informs
http://pubsonline.informs.org/journal/inte

INFORMS JOURNAL ON APPLIED ANALYTICS
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0092-2102, EISSN 1526-551X

Submitted to *INFORMS Journal on Applied Analytics*

Passenger Ferry Operations in the Digital Era: Forecasting and Revenue Management at Molslinjen

Pierre Pinson
Halfspace, Borgergade 28, 1300 Copenhagen, Denmark, pp@halfspace.io
Imperial College London, Dyson School of Design Engineering, Imperial College Road, London SW7 2AZ, United Kingdom
Technical University of Denmark, Department of Technology, Management and Economics, Akademivej, 2800 Kgs. Lyngby, Denmark

Mikkel Bjørn, Simon Kristiansen, Claus B. Nielsen
Halfspace, Borgergade 28, 1300 Copenhagen, Denmark, {mb@halfspace.io, sk@halfspace.io, cbn@halfspace.io}

Lasse Janerka, Jesper Skovgaard, Kristian Durhuus
Molslinjen, 4 Hveensgade, 8000 Aarhus, Denmark, {lja@molslinjen.dk,jesk@molslinjen.dk,krd@molslinjen.dk}

Abstract. Molslinjen, one of the world's largest operators of fast-moving catamaran ferries, based in Denmark, adopted a strategical focus on digitalization to profoundly change their operations and business practice. They partnered with Halfspace, a data, analytics and AI company based in Copenhagen, Denmark, to support that transition. Halfspace and Molslinjen have jointly developed and deployed a successful forecasting and revenue management toolbox for the data-driven operation of ferries in Denmark, rolled out operationally since 2020. This has resulted in $2.6-3.2 million yearly savings, significant reduction in number of delayed departures and average delays, and a 3% reduction in fuel costs and emissions. This toolbox relies on some of the latest advances in machine learning for forecasting and in analytics approaches to revenue management. The potential for generalizing to the global ferry industry is significant, with an impact on both revenues and ESG criteria.

Key words: Ferry operations; demand forecasting; revenue management; machine learning; Franz Edelman award

1. Introduction

Transportation has always been a key application area within operations research and management science, as exemplified by the recent review paper of Petropoulos *et al.* (2023). In principle though, when referring to transportation, one mainly thinks of routing problems, assignment problems, scheduling, etc. Most importantly, emphasis is most often placed on road and air transportation, as well as passenger trains and freight, to a lesser extent. Obviously, there are also a lot of maritime applications, e.g., related to optimal network design (Christiansen *et al.*, 2020), bin-packing (Trivella and Pisinger , 2016) and stowage planning (van Twiller *et al.*, 2023), etc., for shipping companies like Maersk, Hapag-Lloyd and the likes. In contrast, the body of work related to optimal management and operations of passenger ferries is much more limited.

In many countries with islands, e.g., Greece and Denmark, and geographical areas easily connected through maritime transport, e.g. Scandinavia and South-East Asia, passenger ferries are deemed a practical and economical solution. They are also an enjoyable travel experience, as supported by numerous traveller and customer satisfaction surveys around the world. The passenger ferry market size was valued between \$12 and \$15 billions at the end of 2023, with a number of passengers similar to that of the airline industry (between 4 and 5 billions passengers per year). That market is expected to grow substantially in the coming decade. Passenger ferry operators are embracing the change towards more sustainable and efficient operations, by rethinking their fleet, their routes and their approach to operations. Some recent and representative examples are the development of models to predict fuel consumption based on contextual variables, the control of autonomous ships, as well as online booking systems, among others.

Operating ferries involves many decision-making tasks, including scheduling and routing, vehicle packing (also referred to as loading), staff planning, procurement, etc. These tasks are traditionally and still commonly addressed based on experience, expert knowledge, and pragmatic decision-making. However, with the increasing availability of data and the process of digitalization in many areas of our industries and societies, passenger ferry operators also have the opportunity to rethink their operational practice. The pace has been slower compared to other industries, due to its specifics (very localised, highly regulated, etc.) – for an overview, see Wergeland (2013). Molslinjen, a leading passenger ferry operator based in Denmark, decided to engage in that digitalization process in the late 2010s, by partnering with Halfspace, a Danish-based company focusing on data, analytics and AI solutions. The central objective of this paper is to introduce and thoroughly describe the development and result of this partnership, which consists in practical analytics solutions, e.g., for forecasting passenger demand (number and types) as input to packing and other decision-making processes, as well as revenue management and dynamic pricing. Beyond these tools, analytics and AI-based solutions, this partnership has had a broader and profound impact on Molslinjen, which has now become a front-runner in the digitalization of the passenger ferry business worldwide. Such an impact is on revenue, ESG (Environmental, Social and Governance) criteria like emissions, as well as change management in the organisation.

The article is structured as following. The operational practice and challenges of Molslinjen are described in Section 2. We there insist on the specifics of the ferry departure management (demand estimation, packing, etc.) that motivated the move towards digitalization of their operations. This also gives us the opportunity to explain how Halfspace and Molslinjen jointly engaged in the development of optimal packing solutions, of a forecasting engine, and of a revenue management toolbox. Subsequently, emphasis is placed in Section 3 on the forecasting platform developed to predict passenger demand, both in terms of number and types of customers. To improve readability, both methods and results are covered within that section.

Section 4 presents the revenue management and dynamic pricing platform, including methodological components and example results. Based on these bespoke solutions, their development and deployment at Molslinjen, we review and discuss their impact in Section 5, by relying both on a quantitative framework and on statements from the main stakeholders, at both management and operational levels. Finally, the paper ends with Section 6, by gathering conclusions and our outlook for further developments.

2. Digitalizing passenger ferry operations

To set the scene, we introduce in the following the role of Molslinjen as a passenger ferry operator in Denmark, their operational challenges, as well as the way they engaged with digitalization of their operations over the last 5 years. Emphasis is placed on describing the status quo prior to the developments covered in the remainder of the paper. We also present some of the specifics of vehicle packing for passenger ferries.

2.1. Passenger ferries in the Danish context

Molslinjen (www.molslinjen.com) is one of the world's largest operators of fast-moving catamaran ferries, serving 9 routes in Denmark, with the main route crossing Kattegat to connect Western and Eastern Denmark. These ferries can transport up to 1200 passengers and 400 vehicles, at speeds of up to 70 km/h when fully loaded. Denmark is a relatively small country in Scandinavia with many islands – around 1400 islands in total, and 70 of them being populated. Ferries comprise the most convenient approach to move people and goods between mainland and islands, or between islands. Molslinjen is to provide this transportation service on a commercial basis, though under regulatory constraints, while also aiming to reduce their environmental footprint[1].

The strategic partnership with Halfspace was launched in 2019, initially with a focus on optimizing the loading plans for the ferries, at each and every departure, to drive efficient operations. It rapidly appeared that building a state-of-the-art forecasting engine was necessary to provide input to the packing optimization algorithm. Before that, all forecasts and operational decisions were based on expert knowledge and heuristics only. The basic quantities to be predicted consisted in both the expected vehicle count and the mix of vehicles (bikes, cars, trucks, etc.). Such forecasts were to be produced from up to a year in advance (of a given ferry departure), and then updated at regular intervals up to the actual packing of the ferry at departure time. From 2021, when Molslinjen was acquired by EQT Partners, a bespoke revenue management system was added with focus on dynamic pricing. The aim of the revenue management approach was to bring additional flexibility in the pricing strategy and to better accommodate the various categories of clients (i.e., business, standard and low-fare customers).

[1] Electric ferries: https://www.molslinjen.dk/kontakt/presse/pressemeddelelser/25082022-to-vundne-udbud-sikrer-groenne-faergeruter (in Danish)
Investing in green technology https://www.molslinjen.dk/kontakt/presse/pressemeddelelser/14112022-molslinjen-investerer-i-banebrydende-teknologi (in Danish)

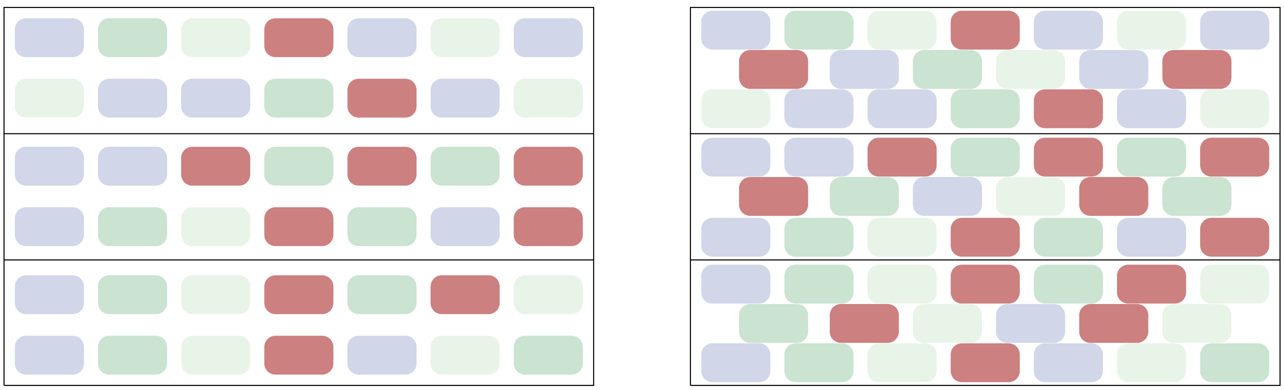
The initial contact between Molslinjen and Halfspace happened after the launch of a new website and of a dedicated app for bookings. This starting point with digitalization was focused solely on UX/UI, and not yet on operations. Both CCO and CEO of Molslinjen at the time came from the airline industry (Scandinavian Airlines – SAS). They had witnessed how revenue management had changed the industry in the 1990s. The CCO set the vision to make good use of all data being collected and to make Molslinjen a digital champion in the passenger ferry industry. This was the start of a fruitful collaboration between Molslinjen and Halfspace, with the subsequent development and deployment of an optimal packing algorithm, a forecasting engine, and a revenue management platform.

2.2. Operational challenges in passenger ferry operations

Obviously, a passenger ferry company needs to make strategic planning decisions in terms of the routes to serve, the ferries to be built or purchased, the timing for departures, etc. These are all planning decisions that are crucial to their business, and to be made under deep uncertainty – though, these are not the type of decisions we focus on here. We concentrate on operational decisions instead. These decisions mainly relate to the demand (in terms of number of passengers and of vehicles) to be accommodated under capacity constraints. The demand varies substantially throughout the year, with seasonality effects, special-event effects (e.g., holidays), weather effects, etc. Such variations are in terms of magnitude, i.e., the overall number of passengers and vehicles, but also in terms of the type of vehicles. Some departures involve many vans and trucks, while other departures may involve small cars only. The reason why it is so important is because all these vehicles are to be packed in the cargo area. Packing these vehicles rapidly and in an efficient manner might be the most significant, and recurrent, operational challenge faced by a passenger ferry operator like Molslinjen. It is like playing a life-size game of Tetris under time pressure.

Passenger ferries are continuously operating during day time (with up to 20-30 departures per day for the main routes), every day of the year. This hence translates to a schedule with back-to-back departures, and with very limited time between arrival at the harbour and the next departure – typically between 20 and 30 minutes. During that period, all passengers have to disembark and all the vehicles have to leave the ferry. After that, all incoming vehicles have to drive onto the cargo area of the ferry and be packed following the chosen packing plan, while passengers and pedestrians board the ferry. If a departure is delayed, this delay will most likely affect all subsequent departures that day. This will force the captain sailing the ferry to try to catch up, by increasing speed and therefore fuel consumption and related emissions. The consequences are both economical (higher operating costs due to increased fuel consumption) and environmental (since increased fuel consumption yields additional emissions).

Let us describe the packing problem in more details. Passenger ferries are built for a certain capacity, both in terms of passengers and in terms of vehicles onboard. However, in reality, the vehicle capacity is not in terms of number of vehicles, but in terms of the space they can use in the cargo area. The vehicle deck is divided by two rows of pillars, leaving three row section to pack the vehicles. Since this area is divided into lanes, the vehicle capacity is then expressed in *lane meters*. When the ferry is not close to capacity, vehicles are packed besides each other, in a "2-by-2" configuration (see Figure 1a). If there are many vehicles to handle though, it may be chosen to pack the vehicles as tight as possible. This so-called "zipper" configuration, while allowing to have more vehicles onboard, requires more time for packing, since all vehicles have to be guided patiently and precisely to end up a few centimeters from each other. A fully loaded ferry uses the zipper approach on all sections in the ferry (see Figure 1b). There are additional considerations related to the spatial weight distribution, which may affect the balance of the ferry and fuel consumption.



(a) 2-by-2 configuration

(b) zipper configuration

Figure 1 Alternative packing strategies for vehicles in the cargo area. The zipper configuration allows for nearly 50% more vehicles to be packed in the cargo area of the passenger ferries.

2.3. The changes brought in by digitalization

Before to engage in that digitalization process, all operations were based on expert knowledge and simple, yet proven, heuristics. For instance, the estimation of upcoming vehicles for the next target departure would use a rough estimate from historical statistics, with some expert-based adjustments. And, the packing process relied on the expert knowledge of the staff members responsible for packing. This required a lot of radio-based communication during the process, while also leading to on-the-fly corrections. Vehicles may have been packed too tight, the spatial weight distribution may have needed to be adjusted, etc. An additional aspect is to ensure that heavy vehicles are packed in the front and middle parts of the cargo area, in order to minimize additional fuel consumption and to have the adjusted weight across the loading space more divided. All these aspects together led to potential time delays, additional stress for the staff members involved, and in extreme cases to decline boarding for some of the vehicles. The quality of service and customer experience was then affected.

The first step with the digitalization of the packing process was for the staff to be able to visualize in advance the portfolio of vehicles that were expected, as well as the alternative options for the packing plan. They eventually got recommendations for optimal packing for each and every departure, in due time. They would then know whether the ferry would have to be packed following a plan that is loose (requiring adequate spatial weight distribution), normal, tight, or extremely tight (the last two using a zipper configuration in one or more lanes). A side benefit of the tool was also an increased consistency in the way the packing was done, since based on recommendations instead of expert knowledge only. This resulted in higher quality of service and increased customer satisfaction.

The optimal packing algorithm and related tools (e.g., including visualisations and recommendations) was the first solution developed. This was followed by the development of the forecasting engine, operational since 2020, and of the revenue management system, rolled out in 2022.

3. Forecasting for bookings and arrivals

One of the key challenges for Molslinjen is to predict demand. Demand is in terms of *(i)* number of passengers, and of *(ii)* number of vehicles of different types. Forecasts are needed at different time scales and to be used as input to different type of decision problems, from planning and staffing to packing of vehicles in the cargo area. Besides these operational problems, forecasts are actually also used for marketing and sales in order to manage advertising campaigns in a timely and efficient manner.

We first start by describing the forecasting setup, as well as the data and the pipeline used to produce operational forecasts. We then dig into the forecasting methodology, as well as aspects related to estimation, re-estimation and forecast verification. Finally, we give an overview of the outcome, both in terms of the capabilities of the forecasting platform and in terms of actual forecast quality.

3.1. Decision-making context

Molslinjen has to make decisions at different time scales and for different purposes. Since we restrict ourselves to operational problems only, all decision-making processess are to be understood as conditional to a fixed set of ferries and open routes in the area served by Molslinjen.

The forecasting and decision-making context for passenger ferry operations is different from what one may be used to for other applications. Typically, we see forecasts as being issued iteratively, with the forecast horizon sliding accordingly. Think of the example of weather forecasts. A forecast is issued on a given day at a given time (say, 1st March, at 6.00am) with hourly resolution for the following 72 hours. And then, at the next forecast update, the following day for instance (so, the 2nd of March at 6.00am), the forecast window is still of 72 hours, but has slid by 24 hours. In contrast, here, the approach to forecast is focused on a given event of interest (i.e., a given ferry departure on a date date and at a given time – say, the 15th of April at 9.30am), which is fixed. And then, one aims to predict its characteristics at different points in time prior to that event.

In the present case, these points in time are *(i)* more than 1 month (and up to a year) in advance , *(ii)* 1 month in advance, *(iii)* 7 days in advance, and *(iv)* 1 hour in advance. Note that these are representative times only, since the forecasts are continuously updated along that time, and until the last-minute before ferry departure.

The first longer-term predictions (more than 1 month in advance) are there to give an initial estimate of a departure's demand based purely on historical data. Such historical data include numbers of passengers and vehicles, for different types, for each and every ferry departure. This initial forecast may be seen as a climatological forecast since based on long-term statistics only. Their main purpose is to inform the process of planning the number of daily departures on a given route (e.g., between 12 and 14 on the link Odden-Aarhus, the main route operated by Molslinjen).

In parallel, the iteratively-updated predictions (from 1 month in advance to a few hours in advance) inform both operational and revenue-focused decision processes. The main driver for these iterative updates is the continuous recording of reservations (and of their specifics). Firstly, it serves as an early warning system for departures performing unusually, potentially weeks in advance based on reservation patterns. As a general indication for the timeline of reservations, 2023 saw 12% of reservations being made 30 days in advance, rising to 46% one week before, with 7% of people making their reservation within the final hour prior to departure.

During that period prior to target departure, example decisions that have to be made, based on these forecasts, primarily include:

Staffing– Number of crew onboard has to be decided, such that the projected passenger capacity can be safely transported. Changing the planned crew within the final week of a departure can be especially costly;

Pricing– A pricing strategy has been planned in the initial phase, but over/under performing departures need tuning to be optimal for the business;

Capacity reservation– Ensuring reserved capacity on departures for business-class customers, who do not need to book in advance. If a business-relevant departure, such as early weekday mornings, exhibits increased interest from non-business customers, it has to be controlled to ensure obligations to business customers.

Eventually, the so-called "real-time" predictions primarily inform operational crews. For the final hour prior to departure, the forecasts tell those who are managing the packing of the ferry and the staff onboard about when they can expect to see vehicles arrive, how many might be late to arrive, etc. The one hour mark is especially important for decision-making, as this is when the operational manager has to decide on the ferry packing strategy. Since very few vehicles are expected to be there one hour prior the departure, the decision is fully based on the forecast. In the case that the ferry is in high-demand, extra time has to be spent ensuring every vehicles is loaded as efficiently as possible (following the zipper configuration), and changing this strategy mid-way through is difficult.

Besides the operational manager, the hands-on-deck crew on the dock directing traffic to the ferry team in charge of packing needs a good overview so that they coordinate well. In that respect, business customers are especially important, as we have to adjust our prediction within the final hour, so that the crew loading knows whether or not all business customers have arrived. If we expect business customers to arrive at the last minute, the ferry crew has to make sure that they still can get their benefits, such as arriving only 5 minutes prior departure, and being loaded in a way that allows them to disembark quickly.

An overview of the forecasting philosophy, timeline and types of decisions, is depicted in Figure 2.

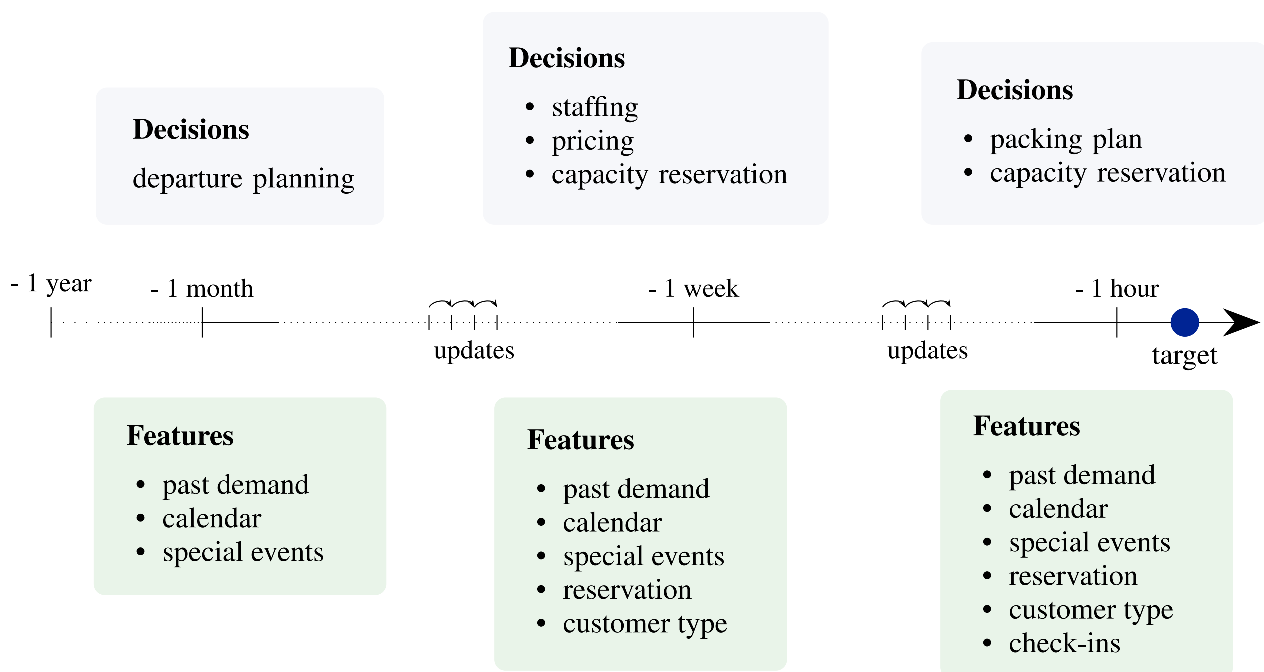
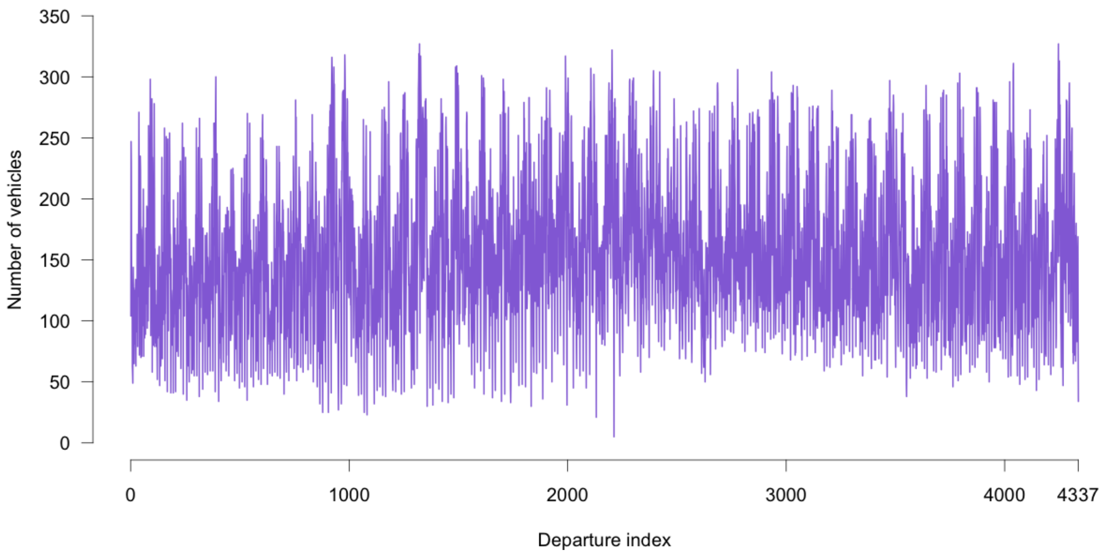


Figure 2 Timeline for the forecasting process, indicating both decisions to be made, as well as features employed at various lead times. The information about past demand and reservations becomes richer when getting closer to target departure.

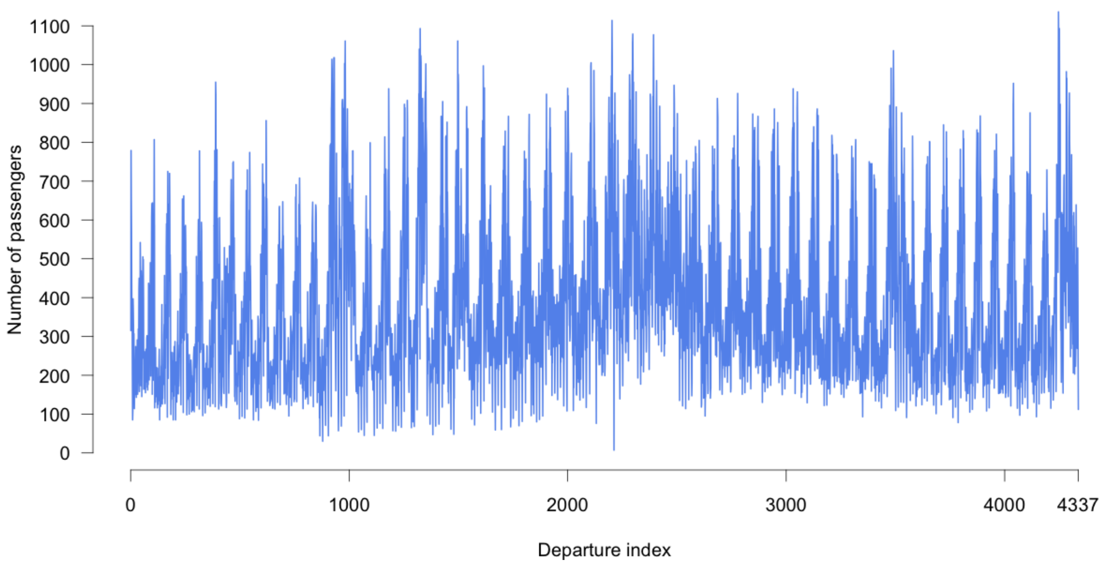
3.2. Data: explanatory and target variables

To illustrate the characteristics of demand for Molslinjen passenger ferries, let us look at the data from 2023. Over that year, forecasts were produced for different routes in Denmark, for a total of 8675 departures. Taking the example of the departure port of Odden in Sjællands Odde (Zealand), Figure 3 depicts the evolution of the number of passengers and vehicles (all types together) over the whole year of 2023. Focusing on that departure port only, there were 4337 departures to predict for (and to handle for the Molslinjen staff involved), i.e., 12 per day on average.

Over that period, the average number of vehicles and passengers per departure were of 154 and 381, respectively. However, the variability around these averages is significant, driven by calendar and seasonal



(a) Vehicles



(b) Passengers

Figure 3 Evolution of the number of passengers and vehicles (all vehicle types together) over all departures of 2023 (4337 of them) with origin Odden, Sjællands Odde in Zealand, Denmark.

effects, holidays, special events, etc. The number of passengers varied from 7 to 1136 per single departure, while the number of vehicles varied from 5 to 327. The variability is more pronounced for the number of passengers than for the number of vehicles. Especially, in Figure 3b, one clearly observes the effect of Easter in Denmark (around departure index 1000), Summer holiday (departure indices between 2200 and 2600), Fall holiday (around departure index 3500) and Christmas holiday (from departure index 4200 onwards). Passenger ferries are very popular among Danes (and possibly tourists) to commute between different parts of the country.

These number of passengers and vehicles are high-level information only, in order to illustrate the temporal evolution of the target variables of interest. However, in practice, these time-series are to be disaggregated if aiming to look at the actual target variables that are modeled and predicted. For the vehicles, there are 9 types of customer groups (8 types of vehicles, as well as pedestrians). And then, for each group, there 5 subgroups of interest: number of reservations, go-shows (arrivals without reservations), no-shows (reservations without arrivals), passengers and cancellations. Obviously these various subgroups do not have the same importance in decision processes. For instance, since business customers do not have have to make reservations, predicting how many will just show up is crucial. Similarly, predicting the number of reservations is key, since limiting the number of available spots for those who will come without a reservation.

The forecasting engine, to be further described in the next section, relies on supervised learning, to learn nonlinear relationships between a set of explanatory variables (also referred to as input features) and the target variables mentioned in the above. Some of these features are classical ones to consider for calendar effects, e.g., time of year, day type (normal, vacation, etc.), dummy variables for special events. Others inform about past demand in similar conditions (analogs). Then, additional input features concentrate on the route specifics, e.g., route ID, departure origin, vehicle group, etc. Additional features could be considered, for instance related to the weather and to the status of other transport infrastructure in Denmark (e.g., bridges). However, it was either found that they were not having a substantial effect (for the weather, except for extreme events, which obviously affect the willingness of people to use ferries), or were more difficult to obtain (like the predicted and real-time usage of Danish bridges).

The way the different explanatory variables are used as input to forecasting, at different lead times, is additionally illustrated in Figure 2.

3.3. Forecasting engine

At the core of the forecasting engine is a regression approach relying on eXtreme Gradient Boosting (XGBoost). XGBoost was originally proposed and described by Chen and Guestrin (2016), and became extremely popular since then. In addition, the use of such an approach and underlying algorithm regularly yielded top performance in data science and forecast competitions (Chen and Guestrin , 2016; Bojer and Meldgaard , 2021). The code and related material was openly released on Github (github.com/dmlc/xgboost), while it is now available as packages in Python, R, etc.

The basic idea of XGBoost is that it relies on, and generalizes, Gradient Boosted Trees (GBTs). In turn, the concept of gradient boosting was proposed 15 years earlier (in 2001), as a basis to obtain universal function approximators (Friedman , 2001). This approach is then tailored to the case of regression trees. This yields a tree ensemble, to be seen as an ensemble of weak learners, for which the final prediction is the sum of the predictions from the individual trees. A gentle introduction to XGBoost is made available in the form of a tutorial at xgboost.readthedocs.io/en/stable/tutorials/model.html.

Based on the decomposition described in Section 3.2, we end up with 135 models to train in parallel. As for other supervised learning approaches, XGBoost models can be readily fitted based on training data consisting of a set of past examples consisting of pairs of input features and corresponding target variable values. However, XGBoost has many hyper-parameters to decide upon. Based on our empirical investigation with historical data, as well as expert knowledge, the 5 hyper-parameters deemed most important are selected based on k-fold cross validation (with $k = 5$). These 5 hyper-parameters, from the most important to the least one, include:

- the number of trees that compose the ensembles (or, equivalently, rounds of boosting),
- the maximum depth of the trees,
- the learning rate,
- the L_1 regularization parameter,
- the subsampling ratio for each feature when constructing the trees.

For both training and cross-validation, the criterion to be minimized is a quadratic criterion, consistent with the idea of predicting the conditional expectation of demand for the departures of interest. As a wrapper to these training and cross-validation exercises, we also test the alternative models, with various sets of competitive hyper-parameters, over a dataset with 5 years of past data, for which 90% of the data is used for training and 10% for testing. Finally, to compare the updated and retrained model with that in production, we implement a form of A/B testing based on the last month of data, in order to assess the potential improvements brought in by the retrained and updated models. This last month of data is left aside for this final step only, and never used for any model selection or training. The retraining of the models and updating of the hyper-parameters is performed on a quarterly basis, unless special events and identified concept drifts (like with COVID-19) motivate dedicated retraining and hyper-parameter tuning.

When used for forecasting, the models issue forecasts that are updated every time new information becomes available. For instance, every time a new reservation is made and recorded, all models update their forecasts for the departure concerned.

From an operational point of view, all the analytics runs on the cloud (i.e., on Molslinjen cloud system), with a user interface (web-based / Power BI) that allows to visualize relevant data, outputs and metrics. Finally, the forecasts are pushed downstream into their proprietary systems, for further decision-making related to operations.

3.4. Forecast verification

Forecasts have been issued since 2020, both offline for model selection and parameter tuning, as well as operationally for Molslinjen. Based on the timeline depicted in Figure 2, from 1 year before the specific departure of interest and until the actual departure, there are specific lead times that are singled out to assess the quality of the forecasts. The quality of the forecasts is to be understood as the objective ability to inform about future events (e.g., assessed based on scores and diagnostic tools). In contrast, the value of the forecasts (to be discussed in a further section), is about the additional benefits resulting for the use of these forecasts in further decision-making. The lead times that are singled out include 6 months ahead, 1 month ahead, 1 week ahead, and 1 hour ahead. Note that the forecasts for lead times between 6 months and 1 year ahead are the same, since relying on long-terms statistics only (or, relying on the same departure the year before, for the baseline models). In terms of forecast verification, we follow best-practice principles for the evaluation of our predictions, as for instance discussed by Bergmeir (2023).

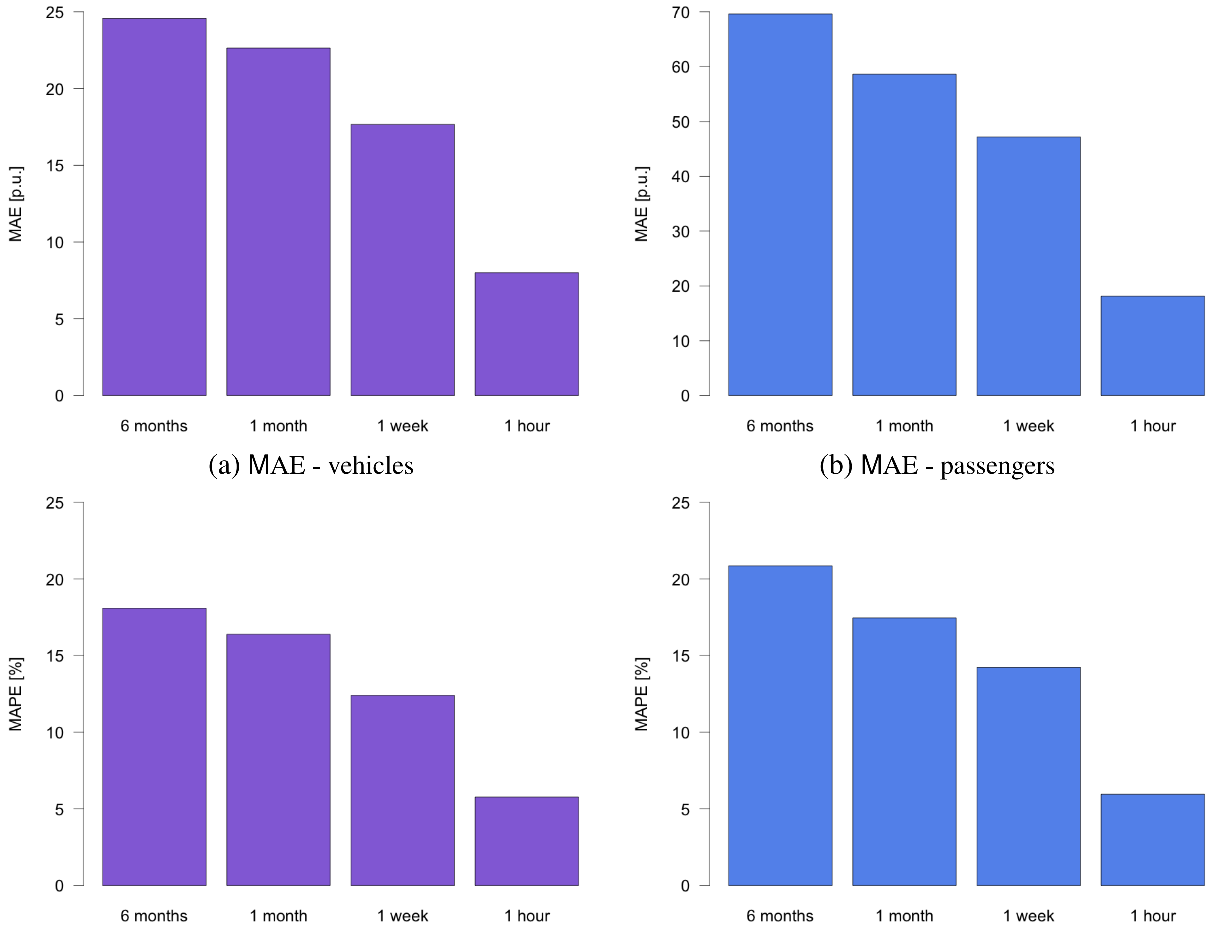
The two scores that were agreed upon (between Molslinjen and Halfspace) to assess the quality of the forecasts are the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). These are very common scores to assess forecast quality, as for instance recently reviewed by Petropoulos *et al.* (2022). We acknowledge that this is not fully aligned with the approach to training of the models (relying on a quadratic loss function). However, in practice, it was observed that using a quadratic loss led to more stable parameter estimates and results in terms of forecast quality. For a given departure at day and time t and lead time k, forecasts are denoted by $\hat{y}_{t|t-k}$ and corresponding observation by y_t. We have an overall number T of departures to consider over 2023 ($T = 8675$). The 2 scores, for a given lead time k are then defined as

$$\text{MAE}_k = \frac{1}{T}\sum_{t=1}^{T} \left| y_t - \hat{y}_{t|t-k} \right| , \tag{1a}$$

$$\text{MAPE}_k = \frac{1}{T}\sum_{t=1}^{T} \frac{\left| y_t - \hat{y}_{t|t-k} \right|}{y_t} . \tag{1b}$$

The interest of these two metrics, for instance compared to the Root Mean Square Error (RMSE, based on a quadratic loss function), is their pragmatic interpretability. MAE can be interpreted as the average deviation, in absolute value, between forecasts and observations. It then allows formulating statements of the type ”Forecasts inform about the number of vehicles, ± 10 units”, for the example of a MAE of 10 units. In parallel, MAPE tell about the average difference, in terms of percentage points, between forecasts and observations. The type of statement that can formulated is ”Forecasts inform about the number of vehicles, ± 5% units”, for the example of a MAPE of 5%.

The forecast quality metrics are depicted in Figure 4, for the various lead times and for the main overall categories of interest (i.e., number of vehicles and number of passengers). As expected, forecast quality consistently improves as we get closer to actual departure. MAE and MAPE are negatively-oriented scores – the lower, the better. Especially, from the 1-week ahead lead time onwards, a lot of new information is made available through online reservations, allowing to substantially improve forecasts. At the crucial lead time of 1 hour between departure, when a decision has to be made about how to optimally pack vehicles in the cargo area, the average deviation between forecasts and actual number of vehicles is of 8 vehicles only. In parallel, in terms of passengers, that deviation is of 18 passengers on average. If thinking is terms of percentage deviation, by using the MAPE score, that translates to a 6% deviation for both of them.



(a) MAE - vehicles (b) MAE - passengers

(c) MAPE - vehicles (d) MAPE - passengers

Figure 4 Overall forecast quality assessment in terms of both MAE and MAPE, for forecasts of number of vehicles (all types together) and number of passengers. Forecast quality metrics are calculated for representative lead times, i.e., 6-month ahead, 1-month ahead, 1 week ahead and 1-hour ahead.

The quality of these forecasts can be readily compared to that of relevant baselines. The baselines are consistent with what the operational practice, before to develop and deploy the bespoke forecasting engine. Here, the baseline typically is the corresponding departure (i.e., on similar day and time) experienced a period of time before the target departure. For instance, considering the departure on Thursday, the 12$^{\text{th}}$ of January 2023, say at 9.00, the first baseline approach to forecast the number of passengers and vehicles is to use the data a year before, on the closest similar day and time (here, that would be Thursday, the 13$^{\text{th}}$ of January 2022 at 9.00). The baseline produced are with lead times of 1 year, 2 weeks and 1 week. Their quality, in terms of both MAE and MAPE, are collated in Table 1. Whatever the lead time of interest, the improvement offered by our approach compared to such baseline forecasts is substantial. Taking the specific example of the 1-week ahead lead time, the MAE for the forecasts for vehicles and passengers is reduced by more than 50%.

Table 1 Performance of baseline approaches based on past similar departures, in terms of MAE and MAPE, with lead times of 1 week, 2 weeks and 1 year.

	Vehicles		Passengers	
Lead time	MAE [p.u.]	MAPE [%]	MAE [p.u.]	MAPE [%]
1 year	36.4	26	97.5	28.4
2 week	32.7	23.3	104.3	30.4
1 week	30.3	21.9	95.22	28.3

In practice, the forecast results are also assessed on a per category basis, e.g., for small and large cars, motorbikes, pedestrians, etc. The forecast quality for these different categories may vary widely, owing to their potential intermittency.

4. Revenue Management through advanced analytics

Revenue management (RM) generally relies on dynamic pricing in one way or another. Variants exist depending on demand modeling, the choice of contextual variables, as well as overall objectives. For a gentle introduction to dynamic pricing, the reader is referred to den Boer (2015). Such revenue management approaches were successfully implemented by the airline industry decades ago. In contrast, passenger ferry operators have only recently opened up to that possibility. Molslinjen and Halfspace have been at the forefront of that evolution, by developing and bringing to operation the first and leading dynamic pricing system for ferries in Denmark. In this section, we start by describing the operational context for the revenue management approach, followed by an extensive description of the methodology. We then gather example results to highlight its workings and related benefits.

4.1. Operational context

The revenue management system was first introduced on Molslinjen's main route over Kattegat, between Odden (Zealand) and Aarhus (Jutland). On that route, Molslinjen operates with three separate fare classes:

Business (referred to as ”Blue class”) – Tickets are sold at a fixed price of 925 DKK ($135). They include guaranteed access to the ferry of their choice without booking, lounge access, and other benefits (e.g., priority in disembarking);

Standard – Tickets are sold at a fixed price of 749 DKK ($109). They provide access to the specific ferry for which the tickets are purchased (with limited additional benefits);

LowFare – Tickets sold at lower prices but have restrictions on cancellation and rebooking. The price can span from 249 DKK to 699 DKK (equivalent $36 and $102, respectively), depending on the day and time of the departure itself.

There are additional conditions for both Standard and LowFare tickets related to re-booking and refunds, as would be expected for such tiered ticket categories. In simple terms, Standard tickets can be modified for free and refundable for a fee, while LowFare tickets are not refundable and can be modified for a fee.

For Molslinjen, the dynamic pricing problem consequently combines two sub-problems. Firstly, for each individual departure, and at any given time to departure, one should decide about how many tickets of each type are made available to customers. In addition, for the specific case of LowFare tickets, for which the price can vary between a lower and an upper bound, one should determine the price to be dynamically updated on a regular basis at any time before departure and based on relevant contextual information.

Prior to developing the current revenue management approach, pricing relied on expert-based methods. Each departure was assigned a pricing template that included a starting price, and a number of fixed rules for when to increase LowFare prices, or to close LowFare sales, depending on the time to departure and on the number of reservations. The price templates were chosen in bulk when planning departure plans, with adjustments based on the previous year's starting prices and subsequent performance.

The overall dynamic pricing problem of Molslinjen shares similarities with that faced by airlines, in which algorithmic revenue management has been commonplace since the 1980's (Talluri and Ryzin , 2004) and where they drove significant revenue gains compared to earlier manual approaches. The approach to dynamic pricing in the airline industry has continuously evolved since then, by adapting to changes in the type of offering and to the competitive context, see e.g. Fiig *et al.* (2018). Similar approaches and techniques have been applied in a number of other industries, including in cases that have been recognized as Edelmann laureates, e.g., in senior housing (Kuyumcu *et al.*, 2018) and car rentals (Guillen *et al.*, 2019).

However, the pricing problem for Molslinjen does differ from the standard airline revenue management problem. On the positive side, dynamic pricing if for a single route (or leg), without any need to account for network effects and many origin-destination possibilities in the optimization. On the negative side, capacity handling is significantly harder in a ferry setting than for airlines, and it is *not* possible to apply standard RM systems out-of-the-box. The reason is that capacity is not measured or sold in integer units of a "ticket", a "seat" or a "room". As described in Section 2.2, ferries have a given surface area available for vehicles, while different vehicles take up significantly different amounts of space. This is *essential* to handle in detail in a ferry-oriented RM system, and must be accounted for before standard techniques could be applied. These idiosyncrasies, along with an already existing forecasting engine tightly integrated with Molslinjen's existing ticketing system, motivated building a bespoke revenue management system.

From a practical implementation point of view, developing a bespoke revenue management system had the advantage that it could easily be integrated into Molslinjen existing system. The revenue management solution could be implemented in Azure Databricks without affecting Molslinjen's existing ticketing system. This kept the development and deployment costs low compared to alternative approaches. Data management within Molslinjen's system was also accommodated within the existing legacy system and a cloud-based data platform via APIs (Application Programming Interfaces).

4.2. Methodology overview

The revenue management engine is composed of three main modules:

Demand model – This module is specifically tailored to estimate demand across the three fare classes at different times before departure, conditioned on availability and price points for each ticket class;

Capacity management – This module provides the expected available remaining tickets available for sale for a given departure, given existing reservations and expected no-shows, and taking into account that reserved capacity is measured in *lane meters*, and must be converted into an equivalent expected number of available tickets, before RM techniques can be applied;

Price optimization – This module sets availability and price points for each departure, given inputs on expected demands and available capacity. The optimisation is made using the EMSRb-MR approach (Fiig *et al.*, 2009).

Each of these components is described in detail in the following sections. The overall data and workflow of the revenue management system is depicted in Figure 5.

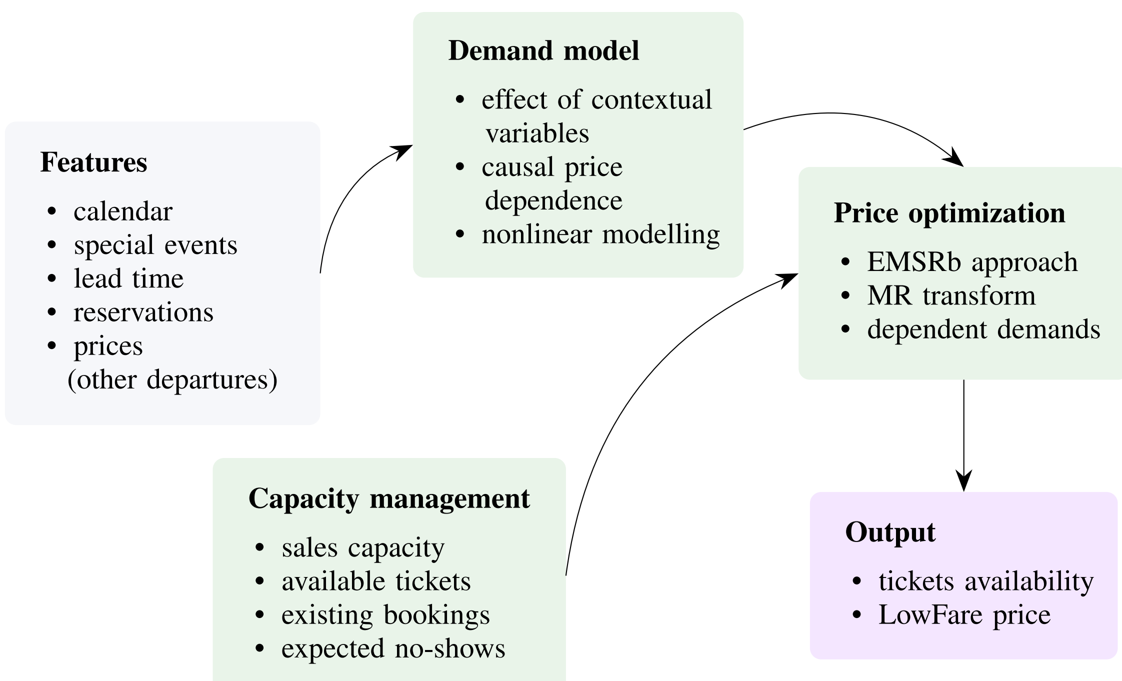


Figure 5 Data and workflow of the pricing system. The process is executed for each individual departure, on a daily basis over the 2 weeks prior to departure, and with less frequent updates up to 12 weeks before departure.

4.2.1. The price-dependent demand model

Any approach to data-driven revenue management is dependent on a forecast of demand for each fare class, should that fare class be offered for sale. This is very similar to the type of (causal) forecasting for pricing performed by online retail companies, e.g., at Zalando

(Schultz *et al.*, 2024). In the case where demand for each fare class cannot be assumed to be independent, it is essential to model the dependencies and take them into account in price optimisation. This is the case for Molslinjen, where the key pricing decision is to choose one among the many potential fares for the LowFare ticket class. Therefore, it was necessary to develop a model for price sensitivity in the customer segment buying LowFare tickets. This required an extension to the forecast engine already developed and presented in Section 3.

In practice, the demand modelling problem was split into two parts. The demand for Business tickets is predicted using the existing forecast engine. In contrast, a dedicated model was developed for the Standard and LowFare classes, with a specific focus on modelling *(i)* price elasticity, and *(ii)* the upsell potential from the LowFare class to the Standard class, in the case where no LowFare tickets were available. Overall, the aim is to model and predict $Y = (n_B, n_S, n_L)$, which are the *daily* ticket sales for the Business, Standard and LowFare categories.

Specifically for the LowFare category, the price elasticity of customers buying such tickets depends significantly on the time to departure, with customers being significantly less sensitive to price in the day(s) immediately before departure, than when booking tickets a long time in advance. This was known prior to the modelling work, and informed the decision to approach this as a regression problem, building a model targeting the daily ticket sales in the LowFare and Standard categories. The modeling approach combines parametric modelling of the price dependency and a Bayesian neural-network model, for the other effects that are not price-related.

As a similar approach is used for both Standard and LowFare categories, we use the latter case as a basis to describe our modelling approach hereafter. The model accommodated 2 types of input features $\mathbf{X} = \{\mathbf{X}_p, \mathbf{X}_{-p}\}$, where $\mathbf{X}_p$ are all features that are related to prices, while $\mathbf{X}_{-p}$ are all the contextual variables that are not price-related. It is implicitly assumed that the features $\mathbf{X}_p$ induce the price-related causal impact on sales, while the other covariates $\mathbf{X}_{-p}$ control for other effects. The price-related features $\mathbf{X}_p$, consist of p_L, the price of LowFare ticket on that day (time-averaged price used in case price changed), as well as an indicator variable I_L for whether LowFare tickets were available for sale or not. Other features included as control variables in the model, $\mathbf{X}_{-p}$, include all the inputs described for the forecast engine in Section 3, the average price ratio to neighbouring departures, and the number of reservations up to and including the prior day (in the form of a time-series).

We model the LowFare sales n_L as a combination of the causal price-induced effects and of the effects from other contextual variables, i.e.,

$$n_L = f\left(p_L; \boldsymbol{\theta}_L\left(\mathbf{X}_{-p}\right)\right) + \varepsilon\,, \tag{2}$$

where ε is a centred noise with finite variance. By using a quadratic criterion at the fitting stage, what we eventually model and predict is the expected LowFare sales. In the above, the function f is of a well-defined parametric form, with $\boldsymbol{\theta}_L$ the parameter vector that parameterizes f. The parameter vector $\boldsymbol{\theta}_L$ is allowed to depend on our contextual covariates $\mathbf{X}_{-p}$ but not on the price-related features. Different models can be considered for f (e.g., exponential and logistic). Our analysis showed that a logistic function was more realistic, while it is also generally deemed well-suited to model price-dependency of demand in choice models (Talluri and Ryzin , 2004). Consequently, we have $\boldsymbol{\theta}_L$ composed of 3 parameters controlling the shape of the logistic function f (the midpoint, the maximum value and the growth rate). Finally, the dependency of $\boldsymbol{\theta}_L$ on $\mathbf{X}_{-p}$ is modelled using a Bayesian neural network. The parameters of the price-demand model are also estimated within a Bayesian inference framework.

Combining the flexibility of the Bayesian neural network with the parametric modelling as in (2) has the benefit of ensuring that, while we can flexibly model the non-linear dependence of demand (on, e.g., time-of-day, time-of-year in a simple setup), the resulting model remains interpretable. Indeed, one can appraise and visualise the relationship between price and demand according to the model, e.g., a single, well-defined expected optimal price for all departures, at any time prior to departure. If going for a non-parametric approach instead, there would be a high risk that the demand model would try to flexibly fit the data, resulting in a price demand relationship that would have oscillations, somewhat indicating that for some price ranges the demand would increase with increasing prices. This would go against basic economic intuition.

The optimisation approach described hereafter depends on potential demand for each departure as a whole, not on daily demands only. Therefore, a simulation engine has been implemented that can estimate the aggregate demand for a departure across multiple days, by continuously sampling predictions from the model, and aggregating reservations and corresponding uncertainties. These simulations are used to calculate the necessary input for the EMSRb-MR setup described hereafter, and are set up to take into account a number of guardrails on prices, e.g., minimum requirements on prices very close to departure.

4.2.2. Capacity management in the revenue management system The available capacity of tickets to sell is a fundamental input to all revenue management optimisation algorithms. In the context of ferries, the number of vehicles that can be accommodated to a large degree depends on the size of the vehicles. Here, capacity on the ferry is measured in lane meters, and the number of vehicles each lane on the ferry can accommodate is determined by the average length of the arriving vehicles. The length varies significantly between vehicle types, and even the average length of a personal car can vary with up to 25% between departures.

The dynamic pricing engine developed by Halfspace and Molslinjen handles this in detail. Since when making a booking, customers provide a registration number, it is possible to find the car in the national registry. The *actual* length of the booked vehicle can hence most often be deduced for both current and historical bookings. Therefore, for a given departure, the lane meters required to accommodate space for already-booked tickets is known with high precision. Likewise, the still-available lane meters can be converted into an equivalent number of tickets for sale using the expected average car length in historical data.[2] In the end, the output is the number of available tickets for sale, which is the standard input required to existing, well-developed revenue management optimisation algorithms.

The capacity management is further optimised by operating with a *sales capacity* that is larger than the *physical capacity* of the ferry. This is done to account for expected cancellations and for expected no-shows, i.e., passengers who reserves a ticket but do not show up. The number of cancellations and no-shows are predicted for each departure by the forecasting engine, following principles laid out in Section 3. The sales capacity is then adjusted by accounting for the historical proportion of tickets being sold (for similar departures and conditions), for which physical capacity was not used in practice (since the corresponding vehicle never boarded the ferry).

4.2.3. The price optimization approach The revenue management engine optimises revenue by setting prices for each departure individually, choosing whether LowFare tickets should be offered and at which price points. The chosen optimisation approach is the Expected Marginal Seat Revenue (often abbreviated EMSRb[3]) optimization, employing a Marginal Revenue (MR) transformation to make the approach suitable to pricing the LowFare class, for which many price points are possible. The combined approach is denoted EMSRb-MR hereinafter.

Expected marginal seat revenue optimization (Belobaba , 1987, 1989, 1992) is a classical revenue management heuristic. It solves the revenue optimisation problem for a situation with a number of fare classes, c_i $(i = 1, \dots, m)$, with corresponding prices, p_i, ordered in decreasing order $(p_1 > p_2 > \dots > p_m)$, assuming that the *independent, unconstrained demands* for each fare class are normally distributed $d_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $\forall i$. Independence, in this context, means for instance that the demand for fare class c_1 does not depend on whether tickets for fare class c_2 are available for sale or not. The output from the revenue optimisation problem takes the form of a number of protection levels, denoted π_i. The protection level π_i is the number of tickets that should be protected for sale in fare classes at price p_i or higher. The demand for all fare classes at these prices is also normally distributed with

$$\mu'_i = \sum_j^{i=1} \mu_j\,, \quad \forall i\,, \tag{3a}$$

$$\sigma_i'^2 = \sum_j^{i=1} \sigma_j^2\,, \quad \forall i\,. \tag{3b}$$

[2] In practice, the capacity management for the revenue management system is handled using so-called *units* of standard vehicle sizes, not lane meters directly. The details are left out for brevity, but the overall principles are as described in the text.

[3] The letter “b” is there to express the fact that this approach is the second attempt at expected marginal seat revenue optimization. The first version was referred to as EMSRa. However, as it led to results that were too conservative, the next version EMSRb was pushed forward. It is the approach that is the most popular and employed in practice today.

By defining the demand-weighted prices $\bar{p}_i = \mu_i'^{-1} \sum_i \mu_i p_i$ $(i = 1, \ldots, m)$, the optimal protection levels are given by

$$\pi_i = \mu_i' + \sigma_i' \Phi^{-1}(1 - p_{i+1}/\bar{p}_i), \quad \forall i. \tag{4}$$

For an extensive description and discussion of the EMSRb approach, the reader is referred to the works of Belobaba (Belobaba , 1987, 1989, 1992).

The basic EMSRb optimisation assumes *independent demands*. This was a reasonable assumption in the early days of revenue management in the airline industry, where great care was taken to differentiate fare classes enough to target non-overlapping customer segments (Talluri and Ryzin , 2004). However, in the case of Molslinjen the assumption completely breaks down for the LowFare fare classes which are *only* differentiated by price. This is not unique to the case of passenger ferries and Molslinjen though. This was also tackled in the airline setting since the advent of budget airlines in the 1990s - see e.g. Weatherford and Ratliff (2010) for a review. This work employs a Marginal Revenue (MR) transformation to overcome the issue, following an approach developed by Fiig *et al.* (2009), in which a dependent-demand problem can be translated into an approximately equivalent independent-demand problem, and standard solvers applied. EMSRb is applied after the MR transform, and the combined approach denoted EMSRb-MR.

In line with the standard EMSRb approach, the output of an EMSRb-MR optimization is a number of protection levels for sets of fare classes, in decreasing price price. An example output could be:[4] 25 tickets are reserved for business tickets only, 35 places are reserved for either standard or business tickets, 45 places are reserved for any tickets costing at least 699 DKK incl. the most expensive LowFare ticket, etc. Given these protection levels and the available remaining capacity on the departure, it is trivial to determine which fare classes should be open for sale, or closed. However, it is *essential* to take into account both the actual size of vehicles that have already booked and the expected vehicle size for future bookings when determining the effective remaining available capacity of tickets for sale – and thus when setting the correct price. This is a significant complication compared to existing systems, and motivate the capacity management module described in Section 4.2.2.

In the EMSRb-MR approach, it may be the case that one or more of the cheapest fare classes remain closed, even if the number of bookings do not reach or exceed the ferry capacity. This happens if the expected marginal revenue of opening such a class is negative, because down-selling from more expensive classes is not compensated by the added volume of ticket sales. In practice, this is taken into account in a step prior to EMSRb-MR optimisation, referred to as direct optimisation. It is done by only considering LowFare prices at-or-above the price that would be optimal if no capacity constraint existed (according to the demand model) when considering potential pricing strategies for which to assign capacity in the EMSRb-MR optimisation. A number of additional guardrails are also applied at this step, e.g., to limit the

[4] The exact numerical values given here are purely for illustrative purposes, and not the outcome of a real optimization step.

potential daily price jumps, or to align the potential pricing strategies with business objectives for given departure segments. These are continuously aligned with the revenue management team at Molslinjen.

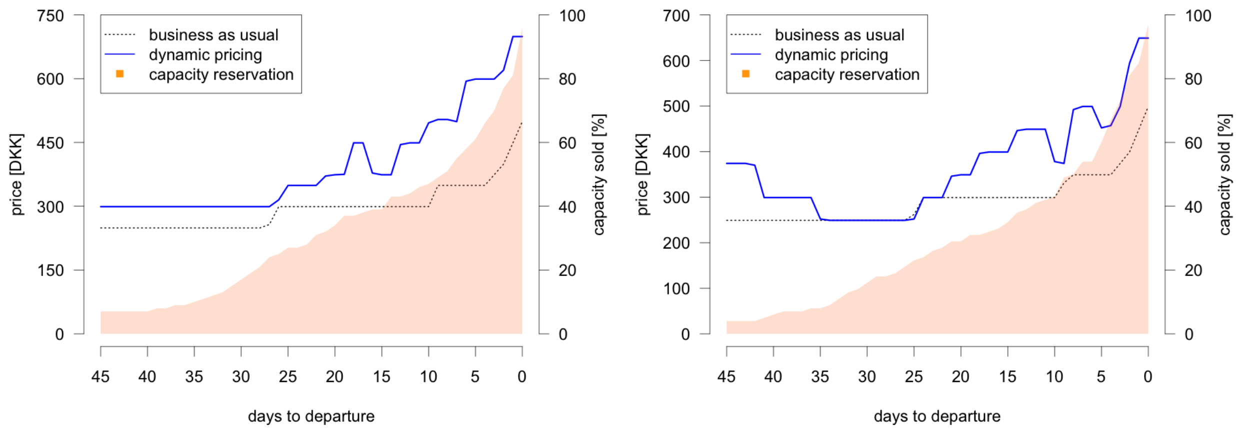
The EMSRb-MR approach was chosen because it provides sufficient performance, while being reasonably simple to implement both in terms of technical complexity, and from a stakeholder-management perspective. This is since EMSRb optimization was well known to a number of key stakeholders within Molslinjen, and because the output in terms of booking limits and protection levels provides some transparency into the algorithmic decisions.

The price optimisation is run on a daily basis for departures within the next two weeks, and at fixed, longer intervals for departures further in the future. The available tickets and prices for a given departure change dynamically within the period between re-optimisations, as new reservations can cause price levels to close down if booking limits are reached.

4.3. Example results

The effect of EMSRb-RM optimisation (and of the whole revenue management approach more generally) can be most easily seen on full departures, where the capacity constraint results in closing the sale of cheaper LowFare tickets. A departure is considered full if more than 90% of the physical capacity was utilized by the arriving vehicles. The initial effect on revenue was estimated by comparing the prices set by the EMSRb-MR system to those of the business-as-usual prices that would have been set by the deterministic pricing templates for such departures.

This effect is depicted for the case of two different target departures in Figure 6. These examples show booking curves for these specific departures, in terms of both price and capacity reservations, from 45 days before departure to the day of departure. The price evolution is compared to the business-as-usual pricing approach, i.e., based on the pricing template that was in use before the roll out of the revenue management toolbox. Following that business-as-usual approach, the price was flat and set to the minimum (249 DKK) for the whole period prior to 4 weeks before departure. Then, for these two examples, the prices were adjusted 10 days before departure, and finally on daily basis over the last 5 days. However, the timing and magnitude of the adjustments were arbitrary, since based on the pricing template. In comparison, the dynamic pricing approach continuously update prices (both upwards and downwards) depending on capacity updates, demand modeling updates, etc. For instance, following a surge in reservations, booking limits determined by EMSRb-RM are hit and the revenue management system closes sale of the cheapest tickets. As a result prices are increased compared to the pricing template. Eventually, the price set through dynamic pricing is consistently superior to the business-as-usual approach and still leading to a full departure. The total revenue gain due price increase from the revenue management system for these 2 specific departures, as compared to the pricing template, are of 22,919 DKK (app. \$3.3k) and 19,220 DKK (app. \$2.8k), respectively.



(a) Departure from Aarhus (Jutland) on 3 September 2023 at 7.15pm

(b) Departure from Odden (Zealand) on 6 October 2023 at 5.05pm

Figure 6 Booking curves, in terms of both prices and capacity reservation, for 2 example departures. The comparison is made with the prices that would have been used if employed the business-as-usual pricing approach. The evolution of prices and capacities goes from 45 days before to actual departure.

During an initial test period spanning July-October 2022, immediately following the launch of the revenue management system, the total revenue increase and the revenue increase for full departures were computed (again, comparing the previous pricing template and the dynamic pricing outcomes). The revenue increased by 7-15% for full departures, and by 1-2 % overall. These number were continuously monitored since then, even after the testing period, with equivalent results achieved during the first three quarters of 2023. During that period, the revenue management system increased prices compared to the baseline on 70% of all full departures.

Assessing the benefits from deploying a new revenue management system always is a subtle task, since the counterfactuals are not available for comparison. Ideally, a well-thought A/B testing setup would have been ideal for the purpose of quantifying the impact with greatest possible confidence. However, it was not feasible from a business perspective. Differentiating pricing approaches, and therefore prices, between customers on a single departure was not deemed acceptable from a business perspective - nor practical to implement. Differentiating pricing approaches between departures would have to be done on a day-by-day basis, or even week-by-week, due to very significant cross-elasticities between departures. The number of days or weekends that would have to be priced with a business-as-usual approach was deemed incommensurate with the objective of rolling out the system fast and realizing revenue gains, especially as launch in July coincided with the high season, where an especially large number of departures are capacity limited. Therefore, the approach to validating input described above was pursued. There is a risk of overestimating the calculated impact, by not taking into account that the volume of sold tickets could have been higher in the business-as-usual setting, in which prices are lower. However, as illustrated in the above, the effect of changing the pricing strategy is mainly visible for full departure, where the impact of the novel dynamic pricing approach is clear. For the other departures, the prices tend to be similar, and the possible impact on demand may be seen as negligible.

5. Impact

The aim of these developments was to have a substantial impact on revenue and ESG criteria, mainly related to emissions. However, impact can also be assessed in terms of changes in operational practice, in terms of staff adoption, and in terms of how this affected customer satisfaction. Therefore, in the following we cover both quantitative and qualitative assessment of impact at Molslinjen.

5.1. Quantitative assessment

Originally, a number of clear criteria were set in order to assess success, from both operational and financial points of view. Obviously, since variations in these criteria may be due to different factors, it was important to clearly state our assumptions and to be rigorous in the way to attribute improvements to the toolbox and solutions developed. On this last point, whenever relevant, we have asked key staff members at Molslinjen to assess how the combination of several factors may have led to the improvements observed, to reach a consensus on the most appropriate and fair attribution. The most important assessment criteria include

- time used for staff planning and ferry packing,
- number of delayed departures and average delays,
- fuel consumption, related costs, and impact on emissions,
- available capacity for vehicles at departure,
- revenues (on full departures, where dynamic pricing is most relevant, and overall).

In practice, some of the relevant inputs were readily available from the operational data. This includes the time used for staff planning and ferry packing, as well as the number of delayed departures and the average delay per departure. Since one may expect to see an effect linked to volumes (i.e., there should be less delays if there are less vehicles to pack at each and every departures), this aspect was accounted for in the comparison. For instance, departures in periods with significant COVID-19 disturbances were removed. In addition, the average numbers of passengers and vehicles were checked over the periods considered to make sure they were comparable. Staff members from Molslinjen reached a consensus such that 70% of the impact (in terms of reducing delays) could be attributed to the forecasting and ferry packing tools.

In parallel, the impact on emissions (denoted ΔE) is directly linked to fuel consumption. It is based on the additional fuel consumption from sailing 1 minute faster (Δf), as well as the average normal fuel consumption per trip ($\bar{f}$), in combination with the decrease in delay (Δd), i.e.

$$\Delta E = (\Delta f \, \Delta d) \, / \, \bar{f}$$

Consequently, the overall cost reduction Δc from less delay was calculated as

$$\Delta c = 0.7\,\Delta f\,\Delta d\,\bar{c}\,n_y$$

where $\bar{c}$ is the average fuel cost and where n_y is the number of yearly departures. The multiplication by 0.7 reflects the attribution factor discussed previously.

Eventually, this led to the result that a the number of delayed departures was decreased by 3.5% and the average departure delay was reduced by 1.5mns. This translates to a 3% reduction in overall fuel consumption, and a fuel cost reduction of 10-12 million DKK per year (across Molslinjen's 8,000+ yearly departures - equivalent to $1.44-1.73 million).

The forecast toolbox has also had a direct impact on revenue: thanks to the efficiency gains in packing, Molslinjen has been able to increase the number of vehicles that can be accomodated on a departure by 6 %. This planned number of vehicles is based on actual capacity of the ferries and their ability to load the ferry without any delay.

Looking at the revenue management toolbox, the sales capacity on full departures has been increased *further* by 5,000 tickets (6%), thanks to the possibility to resell tickets that were expected to turn into cancellations/no-shows. This was calculated based on total expected cancellations/no-shows predicted for full departures in 2022.

The increase in revenue on full departures from the EMSRb-MR approach has been calculated as the difference between actual prices for tickets sold on full departures, and those that would have been with status/quo pricing. The status-quo prices are known, since they follow deterministic pricing templates, and the pricing template relevant for each individual departure is known. Eventually, this led to an increase in revenue of 7-15% on full departures, and a 1-2% increase of overall revenue, compared to business as usual, through dynamic pricing.

Altogether, the analytics toolbox has brought Molslinjen 15-20 million DKK ($2.6-3.2 million) additional profits per year, from a combination of cost savings and increased top line, and a the 3% reduction in fuel consumption and emissions. Since their launches in 2020 and 2022, respectively, the forecasting and dynamic pricing engines have increased profits by a total of $5 million overall. And, as the solutions are currently being improved and added to additional routes, that number is only set to increase in the future.

5.2. Qualitative assessment

Over the last 5 years, Molslinjen and Halfspace have had multiple occasions to reflect on the challenges and successes related to the development of these analytics solutions. Besides the quantitative assessment and underlying indicators discussed in the above, all stakeholders involved have also noticed the positive changes brought in by this process. Some of these elements were already discussed throughout the paper, e.g., how the packing process got easier for the staff involved, in a consistent manner for all ferry departures.

In parallel, it was observed how these new analytics solutions and related changes in practice were widely adopted by all concerned staff in the company. As of now, no one would envisage to go back to previous practice.

We have collected multiple statements from leading stakeholders within Molslinjen, who have been directly involved with the process and in assessing its success.

Kristian Durhuss, CEO of Molslinjen: *"When we embarked on this journey more than five years ago, we envisaged that Molslinjen could fundamentally change how the ferry industry operated by applying advanced technologies across our organisation. Together with Halfspace, I believe that we have redefined operational principles for the ferry industry and have set new benchmarks for impact and innovation in a global scale."*

Jesper Skovgaard, CCO at Molslinjen: *We have had the vision to be the world's most digitalized and AI-driven ferry company. Together with Halfspace, we have achieved that in a very short time. We can already now see the effect of using advanced analytics and artificial intelligence to optimise our services and profits via fully automated forecasts and dynamic pricing of our products, and we only expect the impact to increase in the future."*

Lasse Janerka, Digital Director at Molslinjen: *"The incorporation of AI technology in our operations, in collaboration with Halfspace, marks a revolutionary shift in our practice, from mostly human-based to fully data-driven operations.As a result, we have witnessed substantial improvements, especially when it comes to predict the demand and to optimize the loading of vehicles, that are even beyond our original expectations. The revenue management system additionally brings a paradigm shift in the way we handle ticket sales. This advancement has elevated our operational efficiency and profoundly improved the customer experience, exemplifying the vast potential of digital transformation in our sector.*

Jens Christian Bjeldorf, Chief Captain at Molslinjen: *"From my perspective at the helm, the changes instigated by our partnership with Halfspace are distinctly noticeable. The improvements in our daily operations have been significant. The enthusiastic adoption and integration of these technologies by our team have been remarkable, since recognising their contribution to increased efficiency and improved decision-making processes. The process of preparing departures has become easier and less stressful. It feels that we have fundamentally transformed our approach to maritime service and that we are setting a precedent in the industry."*

Besides the direct recognition from stakeholders within Molslinjen, the work performed received substantial acclaim. The broad external recognition is evident from the many awards within digital innovation, data science, AI, and Advanced Analytics, won in Denmark and the UK, e.g.,

- *2022 Danish Digital Awards*: Award winner in Data Science
- *2022 UK AI & Machine Learning Awards*: Dynamic Pricing case ("highly commended")
- *2021 UK AI & Machine Learning Awards*: Business Transformation of the Year

- *2021 Data Breakthrough Awards*: Halfspace named "Industry Leader"
- *2021 IDC Data Strategy & Innovation Awards* (runner-up)
- *2021 Danish Digital Awards*: Award winner within Analytics & AI, Innovation & Marketing Automation

6. Conclusions and outlook

Digitalization is also changing the way passenger ferries may be operated, in a more data-driven and efficient manner. From a company like Molslinjen, the change has been substantial and happening over a fairly short period of time. Starting with simple data products to improve user engagement, it rapidly evolved into development and rolling out a set of analytic tools including a packing solution, a forecasting engine and a revenue management system. The impact from these new solutions were evidenced, both quantitative and qualitatively. This has brought profound changes at all levels within the company, but also with external stakeholders (e.g., customers, investors and the media).

Molslinjen now is in a comfortable position to continue this digitalization process. Some of the obvious next steps will be to source additional data to improve the outcome of the forecasting and dynamic pricing analytics components. Another one is to more broadly deploy the complete toolbox on all routes. Eventually, since having reached a leading position within the passenger ferry landscape, Molslinjen and Halfspace may help and advise others to engage with similar developments.

Acknowledgements

Over the years, many people at Halfspace and Molslinjen got involved in this partnership, as well as specific development and related activities. Hence, the list of people who should be acknowledged for this work and the success of that partnership extends far beyond the list of authors of this paper. Specifically at Halfspace, Svend Lund Breddam, Marie Opstrup Andersen, Vera Patricio, Uffe Furlig Larsen and Christian Michelsen are to be acknowledged for their input to the forecasting and revenue management works (as well as their input to this manucript). We also want to acknowledge all those who have contributed with both positive and critical feedback over the years, as well as with integration, training, etc. at Molslinjen and Halfspace. Finally, we are grateful for the opportunity given by the Franz Edelman award committee to be finalists of the award competition in 2024. Our coaching team (Mikael Rönnqvist, Kermit Threatte and Manish Bansal) has provided us with continuous and invaluable support – they should be thanked for that.

Bios

Pierre Pinson is Chief Scientist at Halfspace, as well as the Chair of Data-centric Design Engineering at Imperial College London, Dyson School of Design Engineering (UK). He is also an Adjunct Professor of Operations Research and Analytics at the Technical University of Denmark. He is the Editor-in-Chief of the

International Journal of Forecasting, an IEEE Fellow and an IIF Director. He is a ISI/Clarivate highly-cited scientist (2019-2023) and he is seen as a leading thought-leader within predictive analytics.

Mikkel Bjørn is a partner and Associate Director of Data Science in Halfspace. He holds a Ph.D. in Particle Physics from the University of Oxford for work at the CERN laboratory, as well as an MSc in Theoretical Physics from the Niels Bohr Institute. His academic and professional work has received multiple prizes and awards, including the 2021 PhD prize for the best doctoral thesis in the UK within the LHCb collaboration. He won a bronze medal at the International Physics Olympiad finals in Mexico in 2009.

Simon Kristiansen holds a Ph.D. in Operations Research and Applied Mathematics from the Technical University of Denmark. He is a co-founder, partner and Associate Director for Engineering at Halfspace. He received the bronze medal from the unofficial World Cup in Schedule Planning.

Claus Bek Nielsen is founding partner and CEO of Halfspace. He has years of experience in building advanced technical teams to support commercial strategies, and has served as an advisor to senior leaders in well-known Danish and American organisations on AI use case maturity and investments in AI. Prior to founding Halfspace, Claus co-founded two of the most recognised online platforms in Scandinavia, within gaming and fashion blogs. He is a former world top 10 ranked backgammon player by World Series of Backgammon.

Lasse Janerka is Director, Business Optimization & Digital at Molslinjen. He holds both B.Sc. and M.Sc. degrees from Aarhus University, and had a diverse career within marketing and journalism before to lead different projects and initiatives for Molslinjen over the last 10 years.

Jesper Skovgaard is Chief Commercial Officer (CCO) at Molslinjen. He received a degree from the Copenhagen Business School. He has been leading pricing and revenue management projects for different companies in the airline industry, before to take that lead position at Molslinjen 12 years ago.

Kristian Durhuus is Chief Executive Officer (CEO) at Molslinjen. He received a B.Sc. in Business Administration from the University of Southern Denmark and an M.Sc. in International Management and Business Economics from the Copenhagen Business School. He has had management leadership positions with Swiss International Airlines, LSG Sky Chef, Copenhagen Airport, ForSea and Molslinjen.

References

Belobaba P (1987) Air travel demand and airline seat inventory management. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA (USA).

Belobaba P (1989) Application of a probabilistic decision model to airline seat inventory control. *Operations Research* 37:183–197.

Belobaba P (1992) Optimal vs. heuristic methods for nested seat allocation. in *Proceedings of the AGIFORS Reservations and Yield Management Study Group, Brussels.*

Bergmeir C (2023) Common pitfalls and better practices in forecast evaluation for data scientists. *Foresight: The International Journal of Applied Forecasting* 70: 5–12.

Bojer CS, Meldgaard JP (2021) Kaggle forecasting competitions: An overlooked learning opportunity. *International Journal of Forecasting* 37(2): 587–603.

Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.

Christiansen M, Hellsten E, Pisinger D, Sacramento D, Vilhelmsen D (2020) Liner shipping network design. *European Journal of Operational Research* 286(1):1–20.

den Boer AV (2015) Dynamic pricing and learning: Historical origins, current research, and new directions. *Surveys in Operations Research and Management Science* 20(1): 1–18.

Fiig T, Isler K, Hopperstad C, Belobaba P (2009) Optimization of mixed fare structures: Theory and applications. *Journal of Revenue and Pricing Management* 9:152–170.

Fiig T, Le Guen R, Gauchet M (2018) Dynamic pricing of airline offers. *Journal of Revenue and Pricing Management* 17:381–393.

Friedman JH (2001) Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 29(5): 1189–1232.

Guillen J, Ruiz P, Dellepiane U, Maccarone L, Maccioni R, Pinzuti A, Procacci E (2019) Europcar integrates forecasting, simulation, and optimization techniques in a capacity and revenue management system. *INFORMS Journal on Applied Analytics* 49(1): 1–91.

Kuyumcu A, Yildirim U, Hyde A, Shanaberger S, Hsiao K, Donahoe S, Wu S, Murray M, Maron M (2018) Revenue management delivers significant revenue lift for holiday retirement. *INFORMS Journal on Applied Analytics* 48(1): 1–91.

Petropoulos F and co-authors (2022) Forecasting: Theory and practice. *International Journal of Forecasting* 38(3): 705–781.

Petropoulos F and co-authors (2023) Operational research: Methods and applications. Preprint, under review, available at arxiv.org/abs/2303.14217.

Schultz D, Stephan J, Sieber J, Yeh T, Kunz M, Doupe P, Januschowski T (2014) Causal forecasting for pricing. Arxiv preprint, arXiv:2312.15282, available at arxiv.org/2312.15282.

Trivella A, Pisinger D (2016) The load-balanced multi-dimensional bin-packing problem. *Computers & Operations Research* 74(1): 152–164.

Talluri K, Ryzin G (2004) *The Theory and Practice of Revenue Management*. Springer: New York.

van Twiller J, Sivertsen A, Pacino D, Jensen RM (2023) Literature survey on the container stowage planning problem. *European Journal of Operational Research*, in press/available online.

Weatherford L and Ratliff R (2010) Review of revenue management methods with dependent demands. *Journal of Revenue and Pricing Management* 9: 326-340.

Wergeland T (2013) Ferry passenger markets. In: *The Blackwell Companion to Maritime Economics* (Talley WK, ed.), Wiley, pp. 161–183.