Computers & Industrial Engineering 210 (2025) 111573

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

An information-sharing and cost-aware custom loss machine learning framework for 3PL supply chain forecasting

Matteo Gabellini [a,*], Francesca Calabrese [b], Francesco Gabriele Galizia [a], Michele Ronchi [a], Alberto Regattieri [a]

[a] *Department of Industrial Engineering, University of Bologna, viale del risorgimento 4, Bologna 61121, Italy*
[b] *Department of Industrial Engineering, University of Trento, Via Sommarive, 9, Trento, Italy*

ARTICLE INFO

Keywords:
3PL
Supply chain forecasting
Less than truckload
Artificial intelligence
Machine learning

ABSTRACT

Supply chain forecasting methods have traditionally been developed from the perspective of manufacturing companies, which historically held dominant roles within supply chain dynamics. However, the growing importance of third-party logistics providers (3PLs) calls for forecasting approaches tailored to their unique operational needs. This paper presents a novel forecasting framework specifically designed for 3PLs to accurately predict the truck space required for transporting their customers' products. Unlike conventional methods, the proposed approach directly forecasts truck space demand by utilizing data obtained through information-sharing technologies to train machine learning models. Furthermore, a customized loss function is introduced for the first time, explicitly accounting for the asymmetric costs associated with overestimating and underestimating truck utilization. The framework was validated through a real-world case study involving a 3PL operating in the food sector. The results demonstrated significant improvements over traditional forecasting techniques, underscoring the benefits of integrating machine learning, information sharing, and a tailored loss function to enhance both predictive accuracy and cost-efficiency.

1. Introduction

In recent years, 3PLs—responsible for outsourced logistics services such as transportation, warehousing, and distribution—have become increasingly crucial within supply chains (Bian et al., 2021; Lieb & Miller, 2002). The global 3PL market was valued at approximately USD 1095.85 billion in 2023, with an estimated compound annual growth rate of 8.1 % through 2030 (Grand View Research, 2023). As 3PL providers play a more direct role in supply chain efficiency, even minor disruptions in their operations can trigger cascading delays and increased costs across all stakeholders (Abbasi et al., 2021; Fabbe-Costes et al., 2009; Yang & Yu, 2019). This growing complexity has underscored the importance of accurate forecasting as a value-added function in 3PL operations (Wolny & Kmiecik, 2025), particularly for providers that utilize Less-Than-Truckload (LTL) transportation strategies (Qiao et al., 2020; Saka & Salman, 2024; Tyan et al., 2003; Wong et al., 2018; Yoon et al., 2016). The LTL model consolidates smaller shipments from multiple customers into trucks, enabling greater cost efficiency (Kay & Warsing, 2009). However, its operational complexity determines that daily operations in this sector are resource-intensive and require timely and precise data to support decision-making and optimize operational efficiency (Lai et al., 2008). Additionally, these providers often operated within supply chains where power imbalances favored larger stakeholders with greater negotiating leverage. The ability to anticipate fluctuations in demand through accurate forecasting thus determined the capacity of 3PL providers to adapt and compete effectively. Due to their comparatively weaker bargaining power, 3PL providers lacked the authority to compel other supply chain members to share more complete or timely information. As a result, they were often forced to respond to demand variability reactively, relying heavily on their own forecasting capabilities to mitigate uncertainty and maintain service levels.

While the importance of 3PL and forecasting activities within these supply chain entities has thus grown, developing effective forecasting methods remains a significant challenge in this sector (Kmiecik, 2022). Several interrelated factors contribute to this complexity. One of the most critical aspects is that, in this sector, demand forecasting is fundamentally tied to space occupation rather than simply the number of products. Indeed, decision-making processes in 3PL operations

* Corresponding author.
E-mail address: matteo.gabellini5@unibo.it (M. Gabellini).

https://doi.org/10.1016/j.cie.2025.111573
Received 7 April 2025; Received in revised form 11 September 2025; Accepted 29 September 2025
Available online 5 October 2025

typically rely on metrics related to space utilization, as capacity planning, load optimization, and route efficiency depend on accurately predicting how goods will physically occupy available truck space. However, forecasting space occupation is particularly challenging due to the possible overlapping of pallets, which must be accurately modeled to avoid inefficiencies (Park & Kim, 2010; Silva et al., 2016; Terno et al., 2000). Traditional forecasting methods, however, often fail to consider these spatial considerations, leading to suboptimal load planning and increased costs. Additionally, the scale of forecasting required in 3PL operations further complicates the process. Unlike manufacturing firms, which need to generate forecasts related to suppliers or customers, 3PL providers connect producers to retailers, thus requiring forecasts for each pair (producer, retailer), significantly increasing the number of forecasts that need to be made. Automated systems are thus essential to handle complexity and high volume of forecasts effectively. Lastly, beyond the sheer number of forecasts required, another critical challenge arises from the asymmetric impact of forecast errors (Sanders & Graman, 2009). In 3PL logistics, forecast errors have unequal cost consequences. Overestimating the space to reserve generally leads to underutilized transport resources and higher holding or idle-capacity costs, whereas underestimating it can result in unfulfilled shipments, lost sales, and cascading supply chain disruptions. Prior research on inventory and forecasting systems shows that these error types are not equally costly: holding costs associated with overestimation increase gradually, while the costs of lost sales from underestimation escalate much more sharply. Consistent with this view, (Dress et al., 2018) demonstrate that incorporating asymmetric cost functions into forecasting models yields substantially lower decision-related costs compared to symmetric approaches. Taken together, these findings suggest that models which penalize over- and underestimation equally are misaligned with the operational realities of 3PL forecasting, where space optimization critically depends on recognizing the asymmetric nature of forecast error costs. Additionally, recent advancements in deep learning further illustrate the value of asymmetric approaches in logistics contexts. For instance, Che et al., (2023) developed a CNN-based model for precise robot position prediction in logistics environments, employing an Asymmetric Gaussian loss function to better capture the skewed nature of sensor data and achieve high predictive accuracy. Their findings highlight the broader importance of integrating advanced deep learning techniques and tailored loss functions into logistics forecasting and automation, reinforcing the argument that 3PL forecasting models must account for asymmetry and complexity to improve decision-making and operational efficiency.

A significant gap thus emerges in the capabilities of existing approaches developed in supply chain forecasting to cover these challenges. Due to the minor weight assumed by 3PL providers in the supply chain, past methods have been designed from manufacturing firms' perspective. As a result, most existing supply chain forecasting studies do not forecast space utilization metrics or consider critical elements such as pallet overlapping or the distinct operational risks posed by asymmetric forecast errors. Moreover, while information sharing has long been recognized as valuable in supply chain operations, limited research has investigated how advanced data collection techniques can be integrated into a holistic predictive framework that enhances forecasting accuracy in 3PL operations. Based on this evidence, this study argues that addressing these shortcomings and adopting the perspective of 3PL providers is essential for advancing the field of supply chain forecasting. The primary aim of this paper is to support 3PL forecasting activities by introducing a novel, holistic framework for truck space reservation prediction. The main contributions of this study are as follows:

- The proposed framework for 3PLs truck space reservation prediction extends beyond traditional predictive models by incorporating comprehensive data collection methodologies that feed into an integrated system for operative decision-making.
- The proposed framework leverages Machine learning (ML) models for the first time to explicitly learn pallet overlapping possibilities and to differentially balance the consequences of overestimating and underestimating the reserved truck space—thereby mitigating the asymmetric impacts of forecast errors relying on a custom loss.
- The practical applicability of the proposed approach is demonstrated through a real case study, where it is rigorously benchmarked against alternative data collection and model design, underscoring its potential to significantly enhance forecasting accuracy and operational efficiency in the LTL domain.

The remainder of the paper is organized to initially provide a comprehensive review of the literature in Section 2, where existing works are examined and key gaps in current predictive models for 3PL providers are identified. Then, Section 3 introduces the novel forecasting framework and the experiments designed to test its effectiveness in a real case study against several benchmark models. Section 4 presents the empirical results derived from this case study, which are then critically discussed in Section 5. Finally, Section 6 concludes the paper with a summary of the key findings and suggestions for future research avenues.

2. Literature review

This Section begins by exploring key trends in supply chain forecasting, focusing on the increasing emphasis on information sharing, the integration of ML technologies, and the rise of end-to-end predictive models as major advancements in the field. It then reviews the existing literature on predictive models tailored for 3PL operators, revealing a limited uptake of these innovations within the sector. This gap highlights the originality and relevance of the present study, which seeks to address this shortfall by introducing advanced forecasting approaches designed explicitly for 3PL providers.

2.1. Key trends in supply chain forecasting

Supply chain forecasting has always represented a fundamental step in decision-making processes, significantly influencing the effectiveness of operations planning, resource allocation, and risk management across organizations (Xie et al., 2004). Accurate forecasting improved manufacturing capacity and logistics planning and reduced inventory distortion and avoided substantial revenue losses. Indeed, Chen et al., (2007) highlighted that demand forecasting outcomes formed the foundation of various planning activities within demand–supply networks, directly influencing operational effectiveness. Their research emphasized that uncertainty in demand signals propagated through supply chains, magnifying the risks of ineffective operational plans and making accurate forecasting essential. Similarly, Doganis et al., (2008) demonstrated the importance of forecasting integration within Model Predictive Control frameworks, showing that predictive insights significantly enhanced the efficiency of control actions and overall system performance. Additionally, Jauhar et al., (2024) further stressed this aspect, identifying inventory distortion due to inaccurate forecasting as a major cause of substantial revenue losses across large organizations.

Due to its central role in orchestrating complex networks, supply-chain forecasting has thus progressively leveraged richer, multivariate data and enhanced information-sharing mechanisms to improve accuracy and coordination. Early work demonstrated the fundamental benefits of sharing demand signals for suppliers and the wider network (Zhao & Xie, 2002) and highlighted trust as a prerequisite for honest data exchange (Firouzi et al., 2016). More recent studies have incorporated diverse inputs—ranging from real-time order weight and shipment counts (Chien et al., 2023; Shao et al., 2025) to news-based sentiment analysis in crisis settings (Nguyen et al., 2023)—while preserving privacy through federated learning (Zheng et al., 2023) and carefully selecting data sources based on their relative predictive value

(Abolghasemi et al., 2023; Dai et al., 2022; Yang et al., 2021).

Concurrently, the field has shifted from traditional statistical methods toward advanced machine-learning and deep-learning techniques (Ali et al., 2024; Punia et al., 2020). Applications of predictive ML for disruption forecasting (Brintrup et al., 2020) and reinforcement-learning for dynamic inventory control (Chien et al., 2020) have demonstrated substantial accuracy gains. Architectures such as LSTM have been shown to outperform classical approaches under variable demand (Kantasa-ard et al., 2021), while hybrid neural models have further reduced error rates (Hooshmand Pakdel et al., 2025; Shajalal et al., 2023). End-to-end, data-centric resilience frameworks (Li et al., 2025) and synthetic data generation strategies (Long et al., 2025; Vlachos & Reddy, 2025) continue to expand the practical adoption of ML and DL across complex supply-chain environments.

Finally, a significant evolution within forecasting practices was the transition from a "predict-then-optimize" paradigm (Mete Ayhan & Kır, 2024) to integrated, end-to-end forecasting solutions that simultaneously predicted demand and optimized operational costs. Sanders & Graman, (2009) emphasized the critical impact of forecast accuracy and bias on organizational costs, highlighting the necessity of integrated approaches that directly considered cost structures in forecasting models. Goltsos et al., (2022) reinforced this perspective by critically reviewing the fragmented literature and advocating for a unified approach that jointly addressed forecasting and inventory control decisions. They introduced a structured framework that explicitly addressed the interactions between forecasting and operational decision-making, emphasizing the practical benefits and complexities of this comprehensive approach. Babaveisi et al., (2023) further illustrated the value of integrated models, demonstrating how combining forecasting with inventory and resource planning significantly enhanced decision-making efficiency, optimized costs, and improved overall forecasting performance in resource-intensive environments—underscoring the growing importance and practical advantages of simultaneous prediction and optimization methodologies in modern supply chains.

In summary, the recent evolution of supply chain forecasting has been marked by three key trends: the increased emphasis on information-sharing and enriched input data, the growing adoption of advanced ML and DL techniques over traditional statistical models, and the shift toward integrated, end-to-end forecasting approaches that jointly optimize operational outcomes. While these advancements have been explored across various supply chain segments, their application within the 3PL sector remains notably limited as the following Section demonstrates.

2.2. Research gaps in supply chain forecasting for 3PL

Statistical forecasting models had historically represented the primary forecasting practices within the 3PL sector, providing foundational tools for predictions essential to operational management. Kmiecik & Wolny, (2022) explored the operational integration of an ARIMA-based forecasting model within a logistics provider serving a confectionery manufacturer. Their study underscored the tool's practical value in resource planning. Nevertheless, they identified gaps concerning the model's transparency and verifiability—factors that operational managers found crucial yet insufficiently addressed by the statistical approach. Further supporting this initial preference for statistical models, Kmiecik, (2025) reinforced that 3PL companies extensively employed statistical forecasting methodologies to enhance the accuracy of demand and supply forecasts across multiple logistics facilities. The study emphasized the direct correlation between data quality, forecasting accuracy, and the skill level of personnel managing these statistical tools. It also noted specific limitations, such as challenges in accurately forecasting smaller, frequent order volumes typical of e-commerce flows, suggesting the need for advancements beyond purely statistical models. Additionally, in analyzing forecast errors within a 3PL's multiple distribution channels, Wolny & Kmiecik, (2025) emphasized the importance of customized statistical forecasting models tailored to individual distribution channels. Although their approach reinforced the practical benefits of statistical methods, the lack of clarity regarding model inputs and forecasting mechanisms limited comprehensive error source identification, pointing toward the potential benefits of transitioning to more interpretable or advanced predictive methodologies.

A progressive shift toward ML and DL methods emerged in recent years, offering significant advantages over traditional statistical models. Zhou et al., (2006) first highlighted the feasibility and superior predictive capability of Neural Networks in demand forecasting for LTL carriers, demonstrating their effectiveness compared to conventional time-series approaches. Zhou, (2017) extended this perspective by advocating a combined forecasting method integrating exponential smoothing, ARIMA, and Neural Networks, which further enhanced prediction accuracy through optimized weighting. Lorenc et al., (2021) utilized Artificial Neural Networks combined with Big Data analytics and Internet of Things sensors, successfully forecasting temperature disruptions in cold chain logistics and significantly reducing waste and operational disruptions. Similarly, Kramarz & Kmiecik, (2022) developed a forecasting model tailored for distribution networks, linking forecasting capabilities directly to logistics coordination skills within 3PL providers and emphasizing the need for comprehensive, data-driven predictive approaches.Furthermore, Kmiecik, (2023) advanced this line of inquiry by analyzing 29 distribution networks, partially confirming that ML- and ANN-based forecasts outperform traditional methods in complex or food-related networks, while also demonstrating the benefits of modern technologies such as cloud-based WMS, EDI, and tracking standards in enhancing logistics coordination and supporting manufacturers' demand planning.

Following key trends in supply chain forecasting, ML and DL models also began to adopt multivariate data. Lindsey et al., (2014) successfully leveraged multivariate inputs to optimize real-time pricing and capacity sourcing decisions for 3PL providers. Budak et al., (2017) demonstrated the effectiveness of multivariate forecasting frameworks using artificial neural networks and quantile regression for predicting truckload spot market prices, highlighting their applicability and superior accuracy. Polim et al., (2017) introduced a multivariate, data-driven framework employing tree-based quantile regression forests, significantly improving delivery time predictions by integrating operational predictors. Additionally, Gürbüz et al., (2019) applied multivariate Artificial Neural Network and data mining techniques to predict and analyze operational error parameters within logistics services, demonstrating substantial potential for enhancing service quality and operational efficiency through comprehensive, multivariate data analysis.

Despite these advancements in forecasting methodologies, several gaps could still be observed in the current literature. Table 1 categorized the reviewed studies based on their predicted output, the input data provided to the models, and the loss function adopted for their training. According to Table 1, the adoption of loss functions tailored specifically to the operational needs of 3PL providers remained notably absent in the current literature. Existing forecasting frameworks predominantly utilized standard accuracy metrics and lacked metrics based on operational cost functions. Moreover, even though the prediction of truck space utilization represented a critical planning unit within the 3PL domain, it remained a significantly underexplored area. Indeed, only one study, conducted by Grzelak et al., (2019), attempted to forecast space requirements using a multiple regression method. However, this study did not incorporate data from information-sharing technologies or apply custom loss functions to optimize predictions. Based on this evidence, the proposed approach represented the first forecasting model for 3PL providers that leveraged ML techniques trained on custom loss functions and relied on multivariate input data from information-sharing technologies to predict truck space utilization.

Table 1
Literature review summary on 3PL supply chain forecasting models. M: Multivariate, U: Univariate, S: Statistical Model; ML: Machine Learning Model.

Study	Forecasted metric	Forecasting models	Forecasting input type	Custom loss function
Lindsey et al., (2014)	Shipment price	S	M	No
Budak et al., (2017)	Shipment price	ML	M	No
Lorenc et al., (2021)	Cold chain disruptions	ML	U	No
Kramarz & Kmiecik, (2022)	Products demand	ML	U	No
Polim et al., (2017)	Shipment time	ML	M	No
Kmiecik, (2022)	Products demand	ML	U	No
Gürbüz et al., (2019)	Damage parameters	ML	M	No
Kmiecik, (2025)	Products demand	S	U	No
Wolny & Kmiecik, (2025)	Products demand	S	U	No
Grzelak et al., (2019)	Pallet space demand	ML	U	No
Kmiecik & Wolny, (2022)	Products demand	S	U	No
Zhou et al., (2006) Zhou et al., (2006)Zhou et al., (2006) Zhou et al., (2006)	Products demand	ML	U	No
Zhou, (2017)Zhou, (2017)Zhou, (2017)Zhou, (2017)	Products demand	ML	U	No
This study	**Truck space demand**	**ML**	**M**	**Yes**

3. Materials and methods

This Section begins by outlining the problem addressed in this study (Section 3.1). It then introduces the proposed approach, detailing its main components and structure (Section 3.2). Finally, the experimental design used to evaluate the approach is presented, along with a description of the setup and conditions under which the experiments were conducted (Sections 3.3 and 3.4).

3.1. The 3PL truck space reservation forecasting problem

Every day, 3PLs receive multiple shipments from various producers, consolidate them at their warehouses, and dispatch them to the corresponding retailers. Due to the tight time constraints typical of sectors such as food distribution, the time available for unloading, sorting, and reloading products is extremely limited. In addition, 3PLs often receive incomplete or late information regarding upcoming shipments, which hinders their ability to plan truck capacity efficiently in real time. As a result, 3PLs must rely on demand forecasts to anticipate how much space will be needed on each truck for every producer–retailer pair on a given day. Fig. 1 presents an overview of the considered problem. Specifically, a supply chain consisting of a set of Producers P, a set of Retailers R, and a 3PL responsible for goods shipments from producers to retailers operating with LTL transports is considered.

To formalize this forecasting problem, Table 2 introduces the main variables and notation.

According to Fig. 1 and Table 2 the core challenge of forecasting in 3PL environments is the need to generate accurate estimates $D_{ijt}^{forecast}$ before the true space requirement D_{ijt}^{real} is known based on incomplete or inaccurate information exchanged with producers $I_{ijt}^{declared}$. In addition, when solving the forecasting problem it has to be considered that the cost implications of errors are asymmetric:

- If $D_{ijt}^{forecast} > D_{ijt}^{real}$, the truck will be underutilized, incurring cost $C^{under-saturated}$

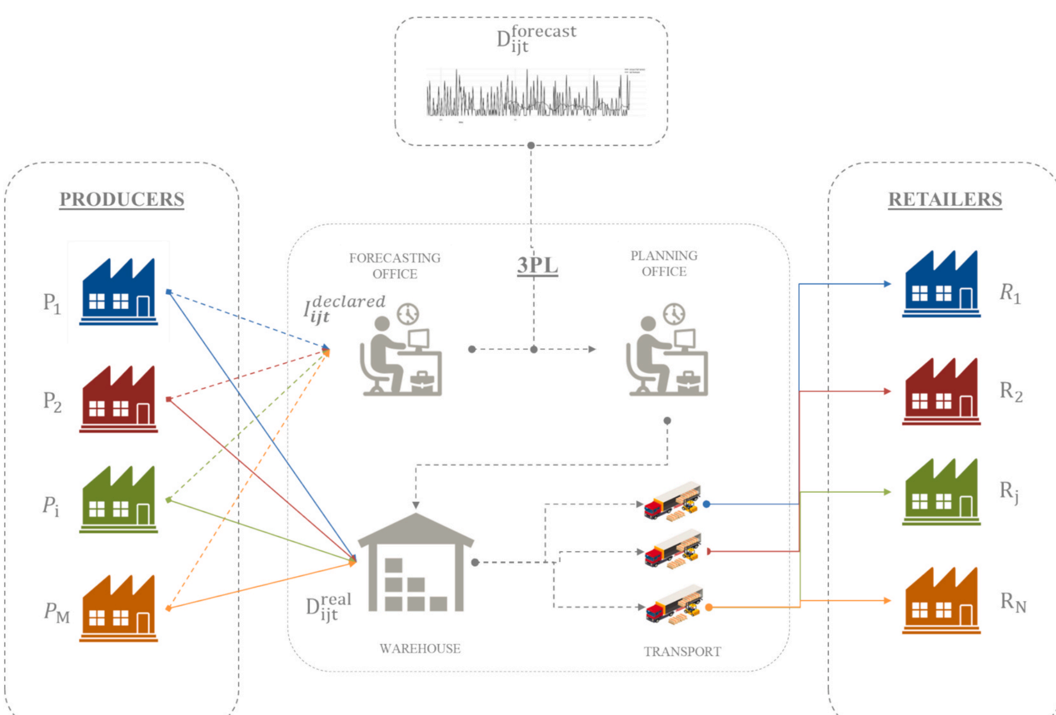

Fig. 1. Problem Overview.

Table 2
Problem nomenclature.

ACRONIM	NOMENCLATURE
$\mathbf{P} = \{1,..\mathbf{i},\cdots\mathbf{M}\}$	Set of producers
$\mathbf{R} = \{1,..\mathbf{j},\cdots\mathbf{N}\}$	Set of retailers
$\mathbf{T} = \{1,..\mathbf{t},\cdots\mathbf{T}\}$	Set of days
$\mathrm{I}_{\mathrm{ijt}}^{\mathrm{declared}}$	Information specified from a producer i related to the request of shipment to retail j in day t
$\mathrm{D}_{\mathrm{ijt}}^{\mathrm{forecast}}$	Forecasted space to reserve on a truck for products of producer i to deliver to retail j in day t
$\mathrm{D}_{\mathrm{ijt}}^{\mathrm{real}}$	Real space to reserve on a truck for products of producer i to deliver to retail j in day t
$\mathrm{C}^{\mathrm{over-saturated}}$	Cost of overestimating the real space to reserve on the truck
$\mathrm{C}^{\mathrm{under-saturated}}$	Cost of underestimating real space to reserve on the truck

- If $\mathrm{D}_{\mathrm{ijt}}^{\mathrm{forecast}} < \mathrm{D}_{\mathrm{ijt}}^{\mathrm{real}}$, the 3PL may lack sufficient truck space, causing missed deliveries or expedited shipments, and thus incurring a higher cost $\mathrm{C}^{\mathrm{over-saturated}}$

Therefore, the forecasting objective is twofold:

1. To estimate the future truck space requirement $\mathrm{D}_{\mathrm{ijt}}^{\mathrm{forecast}}$ for each producer–retailer pair on each day t;
2. To minimize the overall cost associated with forecasting errors, accounting for their asymmetric impacts.

3.2. The new proposed forecasting framework

Fig. 2 illustrates the proposed framework to solve the 3PL truck space reservation forecasting problem detailed in the previous Section 3.1. Specifically, the proposed framework details the forecasting models' design and provides a holistic view of how 3PLs should redesign their operations from a data-driven perspective. In particular, the framework encompasses the data collection, the design of the forecasting model and the data flow related to the new data-driven-based operations.

Starting from the data collection, which represents the core element of each data-driven solution, the proposed framework identifies valuable data to collect and the related technological solutions for data acquisition. Specifically, two distinct moments for data gathering are identified based on their temporal position with respect to the planning phase. The ex-ante data collection refers to the data collected before the planning phase of the truck. Conversely, the ex-post data refers to those data that can be acquired after the planning stage (and thus after the forecast generation) once products are received in the 3PL warehouse.

According to the Fig. 2, the ex-ante data collection should focus on gathering data provided by producers through information-sharing mechanisms. Valuable information to collect in this phase is represented by the total weight of products in the order of producer i destinated to retailer j in period t, the total number of shipment units in the

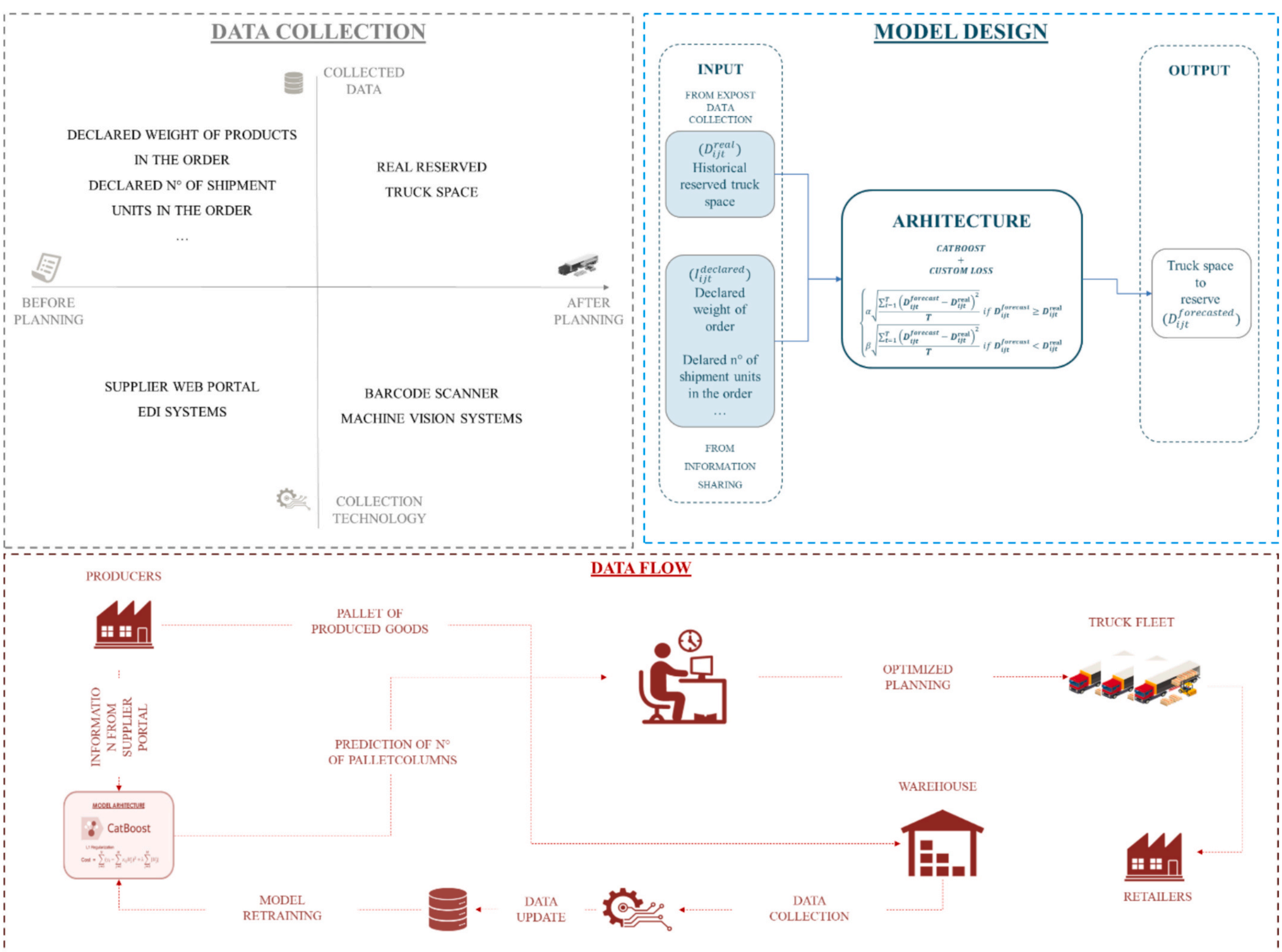

Fig. 2. Proposed framework for 3PL truck space reservation prediction problem.

order of producer i destinated to retailer j in period t and other information such as order frequency or product types. Indeed, such indirect information can help to generate future estimates of the truck space to reserve for products related to the producer i directed to retail j in day t ($D_{ijt}^{forecast}$). Moreover, these data represent information often present in producers' databases due to the need to manage their internal logistics. Potential technological solutions for collecting data at this stage may include supplier portals or EDI protocols. Indeed, 3PL integration has been widely recognized as a fundamental element for value creation in several studies (Lai et al., 2008; Premkumar et al., 2021) and supplier web portals and EDI protocols represent widely recognized technological solutions to address this necessity.

In contrast, ex-post data collection aims to capture the real values of the truck space to reserve for products of producer i to retailer j in day t (D_{ijt}^{real}). Indeed, collecting ex-post data is crucial for training predictive models that use a supervised learning approach that requires historical data reflecting actual values for generating reliable forecasts. A fundamental aspect to capture at this step is the reduction of space in the value of D_{ijt}^{real} due to the possible overlapping of pallets. The moment suggested to capture this effect, for the ex-post data collection, is thus exactly before the filling process of the truck when "columns of pallet" are formed on the ground and prepared to be positioned inside the truck. Indeed, according to Fig. 3, as pallets present standard dimensions on the plant, capturing the height of these "columns" allows to uniquely identify the space occupied within the truck by each product related to specific deliveries from producer i to retailer j in day t.

The technological solutions suggested for this phase may include various levels of automation, starting from adopting barcode scanners, which can help track the number of generated columns of pallets. Alternatively, machine vision algorithms, placed within fixed gates or over forklifts, can be deployed to automatically estimate the necessary values without relying on the human workforce to generate the necessary annotations.

Once both ex-ante and ex-post data are gathered, according to the proposed approach, the use of supervised ML algorithms is adopted to uncover patterns between ex-ante data and the ex-post data and thus to generate the forecasts $D_{ijt}^{forecast}$. Specifically, among the available supervised ML algorithms, in this study, the adoption of CatBoost regressor algorithms is suggested (Dorogush et al., 2018). Gradient-boosting decision trees have demonstrated high predictive performance in a wide set of forecasting competitions (Bentéjac et al., 2021; Gupta et al., 2016; Islam & Amin, 2020). Moreover, among gradient-boosting algorithms, CatBoost has reported the best performance (Dorogush et al., 2018). Although CatBoost, XGBoost, Random Forests, and interpretable models each have advantages for structured data, CatBoost offers specific benefits: it handles categorical features natively, automatically transforming identifiers such as producer and retailer IDs into statistically meaningful representations. This is achieved through target-based encoding, where categorical values are replaced by aggregated statistics (e.g., averages) of the target variable computed over the training data in a way that preserves generalization. As a result, IDs and other categorical inputs can be directly leveraged by the model without extensive manual preprocessing, improving both efficiency and predictive accuracy. Morevore, it reduces overfitting via ordered boosting, and often achieves superior accuracy without extensive tuning. Compared to XGBoost and LightGBM, CatBoost's symmetric tree-building and efficient handling of default parameters simplify model training and can yield faster convergence on heterogeneous datasets. While Random Forests provide robustness and interpretability, they may underperform in predictive accuracy for complex feature interactions relative to gradient-boosting algorithms. Furthermore, although linear or rule-based interpretable models facilitate transparency, they often lack the nonlinear capacity required for capturing intricate patterns in 3PL operations. Therefore, CatBoost represented a balanced choice, combining high predictive power, built-in categorical support, and manageable tuning complexity. Furthermore, to consider the nature of the different effects that forecasting error generates in the problem, implementing a custom loss function within the CatBoost algorithm is suggested. Indeed, the custom loss function determines the objective the algorithm seeks to minimize to reduce prediction error. However, while for regression tasks, the Root Mean Squared Error (RMSE) is a widely adopted loss function, its definition, according to Equation (1), does not allow weight forecasting error since they are over-estimating or under-estimating the predicted metric:

$$RMSE_{ij} = \sqrt{\frac{\sum_{t=1}^{T}\left(D_{ijt}^{forecast} - D_{ijt}^{real}\right)^2}{T}} \tag{1}$$

Therefore, when supporting 3PLs predictions, the custom loss introduced in Equation (2) can be adapted to better capture the imbalance nature of forecasting error in the problem:

$$CUSTOMLOSS = \begin{cases} \alpha\sqrt{\dfrac{\sum_{t=1}^{T}\left(D_{ijt}^{forecast} - D_{ijt}^{real}\right)^2}{T}} if D_{ijt}^{forecast} \geq D_{ijt}^{real} \\ \beta\sqrt{\dfrac{\sum_{t=1}^{T}\left(D_{ijt}^{forecast} - D_{ijt}^{real}\right)^2}{T}} if D_{ijt}^{forecast} < D_{ijt}^{real} \end{cases} \tag{2}$$

In designing the proposed custom loss function, the square root term from the standard RMSE has been retained to preserve the interpretability of errors in the same units as the forecasted metric, ensuring that stakeholders can directly gauge the magnitude of deviations. Compared to a plain MSE, the square root dampens the effect of extreme errors so that large deviations do not dominate the objective unduly. Meanwhile, unlike MAE (Mean Absolute Error), which treats all errors linearly, the square root form still more heavily penalizes larger squared deviations than MAE, providing a balance between sensitivity to outliers and interpretability. The piecewise structure further distinguishes the proposed approach by enabling different scaling factors α and β for underestimation and overestimation. In many 3PL operations, underestimating required truck space (leading to over-saturation) can be costlier—causing missed shipments or expedited deliveries—while overestimation (resulting in under-utilized capacity) leads primarily to idle resources. Traditional symmetric losses (e.g., RMSE or MSE) cannot

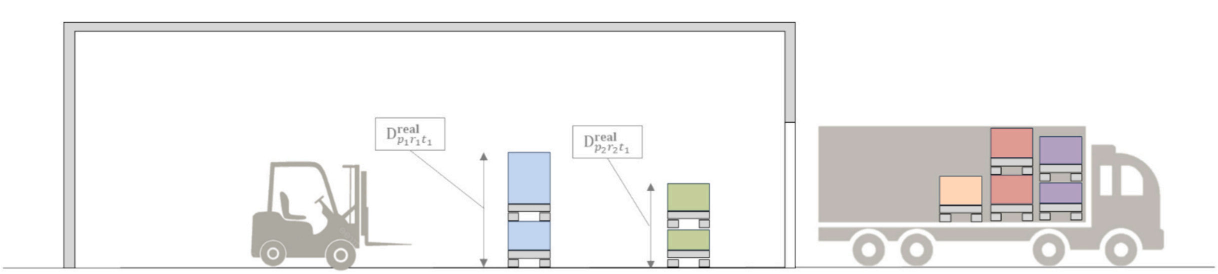

Fig. 3. Suggested moment to perform the ex-post data collection.

capture this asymmetry; even quantile loss functions require specifying a quantile threshold, which does not explicitly relate to cost ratios. By contrast, the proposed piecewise loss allows direct calibration of α and β to real-world cost metrics, providing a more intuitive and operationally aligned objective. As a result, the proposed loss function balances interpretability, robustness to outliers, and cost-aware asymmetry, outperforming standard regression losses in settings where over- and underestimation have very different implications. Indeed, the proposed loss function includes different penalties for overestimating and underestimating predictions, reflecting the real-world operational costs more accurately. Specifically, α and β can be adopted to represent the respective costs of underestimating $C^{under-saturated}$ and overestimating $C^{over-saturated}$ the true values.

In conclusion, once the model has been accurately designed and trained using data collected over an adequate period (for instance, one year), the workflow should start with the daily reception of information from suppliers ($I_{ijt}^{declared}$). This data is then combined with historical records of the actual space to reserve on trucks observed for each specific producer-retailer pairs (D_{ijt}^{real}) and fed into the forecasting model. The resulting predictions $D_{ijt}^{forecast}$ can be promptly transmitted to the planning department to optimize load planning or routing. Moreover, a continuous data collection process must be implemented daily, capturing the new D_{ijt}^{real} generated values each time products are received in the 3PL warehouse. This information is stored in a centralized database, ensuring the predictive models can be periodically retrained and tuned to maintain accuracy.

3.3. Experimental design

Data collected from a real case study of a 3PL in the food sector has been adopted to validate the effectiveness of the proposed framework. The initial dataset consisted initially of 1′308′761 rows and 114 columns. At the finest level, each row represents a single pallet on a particular trip—i.e., one pallet traveling from a specific producer to a specific retailer on a given day. For each pallet, operators recorded its height, while the ERP and supplier portal supplied the declared weight and the number of shipment units together with other additional information. Any row lacking both a valid pallet weight and a declared number of colli was dropped in its entirety (removing the corresponding order from that producer to that retailer) to prevent partial, misleading information when later aggregated. This cleanup affected only 0.7 % of rows, reducing the dataset to 1′299′599 records. Table 3 provides a few key data describing the complexity of the examined case study.

Starting from those 1′299′599 pallet-level records, rows have been aggregated by producer–retailer pair and date. For each pair (i = producer, j = retailer) on day t. After aggregation, the dataset contained 441′801 unique (i, j, t) observations. We also trimmed the number of columns from 114 to the 7 essential fields. The retained field represented the identifier related to producer, retailer and date, the value of the reserved truck space which is the object of prediction and the additional information shared with producers to support forecast.

Specifically, according to the framework described in Fig. 2, the set of information exchanged with producers in the company ($I_{ijt}^{declared}$) included:

Table 3
Case study data overview.

DATA	VALUES
Average number daily of producers	130
Average number of daily retailers	700
Average number of daily trucks	250
Average number of daily orders	1300
Data collection period	2021—2024

- The total weight of products in the order ($W_{ijt}^{declared}$);
- The total number of shipment units in the order ($SU_{ijt}^{declared}$).

Fig. 4 highlights the correlation between these values and the real historically observed space reserved for products on trucks for each pair of producer i and retailer j (D_{ijt}^{real}).

According to the Fig. 4, a high correlation between the target variable (D_{ijt}^{real}) and the variables suggested to collect in the ex-ante phase ($I_{ijt}^{declared}$) can be noticed. This evidence thus provides empirical support for the theoretical statements reported in Section 3.2.

Starting from the data collected in the examined case study, several benchmarks have been compared to the proposed framework to validate its effectiveness. In particular, a benchmark data collection that does not consider the possibility of leveraging information-sharing mechanisms has been adopted to validate the effectiveness of the data collection reported in the framework. In this first benchmark, predictive models are thus trained in a univariate manner, relying only on the historical evolution of the truck space reserved for a specific pair of producer i and retailer j in the past days.

Conversely, benchmark approaches based on different forecasting models have been considered to investigate the effectiveness of the model design proposed in the framework. Specifically, traditional and statistical models such as the Naïve Moving Average (NMA) model and the Autoregressive Integrated Moving Average Model (ARIMA) have been considered here. Moreover, to confirm the benefit of relying on CatBoost within the family of gradient boosting models, the LightGBM model, which represents another model based on gradient boosting logic, has been considered as an additional benchmark.

Lastly, a version of the CatBoost model trained with a classical RMSE loss function has also been considered to evaluate the benefit of relying on the proposed custom loss function.

A summary of the characteristics of the proposed approach and the benchmarks generated from the combination of the different data collection and the model architectures is reported in Table 4.

To ensure a fair comparison, all models were trained on the same dataset and evaluated on the same output (D_{ijt}^{real}), although they had access to different input features based on their design. Univariate benchmarks (NMA, ARIMA, LightGBM, CatBoost) only used historical D_{ijt}^{real}, whereas covariate-based models (ARIMA_COV, LightGBM_COV, CatBoost_COV, and CatBoost_COV_CL) received $W_{ijt}^{declared}$, $SU_{ijt}^{declared}$, D_{ijt}^{real}. Only CatBoost_COV_CL applied the custom loss because it uniquely supports user-defined objectives; other models employed RMSE to provide a consistent baseline. By varying model architecture (e.g., CatBoost vs. LightGBM) or input features (univariate vs. multivariate) while evaluating all approaches on the same data, an ablation analysis has been conducted to isolates the benefits of the proposed approach: the multivariate framework and cost-aware custom loss.

Lastly, widely recognized accuracy metrics have been considered during the experiments to compare the proposed approach's effectiveness against the examined benchmarks. These metrics included the Mean Error (ME), the Root Mean Squared Error (RMSE), the Mean Absolute Percentage Error (MAPE) and the R2 coefficient. The metrics have been computed for each pair (producer i, retailer j) as follows:

$$ME_{ij} = \frac{1}{T}\sum\nolimits_{t=1}^{T} D_{ijt}^{real} - D_{ijt}^{forecasted} \tag{3}$$

$$MAE_{ij} = \frac{1}{T}\sum\nolimits_{t=1}^{T} \left| D_{ijt}^{real} - D_{ijt}^{forecasted} \right| \tag{4}$$

$$MAPE_{ij} = \frac{1}{T}\sum\nolimits_{t=1}^{T} \frac{\left| D_{ijt}^{real} - D_{ijt}^{forecasted} \right|}{D_{ijt}^{real}} \tag{5}$$

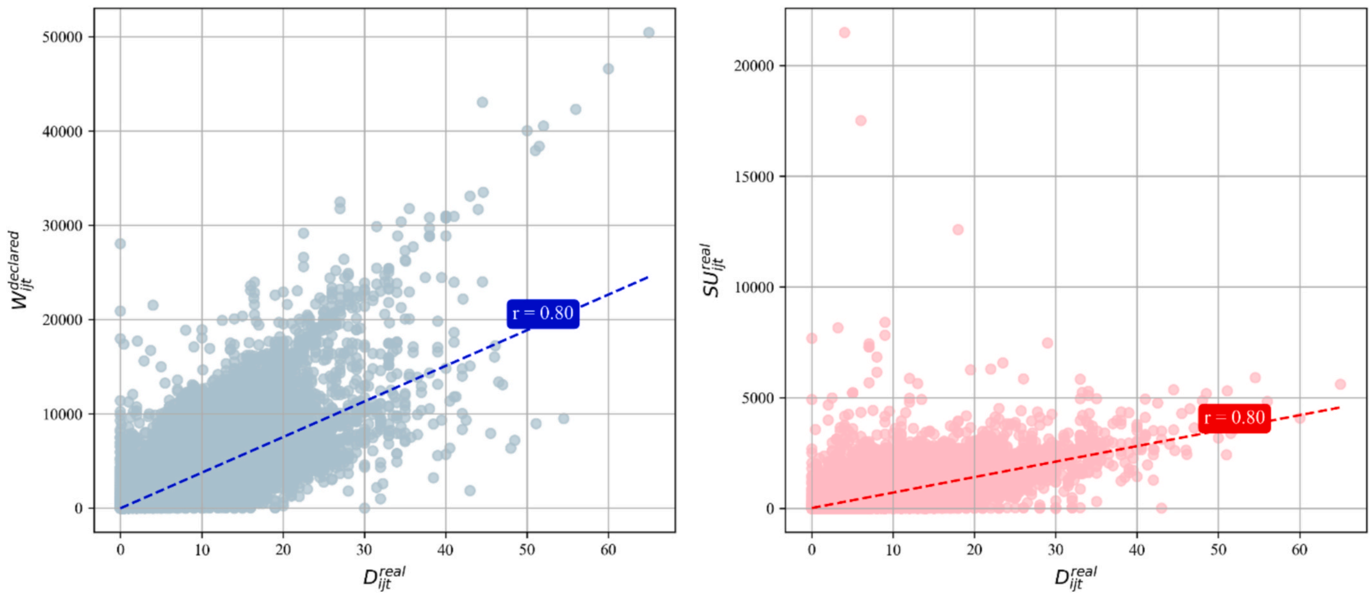

Fig. 4. Pearson correlation observed in the examined case study between data obtained from producers through information-sharing mechanisms and the real space reserved on trucks for its products.

Table 4
Nomenclature of the examined forecasting approaches.

Approach group	Approach nomenclature	Forecasting algorithm	Input Data	Loss function
Proposed	CatBoost_COV_CL	CatBoost	$W_{ijt}^{declared} SU_{ijt}^{declared} D_{ijt}^{real}$	Custom Loss
Benchmark	NMA	Naïve Moving Average	D_{ijt}^{real}	RMSE
Benchmark	ARIMA	Autoregressive Integrated Moving Average	D_{ijt}^{real}	RMSE
Benchmark	ARIMA_COV	Autoregressive Integrated Moving Average	$W_{ijt}^{declared} SU_{ijt}^{declared} D_{ijt}^{real}$	RMSE
Benchmark	LightGBM	LightGBM	D_{ijt}^{real}	RMSE
Benchmark	LightGBM_COV	LightGBM	$W_{ijt}^{declared} SU_{ijt}^{declared} D_{ijt}^{real}$	RMSE
Benchmark	CatBoost	CatBoost	D_{ijt}^{real}	RMSE
Proposed	**CatBoost_CL**	**CatBoost**	$\boldsymbol{D_{ijt}^{real}}$	**Custom Loss**
Benchmark	CatBoost_COV	CatBoost	$W_{ijt}^{declared} SU_{ijt}^{declared} D_{ijt}^{real}$	RMSE

$$RMSE_{ij} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(D_{ijt}^{real} - D_{ijt}^{forecasted}\right)^2} \tag{6}$$

$$\boldsymbol{R_{ij}}^2 = 1 - \frac{\sum_{t=1}^{N}\left(\boldsymbol{D_{ijt}^{real}} - \boldsymbol{D_{ijt}^{forecasted}}\right)^2}{\sum_{t=1}^{N}\left(\boldsymbol{D_{ijt}^{real}} - \overline{\boldsymbol{D_{ijt}^{real}}}\right)^2} \tag{7}$$

Where T is the total number of days considered in the the test set.

Additionally, a Transport Cost Metric (TCM) considering the cost of the unsaturated truck and undelivered goods has been introduced to complement the analysis:

$$TCM_{ij} = \begin{cases} C^{over-saturated}\sqrt{\dfrac{\sum_{t=1}^{T}\left(D_{ijt}^{real} - D_{ijt}^{forecasted}\right)^2}{T}} if D_{ijt}^{real} \geq D_{ijt}^{forecasted} \\ C^{under-saturated}\sqrt{\dfrac{\sum_{t=1}^{T}\left(D_{ijt}^{real} - D_{ijt}^{forecasted}\right)^2}{T}} if D_{ijt}^{real} < D_{ijt}^{forecasted} \end{cases} \tag{8}$$

Where $C^{over-saturated}$ and $C^{under-saturated}$ represent the cost generated when predictions underestimate and overestimate the real space to reserve on trucks.

The cost parameters ($C^{over-saturated}$) and ($C^{under-saturated}$) have been directly extracted from the case study's operational cost records. Table 5 reports the values used in this study, reflecting costs per cubic meter of forecasting error based on historical financial data.

Table 5
Case study cost parameter.

Cost parameter	Value [euro/m^3]
$C^{over-saturated}$	9
$C^{under-saturated}$	15

Lastly, an explainability study based on SHAP values was conducted on both the proposed CatBoost model with covariates and custom loss (CatBoost_COV_CL) and the baseline CatBoost model with standard RMSE loss (CatBoost_COV). The trained models were applied to a held-out test set, and SHAP values were computed for each feature across all observations. SHAP summary plots were generated to assess the overall contribution of each covariate to model outputs. By comparing the distribution and magnitude of SHAP values between the two models, it was possible to determine whether the cost-aware loss function altered the relative importance or interaction patterns of declared weight and shipment units without compromising the stability of feature attributions. This approach preserves transparent, consistent explanations of model behavior while enhancing cost sensitivity. Moreover the time required to train each predictive model is recorded in seconds.

3.4. Experimental setup

Experiments have been developed by simulating the performances of the proposed and benchmark approaches over the historical data collected in the examined case study. Specifically, the overall available dataset has been initially split into three different parts, maintaining the temporal integrity of the entire dataset: the training set (from 2021 to 2022), the validation set (from 2022 to 2023) and the test set (from 2023 to 2024).

Initially, the training and validation sets were adopted to tune the hyperparameter of each forecasting model by selecting the hyperparameter that led to the lowest average RMSE over the validation set. Once the model hyperparameters have been identified, both the training and the validation set have been adopted as a new training set, and the metrics detailed in Section 3.3 have been computed over the test set. Details of the hyperparameter space and the selected hyperparameter for each model have been reported in Table 5. A Bayesian optimization approach with a time limit of 8 h has been adopted to search over the defined hyperparameter space. The definition of this search space was based on prior empirical knowledge, as no universally valid tuning strategy exists and the optimal ranges are typically data-dependent.

Table 6
Model research space and selected hyperparameter.

Model	Hyperparameter	Research space	Selected value
NMA	Window size	[3, 5, 7, 10, 14]	7
ARIMA	(p,d,q) order	p:[0–5], d:[0–2], q:[0–5]	(2,1,2)
LightGBM	Num. of estimators	[50, 100, 200, 500, 1000]	200
	Learning rate	[0.01, 0.05, 0.1, 0.2]	0.05
CatBoost	Num. of estimators	[100, 300, 500, 1000]	500
	Depth	[4, 6, 8, 10]	6
	Learning rate	[0.01, 0.05, 0.1]	0.1

While this represents a limitation, it does not compromise the reproducibility of the study, since both the search space and the optimization procedure are fully documented (See Table 6).

During both the tuning phase and forecast generation over the test set, a rolling forecasting horizon schema was adopted to simulate the day-by-day workflow of a typical 3PL provider. Specifically, the training dataset progressively included the data in the rolling period as time passed, and predictions were generated for a specific forecasting window every day. Specifically, because truck planning is a recurring process for the next day, the rolling period has been considered to be 1 day and the forecasting window has been considered 1 day.

All the experiments have been coded in Python, relying on the Darts library for the forecasting models, while the Optuna packages have been adopted for the Bayesian tuning strategy. All the computational experiments were conducted on a workstation with the following specifications: an Intel® Core™ i9-10900 CPU operating at 2.80 GHz, 64 GB of RAM. The workstation operated on Windows 10 Pro (version 22H2). The operating system was based on a 64-bit architecture.

4. Results

This Section reports the results of the experiment described in the previous paragraph. Specifically, Section 4.1 compares the forecasting accuracy between the proposed and the benchmark models. Section 4.2 illustrates a parametric comparison analysis of the cost produced by each forecasting model when varying the ratio between the $C^{over-saturated}$ and $C^{under-saturated}$. Section 4.3 details the computational efficiency of each model. Finally, Section 4.4 presents the SHAP-based explainability analysis, contrasting feature-attribution patterns under the standard RMSE loss and our cost-aware custom loss.

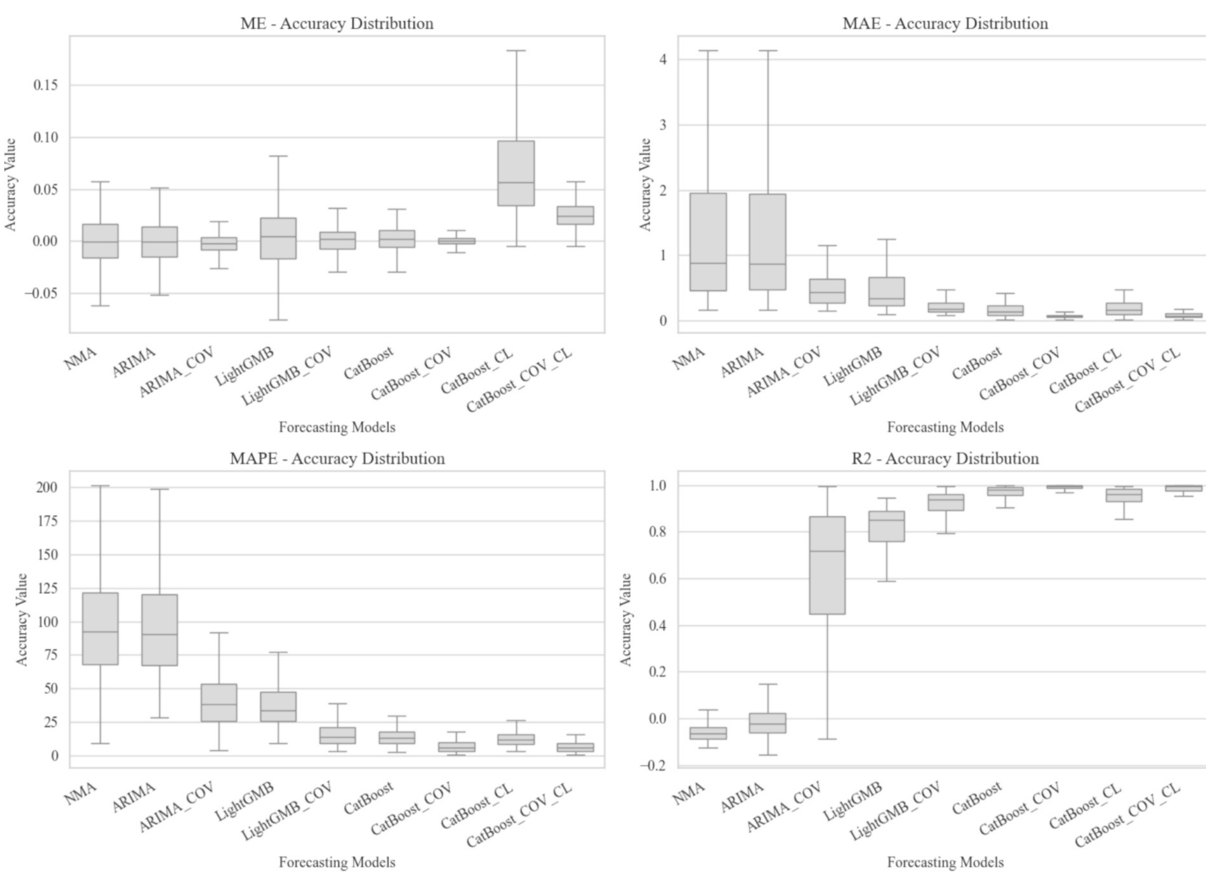

Fig. 5. Accuracy performance comparison in pallet predictions.

4.1. Accuracy performance results

Fig. 5 compares the distributions of the accuracy metrics obtained when leveraging the proposed approach and the benchmarks to predict the future value of the truck space to reserve for each producer i and retailer j in the investigated case study.

According to the Fig. 5, several considerations can be made. First, a progressive improvement in forecasting accuracy is observed as the models advance from traditional statistical methods to ML-based approaches. Traditional models, such as NMA, exhibit a median MAE of approximately 0.85, a MAPE of around 93 %, and an R^2 close to –0.05. Conversely, advanced ML models like CatBoost show a median MAE of approximately 0.01, a MAPE of around 12 %, and an R^2 close to 1 for these metrics. In addition, they achieve a considerably narrower dispersion of error. Indeed, whiskers in ARIMA extend from 26 % to 199 % in terms of MAPE, while the metric's value remains within the range of 3–28 % for the CatBoost model.

In addition to the superiority of advanced models, a significant performance improvement can also be noticed when moving from the benchmarks to the proposed multivariate data collection relying on data acquired through information-sharing mechanisms. Indeed, regardless of the specific forecasting method, shared information allows for consistently generating better predictions independently of the forecasting model considered. Median MAPE decreased from 93 % to 38 % for the ARIMA models, 32 % to 27 % for the LightGBM model, and 15 % to 9 % for the CatBoost model.

Third, within the ML models evaluated, CatBoost demonstrates a slight advantage over LightGBM when relying on both the benchmark and the proposed data collection. Indeed, when considering the univariate data collection, the error decreased from a median MAE of approximately 0.5–0.1 and from a MAPE of around 22 % to a one of 16 %. Similarly, improvements can be noticed when considering the multivariate data collection, such as an increase in R2 from 0.93 to a value of 0.96.

Lastly, it can be noticed that the CatBoost models trained without a custom loss function tend to provide slightly lower accuracy (from a MAE of 0.03 to one of 0.04 when considering the benchmark data collection and from a MAE of 0.01 to one of 0.02 when considering the proposed data collection). However, models trained with a custom loss function show a positive forecast bias (0.06 and 0.03 in terms of ME, depending on the considered data collection), leading to conservative estimates.

The results reported in Fig. 5 are also confirmed by the analysis in Fig. 6. Methods that never outperform any other—i.e., record 0 % "wins" versus all alternatives in Table 4 like ARIMA, ARIMA_COV and LightGBM —are omitted from the chart. Fig. 6 highlights the frequency with which specific combinations of forecasting algorithms and data configurations outperform others across MAE metrics.

The charts illustrate that the CatBoost model trained with the proposed data collection but without custom loss consistently yields the best results across 89 % of the evaluated time series. Instead, when considering the CatBoost with custom loss functions, the models outperform other algorithms only for 2.1 % of the examined time series.

In conclusion, Fig. 7 reports the MAE difference distribution for each considered time series between the proposed approach and the benchmarks.

According to the chart, CatBoost relies on the proposed data collection but without the custom loss and it is the only model reporting a

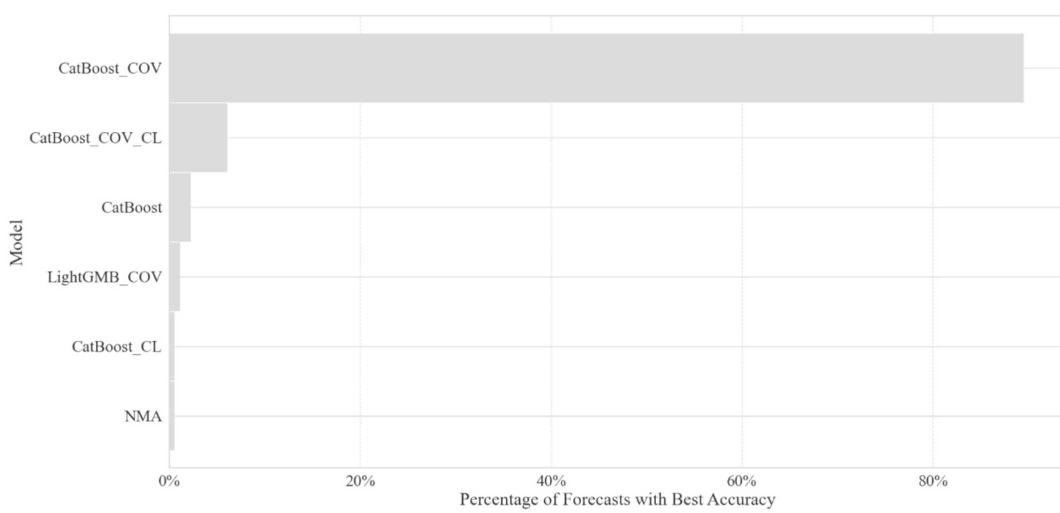

Fig. 6. Percentage of time series for which an examined approach reports the lower MAE error.

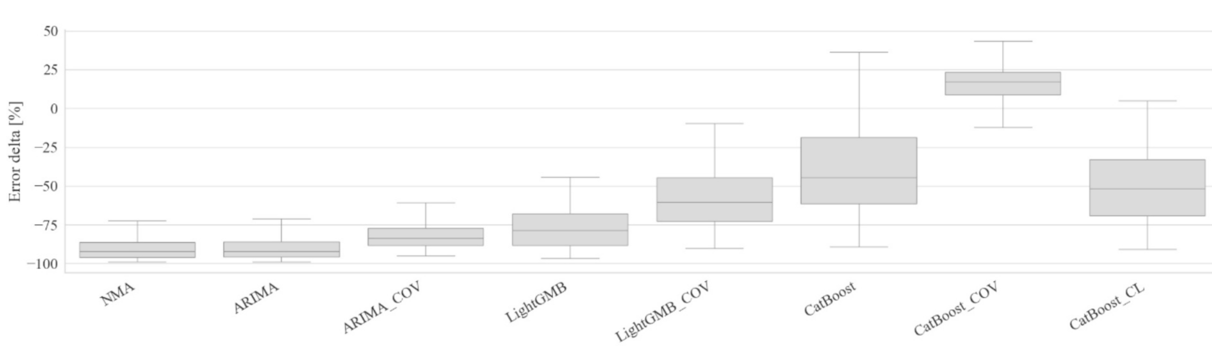

Fig. 7. Error delta between the proposed one and benchmarks.

median accuracy advantage of about 18 % compared to the proposed model. Indeed, compared to the other approach, the proposed one reported minor errors ranging from 92 % with NMA to 42 % with CatBoost when not relying on the proposed data collection.

4.2. Cost performance results

Fig. 8 presents the cost difference between the proposed approach and each benchmark; since there are eight benchmarks in Table 4, eight curves are shown, each representing the cost gap between the proposed model and one benchmarkunder varying ratios between the $C^{over-saturated}$ and the $C^{under-saturated}$.

Fig. 8 clearly shows that all the benchmarks examined, regardless of the ratios between $C^{over-saturated}$ and $C^{over-saturated}$, report an average negative cost difference when compared to the proposed model. The results thus indicate that the proposed approach consistently results in lower costs, with reductions ranging from −20 % to −87 %, depending on the model and ratios considered. Notably, the cost difference increases as we move from ML to baseline models and from the proposed data collection method to a univariate approach. The only exception to this trend is observed when comparing the proposed model to the CatBoost approach, which does not utilize the custom loss function. In this case, for ratios close to 1:1 between $C^{over-saturated}$ and $C^{over-saturated}$, the proposed model can result in costs that are up to 18 % higher.

Support for the results illustrated in Fig. 8 can also be noticed in Fig. 9, representing the number of time series for which a specific approach reports the lower cost depending on the considered cost ratios. Methods that never achieve a lower cost than any other—i.e., record 0 % wins like NMA, ARIMA, ARIMA_COV and LightGBM—are not shown in the plot.

Consistent with Fig. 8, the CatBoost model—without a tailored custom loss function—outperformed the proposed approach in 88 % of the analyzed time series when the cost ratio was close to 1:1. However, as the cost ratio increased, the proposed approach demonstrated superior performance, progressively improving its share of time series with the best cost outcome from 81 % to 97 %.

4.3. Training time results

Fig. 10 illustrates the distribution of training times required for various combinations of forecasting algorithms and data configurations in the examined case study.

According to the plots, the results reveal a clear trend of progressively increasing training time as models shift from traditional statistical approaches (with a median training time of 0.01 s for NMA and 0.9 s for

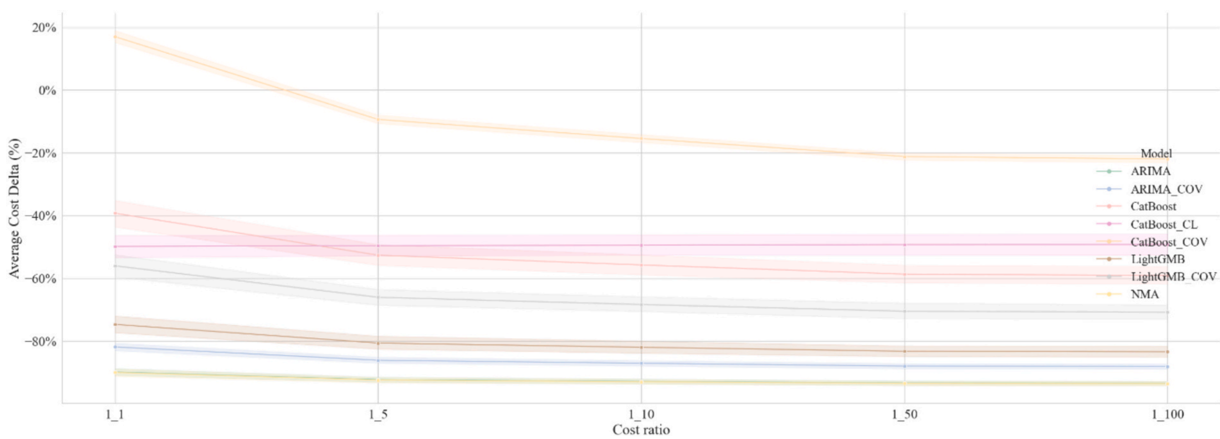

Fig. 8. Cost difference performance comparison.

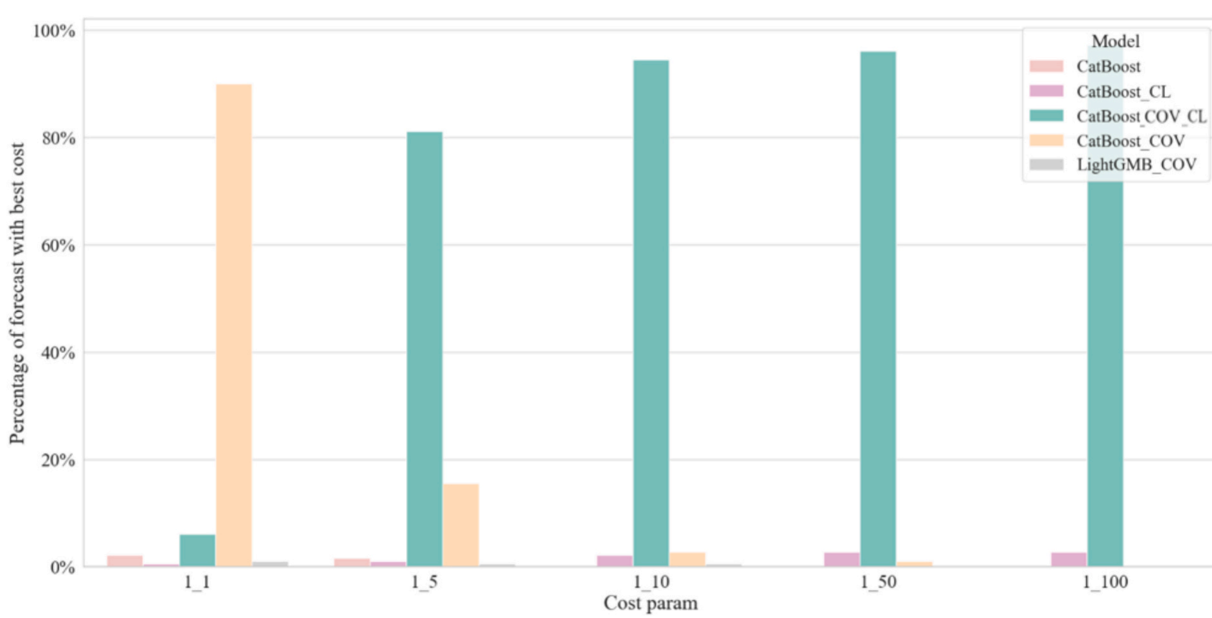

Fig. 9. Percentage of time series for which an examined approach reports the lower cost depending on the considered cost ratio.

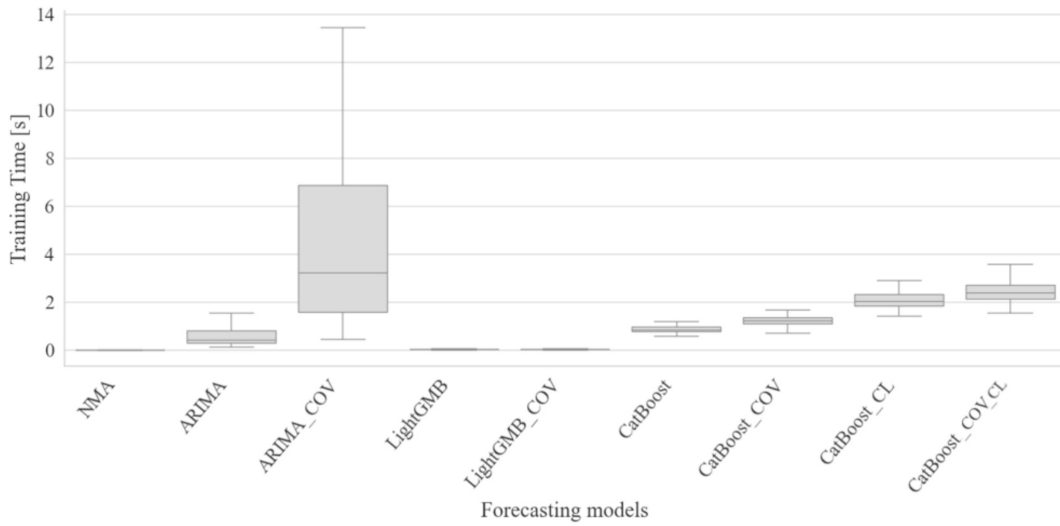

Fig. 10. Training time comparison.

ARIMA) to more advanced ML techniques (with a median training time of 1.1 s for CatBoost). An exception is, however, represented by the ARIMA model trained based on the proposed data collections, which exhibits the highest median training time (3.8 s) among all methods, outperforming even complex ML models. Additionally, the inclusion of covariates consistently leads to an increase in training time across all forecasting models (from 0.4 s to 3.2 s for the ARIMA model, from 0.034 s to 0.037 s for the LightGBM model and from 0.96 s to 1.34 s for the CatBoost model). Moreover, among the ML models, CatBoost demonstrates longer training times than LightGBM (from 0.01 s to 1.4 s when considering the proposed data collection), reflecting the higher computational demands of CatBoost's model structure and optimization process. Lastly, CatBoost models utilizing a custom loss function were shown to require more training time than those using the standard loss function (from 1.1 s to 1.9 s when not considering the proposed data collection and from 1.3 s to 2.2 s when considering the proposed data collection).

4.4. Model explainability analysis

Fig. 11 reports SHAP summary plots for the two CatBoost variants evaluated on the held-out test set. Each violin-style plot visualizes the distribution of SHAP values (impact on model output) for the top 16 features, ordered by mean absolute impact. The left panel corresponds to the covariate-based CatBoost with standard RMSE loss (CatBoost_COV), while the right panel shows the same model trained with our cost-aware custom loss (CatBoost_COV_CL). Feature values are colored from blue (low) to red (high).

In both models, the declared weight ($W_{ij}^{tdeclared}$_futcov_lag0) and the declared shipment units ($SU_{ij}^{tdeclared}$_futcov_lag0) dominate the feature-importance ranking, confirming that real-time covariates are stronger predictors of required truck space than any single historical load. The subsequent positions are occupied by various lags of the true reserved space (D_{real}_target_lag-n), reflecting the residual contribution of past volume trends.

Comparing panels, it can be seen an almost complete overlap in the bulk of each distribution. Only in the extreme upper tails do the covariate SHAP values under the custom loss (right panel) extend marginally farther than under RMSE (left panel), suggesting a very slight increase in sensitivity to high declared weights or unit counts. Crucially, however, these differences are negligible for the vast majority of observations. This close alignment demonstrates that introducing our cost-aware objective leaves the core feature-attribution structure intact—-preserving transparency and stability—while imparting only a subtle

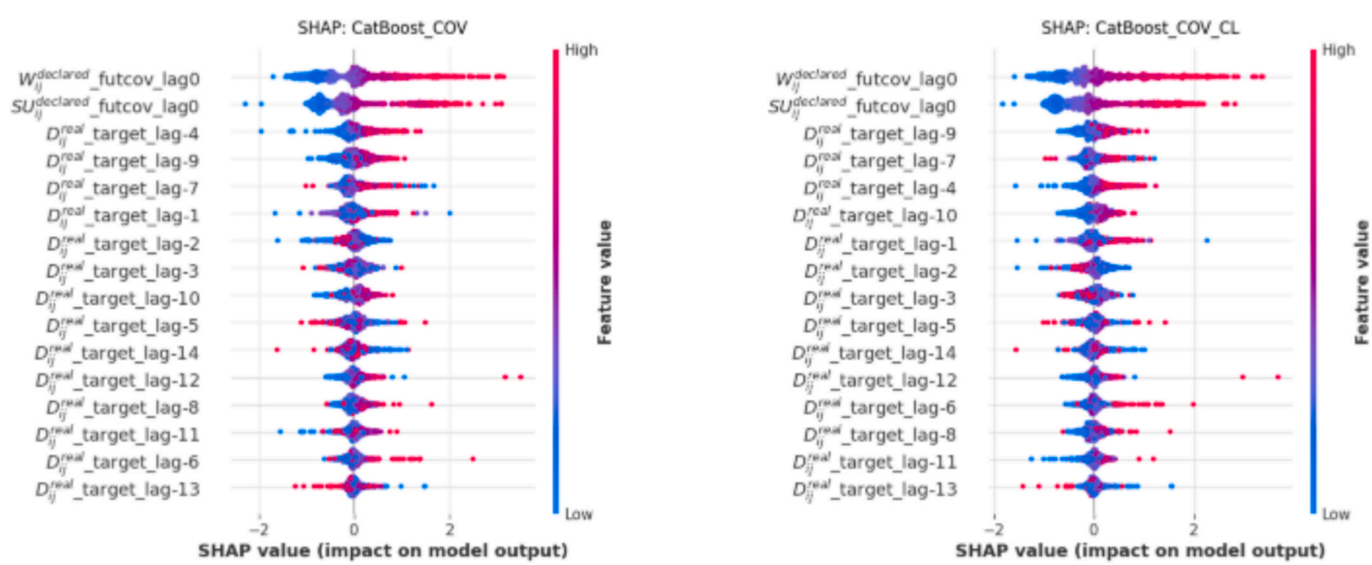

Fig. 11. Comparison of model explainability results of proposed approach with and without custom loss.

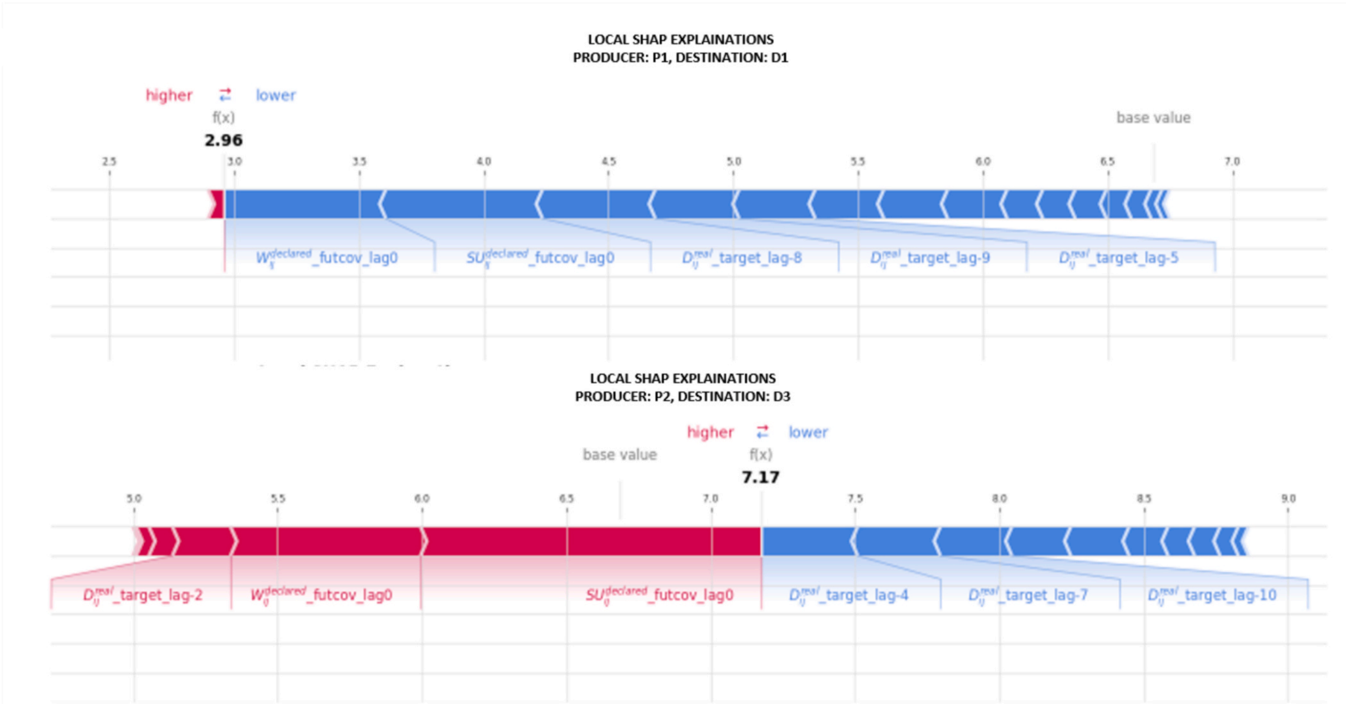

Fig. 12. Local SHAP explainability for 3PL forecasting.

shift toward greater emphasis on cost-critical high-volume cases.

While the global SHAP summaries provide a comprehensive view of the overall feature hierarchy, they do not reveal why a specific prediction takes a particular value. To address this, Fig. 12 presents two local SHAP force plots for anonymized producer–destination pairs. These plots show how the model's baseline forecast is adjusted upward or downward by the contribution of individual features in a single prediction.

In the first case (Producer P1, Destination D1), the model's baseline expectation of around 6.5 units is reduced to approximately 3 units, largely due to the downward influence of recent lags of realized demand and contemporaneous declared weights and units. In contrast, the second case (Producer P2, Destination D3) illustrates a forecast of roughly 7 units, where short-term signals such as demand at lag 2 and higher declared weights and units exert a strong positive influence, partially counterbalanced by longer-term lagged values.

These local explanations are particularly relevant in practice because they provide planners with an interpretable rationale for individual forecasts. When a prediction is driven by recent increases in declared weights or shipment units, planners can verify whether these signals reflect genuine operational changes, such as promotions or large orders, or whether they stem from anomalous data. Conversely, when the model emphasizes long-run demand lags, planners can understand that the forecast is grounded in historical consistency rather than short-term fluctuations. In both cases, the ability to trace forecasts back to their driving factors enhances trust in the system and supports more informed decisions on resource allocation and transport planning.

5. Managerial insights

The results of this study offer both operational guidance and cost trade-off considerations for 3PL providers seeking to improve forecasting performance and reduce logistics costs. By distinguishing these two perspectives, managers can better evaluate how to integrate the proposed forecasting approach into their daily workflows while balancing accuracy, efficiency, and investment.

5.1. Operational use

The case study confirms the practical value of adopting ML models in truck space forecasting. Among the tested approaches, CatBoost consistently outperformed traditional statistical models in terms of predictive accuracy. The introduction of a cost-sensitive loss function further allowed forecasts to reflect business priorities—favoring conservative estimates in cases where underestimation would result in service failures or expedited shipments. The analysis also highlights the importance of leveraging information-sharing mechanisms to enrich input data. Even simple producer-provided attributes (e.g., shipment weight, number of units) were shown to significantly improve forecast precision, suggesting that 3PLs can achieve substantial gains with relatively low-cost data integration technologies such as supplier portals and EDI protocols. Operationally, the proposed model is designed for daily use: forecasts are generated each morning by combining ex-ante data from producers with ex-post data from the warehouse management system. Results are delivered to the planning department by 6 AM, ensuring seamless integration into load planning. Retraining occurs monthly, with additional retraining triggered automatically if forecast errors exceed a threshold, ensuring model robustness over time.

5.2. Cost trade-off

While ML models with multivariate inputs and custom loss functions deliver superior accuracy, they do introduce moderately higher computational requirements. Nonetheless, training times remain well below a few seconds per model, making them fully feasible for daily planning operations. To implement the workflow, the case study 3PL deployed the system on a modest cloud-based server (8 vCPUs, 32 GB RAM) using simple orchestration tools, a centralized database, and a RESTful API to provide forecasts to planners. The infrastructure requires only limited technical expertise and can be scaled with data growth. From a cost perspective, cloud deployment at this scale typically incurs a few hundred euros per month, which is accessible to mid-sized 3PL providers while still offering scalability for larger operations. This demonstrates that the trade-off between accuracy and complexity remains favorable, with tangible cost savings outweighing infrastructure expenses.

6. Conclusions and research perspectives

The proposed study highlighted the critical relevance of a predictive framework to support 3PL providers in estimating their future requirements regarding truck space utilization for each customer. This requirement is indeed compounded by the imperfect and often unreliable information they receive. Given the disparity in bargaining power between 3PL companies and their customers, improving the quality of this incoming information is not always feasible, making accurate forecasting a vital tool for logistics efficiency.

To address this challenge, the study proposed a novel forecasting framework that integrates multivariate data collection with advanced ML techniques. Specifically, the CatBoost model was employed as the primary model due to its robust handling of complex data relationships, and a custom loss function was designed to prioritize forecast reliability in cost-sensitive scenarios. This approach was evaluated within a real-world 3PL case study in the food sector, comparing its performance against multiple benchmark models and data configurations to validate its effectiveness.

The results demonstrate that the proposed approach significantly benefits forecast accuracy and cost efficiency. The proposed solution substantially reduces forecast errors compared to traditional statistical models and univariate data collection. Furthermore, models using the custom loss function show improved cost performance, particularly under scenarios with imbalanced logistic costs related to the saturation and non-saturation of trucks. However, these gains come at the cost of increased computational effort, as the training time for CatBoost models is higher than that of both traditional models and other ML benchmarks.

While the results confirm the utility of the proposed approach, certain limitations must be acknowledged. First, the experimentation was conducted on a single case study, although many time series within that case were analyzed. Second, the study did not consider macroeconomic factors that could influence the space to reserve on trucks, as this was outside the scope of the investigation. Finally, the analysis focused on a single forecasting length, which, while representative of the primary operational horizon used in practice, limits the broader applicability of the findings to other forecasting scenarios. Finally, the definition of the hyperparameter search space was based on empirical experience, as no universally valid tuning strategy exists and the optimal ranges are inherently data-dependent. Although this represents a methodological limitation, it does not compromise the reproducibility of the study since the search space and optimization procedure are explicitly documented. Therefore, the results should be interpreted as domain-specific, and further validation in different 3PL contexts is needed before generalizing conclusions.

Future research could thus focus on validating the proposed approach across multiple case studies to ensure generalizability. In particular, testing the framework in different 3PL industry segments—such as pharmaceuticals, automotive, or high-value goods—could help assess its robustness and adaptability to varying logistics constraints and service levels. Additionally, incorporating macroeconomic indicators (e.g., fuel prices, inflation trends, seasonal demand shocks) into the input data may further enhance the forecasting model's responsiveness to external factors that influence space reservation dynamics. Another promising direction involves exploring hybrid approaches that integrate machine learning with optimization techniques. Such hybrid ML-optimization models could generate forecasts that are not only accurate but also directly actionable within cost-aware planning and routing tools, thus supporting end-to-end decision-making in 3PL operations. Moreover, the proposed framework is designed with scalability in mind: its reliance on modular data integration and cloud-based deployment ensures that it can be extended from single warehouses to multi-site networks without substantial changes in infrastructure. This scalability makes the approach particularly attractive for mid- to large-sized 3PL providers aiming to standardize forecasting practices across geographically distributed facilities. Considering the mentioned limitation, we reinforce the novel contribution of the proposed approach, which addresses the critical problem of truck pace reservation prediction in 3PL operations, demonstrating its effectiveness in a real case study and against several benchmarks.

CRediT authorship contribution statement

Matteo Gabellini: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Francesca Calabrese:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Francesco Gabriele Galizia:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Michele Ronchi:** Writing – review & editing, Writing – original draft, Conceptualization. **Alberto Regattieri:** Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Project administration, Funding acquisition, Conceptualization.

Funding

This study was carried out within the MICS (Made in Italy – Circular and Sustainable) Extended Partnership and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1551.11-10-2022, PE00000004). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Use of Generative AI Statement

Generative AI tools (ChatGPT, OpenAI, version GPT-4) were used solely to assist in improving the clarity and fluency of the manuscript's language. All content, including research findings, data interpretation, and scientific conclusions, was conceived, written, and validated entirely by the authors. The authors reviewed and verified all AI-assisted suggestions to ensure accuracy and integrity.

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.

Data availability

Data will be made available on request.

References

Abbasi, S., Saboury, A., & Jabalameli, M. S. (2021). Reliable supply chain network design for 3PL providers using consolidation hubs under disruption risks considering product perishability: An application to a pharmaceutical distribution network. *Computers and Industrial Engineering, 152*. https://doi.org/10.1016/j.cie.2020.107019

Abolghasemi, M., Rostami-Tabar, B., & Syntetos, A. (2023). The value of point of sales information in upstream supply chain forecasting: an empirical investigation. *International Journal of Production Research, 61*(7), 2162–2177. https://doi.org/10.1080/00207543.2022.2063086

Ali, A., Jayaraman, R., Azar, E., & Maalouf, M. (2024). Maximizing supply chain performance leveraging machine learning to anticipate customer backorders. *Computers & Industrial Engineering, 194*, 110414.

Babaveisi, V., Teimoury, E., Gholamian, M. R., & Rostami-Tabar, B. (2023). Integrated demand forecasting and planning model for repairable spare part: An empirical investigation. *International Journal of Production Research, 61*(20), 6791–6807. https://doi.org/10.1080/00207543.2022.2137596

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review, 54*(3), 1937–1967. https://doi.org/10.1007/s10462-020-09896-5
Bian, W., Yang, X., Li, S., Yang, X., & Hua, G. (2021). Advantages of 3PLs as healthcare supply chain orchestrators. *Computers and Industrial Engineering, 161*. https://doi.org/10.1016/j.cie.2021.107628
Brintrup, A., Pak, J., Ratiney, D., Pearce, T., Wichmann, P., Woodall, P., & McFarlane, D. (2020). Supply chain data analytics for predicting supplier disruptions: a case study in complex asset manufacturing. *International Journal of Production Research, 58*(11), 3330–3341. https://doi.org/10.1080/00207543.2019.1685705
Budak, A., Ustundag, A., & Guloglu, B. (2017). A forecasting approach for truckload spot market pricing. *Transportation Research Part A: Policy and Practice, 97*, 55–68. https://doi.org/10.1016/j.tra.2017.01.002
Che, C., Liu, B., Li, S., Huang, J., & Hu, H. (2023). Deep Learning for Precise Robot Position Prediction in Logistics. *Journal of Theory and Practice of Engineering Science, 3*(10), 36–41. https://doi.org/10.53469/jtpes.2023.03(10).05
Chen, A., Hsu, C.-H., & Blue, J. (2007). Demand planning approaches to aggregating and forecasting interrelated demands for safety stock and backup capacity planning. *International Journal of Production Research, 45*(10), 2269–2294. https://doi.org/10.1080/00207540600690693
Chien, C. F., Ku, C. C., & Lu, Y. Y. (2023). Ensemble learning for demand forecast of After-Market spare parts to empower data-driven value chain and an empirical study. *Computers & industrial engineering, 185*, 109670.
Chien, C.-F., Lin, Y.-S., & Lin, S.-K. (2020). Deep reinforcement learning for selecting demand forecast models to empower Industry 3.5 and an empirical study for a semiconductor component distributor. *International Journal of Production Research, 58*(9), 2784–2804. https://doi.org/10.1080/00207543.2020.1733125
Dai, Y., Dou, L., Song, H., Zhou, L., & Li, H. (2022). Two-way information sharing of uncertain demand forecasts in a dual-channel supply chain. *Computers & Industrial Engineering, 169*, 108162.
Doganis, P., Aggelogiannaki, E., & Sarimveis, H. (2008). A combined model predictive control and time series forecasting framework for production-inventory systems. *International Journal of Production Research, 46*(24), 6841–6853. https://doi.org/10.1080/00207540701523058
Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. http://arxiv.org/abs/1810.11363.
Dress, K., Lessmann, S., & von Mettenheim, H. J. (2018). Residual value forecasting using asymmetric cost functions. *International Journal of Forecasting, 34*(4), 551–565. https://doi.org/10.1016/j.ijforecast.2018.01.008
Fabbe-Costes, N., Jahre, M., & Roussat, C. (2009). Supply chain integration: The role of logistics service providers. *International Journal of Productivity and Performance Management, 58*(1), 71–91. https://doi.org/10.1108/17410400910921092
Goltsos, T. E., Syntetos, A. A., Glock, C. H., & Ioannou, G. (2022). Inventory – forecasting: Mind the gap. In European Journal of Operational Research (Vol. 299, Issue 2, pp. 397–419). Elsevier B.V. https://doi.org/10.1016/j.ejor.2021.07.040.
Grand View Research. (2023). Third-party Logistics Market Size, Share, & Trends Analysis Report By Service (DTM, ITM, VAL), By End-use (Retail, Manufacturing, Automotive), By Transport, By Region, And Segment Forecasts, 2024 - 2030.
Firouzi, F., Jaber, M. Y., & Baglieri, E. (2016). Trust in supply forecast information sharing. *International Journal of Production Research, 54*(5), 1322–1333. https://doi.org/10.1080/00207543.2015.1068961
Grzelak, M., Borucka, A., & Buczyński, Z. (2019). Forecasting the demand for transport services on the example of a selected logistic operator. *Archives of Transport, 52*(4), 81–93. https://doi.org/10.5604/01.3001.0014.0210
Gupta, A., Gusain, K., & Popli, B. (2016). Verifying the value and veracity of extreme gradient boosted decision trees on a variety of datasets. 11th International Conference on Industrial and Information Systems, ICIIS 2016 - Conference Proceedings, 2018-January, 457 – 462. https://doi.org/10.1109/ICIINFS.2016.8262984.
Gürbüz, F., Eski, İ., Denizhan, B., & Dağlı, C. (2019). Prediction of damage parameters of a 3PL company via data mining and neural networks. *Journal of Intelligent Manufacturing, 30*(3), 1437–1449. https://doi.org/10.1007/s10845-017-1337-z
Hooshmand Pakdel, G., He, Y., & Chen, X. (2025). Predicting customer demand with deep learning: an LSTM-based approach incorporating customer information. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2025.2468885
Islam, S., & Amin, S. H. (2020). Prediction of probable backorder scenarios in the supply chain using distributed Random Forest and Gradient Boosting Machine learning techniques. *Journal of Big Data, 7*(1). https://doi.org/10.1186/s40537-020-00345-2
Jauhar, S. K., Jani, S. M., Kamble, S. S., Pratap, S., Belhadi, A., & Gupta, S. (2024). How to use no-code artificial intelligence to predict and minimize the inventory distortions for resilient supply chains. *International Journal of Production Research, 62*(15), 5510–5534. https://doi.org/10.1080/00207543.2023.2166139
Kantasa-ard, A., Nouiri, M., Bekrar, A., Ait el cadi, A., & Sallez, Y. (2021). Machine learning for demand forecasting in the physical internet: a case study of agricultural products in Thailand. *International Journal of Production Research, 59*(24), 7491–7515. https://doi.org/10.1080/00207543.2020.1844332
Kay, M. G., & Warsing, D. P. (2009). Estimating LTL rates using publicly available empirical data. *International Journal of Logistics Research and Applications, 12*(3), 165–193. https://doi.org/10.1080/13675560802392415
Kmiecik, M. (2022). Logistics Coordination based on Inventory Management and Transportation Planning by Third-Party Logistics (3PL). *Sustainability (Switzerland), 14*(13). https://doi.org/10.3390/su14138134
Kmiecik, M. (2023). Supporting of manufacturer's demand plans as an element of logistics coordination in the distribution network. *Production Engineering Archives, 29*(1), 69–82. https://doi.org/10.30657/pea.2023.29.9
Kmiecik, M. (2025). Supply and demand prediction by 3PL for assortment planning. *Management Science Letters, 15*(2), 97–112. https://doi.org/10.5267/j.msl.2024.5.001
Kmiecik, M., & Wolny, M. (2022). FORECASTING NEEDS OF THE OPERATIONAL ACTIVITY OF a LOGISTICS OPERATOR. *Logforum, 18*(2), 197–212. https://doi.org/10.17270/J.LOG.2022.713
Kramarz, M., & Kmiecik, M. (2022). Quality of forecasts as the factor determining the Coordination of Logistics Processes by Logistic Operator. *Sustainability (Switzerland), 14*(2). https://doi.org/10.3390/su14021013
Lai, F., Li, D., Wang, Q., & Zhao, X. (2008). The information technology capability of third-party logistics providers: A resource-based view and empirical evidence from China. *Journal of Supply Chain Management, 44*(3), 22–38. https://doi.org/10.1111/j.1745-493X.2008.00064.x
Li, X., Krivtsov, V., Pan, C., Nassehi, A., Gao, R. X., & Ivanov, D. (2025). End-to-end supply chain resilience management using deep learning, survival analysis, and explainable artificial intelligence. *International Journal of Production Research, 63*(3), 1174–1202. https://doi.org/10.1080/00207543.2024.2367685
Lieb, R., & Miller, J. (2002). The use of Third-party Logistics Services by Large US Manufacturers, the 2000 Survey. *International Journal of Logistics Research and Applications, 5*(1), 1–12. https://doi.org/10.1080/13675560110114270
Lindsey, C., Frei, A., Mahmassani, H., Park, Y., Klabjan, D., Reed, M., Langheim, G., & Keating, T. (2014). Predictive analytics to improve pricing and sourcing in third-party logistics operations. *Transportation Research Record, 2410*, 123–131. https://doi.org/10.3141/2410-14
Long, Y., Kroeger, S., Zaeh, M. F., & Brintrup, A. (2025). Leveraging synthetic data to tackle machine learning challenges in supply chains: challenges, methods, applications, and research opportunities. *International Journal of Production Research*, 1–22.
Lorenc, A., Czuba, M., & Szarata, J. (2021). Big Data Analytics and Anomaly Prediction in the Cold Chain to Supply Chain Resilience. *FME Transactions, 49*(2), 315–326. https://doi.org/10.5937/fme2102315L
Mete Ayhan, H., & Kır, S. (2024). Ml-driven approaches to enhance inventory planning: Inoculant weight application in casting processes. *Computers and Industrial Engineering, 193*. https://doi.org/10.1016/j.cie.2024.110280
Nguyen, A., Pellerin, R., Lamouri, S., & Lekens, B. (2023). Managing demand volatility of pharmaceutical products in times of disruption through news sentiment analysis. *International Journal of Production Research, 61*(9), 2828–2839. https://doi.org/10.1080/00207543.2022.2070044
Park, T. K., & Kim, K. H. (2010). Comparing handling and space costs for various types of stacking methods. *Computers and Industrial Engineering, 58*(3), 501–508. https://doi.org/10.1016/j.cie.2009.11.011
Polim, R., Kumara Harold, S., & Marcus, I. (2017). Real-Time Shipment Duration Prediction.
Premkumar, P., Gopinath, S., & Mateen, A. (2021). Trends in third party logistics–the past, the present & the future. *International Journal of Logistics Research and Applications, 24*(6), 551–580. https://doi.org/10.1080/13675567.2020.1782863
Punia, S., Singh, S. P., & Madaan, J. K. (2020). A cross-temporal hierarchical framework and deep learning for supply chain forecasting. *Computers & Industrial Engineering, 149*, 106796.
Qiao, B., Pan, S., & Ballot, E. (2020). Revenue optimization for less-than-truckload carriers in the Physical Internet: Dynamic pricing and request selection. *Computers and Industrial Engineering, 139*. https://doi.org/10.1016/j.cie.2018.12.010
Saka, O. C., & Salman, F. S. (2024). An improved Freight Transportation Planning System for Less-Than-Truckload Operations of a Third-Party Logistics Carrier. *Networks*. https://doi.org/10.1002/net.22264
Sanders, N. R., & Graman, G. A. (2009). Quantifying costs of forecast errors: A case study of the warehouse environment. *Omega, 37*(1), 116–125. https://doi.org/10.1016/j.omega.2006.10.004
Shajalal, M., Hajek, P., & Abedin, M. Z. (2023). Product backorder prediction using deep neural network on imbalanced data. *International Journal of Production Research, 61*(1), 302–319. https://doi.org/10.1080/00207543.2021.1901153
Shao, J., Hong, J., Wang, M., & Wang, X. (2025). New energy vehicles sales forecasting using machine learning: The role of media sentiment. *Computers & Industrial Engineering*, 110928.
Silva, E., Oliveira, J. F., & Wäscher, G. (2016). The pallet loading problem: A review of solution methods and computational experiments. *International Transactions in Operational Research, 23*(1–2), 147–172. https://doi.org/10.1111/itor.12099
Terno, J., Scheithauer, G., Sommerweiû, U., & Riehme, J. (2000). An efficient approach for the multi-pallet loading problem. www.elsevier.com/locate/orms.
Tyan, J. C., Wang, F.-K., & Du, T. C. (2003). An evaluation of freight consolidation policies in global third party logistics. In. *Omega, 31*, Article www.sciencedirect.comwww.elsevier.com/locate/dsw.
Vlachos, I., & Reddy, P. G. (2025). Machine learning in supply chain management: systematic literature review and future research agenda. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2025.2466062
Wolny, M., & Kmiecik, M. (2025). Unveiling patterns in forecasting Errors: A Case Study of 3PL Logistics in Pharmaceutical and Appliance Sectors. *Sustainability (Switzerland), 17*(1). https://doi.org/10.3390/su17010214
Wong, E. Y. C., Tai, A. H., & Zhou, E. (2018). Optimising truckload operations in third-party logistics: A carbon footprint perspective in volatile supply chain. *Transportation Research Part D: Transport and Environment, 63*, 649–661. https://doi.org/10.1016/j.trd.2018.06.009
Xie, J., Lee, T. S., & Zhao, X. (2004). Impact of forecasting error on the performance of capacitated multi-item production systems. *Computers and Industrial Engineering, 46*(2), 205–219. https://doi.org/10.1016/j.cie.2003.12.020

Yang, J., & Yu, K. (2019). The role of an integrated logistics and procurement service offered by a 3PL firm in supply chain. *Journal of Management Analytics, 6*(1), 49–66. https://doi.org/10.1080/23270012.2019.1568921

Yang, M., Zhang, T., & Wang, C. X. (2021). The optimal e-commerce sales mode selection and information sharing strategy under demand uncertainty. *Computers & Industrial Engineering, 162*, 107718.

Yoon, J., Yildiz, H., & Talluri, S. S. (2016). Risk Management strategies in Transportation Capacity Decisions: An Analytical Approach. *Journal of Business Logistics, 37*(4), 364–381. https://doi.org/10.1111/jbl.12144

Zhao, X., & Xie, J. (2002). Forecasting errors and the value of information sharing in a supply chain. *International Journal of Production Research, 40*(2), 311–335. https://doi.org/10.1080/00207540110079121

Zheng, G., Kong, L., & Brintrup, A. (2023). Federated machine learning for privacy preserving, collective supply chain risk prediction. *International Journal of Production Research, 61*(23), 8115–8132. https://doi.org/10.1080/00207543.2022.2164628

Zhou, L. (2017). Prediction of a service demand using combined forecasting approach. *Journal of Physics: Conference Series, 887*(1). https://doi.org/10.1088/1742-6596/887/1/012075

Zhou, L., Heimann, B., & Clausen, U. (2006). Forecasting a Logistic Service Demand based on Neural Network. 2006 International Conference on Service Systems and Service Management, 1, 530–534. https://doi.org/10.1109/ICSSSM.2006.320518.