



What are the root causes of material delivery schedule inaccuracy in supply chains?

Downloaded from: <https://research.chalmers.se>, 2025-12-04 23:22 UTC

Citation for the original published paper (version of record):

Jonsson, P., Öhlin, J., Shurrab, H. et al (2023). What are the root causes of material delivery schedule inaccuracy in supply chains?. International Journal of Operations and Production Management, 44(13): 34-68. <http://dx.doi.org/10.1108/IJOPM-12-2022-0806>

N.B. When citing this work, cite the original published paper.

What are the root causes of material delivery schedule inaccuracy in supply chains?

Patrik Jonsson

*Department of Technology Management and Economics,
Chalmers University of Technology, Gothenburg, Sweden*

Johan Öhlin

*Department of Manufacturing Planning and Supply Chain, Volvo Car Group,
Gothenburg, Sweden*

Hafez Shurrab

*Sam and Irene Black School of Business, Penn State Erie The Behrend College,
Erie, Pennsylvania, USA*

Johan Bystedt

Meridion AB, Gothenburg, Sweden

Azam Sheikh Muhammad

*Department of Computer Science and Engineering,
Chalmers University of Technology, Gothenburg, Sweden, and*

Vilhelm Verendel

Department of Physics, Chalmers University of Technology, Gothenburg, Sweden

Abstract

Purpose – This study aims to explore and empirically test variables influencing material delivery schedule inaccuracies?

Design/methodology/approach – A mixed-method case approach is applied. Explanatory variables are identified from the literature and explored in a qualitative analysis at an automotive original equipment manufacturer. Using logistic regression and random forest classification models, quantitative data (historical schedule transactions and internal data) enables the testing of the predictive difference of variables under various planning horizons and inaccuracy levels.

Findings – The effects on delivery schedule inaccuracies are contingent on a decoupling point, and a variable may have a combined amplifying (complexity generating) and stabilizing (complexity absorbing) moderating effect. Product complexity variables are significant regardless of the time horizon, and the item's order life cycle is a significant variable with predictive differences that vary. Decoupling management is identified as a mechanism for generating complexity absorption capabilities contributing to delivery schedule accuracy.

Practical implications – The findings provide guidelines for exploring and finding patterns in specific variables to improve material delivery schedule inaccuracies and input into predictive forecasting models.



Originality/value – The findings contribute to explaining material delivery schedule variations, identifying potential root causes and moderators, empirically testing and validating effects and conceptualizing features that cause and moderate inaccuracies in relation to decoupling management and complexity theory literature?

Keywords Supply chain complexity, Information sharing, Material delivery schedule, Mixed method, Complexity absorption capability, Decoupling management, Automotive industry

Paper type Research paper

1. Introduction

Communication of material delivery schedules through electronic data interchange is a widely established supply chain practice in the automotive industry, which is the empirical focus of this study. However, the value of sharing schedule information depends on its stability and accuracy (Holweg, 2005; Jonsson and Myrelid, 2016). According to a recommendation by the German automotive association VDA (2008), anything less than 90% schedule accuracy on the item level is considered poor performance for weekly planning buckets on a 3- to 8-week horizon – a target being far from reached in practice (Shurrab and Jonsson, 2023) [1]. Delivery schedule inaccuracies require corresponding supplier responsiveness and force suppliers to decouple from vehicle manufacturers' demand signals (Holweg, 2005), apply reactive rescheduling and reworking activities and proactively build capacity flexibility and safety stocks. These measures lead to increased costs without necessarily solving delivery service problems (e.g. Myrelid, 2017).

Research has examined suppliers' strategies to mitigate schedule inaccuracies through, for instance, excess capacity and safety stocks (e.g. Atadeniz and Sridharan, 2019; Krajewski *et al.*, 2005; Li and Disney, 2017). However, limited empirical evidence exists on the root causes generated downstream at original equipment manufacturers (OEMs) and their effects on delivery schedule inaccuracy. Most research comprises analytical modelling using various operating variables (parameters) and operating conditions (context) (Pujawan *et al.*, 2014); we refer to such variables and conditions as "features" in this study. Empirical studies that explore or test such features are limited, with some exceptions (e.g. Pujawan *et al.*, 2014; Pujawan and Smart, 2012; Shurrab and Jonsson, 2023). Consequently, we have limited knowledge about how possible "features" affect the processes in which schedules are generated and communicated. Thus, the literature does not give managers clear guidelines about explaining or managing schedule inaccuracies.

To address this knowledge gap, we conducted a mixed qualitative–quantitative empirical study to explore how product and process variables constitute features statistically related to delivery schedule inaccuracy. Drawing from decoupling management (Wikner, 2014) and complexity (Shurrab and Jonsson, 2023) literature, we analyse the predictive difference of features on various time horizons, conducted within a case study representing specific settings of a global car manufacturer (called the OEM).

Section 2 presents a literature review defining delivery schedule inaccuracy, exploring the perspectives of decoupling management and complexity literature on the problem and identifying relevant variables that explain associated inaccuracies. We then describe our mixed-method case study design in Section 3. Finally, the analyses' findings are described and discussed in Section 4, and the conclusions in Section 5.

2. Definition of delivery schedule inaccuracy and independent variables

The literature review defines delivery schedule inaccuracy (dependent variable), presents the decoupling management and complexity theory perspectives on delivery schedule inaccuracy and reviews the literature on independent variables (features).

2.1 Delivery schedule and schedule inaccuracy

A delivery schedule is defined as “the required or agreed time or rate of delivery of goods and services purchased for a future period” (Blackstone, 2013). Delivery schedules contain planned orders and call-off information on various planning horizons. The order information can be expressed in different planning buckets, normally varying among days, weeks and months.

Schedule inaccuracy in a material ordering process refers to the difference between a planned and actual demand for items (Lee and Adam, 1986). The absolute percentage error, or mean absolute percentage error (MAPE), is a common way of measuring forecast inaccuracy (Makridakis *et al.*, 1998), thus a relevant performance measure of schedule inaccuracy.

2.2 Theoretical perspectives on factors causing delivery schedule inaccuracies

This study addresses schedule inaccuracies in material delivery using two theoretical perspectives to understand the various underlying elements contributing: decoupling management and complexity theory. Both decoupling management and complexity theory literature provide essential theoretical lenses through which we can understand, anticipate and perhaps even mitigate the delivery schedule inaccuracies in material ordering processes.

2.2.1 Decoupling management perspective. The decoupling management perspective presents a comprehensive view of how planning horizons, time fences and distinct decoupling points can influence the accuracy of delivery schedules. From this standpoint, schedule inaccuracies may be tied to decoupling points and their respective hybrid zones (Wikner, 2014; Banerjee *et al.*, 2012).

The customer order decoupling point (CODP) denotes a critical juncture between speculation-based flow initiation and commitment to actual customer orders. The decoupling point, therefore, separates two fundamentally different operational paradigms: speculation on future orders and commitment to existing orders. Near-term plans are frozen and primarily customer order-dependent, while long-term plans are less fixed and rely more heavily on market forecasts. The hybrid zone before the CODP (Wikner and Rudberg, 2005) represents a transition phase, characterized by a blend of customer orders, forecasts and modifications of customer orders. In this context, feature effects (i.e. the influence of specific item characteristics on schedule inaccuracy) are expected to be relatively more pronounced when the commitment to customer orders is higher.

Similarly, the customer adaptation decoupling point (CADP) introduces another level of complexity by addressing the standardization–adaptation dichotomy in fulfilling customer orders. This point separates decisions based on standardized procedures from those necessitating unique adaptations to meet specific customer needs (Amaro *et al.*, 1999). In the hybrid zone before the CADP, there is a mixture of standardization and adaptation. For delivery schedules of planned purchase orders, the CADP and its hybrid zone relate particularly to customized and variant items. In such a scenario, feature effects might differ significantly depending on whether schedules represent customized or variant items.

2.2.2 Complexity theory perspective. According to Turner and Baker (2019), complexity science (or, synonymously, complexity theory) is separate from general systems theory – “the behavior of a system in terms of that of its constituent components and the interrelationships between those components” (Koopmans, 2017, p. 21) – and embraces other concepts, such as complex adaptive systems and chaos theory. Evolving from the general and open systems models, complexity in social sciences through, for instance, system dynamics, complex adaptive systems and deterministic chaos theory contributed to establishing modern organization theory (Roehrich and Lewis, 2014).

Complexity theory has been used in various operations and supply chain management domains to explore a system’s non-linearity, emergence, adaptive behaviour, interconnectedness and interdependence and self-organization (e.g. Aitken *et al.*, 2016;

Fernández Campos *et al.*, 2019; Manuj and Sahin, 2011; Maylor and Turner, 2017; Shurrab *et al.*, 2022a, b). It can provide valuable insights into how systemic interdependencies and non-linear dynamics can exacerbate delivery schedule inaccuracies (Shurrab and Jonsson, 2023). Focusing on how the inherent unpredictability and emergent behaviour in complex systems, such as material delivery processes, can lead to a high degree of schedule inaccuracy arguably complements the perspectives provided by decoupling management.

Viewing the causes of material delivery schedule instability as per complexity theory necessitates an in-depth exploration of the intricate process dynamics at play. The instability of delivery schedules can be construed as a function of the static elements or structures in a process, the heterogeneity of these elements and the dynamism of their interactions (Holweg *et al.*, 2018). Accordingly, increased external variety, such as fluctuations in the market or changes in customer preferences, can be addressed or elucidated by a corresponding increase in internal variety. This emergent complexity could manifest as broader product ranges to cater to more diverse customer needs (Serdarasan, 2013).

Underlying delivery schedule variations can be viewed as complexities inherent in the internal process itself. Two crucial aspects come to the fore: process variation and design. Process variation refers to the fluctuations in quality, quantity and timing within the process (Holweg *et al.*, 2018). This variation could result from various factors, including machine downtime, labour inefficiencies, supplier delays or demand fluctuations (Mendoza *et al.*, 2014). In contrast, process design encompasses elements such as buffering strategies, system throughput and process scale and scope (Holweg *et al.*, 2018). Buffering strategies aim to manage variations but can lead to schedule instability if not sufficiently designed or implemented (Atadeniz and Sridharan, 2019). System throughput can be constrained by bottlenecks that limit the rate at which the system can deliver products. The scale of the process, in terms of the number of demands (numerousness), and the scope, in terms of demand variety, can add layers of complexity that affect schedule stability (Shurrab and Jonsson, 2023).

Apart from process dynamics, complexity transfers across organizational boundaries (Huatuco *et al.*, 2021), where intricacies in one area can create ripples of complexities in others, leading to delivery schedule variations. According to Shurrab and Jonsson (2023), identifying the root causes of delivery schedule variations involves untangling the network of their underlying effects.

Firstly, interactions that induce delivery schedule instability can manifest as the importation of demand and supply chain complexities. These might arise from external factors, such as the need to respond to evolving market demand or the complexity associated with managing multiple suppliers. Both these factors can lead to frequent changes in gross requirements, thus contributing to schedule inaccuracies.

Secondly, the exportation of product and process complexities upstream and downstream in the supply chain can also generate inaccuracies in delivery schedules. An example is the decision to offer customizations in products, which can considerably increase product complexity (Chand *et al.*, 2022). As more customization options are made available, the process of estimating accurate delivery schedules can become more challenging due to the more significant variability in production times, resource needs and coordination efforts (Shurrab *et al.*, 2022a).

Finally, the generation and absorption of complexity within the internal environment of a supply chain member can lead to delivery schedule inaccuracies. If a firm's internal systems have a limited capability to handle variations, the result could be an inability to maintain stability in delivery schedules (Shurrab and Jonsson, 2023). This limitation could be due to a lack of robust processes to handle the dynamism stemming from product complexity, capacity constraints or other factors.

2.3 Causes of delivery schedule inaccuracy

The terms schedule instability and nervousness have been used to define and understand variations in delivery schedules, leading to schedule inaccuracies. [Steele \(1975\)](#) uncovered the phenomenon of schedule instability and related effects on performance. Over several decades, researchers have suggested scheduling configurational variables or parameters that reduce or dampen schedule instability (e.g. [Atadeniz and Sridharan, 2019](#); [Herrera et al., 2016](#); [Ho, 1989](#); [Lalami et al., 2017](#); [Lee and Adam, 1986](#); [Li and Disney, 2017](#); [Zhao et al., 1995](#); [Zhao and Lee, 1996](#)).

Primary scheduling parameter variables include planning horizon (e.g. [Lee and Adam, 1986](#)), time fence management/frozen period (e.g. [Zhao et al., 1995](#)) and re-planning periodicity or frequency (e.g. [Ho, 1989](#)). These parameters can moderate generated instability in delivery schedules ([Shurrab and Jonsson, 2023](#)).

Schedule instability is also contingent on contextual planning environment variables ([Harrison, 1997](#); [Inman and Gonsalvez, 1997](#); [Jonsson and Mattsson, 2003](#)). Previous studies have considered several contextual variables potentially impacting delivery schedule inaccuracy: external (demand- and supply-related) and internal (product- and manufacturing-related) variables.

External demand-related variables include customer input (e.g. [Pujawan, 2008](#)), demand pattern (e.g. [Zhao et al., 1995](#)), sales forecasting errors (e.g. [Lee and Adam, 1986](#)) and stock-out cost (e.g. [Ho and Carter, 1996](#)). Demand patterns – featuring frequent low actual demands and increased product life cycle uncertainty – and forecasting errors – leading to late increases in demand – can cause schedule instability. In contrast, late changes in end-item specifications can moderate such generated instability [Shurrab and Jonsson \(2023\)](#).

External supply-related variables include supply relationships and supplier profiles (e.g. [Krajewski et al., 2005](#)). A lack of readiness for supply chain disruptions and transportation dynamics characterizing suppliers or relationships with suppliers can cause delivery schedule instability. Alignment among engineer-to-order's (ETO's) and suppliers' operations times, pick-up frequency, lot sizing rules, unit packaging (unit load) or any other delivery flexibility forms can moderate schedule instability [Shurrab and Jonsson \(2023\)](#).

Internal product-related variables (e.g. product structure and item commonality) and manufacturing-related variables (e.g. [Ho, 1989](#)) can also generate delivery schedule instability (e.g. [Lee and Adam, 1986](#)) through causal and moderating effects ([Shurrab and Jonsson, 2023](#)). Manufacturing-related causal variables could include, for example, production disruptions and inventory stock discrepancies. Related moderators include capacity scalability, transportation optimization, schedule miscommunication, safety stock policies and policies penalizing suppliers for delivery quantity and timing deviations.

3. Methodology

3.1 Study design

The data analysis is based on qualitative and quantitative data collected at a European personal car manufacturer (here, called the OEM). A single case was considered appropriate due to the massive amount of data that needed collection. We wanted to understand the case company's specific operating variables and conditions, identify relevant explanatory/predictive variables, collect variable data and analyse and interpret the findings in detail. Therefore, a qualitative explorative study preceded a quantitative data analysis. The qualitative study identified possible causes and specified variables with potential causal effects on delivery schedule accuracy. We then collected quantified data corresponding to some of these variables and built a database of variables to further analyse the predictive difference of these variables on delivery schedule inaccuracies. As such, we apply a mixed method design to use the findings of the first study to inform the second study and expand the insights generated in a developmental manner (e.g. [Davis et al., 2011](#)). The following eight data collection and

analysis phases are conducted: (1) mapping the material planning process, (2) collecting/extracting two years of delivery schedule data from the OEM's enterprise resource planning system and forming delivery schedule groups, (3) measuring delivery schedule accuracy on different horizons, (4) exploring features and proposing how OEM-specific features explain schedule inaccuracies, (5) defining quantitative feature measures and collecting quantitative data of identified features, (6) exploring relationships between features and inaccuracy (plotting and correlations) and interpreting the size of predictive differences, (7) analysing feature reduction, logistic regression and random forest and (8) interpreting data model findings.

The qualitative and quantitative studies were jointly conducted by a team of four researchers and two practitioners. Two researchers focused on data analytics, and the other two framed and led the data collection and analysis. One of the practitioners represented the case company, focusing on material planning and delivery schedules. The other practitioner is a specialist in delivery scheduling and information sharing in the automotive industry (i.e. the case company's industry). Both practitioners actively participated from the beginning to the end of the study, including problem formulation, research design and research analysis.

3.2 Delivery schedule process and delivery schedule data collection (Phases 1–2)

The OEM has 10 assembly factories globally. Most purchased components are delivered in sequence. The material requirements planning (MRP) and delivery schedules are generated on daily planning buckets with a 60-week horizon and are updated daily. Call-off volumes required for the upcoming two days are frozen, and a safety lead time of two days is used for inbound material from suppliers. The last day of (customer) order confirmation (LDOC) of a car is 3–4 weeks before the delivery date. Four weeks of the order book is frozen every Thursday. In other words, on Wednesday, the production schedule contains a three-week frozen period; on Thursday, the fourth week on the horizon becomes frozen. Accordingly, the end product demand is perfectly known during 3–4 weeks. The assembly sequence is fixed for the next two weeks, meaning that the company knows exactly what car variants to assemble each day during this period. Consequently, we presume the gross requirement of purchased items to be close to fixed during 2–4 weeks.

Schedule accuracy (the dependent variable) was measured based on historical delivery schedule data. Therefore, the first quantitative data collection phase was to collect historical delivery schedule data from the OEM and build a two-year delivery schedule database (2017–2019). In the empirical analysis, we aggregated the data into weekly planning buckets. The database of delivery schedules contained 16.5 million schedule records (rows). The records were grouped, and each group included records with identical combinations of (1) item number, (2) ship-to-gate address (unique delivery address) and (3) demand (delivery) week. Consequently, each group included all planned/required order volumes of the unique items shared by the OEM with the individual suppliers over two years, specifying the unique delivery addresses and weeks. The grouping resulted in 0.63 million unique schedule groups. These groups were called “delivery schedule groups”.

3.3 Exploring possible root cause effects at the OEM (Phase 4)

Based on the literature review on delivery schedule inaccuracy causes – in combination with workshops and dialogues within the OEM – we identified features with expected explanatory effects on schedule inaccuracy. The OEM representative (i.e. the co-author mentioned earlier) led this exploratory phase. The identified causes were discussed in a larger research project involving two more automotive OEMs. Therefore, although the variables were generated from one specific OEM, they were validated as generally relevant possible causal variables in other automotive settings.

In total, six interviews were conducted with the senior logistics planning manager responsible for production planning and material ordering. Archival data, in the form of detailed descriptions and maps of production and material ordering process steps, subprocesses and systems, were also collected and analysed. Figure 1 summarizes the identified categories of features and individual feature variables at the OEM.

The LDOC is an important decoupling point. Inside the LDOC, the order book is fixed. Its data consist of customer orders (used as actual demand data within this period) and a mix of customer orders and sales forecasts (used outside this point). Therefore, we can distinguish between variables affecting the gross requirement outside (1 in Figure 1) and after (2 in Figure 1) the LDOC. We also distinguish between variables as follows: *variables affecting the net requirement before and after the LDOC* (3 in Figure 1), *variables not acting as possible root causes of schedule variations but that can potentially amplify (moderate) them* (4 in Figure 1) and *variables affecting the variations by interrupting the levelled schedule* (not in Figure 1).

The assembly sequence is fixed for two weeks, meaning that, combined with a 100% accurate market demand (fixed-order book), the gross requirement within the two-week horizon should be close to 100% accurate. A prerequisite for such accuracy is that the production capacity is unchanged within this two-week horizon, the scrap levels do not exceed set parameters and the daily production volumes do not deviate from planned volumes.

If the production output deviates from the plan, the OEM's strategy is not to add extra capacity that day to ensure that the actual daily production output does not deviate from the planned. Instead, the required compensatory capacity is scheduled for a later date. As the assembly sequence is frozen for two weeks, if the production volume plan is not met on the first day, the remaining unmet volume of cars planned for assembly will be assembled at the beginning of the next day. Production deviations entail changes in daily gross requirements at the item level if another mix of car variants is built on the second day compared to previously planned. However, capacity solutions such as overtime may be added later that week, leading to changes in gross requirements since more cars will be assembled per day than previously planned.

Other variables affecting the gross requirement on short horizons are phase-in or phase-out decisions, as the exact date may change on short notice. The deviation between the planned and fixed order sequencing on the assembly line (during the 3–4-week frozen period)

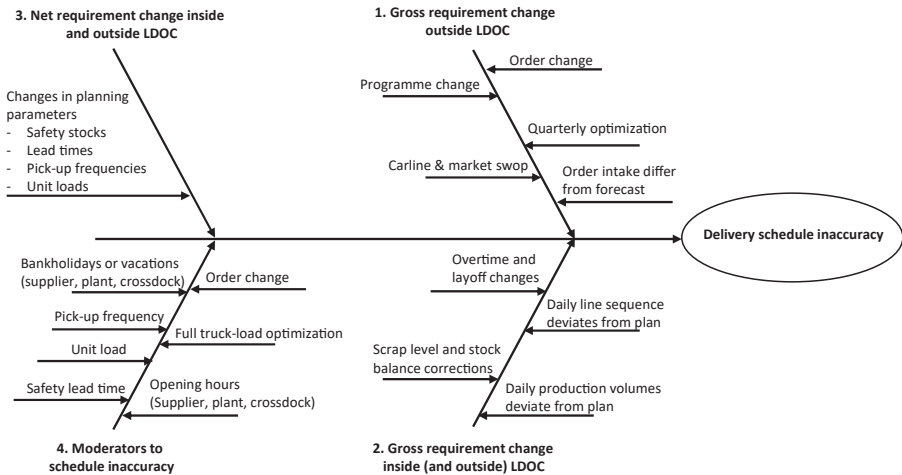


Figure 1.
Proposed feature
categories and feature
variables explaining
schedule variation

also changes the short-term gross requirement, correction of scrap levels and inventory balance records.

Outside the LDOC, market forecast accuracy (i.e. if order intake deviates from the forecast) is expected to explain a significant proportion of the gross requirement changes. Market forecast accuracy is also affected by overall production programme decisions at the car model level, such as shifting volumes between models or markets. Sometimes, the production programme requires modifications towards optimized levels in line with sales initiatives to meet quarterly sales targets in the sales organization. As for changes in the net requirement (inside and outside the LDOC), variables leading to changes in material ordering parameters (e.g. safety stock, order quantities, unit loads and pick-up frequencies) come into play.

Variables with presumed moderation influence on delivery schedule variations include suppliers' local bank holidays, vacations and opening hours, unit loads, full truckload optimizations, transport lead time and pickup frequency. The transport lead time is from pickup at the supplier until the goods are received at the OEM. The inaccuracy is expected to be higher the more extended the transport lead time since material ordering becomes more dependent on forecasts. Larger unit loads and more infrequent pickups may increase the lumpiness of schedules, thus increasing inaccuracies. However, they may, at the same time, stabilize the schedules on the short horizon, as minor daily variations may be evened out when using larger unit loads and weekly instead of daily deliveries. Opening hours at suppliers, cross-docks and plant goods receptions may impact daily schedules but are not expected to affect weekly schedules. Local plant shutdowns associated with bank holidays and vacations are other potential variables that may lengthen delivery lead times and cause multiplication effects.

Take rate is a variable not mentioned in [Figure 1](#), but it has a presumed effect. It refers to the proportion of assembled vehicles that contain an item. The highest take rate is achieved for items included in all vehicle models and all model variants. The lowest take rate is for variant items included in one or a few models. A low take rate is expected to affect gross requirement uncertainty due to increased market forecast inaccuracy.

3.4 Dependent and independent variable definitions

3.4.1 Dependent variable definition and measurement (Phase 3). The empirical analysis explores and empirically tests (1) which of the independent variables (features) influence delivery schedule inaccuracies and (2) develops knowledge about how to assess the predictive difference of these features given different time lags (horizons). For (1), we performed the feature analysis and reduction (association/correlation analyses), as presented in [Section 3.5](#), in which we defined two sets of numerical performance variables of delivery schedule inaccuracy, each measured on 2- and 8-week horizons. The first measures symmetric MAPE (sMAPE) to overcome the asymmetry issue of MAPE, which favours under-forecasted items ([Kim and Kim, 2016](#)). The second measure is the "log accuracy ratio" measure, which is suggested for fitting features to prediction models ([Tofallis, 2015](#)). In the setting of conducting variable reduction by assessing their impact on delivery schedule inaccuracies, we use the strength of the association/correlation of the performance variables to all other variables in order to identify the relevant variables that explain associated inaccuracies.

Similarly, for (2), with the objective of estimating the predictive difference of variables within the data using logistic regression model analysis, we make use of binary dependent variables associated with various percentage levels of delivery schedule inaccuracies and perform explanatory analysis using logistic regression model fitting to achieve our objective. We split the delivery schedule data for a given horizon based on the threshold defined on the percentage of deviation of the absolute error with respect to a reference volume. For each schedule, the absolute percentage difference between the scheduled volume and the reference

volume (i.e. the actual final order volume) is calculated. When the reference volumes are zero, the percentage error is undefined. If the reference volume is zero and the scheduled volume is greater than zero, we set those errors as 100%.

3.4.2 Independent variable definition and measurement (Phases 5–6). Identifying independent variables started with qualitative mapping and analysing the material delivery scheduling process at the OEM, as described in [Section 3.3](#). We used data from the delivery schedule database to quantitatively measure the independent variables. We also collected additional quantitative data from other internal databases. Variables with expected effects on the gross requirement outside the LDOC were omitted due to not having access to market demand data (e.g. sales forecasts).

We collected quantified data from the OEM representing the following variables: unit loads, pick-up frequency, transport lead time, planned and actual production volumes per day, take rate, phase-in/out dates, vacation weeks and car model. In addition, we developed a variable named “life cycle phase” that measures whether an item is early or late in its life cycle. The expected item life cycle effect is mainly due to the lack of demand history early in the life cycle, which may negatively affect forecast accuracy. Late in the life cycle, when items are planned to be phased out, parameters such as safety stocks and unit loads may be changed, affecting schedule fluctuations inside and outside the LDOC.

Altogether, we defined 23 measurable variables as independent variables ([Table 1](#)). These variables represent categories 2–4 of [Figure 1](#). Seventeen were categorical and measured on an ordinal scale, while six were numerical on a ratio scale. For the numerical variables, the pick-up frequency is at least weekly for all deliveries, while 64% have daily pickup frequencies. Most suppliers are located in Europe, with transport lead times no longer than a few days, in addition to longer inbound flows from distant suppliers (e.g. in Asia). Twelve per cent of the transport requires lead times longer than a week, while the rest occur within a week. The component commonality is relatively low, and/or the variant breadth is relatively large. Exactly 57% of purchased items have a take rate lower than 5%, and 86% have take rates lower than 20%.

Thirteen of all selected variables progressed to the regression analysis as a result of correlation tests. [Table 1](#) shows the predictive difference of these variables on material delivery schedule variations according to terminology in the related literature (e.g. [Shurrab and Jonsson, 2023](#)), including four types of relationships: independent causes and moderators and contingent causes and moderators.

Independent causes are variables that could cause variations independently of any preconditions, and independent moderators are variables that could moderate the effect of variables on variations without requiring a specific cause source. On the other hand, contingent causes are variables with particular prerequisites for variations, and contingent moderators are cause-dependent variables that require specific conditions to moderate the effect of possible causes on variations.

[Table 1](#) also presents what the variables eligible for regression analysis represent as sources of complexity, including phenomena representing issues either in the process design (scale and scope dynamics and insufficient buffering of time) or the process context (variations in quality and timing). Accordingly, the selected variables represent these process phenomena that eventually generate schedule variations.

3.5 Independent variable reduction and data analysis

3.5.1 Reduction of independent variables using correlation analyses (Phases 6–7). Three sets of correlation analyses (categorical–categorical, numerical–numerical and numerical–categorical variables) were conducted to identify highly correlated variables and reduce the 23 variables for the logistic regression analysis.

Variables	Description	Range of values	In regression analysis	Proposed effect on schedule variations	Sources of complexity
Categorical	Item	4,977 items			
	Item group	Nominal categories			
	Supplier	Nominal categories			
	Material receiving address	14 addresses			
	Scheduled demand date (3 variables)	76 weeks: between 2017 (week 31) and 2018 (week 52) Two years: 2017 and 2018			
	Year	52 weeks			
	Week number	13 categories: 0-9, 11, 13, 41, 45, 46	X	Contingent cause	Insufficient buffering of time, process design
	Transport lead time in days	The standard time required for delivering items from the supplier to the receiving address			Scale dynamics, process design
	Pick-up frequency in times per week	The number of days per week when picking up material from the supplier is possible	X	Independent moderator	
	Phase-in/-out week	Does the schedule belong to the fixed phase-in/out weeks (two weeks per year)?			
	Shifts in scheduled volumes from or to zero	Does the scheduled volume change from any value to zero or vice versa?			
	Holiday week	Weeks per year when the plants are closed			
	Car Model A	Do the items exist in the model's BOM?	X	Independent cause	Scope dynamics, process design
	Car Model B		X	Independent cause	Scope dynamics, process design
	Car Model C		X	Independent cause	Scope dynamics, process design
	Car Model D		X	Independent cause	Scope dynamics, process design
	Car Model E		X	Independent cause	Scope dynamics, process design
	Car Model F		X	Independent cause	Scope dynamics, process design

(continued)

Material delivery schedule inaccuracy

Table 1.
Feature variables considered in the correlation tests

Table 1.

Variables	Description	Range of values	In regression analysis	Proposed effect on schedule variations	Sources of complexity
Numerical	Forecasted volume	Forecasted volume X weeks before the delivery date, $X = \{2, 4, 8, 12\}$	X	Contingent cause	Variations in timing, process context
	Unit load per week	Unit load divided by the average weekly demand or reference volume	X	Contingent moderator	Scale dynamics, process design
	Aggregated production deviation	Number of cars behind or ahead of actual demand	X	Contingent cause	Variations in quality, process context
	Item's order life cycle ratio	The number of weeks from the first delivery date in the dataset to present over the number of weeks from the first delivery date to the last planned date on the horizon	X	Contingent cause	Scope dynamics, process design
	Take rate at the plant level	Percentage of manufactured cars at a plant in which the item is used	X	Contingent cause	Scope dynamics, process design

The first correlation analysis was conducted between all possible pairs of categorical variables using the uncertainty factor (Theil's U), a widely used method for finding pairs of highly correlated categorical variables (White *et al.*, 2004). The analysis identified the variables "item" and "schedule delivery date", measured as demand year and demand week, to have a full association (coefficient = 1), with at least one other variable. Therefore, these two variables were excluded from the subsequent analysis. Table 2 shows the correlations between the remaining 15 categorical variables. Furthermore, transport lead time and pick-up frequency are fully associated with the supplier variable (i.e. a specific supplier has a fixed transport frequency and lead time). The analysis also shows full, or close to full, associations between the phase-in/out and holiday week variables with the schedule delivery week variable. Because of these correlations, the supplier, phase-in/out week and holiday week variables could be excluded in the regression analysis if the supplier and schedule delivery week variables are included.

Pearson correlation (Table 3) was used for all numerically scaled variables (six feature variables, four performance variables and the reference volume variable). As expected, the forecasted volumes and the reference volume are highly correlated. Thus, the subsequent analysis could include only one of the two measures. Similarly, the item's order life cycle and production deviation are highly correlated.

The third correlation analysis (Table 4) was conducted between the remaining categorical and numerical variables by calculating the correlation ratio (Kenney and Keeping, 1951). Some relatively high correlations were identified, but none were close to zero. As a result of the three correlation tests, 13 variables proceeded to the regression analysis step.

3.5.2 Statistical analysis using logistic regression analysis (Phase 7). The most straightforward approach to model delivery schedule inaccuracy is to view it as a binary classification problem, whereby a schedule is either inaccurate or not. A natural first model to explain feature influence is a logistic regression with transparency and the possibility of interpreting the predictive differences of varying model coefficients. Other possible options such as count regression modelling or multi-class classification problems would have needed potential higher-order interactions to account for differences over different time horizons or have required restrictions and/or imputation of the data. An exploration of the target variable

Categorical variables 1–15	Categorical values 2–15														
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1. Supplier	0.58	0.75			1.00	1.00			0.49	0.45	0.45	0.39	0.45	0.43	
2. Receiving address		0.19				0.13			0.14	0.12	0.12	0.11		0.10	
3. Item group					0.25	0.22			0.15	0.10	0.16			0.11	
4. Schedule demand year															
5. Schedule demand week							1.00	0.86							
6. Transport lead time						0.21									
7. Pickup frequency															
8. Phase-in/out-week															
9. Holiday week															
10. Car Model A										0.62	0.10				
11. Car Model B												0.10			
12. Car Model C												0.44			
13. Car Model D															
14. Car Model E															
15. Car Model F															

Note(s): Correlation values < 0.10 are excluded

Table 2.
Correlation of
categorical–categorical
variables using the
uncertainty factor

Table 3.
Correlation of
numerical-numerical
variables using
Pearson correlation

Numerical variables 1–11	Numerical variables 2–11									
	2	3	4	5	6	7	8	9	10	11
1. Unit load per week										
2. Production deviation		–0.71	0.10							
3. Order life cycle										
4. Take rate				0.52	0.54	0.53				
5. Forecasted volume (8 weeks)					0.94	0.94				
6. Forecasted volume (2 weeks)						0.97				
7. Reference volume										
8. sMAPE (8 weeks)								–0.19	0.28	0.13
9. Log accuracy ratio (8 weeks)									0.18	0.39
10. sMAPE (2 weeks)										0.15
11. Log accuracy ratio (2 weeks)										
Note(s): Correlation values < 0.10 are excluded										

Table 4.
Correlation categorical-
numerical variables
using correlation ratio

Categorical variables 1–15	Numerical variables 1–11										
	1	2	3	4	5	6	7	8	9	10	11
1. Supplier	0.62	0.21	0.19	0.83	0.57	0.58	0.57	0.24	0.18	0.21	
2. Receiving address	0.22	0.14	0.15	0.52	0.27	0.28	0.28	0.17		0.13	
3. Item group	0.26			0.47	0.23	0.25	0.24				
4. Schedule demand year		0.71	0.74								
5. Schedule demand week		0.43	0.34					0.14	0.11	0.16	0.12
6. Transport lead time	0.12			0.24	0.15	0.13	0.13		0.11		
7. Pick-up frequency				0.35	0.24	0.21	0.21				
8. Phase-in/out-week											
9. Holiday week											
10. Car Model A				0.37	0.26	0.27	0.26				
11. Car Model B				0.48	0.30	0.31	0.31				
12. Car Model C				0.44	0.24	0.25	0.25				
13. Car Model D				0.52	0.27	0.28	0.28				
14. Car Model E				0.47	0.23	0.23	0.23				
15. Car Model F				0.51	0.25	0.26	0.26				
Note(s): Correlation values < 0.10 are excluded											

related to empirical logits suggested a reasonable fit for logistic regression. Taken together, this speaks in favour of logistic regression.

The 13 variables included: (1) the item life cycle measured on a scale from 1 to 2, where 1 indicates the first delivery date in the dataset and 2 indicates the last date, (2) take rate at the plant level measured as a percentage expressing the proportion of finally produced unique cars the item represents, (3) forecasted volume is for the actual planning horizon, (4) unit load per weekly demand is the standard unit load divided by the average weekly demand, (5) production deviation is the actual weekly production volume divided by the planned volumes, (6) pick-up frequency, (7) transport lead time and (8)–(13) Car Models A–F are on ordinal scales. See Table 4. Variables 4, 5 and 6 have expected multiplicative influence on variations. The other variables are expected to have direct effects.

To test the sensitivity of the features on different levels of schedule inaccuracies, we developed five regression models with different inaccuracies (thresholds/levels for what is considered an error: 5%, 10%, 30%, 50% and 100%) for each of the four planning horizons (2, 4, 8 and 12 weeks). Before including the features in the model, we tested the linear

relationship between the independent and dependent variables. Consequently, we found that the relationship between the unit load per week variable and the dependent variable did not show acceptable levels of linearity for the 50% and 100% models. We also identified weak levels of linearity for several forecasted volume variables. All other variables showed acceptable linearity patterns. The unit load per week variable was transformed into $1/(1 + \text{unit load per week})$, an approximate relationship that was judged from the data exploration. This transformation clearly improved linearity in the empirical logits of the target variable. Similarly, the forecasted volume variables were transformed into $\log(\text{forecasted volume} + 1)$ to reduce the impact of some outliers with very large volume forecasts. Because of the large volume of data, we relied on the integrated nested Laplace approximation (INLA) methodology for fast approximate Bayesian inference ([Gomez-Rubio, 2020](#)). Logit plots in [Appendix 1](#) show acceptable linearity for most of the transformed variables described above.

We fit $20(5 \times 4)$ logistic regression models with a binary MAPE as the dependent delivery schedule variable (one group representing cases with MAPE values above a defined threshold value and one group with values below) and the 11 covariate variables as independent variables. The target binary variable was defined as a schedule error above the corresponding threshold. These models were used to analyse the predictive difference of independent variables on delivery schedule accuracy and to test the size of the predictive difference on the four different horizons and for the five different schedule inaccuracy levels. We also included calendar data (demand year and week, demand week, demand year and holiday weeks) as control variables for systematic external variations. Consequently, a logistic regression model for the horizon X for the log odds of binary outcome Y (error on a 5%, 10%, 30%, 50% or 100% threshold level) was specified as follows:

Forecasted volume (X) + demand year + demand week + receiving address + order item life cycle + take rate + unit load + pickup frequency + transport lead time + production deviation + Car Model A + Car Model B + Car Model C + Car Model D + Car Model E + Car Model F, where forecasted volume (x) is the quantity for a given order on a horizon of x weeks in advance.

To complement the logistic regression model findings, we developed random forest models ([Hastie et al., 2009](#)) for the same problem. Fitting random forest models allows for exploring what features are most useful for explaining the classification in the data with decision trees. One advantage of random forest and decision tree models over logistic regression is the ability to pick up non-linear patterns in the data. Therefore, we can expect interactions with other variables to be useful in fitting non-linear models. For each random forest model (on a given schedule horizon and threshold value), we evaluated what features are most important with regard to whether schedule inaccuracies are correctly classified found in a held-out evaluation test set which is a random sub-selection of the data. We varied the random forest model hyper-parameters in terms of number of trees and other hyper-parameters. The results of the logistic regression models were compared with the most useful features in the random forest models. Further details are found below and in [Appendix 2](#).

As described earlier, the LDOC is 3–4 weeks, and the assembly sequence is frozen for two weeks. Consequently, the CODP and CADP could be considered positioned with a 3- to 4-week horizon, and schedule instabilities within this horizon likely originate from internal product-, manufacturing- and supply-related operating conditions. Consequently, we defined 2- and 4-week planning horizons, within which internal variables are expected to cause schedule inaccuracies as the market demand and gross requirement are not changed within these horizons. Furthermore, we identified 8- and 12-week planning horizons, representing periods where variables can be expected to have a larger influence.

4. Findings and discussion

4.1 Identified differences related to individual features of delivery schedule inaccuracies

Table 5 presents the results of the 20 logistic regression analyses. The table lists the significant features of the various models. Appendix 2 presents the five most important features of random forest models as a complement.

Our regression findings (Table 5) show that the forecasted volume, item's order life cycle, take rate, unit load, production deviation, pick-up frequency and transport lead times are significant variables in most models. The estimated values of the corresponding coefficients (the influence) vary in significance and direction between models.

Order life has a significant positive influence on the longer horizon for all models. As proposed, order life significantly causes gross requirement accuracy on longer horizons since demand forecast accuracy is expected to be lower early in the order life cycle when there are lower sales. On the 2-week (5, 10%, 30% and 50% models) and 4-week horizons (5%, 10% and 30% models), the order life variable significantly affects in the opposite direction. Accordingly, the later an item is in the life cycle, the higher the probability of inaccuracy on shorter horizons. When exploring the data, we found that items close to the phase-out usually have inaccurate plans due to, for instance, changed safety stocks and late phase-out rescheduling, which has also been reported in the literature (Teunter *et al.*, 2011; Wänström and Jonsson, 2006). Consequently, the findings verify that order life cycle data can explain schedule inaccuracies. However, the way it impacts varies over short and long horizons.

Our order life cycle measure indicates how late an item is in a life cycle. Although the current measure does not cover phase-out planning and its influence on inaccuracy, we expect the inaccuracy to be more significant during phase-out than during periods of continuous demand. Perhaps the size on a 2-week horizon indicates this significance, but we need a more accurate phase-out variable to measure the phase-out phenomenon. Alternative life cycle measures may cover additional life cycle patterns, for example, separating phase-in, steady-state and phase-out periods (Nepal *et al.*, 2012).

Take rate and forecasted item volume are significant in almost all models. These are measures of item commonality and product complexity and are expected to directly impact the demand forecast (i.e. the gross requirement). Items with low take rates, and corresponding low volumes, pose a significant forecasting challenge and in turn intensify schedule inaccuracies by causing lumpiness and intermittency in the requirement. A decreased item commonality could further destabilize the schedule, as outlined by Meixell (2005). Therefore, the influence of such commonality on schedule inaccuracies was anticipated across both short and long horizons. The two features are correlated, but they also differ. The forecasted item volume is affected by planning parameters (e.g. production sequencing and batch sizing) and the number of units of the respective item per assembled car. In addition, some of the transformed forecasted volume variables show partially limited levels of linearity, which may explain the lack of significance in some models (e.g. for the 100% models on 8- and 10-week horizon – See Appendix 1).

Unit load is significant in all models, with a reverse influence for the 5% and 10% models compared to the 30%, 50% and 100% models. The findings indicate that larger unit loads (i.e. lower values of the transferred unit load variable) create stability in the face of larger variations. However, these same larger unit loads can also lead to more significant inaccuracies when faced with smaller variations, suggesting the paradoxical nature of large unit loads. On the one hand, they confer responsiveness when there are larger fluctuations. On the other hand, when schedule variations are considerably smaller, larger unit loads contribute to lumpiness in demand (as suggested by Teunter *et al.* (2011) and Wänström and Jonsson (2006)), which amplifies schedule variations. This dichotomy underscores the delicate balance involved in managing unit loads to optimize schedule accuracy.

Accuracy level Horizon (weeks) N total N accurate	5% models				10% models				30% models				50% models				100% models			
	2	4	8	12	2	4	8	12	2	4	8	12	2	4	8	12	2	4	8	12
	211,092 132,757	212,087 111,935	204,116 82,064	201,129 88,401	211,092 146,301	212,087 125,218	204,116 89,630	201,129 100,236	211,092 178,907	212,087 163,730	204,116 123,152	201,129 136,917	211,092 197,312	212,087 188,859	204,116 162,780	201,129 167,681	211,092 205,121	212,087 204,116	204,116 189,049	201,129 187,665
Features																				
Intercept	-0.91*	1.01*	1.11*	-0.44*	-1.61*	0.63*	1.08*	-0.54*	-2.64*	-0.77*	0.40*	-1.09*	-4.52*	-2.50*	-1.12*	-2.11*	-6.49*	-4.27*	-3.76*	
Car Model A	-0.03	-0.06*	-0.10*	-0.06*	-0.03	-0.07*	-0.11*	-0.06*	-0.03	-0.08*	-0.11*	-0.05*	-0.09*	-0.12*	-0.12*	-0.02	-0.17*	-0.18*	-0.20*	
Car Model B	-0.17*	-0.10*	-0.11*	-0.07*	-0.17*	-0.07*	-0.10*	-0.06*	-0.07*	-0.03	-0.07*	-0.05*	-0.06	-0.04	-0.09*	-0.10*	-0.03	-0.06	-0.06	
Car Model C	0.00	0.03	0.01	0.02	0.02	0.07*	0.03	0.04*	0.06*	0.14*	0.09*	0.08*	0.16*	0.20*	0.12*	0.08*	0.17*	0.21*	0.12*	
Car Model D	-0.08*	0.00	-0.06*	0.09	-0.14*	-0.03	-0.08*	0.08*	-0.15*	-0.04*	-0.09*	0.08*	-0.03	0.03	-0.02	0.13*	0.16*	0.18*	0.16*	
Car Model E	0.12*	0.04*	0.02	0.04*	0.13*	0.02	0.01	0.03*	0.19*	-0.04*	-0.05*	0.00	0.24*	-0.13*	-0.09*	-0.09*	0.20*	-0.11*	-0.03	
Car Model F	-0.08*	0.02	0.05*	0.12*	-0.06*	0.04	0.04	0.14*	0.05*	0.15*	0.08*	0.17*	0.06*	0.15*	0.03	0.13*	0.11*	0.20*	0.00	
Forecasted volume	0.21*	0.07*	0.06	-0.11*	0.07*	-0.05*	-0.02	-0.12*	-0.28*	-0.38*	-0.23*	-0.52*	-0.14*	-0.21*	0.05	-0.17*	-0.13*	-0.29*	0.25*	
Order life cycle	0.27*	0.16*	-0.14*	-0.50*	0.22*	0.15*	-0.05*	-0.41*	0.23*	0.11*	0.06*	-0.32*	0.13*	-0.11*	-0.10*	-0.38*	-0.04	-0.30*	-0.35*	
Take rate	-0.20*	-0.11*	-0.06*	-0.14*	-0.19*	-0.11*	-0.05*	-0.14*	-0.07*	-0.02*	-0.01	-0.05*	-0.19*	-0.13*	-0.28*	-0.22*	-0.25*	-0.18*	-0.51*	
Unit lead	-0.41*	-0.32*	-0.41*	-0.32*	-0.21*	-0.15*	-0.30*	-0.17*	0.20*	0.25*	0.02*	0.19*	0.45*	0.48*	0.27*	0.34*	0.67*	0.70*	0.56*	
Production deviation	0.28*	0.22*	0.05*	0.09*	0.29*	0.21*	0.05*	0.10*	0.21*	0.18*	0.04*	0.08*	0.12*	0.13*	0.00	0.04*	0.08*	0.08*	-0.04*	
Pick-up frequency	0.14*	0.14*	0.25*	0.13*	0.13*	0.13*	0.27*	0.12*	0.15*	0.13*	0.28*	0.15*	0.12*	0.09*	0.23*	0.16*	0.12*	0.10*	0.22*	
Transport lead time	0.09*	0.06*	-0.01	0.08*	0.07*	0.04*	-0.04*	0.09*	0.09*	0.06*	-0.09*	0.07*	0.10*	0.07*	0.03*	0.07*	0.11*	0.08*	0.09*	

Note(s): Mean coefficient estimates in the 2, 4, 8 and 12 weeks regression models. *Indicates significant ($p < 0.05$) features in the respective model, (-) indicates a negative effect on inaccuracy

Note(s): Mean coefficient estimates in the 2, 4, 8 and 12 weeks regression models, *Indicates significant ($p < 0.05$) features in the respective model, (-) indicates a negative effect on inaccuracy

Table 5.
Coefficient estimates of
features in the 20
logistic regression
models

When there is a deviation between planned and actual production volumes, item demand fulfilment in the imminent period is expected to be affected (Atadeniz and Sridharan, 2019). The effect is expected to exist on all horizons but particularly on short horizons when the assembly sequence is fixed. The findings confirm that production deviation has a short-term influence (Pujawan and Smart, 2012) during the period of a two-week fixed assembly sequence. Most production capacity losses at the studied OEM are solved by overtime on the weekends during the same weeks. Therefore, a daily production deviation affects the daily schedule accuracy but may not be visible in the weekly bucket studies. Consequently, the short-term influence may be stronger when considering daily buckets. The impact of production deviation may also be further conditioned by features not studied here; for example, the safety stock policies, capacity scalability (e.g. maximum overtime and space), suppliers' limited flexibility (Shurrab and Jonsson, 2023) and take rate with larger expected influence for items with low take rate.

While pick-up frequency might intuitively seem to enhance accuracy due to the potential for a rapid response to changing material requirements (Shurrab and Jonsson, 2023), our findings reveal a nuanced picture. Higher frequencies show a statistically significant predictive difference in all regression models, with more considerable inaccuracies associated with higher frequencies. The distribution of frequencies at the OEM ranges from daily to weekly. Because smaller production deviations are normally mitigated by weekend overtime shifts within the same week, day-to-day variations do not impact schedule accuracy as long as the weekly requirement is consolidated into weekly pickups. The aggregation of requirements on a weekly basis consequently provides robustness, leading to schedule accuracy. However, the story is different regarding daily pickups: daily variations directly impact the accuracy of these schedules.

The findings indicate that a longer transport lead time is associated with higher inaccuracies. The feature was significant in all 50% and 100% models but not in all 5%, 10% and 30% models. We expected a stronger and more significant predictive difference in transport lead time. This was not identified, probably because almost all transportation lead times are within a week and almost none is longer than two weeks.

Car models emerged as significant variables in some regression models, irrespective of the horizon or the degree of inaccuracy. Notably, how these car model variables influence schedule variations differs depending on the magnitude of the MAPE thresholds. Both the significance and, in certain instances, the directionality (whether positive or negative) of their impact on schedule variations differ. Hence, we can infer that car model variables influence schedule accuracies differently, contingent on the severity of schedule inaccuracies. Despite these findings, the interpretation of how and why different car models affect schedule accuracies is multifaceted and extends beyond the scope of this study. This complex interplay is mainly attributable to several items being utilized across multiple car models, suggesting various potential differences and characteristics that could impact variations. Consequently, an in-depth analysis of features specific to different car models is essential for a more nuanced understanding of their influence on schedule accuracy.

The influence of the item's order life cycle on delivery schedule instability becomes increasingly notable for longer, compared to shorter, planning horizons. Similarly, the production deviation influence is more notable on shorter horizons. This distinction contrasts with the take rate, forecasted volume and pickup frequency, demonstrating a relatively stable impact across all planning horizons. An inverse relationship emerges when examining the order life cycle on shorter horizons, potentially indicating a phase-out influence. Consequently, our analysis substantiates the claim that features contributing to delivery schedule instability vary in their influence, depending on the planning horizon relative to the LDOC decoupling point and the frozen production plan.

The random forest models ([Appendix 2, Figure A9](#)) identify unit load, transport lead time, order life, take rate and production deviation as the five most influential features. This roughly corroborate the general findings of the logistic regressions regarding what features affect schedule accuracy. Unit load stands out as the most influential feature in random forest models. However, it should be noted that the impact of features is not as clearly identified in random forests as in logistic regression models. Feature importance in random forest models reflects what variables are useful for classifying/predicting scheduling inaccuracies, but not how these affect its likelihood (as in logistic regression). An exploration of the unit load suggests that there are subregions in the variable range that do not follow a straightforward linear fit. This is an interesting case for how other models (i.e. non-linear) can pick up useful feature patterns to explain delivery-inaccuracy classifications.

4.2 Conceptualizing the influences on delivery schedule inaccuracy

Here, features of delivery schedule inaccuracies are discussed with the ambition of conceptualizing the causality of delivery schedule inaccuracies ([Figure 2](#)). Firstly, the findings show that features can be associated with delivery schedule inaccuracies in different ways, depending on whether the variable affects (1) before or (2) after a decoupling point separating forecasts and customer orders (here, defined as the LDOC), if it affects the gross or (3) net material requirement and if it is a direct or (4) moderating cause. We also propose a category of causes (5) related to the disruption of a levelled plan from the qualitative data analysis, but this category is not empirically analysed quantitatively here.

The first two categories cause delivery schedule inaccuracies by directly affecting gross material requirements. The LDOC operationalizes the CODP ([Wikner, 2014](#)). Before the LDOC, the market demand is known, as it consists of customer orders, not forecasts. Production capacity losses resulting in lower production output than planned and extra/exceptional customer orders accepted before the LDOC are features identified in the qualitative analysis. The features directly affect the gross requirement before the LDOC. After the LDOC, the gross requirement is more uncertain, especially for variant items with low take rates, as it is increasingly dependent on the market forecast further out in time from the LDOC. Our quantitative analysis identified a large and significant influence of the take rate feature on schedule inaccuracy before and after the LDOC. It also verifies the expected impact of the production deviation before the LDOC.

Looking at the variables describing delivery schedule inaccuracies through the lens of complexity theory reveals a network of interconnected variables. At the core of this network, we find product complexity, a determinant that various researchers have singled out for its

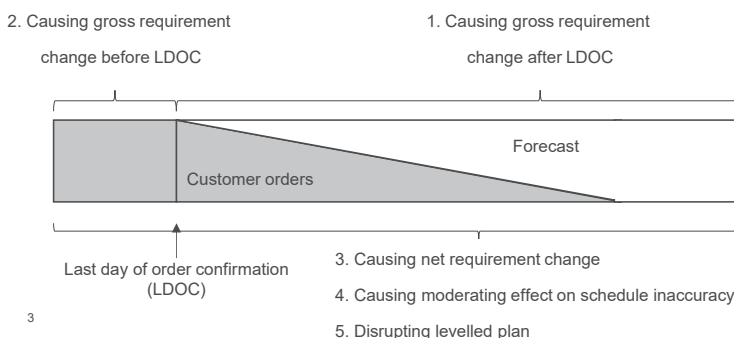


Figure 2.
Five categories of
causal effects on
delivery schedule
inaccuracies

significant influence on delivery schedules (e.g. Kabak and Ornek, 2009; Pujawan *et al.*, 2014; Shurrab and Jonsson, 2023; Sivadasan *et al.*, 2013).

The take rate and forecasted volume features serve as tangible measures of product complexity. As proposed by Serdarasan (2013) and Shurrab and Jonsson (2023), product variety can lead to increased internal variety, causing fluctuations in demand and, subsequently, delivery schedules. Incorporating the perspective of decoupling management (Banerjee *et al.*, 2012; Wikner, 2014), it is clear that product complexity management is critical. As CODPs and CADPs shift, the impact of product complexity on delivery schedules can differ. Particularly, in a non-make-to-stock environment, where the CADP plays a significant role, the ability to manage and absorb product complexity can become a decisive factor in maintaining accurate delivery schedules.

Looking at the complexity absorption capability of a firm, Huatuco *et al.* (2021) reinforce the idea that an organization's internal environment needs to be well-equipped to handle high product complexity. If there is a lack of capacity, flexibility or adequate information systems, there is a greater likelihood of schedule inaccuracies.

Echoing this, Holweg *et al.* (2018) suggest that the design and variations of processes within a firm can contribute to delivery schedule inaccuracies. The firm's internal process variations and design, such as insufficient buffering of variations and bottlenecks, can further amplify the impact of product complexity on delivery schedule inaccuracies. Summarizing these insights leads us to the following proposition:

- P1. Product complexity*, due to its multi-dimensional interactions with demand complexity and the consequential challenges it exports to both the demand environment and a firm's internal environment, is a pivotal determinant of delivery schedule inaccuracies across all time horizons. Its influence is magnified by the dynamism of the decoupling points, the variety in product demand and the firm's internal capacity to absorb complexity. This complexity triggers changes in gross requirements, thereby posing a significant challenge to maintaining accurate delivery schedules.

Figure 2 illustrates the third category, wherein changes in the net requirement emerge from process variations tied to alter planning parameters. The concept that modifications in planning parameters, including safety stocks and lot sizes, can instigate net requirement fluctuations is reinforced by studies from Atadeniz and Sridharan (2019) and Li and Disney (2017). They proposed that these parameters could serve as moderators, buffering the effects of variances on delivery schedules. Although we examined these impacts in our qualitative study, the quantitative analysis did not directly address how they altered the net requirement. Nonetheless, we indirectly deduced their influence from the order life cycle variable results, such as decreased safety stock volumes during phase-out periods (Wänström and Jonsson, 2006).

Our quantitative analysis pinpoints the order life cycle as a significant determinant of delivery schedule inaccuracies. It exposed that these inaccuracies could either be amplified or mitigated at different order life cycle stages due to interacting variables. For instance, items in the later stages of the life cycle might influence the net requirement before the LDOC due to these process variations, thereby enlarging schedule inaccuracies. As the product transitions towards its phase-out period, items tend to become slow-moving, leading to a lumpiness in their demand (Andersson and Jonsson, 2018).

A decoupling management perspective (Banerjee *et al.*, 2012; Wikner, 2014) can help elucidate further complexities. In the phase-out stage, unpredictable and declining demand could shift the CODP or extend the hybrid zone closer to the end customer, implying a transition from forecast-driven to actual demand-driven production and delivery. This move minimizes surplus stock risks. Simultaneously, during the phase-out period, companies

might opt to customize the remaining units for maintaining appeal or managing dwindling resources, moving the CADP and hybrid zone towards the end customer. This switch escalates customization while decreasing standardization. These CODP and CADP shifts could sway planning parameters, processes and scheduling decisions. The intensified dependence on actual orders (due to CODP shifts) and customer-specific adaptations (due to CADP shifts) could enlarge schedule uncertainties, potentially exacerbating inaccuracies, primarily if not adequately managed.

In contrast, items at the early life cycle stage pose distinct challenges. Holweg (2005) suggested that larger variation and complexity in early life cycle stages – owing to new product introductions and ramp-ups – spur more extensive gross requirement changes, hence inflating delivery schedule inaccuracies across the time horizon. From the complexity perspective, Shurrab and Jonsson (2023) argued that interactions between product and process complexities and demand and supply chain complexities at various life cycle stages can amplify delivery schedule instability. This perspective reaffirms complexity theory's stance on the dynamic and non-linear attributes of the variables involved. Thus, synthesizing our findings, we propose the following:

- P2. The various stages of an *item's order life cycle* significantly impact the accuracy of delivery schedules. At each stage, it presents unique challenges: Late in the cycle, increased process variations contribute to inaccuracies before the LDOC through net requirement changes due to alterations in planning parameters and lumpiness in demand. Conversely, early in the cycle, the dynamism of new product introductions contributes to inaccuracies throughout the entire time horizon via gross requirement changes. This multi-stage, multi-effect interaction reflects the complexity of the item's order life cycle in shaping delivery schedule accuracy.

The order life cycle and production deviation variables show the decoupling role, in which variables' influence before and after decoupling points separate forecast-driven from non-forecast-driven requirements (Wikner, 2014) and frozen from non-frozen production plans. Several other variables can directly (as possible causes) or indirectly (as moderators) affect schedule inaccuracy, regardless of the time horizon (Shurrab and Jonsson, 2023). However, features impacting production process performance may affect the gross requirements and their fulfilment before the LDOC (i.e. before and in the hybrid decision zone of LDOC), as referred to by Wikner (2014). In contrast, process variables generating changes in planning parameters can change net requirements, affecting schedule accuracy before or even after the LDOC, depending on the variable in play. Such changes could be adaptations to customer orders but are typically internally generated and, therefore, unrelated to a CADP (Wikner, 2014). When they occur depends on internal time fences and planning policies. Previous studies have indirectly supported the contingency effect of the CODP. For instance, although delayed differentiation to final assembly operations (i.e. pushing back the CODP) enhances competitiveness (Blecker and Abdelkafi, 2006), Shurrab and Jonsson (2023) and Pujawan *et al.* (2014) found it also causes late changes to end-item specifications. Atadeniz and Sridharan (2019) found a negative effect of late demand increases and short frozen periods – a challenge also emphasized by Holweg (2005) – highlighting the criticality of decoupling point and time fence choices for delivery schedule accuracies. Consequently, we introduce the LDOC to represent a critical decoupling point and emphasize the importance of time fences for parameter changes in delivery schedule accuracy. As such, we propose the following:

- P3. The influence of variables impacting delivery schedule inaccuracies is contingent on (a) the LDOC *decoupling point* separating forecast from non-forecast-driven requirements and (b) the internally generated *time fences* for parameter revision.

The fourth and fifth categories in Figure 2 do not directly affect gross and net requirements but contribute to inaccuracies in other ways. The fourth type moderates the influence of inaccuracies multiplicatively. We identified some variables in this category as process design features, such as large unit loads, infrequent pick-up frequencies, local vacations, holidays and opening hours. They usually amplify generated (caused) variations, aligning with previous studies' results (e.g. Inman and Gonsalvez, 1997; Krajewski *et al.*, 2005). Strikingly, we identified a significant predictive difference between these variables and that they may stabilize schedules by reducing variations on the short horizon while amplifying on the long horizon, or vice versa. Both the complexity and process theory perspectives motivate a combined direct and indirect effect of features on schedule inaccuracies. Still, the combined amplifying and stabilizing effects of the same variable have not been emphasized in the literature. This indicates a combined complexity absorbing and generating effect for the same variables. Consequently, we propose the following paradox:

- P4. A variable may have a combined amplifying (complexity generating) and stabilizing (complexity absorbing) moderating effect on delivery schedule accuracies.*

The fifth category causes inaccuracies by disrupting the levelling of a plan. This effect is specific to the manufacturing strategy environment of our empirical study (i.e. levelled production in repetitive high-volume manufacturing). We did not directly study this fifth effect in the quantitative study. Still, examples of process design features were identified in the qualitative study: unit loads, pick-up frequencies and planned seasonality build-ups. The disrupting influence of these features is primarily due to synchronization and compatibility issues with the supply chain. These issues are typically attributed to process design features, such as increasing product complexities (Bozarth *et al.*, 2009; Fernández Campos *et al.*, 2019), constraints on capacity scalability and strict delivery terms (Shurrab and Jonsson, 2023) and process context features, for example, limited suppliers' flexibilities (Ponomarov and Holcomb, 2009). Accordingly, we propose the following:

- P5. Variables contributing to disrupting a levelled production plan are significant possible causes of delivery schedule inaccuracies, regardless of the time horizon. They could directly cause inaccuracy by generating gross requirement changes.*

5. Conclusions

The study contributes by exploring variables with potential direct causal and moderating effects, empirically testing and validating previous research outcomes on the material delivery scheduling process and conceptualizing how features can influence and moderate delivery schedule inaccuracies. We explored how variables are related to schedule variations in five ways: affecting the gross requirement before or after the LDOC decoupling point, affecting the net requirement, adding a multiplicative effect on a schedule variation and disrupting a levelled plan. The empirical testing verified that the features relate to the planning horizon (related to the LDOC and frozen production plans). It also identified that the features have varying influence between low and high inaccuracies, may have reversed predictive differences on the inaccuracy given various planning horizons and may have combined amplifying and stabilizing influence.

The study also contributes by identifying take rate and forecasted volume (expressing item volume/commonality), item's order life cycle, unit load, production deviation and pickup frequency as essential variables to explain delivery schedule inaccuracies. Higher take rates and less frequent pickup frequencies positively relate to schedule accuracy regardless of the planning horizon, while the order life cycle is related to schedule accuracy differently early and late in the life cycle. Production deviation has a relatively stronger effect on the short

horizon. Larger unit loads have stabilizing effects for larger variations but may have amplifying effects for smaller variations. However, further studies are needed to understand more deeply how individual variables cause delivery schedule inaccuracies.

From a complexity theory perspective (e.g. Aitken *et al.*, 2016; Holweg *et al.*, 2018; Huatuco *et al.*, 2021; Ateş *et al.*, 2022; Shurrab and Jonsson, 2023), we see how product-related complexity has a direct negative effect on delivery schedule accuracy and thus also on operational performance. However, as emphasized in the literature (e.g. Aitken *et al.*, 2016; Ateş *et al.*, 2022), product complexity may have a positive effect on both innovation and financial performance, so reducing product complexity is not a generally feasible strategy, especially for companies and supply chains competing with design, quality and innovation capabilities, as is the case for the automotive OEM empirically studied here. A certain level of product complexity may be required by the firm's business strategy. This complexity should then be accommodated rather than reduced. Therefore, it is important to understand how product complexity is contingent on business strategy and how complexity absorption capabilities can be generated and adopted and have effects. Traditionally, absorption capabilities have been related to slack resources, information systems and relational mechanisms (Galbraith, 1974; Aitken *et al.*, 2016). We identify such mechanisms (e.g. capacity strategies to manage production disturbances), but we also identify the key role of decoupling management and related planning policies and time fences for generating process stability and absorptive capabilities contributing to delivery schedule accuracy.

Several managerial implications can be derived from this analysis. The five categories of possible causes of delivery schedule variations could help us understand how internal processes, conditions and parameters generate variations. The impact of the item's order life cycle motivates differentiated item management according to life cycle phases. The take rate impact suggests an effect of item complexity reduction, as it reduces variations and/or the generation of complexity absorption capabilities to manage the implications of item-related complexity. The stabilizing influence of large unit loads and weekly pickup frequencies is interesting. This indicates a possible tipping point where the lumpiness generated by large unit loads is related stronger and more negatively to schedule inaccuracy than a positive stabilizing influence.

The general propositions are also expected to be relevant to environments other than automotive OEMs, but some results are affected by our case-specific environment. The product complexity in our studied environment is high in terms of bill-of-material width and breadth, component variant breadth (many items with very small take rates) and frequent phase-in/out of components with high technology clock speed. The product complexity characteristics are a result of the type of product as well as the OEM's business strategy. The highly significant features take rate and the item's order life cycle relate to product complexity. The identified predictive difference of take rate on delivery schedule accuracy is expected to be general and exists in situations with low product complexities. The item's order life cycle effect, however, is less likely to be transferrable to situations with low technology replacement rates and low levels and frequencies of phasing-in/out components in product models. The assemble-to-order and lean manufacturing environment, characterized by levelled production and frequent deliveries of supplied items in sequence, or small batches with short transport lead times, may also affect the findings. In environments with lumpier and larger proportions of intercontinental supply and long transport lead times, some features may have stronger and more significant coefficients than in this study.

Consequently, a comparison of causal effects and/or predictive differences across various planning environments is left for further research. Another limitation of the study design is the weekly bucketing of schedules. This aggregation may have eliminated some of the daily schedule variations, thereby eliminating some significant relationships to variables on a daily

level. Using daily data would consequently be interesting. The next research step is to test the extent to which predictive forecasting models could be developed using the variables presented in this study to propose alternative reference volumes to what is expressed in delivery schedules. Findings from the random forest models (Appendix 2) can serve as first steps towards this and suggest that the data could be successfully used for out-of-sample prediction and forecasting.

Notes

1. They reported that one-third of the schedules from three automotive OEMs have less than 90% accuracy on a three-week horizon, and half of them have less than 90% accuracy on an 8-week horizon. The average accuracy for all schedules was 70% on a three-week horizon and 50% on an 8-week horizon.
2. James, G., Witten, D., Hastie, T., Tibshirani, R. (2021), *An introduction to statistical learning: With applications in R*, Springer US.
3. Breiman, L. (2017), *Classification and regression trees*, CRC Press.
4. Breiman, L. (2001), "Random forests", *Machine Learning*, Vol. 45, No. 5, pp. 5–32.

References

- Aitken, J., Bozarth, C. and Garn, W. (2016), "To eliminate or absorb supply chain complexity: a conceptual model and case study", *Supply Chain Management: An International Journal*, Vol. 21 No. 6, pp. 759-774, doi: [10.1108/scm-02-2016-0044](https://doi.org/10.1108/scm-02-2016-0044).
- Amaro, G., Hendry, L. and Kingsman, B. (1999), "Competitive advantage, customisation and a new taxonomy for non make-to-stock companies", *International Journal of Operations and Production Management*, Vol. 19 No. 4, pp. 349-371, doi: [10.1108/01443579910254213](https://doi.org/10.1108/01443579910254213).
- Andersson, J. and Jonsson, P. (2018), "Big data in spare parts supply chains", *International Journal of Physical Distribution and Logistics Management*, Vol. 48 No. 5, pp. 524-544, doi: [10.1108/ijpdlm-01-2018-0025](https://doi.org/10.1108/ijpdlm-01-2018-0025).
- Atadeniz, S.N. and Sridharan, S.V. (2019), "Effectiveness of nervousness reduction policies when capacity is constrained", *International Journal of Production Research*, Vol. 57 No. 1, pp. 1-17, doi: [10.1080/00207543.2019.1643513](https://doi.org/10.1080/00207543.2019.1643513).
- Ateş, M.A., Suurmond, R., Luzzini, D. and Karuse, D. (2022), "Order from chaos: a meta-analysis of supply chain complexity and firm performance", *Journal of Supply Chain Management*, Vol. 58 No. 1, pp. 3-30, doi: [10.1111/jscm.12264](https://doi.org/10.1111/jscm.12264).
- Banerjee, A., Sarkar, B. and Mukhopadhyay, S.K. (2012), "Multiple decoupling point paradigms in a global supply chain syndrome: a relational analysis", *International Journal of Production Research*, Vol. 50 No. 11, pp. 3051-3065, doi: [10.1080/00207543.2011.588624](https://doi.org/10.1080/00207543.2011.588624).
- Blackstone, J.H. Jr (2013), *APICS Dictionary: the Essential Supply Chain Reference*, APICS, Chicago.
- Blecker, T. and Abdelkafi, N. (2006), "Complexity and variety in mass customization systems: analysis and recommendations", *Management Decision*, Vol. 44 No. 7, pp. 908-929, doi: [10.1108/00251740610680596](https://doi.org/10.1108/00251740610680596).
- Bozarth, C.C., Warsing, D.P., Flynn, B.B. and Flynn, E.J. (2009), "The impact of supply chain complexity on manufacturing plant performance", *Journal of Operations Management*, Vol. 27 No. 1, pp. 78-93, doi: [10.1016/j.jom.2008.07.003](https://doi.org/10.1016/j.jom.2008.07.003).
- Chand, P., Kumar, A., Thakkar, J. and Ghosh, K.K. (2022), "Direct and mediation effect of supply chain complexity drivers on supply chain performance: an empirical evidence of organizational complexity theory", *International Journal of Operations and Production Management*, Vol. 42 No. 6, pp. 797-825, doi: [10.1108/ijopm-11-2021-0681](https://doi.org/10.1108/ijopm-11-2021-0681).

- Davis, D.F., Golobic, S.L. and Boerstler, C.N. (2011), "Benefits and challenges of conducting multiple methods research in marketing", *Journal of the Academy of Marketing Science*, Vol. 39 No. 3, pp. 467-479, doi: [10.1007/s11747-010-0204-7](https://doi.org/10.1007/s11747-010-0204-7).
- Fernández Campos, P., Trucco, P. and Huatuco, L.H. (2019), "Managing structural and dynamic complexity in supply chains: insights from four case studies", *Production Planning and Control*, Vol. 30 No. 8, pp. 611-623.
- Galbraith, J.R. (1974), "Organization design: an information processing view", *Interfaces*, Vol. 4 No. 3, pp. 28-36, doi: [10.1287/inte.4.3.28](https://doi.org/10.1287/inte.4.3.28).
- Gomez-Rubio, V. (2020), *Bayesian Inference with INLA*, CRC Press, Boca Raton.
- Harrison, A. (1997), "Investigating the sources and causes of schedule instability", *The International Journal of Logistics Management*, Vol. 8 No. 2, pp. 75-82, doi: [10.1108/09574099710805682](https://doi.org/10.1108/09574099710805682).
- Hastie, T., Tibshirani, R. and Friedman, J. (2009), *Elements of Statistical Learning*, Springer, New York.
- Herrera, C., Belmokhtar-Berraf, S., Thomas, A. and Parada, V. (2016), "A reactive decision-making approach to reduce instability in a master production schedule", *International Journal of Production Research*, Vol. 54 No. 8, pp. 2394-2404, doi: [10.1080/00207543.2015.1078516](https://doi.org/10.1080/00207543.2015.1078516).
- Ho, C.-J. (1989), "Evaluating the impact of operating environments on MRP system nervousness", *International Journal of Production Research*, Vol. 27 No. 7, pp. 1115-1135, doi: [10.1080/00207548908942611](https://doi.org/10.1080/00207548908942611).
- Ho, C.-J. and Carter, P.L. (1996), "An investigation of alternative dampening procedures to cope with MRP system nervousness", *International Journal of Production Research*, Vol. 34 No. 1, pp. 137-156, doi: [10.1080/00207549608904895](https://doi.org/10.1080/00207549608904895).
- Holweg, M. (2005), "An investigation into supplier responsiveness", *The International Journal of Logistics Management*, Vol. 16 No. 1, pp. 96-119, doi: [10.1108/09574090510617376](https://doi.org/10.1108/09574090510617376).
- Holweg, M., Davies, J., de Meyer, A., Lawson, B. and Schmenner, R.W. (2018), *Process Theory: the Principles of Operations Management*, Oxford University Press, Oxford.
- Huatuco, L.H., Smart, J., Calinescu, A. and Sivadasan, S. (2021), "Complexity transfer in supplier-customer systems", *Production Planning and Control*, Vol. 32 No. 9, pp. 747-759, doi: [10.1080/09537287.2020.1762135](https://doi.org/10.1080/09537287.2020.1762135).
- Inman, R.R. and Gonsalvez, D.J.A. (1997), "The causes of schedule instability in an automotive supply chain", *Production and Inventory Management Journal*, Vol. 38 No. 2, pp. 26-31.
- Jonsson, P. and Mattsson, S.-A. (2003), "The implications of fit between planning environments and manufacturing planning and control methods", *International Journal of Operations and Production Management*, Vol. 23 No. 8, pp. 872-900, doi: [10.1108/01443570310486338](https://doi.org/10.1108/01443570310486338).
- Jonsson, P. and Myrelid, P. (2016), "Supply chain information utilisation: conceptualisation and antecedents", *International Journal of Operations and Production Management*, Vol. 36 No. 12, pp. 1769-1799, doi: [10.1108/ijopm-11-2014-0554](https://doi.org/10.1108/ijopm-11-2014-0554).
- Kabak, K.E. and Ornek, A.M. (2009), "An improved metric for measuring multi-item multi-level schedule instability under rolling schedules", *Computers and Industrial Engineering*, Vol. 56 No. 2, pp. 691-707, doi: [10.1016/j.cie.2006.11.001](https://doi.org/10.1016/j.cie.2006.11.001).
- Kenney, J.F. and Keeping, E.S. (1951), *Mathematics of Statistics*, Van Nostrand, NY.
- Kim, S. and Kim, H. (2016), "A new metric of absolute percentage error for intermittent demand forecasts", *International Journal of Forecasting*, Vol. 32 No. 3, pp. 669-679, doi: [10.1016/j.ijforecast.2015.12.003](https://doi.org/10.1016/j.ijforecast.2015.12.003).
- Koopmans, M. (2017), "Perspectives on complexity, its definition and applications in the field", *Complicity: An International Journal of Complexity and Education*, Vol. 14 No. 1, pp. 16-35, doi: [10.29173/cmplt27611](https://doi.org/10.29173/cmplt27611).
- Krajewski, L., Wei, J.C. and Tang, L.-L. (2005), "Responding to schedule changes in build-to-order supply chains", *Journal of Operations Management*, Vol. 23 No. 5, pp. 452-469, doi: [10.1016/j.jom.2004.10.006](https://doi.org/10.1016/j.jom.2004.10.006).

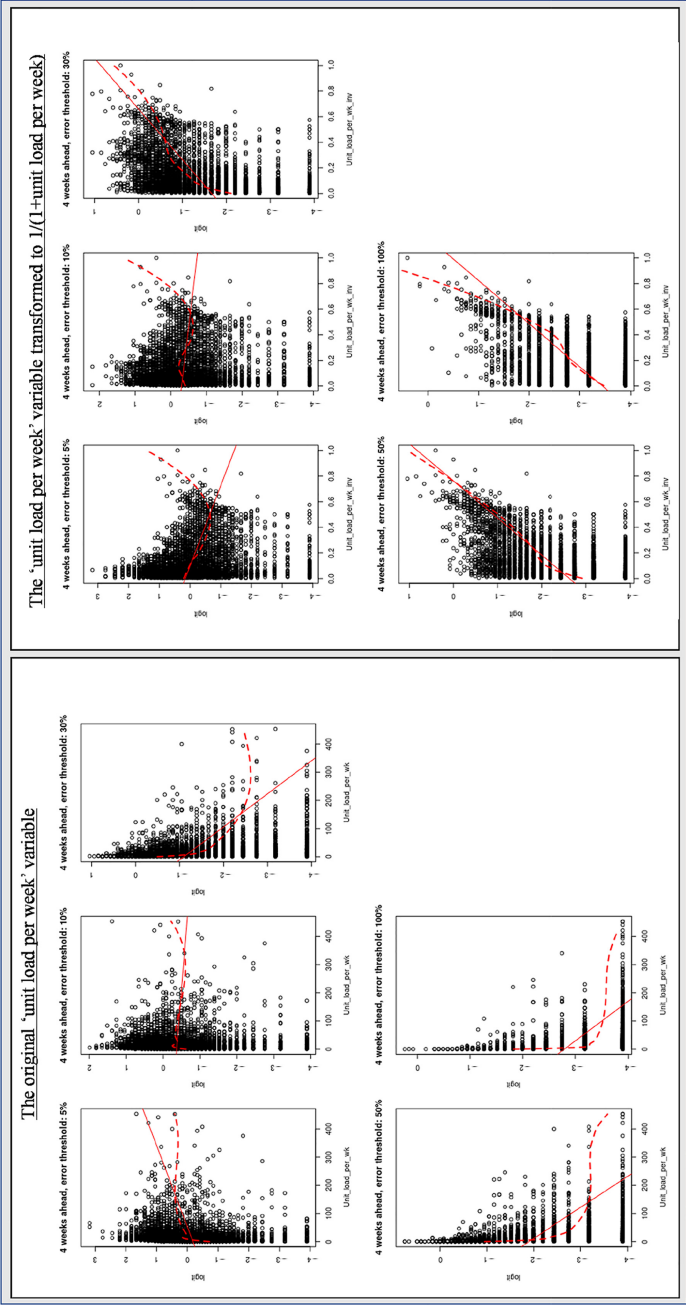
- Lalami, I., Frein, Y. and Gayon, J.-P. (2017), "Production planning in automotive powertrain plants: a case study", *International Journal of Production Research*, Vol. 55 No. 18, pp. 5378-5393, doi: [10.1080/00207543.2017.1315192](https://doi.org/10.1080/00207543.2017.1315192).
- Lee, T.S. and Adam, E.E. (1986), "Forecasting error evaluation in material requirements planning (MRP) production-inventory systems", *Management Science*, Vol. 32 No. 9, pp. 1186-1205, doi: [10.1287/mnsc.32.9.1186](https://doi.org/10.1287/mnsc.32.9.1186).
- Li, Q. and Disney, S.M. (2017), "Revisiting rescheduling: MRP nervousness and the bullwhip effect", *International Journal of Production Research*, Vol. 55 No. 7, pp. 1992-2012, doi: [10.1080/00207543.2016.1261196](https://doi.org/10.1080/00207543.2016.1261196).
- Makridakis, S., Wheelwright, S.C. and Hyndman, R.J. (1998), *Forecasting: Methods and Applications*, Wiley, New York.
- Manuj, I. and Sahin, F. (2011), "A model of supply chain and supply chain decision-making complexity", *International Journal of Physical Distribution and Logistics Management*, Vol. 41 No. 5, pp. 511-549, doi: [10.1108/09600031111138844](https://doi.org/10.1108/09600031111138844).
- Maylor, H. and Turner, N. (2017), "Understand, reduce, respond: project complexity management theory and practice", *International Journal of Operations and Production Management*, Vol. 37 No. 8, pp. 1076-1093, doi: [10.1108/ijopm-05-2016-0263](https://doi.org/10.1108/ijopm-05-2016-0263).
- Meixell, M.J. (2005), "The impact of setup costs, commonality, and capacity on schedule stability: an exploratory study", *International Journal of Production Economics*, Vol. 95 No. 1, pp. 95-107, doi: [10.1016/j.ijpe.2003.11.010](https://doi.org/10.1016/j.ijpe.2003.11.010).
- Mendoza, D.J., Mula, J. and Campuzano-Bolarin, F. (2014), "Using systems dynamics to evaluate the tradeoff among supply chain aggregate production planning policies", *International Journal of Operations and Production Management*, Vol. 34 No. 8, pp. 1055-1079, doi: [10.1108/ijopm-06-2012-0238](https://doi.org/10.1108/ijopm-06-2012-0238).
- Myrelid, P. (2017), "Information quality deficiencies in delivery schedules and their impact on production scheduling", *Production Planning and Control*, Vol. 28 No. 3, pp. 232-243, doi: [10.1080/09537287.2016.1262079](https://doi.org/10.1080/09537287.2016.1262079).
- Nepal, B., Murat, A. and Babu Chinnam, R. (2012), "The bullwhip effect in capacitated supply chains with consideration for product life-cycle aspects", *International Journal of Production Economics*, Vol. 136 No. 2, pp. 318-331, doi: [10.1016/j.ijpe.2011.12.018](https://doi.org/10.1016/j.ijpe.2011.12.018).
- Ponomarev, S.Y. and Holcomb, M.C. (2009), "Understanding the concept of supply chain resilience", *The International Journal of Logistics Management*, Vol. 20 No. 1, pp. 124-143, doi: [10.1108/09574090910954873](https://doi.org/10.1108/09574090910954873).
- Pujawan, I.N. (2008), "Schedule instability in a supply chain: an experimental study", *International Journal of Inventory Research*, Vol. 1 No. 1, pp. 53-66, doi: [10.1504/ijir.2008.019208](https://doi.org/10.1504/ijir.2008.019208).
- Pujawan, I.N. and Smart, A.U. (2012), "Factors affecting schedule instability in manufacturing companies", *International Journal of Production Research*, Vol. 50 No. 8, pp. 2252-2266, doi: [10.1080/00207543.2011.575095](https://doi.org/10.1080/00207543.2011.575095).
- Pujawan, I.N., Er, M., Kritchanai, D. and Somboonwivat, T. (2014), "Uncertainty and schedule instability in supply chain: insights from case studies", *International Journal of Services and Operations Management*, Vol. 19 No. 4, pp. 468-490, doi: [10.1504/ijsum.2014.065670](https://doi.org/10.1504/ijsum.2014.065670).
- Roehrich, J. and Lewis, M. (2014), "Procuring complex performance: implications for exchange governance complexity", *International Journal of Operations and Production Management*, Vol. 34 No. 2, pp. 221-241, doi: [10.1108/ijopm-01-2011-0024](https://doi.org/10.1108/ijopm-01-2011-0024).
- Serdarasan, S. (2013), "A review of supply chain complexity drivers", *Computers and Industrial Engineering*, Vol. 66 No. 3, pp. 533-540, doi: [10.1016/j.cie.2012.12.008](https://doi.org/10.1016/j.cie.2012.12.008).
- Shurrah, H. and Jonsson, P. (2023), "Untangling the complexity generating material delivery 'schedule instability' insights from automotive OEMs", *International Journal of Operations and Production Management*, Vol. 43 No. 2, pp. 235-273, doi: [10.1108/ijopm-02-2022-0105](https://doi.org/10.1108/ijopm-02-2022-0105).

-
- Shurrab, H., Jonsson, P. and Johansson, M.I. (2022a), "Managing complexity through integrative tactical planning in engineer-to-order environments: insights from four case studies", *Production Planning and Control*, Vol. 33 Nos 9-10, pp. 907-924, doi: [10.1080/09537287.2020.1837937](https://doi.org/10.1080/09537287.2020.1837937).
- Shurrab, H., Jonsson, P. and Johansson, M.I. (2022b), "A tactical demand-supply planning framework to manage complexity in engineer-to-order environments: insights from an in-depth case study", *Production Planning and Control*, Vol. 33 No. 5, pp. 462-479, doi: [10.1080/09537287.2020.1829147](https://doi.org/10.1080/09537287.2020.1829147).
- Sivadasan, S., Smart, J., Huatuco, L.H. and Calinescu, A. (2013), "Reducing schedule instability by identifying and omitting complexity-adding information flows at the supplier-customer interface", *International Journal of Production Economics*, Vol. 145 No. 1, pp. 253-262, doi: [10.1016/j.ijpe.2013.04.043](https://doi.org/10.1016/j.ijpe.2013.04.043).
- Steele, D.C. (1975), "The nervous MRP system: how to do battle", *Production and Inventory Management*, Vol. 16 No. 4, pp. 83-89.
- Teunter, R.H., Syntetos, A.A. and Zied Babai, M. (2011), "Intermittent demand: linking forecasting to inventory obsolescence", *European Journal of Operational Research*, Vol. 214 No. 3, pp. 606-615, doi: [10.1016/j.ejor.2011.05.018](https://doi.org/10.1016/j.ejor.2011.05.018).
- Tofallis, C. (2015), "A better measure of relative prediction accuracy for model selection and model estimation", *Journal of the Operational Research Society*, Vol. 66 No. 8, pp. 1352-1362, doi: [10.1057/jors.2014.103](https://doi.org/10.1057/jors.2014.103).
- Turner, J.R. and Baker, R.M. (2019), "Complexity theory: an overview with potential applications for the social sciences", *Systems*, Vol. 7 No. 1, p. 4, doi: [10.3390/systems7010004](https://doi.org/10.3390/systems7010004).
- VDA (2008), *Forecast-Qualitätskennzahl: Definition und Anwendung*, Verband der Automobilindustrie, Frankfurt.
- Wänström, C. and Jonsson, P. (2006), "The impact of engineering changes on materials planning", *Journal of Manufacturing Technology Management*, Vol. 17 No. 5, pp. 561-584, doi: [10.1108/17410380610668522](https://doi.org/10.1108/17410380610668522).
- White, J., Steingold, S. and Fournelle, C. (2004), "Performance metrics for group-detection algorithms", *Proceeding of the 36th Symposium on the Interface*, Baltimore, MD, pp. 26-29.
- Wikner, J. (2014), "On decoupling points and decoupling zones", *Production and Manufacturing Research*, Vol. 2 No. 1, pp. 167-215, doi: [10.1080/21693277.2014.898219](https://doi.org/10.1080/21693277.2014.898219).
- Wikner, J. and Rudberg, M. (2005), "Integrating production and engineering perspectives on the customer order decoupling point", *International Journal of Operations and Production Management*, Vol. 25 No. 7, pp. 623-641, doi: [10.1108/01443570510605072](https://doi.org/10.1108/01443570510605072).
- Zhao, X. and Lee, T.S. (1996), "Freezing the master production schedule in multilevel material requirements planning systems under deterministic demand", *Production Planning and Control*, Vol. 7 No. 2, pp. 144-161, doi: [10.1080/09537289608930337](https://doi.org/10.1080/09537289608930337).
- Zhao, X., Goodale, J.C. and Lee, T.S. (1995), "Lot-sizing rules and freezing the master production schedule in material requirements planning systems under demand uncertainty", *International Journal of Production Research*, Vol. 33 No. 8, pp. 2241-2276, doi: [10.1080/00207549508904814](https://doi.org/10.1080/00207549508904814).

Corresponding author

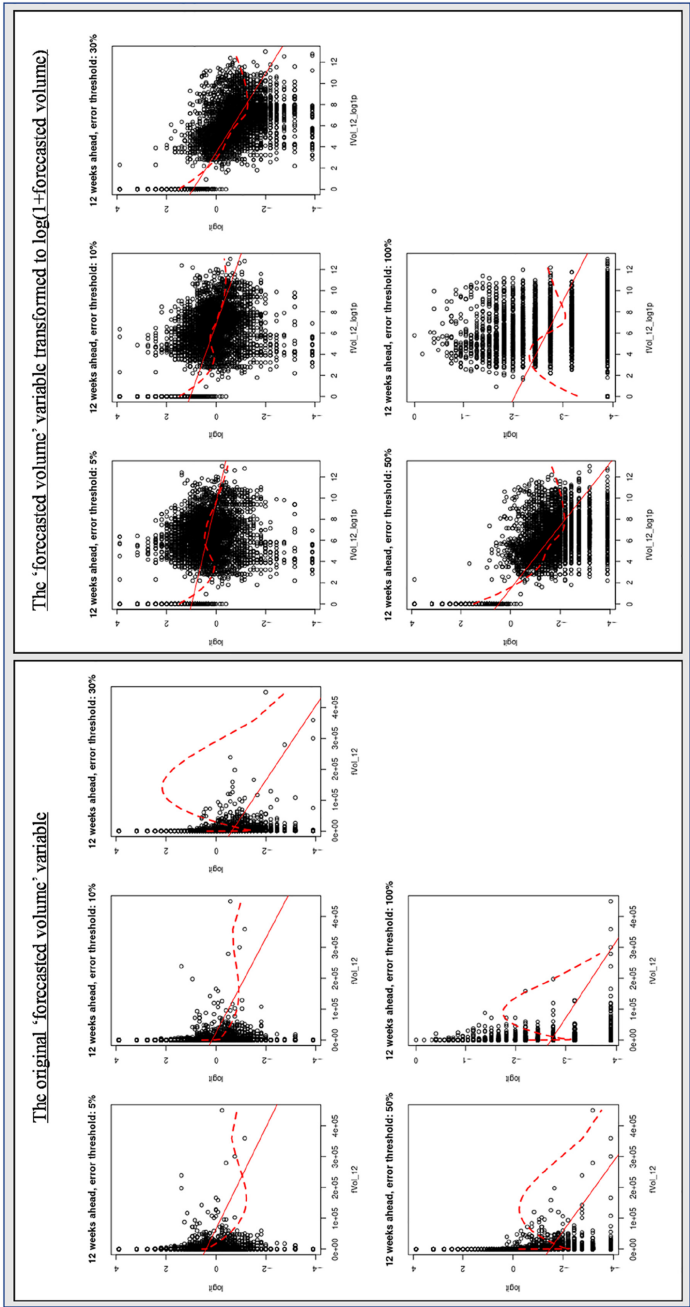
Patrik Jonsson can be contacted at: patrik.jonsson@chalmers.se

Figure A1.
Logit plots for the “unit
load per week”
variables



Note(s): The five graphs to the left present logit plots for the original ‘unit load per week’ variable. Non-linearity was identified for the 50 and 100% models. The five graphs to the right represent logit plots for the ‘unit load per week’ variable transformed to $1/(1+\text{unit load per week})$. Acceptable linearity is shown for all models with this variable

Source(s): Figure created by authors



Note(s): The five graphs to the left present logit plots for the original 'forecasted volume' variable on 12 weeks horizon. Non-linearity was identified for all models. The five graphs to the right represent logit plots for the 'forecasted volume' variable transformed to log(1+ forecasted volume). Acceptable linearity is shown for all models except for the 100% model's

Source(s): Figure created by authors

Figure A2.
Logit plots for the
"forecasted volume"
variables

Motivation

Our primary approach in the paper uses logistic regression, which provides an interpretable explanatory fit using the estimated coefficients. This leaves out-of-sample prediction aside: For logistic regression models, the sign and size of the estimated coefficients describe the average predictive difference per unit of variable within the training data (Given that the linearity assumptions hold sufficiently well.). It is reasonable, given that such simple models have high bias and low variance.

Turning to other methods in machine learning, several ways exist to assess variable/feature importance for the scheduling inaccuracy classification task, besides estimating the size of model coefficients. It is also conceivable that other non-linear relationships exist between one or more of our features and the target variable. It should not be a surprise that training variables based on variable interactions may improve the prediction of delivery schedule inaccuracies, but at the cost of making model interpretation more complicated by including non-linearities and interactions (low bias).

As the first step to explore this direction, we fit alternative models for the classification task using random forest classification models (boosting regression trees, lower bias/higher variance models) [2]. Random forest models fit relatively straightforward to a classification task where the aim is prediction rather than statistical explanation. For random forests, measures of variable importance typically do not provide a straightforward interpretation of direction. However, using random forest models also allows some exploration of what variables are most important for predictive performance on data that the model was trained for. Moreover, the low bias/high variance property of random forest means that the model may change much across parts of the data set. We address this by using cross-validation.

Method: Random forest variable permutation importance

The technical setup is as follows: Using the same explanatory variables as for logistic regression, we instead, fit separate random forest models to binary classification tasks (one binary classification task per unique scheduling horizon/inaccuracy threshold definition). We split the data into a 75% training set (at random) and a 25% test set (model testing out of sample). After a given random forest model was fit to the training data, we evaluate variable importance by how useful the included variables are used by the random forest models on the test set, when classifying delivery schedule inaccuracies. Precisely, we measure the impact on classification *recall*: The share of true delivery schedule inaccuracies that also get classified as such by the model.

Two aspects of our data are noteworthy to take into account: Firstly, we have possibly a clearly skewed classification task at hand (when the inaccuracy threshold in these events is rare: reflected in our target variable). A classifier may take shortcuts using something similar to the majority rule, and other metrics than classification accuracy become relevant. Here, we study the precision-recall trade-offs as well as the trade-off between true positive rate (recall) and false positive rate (the share of delivery schedule inaccuracy classifications by the model, that are in fact no inaccuracies). To handle skewness in the target variable, when splitting the data into train and test parts, we use stratified sampling.

Secondly, we have high-cardinality variables: Some numerical features have many unique values. This may introduce a bias against variables when directly using the mean decrease of the Gini impurity [3], which is another frequently used approach for random forest variable importance. Moreover, it may inflate the importance of variables (unique values) if computed on the training data. We handle these two concerns by random permutations of variable values, the so-called permutation importance [4].

As a baseline dummy majority classifier, it observes the majority class (for most planning horizons and inaccuracy thresholds, there is no delivery schedule inaccuracy) and classifies/predicts all cases in the test data as the majority class.

We vary random forest hyper-parameters for each horizon/threshold scenario and 10-fold cross-validation with stratified sampling. More specifically, we vary the number of estimator decision trees, tree depth, number of random variables per split and minimum number of samples per tree node. As keeping both recall high and false positive rate (and precision) low are important in this scheduling problem, we optimized for both the Area under the ROC Curve and F1 scores. These give similar results for variable importance (reported below).

The results presented in [Figures A3–A10](#) show and indicate that:

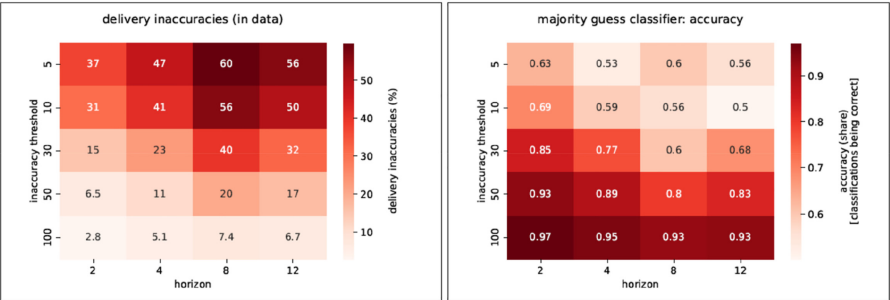
- (1) Skewed data: For higher inaccuracy thresholds, we have imbalanced data (target variable) where nearly all scheduled volumes are considered correct. This is the case for at least 50% and 100% inaccuracy thresholds, shown in [Figure A3](#). It is reasonable to use the majority dummy classifier to compare the performance of the random forest models, as the data is skewed.
- (2) The largest shares of delivery inaccuracies are obtained for low inaccuracy thresholds (natural and per definition) but also for longer horizons, which is not fully unexpected. However, interestingly, a horizon of 8 weeks ahead is consistently more associated with inaccuracies than 12 weeks ahead. For low inaccuracy thresholds and long horizons, we have that a majority of the scheduled volumes are inaccurate. This is shown in [Figures A3 and A4](#).
- (3) Learning and evaluation: A random forest fit to the data shows good scores for accuracy for lower schedule inaccuracy thresholds. However, performance is quite similar to the majority classifier for higher thresholds ([Figure A4](#)). Evaluating precision and recall ([Figure A5](#)) suggests that it could be easier to have both high precision and recall for longer time horizons.
- (4) A further evaluation of precision and recall (with the F1 score) and the trade-off between recall and a false positive classification rate ([Figure A6](#)): This shows that delivery schedule inaccuracies on the 8-week horizon are comparably easier to predict. A more detailed look into decision thresholds ([Figure A7](#)), by varying the decision threshold (the share of classification decision trees required for a positive label) also suggests a pattern. An 8-week horizon looks more promising for finding relatively high scores for precision and recall. This is consistent with [Figure A9](#). For further work, we note that prediction out of sample for 8 weeks ahead seems more promising than the other cases.
- (5) Forecasting is left for further work: The capacity to predict delivery schedule inaccuracies out of sample (with in some cases, simultaneously high precision and recall) are promising suggestions for forecasting. However, we should note that our results are no pure forecasts (predicting out of sample to examine variable importance, not ordered by time). A full setup for delivery schedule inaccuracy forecasting would not permit learning from future data. Note that in our setup, we have randomly split the data to estimate variable importance for prediction/explanation, and not yet estimated forecasting.
- (6) Variable importance: We report on the effect of permuting individual features, then estimate the drop on the recall score (share of delivery inaccuracies in the test data, correctly labelled by the classifier). [Figure A10](#) shows the mean effects of variable importance. In the aggregate, the five most important variables are Unit load, transport lead time, take rate, order life cycle, and production deviation.

The forecasted volume is also important, but a few negative scores here suggest that learning and generalizing from forecasted volume can in a few cases be tricky (left for further work). An exploration of the data suggests that the forecasted volume variable contains a small share of clear outliers: [Figure A10](#) shows that on, for example, a 12-week horizon scrambling the forecasted volume variable improves recall for smaller inaccuracy thresholds. There are different ways to understand this, depending on where said variable is typically used in the decision trees. In the following, we briefly consider possibilities.

In terms of decision trees (the base estimator), this can be understood as reversing the majority voting improves recall to permute the variable. Several explanations are possible here. If the variable is used for splits close to tree leaves, considering how the variable is directly related to the target variable becomes important. An inspection of the data suggests that this could have to do with outliers: For lower inaccuracy thresholds, the relationship between the target variables for outlier values of the variable is much more balanced. Moreover, it is also possible that there is a non-trivial amount of scheduled orders with nearly identical predictors, but where different outcomes are related to precisely this variable.

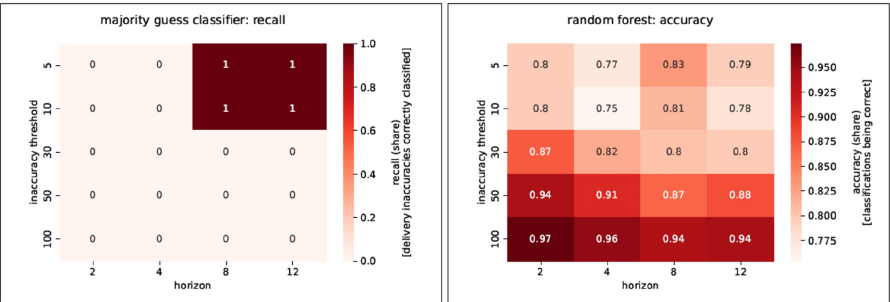
It is also possible that a negative score (positive effect on recall score) could reflect limits with greedy training algorithms (that determine decision tree splits). More specifically, a variable with negative

Figure A3.
Delivery schedule
inaccuracy description



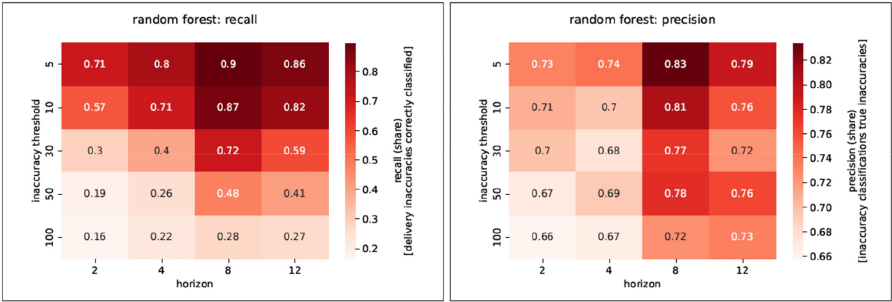
Source(s): Figure created by authors

Figure A4.
Delivery schedule
inaccuracy
classification fit

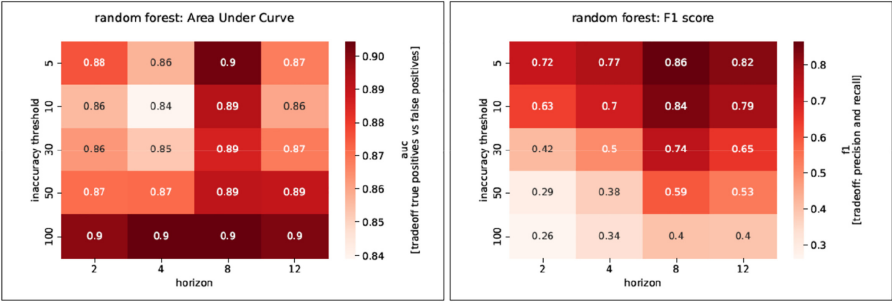


Source(s): Figure created by authors

Figure A5.
Recall and precision
per inaccuracy
threshold and schedule
horizon



Source(s): Figure created by authors



Source(s): Figure created by authors

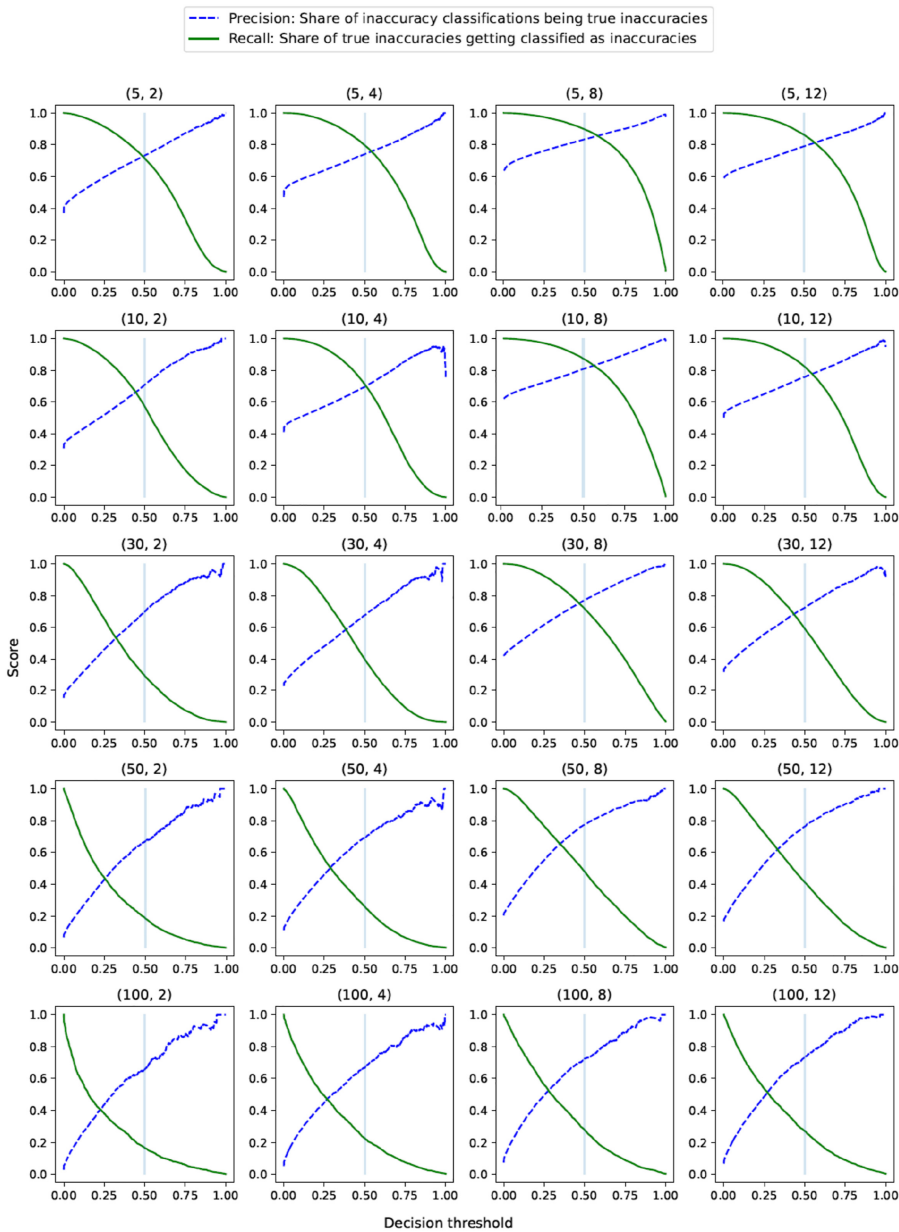


Figure A7.
Precision and recall
with varying decision
threshold

Source(s): Figure created by authors

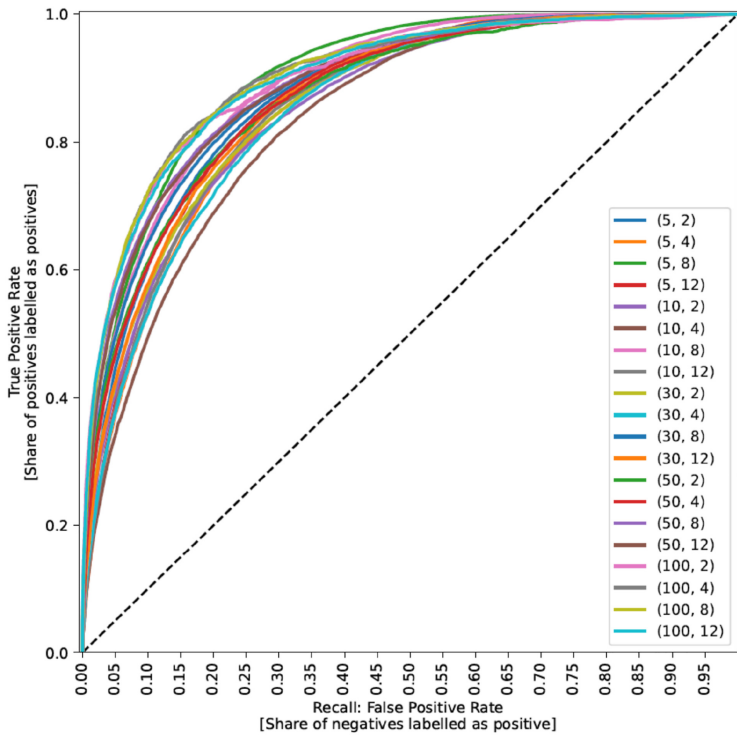


Figure A8.
ROC curves: trade-off
for random forests

Source(s): Figure created by authors

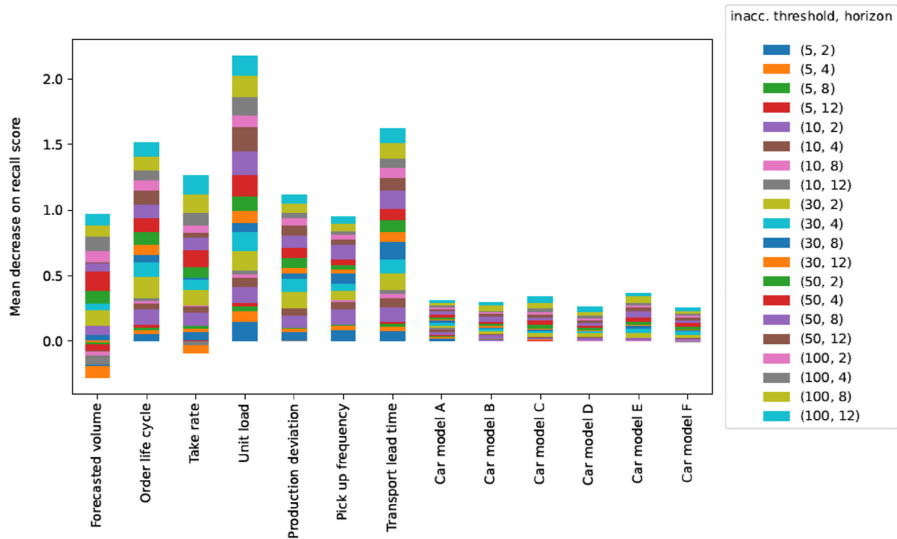


Figure A9.
Variable importance:
the mean effect on
recall score

Source(s): Figure created by authors

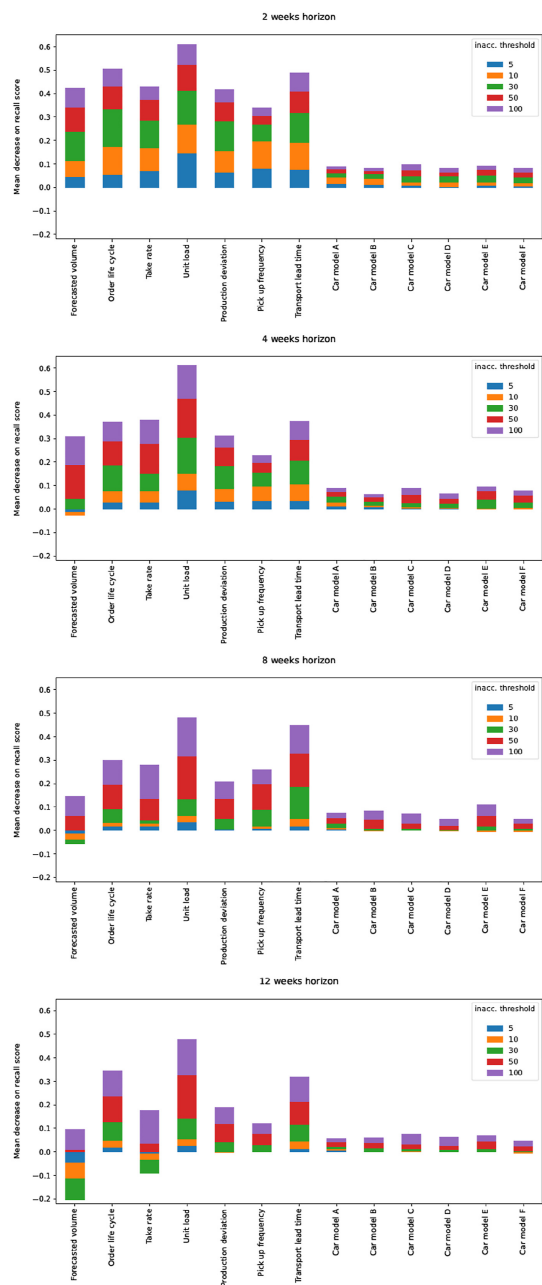


Figure A10.
Variable importance:
mean effects

Source(s): Figure created by authors