Expert Systems With Applications 271 (2025) 126586

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Machine Learning for Master Production Scheduling: Combining probabilistic forecasting with stochastic optimisation

Sebastian Paull *, Alexander Bubak, Heiner Stuckenschmidt

University of Mannheim, B6, 26, Mannheim, 68159, Baden-Württemberg, Germany

ARTICLE INFO

Keywords:
Master Production Scheduling
Stochastic optimisation
Supply chain optimisation
Probabilistic forecasting
Machine learning forecasting
Master planning

ABSTRACT

This research paper delves into the challenges of Supply Chain (SC) planning under demand uncertainty, focussing on Master Production Scheduling (MPS) with capacity constraints. Traditional SC planning methods, often based on point forecasts and basic safety stock calculations, overlook the distinct demand distribution of each product in each planning period. Considering the demand uncertainty is crucial for robust decision making, especially for time series with seasonal and non-stationary demand. To address this gap, the paper introduces a pioneering Separated Estimation and Optimisation (SEO) approach that estimates demand uncertainty using Machine Learning (ML)-based probabilistic forecasting, subsequently solved by stochastic optimisation. Through a comprehensive analysis involving 17 datasets with a total of 303 products, the study confirms the robustness of stochastic optimisation approaches. It also demonstrates the superiority of ML-based forecasting, which is particularly adept at capturing the intricacies of complex demand patterns. The research challenges the conventional reliance on the Gaussian distribution, instead advocating for the adoption of more flexible parametric distributions such as the Negative Binomial (Neg.-Bin.) distribution. Furthermore, it illustrates how to leverage advances in the research areas of ML-based probabilistic demand forecasting and stochastic MPS, as well as providing a basis for future research. Such avenues include the exploration of testing these approaches with a rolling planning horizon and incorporating both demand and supply uncertainties.

1. Introduction

Stemming from the extreme global disruptions caused by the Covid-19 pandemic, recent years have seen a growing awareness around the unpredictable nature of Supply Chains (SCs). Academia and industry alike are realising how crucial it truly is for SC planning to mitigate the effects of demand and supply uncertainties; now, more than ever before, the pressure is on to find and employ planning strategies that factor uncertainties into decision making. Furthermore, with the rise of artificial intelligence and data-driven decision-making, there is a pressing need for methodologies that can seamlessly integrate probabilistic insights into optimisation processes. This research directly responds to this demand by proposing a hybrid approach that unifies advanced forecasting with operational decision-making.

The SC planning process of Master Planning (MP) plays a key role in SCs as it 'coordinates the material flow of the supply chain as a whole for a mid-term planning horizon' (Fleischmann & Meyr, 2003, 481). Similarly, recent research in the medical field confirms the critical role of comprehensive supply chain coordination during pandemic situations (Fallahi et al., 2024). The subprocess of Master Production Scheduling (MPS), which is the focus of this paper, ensures that the production plan is aligned with broader business goals. MPS determines production quantities and inventory levels by taking demand fluctuations into account and respecting capacity constraints (Stadtler & Kilger, 2005). Depending on the specific setup of the process in a company, the results of the MPS are then synchronised with the upstream processes of personnel planning and Material Requirements Planning (MRP), and downstream with the distribution planning process, illustrated in the SC Planning Matrix by Stadtler and Kilger (2005, 87). On a short-term horizon, operational production planning builds on the results of MPS and refines the decisions by specifying, for example, which product should be manufactured on which machine and with which tools. Consequently, MPS represents a cornerstone for efficient SC planning. Most notably, MPS helps to align production schedules with resource availability, allowing an efficient allocation of manpower, machinery, and materials while enhancing production efficiency (Serrano-Ruiz et al., 2021). By balancing supply and demand, MPS optimises inventory management, avoiding excess inventory and reducing the risk of stockouts (Bagheri et al., 2022). In addition,

* Corresponding author.
E-mail addresses: sebastian.paull@uni-mannheim.de (S. Paull), alexander.bubak@uni-mannheim.de (A. Bubak), heiner.stuckenschmidt@uni-mannheim.de (H. Stuckenschmidt).

https://doi.org/10.1016/j.eswa.2025.126586
Received 24 September 2024; Received in revised form 10 January 2025; Accepted 14 January 2025
Available online 1 February 2025

the accurate incorporation of mid-term demand uncertainties ensures customer orders are met on time, improving customer satisfaction and loyalty (Bagheri et al., 2022). From the perspective of holistic SC planning, MPS provides the basis for inter-departmental coordination, ensuring that all parts of the production process are aligned, thus improving planning and control along the entire SC (Serrano-Ruiz et al., 2021). Inefficient MPS can lead to a range of negative consequences, *e.g.* causing bottlenecks and disrupting the entire production process (Bagheri et al., 2022). Production inefficiencies may lead to increased cost in terms of expedited orders or overtime, reduce customer satisfaction in case of production delays, overburden production equipment, and reduce product quality. Inaccurate MPS may also result in frequent scheduling changes that can strain supplier relationships (Fan et al., 2020). In a nutshell, efficient MPS is vital for many reasons, and is particularly crucial in modern, complex supply chains that are exposed to many disruptions and uncertainties.

In this paper, we consider a MPS problem where production quantities for a set of products must be determined for each period of the planning horizon. The maximum production capacity cannot be increased in the planning horizon. Demand follows an unknown non-stationary distribution and is subject to seasonal fluctuations, which poses the problem of how to incorporate demand uncertainty into the decision-making. The seasonal and safety stock causes inventory holding costs, while excess demand results in lost sales as backorders are not allowed. Supply chain planning involves multiple sources of uncertainty, including supply variability, process disruptions, and lead time variability, however, this study primarily focuses on demand uncertainty. This decision stems from the critical role demand fluctuations play in driving production planning decisions. Unlike other uncertainties, which can often be mitigated through robust internal controls or operational adjustments, demand variability is inherently external and less predictable, making it a pivotal challenge for effective MPS. Moreover, a significant body of literature prioritises demand uncertainty as the dominant factor impacting production planning (*e.g.*, Ciarallo et al. (1994), Englberger et al. (2016)).

Stochastic SC optimisation problems of this kind are traditionally solved by two-step approaches, also referred to as Separated Estimation and Optimisation (SEO) approaches (Huber et al., 2019). The first step constructs a model to estimate the uncertain quantity, *i.e.* the demand, while the second step receives the uncertainty estimate as an input and solves the optimisation problem. Algorithms that are popular in practice, *i.e.* within Enterprise Resource Planning systems (ERPs), and in related literature often rely on simplistic techniques to incorporate demand uncertainty into the optimisation process. Forel and Grunow (2021) and Feng et al. (2011) construct point predictions and account for uncertainty using a simple safety stock rule. Bollapragada and Rao (2006) and Englberger et al. (2016) consider multiple demand scenarios, but rely on rigid distribution assumptions.

Both types of approaches commonly suffer from two major disadvantages. Firstly, they often rely on traditional but overly simplistic forecast models. For instance, Forel and Grunow (2021) adopt predictions from their industry partner, and Englberger et al. (2016) resort to Exponential Smoothing (ETS), a basic yet widely accepted statistical forecasting method. Recent research on demand forecasting demonstrates the superiority of Machine Learning (ML) methods over traditional techniques (Makridakis et al., 2021). Unlike the latter, ML methods allow the incorporation of contextual information in form of features and the exploitation of cross-time series learning. Consequently, in our study, we adopt the best-performing ML forecasting techniques as reported by the *M5 competition*, a leading global forecasting competition on sales data (Makridakis et al., 2021).

Secondly, existing approaches in MPS do not accurately incorporate the distribution of forecast errors. Forel and Grunow (2021) and Feng et al. (2011) rely on a simple safety stock rule to account for prediction errors, while Bollapragada and Rao (2006) and Englberger et al. (2016) assume errors to follow a stationary, Gaussian distribution. Feng et al. (2011) demonstrate on stationary synthetic data that 'when demand is not normal, the gamma approximation significantly outperforms the existing normal approximation from Bollapragada and Rao (2006)' (Feng et al., 2011, 4007), highlighting the need to challenge the traditional assumption of Gaussian demand.

Other recent research corroborates the above observations by emphasising how data-driven methodologies can optimise outcomes amidst uncertainties (Liu et al., 2023), and how Neural Networks (NNs) can improve results (Huang et al., 2024). Consequently, in this work, we strive to close the identified research gaps by introducing a novel SEO framework that combines cutting-edge ML-based probabilistic forecasting with stochastic optimisation. This framework is designed to handle real-world challenges, such as non-stationary demand and flexible distribution modelling, which remain underexplored in current research and practice. By leveraging ML-based methods, our approach not only surpasses the accuracy of statistical point forecasts but also demonstrates robustness across diverse scenarios, laying the groundwork for practical application in modern SCs. While the current study focuses on demand uncertainty, future research could explore integrating other types of uncertainties, such as supply disruptions and production delays, into the proposed framework. Such extensions could provide a more holistic approach to stochastic MPS, further enhancing its robustness in complex supply chain environments.

In conclusion, our key contributions to the current research are as follows:

- We propose a novel SEO approach that employs cutting-edge ML for probabilistic forecasting to accurately incorporate demand uncertainty into MPS decision-making.
- We extensively evaluate our approach using a large real-world data set encompassing 303 products, thereby corroborating and extending findings that have so far only existed on synthetic data. Our approach outperforms two traditional, state-of-the-art benchmark approaches.
- We derive valuable insights for SC practitioners on how to incorporate demand uncertainty into MPS decisions, thus challenging the traditional assumption of stationary, normally distributed demand.

The implications of this research extend beyond theoretical advancements. By providing an adaptable framework, this work contributes to bridging the gap between academic methodologies and their implementation in industry, particularly under conditions of high uncertainty. This innovation aligns with the growing demand for resilient SC solutions in an era marked by disruptions like the COVID-19 pandemic.

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature on stochastic MPS and demand estimation. In Section 3, we give a detailed presentation of the Mixed-Integer Linear Program (MILP) formulation of the MPS problem considered in this work. Section 4 presents the two baseline methods and our proposed solution approach. In Section 5, we detail the data and parameters used in the experimental setup and show in the numerical results the effectiveness of combining probabilistic forecasting with stochastic optimisation. Section 6 summarises the findings and highlights the practical guidance this research provides. This section also outlines opportunities for future research: Implementing rolling-horizon planning, alternating parameters, or fitting distributions on residuals.

2. Related literature

We commence with a brief summary of relevant research addressing the MPS problem, beginning by reviewing studies on optimisation for MPS, and then delving into demand estimation.

2.1. Optimisation

The study of MPS has evolved significantly over the years. Earlier works by McCormick (1980) and Berry et al. (1979) established the fundamental role of MPS in optimally allocating manufacturing resources to meet customer demand. Extensive reviews by Vollmann et al. (1997) and Zipkin (2000) explored inventory management, addressing both non-capacitated and capacitated versions of MPS. Several authors have followed the latter research stream (*e.g.* Ciarallo et al. (1994), DeCroix and Arreola-Risa (1998), Metters (1997)), which is also the subject of this research. Subsequently, Lee and Adam (1986), Lin and Krajewski (1992), and Koh et al. (2002) highlighted the impact of uncertainties in SC environments, emphasising the interplay between forecast errors and system performance. These foundational studies provided a framework for addressing the complexities of MPS optimisation.

Although stochastic models are more suitable for capturing demand uncertainty in MPS, deterministic approaches are still widely adopted and hence provide a baseline for comparison with our research. For example, Krüger and Koberstein (2023) proposed a MILP model for short-term MPS in the automotive industry, incorporating plant structure, state, and order due dates, demonstrating its practical applicability through a simulation study. Similarly, Trost et al. (2023) introduced a deterministic MPS model that considers utilisation-dependent exhaustion and capacity load, addressing social and economic aspects of hierarchical production planning. Furthermore, Martín et al. (2020) developed a deterministic model with robust optimisation techniques to improve production, inventory, and backlogging costs for an automobile second-tier supplier, highlighting its superiority over heuristic approaches. Lastly, Mohammed et al. (2023) explored dynamic resource availability in product mix problems, comparing linear and non-linear models for optimising profit and inventory costs. These examples illustrate the continued utility of deterministic MPS models in addressing real-world production scheduling challenges.

Overall, in the last decade, the focus has shifted towards stochastic optimisation approaches, incorporating demand uncertainty into MPS models. Englberger et al. (2016) introduced a two-stage stochastic programming model with recourse, tailored for rolling planning environments. This model mitigates tardiness in customer orders by considering numerous demand scenarios while increasing inventory levels and capacity utilisation. Similarly, Forel and Grunow (2021) developed a framework to overcome barriers to industrial adoption of stochastic programming, employing a two-stage stochastic model in a real-world case study. Their approach demonstrated substantial reductions in planning nervousness and improved stability, demand satisfaction, and inventory costs. Englberger et al. (2022) proposed stochastic models for MPS that integrate scenario-based capacity-load factors. By iteratively building realistic capacity-load factor scenarios in a rolling horizon environment, their models address capacity bottlenecks effectively, reducing production order tardiness.

These contemporary studies underscore the importance of modelling uncertainty and solving the MPS problem with robust optimisation techniques. Leveraging scenario-based and adaptive approaches, these methods yield more resilient production schedules, supporting practical applications in industrial settings. Employing flexible distributions such as the Negative Binomial (Neg.-Bin.) one allows for a more accurate estimation of the actual demand. In this research, we build on these advancements by opting for stochastic optimisation as it ensures an optimal solution irrespective of the demand distribution. Given today's computing power and big data handling techniques, we do not expect the solution time to be a limiting factor.

2.2. Estimation

The MPS field of study typically focusses on optimisation of the MPS problem on simulated demand data rather than including demand estimation in their studies. Chu (1995), Forel and Grunow (2021), and Tang and Grubbström (2002) are an exception as they are based on given point forecasts. Consequently, we chose to explore the literature on demand forecasting that does not concentrate on the MPS issue. To gain an understanding of the best-in-class forecasting techniques for sales forecasting, we examine the results of the 2020 *M5 Accuracy competition*. The *M competitions* are described as 'probably the most influential and widely cited in the field of forecasting' (Makridakis et al., 2022, 2). The *M5 competition* in both the *M5 Accuracy* and *M5 Uncertainty* variant used a SC dataset, namely sales data. This data represented the hierarchical unit sales of Walmart, the world's largest retail company. The goal of the *M5 Uncertainty competition* was to generate the most precise point forecasts for 42,840 time series based on the accuracy metric WRMSSE (Makridakis et al., 2022). The time series were typical of a retail company, with grouped, highly correlated, and cross-sectional series, and focused on intermittent demands, with explanatory features provided. The key insights of the *M5 Accuracy competition*, according to Makridakis et al. (2022), were:

- Most contestants used Light Gradient-Boosting Machine (LGBM), a ML algorithm for non-linear regression based on Gradient Boosting Decision Trees (GBDTs).
- Among the best performing techniques were basic NNs and deep learning approaches, which yielded significantly better results than the statistical benchmarks.
- Training global models and including features was found to be successful.

These point forecasts can be either used directly or adjusted by the empirical forecast errors to create demand distributions. Instead of building on empirical forecast errors, one could also estimate the parameters of demand distributions or estimate the prediction intervals based on past demands and explanatory features. The *M5 Uncertainty competition* in 2020 evaluated state-of-the-art solutions for probabilistic forecasting (Makridakis et al., 2021). The contestants were asked to estimate the prediction intervals 50%, 67%, 95%, and 99% and were evaluated using the WSPL metric. Most of the participants used LGBM, followed by Long Short-Term Memory Neural Networks (LSTM NNs). The top three contestants used extremely large GBDTs instances applying LGBM and trained them separately as point forecasts for the prediction intervals. The fifth and seventh ranked teams based their uncertainty estimates on forecast residuals, the fourth team estimated parameters for Neg.-Bin. distributions and Student's T (Stud. T) distributions for higher levels using NNs. The sixth-ranked team used Neg.-Bin. distributions generated by a multi-stage local level state-space model.

Few researchers in the domain of stochastic MPS illustrate how demand forecasts can be used. Vargas and Metters (2011) simulate data with a Neg.-Bin. distribution. They then directly use this as input to the MPS problem. Similarly, Feng et al. (2011) use a Gamma (Gamma) distribution and include the approach of Bollapragada and Rao (2006) with a Gaussian distribution. Vogel et al. (2017) are one step closer to practical application with simulated data to generate forecasts with Gaussian distributed forecast errors but fit the parameters accordingly. Englberger et al. (2016), Körpeolu et al. (2011) and Bollapragada and Rao (2006) all use demand scenario paths based on either simulation or human estimation. Forel and Grunow (2021) show multiple ways of generating probabilistic forecasts. In a rolling-horizon setting, they directly use past demands as distributions, determine the empirical distribution from the forecast errors, and estimate distribution parameters of Gaussian and uniform distributions on the forecast errors.

We incorporate the insights gained from the *M5 competition* into our research by deciding to use the best-performing ML approaches from this competition. GBDTs, like LGBM, can only be used for parametric distributions with one parameter or to estimate prediction intervals but not whole distributions. Like the team ranked fourth in the *M5 Uncertainty competition*, we are using NNs with the Neg.-Bin. and Stud. T distributions. Additionally, similarly to the fifth and seventh ranked team, we also tested the Residual-based (RB) demand estimate approaches. To the best of our knowledge, we are the first to use probabilistic forecasting in the context of MPS.

3. Problem formulation

In this article, we study the problem of MPS, a multi-period inventory management problem (Fleischmann & Meyr, 2003). We make the following assumptions in the formulation of the model: The demand in each period is unknown and stochastic, the distribution is non-stationary and independent across periods. We assume to have access to contextual information, or feature values, that have predictive power towards the demand. Backlogging is not allowed, so any excess demand is lost. In this research, lost sale penalty costs are not included explicitly in the objective function due to the lack of scientific basis for estimating their effect. Commonly, a lost sales cost per unit l is assumed (Liao et al., 2011). This adjustment could be implemented simply by incorporating $-f \cdot m_j \cdot U_{t,j,s}$ in the objective function (see Eq. (1)). Built-up inventory is penalised by linear holding costs. Aggregated production quantities of all products are subject to a period-dependent maximum capacity. The sequence of events in each period is similar to Feng and Shanthikumar (2018): (1) Receive planned deliveries and (2) add them to the end inventory of the previous period, (3) observe and satisfy demand, (4) determine the end inventory, and (5) record profit margin and inventory holding costs. The notation, consisting of indices, parameters and decision variables, together with Eq. (1) to Eq. (4), present the MILP formulation of the stochastic MPS problem. Since the deterministic MPS formulation is equivalent to the stochastic MPS formulation with exactly one scenario, we do not explicitly present it in this paper to avoid redundancy.

Indices

T	Set of planning periods t
J	Set of products j
S	Set of scenarios s

Parameters

m_j	product margin per unit and product $j \in J$
h_j	inventory holding cost per unit, and product $j \in J$, and period $t \in T$
$i_{0,j}$	initial inventory at period 0 for product $j \in J$
$\hat{d}_{t,j,s}$	demand of scenario $s \in S$ of product $j \in J$ in period $t \in T$
c_t	available production capacity in period $t \in T$

Decision variables

$X_{t,j} \in \mathbb{N}_0^+$	production quantity of the product $j \in J$ in the period $t \in J$
$U_{t,j,s} \in \mathbb{N}_0^+$	unmet demand in the scenario $s \in S$ of the product $j \in J$ in the period $t \in T$

Maximise

$$\sum_{t\in T, j\in J, s\in S} m_j \cdot (\hat{d}_{t,j,s} - U_{t,j,s}) - h_j \cdot \left(i_{0,j} + \sum_{\tau \le t} \left(X_{\tau,j} - \hat{d}_{\tau,j,s} + U_{\tau,j,s} \right) \right) \tag{1}$$

subject to

$$\sum_{j\in J} X_{t,j} \le c_t \qquad \forall t \in T \tag{2}$$

$$\sum_{\tau \le t} \left(X_{\tau,j} - \hat{d}_{\tau,j,s} + U_{\tau,j,s} \right) + i_{0,j} \ge 0 \qquad \forall t \in T, j \in J, s \in S \tag{3}$$

$$X_{t,j}, U_{t,j,s} \ge 0 \qquad \forall t \in T, j \in J, s \in S \tag{4}$$

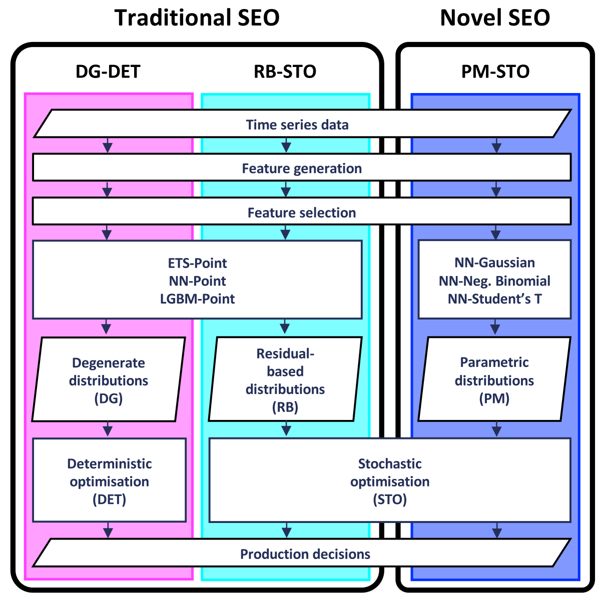

Fig. 1. Three categories of approaches to solve the MPS problem: DG-DET, RB-STO, and PM-STO.

The objective is to maximise the expected net profit over all periods and products (see Eq. (1)). As is common in the literature, the expectation is approximated by the sample average over a number of demand scenarios S (Englberger et al., 2016). As the objective is maximised, we consider the sum instead of the average for notational simplicity. Gross profit is calculated by multiplying the product margin m with the fulfilled demand, which is the scenario demand $\hat{d}_{t,j,s}$ reduced by the unmet demand $U_{t,j,s}$. To obtain net profit, inventory holding costs are deducted, which are computed by multiplying the unit inventory cost h_j with the ending inventory $I_{t,j,s}$. As $I_{0,j,s} = i_{0,j}$ and $I_{t,j,s} = I_{t-1,j,s} + X_{t,j} - \hat{d}_{j,t,s} + U_{t,j,s}$, inventory is completely determined by production quantity, demand and unmet demand, so we can substitute $I_{t,j,s}$ with $i_{0,j} + \sum_{\tau \le t} \left(X_{\tau,j} - \hat{d}_{\tau,j,s} + U_{\tau,j,s} \right)$. Profit maximisation is subject to a number of constraints. Constraint (2) ensures that the aggregate production of products in each period t does not exceed the production limit c_t. Constraint (3) specifies that ending inventory must be positive, which guarantees that backlogging is prevented and unmet demand recorded accordingly. Constraint (4) specifies the variables' domains, ensuring that production quantities and unmet demand are non-negative.

4. Solution approaches

Inventory management problems such as MPS are frequently addressed using SEO techniques (Ban, 2020; Bertsimas & McCord, 2019; Feng et al., 2011; Forel & Grunow, 2021; Huber et al., 2019; Müller et al., 2020). These SEO techniques typically involve two stages. In the first stage, the demand distributions are estimated, generating the demand samples $\hat{d}_{t,j,s}$ needed as input for the MPS MILP problem (Eq. (1) to Eq. (4)). In the second stage, the latter is solved, generating production decisions. The approaches applied in this paper differ mainly in the first stage. Following different methods of distribution estimation, all approaches rely on the same linear programming solver to solve the MPS problem. Focusing on the estimation stage, this section presents these approaches. Our explanations are supported by the graphical overview provided in Fig. 1.

4.1. Traditional SEO baselines

The estimation step of conventional SEO approaches revolves around a demand prediction model and can be further divided into *demand* *prediction* and *distribution estimation*. We present the demand prediction methods applied in this work in Section 4.1.1, followed by the distribution estimation techniques in Section 4.1.2.

4.1.1. Demand prediction

Demand forecasting has long been a key focus in operations research, leading to numerous prediction techniques. We have chosen the state-of-the-art methods LGBM, and NN, and for reference ETS. The ML models were selected because they ranked highest in the *M5 accuracy competition* (Makridakis et al., 2021), a prestigious sales forecasting challenge. More details on our selection can be found in Section 2.2. As MPS is a multiperiod problem, multiperiod forecasting is required, such that the output of our forecasting methods are not point forecasts but instead path forecasts $(\hat{d}_{t,j})_{t\in T}$ for the entire optimisation horizon T. Since a point forecast for every planning period is equivalent to a path forecast for the planning horizon, the terms point forecast and path forecast are used interchangeably in this paper.

The ETS methods generate forecasts by calculating weighted averages of past data points, with the weights decreasing as the data points become older (Hyndman et al., 2008, 5). The set of 18 ETS models includes various types of errors (additive or multiplicative), trends (none, additive, additive damped), and seasonal patterns (none, additive, or multiplicative) (Hyndman & Khandakar, 2008), from which we have selected the variant (A, Ad, A) as it minimises Mean Absolute Percentage Error (MAPE) in the test set (see Supplementary material C) and optimises the likelihood-based Akaike information criterion (Hyndman et al., 2008, 27). In our implementation, we rely on the class `ExponentialSmoothing` of the Python module `statsmodels.tsa.holtwinters` module.

Both NNs and LGBMs are ML models. ML models provide two main advantages over traditional statistical forecasting methods like ETS (Makridakis et al., 2021). First, a single ML model can be trained on multiple time series, leveraging cross-time series learning. Second, in addition to the time series data, explanatory features, whether externally sourced or derived from the time series itself, can be used to enhance the forecasts.

A NN is a network of computing nodes connected by directed links. Input is provided to an input layer, output is obtained from an output layer, and the remaining nodes are contained in hidden layers (Gamboa, 2017). NNs with multiple layers are capable of recognising intricate patterns and are called Deep Neural Networks (DNNs). For point predictions, we employ a feed-forward NN relying on the `GluonTS` package (see Section 4.2) considering the features over the past $3|T|$ periods as input, using one hidden layer of 40 neurons and the ReLU activation function. We train the network with the Adam optimiser (Kingma & Ba, 2015) minimising the Mean Absolute Error (MAE), employing mean scaling, 100 batches per epoch and a total of 1000 epochs. The learning rate was tested in $\{10^{-4}, 10^{-3}, 10^{-2}\}$ and tuned to 10^{-3}, while the *context length*, which determines the number of periods included in the training window, was tested in $\{2|T|, 3|T|, 4|T|\}$ and tuned to $3|T|$ (see Supplementary material C).

LGBM is a tree-based ML framework. Decision trees are a type of supervised learning algorithm used for classification and regression tasks, where the model makes decisions based on a series of binary questions about input features, leading to a final decision at each leaf node (Hastie et al., 2009). Given their intelligibility and simplicity, decision trees are among the most popular machine learning algorithms. Gradient Boosting Decision Trees (GBDTs) are a special kind of tree-based model ensemble, which have proven to achieve state-of-the-art performance for many ML learning tasks like multiclass prediction, click prediction, and learning-to-rank (Ke et al., 2017; Li et al., 2020; Zhang & Jung, 2021). GBDTs are composed of a series of decision trees that are trained in succession. At each step, the model fits a new tree to the negative gradients of the current model, thereby correcting the error of its predecessors (Ke et al., 2017). In 2017, Microsoft introduced LGBM, an implementation of a GBDT with two novel techniques that significantly improve performance: *Gradient-based One-Side Sampling* and *Exclusive Feature Bundling*. In the *M5 accuracy competition*, all 50 top contestants used LGBM (Makridakis et al., 2021). We use LGBM through the `TreeEstimator` class within the `GluonTS` framework (see Section 4.2). The default number of leaves is set to 31 and we enforce column-wise training to maintain stability. Otherwise, we apply the same hyperparameter tuning as for NNs (see Supplementary material C).

$\forall j \in J$										
$d_{t,j}$	8	18	19	2	11	7	11	8	16	2
$s = 1$										
t	−3	−2	1	0	1	2	2	4		
$d_{t,j,1}$	8	18	19	2	11	7	11	8		
$\hat{d}_{t,j,1}$					1	12	4	19		
$\mathbf{e}_{t,j,1}$					10	−5	7	−11		
$s = 2$										
t	−4	−3	−2	−1	0	1	2	3	4	
$d_{t,j,2}$	8	18	19	2	11	7	11	8	16	
$\hat{d}_{t,j,2}$						19	2	17	18	
$\mathbf{e}_{t,j,2}$						−12	9	−9	−2	
$s = 3$										
t	−5	−4	−3	−2	−1	0	1	2	3	4
$d_{t,j,3}$	8	18	19	2	11	7	11	8	16	2
$\hat{d}_{t,j,3}$							3	20	13	17
$\mathbf{e}_{t,j,3}$							8	−12	3	−15

Fig. 2. Calculation of the forecast errors $e_{t,j,s}$.

4.1.2. Distribution estimation

The pertinent literature presents two common methods to get from point predictions to production decisions.

The most simplistic approach, which (Feng et al., 2011) call the traditional MRP approach, only uses information on the conditional expectation of demand and neglects demand variability (Krüger & Koberstein, 2023; Martín et al., 2020; Mohammed et al., 2023; Trost et al., 2023). The predicted demand path is fed unmodified to the MPS problem, *i.e.* problem (1)–(4) is solved with only one scenario path by setting $S = 1$ and $\hat{d}_{t,j,1} = \hat{d}_{t,j}$. As demand variability is neglected, this is commonly referred to as deterministic optimisation (Härdle et al., 2005). Since the approach relies on point forecasts and a DG distribution estimate, subsequently solved by deterministic optimisation, we refer to it as Point forecast with a degenerate distribution solved by deterministic optimisation (DG-DET) (see Fig. 1).

A point forecast-based method which accounts for demand variability by utilising multiple demand scenarios can be developed by incorporating empirical forecast errors (Forel & Grunow, 2021). This process is illustrated in Figs. 2 and 3. For each product j, predictions are made on the training data in a rolling horizon manner, and each prediction generates a scenario path $\hat{d}_{t,j,s}$ of length $|T|$. Using actual demand $d_{t,j}$, the forecast error $e_{t,j,s}$ can be calculated as

$$e_{t,j,s} = d_{t,j,s} - \hat{d}_{t,j,s}, \forall t \in T, j \in J, s \in S \tag{5}$$

The toy example shown in Fig. 2 illustrates this procedure using $|S| = 3$ scenarios and a planning horizon of $T = \{1, 2, 3, 4\}$. This methodology is based on Forel and Grunow (2021) with two modifications due to the characteristics of our data. First, we assume that error distributions are dependent on the product, so we compute product-dependent distributions. Second, we do not assume season-dependent distributions but calculate errors independent of the season instead. To construct the distribution estimate $\hat{d}_{t,j,s}$ on the test period, we add the scenario errors to the demand forecasts $\hat{d}_{t,j}$ made for the test period as shown in Fig. 3. All scenario demand values below zero are adjusted:

$$\hat{d}_{t,j,s} = \max(\hat{d}_{t,j} + e_{t,j,s}, 0) \forall t \in T, j \in J, s \in S \tag{6}$$

$\forall j \in J$

t	−4	−3	−2	−1	0	1	2	3	4
s						$e_{t,j,s}$			
1						10	−5	7	−11
2						−12	9	−9	−2
3						8	−12	3	−15
$d_{t,j}$	18	7	14	3	2				
$\hat{d}_{t,j}$						1	4	1	19
s						$d_{t,j,s}$			
1						11	0	8	8
2						0	13	0	17
3						9	0	4	4

Fig. 3. Adding forecast errors $e_{t,j,s}$ to demand forecasts $\hat{d}_{t,j}$.

Finally, these scenario paths are fed to the MPS MILP for stochastic optimisation (Birge & Louveaux, 2011; Härdle et al., 2005, pp. 173–201). As this approach relies on point demand forecasts and a residual-based distribution estimate, subsequently solved by stochastic optimisation, we refer to it as Point forecast with a residual-based distribution solved by stochastic optimisation (RB-STO) (see Fig. 1).

4.2. Novel SEO approach

Instead of the two-step approach of point prediction and subsequent distribution approximation, the method we propose estimates the demand distribution directly. Despite the popularity of ETS and LGBM as forecasting models, they are not suitable for probabilistic forecasting with the exception of quantile forecasting or parametric distributions with only one parameter. Therefore, the parameters of these distributions are estimated by NNs.

To this aim, as mentioned in Section 4.1.1, we rely on `GluonTS` (Alexandrov et al., 2020), a toolkit for time series modelling with DNNs a special focus on probabilistic forecasting, launched by Amazon Web Services in 2019. To this aim, a parametric distribution is fixed and a function mapping input features to the time-dependent distribution parameters is learnt by minimising the negative log-likelihood. The available models are listed on the `GluonTS` website (Amazon, 2023), including advanced specialised models like `DeepAR` by Salinas et al. (2020) and simpler customisable models like the `SimpleFeedForward` estimator, which we adopt in our work. We employ a feed-forward NN of one hidden layer with $10|T|$ neurons, which the `GluonTS` network reshapes to a $|T| \times 10$ matrix. For a distribution with n parameters, the network then applies n distinct output layers to obtain a $|T| \times n$ matrix of $|T|$ time-dependent distribution parameterisations. Distribution-specific domain maps ensure the validity of the output, such as positive values for standard deviations. `GluonTs` provides various probability distributions, including common parametric distributions such as the three we adopt: Gaussian, Stud. T, and Neg.-Bin. (Alexandrov et al., 2020). Except for the number of hidden neurons and the loss function, all hyperparameters coincide with those used for the point prediction NNs (see Section 4.1.1) and are tuned accordingly (see Supplementary material C). After predicting the parameters for the test period, we sample scenario paths and feed these scenarios to the MPS MILP (1)–(4) for stochastic optimisation (Härdle et al., 2005). As this approach makes probabilistic forecasts coupled with stochastic optimisation, we refer to it as Probabilistic forecast with a parametric distribution solved by stochastic optimisation (PM-STO) (see Fig. 1).

4.3. Justification of the proposed approaches

The different point and probabilistic prediction models coupled with different distribution estimation techniques and optimisation paradigms result in a total of nine different SEO methods, as visualised in Fig. 1. The integration of the three demand forecasting models (ETS, NN, and LGBM) with the deterministic MILP formulation that results from (1)–(4) when considering only one demand scenario produces three DG-DET models, which we refer to as ETS-DG, LGBM-DG, and NN-DG. Combining the three demand forecasting models with a residual-based distribution estimate and solving the stochastic MILP produces three RB-STO methods, which we refer to as ETS-RB, LGBM-RB, and NN-RB. Lastly, testing three different parametric distributions (Gaussian, Stud. T, and Neg.-Bin.) and solving the stochastic MILP generates the three techniques categorised into the PM-STO group: NN-Gaussian, NN-Stud. T, and NN-Neg.-Bin.

The latter group of approaches is proposed in this work due to the increased flexibility, scalability, and superior performance of these approaches (see Section 5.4). Recently, Operations Research (OR) researchers have started to question the conventional method of keeping estimation and optimisation separate. Instead, Integrated Estimation and Optimisation (IEO) methods have been introduced that combine both steps, directly predicting the solutions to the optimisation problem (Ban & Rudin, 2019; Huber et al., 2019; Müller et al., 2020). However, decoupling demand estimation from the MPS problem makes our approach flexible to different problem variations, extending even beyond MPS. In addition, traditional optimisation techniques such as MILPs permit explicit formulation and adherence to hard constraints, which are not inherently recognised by conventional IEO methods like NNs. Lastly, the SEO methods are more resilient to external disruptions such as changes in profit margin, cost factor, or production capacity. Regarding scalability, we stress that, on average, solving the stochastic MPS MILP took 519 s using 10 CPU cores, demonstrating that the need to solve the MILP repeatedly is not a real concern. Moreover, SEO methods offer explicit control over computational effort, as practitioners can choose to reduce the number of sample paths or terminate the MILP solver early to save computation time. Compared to traditional forecasting methods like ETS, our approach leverages the well-known advantages of ML. For instance, ML is known for improving with more training data. Moreover, ML exploits cross-product relationships by fitting a global model, whereas traditional methods like ETS fit one model per product, which scales poorly for larger assortments. The high effort in retraining ML models can be mitigated by warm-starting the training process. Finally, our proposed approach directly predicts the demand distribution, which avoids the additional steps required under RB-STO and scales better. In addition, it allows explicit control over the predicted distributions and can be particularly valuable under non-stationary error distributions.

5. Numerical study

5.1. Data and parameters

Our evaluation of the solution approaches uses a real-world weekly demand data set from a leading global manufacturer of household appliances. This ensures the results reflect practical conditions without artificial constraints. The data set encompasses a diverse product portfolio, including small and large kitchen appliances, refrigeration units, and cleaning devices. Products are manufactured to stock and can be stored for extended durations exceeding one year. The products are categorised into 17 distinct assembly groups and are distributed to 19 distribution centres located across multiple continents. Since not all assembly groups are supplied to every distribution centre, the data represents 303 unique location-product combinations. Each assembly group constitutes an independent data set, with no overlap or dependencies between groups, allowing the optimisation problem to be solved separately.

On average, the weekly demand across all products is 271 units, with a Coefficient of Variation (CV) of 2.53. This high CV reflects significant variability in demand, highlighting the challenge of accurate

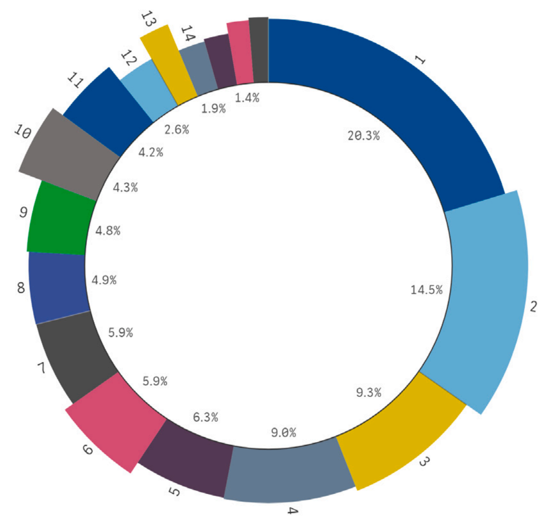

Fig. 4. Demand per assembly group, the radius indicating the CV.

forecasting and robust optimisation. Fig. 4 illustrates the demand patterns for each assembly group, with the radius indicating the respective CV.

The planning horizon covers the latest $T = 26$ weeks of the data set, during which fixed capacities cannot be increased. These 26 weeks are used as the test set for evaluating model predictions and production decisions. All data preceding this test period is used for training the forecast models in a rolling manner, ensuring that only past information is available for model training.

We use a minimum number of 104 periods for training, as this is required by an ETS method with a seasonality of 52. With a total of 425 weeks in our data and a test period of 26 weeks, we obtain 270 forecast rolls, and hence 270 scenario paths for RB-STO. For comparability, we also sample 270 scenarios when evaluating PM-STO.

For our numerical study, we have to make assumptions and define exogenous parameters. Similar to Feng and Shanthikumar (2018), transportation and goods issue lead times are assumed to be deterministic and fixed and are incorporated in the data by shifting the time series accordingly. We set the initial stock level $i_{0,j}$ to 0 for each product $j \in J$. We fix product margins at $m_j = 100$ and vary inventory holding cost $h_j \in \{100, 66.67, 42.86, 25, 11.11, 5.26\}$ to cover the corresponding Newsvendor service levels $SL = \frac{m_j}{m_j+h_j} \in \{50\%, 60\%, 70\%, 80\%, 90\%, 95\%\}$. The production capacity limit c_t is determined by the mean of the weekly aggregated demand for each assembly group across all periods $t \in T$, a practical approach that overlooks factors such as factory holidays.

5.2. Feature engineering

To prepare for the feature extraction and selection process, we first identify and utilise static features, such as product attributes and country-specific information, in conjunction with multiple levels of product hierarchy. The time series are then aggregated according to these features, effectively increasing the number of derived time series on which features can be computed.

For automated feature extraction, we employ the Python package `tsfresh` by Christ et al. (2018), which is specifically designed to handle time series data. To enhance scalability and handle large data sets efficiently, we reimplemented several components of `tsfresh` using `Apache Spark`. The feature selection process involves three primary steps:

1. Feature Extraction: A comprehensive set of features is extracted using a predefined configuration, in this case, the `MinimalFCParameters` class of `tsfresh`. These features include statistical measures (*e.g.*, mean, variance), temporal patterns, and other domain-independent characteristics.
2. Significance Testing: Each extracted feature is evaluated for its relevance to the regression task by calculating p-values through statistical hypothesis tests. This step assesses the relationship between each feature and the target variable to identify meaningful predictors.
3. Feature Selection: Based on the calculated p-values, a multiple testing procedure is applied to select statistically significant features. This ensures that only the most relevant features are used in the subsequent modelling stage, improving model performance and interpretability.

5.3. Evaluation metrics

Calculation of forecast error metrics

$$\text{MAPE} = \frac{1}{|T \parallel J|} \sum_{t\in T, j\in J} \frac{|d_{t,j} - \hat{d}_{t,j}^{0.5}|}{|d_{t,j}|} \times 100 \tag{7}$$

$$\text{SMAPE} = 2 \cdot \frac{1}{|T \parallel J|} \sum_{t\in T, j\in J} \frac{|d_{t,j} - \hat{d}_{t,j}^{0.5}|}{|d_{t,j}| + |\hat{d}_{t,j}^{0.5}|} \times 100 \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{1}{|T \parallel J|} \sum_{t\in T, j\in J} (d_{t,j} - \hat{d}_{t,j}^{0.5})^2} \tag{9}$$

$$\text{SEC}_f = \frac{1}{|T \parallel J|} \sum_{t\in T, j\in J} |d_{t,j} - d_{t-f,j}| \tag{10}$$

$$\text{MASE}_f = \frac{1}{|T \parallel J|} \sum_{t\in T, j\in J} |d_{t,j} - \hat{d}_{t,j}^{0.5}| / \text{SEC}_f \tag{11}$$

$$\text{COV}_\alpha = \frac{1}{|T \parallel J|} \sum_{t\in T, j\in J} I[d_{t,j} \le \hat{d}_{t,j}^{\alpha}] \times 100 \tag{12}$$

$$\text{QL}_\alpha = 2 \cdot \sum_{t\in T, j\in J} |(d_{t,j} - \hat{d}_{t,j}^{\alpha}) \cdot (I[d_{t,j} \le \hat{d}_{t,j}^{\alpha}] - \alpha)| \tag{13}$$

To evaluate prediction quality of both point predictions and probabilistic forecasts, we use multiple common error metrics, see (7)–(13). The metrics are calculated according to the definitions provided by Hyndman and Athanasopoulos (2018) and Makridakis et al. (2020). Each metric has its own strengths and weaknesses (Makridakis et al., 2021). The MAPE (7), Symmetric Mean Absolute Percentage Error (SMAPE) (8), and Mean Absolute Scaled Error (MASE) (11) are scale-independent metrics that allow for comparisons between different time series. However, they may not provide much insight when the demand is low. On the other hand, the metric Root Mean Square Error (RMSE) (9) gives more weight to large errors, penalising outliers. The seasonally adjusted metric MASE has a seasonal index f, which we set to 52. We also evaluate the probabilistic forecast error metrics Coverage (COV) (12) and Quantile Loss (QL) (13) on the $\alpha = 0.5$ quantile. For most forecast error metrics, the optimal value is 0; however, for the COV, the optimal value corresponds to the α quantile, so in our case the ideal for COV is 50%. QL measures how well a model predicts a given quantile of the target distribution, similarly to COV which evaluates how well the predicted intervals capture the actual outcomes. For the probabilistic forecasts, $\hat{d}_{t,j}^{\alpha}$ denotes the α-quantile of the distribution estimate $\hat{d}_{t,j,s}$.

We may also evaluate point accuracy metrics on probabilistic forecasts, in which case we take the median forecast $\hat{d}_{t,j}^{0.5}$ to be the point forecast. Conversely, we can apply the probabilistic metrics to point forecasts by interpreting the point forecasts as degenerate probability distributions with a 100% estimated probability for one single value.

In addition to forecast accuracy, we monitor Key Performance Indicators (KPIs) directly derived from the optimisation problem, as shown in (15)–(19). These KPIs measure critical aspects of the system's performance, ensuring a comprehensive evaluation of the solution approaches. Fulfilled demand $F_{t,j}$ corresponds to the β-service level, which is the main KPI for demand fulfilment. It is defined as the portion of demand satisfied by starting inventory plus production, see (16).

The profit margin $M_{t,j}$ is computed as the product of the product-specific profit margin m_j and the fulfilled demand $F_{t,j}$, see Eq. (17). Inventory holding costs $C_{t,j}$ are derived from product-specific holding costs h_j and ending inventory $I_{t,j}$, see Eq. (18). Together, these components define the objective value $O_{t,j}$, which reports overall profitability, see Eq. (19). For comparability across assembly groups, we introduce the relative objective value R_j, see Eq. (20).

To compare the nine solution approaches, these KPIs are aggregated over all periods $t \in T$ and products $j \in J$. For example, the aggregated objective function value is denoted as $O = \sum_{t\in T, j\in J} O_{t,j}$. In addition, we track the aggregated production quantity $X = \sum_{t\in T, j\in J} X_{t,j}$ to assess overall production activity, the aggregated fulfilled demand $F = \sum_{t\in T, j\in J} F_{t,j}$, the aggregated product margins $M = \sum_{t\in T, j\in J} M_{t,j}$, and to capture forecasting, the aggregated demand forecast $\hat{d} = \frac{1}{|S|} \sum_{t\in T, j\in J, s\in S} \hat{d}_{t,j,s}$.

Key Performance Indicators (KPIs)

$\hat{d}_{t,j}$ demand forecast of the product $j \in J$ in the planning period $t \in T$, averaged over $s \in S$

$I_{t,j}$ end inventory of the product $j \in J$ in the planning period $t \in T$

$F_{t,j}$ fulfilled demand of the product $j \in J$ in the planning period $t \in T$; Expressed as share of demand d and therefore equivalent to the β-service level

$M_{t,j}$ profit margin of the product $j \in J$ in the planning period $t \in T$

$C_{t,j}$ inventory holding costs of the product $j \in J$ in the planning period $t \in T$

$O_{t,j}$ objective value of the product $j \in J$ in the planning period $t \in T$

Calculation of KPIs

$$\hat{d}_{t,j} = \frac{1}{|S|} \sum_{s\in S} \hat{d}_{t,j,s} \quad \forall t \in T, j \in J \tag{14}$$

$$I_{t,j} = i_{0,j} + \sum_{\tau \le t} X_{\tau,j} - d_{\tau,j} + U_{\tau,j} \quad \forall t \in T, j \in J \tag{15}$$

$$F_{t,j} = \begin{cases} d_{t,j}, & d_{t,j} \le I_{t-1,j} + X_{t,j} \\ I_{t-1,j} + X_{t,j} \end{cases} \quad \forall t \in T, j \in J \tag{16}$$

$$M_{t,j} = m_j \cdot F_{t,j} \quad \forall t \in T, j \in J \tag{17}$$

$$C_{t,j} = h_j \cdot I_{t,j} \quad \forall t \in T, j \in J \tag{18}$$

$$O_{t,j} = M_{t,j} - C_{t,j} \quad \forall t \in T, j \in J \tag{19}$$

$$R_j = \sum_{t\in T} \frac{O_{t,j}}{d_{t,j}} \quad \forall j \in J \tag{20}$$

5.4. Results

This section gives a detailed analysis of our results. To enhance readability, we make extensive use of abbreviations, which can be found in the acronyms section at the end of the article.

Forecast accuracy. Table 1 shows the forecast error for each approach. No forecast method performs best across all error metrics. We give a rank $r \in 1, 2, \ldots, 9$ to each approach in each metric separately and then find the average over the metrics to get $\bar{r}$. The best-performing approaches in terms of average forecast error rank $\bar{r}$ are LGBM-DG (2.3), NN-RB (2.5), NN-Neg.-Bin. (4.2) and NN-Stud. T (4.5).

Objective function. After determining the forecast error metrics as our initial indicator, we assess the models based on the KPIs. The outcomes are shown in Table 2 for the case of $m_j = 100$ and $h_j = 25$, the parameters at which the optimal service level of the Newsvendor problem would be 80%. Results for alterations of the inventory holdings costs h_j can be found in Supplementary material A.

Using aggregated demand d as a reference, we express aggregated forecast $\hat{d}$, production X, as well as the fulfilled demand F as a percentage of d. The profit margins M, inventory holding costs C, and objective function values O are expressed as a percentage of the highest objective function value. Table 2 highlights the approach with the best result, outcomes that are within 5% of the best result are printed in bold.

Interestingly, the top three performing approaches originate from three different groups, suggesting that no single demand estimation technique is superior. The best results, measured by the value of the objective function, is obtained by NN-Neg.-Bin., closely followed by NN-RB with 99%. The third highest result, with 91%, comes from the NN-DG. The three approaches have in common that they leverage NNs in the prediction step. With 78% to 68%, most other approaches achieve moderate performance. Only NN-Gaussian performs considerably worse, achieving only 8% of the objective value attained by NN-Neg.-Bin. When comparing the PM-STO approaches that use the same NN, it appears that the choice of distribution makes a considerable difference. Furthermore, with the exception of the underperforming Neural Network with a Gaussian distribution (NN-Gaussian), the approaches perform better the more powerful the prediction model.

The last column of Table 2 shows the Interquartile range (IQR) of the relative objective function R_j, which serves as a measure of the robustness of a method across products. The NN-Gaussian approach exhibits the lowest IQR at 5.6, but this is not particularly informative as it simply indicates uniformly poor results across all products. For the top three methods, the IQR provides valuable insights: Although Neural Network with a Negative Binomial distribution (NN-Neg.-Bin.) and Neural Network with a residual-based distribution (NN-RB) have very similar objective function values, the robustness of NN-Neg.-Bin. with an IQR of 40.1 is superior to NN-RB with an IQR of 48.9. The third-best method, Neural Network with a degenerate distribution (NN-DG), has a 9% point difference in the objective function value compared to NN-Neg.-Bin. and a distinctly worse IQR of 63.8.

Other key performance indicators (KPIs). The objective function value O is the primary KPI for which we are optimising. However, it is essential to discuss other KPIs as well, as they provide insights into how the objective function value is derived.

ETS-DG, ETS-RB, and NN-Gaussian provide nearly perfect predictions if the aggregated demand across all periods and products was relevant. The models with the lowest total demand forecasts are LGBM-RB (56%), NN-Stud. T (64%), and LGBM-DG (73%). Nonetheless, it is vital to note that the total demand forecast does not necessarily reflect the overall performance, as it is highly aggregated. Indeed, we see little correlation between aggregated forecast accuracy and objective value.

In terms of production, ETS-DG and NN-DG lead to the highest production quantities X of 86%. This is followed at distance by NN-Neg.-Bin. (73%), LGBM-DG (72%), ETS-RB (70%) and then NN-RB (68%). LGBM-RB (48%), and NN-Stud. T (42%) have fairly low aggregated production quantities X, while the optimisation of NN-Gaussian led to only a production of 5% of actual demand. We can observe from Table 2 that higher production generally leads to higher profit margins, but the decisive factor to achieve a high objective is to increase margins while maintaining a reasonable cost level.

Considering that the fulfilled demand F is expressed as a percentage of the total demand d, it equates to the β-service level. The methods NN-DG and ETS-DG show the highest levels of fulfilled demand F, reaching 77% and 76% respectively. A moderate result is obtained by NN-Neg.-Bin. (68%), LGBM-DG (65%), NN-RB (64%), and ETS-RB

Table 1
Forecast error metrics, averaged across the assembly groups. Avg. rank $\bar{r}$, averaged across the forecast error metrics.

Model	Distribution	Group	MAPE	SMAPE	MASE	RMSE	$COV_{\alpha=0.5}$	$QL_{\alpha=0.5}$	$\bar{r}$
ETS	Degenerate (DG)	DG-DET	203%	80%	11.5	207	55%	$4.6 \cdot 10^3$	5.5
	RB	RB-STO	225%	79%	55.5	217	58%	$4.2 \cdot 10^3$	6.3
LGBM	DG	DG-DET	**108%**	**73%**	**6.3**	202	44%	$4.0 \cdot 10^3$	2.3
	RB	RB-STO	113%	78%	44.5	216	46%	$4.5 \cdot 10^3$	5.0
NN	DG	DG-DET	109%	122%	6.8	233	31%	$4.6 \cdot 10^3$	6.5
	RB	RB-STO	109%	**77%**	41.6	**170**	**50%**	$\mathbf{3.3} \cdot 10^3$	2.5
	Gaussian	PM-STO	328%	114%	9.4	381	61%	$8.7 \cdot 10^3$	7.7
	Neg.-Bin.	PM-STO	**103%**	81%	6.6	**176**	29%	$4.1 \cdot 10^3$	4.2
	Stud. T	PM-STO	120%	82%	**6.2**	206	36%	$4.0 \cdot 10^3$	4.5

Table 2
KPIs of the evaluation, averaged across the assembly groups. Spread of R_j amongst products $j \in J$, quantified by the IQR.

Method	Distribution	Group	Forecast $\hat{d}$	Production X	Fulfilled F	Margins M	Costs C	Objective O	IQR of R_j
ETS	DG	DG-DET	100%	86%	**76%**	145%	75%	70 %	53.5
	RB	RB-STO	97%	70%	63%	120%	53%	68 %	40.4
LGBM	DG	DG-DET	73%	72%	65%	124%	46%	78 %	40.5
	RB	RB-STO	56%	48%	46%	88%	17%	71 %	30.9
NN	DG	DG-DET	90%	86%	**77%**	147%	56%	91 %	63.8
	RB	RB-STO	83%	68%	64%	123%	24%	**99 %**	48.9
	Gaussian	PM-STO	99%	5%	5%	9%	1%	8 %	5.6
	Neg.-Bin.	PM-STO	113%	73%	68%	131%	31%	**100%**	40.1
	Stud. T	PM-STO	64%	42%	41%	79%	5%	74 %	31.3

Table 3
Objective values by assembly group; normalised to the highest value of each assembly group.

Group	DG-DET	RB-STO	DG-DET	RB-STO	DG-DET	RB-STO	PM-STO	PM-STO	PM-STO
Assembly group	**ETS-DG**	**ETS-RB**	**LGBM-DG**	**LGBM-RB**	**NN-DG**	**NN-DG**	**NN-Gaussian**	**NN-Neg.-Bin.**	**NN-Stud. T**
1	42 %	88 %	32 %	56 %	61 %	**100%**	76 %	**98%**	69 %
2	67 %	**100%**	50 %	52 %	71 %	**99%**	91 %	**99%**	78 %
3	77 %	85 %	65 %	77 %	72 %	**100%**	74 %	**100%**	86 %
4	41 %	**100%**	62 %	67 %	76 %	93 %	91 %	**100%**	68 %
5	5 %	71 %	0 %	88 %	27 %	**95%**	72 %	**100%**	86 %
6	58 %	65 %	73 %	84 %	66 %	86 %	93 %	**100%**	79 %
7	48 %	50 %	65 %	71 %	58 %	85 %	**99%**	**100%**	82 %
8	59 %	75 %	68 %	92 %	55 %	95 %	82 %	**100%**	94 %
9	59 %	**96%**	55 %	74 %	30 %	79 %	84 %	**100%**	81 %
10	49 %	91 %	76 %	85 %	71 %	**98%**	82 %	**100%**	64 %
11	-30 %	80 %	58 %	**100%**	42 %	**99%**	86 %	**97%**	93 %
12	67 %	89 %	70 %	93 %	73 %	**98%**	87 %	**100%**	86 %
13	80 %	91 %	62 %	68 %	75 %	57 %	**100%**	89 %	48 %
14	-25 %	93 %	35 %	**100%**	-18 %	81 %	76 %	**100%**	83 %
15	40 %	87 %	69 %	**100%**	63 %	95 %	74 %	**98%**	82 %
16	-6 %	**100%**	39 %	86 %	-5 %	60 %	**99%**	84 %	58 %
17	-111 %	39 %	-4 %	**96%**	53 %	**97%**	66 %	**100%**	92 %

(63%). The poorest results are seen with LGBM-RB (46%) and NN-Stud. T (41%), and significantly lower, the NN-Gaussian (5%) approach. Although the fulfilled demand, or β-service level, is a typical goal for SC practitioners, it is merely an indicator and not the primary objective of our optimisation.

With 147% and 145% respectively, NN-DG and ETS-DG achieve the largest profit margins. However, while the latter does this at the expense of the largest observed inventory costs, NN-DG matches actual demand much better, keeping inventory levels lower and achieving the third highest objective. This pattern repeats for other groups of approaches. ETS-RB, LGBM-DG, NN-RB, and NN-Neg.-Bin. all achieve similar profit margins, but while the first two require substantial levels of inventory to cover demand, the second two operate much more efficiently, eventually achieving a higher objective. LGBM-RB and NN-Stud. T fully commit to the latter strategy, keeping inventory levels considerably lower than all other approaches. While it saves holding costs, it leads to the lowest fulfilled demand and profit rates, which is eventually an unfavourable tradeoff.

Analysis on assembly group level. In Table 3, we show the objective function values evaluated at the assembly group level. The primary takeaway is that the trends observed at the aggregated level are also evident at the assembly group level for the majority of assembly groups. For 15 out of 17 assembly groups, the NN-Neg.-Bin. method is either the best or within 3% of the best solution, consolidating its strong performance on the aggregate level. For the NN-DG method, this holds true for 7 out of 17 groups. Notably, for three assembly groups, the NN-Gaussian, ETS-DG, and LGBM-RB methods perform exceptionally well, being the best or within 3% of the top solution, but these groups do not overlap. This suggests that for further improvements, the approaches, primarily the forecast methods, should be tailored to specific assembly groups or even individual products. The remaining KPIs at the assembly group level can be found in the Supplementary material A.

Correlation of forecast accuracy and objective. After analysing the forecast errors and the optimisation results independently, it is crucial to consider their correlation. We evaluated the correlation between various forecast error metrics and the relative objective function value R_j using Spearman's rank correlation coefficient (Spearman's ρ), which spans from -1 to 1. In this context, a perfect correlation would be -1, indicating a linear relationship in which a reduction in the forecast error metric corresponds to an improvement in the value of the objective function. Table 4 presents the findings of the correlation analysis.

Unsurprisingly, the relative forecast error metrics MAPE and SMAPE exhibit moderate to strong negative correlations of -0.53 and -0.69 on average, respectively. This indicates that across all models (with the exception of the poorly performing NN-Gaussian), an increase in

Table 4
Correlation between the forecast error metrics and R_j measured by Spearman's ρ.

Model	Distribution	Group	MAPE	SMAPE	MASE	RMSE	$COV_{q=0.5}$	$QL_{q=0.5}$	$\bar{\rho}$
ETS	DG	DG-DET	-0.64	-0.72	-0.03	0.00	-0.44	-0.01	-0.31
	RB	RB-STO	-0.54	-0.75	-0.32	-0.10	-0.23	-0.07	-0.34
LGBM	DG	DG-DET	-0.57	-0.69	-0.02	-0.09	-0.29	0.01	-0.28
	RB	RB-STO	-0.54	**−0.82**	-0.29	-0.12	0.00	-0.07	-0.31
NN	DG	DG-DET	-0.66	**−0.79**	0.05	-0.04	-0.12	0.08	-0.25
	RB	RB-STO	-0.64	**−0.80**	-0.18	0.02	-0.08	0.07	-0.27
	Gaussian	PM-STO	0.08	-0.24	0.05	0.17	0.26	0.18	0.08
	Neg.-Bin.	PM-STO	-0.68	-0.67	-0.05	-0.04	-0.51	0.08	-0.31
	Stud. T	PM-STO	-0.55	-0.76	-0.36	0.05	-0.36	0.11	-0.31
		$\bar{\rho}$	-0.53	-0.69	-0.13	-0.02	-0.20	0.04	-0.25

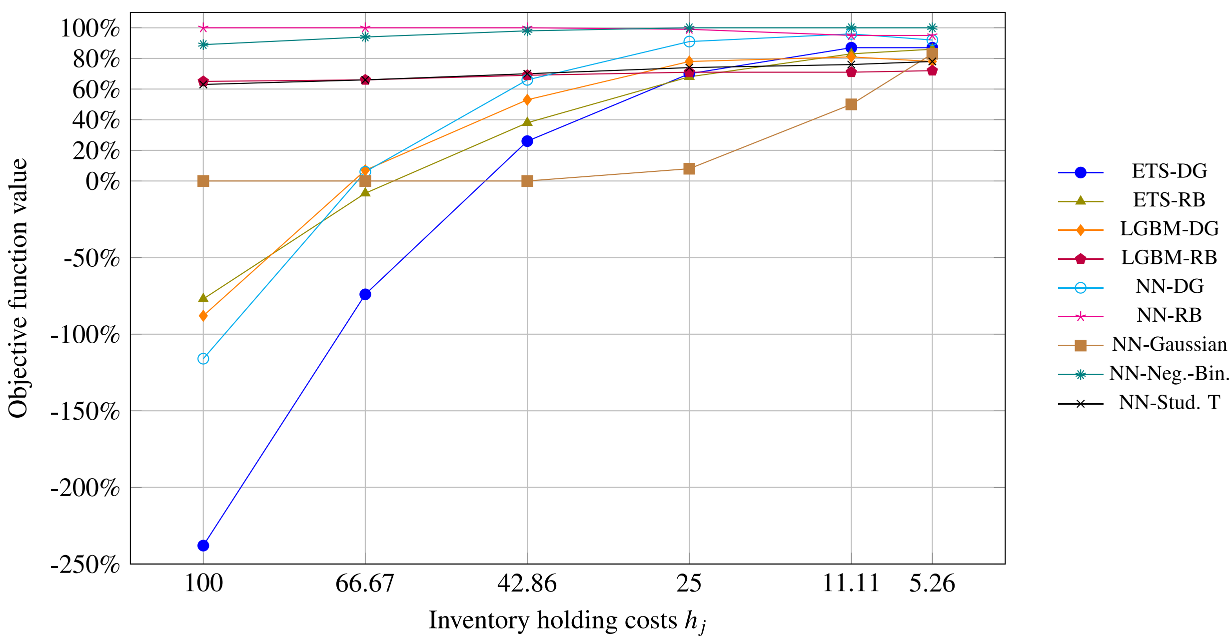

Fig. 5. Objectives function values by inventory holding costs h_j, normalised to maximum per target β-service level.

relative prediction accuracy improves relative net profit. The correlations for all other forecast error metrics are approximately 0, suggesting that they are not correlated with the objective. The reason is that one of the observed quantities is relative (relative profit), while the other is absolute (absolute errors). Products with larger demand volumes often exhibit more absolute volatility and hence cause a larger absolute forecast error. While absolute profit also increases simply due to the larger absolute volumes, this is not necessarily the case for relative profit, leading to the absence of correlation. Since the same logic applies to relative errors and absolute profit, this shows that for groups with different target volumes, both the error metric and objective function must be expressed in relative terms.

Other service levels. Finally, in Fig. 5, we vary the inventory holding costs h_j to examine the impact on model performance. At $h_j = 100$, the three DG forecasts ETS-DG, NN-DG, and LGBM-DG, along with ETS-RB, perform so poorly that the objective function values become severely negative. This situation improves significantly for $h_j = 66.67$ and $h_j = 42.86$, although these approaches still rank among the worst performing. The best results for $h_j = 100$, $h_j = 66.67$, and $h_j = 42.86$ are achieved by NN-RB and NN-Neg.-Bin., followed by LGBM-RB and NN-Stud. T, which interestingly show almost identical values across all inventory holding cost variations. The demand forecast by NN-Gaussian is so low that for $h_j = 100$, $h_j = 66.67$, and $h_j = 42.86$, the optimiser decides not to produce anything. In the chosen problem formulation in this paper, this implies that there are neither profit margins nor inventory holding costs, leading to an objective function value of 0. For $h_j = 11.11$, and $h_j = 5.26$, the values of all approaches converge. This is because, with such low inventory holding costs, the optimiser decides to produce as much as possible. Given the limited production capacity, it is not feasible for the methods to outperform others using more accurate forecasts; they can only surpass others if the distribution of forecasts among the products is more precise.

Computational effort. To conclude this section, we compare the approaches based on the time complexity of solving the optimisation problem. On the prediction side, there is no notable time difference between probabilistic and point forecasts, other than the additional effort required to process larger datasets. In our numerical study, the deterministic optimisation model was solved in an average of 2.4 s. In contrast, the stochastic model, which incorporates demand variability, required an average of 8 min and 39 s. These computations were carried out using the `Gurobi` solver with 10 threads on a shared compute pool with servers of different capacities. While the stochastic model took longer, the solution time of under ten minutes is still practical for real-world applications, particularly given its superior results.

6. Conclusion

Our study has primarily focused on different approaches to combining probabilistic forecasting, enhanced by state-of-the-art ML methods. These approaches were then used with stochastic optimisation to tackle the MPS problem. We have demonstrated the efficacy of novel techniques through extensive experimentation. The following key points summarise our contributions and the implications of our research:

- *Established Effectiveness of Probablistic Forecasting with Stochastic Optimisation:* The RB-STO and PM-STO approaches yielded the best objective function values in the optimisation, demonstrating the superiority of combining probabilistic forecasts with stochastic optimisations. These innovative approaches outperformed those relying solely on DG forecasts with deterministic optimisation. By experimenting with the combination of probabilistic forecasting with stochastic optimisation, we also count the following sub-points among the contributions of this paper: (1) Pioneering the integration of probabilistic forecasting to solve the MPS problem and (2) Confirming the robustness of stochastic optimisation.
- *Challenged the Use of the Gaussian Distribution:* The results of our experiment are also a crucial challenge to the established norm of assuming a Gaussian distribution, which is foundational for many traditional approaches. We realised that instead of blindly using a Gaussian distribution, it was critical to choose a distribution that allows the NN to accurately approximate the underlying data. By doing so, we observed a better performance in solving the MPS problem.
- *Confirmed Superiority of ML Forecasting Methods:* Our results highlight that the top ML predictions surpass the non-ML in capturing non-stationary and seasonal demand patterns. Moreover, they maintain a minimal variation in the relative objective function value, as indicated by the IQR, demonstrating increased robustness.
- *Breadth and Scope of Experiment Results:* This point is based on two aspects that demonstrate the paper's considerable depth: the extensive range of real-world data; and the evaluation of multiple forecasting methods.
  Firstly, using the huge range of seventeen datasets allowed us to conduct our experiment within real-world conditions. The inherent limitation of real-world data is its obfuscation of true demand distributions; each product has only one sales value available, equivalent to demand only in the absence of stockouts or other impediments. By including such limitations, we are able to provide more comprehensive and truer-to-reality results.
  Secondly, with a total of nine approaches being used, we conducted a wide-ranging analysis of the benefits and limitations of using the DG-DET, RB-STO, and PM-STO methods.
  By combining an ambitious range of approaches with large and multiple datasets, we were able to make robust and reliable observations. We observed consistent patterns in our results when assessed at the assembly group level. However, certain data sets show excellent performance with the ETS with a residual-based distribution (ETS-RB) method. Thus, we conclude that, while general guidelines can be provided, experimenting with different approaches is crucial to achieve the best results.

Practical implications. This research bridges the gap between advanced forecasting methodologies and their practical application in SCs. By using real-world datasets, our approach tackles the core inventory management challenge of accurately managing over- and underage costs without relying on proxy metrics like service levels. We provide a roadmap for practitioners to incorporate probabilistic forecasting seamlessly for diverse operational contexts, offering a clear pathway from data preparation to optimisation.

While the guidance provided by this research allows the individual to seek out the correct approach for themselves, we also provide evidence that stochastic optimisations yields generally superior results. Owing this superiority to the comprehensive analysis of diverse scenarios, the research underlines the importance of accurately predicted distributions. Future practitioners should pay heed to this finding.

When using point forecasts, it is necessary for most SC scenarios to modify negative demand values to 0 after the prediction through post-processing. This approach is suboptimal, as manually adjusting the forecast values introduces a two-step forecasting method, potentially impacting the accuracy of the forecasts in unpredictable manners. Therefore, it would be better for demand forecasting to choose discrete and nonnegative probability distributions, with parameters predicted by NNs. Fortunately, training NNs for probabilistic forecasts requires roughly the same effort as training for point forecasts. Although the management and processing of such data volumes does require more storage and computational power, modern big data technologies can mitigate this drawback.

To guide practitioners, we recommend first using the simpler RB-STO approaches. By using data from previous forecasts stored in demand planning software, these approaches provide results comparable to the more complicated PM-STO approaches. However, the results are only similar in cases where long time series are available. As such, we advise using PM-STO approaches that work on both long and short time series. Additionally, training NNs for probabilistic forecasting on historical data is more efficient than retraining point forecasts in a rolling manner on the whole history. In the latter option, one risks incurring huge computational expenses, for instance while testing new features. Instead, the former option allows data scientists to shorten learning and optimisation cycles and thereby improve more rapidly. This is crucial, as the outcomes of the MPS are significantly dependent on the accuracy of the forecasts. A further benefit of PM-STO approaches is their quick adjustment to major demand pattern changes, such as those experienced during the pandemic, as the NNs can learn during training from features to disregard outdated observations.

Future direction. This paper provides numerous opportunities for further exploration. A promising avenue involves evaluating the outlined approaches with a rolling planning horizon. Other interesting directions would be to alternate capacity constraints to simulate both under- and overage situations. An alternative to the RB-STO group of approaches could be fitting distributions on the forecast residuals and subsequently sampling them. The ML-based forecast models could further improve with more sophisticated feature engineering.

Our work lays the foundation for broader exploration into adaptive SC solutions. Exploring alternative forecasting models and different distributions for the PM-STO group of approaches could enhance overall results. In particular, this paper did not delve into the complexities of probabilistic lead times or supply uncertainty, both of which add difficulty to the problem but could significantly enhance robustness if addressed correctly.

We strongly encourage research based on real datasets, avoiding simplistic assumptions about demand distributions. This convergence of the research areas operations research and time series forecasting is essential. Only through this integration can the outcomes of these research topics become relevant for practical applications. Ultimately, this research enhances the resilience of modern SCs by providing a flexible and robust framework for managing uncertainty, empowering organisations to maintain efficiency and agility in the face of unforeseen challenges.

CRediT authorship contribution statement

Sebastian Paull: Writing – original draft, Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualization. **Alexander Bubak:** Conceptualization, Validation, Writing – review & editing. **Heiner Stuckenschmidt:** Conceptualization, Writing – review & editing, Supervision.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used the softwares *Writefull* as well as *ChatGPT* to improve the language and readability of their paper, mainly for spelling checks, synonym suggestions, and sentence rephrasing. Furthermore, *TexGPT* and *ChatGPT* were used to generate ideas for the title and abstract. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

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

Acknowledgement

We would like to extend our sincere gratitude to Prof. Dr. Fleischmann of the University of Mannheim for his valuable insights and recommendations during the initial phases of this study.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.eswa.2025.126586.

References

Alexandrov, A., Benidis, K., Bohlke-Schneider, M., Flunkert, V., Gasthaus, J., Januschowski, T., Maddix, D. C., Rangapuram, S., Salinas, D., Schulz, J., Stella, L., Türkmen, A. C., & Wang, Y. (2020). Gluonts: Probabilistic and neural time series modeling in python. *Journal of Machine Learning Research*, *21*, http://dx.doi.org/10.5555/3455716.3455832.

Amazon (2023). Available models. URL https://ts.gluon.ai/stable/getting_started/models.html.

Bagheri, F., Demartini, M., Arezza, A., Tonelli, F., Pacella, M., & Papadia, G. (2022). An agent-based approach for make-to-order master production scheduling. *Processes*, *10*(921), http://dx.doi.org/10.3390/pr10050921.

Ban, G.-Y. (2020). Confidence intervals for data-driven inventory policies with demand censoring. *Operations Research*, *68*(2), 309–326. http://dx.doi.org/10.1287/opre.2019.1883.

Ban, G. Y., & Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research*, *67*(1), 90–108. http://dx.doi.org/10.1287/opre.2018.1757.

Berry, W., Vollmann, T., & Whybark, D. (1979). *Master production scheduling: principles and practice* (5th ed.). American Production and Inventory Control Society.

Bertsimas, D., & McCord, C. (2019). From predictions to prescriptions in multistage optimization problems.

Birge, J. R., & Louveaux, F. (2011). Introduction to stochastic programming. *Springer science & business media*. (2).

Bollapragada, R., & Rao, U. S. (2006). Replenishment planning in discrete-time, capacitated, non-stationary, stochastic inventory systems. *IIE Transactions (Institute of Industrial Engineers)*, *38*(7), 605–617. http://dx.doi.org/10.1080/07408170600634533.

Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package). *Neurocomputing*, *307*, 72–77. http://dx.doi.org/10.1016/j.neucom.2018.03.067.

Chu, S. C. (1995). A mathematical programming approach towards optimized master production scheduling. *International Journal of Production Economics*, *38*(2–3), 269–279. http://dx.doi.org/10.1016/0925-5273(95)00015-G.

Ciarallo, F. W., Akella, R., & Morton, T. E. (1994). A periodic review, production planning model with uncertain capacity and uncertain demand—Optimality of extended myopic policies. *Management Science*, *40*(3), 320–332. http://dx.doi.org/10.1287/mnsc.40.3.320.

DeCroix, G. A., & Arreola-Risa, A. (1998). Optimal production and inventory policy for multiple products under resource constraints. *Management Science*, *44*(7), 950–961. http://dx.doi.org/10.1287/mnsc.44.7.950.

Englberger, J., Herrmann, F., & Manitz, M. (2016). Two-stage stochastic master production scheduling under demand uncertainty in a rolling planning environment. *International Journal of Production Research*, *54*(20), 6192–6215. http://dx.doi.org/10.1080/00207543.2016.1162917.

Englberger, J., Herrmann, F., & Manitz, M. (2022). Master production scheduling with scenario-based capacity-load factors in a rolling planning environment. *Logistics Research*, *15*(12), 1–16.

Fallahi, A., Mousavian Anaraki, S. A., Mokhtari, H., & Niaki, S. T. A. (2024). Blood plasma supply chain planning to respond COVID-19 pandemic: a case study. *Environment, development and sustainability*: *vol. 26*, (pp. 1965–2016). Springer Netherlands, http://dx.doi.org/10.1007/s10668-022-02793-7.

Fan, Y., Stevenson, M., & Li, F. (2020). Supplier-initiating risk management behaviour and supply-side resilience: the effects of interpersonal relationships and dependence asymmetry in buyer-supplier relationships. *International Journal of Operations and Production Management*, *40*(7–8), 971–995. http://dx.doi.org/10.1108/IJOPM-06-2019-0497.

Feng, K., Rao, U. S., & Raturi, A. (2011). Setting planned orders in master production scheduling under demand uncertainty. *International Journal of Production Research*, *49*(13), 4007–4025. http://dx.doi.org/10.1080/00207543.2010.495955.

Feng, Q., & Shanthikumar, J. G. (2018). How research in production and operations management may evolve in the era of big data. *Production and Operations Management*, *27*(9), 1670–1684. http://dx.doi.org/10.1111/poms.12836.

Fleischmann, B., & Meyr, H. (2003). Planning hierarchy, modeling and advanced planning systems. *Handbooks in Operations Research and Management Science*, *11*(C), 455–523. http://dx.doi.org/10.1016/S0927-0507(03)11009-2.

Forel, A., & Grunow, M. (2021). Stochastic programming in master production scheduling: overcoming barriers to industry application. URL https://www.ot.mgt.tum.de/fileadmin/w00cjr/scm/ScientificPublications/forel_stoch_mps_working_paper.pdf.

Gamboa, J. C. B. (2017). Deep learning for time-series analysis. http://dx.doi.org/10.48550/arXiv.1701.01887.

Härdle, W. K., Gentle, J. E., & Mori, Y. (2005). Handbook of computational statistics. In *Technometrics, no. 3* (pp. 383–384). http://dx.doi.org/10.1198/tech.2005.s314.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). (pp. 1–758). New York: Springer New York, http://dx.doi.org/10.1007/b94608.

Huang, K., Ying, X., Liu, X., Wu, D., Yang, C., & Gui, W. (2024). One network fits all: A self-organizing fuzzy neural network based explicit predictive control method for multimode process. *IEEE Transactions on Fuzzy Systems*, *PP*, 1–14. http://dx.doi.org/10.1109/TFUZZ.2024.3362792.

Huber, J., Müller, S., Fleischmann, M., & Stuckenschmidt, H. (2019). A data-driven newsvendor problem: From data to decision. *European Journal of Operational Research*, *278*(3), 904–915. http://dx.doi.org/10.1016/j.ejor.2019.04.043.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *27*(3), 1–22. http://dx.doi.org/10.18637/jss.v027.i03.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). Forecasting with exponential smoothing: The state space approach. (p. 350). http://dx.doi.org/10.1007/978-3-540-71918-2.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems*: *vol. 2017-Decem*, (pp. 3147–3155). Long Beach.

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. In *3rd International conference on learning representations, ICLR 2015 - conference track proceedings* (pp. 1–15).

Koh, S. C., Saad, S. M., & Jones, M. H. (2002). Uncertainty under MRP-planned manufacture: Review and categorization. *International Journal of Production Research*, *40*(10), 2399–2421. http://dx.doi.org/10.1080/00207540210136487.

Körpeolu, E., Yaman, H., & Selim Aktürk, M. (2011). A multi-stage stochastic programming approach in master production scheduling. *European Journal of Operational Research*, *213*(1), 166–179. http://dx.doi.org/10.1016/j.ejor.2011.02.032.

Krüger, T., & Koberstein, A. (2023). Plant-Wide Master Production Scheduling in the Automotive Industry: A MILP-approach and a simulation study. SSRN.

Lee, T. S., & Adam, E. E. (1986). Forecasting error evaluation in material requirements planning (MRP) production-inventory systems. *Management Science*, *32*(9), 1186–1205. http://dx.doi.org/10.1287/mnsc.32.9.1186.

Li, Q., Wu, Z., Wen, Z., & He, B. (2020). Privacy-preserving gradient boosting decision trees. In *AAAI 2020 - 34th AAAI conference on artificial intelligence* (pp. 784–791). http://dx.doi.org/10.1609/aaai.v34i01.5422.

Liao, Y., Banerjee, A., & Yan, C. (2011). A distribution-free newsvendor model with balking and lost sales penalty. *International Journal of Production Economics*, *133*(1), 224–227. http://dx.doi.org/10.1016/j.ijpe.2010.04.024.

Lin, N.-P., & Krajewski, L. (1992). A model for master production scheduling in uncertain environments. *Decision Sciences*, *23*(4), 839–861. http://dx.doi.org/10.1111/j.1540-5915.1992.tb00422.x.

Liu, Y., Liu, W., Lin, S., Yang, C., Huang, K., & Shen, Z. J. M. (2023). Data-driven raw material robust procurement for non-ferrous metal smelter under price and demand uncertainties. *IEEE Transactions on Automation Science and Engineering*, *PP*, 1–14. http://dx.doi.org/10.1109/TASE.2023.3319390.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74. http://dx.doi.org/10.1016/j.ijforecast.2019.04.014.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, (xxxx), http://dx.doi.org/10.1016/j.ijforecast.2021.11.013.

Makridakis, S., Spiliotis, E., Assimakopoulos, V., Chen, Z., Gaba, A., Tsetlin, I., & Winkler, R. L. (2021). The M5 uncertainty competition: Results, findings and conclusions. *International Journal of Forecasting*, http://dx.doi.org/10.1016/j.ijforecast.2021.10.009.

Martín, A. G., Díaz-Madroñero, M., & Mula, J. (2020). Master production schedule using robust optimization approaches in an automobile second-tier supplier. *Central European Journal of Operations Research*, *28*(1), 143–166. http://dx.doi.org/10.1007/s10100-019-00607-2.

McCormick, R. (1980). Master Production Scheduling–the fundamental concepts. In *Proceedings of 23rd APICS annual international conference* (pp. 400–403).

Metters, R. (1997). Production planning with stochastic seasonal demand and capacitated production. *IIE Transactions (Institute of Industrial Engineers)*, *29*(11), 1017–1029. http://dx.doi.org/10.1080/07408179708966420.

Mohammed, A. R., Sleem, A. A., & Abdel-Aal, M. A. (2023). Linear and non-linear models for master scheduling of dynamic resources product mix. *Proceedings of International Conference on Computers and Industrial Engineering, CIE*, *2*, 917–928.

Müller, S., Huber, J., Fleischmann, M., & Stuckenschmidt, H. (2020). Data-driven inventory management under customer substitution. (pp. 1–29).

Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(3), 1181–1191. http://dx.doi.org/10.1016/j.ijforecast.2019.07.001.

Serrano-Ruiz, J. C., Mula, J., & Poler, R. (2021). Smart master production schedule for the supply chain: A conceptual framework. *Computers*, *10*(12), http://dx.doi.org/10.3390/computers10120156.

Stadtler, H., & Kilger, C. (2005). *Supply Chain Management and Advanced Planning*. Springer Berlin Heidelberg New York, http://dx.doi.org/10.1007/978-3-540-74512-9.

Tang, O., & Grubbström, R. W. (2002). Planning and replanning the master production schedule under demand uncertainty. *International Journal of Production Economics*, *78*(3), 323–334. http://dx.doi.org/10.1016/S0925-5273(00)00100-6.

Trost, M., Claus, T., & Herrmann, F. (2023). Master production scheduling with consideration of utilization-dependent exhaustion and capacity load. *Sustainability (Switzerland)*, *15*(8), 1–20. http://dx.doi.org/10.3390/su15086816.

Vargas, V., & Metters, R. (2011). A master production scheduling procedure for stochastic demand and rolling planning horizons. *International Journal of Production Economics*, *132*(2), 296–302. http://dx.doi.org/10.1016/j.ijpe.2011.04.025.

Vogel, T., Almada-Lobo, B., & Almeder, C. (2017). Integrated versus hierarchical approach to aggregate production planning and master production scheduling. *OR Spectrum*, *39*(1), 193–229. http://dx.doi.org/10.1007/s00291-016-0450-2.

Vollmann, T. E., Berry, W. L., & Whybark, D. C. (1997). *Manufacturing planning and control systems* (4th ed.). Irwin/McGraw-Hill..

Zhang, Z., & Jung, C. (2021). GBDT-MO: Gradient-Boosted Decision Trees for Multiple Outputs. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(7), 3156–3167. http://dx.doi.org/10.1109/TNNLS.2020.3009776.

Zipkin, P. (2000). *Foundations of inventory management*. New York: McGraw-Hill.

Glossary

Acronyms

COV: Coverage
CV: Coefficient of Variation
DG: Degenerate
DG-DET: Point forecast with a degenerate distribution solved by deterministic optimisation
DNN: Deep Neural Network
ERP: Enterprise Resource Planning
ETS: Exponential Smoothing
ETS-DG ETS: with a degenerate distribution
ETS-RB ETS: with a residual-based distribution
Gamma: Gamma
Gaussian: Gaussian
GBDT: Gradient Boosting Decision Tree
IEO: Integrated Estimation and Optimisation
IQR: Interquartile range
KPI: Key Performance Indicator
LGBM: Light Gradient-Boosting Machine
LGBM-DG: Light Gradient-Boosting Machine with a degenerate distribution
LGBM-RB: LGBM with a residual-based distribution
LSTM NN: Long Short-Term Memory Neural Network
MAE: Mean Absolute Error
MAPE: Mean Absolute Percentage Error
MASE: Mean Absolute Scaled Error
MILP: Mixed-Integer Linear Program
ML: Machine Learning
MP: Master Planning
MPS: Master Production Scheduling
Neg.-Bin.: Negative Binomial
NN: Neural Network
NN-DG: Neural Network with a degenerate distribution
NN-Gaussian: Neural Network with a Gaussian distribution
NN-Neg.-Bin.: Neural Network with a Negative Binomial distribution
NN-RB: Neural Network with a residual-based distribution
NN-Stud. T: Neural Network with a Student's T distribution
OR: Operations Research
PM-STO: Probabilistic forecast with a parametric distribution solved by stochastic optimisation
QL: Quantile Loss
RB: Residual-based
RB-STO: Point forecast with a residual-based distribution solved by stochastic optimisation
RMSE: Root Mean Square Error
SC: Supply Chain
SEC: Seasonal Error Component
SEO: Separated Estimation and Optimisation
SMAPE: Symmetric Mean Absolute Percentage Error
Spearman's: ρ Spearman's rank correlation coefficient
Stud. T: Student's T