.2021)

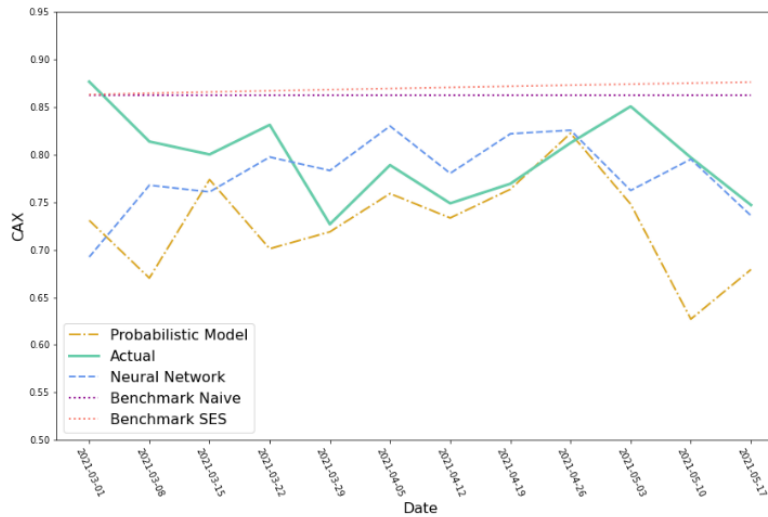


Source: own figure

Figure 9 and Figure 10 show the twelve-week container availability forecasts of the four approaches starting from the forecasting origin 1st March 2021 for the ports of Qingdao and Tokyo.

Over the depicted time period, the container availability index at the port of Qingdao increased from 0.21 to 0.35. The neural network approach and Benchmark SES quite accurately anticipate the increase while the other two models fail to forecast the observed pattern accurately. The observation is also captured by the MAE (cf. Table 3): The neural network approach and Benchmark SES have a MAE of 0.0160 resp. 0.0202 followed by Benchmark Naïve (0.0426) and the probabilistic model (0.0460).

Figure 10: Predicted container availability for the port of Tokyo (forecasting origin: 01.03.2021)



Source: own figure

The development of the container availability index at the port of Tokyo is more volatile and does not follow a strict pattern. All four approaches do not accurately forecast container availability in Tokyo. The predictions made by the neural network approach are the most accurate according to the MAE (0.0499), followed by the naïve benchmark (0.0680). The weekly container availability index at the port of Tokyo is calculated on average from 115 containers of the evaluation data set. In comparison, on average, 795 containers of the evaluation data set are dispatched or returned each week at the port of Qingdao. Hence, the volatility of the container availability index at the port of Tokyo might originate from the relatively small number of events at this port.

Table 3: MAE of the prediction models for the port of Qingdao and Tokyo

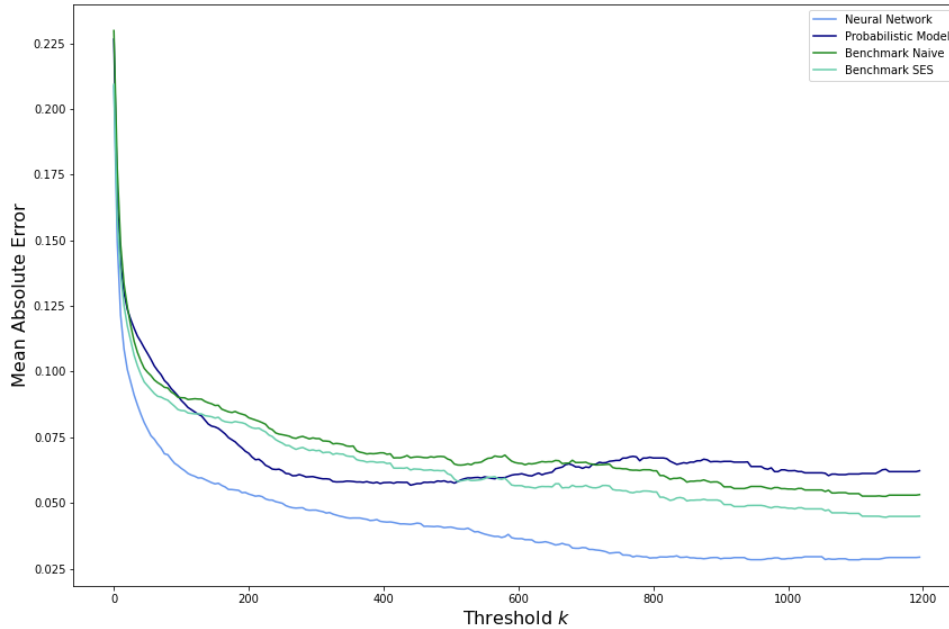
Approach	Mean Absolute Error	
	Qingdao	Tokyo
Neural Network	0.0160	0.0499
Probabilistic Model	0.0460	0.0712
Benchmark Naive	0.0426	0.0680
Benchmark SES	0.0202	0.0752
Total container	12,729	1,843

Source: authors

To test the hypotheses that the forecast quality increases with a higher number of container events at the locations, the MAE was computed for different $subsets_k$ of the evaluation data set. In each $subsets_k$ locations with less than k weekly events were excluded from the evaluation data set. For example, $subsets_{50}$ only includes the ports and countries where at least 50 container events occur

weekly. Figure 11 depicts the MAE of subsets of the evaluation data set with different thresholds k , in such a way that it holds that $0 \leq k \leq 1200$.

Figure 11: MAE for different subsets of the evaluation data set



Source: own figure

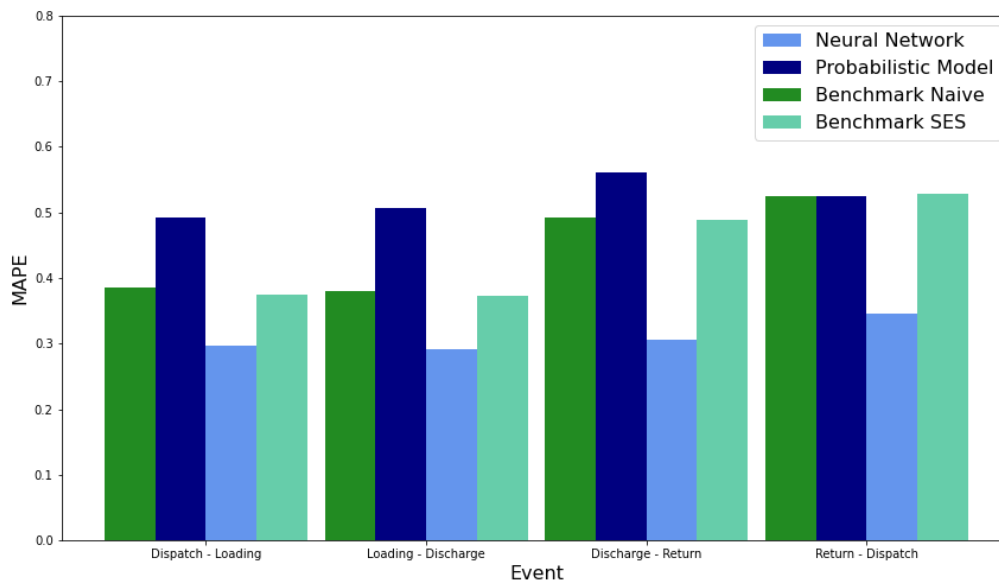
It can be seen in Figure 11 that the predictive performance of the neural network approach, Benchmark Naïve and Benchmark SES – expressed by the MAE – nearly constantly increases with a higher number of container events at the locations of the evaluation data set. Only the probabilistic model does not follow this pattern. Instead, the prediction quality of the probabilistic model slightly decreases if k is greater than 400. Additionally, it can be observed that the MAE of all four approaches is relatively high if all locations are included in the evaluation set, indicating that the models do not accurately predict the container availability for locations with only a few weekly container events. This can mainly be attributed to the volatility of the container availability index at these locations. However, it can be seen that the predictive performance significantly increases if low-volume locations are excluded from the evaluation data set. At locations with more than 200 weekly container events, the neural network approach achieves an MAE of approximately 0.05. The metric even improves to 0.03 if only locations with more than 800 weekly events are included.

Furthermore, it can be seen that the model using neural networks outperforms the other three approaches for each threshold k . This observation also indicates that forecasts made on the individual container level can outperform predictions on the port level. The advantage of the neural network approach compared to predictions on port level especially is that the presented approach can forecast the movements of the containers given the current state of the global container fleet. For example, if only a relatively small number of containers were loaded on vessels in Chinese ports over the last weeks, the presented prediction approach can anticipate that this situation will result in a decrease of discharged containers at ports that mainly import goods from China. Furthermore, the prediction approach can estimate the implications of this situation over the following weeks at every considered port and country.

In addition to the MAE of the forecasted container availability index over the 18 forecast-origins, the mean absolute percentage error (MAPE) was used to quantify the ability of four models to forecast the absolute number of containers dispatched and returned at the locations around the world. The

MAPE indicates the relative difference between the absolute and predicted number of events of one event type (dispatch, loading, discharge, return). For example, if 10 containers are dispatched from container depot in one week at a specific location and 20 containers at another and the model forecasts 12 dispatches for the first location and 15 for the second location, the absolute percentage error of the first location corresponds to $\frac{|12-10|}{10} = 0.2$, and the absolute percentage error of the second location equals $\frac{|15-20|}{20} = 0.25$. Hence, the MAPE of the dispatches in this example is $\frac{0.2+0.25}{2} = 0.225$. Figure 12 shows the MAPE of the four models for each of the four event types over the 18 forecasting origins for locations with more than 50 weekly events.

Figure 12: MAPE of the prediction models over the 18 forecasting origins



Source: own figure

It can be seen in Figure 12 that the neural network approach outperforms the other three prediction approaches regarding the prediction of the absolute number of container events at the locations in the evaluation set with a MAPE between 0.2909 and 0.3460, followed by Benchmark Naïve with a MAPE between 0.3797 and 0.5246. However, the probabilistic model is last for three of the four event types with a MAPE between 0.4929 and 0.5606. Even though the neural network performs better than the other three models, the MAPE of all models is still relatively high: a MAPE of approximately 0.3 for the neural network approach indicates that the model either over or underestimates the absolute number of container events at the locations by 30%. Similarly, the observation was made for the container availability index that the forecasts improve for locations with more data.

5. CONCLUSIONS

Liner shipping companies frequently have to transport empty containers from import-dominant regions to export-dominant regions to match the local demand for empty containers. These empty container relocations cause significant financial and capacity-related expenses for the carriers. A novel data-driven prediction approach to support the decision-making process related to empty container logistics was presented in this paper. While previous related work estimated the future container throughput on a port level, the container availability is derived from forecasts of movements of individual containers in the presented approach. Thereby, the models can incorporate knowledge



of the current situation of the global container fleet when forecasting the container throughput at the port around the world.

Two prediction models implementing the novel prediction approach were introduced and compared to two conventional methods which estimate the container throughput based on historical data on a port level. The neural network model outperforms the other three approaches, which underlines the applicability of the introduced prediction approach. The prediction quality of the probabilistic model instead is similar to the conventional approaches.

The presented approach using neural networks can approximate the container availability for the selected locations with an average MAE of around 0.05. It was shown that predictions tend to be more accurate if more training and evaluation data is available. While the neural network approach can predict a container availability index relatively accurately, all models fail to estimate empty containers' absolute weekly supply and demand. The forecasts of the best performing model differ from the actual values by 30% on average. However, the prediction quality is expected to increase with more high-quality data and less exceptional circumstances (i.e. COVID pandemic) represented in the training and evaluation data set.

In future research, the use of additional characteristic features, e.g. financial indicators or time-related information, is expected to improve the prediction quality of the forecasting approaches. Additionally, a more extensive hyperparameter optimisation, primarily focusing on the distribution used by the mixture density network, might amend the accuracy of the models. Furthermore, it would be interesting to include more individual ports in the training and evaluation process. However, the number of ports that can be used for training and evaluation depends on the availability of sufficient data.

The forecasts of container availability can be used by various parties participating in the maritime supply chain. Liner shipping companies can, for example, use forecasts as a tool to support the planning process related to empty container relocations. Additionally, transportation prices can be adjusted based on the future demand and supply of empty containers to optimally utilise the container fleet or maximise earnings.

ACKNOWLEDGMENTS

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany in the framework of the project C-TIMING (Container Availability Index Made in Germany) (project number: 01IS20004).

REFERENCES

- Baxter, Rohan A. (2010): Mixture Model. In Claude Sammut, Geoffrey I. Webb (Eds.): Encyclopedia of Machine Learning. Boston, MA: Springer US, pp. 680–682.
- Bishop, Christopher M. (1994): Mixture density networks.
- Bishop, Christopher M. (2006): Pattern recognition and machine learning: Springer.
- Borchers, Oliver (2019): A Hitchhiker's Guide to Mixture Density Networks. Edited by Towards data science. Medium. Available online at <https://towardsdatascience.com/a-hitchhikers-guide-to-mixture-density-networks-76b435826cca>, updated on 2/15/2019, checked on 7/5/2021.
- Box, George E. P.; Jenkins, Gwilym M.; Reinsel, Gregory C.; Ljung, Greta M. (2015): Time series analysis: forecasting and control: John Wiley & Sons.



- Chan, Hing Kai; Xu, Shuojiang; Qi, Xiaoguang (2019): A comparison of time series methods for forecasting container throughput. In *International Journal of Logistics Research and Applications* 22 (3), pp. 294–303. DOI: 10.1080/13675567.2018.1525342.
- Che, Chan Hou; Huang, Weili; Lim, Andrew; Zhu, Wenbin (2011): The multiple container loading cost minimization problem. In *European Journal of Operational Research* 214 (3), pp. 501–511. DOI: 10.1016/j.ejor.2011.04.017.
- Chen, Peter Pin-Shan (1976): The entity-relationship model—toward a unified view of data. In *ACM transactions on database systems (TODS)* 1 (1), pp. 9–36.
- Clevert, Djork-Arné; Unterthiner, Thomas; Hochreiter, Sepp (2015): Fast and accurate deep network learning by exponential linear units (elus). In *arXiv preprint arXiv:1511.07289*.
- Crainic, Teodor Gabriel (2003): Long-haul freight transportation. In : *Handbook of transportation science*: Springer, pp. 451–516.
- Di Francesco, Massimo; Crainic, Teodor Gabriel; Zuddas, Paola (2009): The effect of multi-scenario policies on empty container repositioning. In *Transportation Research Part E: Logistics and Transportation Review* 45 (5), pp. 758–770. DOI: 10.1016/j.tre.2009.03.001.
- Dragan, Dejan; Kramberger, Tomaž; Intihar, Marko (2014): A comparison of methods for forecasting the container throughput in north Adriatic ports. In : *Conference IAME*.
- Fagerholt, Kjetil; Gausel, Nora T.; Rakke, Jørgen G.; Psaraftis, Harilaos N. (2015): Maritime routing and speed optimization with emission control areas. In *Transportation Research Part C: Emerging Technologies* 52, pp. 57–73. DOI: 10.1016/j.trc.2014.12.010.
- Hyndman, Rob J.; Athanasopoulos, George (2018): *Forecasting: principles and practice*: OTexts.
- Jeon, Su Min; Kim, Kap Hwan; Kopfer, Herbert (2011): Routing automated guided vehicles in container terminals through the Q-learning technique. In *Logist. Res.* 3 (1), pp. 19–27. DOI: 10.1007/s12159-010-0042-5.
- Karmelić, Jakov; Dundović, Čedomir; Kolanović, Ines (2012): Empty container logistics. In *PROMET-Traffic & Transportation* 24 (3), pp. 223–230.
- Klöver, Steffen; Kretschmann, Lutz; Jahn, Carlos (2020): A first step towards automated image-based container inspections.
- Konings, Rob; Thijs, Remmelt (2001): Foldable containers: a new perspective on reducing container-repositioning costs. In *European journal of transport and infrastructure research* 1 (4).
- Moon, Ilkyeong; Do Ngoc, Anh-Dung; Konings, Rob (2013): Foldable and standard containers in empty container repositioning. In *Transportation Research Part E: Logistics and Transportation Review* 49 (1), pp. 107–124. DOI: 10.1016/j.tre.2012.07.005.
- Nair, Vinod; Hinton, Geoffrey E. (2010): Rectified linear units improve restricted boltzmann machines. In : *Icml*.
- Nwankpa, Chigozie; Ijomah, Winifred; Gachagan, Anthony; Marshall, Stephen (2018): Activation functions: Comparison of trends in practice and research for deep learning. In *arXiv preprint arXiv:1811.03378*.
- Pang, Gu; Gebka, Bartosz (2017): Forecasting container throughput using aggregate or terminal-specific data? The case of Tanjung Priok Port, Indonesia. In *International Journal of Production Research* 55 (9), pp. 2454–2469. DOI: 10.1080/00207543.2016.1227102.
- Sanders, Ulrik; Kloppsteck, Lars; Roeloffs, Christian; Schlingmeier, Johannes; Riedl, Jens (2015): *Think Outside Your Boxes: Solving the Global Container-Repositioning Puzzle*. Edited by Boston Consulting Group. Available online at <https://www.bcg.com/de-de/publications/2015/transportation-travel-logistics-think-outside-your-boxes-solving-global-container-repositioning-puzzle>, updated on 11/17/2015, checked on 8/7/2021.
- Schlingmeier, Johannes (2017): Kooperationen in der maritimen Leercontainerlogistik. eine empirische Untersuchung mit Fokus auf Equipment Interchange und Container Pools. In : *Innovationen für die maritime Logistik*, vol. 3. Hamburg: Fraunhofer Verlag (3).
- Seward, Calvin (2015): *Optimizing Warehouse Operations with Machine Learning on GPUs*. Edited by NVIDIA Developer Blog. NVIDIA. Available online at <https://developer.nvidia.com/blog/optimizing-warehouse-operations-machine-learning-gpus/>, updated on 12/13/2015, checked on 9/9/2021.
- Shankar, Sonali; Ilavarasan, P. Vigneswara; Punia, Sushil; Singh, Surya Prakash (2020): Forecasting container throughput with long short-term memory networks. In *IMDS* 120 (3), pp. 425–441. DOI: 10.1108/IMDS-07-2019-0370.



- Shen, Yifan; Zhao, Ning; Xia, Mengjue; Du, Xueqiang (2017): A Deep Q-Learning Network for Ship Stowage Planning Problem. In *Polish Maritime Research* 24 (s3), pp. 102–109. DOI: 10.1515/pomr-2017-0111.
- Theofanis, Sotirios; Boile, Maria (2009): Empty marine container logistics: facts, issues and management strategies. In *GeoJournal* 74 (1), p. 51.
- Tsai, Feng-Ming; Huang, Linda J.W. (2017): Using artificial neural networks to predict container flows between the major ports of Asia. In *International Journal of Production Research* 55 (17), pp. 5001–5010. DOI: 10.1080/00207543.2015.1112046.
- Tukey, John W. (1977): *Exploratory data analysis*: Reading, Mass (2).
- van Meldert, Bram; Boeck, Liesje de (2016): Introducing autonomous vehicles in logistics: a review from a broad perspective. In *FEB Research Report KBI_1618*.
- Xiao, Yi; Xiao, Jin; Wang, Shouyang (2012): A hybrid model for time series forecasting. In *HSM* 31 (2), pp. 133–143. DOI: 10.3233/HSM-2012-0763.

AUTHOR(S) BIO NOTE(S)

Christoph Martius works as a research scientist at Fraunhofer Center for Maritime Logistics and Services CML in Hamburg, Germany, where he is engaged in applied research and consulting projects in maritime logistics and shipping. A focus of his work is the design and implementation of novel machine learning approaches across the maritime supply chain and the analysis of big data. His own research interest is the application of deep learning approaches to use cases in various domains ranging from biology and chemistry to logistics and engineering.

Lutz Kretschmann works as a senior scientist at Fraunhofer Center for Maritime Logistics and Services CML in Hamburg, Germany, where he leads the research team “Marine Operations Management”. A focus of his work is the digital transformation of the maritime industry, new data driven business models, and the development of data- and algorithm-based decision support systems for shipping and maritime logistics. His own research interests lie in the field of machine learning and specifically applying deep learning to current problems in the maritime industry.

Carlos Jahn is head of the Institute of Maritime Logistics at Hamburg University of Technology as well as of the Fraunhofer Center for Maritime Logistics and Services in Hamburg. At the beginning of his professional career Carlos Jahn served as a sailor and later on as an officer in the German Navy. He studied mechanical engineering at the Helmut Schmidt University in Hamburg and economics at the University of Hagen. As a research assistant at the Fraunhofer IFF Carlos Jahn obtained a doctorate at the Otto von Guericke University of Magdeburg. Carlos Jahn held various management and staff functions in science and industry. In 2009, he accepted the offer of a full professorship at Hamburg University of Technology and was appointed head of the Institute of Maritime Logistics. The Fraunhofer Center for Maritime Logistics and Services was founded in 2010.

Ole John leads the department Ship & Information Management at the Fraunhofer Center for Maritime Logistics and Services CML in Hamburg. In his research he focusses on human resources, ship operation as well as maritime logistics. During his career as a Navy Officer he gained experience as navigational officer on miscellaneous types of ships. Mr. John has a Diploma in Education from the Helmut-Schmidt-University in Hamburg as well a Master of Business Administration from the University of the Federal Armed Forces in Munich in collaboration with the ESB Business School Reutlingen. During his time as a research assistant Ole John received his doctorate at the at the Hamburg University of Technology

With the Sponsorship of



With the Support of

