International Journal of Industrial Engineering, 30(6), 1558-1577, 2023

AN APPROACH BASED ON MACHINE LEARNING AND DISCRETE EVENT SIMULATION FOR SUPPLY CHAIN OPTIMIZATION: THE CASE OF ON-STOCK CHAINS

Zineb Nafi*, Fatima Ezzahra Essaber, Fatine Elharouni, and Rachid Benmoussa

Department of Industrial and Logistic Engineering
National School of Applied Sciences, Cadi Ayyad University
Marrakech, Morocco
*Corresponding author's e-mail: zineb.nafi@ced.uca.ma

The complexity of supply chain problems, more specifically the case of on-stock chains, is due to performance indicators variety, antagonism, and the difficulty of understanding the effects and interactions of different performance drivers with regard to these indicators. As mathematical formalization is essential to optimize the performance of these chains, this paper generally aims to study the contribution of Machine Learning to mathematically link the evaluation parameters of an on-stock supply chain to its action parameters. This work is based on an academic case study that seeks to mathematically formalize the problem of delivery delay in an on-stock supply chain. To this end, several Machine Learning algorithms have been tested and compared. This experience highlighted the impossibility of obtaining a labeled dataset through data collection from the real system. It thus demonstrates the necessity to use a simulation system, in particular, discrete event simulation, to generate this dataset.

Keywords: Machine Learning, On-stock Supply Chain, Mathematical Formalization, Simulation, Optimization.

(Received on October 5, 2022; Accepted on October 27, 2023)

1. INTRODUCTION

Supply chain management is a complex field with various challenges, including cost optimization, quality control, and meeting strict delivery schedules. In the context of on-stock supply chains, a specific procurement strategy is employed where purchased items are stored before being sent to customers. However, this approach introduces a phenomenon known as "decoupling," creating a disconnect between the procurement and expedition processes. While individual procurement actions may not directly affect specific expeditions, their cumulative impact significantly influences overall performance. To address these complexities, we utilize Discrete-Events Simulation (DES), a powerful analytical tool capable of dissecting and understanding each link of the supply chain independently, without requiring predefined connections, instead of Machine Learning (ML), which relies on labeled data and direct links. A labeled dataset consists of data points that are explicitly categorized or "labeled" to indicate their characteristics or outcomes.

Moreover, due to the inherent decoupling problem within on-stock supply chains, collecting such labeled datasets becomes challenging. The decoupling creates discontinuities and uncertainties that make it difficult to directly label data. As a result, applying machine learning techniques, which rely on these labeled datasets for training and prediction, becomes a complex endeavor in this context. To address this challenge, we turned to discrete event simulation (DES) as part of our methodology. DES allows us to model and understand each individual link within the supply chain, offering a clear and accurate representation without the need for pre-labeled data. By simulating various scenarios and events within the supply chain, DES generates data that inherently carries labels associated with each step and outcome. These outcomes serve as the labeled data points we require for training and informing our machine-learning models.

To comprehensively address the challenges within the global on-stock supply chain, we recognize the need for a dual approach. DES allows us to precisely represent and understand each link of the supply chain individually, offering a clear and accurate depiction. On the other hand, ML thrives when provided with labeled data sets. By combining the strengths of DES and ML, we can create a more holistic solution. While DES helps us map out each step in the supply chain, ML can leverage labeled data to generate formulas that aid in managing the discontinuities inherent to on-stock supply chains. These formulas facilitate the ability to optimize and streamline on-stock supply chain operations effectively, thereby enhancing overall performance (Benmoussa, 2021).

DOI: 10.23055/ijietap.2023.30.6.8613 ISSN 1943-670X

While we recognize the importance of on-stock supply chains, it's crucial to emphasize their unique challenges, particularly related to inventory management and delivery delays. These delays often stem from intricate inventory management complexities, supplier reliability issues, transportation challenges, and unexpected disruptions. This research paper aims to offer practical solutions and insights to mitigate the sources of delivery delays within on-stock supply chains. This approach involves considering manufacturing and expedition variables that impact these delays. By doing so, we aim to examine the entire on-stock supply chain and take into account the decoupling issues in these chains. By exploring the complexities of the on-stock chain, we aim to provide practical solutions that not only optimize logistic performance but also directly target the root causes of the decoupling problem within on-stock supply chains.

The primary objective is to create a mathematical model that bridges the gap between discrete-events simulation and machine-learning techniques. This model will serve to formalize the intricate connections between evaluation and action parameters within the global on-stock supply chain. By doing so, we intend to tackle the issue of decoupling, which is the disconnection between the procurement, manufacturing, and expedition processes in on-stock supply chains. By developing this model, we aim to bring more continuity and coherence to the supply chain's operations. This continuity is crucial because it enables us to optimize the flow of goods more effectively. By understanding the relationships between various supply chain components, we can identify bottlenecks and inefficiencies contributing to delays. Consequently, our formalized approach will facilitate more precise interventions, allowing for improved coordination of procurement, manufacturing and expedition processes. Ultimately, this will lead to reduced delivery delays and enhanced overall efficiency in on-stock supply chains.

In our quest, we present a structured approach that blends theoretical insights with practical testing. While our theories hold promise, it's essential to confirm their usefulness. To bridge this gap, we employ a method rooted in practical experience. Through an academic case study involving simulation and machine learning techniques, we aim to demonstrate how well our approach works in dealing with the challenges of on-stock supply chains. This case study seeks to develop an equation that links the theoretical delay created in an on-stock supply chain and the various action variables.

However, our goal extends beyond this academic case study. We intend to create a versatile model that can be adapted to different situations beyond our chosen case study. This model serves as a blueprint for understanding the relationship between action and evaluation parameters in on-stock supply chains. By combining theory with practical usefulness, this model becomes a valuable tool for improving on-stock supply chain performance and dealing with the decoupling problem. Through this effort, we aim to provide practical solutions that not only address the specific challenges of our academic case study but also have the potential to improve a wide range of on-stock supply chain situations. Hence, via this academic case study, this paper seeks to provide responses to the following research question:

How to combine machine learning and discrete event simulation in order to mathematically formalize the action and evaluation parameters and solve the decoupling problem in an on-stock supply chain so as to optimize its logistic performance?

This paper is organized as follows. The first section describes the literature review. The following section reveals the methodology adopted. Then, a third section is presented that discusses the academic case study's findings. The discussion and results section compares the different algorithms presented in the case study and eventually compares them with the other research works. The case study's results are also presented in the form of an equation that links the theoretical delay to the different evaluation variables. The paper ends with crucial conclusions and suggestions for further research.

2. LITERATURE REVIEW:

2.1 Supply Chain Optimization

Several methods have been used to optimize the supply chain's performance. In the case of a sustainable supply chain, a multi-objective optimization model was developed to optimize supply chain activities (Liu *et al*., 2019; Attiaa *et al*., 2019); (Gupta *et al*., 2022). First, it was applied in order to minimize the cost and the emissions and maximize the social benefit within the four phases of the supply chain (Gupta *et al*., 2022). To solve the optimization model, the weighted sum approach has been employed (Gupta *et al*., 2022). The two-stage multi-objective equilibrium optimization approach was utilized to evaluate demand uncertainty and manage distribution demand ambiguity (Liu *et al*., 2019). Using the same methodology, an optimization model that minimizes the total cost and maximizes revenue in a hydrocarbon supply chain (supply chain of oil and gas) was developed. Thus, sustainability and environmental constraints were used to reduce the depletion rate and restrict CO2 release (Attiaa *et al*., 2019).

Furthermore, a literature review has been conducted to investigate the application of optimization and simulation models, machine learning methods, and fuzzy techniques in sustainable transportation systems. Taking into account the complexity of sustainable transportation systems, it was asserted that hybrid methods, such as simulation with fuzzy optimization methods or optimization and machine learning methods, are suitable solutions in this case (Torre *et al*., 2021). On the other hand, the multi-objective linear programming optimization model was employed to provide a three objectives model concurrently optimizing stock, economic, and environmental issues in designing and handling modern sustainable supply chain networks (Yavari and Geraeli, 2019). Additionally, the mixed-integer linear programming (MILP) model has been used to minimize the cost and environmental pollutants. As well as a robust MILP model was developed for the problem under uncertainty (Bortolini *et al*., 2022). Besides, the reverse supply chain model with demand disruptions has been explored to drive optimal pricing, sustainability level, and the decisions of corporate social responsibility (CSR) (Hosseini-Motlagha *et al*., 2019).

In addition, stochastic mixed-integer programming has been used so as to improve the resilience and optimize supply chain operations under ripple effects driven by risks of regional pandemic disruption spreading from a single primary source location and inducing delayed regional disruptions of different durations in other regions (Sawik, 2022). An optimal solution was provided by using Mixed Integer Linear Programming and two other optimization methods, multiple regression and autoregressive integrated moving average (ARIMA) forecasting. The study seeks to estimate the quantity and size of Liquefied Natural Gas (LNG) bunker barges, as well as the ideal allocation and distribution network inside a ship-to-ship bunkering framework. Additionally, a robust mixed-integer linear programming model was used to reduce the vendor's expenses while forecasting LNG sales over a particular time horizon. (Doymus *et al*., 2022)

The bi-objective optimization model has been adopted for forest-based biomass supply chains and a blood supply chain simultaneously (Ahmadvand *et al*., 2021; Hosseini-Motlagh *et al*., 2019). While the objective of the first study was to minimize the upstream supply chain costs and the negative deviations of monthly inventory from the safety stock (Ahmadvand *et al*., 2021), the goal of the other research was to determine the optimal location-allocation as well as inventory management decisions and to reduce the overall cost of the supply chain, which include fixed costs, operating costs, inventory holding costs, wastage costs, and transport costs, in addition to minimizing the substitution levels to provide safer blood transfusion services. The robust optimization approach was combined with the TH method in order to reduce the uncertainty of the blood supply chain environment (Hosseini-Motlagh *et al*., 2019).

To enhance resilience in a healthcare supply chain during the COVID-19 pandemic, the multi-period multi-objective distributional robust optimization framework was applied. The goal was to provide reliable solutions over the trade-off between cost minimization and service level maximization by applying the ∊-constraint approach. (Ash *et al*., 2022). On the other hand, a non-linear and multi-objective optimization has been implemented to improve the resilience of a green supply chain (Hasani *et al*., 2020).

Robust optimization and Monte Carlo Simulation were combined, intending to optimize the forest-based biomass supply chain for syngas production at the tactical level considering uncertainties (Ahmadvand and Sowlati, 2022). To improve the efficiency of inventory management and, implicitly, an optimal replenishment time, the Q-learning was implemented (Wang and Lin, 2021). Moreover, the 5G Network and Markov Model have been utilized so as to optimize an industrial supply chain (Li *et al*., 2021). In addition, in order to design routes in the first place and to compute products and raw material flows, a standard decomposition technique was employed (Vitale *et al*., 2022). A framework has been developed to optimize the modular manufacturing supply chain.

The first step in the approach was to collect the data. The gathered data will be given as input to a classifier (Support Vector Machine (SVM), Decision trees). Then, after the assessment step, the algebraic equation of the predictor will be determined. In the final step, the feasibility of the process by incorporating all the classifiers is represented (Bhosekar and Ierapetritou, 2020). Machine learning was also used in a pharmaceutical supply chain, in which a comparison was carried out (Konovalenko and Ludwig, 2021). The projected stochastic gradient (PSG) method was employed to improve the supply chain management analysis efficiency (Alkahtani, 2022). A literature review has been carried out in order to present the research gap that concerns the simulation-optimization techniques for the design and evaluation of robust supply chain networks in unpredictable environments (Tordecilla *et al*., 2020). The artificial neural network was combined with a genetic algorithm aiming to optimize manufacturing resource configuration for small and medium-sized enterprises and also for model simulation and data relationship recognition simultaneously (Teerasoponpong and Sopadang, 2020). Edge computing was applied on agriculture supply chain architecture to optimize efficiency (Cui, 2021). Based on the fuzzy decision-making model in the internet of things, an optimization method was proposed to solve the Internet of Things (IoT) technology's fluency in the supply chain operation and allocation of resources that affect the supply chain and create management problems (Yue and Chen, 2018). Simulation-based optimization was deployed in the context of inventory management to control multi-echelon inventory taking into consideration the order uncertainty (Zhao and Wang, 2018).

So as to minimize the supply chain cost, the production capacity, batch size in each delivery, number of shipments, lead time, the chance of transition from uncontrolled to controlled state, safety factors, and backorder price reduction have been optimized (Sarkar and Chung, 2021). To solve the proposed model numerically, three different algorithms were deployed (Sarkar and Chung, 2021). Deep reinforcement learning was used in a blockchain-based Agri-Food supply chain so as to increase the company's profit and improve the effective traceability and management of agri-food products (Chen *et al*., 2021). A two-part advanced shipping system was proposed aiming to optimize the order dispatch operations and delivery time prediction in an intelligent logistics environment (Issaoui *et al*., 2022). On the other hand, an algorithm based on the

Generalized Benders Decomposition (GBD) method was employed in order to resolve the optimization model of routing, inventory, and location in the supply chain network design (SCND) problem (Zheng *et al*., 2019).

2.2 On-stock Supply Chain Optimization

We performed a literature review of the various methods that have been applied in on-stock supply chain optimization in the previous research.

The use of reinforcement learning (RL) was investigated to optimize the safety stock level and the order quantity rule in a linear chain of an independent agent by utilizing critic and actor neural networks. The analytical-based method was adopted to provide the optimal safety stock allocation and how much inventory should be kept in each location. Then, three algorithms of RL were tested to find the safety stock level from the agent states, agent actions, and environment model. The results suggest that Q-Learning is a successful method; nevertheless, this method expects a discrete action space, which does not exist in this case (Kosasih and Brintrup, 2021).

Several researches have been done concerning supplier selection and order allocation planning. In order to accomplish that, a forecasting procedure was employed with an optimization model. All the methods reviewed integrated various machine learning techniques in the supplier selection process and have focused on the supplier selection process (Islam *et al*., 2021). However, supplier selection and order allocation have been studied, and the Relational Regressor Chain (RRC) method, Holt's Linear Trend, and the Auto-Regressive Integrated Moving Average methods were integrated to forecast demand. The methods were tested and compared by applying The Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE) in order to measure the errors. In the second step, the forecasted demands are fed into the optimization model, which considers two different optimization techniques: the weighted-sum method and the ε-constraint technique. The two approaches were compared, and the one with a higher efficiency was chosen. Consequently, the RRC method outperformed two other regression methods, namely polynomial regression and support vector machine (Islam *et al*., 2021).

The artificial neural networks (ANNs) and a robust metamodel-based simulation-optimization approach were combined to determine near-optimal safety stock levels in a multi-product supply chain concerning deviations of its overall cost. The metamodel-based simulation optimization method was applied to find robust optimal inventory levels in the first step. Then, the ANN was used in the second step, where the decision and environment variables are defined as the coefficients of predefined inventory levels and the coefficients of demand variations, respectively (Sharifnia *et al*., 2021).

The method ARIMA was used in order to forecast the price of raw materials based on historical data and identify opportunities to buy the stock at a lower cost and sell a portion of the unused stock to generate additional profits for the organization. The eXtreme Gradient Boosting (XGBoost) regression model was employed to perform demand forecasting. A combinatorial optimization model is also used to decide whether to order or sell based on the forecasted price, demand, and other variables (Namir *et al*., 2021).

Rather than machine learning, for example, the Q-model was employed to predict better planning for the production or sale of other parts of the supply chain. Based on the cost of the non-availability of product in stock, the probability of demand, the unit cost of product shortage, and the number of orders per year, the Supplemental Inventory Cost can be recognized (Atyeh, 2020). However, A literature review was derived to determine the inventory drivers, and then choose the case study based on the pre-determined criteria to collect data. Based on the case study, the inventory drivers were determined. The simulation program was configurated to recommend the best supply chain configuration. The input data analysis of the simulation program was daily demand, input safety stock, and shipping quantities, and the output data were two key performance indicators: actual daily safety stock and actual daily cycle stock. According to this research findings, establishing a product classification yields more benefits than minimizing the lead time or boosting delivery frequency. Product classification is particularly beneficial if the goal is to reduce the amount of stock on hand (Chinello *et al.*, 2020).

Furthermore, a mixed-integer linear programming model is employed to optimize supply chains, specifically to model and optimize a food supply chain that considers a circular economy. The coffee cherries consumption, the consumption of water, the waste generation and CO2 emissions were included to evaluate the mathematical model (Baratsas *et al*., 2021). However, an approach was developed to optimize procurement, inventory, production, and distribution decisions for each period of the planning horizon. A set of cost variables was taken into account, raw material costs, pickup costs, delivery costs, production costs, and inventory costs, in order to precise another set of quantities: the quantity of raw material to be purchased from each raw material source, the quantity of each type of product to be manufactured in each plant, the pickup, and delivery routes to be used, the quantity of raw material to be collected by each pickup vehicle from any visited raw material-source, the number of products to be unloaded at any visited retailer by each delivery vehicle. As a result, this study proposes a decomposition approach for determining raw material and product flows in a linear multi-product supply chain (Cóccola *et al*., 2022).

Performance drivers are the variables that can identify, clarify, create, and drive final results. Hence, to build visibility on the supply chain and optimize its performance, it is crucial to understand the complexity of the interaction between different performance drivers. Several of the previously reviewed methods have focused on how to study these interactions and enhance the supply chain performance (Vignieri, 2016).

The suggested methods do not cover all the on-stock supply chain blocks; only the state elements of the global chains have been taken into account. We conclude that these methods are non-generalized. They concentrate on the specific rather than the global.

2.3 Delivery Delays in On-stock Supply Chains

The intricate issue of inventory control in a challenging environment marked by supply delays, concurrent orders, and an emergency supplier with no delays is addressed in one of the previous researches. The study's key objectives are motivated by the growing significance of effectively managing inventory within the context of just-in-time manufacturing and the prevailing trend toward smaller inventory orders. A stochastic control framework for inventory management is successfully established in the paper, illustrating that modeling instant supplies with zero lead times as a stochastic control problem with a reflection boundary equal to the cost of such supplies when the inventory is depleted (Agnes and Charles, 1995).

A comprehensive analysis is conducted on the pervasive supply-chain disruptions and extended delivery times observed in the post-COVID era. The primary focus lies in quantifying the far-reaching effects of these disruptions on the overall economy, utilizing a dynamic general equilibrium model enriched with input-output linkages and supply chain complexities. Significantly adverse effects are observed, with consumer prices experiencing a notable rise of approximately seven percent when deflated by wages. This surge is attributed to firms adjusting prices to cope with constrained supply. Firms, in response to the challenges, adapt their inventory strategies, influencing order timing and the balance between domestic and imported inputs (George *et al*., 2023).

In order to investigate strategies aimed at enhancing supply chain delivery performance, a cost-based analytical model is employed. This model serves to assess the expected penalty cost associated with both early and late deliveries. The central focus of the research is the optimization of the delivery window's position within supply chains, taking into consideration factors such as the width of the delivery window, penalty costs for untimely deliveries, and the parameters of the delivery time distribution. Through the development of analytical propositions, the paper analyzes the impact of these factors on delivery performance, with a specific emphasis on cost reduction strategies. The results demonstrate that increasing the width of the on-time portion of the delivery window, lowering penalties for untimely deliveries, and reducing the variance of the delivery time distribution can significantly improve delivery performance. These findings also provide valuable insights for negotiations in supplier and buyer contracts, as well as informed decision-making in investment initiatives focused on enhancing delivery performance (Maxim, 2018).

In order to tackle the challenge of optimizing the vendor-buyer cooperation strategy within a supply chain network characterized by constant demand and a variable delivery lead time distribution, the paper aims to minimize the integrated expected cost. The research takes into account key decision variables, namely the reorder point, delivery lot size, number of deliveries, and delivery time thresholds. Given the absence of readily available closed-form solutions, the paper utilizes diverse search procedures to identify integer solutions. Through the presentation of numerical results across various delivery lead time distributions, including uniform, exponential, and normal, the paper seeks to establish the broad applicability of the proposed model (Monami and Bhaba, 2021).

Supply chain optimization is a critical area of study, and various mathematical modeling and analysis techniques have been employed to address its complexity. Mathematical optimization models, such as linear programming and mixed-integer linear programming, are effective for cost minimization and resource allocation but face challenges in handling uncertainty and intricate interactions within supply chains. Multi-objective optimization, on the other hand, offers a holistic view by balancing conflicting goals like cost reduction and sustainability. However, it can be computationally intensive and challenging to implement. Machine Learning techniques, including regression analysis, decision trees, support vector machines, and neural networks, provide flexibility in handling complex data but often require substantial labeled datasets. Discrete event simulation is used to model supply chain processes, capturing real-world complexity, yet it demands significant computational resources. Some studies propose hybrid approaches, combining optimization, simulation, and machine learning to address diverse challenges. Transportation and distribution network optimization studies focus on route optimization, vehicle allocation, and demand forecasting. Inventory management researches delve into optimizing safety stock levels, order quantities, and supplier selection, a crucial aspect of on-stock supply chains. Finally, a subset of studies investigates strategies for supply chain resilience, particularly in the face of disruptions like pandemics, with the goal of enhancing adaptability to unforeseen events.

In conclusion, the field of supply chain optimization is marked by its complexity, data challenges, and computational intensity. Hybrid approaches, combining optimization, simulation, and Machine Learning, demonstrate the potential in providing holistic solutions to multifaceted supply chain problems. These insights collectively advocate for the integration of Machine Learning into the realm of on-stock supply chain optimization, particularly in the context of optimizing delivery performance and mitigating delivery delays, paving the way for more informed and adaptive decision-making in these intricate systems.

One notable gap in the existing body of research on on-stock supply chain optimization, which includes various aspects such as sustainability, inventory management, and transportation, is the limited attention given to the holistic and interconnected nature of these supply chains. While these studies offer valuable insights into individual components of supply chain management, they often overlook the critical issue of decoupling within the broader supply chain system. The decoupling problem, which arises when attempting to model and optimize the entire on-stock supply chain, becomes particularly evident when addressing delivery delays. Existing research tends to isolate specific aspects of the supply chain for optimization, thus failing to capture the intricate interactions and dependencies that emerge in a real-world, end-to-end supply chain scenario.

After a thorough examination of the research papers that delve into the enhancement of delivery performance within on-stock supply chains, it becomes evident that the majority of these studies have neglected to address the decoupling problem. These investigations predominantly concentrate on improving delivery efficiency from the supplier's perspective, emphasizing the reduction of delays, enhancement of delivery accuracy, and mitigation of associated penalty costs. In pursuit of these objectives, researchers often resort to simplifying assumptions to create manageable mathematical models and conduct empirical analyses. These assumptions typically involve envisioning a simplified supply chain structure with a linear progression from procurement to delivery. This simplified perspective intentionally sidesteps the complexities inherent in multi-stage supply chains, decision interdependencies, and dynamic coordination, complexities that are intrinsically linked to the decoupling problem but do not align with the immediate objectives of these studies.

Furthermore, a critical oversight in these studies is the lack of consideration for the inherent misalignment and absence of coordination between procurement and expedition processes within on-stock supply chains. The strategies required to synchronize these processes or explore the implications of their independence are rarely explicitly incorporated. Instead, the research predominantly centers around optimizing individual elements of the supply chain, such as lead time variability or delivery schedules, without examining their intricate interplay. Consequently, the primary aim of these studies is to provide practical solutions for enhancing delivery performance, which is undoubtedly a vital aspect of on-stock supply chain management. Nevertheless, it is apparent that addressing the decoupling problem presents a separate research challenge due to its distinct nature and considerable complexity. Therefore, there is a pressing need for dedicated research activities to explore the intricacies and obstacles associated with decoupling within the realm of on-stock supply chain management.

3. MATERIALS AND METHODS

In order to tackle the intricate decoupling challenge within on-stock supply chains, particularly the prevalent issue of delivery delays, our research endeavors to construct a comprehensive methodology that harnesses the power of ML techniques. This methodology serves as a bridge between DES and ML, facilitating the linkage of evaluation parameters within the on-stock chain, with a specific focus on delivery delays, to their corresponding action parameters. In this section, we delineate our systematic approach for achieving these research objectives. By establishing a robust connection between evaluation and action parameters in the context of on-stock supply chains, we gain the ability to discern the variables that contribute to the occurrence of delivery delays, and we also solve the decoupling problem. More importantly, this linkage empowers us to formulate a mathematical equation, a theoretical framework, which encapsulates the relationships among these parameters. This equation becomes a valuable tool for automating the optimization of on-stock chain activities. To clarify further, our mathematical equation will serve as a quantitative representation of the on-stock chain's dynamics, where various factors and operational constraints are mathematically defined. Leveraging mathematical optimization methods, we can fine-tune this equation based on real-world constraints. In essence, our research seeks to create a theoretical framework that integrates ML techniques with DES, enabling us to establish a mathematical equation representing on-stock supply chain dynamics. This equation, once calibrated with real-world constraints, facilitates the automatic optimization of on-stock chain activities and empowers proactive decision-making to mitigate delivery delays.

Step 1: Problem specification consists of describing the processes in the on-stock supply chain. Thus, it presents the problem to solve in this chain and specifies the evaluation and the decision variables.

Step 2: In the context of on-stock supply chains, obtaining labeled datasets, which are vital for machine learning, can be challenging due to the indirect linkage between action parameters and evaluation parameters. To overcome this challenge, we turn to discrete event simulation (DES). DES serves as a foundational step in our methodology, enabling us to model, simulate, and analyze the intricate dynamics of the supply chain system. By constructing a conceptual framework through DES, we create a simulated environment that closely mirrors the real-world supply chain.

The discrete-event simulation (DES) comprises modeling, simulating, and analyzing systems using computational and mathematical methodologies while constructing a model construct of a conceptual framework that explains a system. The system is simulated by executing experiments using a computer simulation of the model and analyzing the results to generate findings that aid in decision-making. Industry and academics have embraced discrete event simulation technologies to solve several industrial difficulties. By the conclusion of the last decades, the simulation software business is going through a phase of consolidation (Babulak and Wang, 2008).

Conceptual modeling: Conceptual modeling is critical to establish a formalism for modeling the principles of operating production and distribution processes (Babulak and Wang, 2008).

Developing the simulation model: The modeling phase includes the simulation project based on the parameters defined in the first context to gather the labeled data set required in the next step.

Data overview: This phase entails defining the set of parameters that would be used to implement the simulator program (Benmoussa, 2021).

Simulator programming: Programming the simulator ARENA consists of designing a model of the existing system and conducting experiments on this model, then interpreting the observations provided by the running of the model and formulating decisions relating to the system (Bouhenni, 2022).

Step 3: In this phase, it is fundamental to accurately examine the data set in order to understand their role and the impacts they can have on our prediction objective. This study involves a description of the data (name, type), as well as various processes such as cleaning (deleting useless data, searching for missing data). Finally, the combination between them, also called aggregation, in order to have a set of knowledge (observations) usable and appropriate for learning and achieving our goal (Vannieuwenhuyze, 2019).

The visualization of the data is one of the data exploration tools. (Sansen, 2017) The visual representation of the data allows an understanding of the distribution of the data set. Thereby, the data analysis helps understand the interaction between the variables before the data exploration comes.

Step 4: In this pivotal stage of our research, our primary objective is to identify the most accurate prediction model for our generated data, a critical step in our overarching goal of formalizing the on-stock supply chain. To accomplish this, we turn to the realm of ML, a field deeply rooted in pattern recognition and computational learning. Machine Learning offers us a potent toolkit to glean valuable insights from data and generate predictions, an essential endeavor in the complex world of supply chain management (Rupasinghe, 2017). Given the intricate nature of our research, which revolves around bridging the gap between evaluation parameters (independent variables) and action parameters (dependent variables) within the on-stock supply chain, selecting the appropriate category of ML algorithms becomes paramount. ML algorithms are categorized into three broad types: supervised, unsupervised, and reinforced learning models (Lickert *et al*., 2021).

In our specific case, the choice of supervised learning shines as the most fitting approach because of several causes. The first one is that supervised learning is explicitly engineered for predictive modeling, aligning seamlessly with our core objective of establishing a mathematical relationship between independent variables and dependent variables. Therefore, supervised learning thrives on labeled datasets comprising both input features and their corresponding target outputs. In our context, this means we can leverage historical data that intimately connects evaluation parameters with actual actions and their resulting outcomes. It also excels in crafting predictive models that can be expressed as mathematical formulas or equations. This perfectly aligns with our mission to derive a formula linking evaluation and action variables, enabling data-driven decision-making and optimization.

In our pursuit of establishing a mathematical formula that precisely links evaluation parameters to action parameters within the on-stock supply chain, regression algorithms emerge as the optimal choice over classification and decision tree algorithms. Regression techniques are inherently designed to model relationships between variables by identifying patterns, trends, and dependencies within data. Unlike classification, which is primarily concerned with assigning data points to discrete categories, and decision trees, which focus on classification and branching decision paths, regression algorithms are expressly geared towards the creation of mathematical equations that express the quantitative relationship between independent and dependent variables. This alignment with our goal of formulating a mathematical formula makes regression algorithms the preferred choice, as they enable us to express the nuanced, continuous connection between evaluation and action variables in a manner that classification and decision trees simply cannot replicate.

Regression: Regression analysis performs a sequence of parametric or non-parametric estimations. The method finds the causal relationship between the input and output variables. The estimation function can be determined by experience using a priori knowledge or visual data observation. Regression analysis aimed to understand how the typical values of the output variables change while the input variables are held unchanged (Franchitti, 2022).

Linear Regression: The linear equation describes the connection of dependent and independent variables. Thus, both input and output are numeric data, which must be cardinal to complete mathematical equations. This algorithm class is sensitive to data outliers and frequently fails when applied to noncorrected real data. The models for interpretability can be seen. (Lickert *et al.*, 2021)

Ridge/Lasso Regression: Ridge and lasso regression is a linear regression model; hence, the prediction formula is the same as conventional least squares. They use coefficients (w) that predict well on training data and fit an additional restriction. (Müller and Guido, 2016)

Polynomial Regression: Polynomial regression is a regression-type analysis that models the link between independent and dependent variables using nth-degree polynomials (Maulud and Abdulazeez, 2020).

Step 5: In this phase, we aim to present the results of our learning model. Moreover, it consists of adjusting and correcting the learning model if necessary.

Learning assessment: The first step entails the training algorithm by giving it a data set; then, we evaluate it on the second set of data to ensure the non-existence of an overlearning (Benmoussa, 2021).

Prediction: The model is deployed in production to make predictions and, if necessary, to retrain and improve the model using new input data (Benmoussa, 2021).

Overfitting and underfitting: Overfitting is a phenomenon where the solution is too well adapted to the training data and does not generalize to new and unknown data. Thus, if, for an algorithm, we obtain an accuracy of 99% on the training data and we obtain a value of 20% on the test data, it is likely that we are in the presence of overfitting. The phenomenon of underfitting occurs when the algorithm fails to find a correlation between the training data and, therefore, fails to make good predictions (Vannieuwenhuyze, 2019). Figure 1 represents the methodology employed.

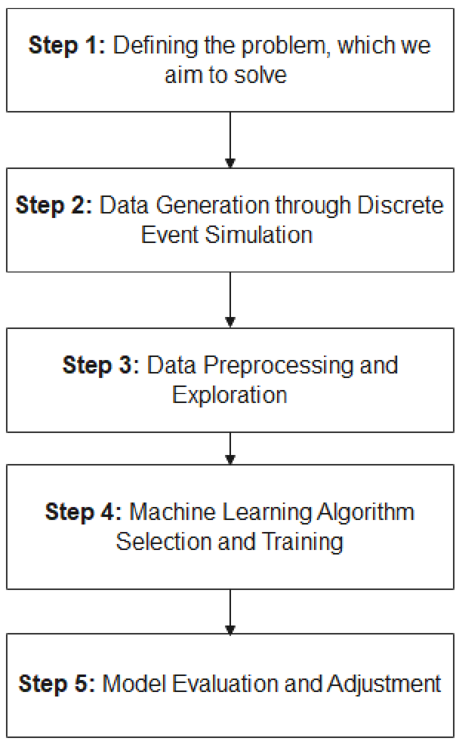

Figure 1. methodology

4. CASE STUDY

4.1 Problem Specification

We consider the situation of a corporation that manufactures on stock and distributes products on demand. The studied supply chain consists of two main processes: Manufacturing and Distribution. The pieces go through three steps of production during the manufacturing process. Each station is equipped with an operator. The pieces must be transported from station 1 to station 2 and then from station 2 to station 3 by a forklift; they must be assembled into a batch before moving from station 1 to station 2 or from 2 to 3.

An order triggers the distribution process. When the stock is depleted, the orders are put on hold until the stock is replenished. First, the orders would be prepared by an operator. The time it takes is proportional to the quantity ordered. Secondly, the delivery is done by two trucks. The time it takes is proportional to the round-trip distance to and from the customer. Due to the theoretical delay in delivery, customers do not receive their orders on time. This delay is caused by production and delivery-related factors. The objective is to develop a model that establishes a relationship between the theoretical delay and the other variables using machine learning to find mathematical modeling of the theoretical delay created in an on-stock supply chain.

This scenario illustrates the decoupling problem, where the disconnection between production and distribution processes leads to delays and operational challenges in meeting customer demands, and how the mathematical modeling via the DES and ML is a suitable approach not only to identify and analyze these delays but also to devise optimized strategies and interventions that can effectively mitigate these challenges, ultimately resulting in improved supply chain performance and enhanced customer satisfaction.

4.2 Data Generation Through Discrete Event Simulation

4.2.1 Conceptual Modeling

In the conceptual models depicted in figures 2 and 3, it becomes evident that the two fundamental processes within the on-stock supply chain, namely Manufacturing and Distribution, operate in a decoupled manner. This decoupling is of paramount significance as it highlights the distinctive objectives of each process. In the manufacturing process, the primary aim is to produce products and push them into inventory. This means that the production stations are geared towards generating products without direct consideration for specific customer orders. On the other hand, the distribution process focuses on handling customer orders, decreasing inventory, and preparing orders for delivery. The significance lies in the fact that these two processes are decoupled, meaning they operate independently and are not explicitly linked since the products will be manufactured whether a customer order has been received or not. This separation can lead to potential challenges. For instance, if customer orders exceed the quantity of products manufactured, there may be stock shortages, potentially resulting in unfulfilled orders and customer dissatisfaction. Conversely, if the manufacturing process produces more than the incoming orders, overstocking may occur, leading to increased carrying costs and potential obsolescence. Thus, the decoupling problem in this on-stock supply chain stems from the lack of synchronization between production and demand, underscoring the need for a comprehensive model to bridge this gap and optimize the supply chain's logistics performance.

Figures 2 and 3 illustrate the conceptual models associated with the manufacturing and distribution processes that are appropriate for our research:

4.2.2 Data Overview

For the production process, three resources are used: Operator Station1, Operator Station2, and Operator Station3, that are separated sequentially at the three manufacturing stations. Each station is executed according to the regular law Normal (10,2) minute. Between the three stations, two transfers are positioned, with each being performed batch by batch by a forklift. The first transfer occurs at a Normal (45,5) minute, while the second transfer occurs at a Normal (30,5) minute.

The integer values of the ordered quantities follow the triangle law Triangularine (1, 10, 15). Each order's preparation time is proportional to the quantity ordered: 0.1 hours * Quantity. The distribution is handled by two trucks, with a capacity of 20 products. The distance follows the triangle law Triangularine (5, 100, 400 km). The truck's average speed is 70 kilometers per hour.

4.2.3 Simulator Programming

The model created in figure 4 was converted into a simulation program in the ARENA simulator. Figure 4 shows the program that was built. In the simulation program using the ARENA simulator, the manufacturing process begins with the "Planification" module, which generates pieces at regular interval of every 10 minutes. These pieces then proceed to "Station 1," where each individual piece undergoes manufacturing with an "Operator Station1". This station follows a regular normal distribution (10,2) for its processing time, meaning that each product takes an average of 10 minutes to complete, with a standard deviation of 2 minutes. After completion at "Station 1," the pieces are individually sent to the "Batch T1" module. In the "Batch T1" module, these individual pieces are aggregated into a batch. The batched products are subsequently transferred batch by batch in the "Transfer 1" module using a forklift. The transfer time follows a normal distribution (45,5) minute.

The batches then progress to "Separate T1," where batches are separated into individual pieces. Each individual piece is subsequently processed in "Station 2". Each individual piece is processed with an “Operator Station2”, with processing times following the normal distribution (10,2) minutes. After processing, individual pieces are aggregated into batches in "Batch T2". The batched products are subsequently transferred batch by batch in the "Transfer 2" module using a forklift. The transfer time follows a normal distribution (30,5) minute.

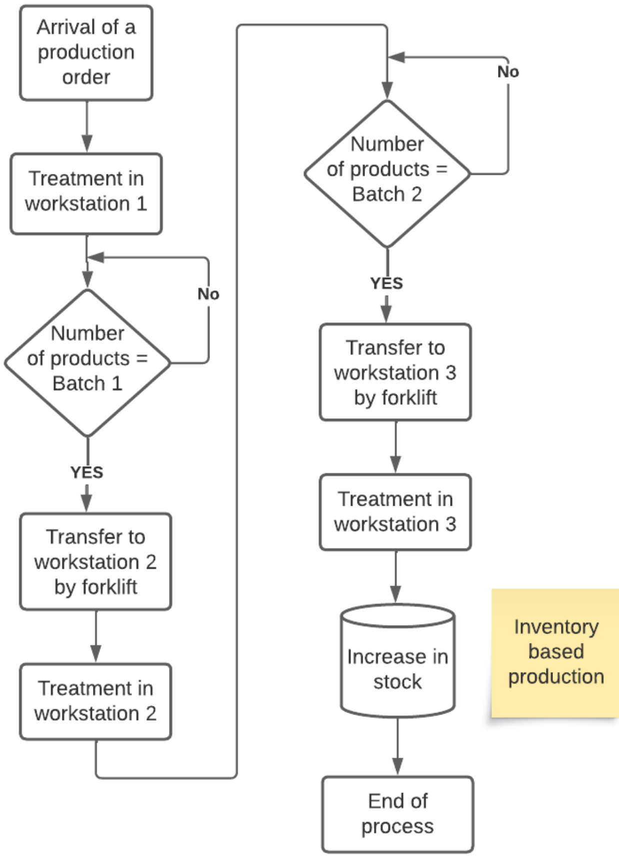

Figure 2. Manufacturing process

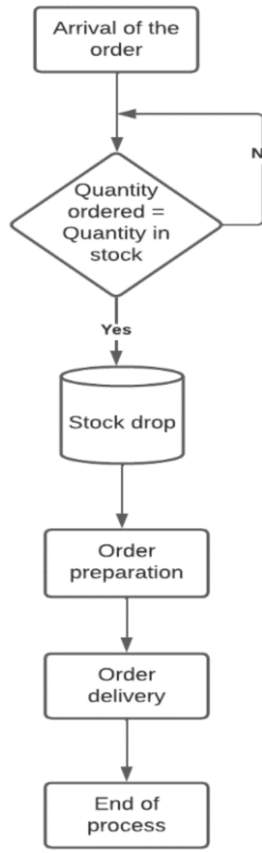

Figure 3. Distribution process

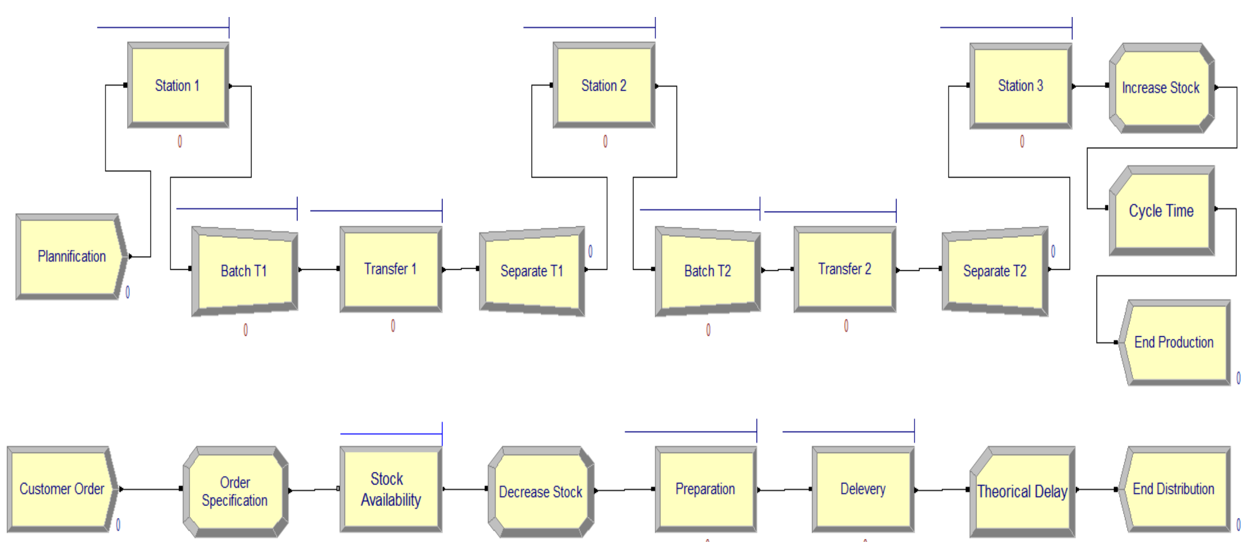

Figure 4. ARENA Flowchart

Following "Transfer 2," the process proceeds to "Separate T2," where batches are once again separated into individual pieces. Individual pieces then undergo manufacturing at "Station 3," with processing times following a normal distribution (10,2) minutes. The products move to an "Assign" module, which increments the stock levels, reflecting the completion of the manufacturing process. Subsequently, a "Record" module calculates the cycle time, and the process concludes with "Dispose," which ends the manufacturing process.

The delivery process begins with the "Customer Order" module, which generates customer orders each hour. These orders exhibit variability in terms of quantity, following a triangular distribution with parameters (1, 10, 15) reflecting diverse order sizes. Following by an "Assign" module in which we define the preparation time for each order and quantity, with the preparation time being directly proportional to the quantity ordered. Specifically, it is determined as 0.1 hours multiplied by the quantity. Afterward, a " hold " module is employed to conduct a capacity check. Here, the program verifies whether the quantity ordered is within the bounds of available stock, ensuring that orders can only be fulfilled if there is sufficient stock. Followed by an "Assign" module, which decrements the stock levels. The process enters the "Preparation" module, where a dedicated operator prepares the orders based on their specified quantities. Once the preparation is complete, the orders move on to the "Delivery" module. In this module, the expedition is carried out using two trucks. The distance traveled follows a triangular distribution with parameters (5, 100, 400 km), while the average speed of the trucks is set at 70 kilometers per hour.

Finally, the process concludes with "End Distribution" after a "Record" module, where the theoretical delay is calculated. This module provides insights into the system's performance and enables the evaluation of potential delays in fulfilling customer orders, contributing to a comprehensive understanding of the delivery process within the on-stock supply chain simulation program.

4.3 Data Preprocessing and Exploration

4.3.1 Data Preparation and Cleaning

We possess nine variables: Truck, Size Batch 1, Size Batch 2, Operator Preparation, Operator Station 1, Operator Station 2, Operator Station 3, Forklift, and Preparation Time; each variable has a minimum and maximum value as well as a step in between. Table 1 lists the nine variables and their associated values.

Table 1. Decision variables and their associated values

Category	Name	Element sort	Sort	Minimal value	Suggested value	Maximal Value	Step
Resources	Truck	Resource	Discrete	1	2	4	1
User Specified	Size Batch 1	Variable	Discrete	4	5	10	1
User Specified	Size Batch 2	Variable	Discrete	4	5	10	1
Resources	Operator Preparation	Resource	Discrete	1	1	2	1
Resources	Operator Station 1	Resource	Discrete	1	1	2	1
Resources	Operator Station 2	Resource	Discrete	1	1	2	1
Resources	Operator Station 3	Resource	Discrete	1	1	2	1
Resources	Forklift	Resource	Discrete	2	2	4	1
User Specified	Preparation Time	Variable	Discrete	0.05	0.1	0.3	0.05

We used OptQuest to explore all possible scenarios. Within OptQuest, we specified the variables, their potential values and theoretical delay. Moreover, we imposed constraints on the optimization process. These constraints are Operator Preparation + Operator Station 1 + Operator Station 2 + Operator Station 3 <= 6 and Truck + Forklift<= 7, as well as ensuring that truck utilization remains above 70% and forklift utilization above 60%. We obtained thus a CSV file that contains a total of 2602 combinations of the nine variables, each associated with its corresponding theoretical delay.

The results were exported, providing a comprehensive dataset for further analysis and modeling in our ML algorithms. In the first step, we separate the data that are going to be helpful in the ML regression algorithms. Thus, we attempted to construct a CSV file including the variables and the theoretical delay.

As mentioned in Scikit-learn documentation, the preprocessing package contains various standard utility functions and transformer classes that allow changing the raw feature vectors into a more suitable representation for the downstream estimators. We employed the StandardScaler utility class in the preprocessing module to normalize the data set to unify the scaling individual samples norm.

4.3.2 Exploring and Analyzing The Data

Mean, Standard Deviation, Min, Max and Quartiles

We calculated the mean and standard deviation of our evaluation variable to analyze the distribution of our data:

Table 2. Mean, standard deviation, min, max and quartiles

Variable	Number	Mean	Min	Max	Standard Deviation	Q1	Median	Q3
Theoretical Delay	2602	24.244	6.98	53.02	13.340	14.049	17.315	39.322

We conclude that 25% of the theoretical delay values are lower or equal to 14.049, also 50% and 70% are lower or equal to 17,315 and 39,322. The standard deviation value is equal to 13.340, which can be considered a high value. We conclude that the data are well dispersed. We notice that the minimum value is too small compared with the maximum, which returns to the fact that we took all the combinations possible of the variable's values and the corresponding theoretical delay value while respecting the constraint system.

Correlation of The Variables

We aim to study the relationship between each variable and the theoretical delay. In order to achieve that, we calculated the theoretical delay average for each variable's value. The following table represents the results obtained:

Table 3. The theoretical delay average for each variable values

Variables	Values	Theoretical delay
Truck	1	44,41699919
	2	20,81934568
	3	19,82843745
	4	21,1167013
Size Batch 1	4	27,7289564
	5	19,67499367
	6	19,88569968
	7	23,19078234
	8	22,66272141
	9	26,49546103
	10	27,43653814
Size Batch 2	4	27,62718618
	5	20,71853365
	6	21,2419297
	7	21,00666157
	8	22,88238834
	9	25,87225323
	10	27,33665876
Operator Preparation	1	29,12486912
	2	18,50694215
Operator Station 1	1	22,58128501
	2	26,47630319
Operator Station 2	1	24,21675438
	2	24,28848482
Operator Station 3	1	23,56914054
	2	24,97471264
Forklift	2	23,21891016
	3	21,9177095
	4	27,28788002

Preparation Time	0,05	19,39582642
	0,1	16,47386245
	0,15	16,52593282
	0,2	24,95410324
	0,25	30,51031009
	0,3	36,38465891

In our analysis of the nine variables and their connection to theoretical delay, several interesting trends emerged. To begin with, we observed a significant and strong negative correlation (approximately -0.767) between the number of trucks and theoretical delay. This indicates that as the number of trucks increased, the theoretical delay decreased. Furthermore, when we delved into Size Batch 1 and Size Batch 2, we found relatively weak positive correlations (about 0.351 and 0.279, respectively) with theoretical delay. Moreover, it was unnecessary to compute the correlation coefficient for the Operator Preparation, Operator Station 1, Operator Station 2, and Operator Station 3 since they take only two values, but it was clear that as the operator's value increased, there was a significant reduction in theoretical delay. Besides, our examination of Forklift and Preparation Time revealed strong positive correlations (approximately 0.726 and 0.891, respectively) with theoretical delay. This suggests a huge influence on the theoretical delay, implying that as Forklift and Preparation Time values increased, the theoretical delay also increased significantly.

Additionally, it's worth noting that while analyzing the theoretical delay, we observed random results for each variable; the average theoretical delay is supposed to decrease when the truck value increases, which was not our case; the values decrease, and when we reach four trucks, it increases to 21.11. These unexpected results can be attributed to the constraints we applied in our analysis. Specifically, we imposed constraints such as Operator Preparation + Operator Station 1 + Operator Station 2 + Operator Station 3 <= 6 and Truck + Forklift <= 7, as well as ensuring that truck utilization remains above 70% and forklift utilization above 60%. Therefore, when, for instance number of trucks is equal to 4, the number of forklifts should be less or equal to 3, which results in increasing the theoretical delay. These constraints likely introduced additional dynamics and complexities that influenced the relationship between variables and theoretical delay.

In conclusion, our analysis has highlighted several important relationships between these variables and theoretical delay. It is evident that the number of trucks, forklifts, and the values of operator preparation, operator station 1, operator station 2, and operator station 3 all play crucial roles in influencing theoretical delay. Understanding these relationships is essential for optimizing processes and minimizing theoretical delay in relevant scenarios.

4.4 Machine Learning Algorithm Selection and Training

In order to create the prediction model, the first step is to read the CSV file and define the model variables. When we use a random sampling method, for example, the bootstrap, with a satisfying number of repeats ($t \geq 100$) and a reasonable balance between training and test set (50–70% for training), we have a huge probability of getting a good model (Xu and Goodacre, 2018). Consequently, we have split the data set into two categories: 70% of the data was used for the training, and we left 30% for the test.

x_train: Parameter matrix (2000 rows, 9 columns)

y_train: Evaluation parameter vector (2000 rows, one column)

With x_train is the matrix that contains the set of nine variables Forklift, Preparation Time, Truck, Size Batch 1, Size Batch 2, Operator Preparation, Operator Station 1, Operator Station 2, Operator Station 3 and y_train represents the corresponding theoretical delays. We tested and compared four supervised machine learning algorithms to predict the theoretical delay value for each variable combination. The regression algorithms were manually developed using the linalg lstsq function, a defined function in the Numpy library, and it returns the least-squares solution to a linear matrix equation (Lathiya, 2022).

4.5 Visualizing The Results, Adjustment or Modification of The Learning Model

4.5.1 Learning Assessment

The training error was calculated to evaluate our learning algorithms: Table 4 summarizes the findings:

Table 4. The cumulative training errors of all created ML algorithms

	Training error in hours (h)
Lasso Regression	6,95
Linear Regression	6,93
Ridge Regression	6,80
Polynomial Regression	3,72

4.5.2 Prediction

The model was tested by 30% of the combinations, which remained after the training, to push it to predict new values with new inputs.

x_test: Parameter matrix (602 rows, 9 columns)

y_test: Evaluation parameter vector (602 rows, one column)

With x_test is the matrix that contains the set of nine variables: Forklift, Preparation Time, Truck, Size Batch 1, Size Batch 2, Operator Preparation, Operator Station 1, Operator Station 2, Operator Station 3 and y_train represents the corresponding theoretical delays. We have calculated the test error, which gives the following results.

Table 5. Test errors of all created ML algorithms

	Test Error (h)
Lasso Regression	7,52
Linear Regression	7,49
Ridge Regression	7,48
Polynomial Regression	3.90

4.5.3 Overfitting and Underfitting

In order to be precise if there is overfitting or underfitting in our models, we have measured the accuracy of each algorithm on both test and training data. We employed r2_score, defined in the Scikit-learn library, and calculated the coefficient of determination depending on Scikit-learn documentation. (Vannieuwenhuyze, 2019) The following table presents the findings:

Table 6. Training and test accuracy

	Training accuracy	Test accuracy
Lasso Regression	0.56	0.52
Linear Regression	0.59	0.51
Ridge Regression	0.60	0.88
Polynomial Regression	0.88	0.86

Furthermost, the training and test accuracy take near values. There is not a noticeable difference between the values. Besides, we have mentioned that overfitting is a phenomenon revealed that our model had only memorized our data set, which let it have a satisfying result in terms of the training data and very bad results with the test data, which proves the non-existence of overfitting. We also conclude that the program makes good predictions since we found a higher test accuracy in all algorithms. We could claim that there is no underfitting in our model.

5. DISCUSSION AND RESULTS:

The contribution of this paper is mainly not to artificial intelligence; we add nothing to this discipline. The contribution consists of a solution to a problem that conventional approaches (simulation) cannot address on its own and which the application of artificial intelligence in combination with these conventional methods can be of tremendous use. So as to optimize the on-stock supply chain, there are various approaches. However, all the cases of optimization took into account the optimization of the safety stock level and the order quantity, supplier's selection and order allocation planning, finding the near-optimal safety stock levels in a multi-product supply, and identifying the opportunities to buy the stock at a lower cost and to sale a portion of the unused stock, but none of these studies involves the optimization of the delivery delays with taking into account the global on-stock supply chain. Even those who studied the delivery delays they didn't take into account variables that cover the decoupling issues in the on-stock chains, like the variables that are defined in the production and procurement processes. It is essential to examine the entire chain while studying delivery delays in order to take into account the decoupling problem and optimize the logistic performance in the global on-stock supply chain. The existing studies did not cover the on-stock supply chain in its entirety, nor did they address the decoupling issues.

On the other hand, optimization methods that have been mostly used in the literature are the mixed-integer linear programming model, Machine Learning, simulation, and Q-model. These methods can be classified in three categories: machine learning, mathematical optimization approaches, or simulation. In some cases, authors used hybrid models combining both simulation and machine learning, where machine learning is used to predict variables, and the simulation is employed for optimization. The on-stock supply chains exhibit a unique decoupling phenomenon, where production and distribution processes operate independently until customer orders trigger the distribution phase.

Our research aims to fill this critical gap by examining the entire on-stock supply chain and addressing the decoupling issues inherent in it. To achieve this, we adopt a two-stage methodology that combines theoretical analysis and empirical validation. In the first stage, we conducted a thorough theoretical analysis that led us to develop an approach based on DES and ML so as to solve the decoupling problem in the on-stock chain and automatically create a link between evaluation and action variables (delivery delays) within this chain, and by doing so, we will optimize their logistic performance. This theoretical foundation forms the basis for our subsequent empirical investigation. In the second stage, we utilize an academic case study and apply the DES to have the labeled data set which we will need in the regression algorithms. Then, we applied these algorithms to mathematically formalize the relationship between the identified variables and delivery delays within on-stock supply chains. While the case study is not derived from real-world industrial data, it serves as a structured environment to explore the effectiveness of our approach to solve the decoupling problem in the on-stock supply chain. The mathematical formalization of the theoretical delay and the nine influencing variables in an on-stock chain contributes significantly to our research objectives in several ways:

Theoretical Understanding: It provides a theoretical foundation for understanding the complex interplay of variables that impact delivery delays. By mathematically modeling these relationships, we gain insights into the underlying mechanisms within on-stock supply chains.

Optimization Insights: The formalization offers actionable insights for optimizing inventory management and logistics activities. By quantifying the impact of variables like forklift availability, batch sizes, operator numbers, and more, we can make informed decisions to reduce delivery delays.

In essence, the formalization of the relationship between variables and delivery delays in on-stock supply chains not only demonstrates the applicability of DES and ML as valuable tools for tackling the decoupling problem in on-stock chains but also offers theoretical and practical contributions to the broader field of supply chain management by enhancing our understanding of delivery delay dynamics and optimization possibilities.

In terms of results, we picked the polynomial regression algorithm, which had the best accuracy of modeling. We define Y and C as follows:

$Y = (1, X_1^2, X_2^2, X_3^2, X_4^2, X_5^2, X_6^2, X_7^2, X_8^2, X_9^2, X_1X_2, X_1X_3, X_1X_4, X_1X_5, X_1X_6, X_1X_7, X_1X_8, X_1X_9, X_2X_3, X_2X_4, X_2X_5, X_2X_6, X_2X_7, X_2X_8, X_2X_9, X_3X_4, X_3X_5, X_3X_6, X_3X_7, X_3X_8, X_3X_9, X_4X_5, X_4X_6, X_4X_7, X_4X_8, X_4X_9, X_5X_6, X_5X_7, X_5X_8, X_5X_9, X_6X_7, X_6X_8, X_6X_9, X_7X_8, X_7X_9, X_8X_9, X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9)$

$C = (C_{1,1}, C_{2,2}, C_{3,3}, C_{4,4}, C_{5,5}, C_{6,6}, C_{7,7}, C_{8,8}, C_{9,9}, C_{10,10}, C_{1,2}, C_{1,3}, C_{1,4}, C_{1,5}, C_{1,6}, C_{1,7}, C_{1,8}, C_{1,9}, C_{1,10}, C_{2,3}, C_{2,4}, C_{2,5}, C_{2,6}, C_{2,7}, C_{2,8}, C_{2,9}, C_{2,10}, C_{3,4}, C_{3,5}, C_{3,6}, C_{3,7}, C_{3,8}, C_{3,9}, C_{3,10}, C_{4,5}, C_{4,6}, C_{4,7}, C_{4,8}, C_{4,9}, C_{4,10}, C_{5,6}, C_{5,7}, C_{5,8}, C_{5,9}, C_{5,10}, C_{6,7}, C_{6,8}, C_{6,9}, C_{6,10}, C_{7,8}, C_{7,9}, C_{7,10}, C_{8,9}, C_{9,10}, C_{10,10})$
$= (2.6, 5.3, -4.49, 5.48, -8.45, 4.55, -3.04, -2.45, 1.19, 2.45, -3.28, 1.92, 6.03, 1.71, 1.58, 1.63, 1.65, -8.27, 1.3, -3.55, -1.22, 3.08, 2.35, -6.13, 4.54, 7.81, 2.06, -1.01, 6.72, -5.19, -8.06, -2.27, 6.52, -4.44, 2.09, 2.47, -9.78, -1.81, -1.08, -1.13, 8.22, -5.67, -7.80, -2.02, -9.43, -2.14, -1.76, 4.18, 9.13, -3.61, -2.47, 1.01, -3.9, 1.13, 5.31)$

The values of the vector C are obtained from the outputs of the polynomial regression model. The polynomial function is expressed in the provided form, as depicted in equation 1 below:

$$P(X) = C^T Y \tag{1}$$

In order to simplify the expression of the equation, we chose to write it in the matrix format, and we added the constraint system. We define X as follows:

$X = (1\ X_1\ X_2\ X_3\ X_4\ X_5\ X_6\ X_7\ X_8\ X_9)$

Matrix A, on the other hand, is a coefficient matrix used in the polynomial regression equation to predict the theoretical delay. In polynomial regression, matrix A typically represents the coefficients that are multiplied by the input variables (X1 to X9 in this case) to estimate the output variable (theoretical delay). Each element in matrix A corresponds to a specific combination of input variables and their impact on the predicted variable in the described way:

$$A = \begin{pmatrix} 2.6 & -3.28 & 1.92 & 6.03 & 1.71 & 1.58 & 1.63 & 1.65 & -8.27 & 1.3 \\ 0 & 5.3 & -3.55 & -1.22 & 3.08 & 2.35 & -6.13 & 4.54 & 7.81 & 2.06 \\ 0 & 0 & -4.49 & -1.01 & 6.72 & -5.19 & -8.06 & -2.27 & 6.52 & -4.44 \\ 0 & 0 & 0 & 5.48 & 2.09 & 2.47 & -9.78 & -1.81 & -1.08 & -1.13 \\ 0 & 0 & 0 & 0 & -8.45 & 8.22 & -5.67 & -7.80 & -2.02 & -9.43 \\ 0 & 0 & 0 & 0 & 0 & -4.55 & -2.14 & -1.76 & 4.18 & 9.13 \\ 0 & 0 & 0 & 0 & 0 & 0 & -3.04 & -3.61 & -2.47 & 1.01 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2.45 & -3.9 & 1.13 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.19 & 5.31 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.45 \end{pmatrix}$$

Table 8. Endogenous variables' signification

Endogenous variable	Meaning
X_1	Number of resources (Truck)
X_2	The size of the first batch (Batch Size 1)
X_3	The size of the second batch (Batch Size 2)
X_4	Number of resources (Operator Preparation)
X_5	Number of resources (Operator Station 1)
X_6	Number of resources (Operator Station 2)
X_7	Number of resources (Operator Station 3)
X_8	Number of resources (Forklift)
X_9	Time to prepare an order (Preparation Time)

Table 9. Exogenous variables' signification

Exogenous variable	Meaning
O_1	The Forklift's occupation rate in the first transfer
O_2	The Forklift's occupation rate in the second transfer
B_1	The limit of the company's budget for the Forklift
B_2	The limit of the company's budget for truck
Lop	The boundary of the existed resources (Operator Preparation)
Los_1	The boundary of the existed resources (Operator Station 1)
Los_2	The boundary of the existed resources (Operator Station 2)
Los_3	The boundary of the existed resources (Operator Station 3)

Tables 8 and 9 provide a breakdown of the variables and their meanings in the context of the supply chain optimization model. These variables are divided into two categories: endogenous and exogenous. Endogenous variables represent parameters that are determined within the system or model itself, and their values depend on decisions or processes within the supply chain. For example, "Number of resources (Truck)" represents the number of trucks used in the supply chain, while "Number of resources (Operator Preparation)" signifies the count of operators involved in the preparation phase. On the other hand, exogenous variables are external factors that influence the system but are not determined within the model. These can include budget constraints like "The limit of the company's budget for the forklift (B1)" or resource limits such as "The boundary of the existing resources (Operator Station 1) (Los1)." These variables collectively define the environment and constraints in which the supply chain operates. We have not specified specific values for these constraints, as they can vary depending on the specific needs and conditions of the company, including, for example, the financial constraints. We present the results in the form of equation (2) along with its associated constraint system as follows:

$$\text{Theoretical delay} = X^T A X \quad (2)$$

where

$X_2 \leq O_1$,
$X_3 \leq O_2$,
$X_8 \leq B_1$,
$X_1 \leq B_2$,
$X_4 \leq Lop$,
$X_5 \leq Los_1$,
$X_6 \leq Los_2$,
$X_7 \leq Los_3$.

6. CONCLUSION:

This research paper addresses the complexities of on-stock supply chains, highlighting the challenge of the decoupling between manufacturing, procurement and expedition processes, with special attention to delivery delays. The primary objective is to bridge the gap between discrete-events simulation (DES) and machine learning (ML) and create a mathematical model that formalizes the relationship between action and evaluation parameters in on-stock supply chains. This model aims to optimize logistical performance by reducing delivery delays and enhancing overall efficiency. The paper employs a structured approach, combining theoretical insights with practical testing through an academic case study that demonstrates the model's applicability. The ultimate goal is to improve the on-stock supply chain performance in an ever-changing global landscape. To achieve this, we involved the use of an academic case study as empirical validation of our approach. Here, we applied regression algorithms to mathematically formalize the relationship between the identified variables and delivery delays. By applying data-driven modeling techniques, we aimed to bridge the gap between theory and practical application, offering actionable insights for supply chain optimization. In future research, there is potential to expand upon this study by integrating the developed algorithms into an optimization framework aimed at minimizing the theoretical delay. This optimization approach would take into account the financial constraints specified by the company and incorporate the equation derived from the regression algorithms. This effort should include a thorough investigation into the interplay between the variables involved. It's important to note that within the equation, there may be various terms that can be pruned, and determining a purification factor becomes essential to guide decisions about which terms to retain or eliminate. Additionally, it would be advantageous to explore and compare various other algorithms to increase the chances of identifying the most effective technique for modeling and addressing delivery delays. On the other hand, real-world scenarios may be explored as they can consider the non-technical aspects of our method concerning the challenges that may arise in project management.

REFERENCES :

Agnes, S., and Charles, S. T. (1995). Inventory Control with Supply Delays, on Going Orders and Emergency Supplies. Ifac Control Applications of Optimization. DOI: https://doi.org/10.1016/s1474-6670(17)43685-8.

Ahmadvand, S., Khadivi, M., Arora, R., and Sowlati, T. (2021). Bi-Objective Optimization of Forest-Based Biomass Supply Chains for Minimization of Costs and Deviations from Safety Stock. Energy Conversion and Management: X, 11: 14.

Ahmadvand, S., and Sowlati, T. (2022). A Robust Optimization Model for Tactical Planning of The Forest-Based Biomass Supply Chain for Syngas Production. Computers and Chemical Engineering, 159.

Alkahtani, M. (2022). Supply Chain Management Optimization and Prediction Model Based on Projected Stochastic Gradient. Sustainability, 14(6).

Ash, C., Diallo, C., Venkatadri, U., and Vanberkel, P. (2022). Distributionally Robust Optimization of A Canadian Healthcare Supply Chain to Enhance Resilience During The Covid-19 Pandemic. Computers & Industrial Engineering, 168.

Attiaa, A. M., Ghaithanb, A. M., and Duffuaaa, S. O. (2019). A Multi-Objective Optimization Model for Tactical Planning of Upstream Oil & Gas Supply Chains. Computers and Chemical Engineering. DOI: https://doi.org/10.1016/j.compchemeng.2019.06.016.

Atyeh, M. A. (2020). Supply Chain Management, Optimization and Forecasting Techniques. Journal of Business Sciences

(Jbs).
Babulak, E., and Wang, M. (2008). Discrete Event Simulation: State of The Art. International Journal of Online Engineering, 60-63.

Baratsas, S. G., Pistikopoulos, E. N., and Avraamidou, S. (2021). A System Engineering Framework for The Optimization of Food Supply Chains Under Circular Economy Considerations. Science of The Total Environment, 794.

Benmoussa, R. (2021, 23 Mar 2021). De La Simulation A Evénement Discret Au Machine Learning Pour La Chaîne Logistique 13ème Conference Internationale De Modelisation, Optimisation Et Simulation (Mosim2020), Maroc-Agadir. DOI: https://hal.archives-ouvertes.fr/hal-03177538/.

Bhosekar, A., and Ierapetritou, M. (2020). A Framework for Supply Chain Optimization for Modular Manufacturing with Production Feasibility Analysis. Computers and Chemical Engineering, 145.

Bortolini, M., Calabrese, F., Galizia, F. G., and Mora, C. (2022). A Three-Objective Optimization Model for Mid-Term Sustainable Supply Chain Network Design. Computers & Industrial Engineering, 168(164). DOI: https://doi.org/10.1016/j.cie.2022.108131. Bouhenni, I. Memoire Online

Chen, H., Chen, Z., Lin, F., and Zhuang, P. (2021). Effective Management for Blockchain-Based Agri-Food Supply Chains Using Deep Reinforcement Learning. IEEE Access, 9.

Chinello, E., Herbert-Hansen, Z. N. L., and Khalid, W. (2020). Assessment of The Impact of Inventory Optimization Drivers in A Multi-Echelon Supply Chain: Case of A Toy Manufacturer. Computers & Industrial Engineering, 141.

Cui, H. (2021). Research on Agricultural Supply Chain Architecture Based on Edge Computing and Efficiency Optimization. Special Section on Deep Learning for Internet of Things, 10.

Cóccola, M. E., Basán, N., Méndez, C. A., and Dondo, R. G. (2022). Optimization of Resource Flows Across The Whole Supply Chain. Application to A Case Study in The Dairy Industry. Computers and Chemical Engineering, 158.

Doymus, M., Sakar, G. D., Yildiz, S. T., and Acik, A. (2022). Small-Scale Lng Supply Chain Optimization for Lng Bunkering in Turkey. Computers and Chemical Engineering, 106.

Franchitti, J.-C. (2022). Cloud-Based Machine & Deep Learning, Supervised Machine Learning Algorithms.

George, A. A., Shafaat, Y. K., Armen, K., Carter, B. M., and Kim, J. R. (2023). The Aggregate Effects of Global and Local Supply Chain Disruptions: 2020–2022. Journal of International Economics. DOI: https://doi.org/10.1016/j.jinteco.2023.103788.

Gupta, P., Mehlawat, M. K., Aggarwal, U., and Khan, A. Z. (2022). An Optimization Model for A Sustainable and Socially Beneficial Four-Stage Supply Chain. *Information Sciences*. DOI: https://doi.org/10.1016/j.ins.2022.02.032.

Hasani, A., Mokhtari, H., and Fattahi, M. (2020). A Multi-Objective Optimization Approach for Green and Resilient Supply Chain Network Design: A Real-Life Case Study. *Journal of Cleaner Production*, 278.

Hosseini-Motlagha, S.-M., , Nouri-Harzvili, M., Choi, T.-M., and Ebrahimi, S. (2019). Reverse Supply Chain Systems Optimization with Dual Channel and Demand Disruptions: Sustainability, CSR Investment and Pricing Coordination. *Information Sciences*, 503. DOI: https://doi.org/10.1016/j.ins.2019.07.021.

Hosseini-Motlagh, S. M., Samani, M. R. G., and Homaei, S. (2019). Blood Supply Chain Management: Robust Optimization, Disruption Risk, and Blood Group Compatibility (A Real-Life Case). *Journal of Ambient Intelligence and Humanized Computing*, 11(4).

Islam, S., Amin, S. H., and Wardley, L. J. (2021). Machine Learning and Optimization Models for Supplier Selection and Order Allocation Planning. *International Journal of Production Economics*, 242.

Issaoui, Y., Khiat, A., Haricha, K., Bahnasse, A., and Ouajji, H. (2022). An Advanced System to Enhance and Optimize Delivery Operations in A Smart Logistics Environment. IEEE Access, 10.

Konovalenko, I., and Ludwig, A. (2021). Comparison of Machine Learning Classifiers: A Case Study of Temperature Alarms in A Pharmaceutical Supply Chain. *Information Systems*, 100.

Kosasih, E. E., and Brintrup, A. (2021). Reinforcement Learning Provides A Flexible Approach for Realistic Supply Chain Safety Stock Optimisation

Lathiya, A. (2022). How to Calculate Square of Numpy Array.

Li, L., Wu, H., Yuan, Y., and Zhou, L. (2021). Industrial Supply Chain Optimization Based on 5g Network and Markov Model. *Computers & Industrial Engineering*, 80.

Lickert, H., Wewer, A., Dittmann, S., Bilge, P., and Dietrich, F. (2021). Selection of Suitable Machine Learning Algorithms for Classification Tasks in Reverse Logistics Cirpe 2020 – 8th Cirp Global Web Conference – Flexible Mass Customisation,

Liu, Y., Chen, Y., and Yang, G. (2019). Developing Multiobjective Equilibrium Optimization Method for Sustainable Uncertain Supply Chain Planning Problems. IEEE Transactions on Fuzzy Systems DOI: https://doi.org/10.1109/tfuzz.2018.2851508.

Maulud, D. H., and Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, 140-147.

Maxim, A. B. (2018). Delivery Performance Improvement in Two-Stage Supply Chain. International Journal of Production Economics, 195, 66-73. DOI: https://doi.org/10.1016/j.ijpe.2017.10.007.

Monami, D. R., and Bhaba, R. S. (2021). Optimizing A Supply Chain Problem with Nonlinear Penalty Costs for Early and Late Delivery Under Generalized Lead Time Distribution. Computers & Industrial Engineering, 160. DOI: https://doi.org/10.1016/j.cie.2021.107536.

Müller, A. C., and Guido, S. (2016). Introduction to Machine Learning with Python.

Namir, K., Labriji, H., and Lahmar, E. H. B. (2021). Decision Support Tool for Dynamic Inventory Management Using Machine Learning, Time Series and Combinatorial Optimization International Workshop on * Information Sciences and Advanced Technologies: Edge Big Data - AI- IoT (Isat 2021),

Rupasinghe, T. D. (2017). Machine Learning-Based Parameter Tuned Genetic Algorithm for Energy Minimizing Vehicle Routing Problem. *Journal of Industrial Engineering*.

Sansen, J. (2017). La Visualisation D'information Pour Les Données Massives : Une Approche Par L'abstraction De Données L'université De Bordeaux Ecole Doctorale De Mathématiques Et D'informatique]. Hal.

Sarkar, M., and Chung, B. D. (2021). Controlling Random Lead Time Demand Through A Flexible Production System Under Supply Chain Management. IEEE Access, 9.

Sawik, T. (2022). Stochastic Optimization of Supply Chain Resilience Under Ripple Effect: A Covid-19 Pandemic Related Study. Omega, 109.

Sharifnia, S., Mohammad, E., Sajjad, A., Rupy, S., and Hwangbo, H. (2021). Robust Simulation Optimization for Supply Chain Problem Under Uncertainty Via Neural Network Metamodeling. *Computers & Industrial Engineering*, 162.

Teerasoponpong, S., and Sopadang, A. (2020). A Simulation-Optimization Approach for Adaptive Manufacturing Capacity Planning in Small and Medium-Sized Enterprises. *Expert Systems with Applications*, 168.

Tordecilla, R. D., Juana, A. A., Montoya-Torres, J. R., Quintero-Araujod, C. L., and Panadero, J. (2020). Simulation-Optimization Methods for Designing and Assessing Resilient Supply Chain Networks Under Uncertainty Scenarios: A

Review. *Simulation Modelling Practice and Theory*, 106.

Torre, R. D. L., Corlu, C. G., Faulin, J., Onggo, B. S., and A., A. (2021). Simulation, Optimization, and Machine Learning in Sustainable Transportation Systems: Models and Applications. Sustainability, 13(13). DOI: https://doi.org/10.3390/su13031551.

Vannieuwenhuyze, A. (2019). Intelligence Artificielle Vulgarisée Le Machine Learning Et Le Deep Learning Par La Pratique (Eni Ed.).

Vignieri, V. (2016). Framing The Sources of Image of A Local Area Through Outcome-Based Dynamic Performance Management Ijpa Symposium, Palermo.

Vitale, I., Dondo, R. G., Gonzalez, M., and Coccola, M. E. (2022). Modelling and Optimization of Material Flows in The Wood Pellet Supply Chain. *Applied Energy*, 313.

Wang, F., and Lin, L. (2021). Spare Parts Supply Chain Network Modeling Based on A Novel Scale-Free Network and Replenishment Path Optimization with Q Learning. *Computers & Industrial Engineering*, 157.

Xu, Y., and Goodacre, R. (2018). On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating The Generalization Performance of Supervised Learning. *Journal of Analysis and Testing*, 249–262.

Yavari, M., and Geraeli, M. (2019). Heuristic Method for Robust Optimization Model for Green Closed-Loop Supply Chain Network Design of Perishable Goods. Journal of Cleaner Production, 226(226). DOI: https://doi.org/10.1016/j.jclepro.2019.03.279.

Yue, X., and Chen, Y. (2018). Strategy Optimization of Supply Chain Enterprises Based on Fuzzy Decision-Making Model in Internet of Things. *Special Section on Deep Learning for Internet of Things*, 6.

Zhao, W., and Wang, D. (2018). Simulation-Based Optimization on Control Strategies of Three-Echelon Inventory in Hybrid Supply Chain with Order Uncertainty. 6.

Zheng, X., Yin, M., and Zhang, Y. (2019). Integrated Optimization of Location, Inventory and Routing in Supply Chain Network Design. *Transportation Research Part B*, 121.