This study examines planning in mass customization contexts that face challenges, due to high product variety, sparse configuration-level demand, and long supplier lead times. Traditional Configure-to-Order and Assemble-to-Order (CTO/ATO) planning approaches often rely on late procurement and full postponement, leading to high and unstable customer lead times. To address this problem, a lead-time-first planning approach is developed to translate historical demand information into executable planning decisions without relying on finished-goods inventory. The approach operates across three levels: feature-level Component Readiness Tiering for upstream component pre-positioning, segment-level Mix Guardrails to control demand heterogeneity, and configuration-level Top-K partial pre-kitting to exploit demand concentration while preserving flexibility through postponement. The approach stands out because it directly links demand variability metrics to operational readiness thresholds. This link enables structured staging and coverage-based configuration selection. The approach is evaluated using a synthetic dataset representing one year of demand for customized laptops. Performance is assessed using lead-time-oriented indicators, including the 95th percentile customer lead time and instant-start rate. Results show improved responsiveness, with the worst-case customer lead time reduced from 12 days to approximately 7 days and immediate production enabled for a significant share of orders. These improvements are explained by early readiness of high-demand components combined with postponed final differentiation. The approach suits modular CTO and ATO environments with clear demand segments, stable high-volume components, and regular planning cycles

Keywords: lead-time reduction, multi-level planning, postponement strategy, component readiness, demand variability

UDC 658.7:658.5

DOI: 10.15587/1729-4061.2026.345253

DEVELOPMENT OF A LEAD-TIME-FIRST MULTI-LEVEL PLANNING APPROACH FOR CTO/ATO MASS CUSTOMIZATION SUPPLY CHAINS

Nouhaila El Assad
Corresponding author
Master's Degree in Logistics Engineering*
E-mail: nouhaila.elassad99@gmail.com
ORCID: https://orcid.org/0009-0004-4078-6703

Salah-eddine Mokhlis
Doctor of Electrical Engineering and Automatic Control*
ORCID: https://orcid.org/0009-0008-3107-4295

Kawtar El Haouti
Doctor of Automatic and Electrical Engineering, Professor*
ORCID: https://orcid.org/0000-0002-4309-0718

Najat Messaoudi
Doctor of Automatic and Industrial Computer, University Professor*
ORCID: https://orcid.org/0000-0002-2554-2020

*Department of Physics
Hassan II University
Route d'El Jadida, Km 8, Casablanca, Morocco, 20100

Received 03.12.2025
Received in revised form 26.01.2026
Accepted 10.02.2026
Published 27.02.2026

How to Cite: *El Assad, N., Mokhlis, S., El Haouti, K., Messaoudi, N. (2026). Development of a lead-time-first multi-level planning approach for CTO/ATO mass customization supply chains. Eastern-European Journal of Enterprise Technologies, 1 (3 (139)), 48–60. https://doi.org/10.15587/1729-4061.2026.345253*

1. Introduction

In the recent years, manufacturing supply chains have been operating in an increasingly complex environment characterized by growing demand uncertainty, shorter product life cycles, and stronger pressure to respond rapidly to customer needs. All these characteristics can be traced to the shift from the earlier mass production mode to the mass customization mode. In this new paradigm, the products are more modular, options have multiplied, and customer preferences directly shape what must be sourced, produced, and delivered. While mass production relied on stable product structures, aggregated demand, and large inventory buffers to absorb uncertainty, contemporary supply chains face fragmented and uneven demand patterns that are far more difficult to predict and manage. As a result, planning and execution are no longer routine operational tasks but have become central determinants of supply chain performance, influencing lead times, service levels, resource utilization, and overall competitiveness.

Even though forecasting techniques, inventory models, and planning systems have significantly advanced, many of these tools were originally designed for more stable mass production environments and do not fully capture the realities of modern, highly customized supply chains. So as a consequence, companies continue to struggle to translate demand information into effective operational decisions, thus creating a persistent gap between planning at an aggregate level and actual execution in practice. This unresolved challenge has both practical and scientific importance: practitioners need better approaches to manage complexity and uncertainty, while researchers must develop more suitable models and frameworks for contemporary supply chains.

The challenge becomes even more complex in the context of mass customization [1]. Mass customization aims to strike a balance between efficiency and flexibility, and this contrasts with mass production where companies produce standardized goods in large quantities, or mass personalization where each product is unique [2]. In industries, such as electronics, companies allow consumers to select and integrate a variety of predetermined components, such as the Central Processing Unit (CPU), Graphics Processing Unit (GPU), storage capacity, screen size, and other components into a variety of possible and valid combinations [3]. For example, a company making laptops can offer consumers the option of using either Intel or Advanced Micro Devices (AMD) processors, a variety of GPU options, and varying screen sizes. This approach certainly ensures greater customer satisfaction but creates a forecasting problem where the number of possible combinations of products explodes, while demand for each option remains highly variable [4].

Traditional planning and forecasting approaches are found to be inadequate in this context [5]. These methods are usually effective in scenarios with stable, repetitive demand patterns. However, they do not adequately manage the variability and interdependence of configurable product features [6]. This is a critical challenge in mass customization. In mass customization, it is not just a matter of volume forecasting, rather it is about understanding the structure of demand in terms of product features, customer segments, and full product configurations. Therefore, firms may either overstock unpopular configurations or struggle in timely delivery of popular ones [7].

At the same time, new opportunities are presented by the technologies of Industry 4.0. Typically, big data analytics, machine learning, and real-time customer data collection provide opportunities to overcome this challenge [8, 9]. This is because these technologies allow firms to go beyond aggregate-level demand analysis, and enable them to analyze patterns at multiple levels: the demand for individual features (e.g.; Solid-State Drive (SSD) storage), for customer segments (e.g., students vs professionals), and for complete configurations. Forecasting can then accurately reflect actual customer purchasing behavior and provide to firms useful information for procurement and production planning [10, 11].

Therefore, study on planning approaches that address how multi-level demand information can be translated into effective operational decisions in mass customization supply chains remains a relevant and unresolved scientific topic.

2. Literature review and problem statement

The paper [12] presents a comprehensive review of postponement strategies and their role in improving responsiveness and reducing inventory risk in supply chains. It is shown that postponement and the positioning of the Customer-Order Decoupling Point (CODP) play a critical role in Configure-to-Order (CTO) and Assemble-to-Order (ATO) environments. However, unresolved issues remain related to the translation of these strategic concepts into operational planning decisions. The reason for this may be the predominantly conceptual nature of the reviewed contributions, which focus on strategic positioning rather than on execution-level planning rules applicable to mass customization contexts.

The paper [13] presents an analysis of the strategic positioning of the order penetration point and its impact on efficiency and responsiveness. It is shown that the CODP defines the boundary between forecast-driven and order-driven processes. However, unresolved issues remain regarding how demand information should be used to coordinate planning decisions upstream and downstream of the CODP. The reason for this may be the complexity of linking demand signals to concrete sourcing and staging decisions in high-variety production systems.

The paper [14] presents an integrated view of production and engineering perspectives on the customer order decoupling point. It is shown that better integration across the CODP can improve planning coherence. However, unresolved issues remain related to the lack of operational mechanisms for converting demand signals into executable pre-order and post-order actions. The reason for this may be the difficulty of operationalizing CODP concepts under sparse and heterogeneous demand conditions.

The paper [15] presents an integrated framework for postponement and discusses its strategic and organizational implications. It is shown that postponement can enhance flexibility in supply chains. However, unresolved issues remain regarding which components should be pre-positioned, how finishing capacity should be allocated, and how partial build reassignability can be maintained. The reason for this may be the absence of quantitative decision rules linking postponement strategies to day-to-day planning processes.

The paper [16] presents an analysis of intermittent demand estimation methods and evaluates their forecasting accuracy. It is shown that traditional forecasting approaches perform poorly under intermittent and lumpy demand conditions. However, unresolved issues remain related to how such forecasts should be used in operational planning decisions. The reason for this may be that forecasting research has historically focused on accuracy metrics rather than on execution-oriented outcomes.

The paper [17] presents a critical discussion of intermittent demand forecasting and highlights persistent challenges despite methodological progress. It is shown that intermittent demand remains difficult to forecast reliably in practice. However, unresolved issues remain regarding how planning systems should adapt to such demand patterns in mass-customized supply chains. The reason for this may be the separation between forecasting research and operational planning design.

A way to overcome these difficulties can be the identification of additional structure in sparse demand series. This approach was used in the paper [18], which presents methods for elucidating structural patterns in intermittent demand. It is shown that incorporating such structure can improve demand characterization. However, unresolved issues remain because this approach is not extended to sourcing, staging, or execution-level planning decisions.

Another attempt to bridge forecasting and operations is presented in the paper [19], which links intermittent demand forecasting to inventory obsolescence. It is shown that forecast errors have significant cost implications. However, unresolved issues remain related to broader planning decisions such as component readiness, mix control, and lead-time reduction in CTO/ATO environments.

All this suggests that it is advisable to conduct a study on the development of an operationally usable planning approach for translating multi-level demand signals into coherent planning decisions in CTO/ATO mass customization supply chains. In particular, there is a lack of approaches that combine feature-level demand information, segment-level choice patterns, and configuration-level demand concentration to support execution-oriented planning decisions aimed at lead-time reduction.

3. The aim and objectives of the study

The aim of the study is to develop an approach by which multi-level demand signals, at the feature, customer-segment and configuration levels, can be translated into coherent planning decisions in CTO/ATO mass customization environments. This will allow to reduce customer lead time, increase the Instant-Start Rate (ISR), and limit reliance on expedites while avoiding finished-goods inventory.

To achieve this aim, the following objectives were accomplished:

– to specify the operational policy and planning views required to combine feature, segment and configuration level demand information within a unified planning process;

– to construct and apply Component Readiness Tiering (CRT) based on demand volume and stability and link them to appropriate supplier strategies;

– to develop Segment Mix Guardrails by identifying segment-specific choice patterns and determining their implications for upstream staging and midstream placement;

– to define Top-K partial pre-kits and the associated short-notice finishing capacity while postponing late-differentiating features;

– to ensure cross-level coherence by establishing a multi-horizon planning cadence and assessing its operational behavior using the synthetic dataset.

4. Materials and methods

4. 1. Object and hypothesis of the study

The object of this study is an operations planning system for a Configure-to-Order/Assemble-to-Order (CTO/ATO) laptop plant. The system translates three layers of demand information, features/components, customer segments, and full configurations, into pre-order and post-order decisions designed to address customer lead time while avoiding finished-goods inventory.

The main research hypothesis is that a lead-time-first, multi-level planning approach that is structured around the analysis of historical demand signals at the feature, segment, and configuration levels, can improve responsiveness in CTO/ATO environments by guiding pre-order and post-order decisions.

The study adopts the following assumptions. The Customer-Order Decoupling Point (CODP) is located before final assembly and finishing. Upstream components are modular and interchangeable across configurations. Finishing steps, including operating system imaging, keyboard layout, and packaging, are postponable and short relative to procurement lead times. Component lead times dominate assembly and finishing times. Production capacity is stable within the planning horizon. Customer segments are identifiable, or can be proxied by region or sales channel, and exhibit systematic option preference patterns.

Simplifications were also made to keep the study auditable: a single-year synthetic dataset with monthly resolution is used; segment proportions are balanced; supply disruptions beyond lead time are not modeled; capacity is fixed; costs are not optimized; and returns/cancellations are excluded.

The synthetic dataset used in this study was generated and validated following a structured methodology previously presented in a peer-reviewed publication [20]. The dataset is used in this study as an experimental setup to test the practical impact of the proposed planning approach, which is explained later.

The used synthetic dataset contains all laptop orders for a single calendar year (12 months). Each entry in the dataset corresponds to one sold laptop and has the following attributes: month, customer profile (five segments), CPU type, GPU type, random access memory (RAM), storage capacity, screen size, battery type, keyboard type, operating system (OS), laptop color, touchscreen display, as well as demographic characteristics, such as: age, income, gender, in order to reflect realistic co-variations between segments and choices.

The dataset was generated through a process of sequential sampling from a set of constraints. The profiles of the customer segments were sampled approximately equally across all five segments. The laptop components are sampled from conditional probability tables (CPTs) that link customer segments to component preferences. For example, gamers prefer high Thermal Design Power (TDP) GPU products, students prefer mid-tier CPUs and 14″ display size. The co-occurrence constrains are imposed through a process of resampling combinations of components that were deemed infeasible or unrealistic. The laptop components' options are sampled subsequently to reflect a process of postponing decisions: keyboards, operating systems, colors, touchscreen displays. Demographic attributes, including age and income, were generated using weakly informative distributions, ensuring a certain degree of heterogeneity within each customer profile. The dataset generation process was designated to produce high variety, long-tail effects, stable customer segments' shares, and segment-specific choice maps. No external demand shocks, such as promotions, were included in the dataset generation process.

4. 2. Research design: from demand information to planning decisions

Historical demand information was processed through a process that was structured and rule-based, aiming to convert multi-level demand information into actual planning decisions. This procedure was carried out through three structured and interdependent levels, which are: feature-level readiness, segment-level mix control, and configuration-level partial pre-kitting. The procedure and its levels, were uniformly applied across the 12 months period using predefined rule-based thresholds.

Feature layer: Component Readiness Tiering (*CRT*).

For each component f, annual demand volume V_f and monthly coefficient of variation CV_f are computed over the 12-month horizon.

Input: monthly demand counts for each feature (e.g., CPU model, GPU family, battery class, panel size).

Processing: for each feature f, the following statistics are computed over the annual horizon:

– total annual demand volume V_f;

– monthly coefficient of variation CV_f.

Decision rule: each feature is assigned to one of three readiness tiers:

– Tier 1 (pre-positioned): high-volume, low-variability features

$$V_f \geq P_{75}(V) \text{ and } CV_f \leq 0.35;$$

– Tier 2 (conditional): moderate-volume or moderately variable features

$$P_{25}(V) \leq V_f < P_{75}(V) \text{ or } 0.35 < CV_f \leq 0.60;$$

– Tier 3 (finish-to-order): low-volume and/or high-variability features

$$V_f < P_{25}(V) \text{ or } CV_f > 0.60,$$

where $P_{25}(V)$ and $P_{75}(V)$ denote the 25[th] and 75[th] percentiles of annual demand volumes across all features.

Thresholds are fixed and not tuned to performance outcomes.

Output: a component readiness tiering (CRT) plan specifying the readiness tier assigned to each feature before order arrival.

Segment layer: Mix Guardrails and Placement (*MGP*).

Input: segment-specific monthly choice distributions derived from historical orders.

Processing: for each customer segment *s*, empirical choice shares are computed for key feature families (e.g., CPU class, GPU family, screen size).

Decision rule: segment mix guardrails are defined using percentile-based bounds on observed choice shares:

$$\min_{s,f} = P_{10}(p_{s,f}), \quad (1)$$

$$\max_{s,f} = P_{90}(p_{s,f}), \quad (2)$$

where P_{10} and P_{90} denote the 10th and 90th percentiles of monthly choice shares for feature *f* within segment *s*.

Guardrails are applied uniformly across the planning horizon and are not tuned to performance outcomes.

Output: segment mix guardrails specifying admissible feature proportions for upstream staging and midstream allocation.

Configuration layer: Top-K Partial Pre-Kitting.

Input: monthly frequency of partial configuration signatures defined by core components (CPU, GPU, RAM, Storage).

Processing: partial-kit signatures are ranked by descending frequency for each month.

Decision rule: the value of *K* is selected as a fixed planning policy parameter based on cumulative demand coverage rather than performance optimization. Specifically, *K* is chosen as the smallest integer satisfying

$$\sum_{i=1}^{K} freq_i \geq \alpha \cdot \sum_{j} freq_j, \quad (3)$$

where $freq_i$ denotes the demand frequency of the *i*-th most common partial configuration and α is a predefined coverage target.

In this study, $\alpha = 0.33$ is used as the reference coverage level, corresponding to a practical balance between readiness benefits and pre-kitting workload in CTO/ATO environments. Sensitivity is assessed using additional fixed values of *K* to illustrate diminishing returns, without retuning the coverage target.

Output: a Top-K partial pre-kitting list and its cumulative demand coverage.

Cross-layer coherence and reconciliation.

Input: CRT assignments, segment guardrails, and Top-K partial kits.

Processing: a configuration-to-feature incidence matrix is used to project configuration plans into feature space.

Decision rule:

– Tier 1 component commitments are treated as fixed;

– when violations occur, reconciliation is performed by adjusting Top-K composition;

– Tier 2 readiness quantities are modified only if conflicts persist.

Output:

A coherent planning plan intended to maintain consistency across feature readiness, segment mix, and configuration coverage.

4. 3. Experimental procedure

Prior to analysis, the synthetic dataset was first conditioned to ensure integrity and consistency. Month labels were verified to cover the complete annual horizon (1–12), component and option labels were standardized, and a configuration-to-feature incidence matrix was built. This mapping links each configuration to its underlying features and supports feasibility verification.

Once the dataset was cleaned, demand series were constructed. Transactions were aggregated into monthly counts for each feature (CPU, GPU, RAM, storage, screen size, etc.) and for each customer segment. At the same time, configuration records were reduced to partial-kit signatures defined by CPU, GPU, RAM, and Storage. The frequency of each signature was tabulated monthly, producing a long-tailed frequency distribution used to identify Top-K kits.

The feature layer was then applied by computing Component Readiness Tiering through direct application of the CRT rules. to the aggregated demand series. The CRT rules were applied to derive tier assignments, which were subsequently associated with predefined supplier sourcing strategies.

At the segment layer, monthly choice maps were constructed and transformed into segment-specific mix guardrails following the percentile-based rules. Placement rules were then derived to guide component staging and reservation of localized finishing capacity.

At the configuration layer, partial-kit signatures were ranked by frequency and Top-K pre-kitting lists were generated according to the fixed coverage-based policy. Coverage curves were computed as part of the research procedure and retained for performance evaluation, and finishing capacity requirements were derived based on the expected share of Top-K orders.

After these steps, coherence checks were performed to ensure consistency across layers. Top-K plans were projected back to feature space using the configuration-to-feature mapping, and feasibility was verified against the CRTs. Compliance with segment guardrails was also tested. Where violations appeared, adjustments were made: marginal kits were swapped within the Top-K list, and Tier-2 readiness targets were revisited if needed. Tier-1 commitments remained stable throughout.

Finally, lead-time parameters representative of CTO/ATO environments were specified to support evaluation. Procurement times were set as the dominant contributor, with shorter assembly and finishing times layered on top. These parameters are used as fixed inputs in the comparative analysis.

The procedure generated CRT tables, segment-level guardrails and placement definitions, Top-K pre-kitting lists, coverages curves and coherence diagnostics.

4. 4. Software, equipment, and reproducibility

Analysis was conducted in Python (3.10+) on a standard workstation using pandas, NumPy, SciPy, and matplotlib. Statsmodels supported basic time-series baselines. Random draws used fixed seeds for reproducibility. All thresholds for tiering, guardrails, and K values were predefined and held constant.

4. 5. Performance indicators and evaluation criteria

In order to define an evaluation framework for the proposed planning approach, performance is determined with lead-time-oriented indicators that are commonly used in CTO/ATO environments. The objective is to characterize responsiveness and execution readiness using lead-time-oriented indicators.

The following indicators are defined:

1. P95 customer lead time.

The 95th percentile of customer order lead time, measured from order confirmation to shipment. This indicator captures worst-case service performance and reflects delays caused by component unavailability and upstream constraints.

2. Instant-start rate (ISR).

The Instant-start rate measures the proportion of orders for which production can begin immediately at order release due to the availability of all required Tier-1 components. This indicator reflects the effectiveness of Component Readiness Targets (CRTs) in removing supplier-induced delays.

3. Partial-kit coverage rate (PKC).

The Partial-Kit Coverage Rate represents the share of total orders, which core configuration (CPU, GPU, RAM, Storage) matches one of the Top-K pre-kitted partial builds. This metric quantifies the contribution of partial pre-kitting to lead-time compression under postponement.

4. Fill rate from on-hand components.

The Fill Rate from On-Hand Components is defined as the proportion of required components that can be fulfilled directly from pre-positioned inventory at order release. This indicator reflects the effectiveness of upstream readiness decisions.

5. Expedite dependency (ED, proxy).

Expedite Dependency measures the proportion of orders requiring expedited procurement or exceptional handling due to component unavailability. Lower values indicate better alignment between demand signals and planning decisions.

These indicators are computed at monthly resolution to match the planning cadence of the proposed planning approach. Performance improvements are evaluated by comparing the proposed multi-level planning policy against a baseline CTO/ATO policy.

4. 6. Baseline planning policy

To provide a comparison basis for the proposed lead-time-first, multi-level planning approach, performance is evaluated relative to a baseline planning policy representative of traditional CTO/ATO procedures.

Under the baseline policy:

– no components are pre-positioned based on demand analysis;

– all components are procured after order confirmation following a pure make-to-order logic;

– no segment-based mix guardrails are applied;

– no partial pre-kitting of configurations is performed;

– all product differentiation steps are executed strictly after order arrival.

This means that the production of each order can only begin after all the components needed for its production have been sourced for. This makes the lead time under the baseline schedule dominated by supplier lead times while assembly and finishing are marginal.

Such baseline corresponds to a conventional planning strategy for CTO/ATO with an emphasis on inventory minimization over responsiveness, which is used as a reference point for comparison of lead time and readiness performance under identical lead-time parameters.

4. 7. Specification of lead-time parameters

To support the evaluation of responsiveness under the proposed planning approach, representative lead-time parameters consistent with CTO/ATO manufacturing environments are specified.

Lead time is decomposed into three components:

– procurement lead time (LT_{proc}): time required to source components from suppliers;

– assembly lead time (LT_{assm}): time required for core assembly operations;

– finishing lead time (LT_{fin}): time required for late-differentiation steps (OS imaging, keyboard layout, packaging).

The following values are used uniformly across all experiments:

– $LT_{proc} = 10$ days for components not pre-positioned;

– $LT_{assm} = 1$ day;

– $LT_{fin} = 1$ day.

For orders which required components are fully available at order release (Tier-1 components and/or Top-K partial kits), procurement lead time is bypassed, and customer lead time is given by

$$LT = LT_{assm} + LT_{fin}. \quad (4)$$

For baseline CTO/ATO orders and non-covered configurations, customer lead time is given by

$$LT = LT_{proc} + LT_{assm} + LT_{fin}. \quad (5)$$

These values are not calibrated to a specific firm but are chosen to reflect the dominance of supplier lead times in high-variety CTO/ATO systems. They are applied consistently across the baseline and proposed planning policies to enable relative comparison of lead-time outcomes.

5. Results of the lead-time-first multi-level planning approach

5. 1. Specification of the lead-time-first multi-level planning approach and evaluation context

This study follows a design-oriented research logic in which a rule-based planning approach is formalized and evaluated under controlled experimental conditions. The first result of the study is the conceptual substantiation of a lead-time-first, multi-level planning approach for Configure-to-Order/Assemble-to-Order (CTO/ATO) environments.

The approach is structured around the transformation of multi-level historical demand signals into coordinated planning decisions. It operates across three interdependent layers:

a) feature-level demand patterns define Component Readiness Tiering (CRTs) upstream;

b) segment-level choice distributions impose Mix Guardrails and Placement (MGP) midstream;

c) configuration-level frequency signals drive Top-K partial pre-kitting and short-notice capacity downstream.

Each layer produces a specific planning artifact and addresses a distinct source of lead-time risk in CTO/ATO systems.

The overall logic follows a lead-time-first principle: upstream procurement delays are mitigated through selective component readiness, while final differentiation is postponed to preserve flexibility. This logic can be summarized as "prepare early, finish late", combining early positioning of structurally stable components with deferred execution of volatile attributes.

The structural logic of the Lead-Time-First approach can be summarized as a five-step transformation from demand inputs to coordinated planning artifacts:

Step 1. Input processing and demand signal construction:

a) input: historical demand records (synthetic), including component-level choices, customer segment profiles, and full configurations over 12 months;

b) processing: clean and standardize inputs. Construct three aggregated demand signals:

– feature-level: monthly counts per CPU, GPU, battery, screen, etc.;

– segment-level: monthly choice maps (% of features per customer profile);

– configuration-level: frequency of CPU+GPU+RAM+Storage combinations;

c) output: clean dataset with structured signals by layer.

Step 2. feature layer: component readiness tiering (CRT):

a) input: monthly demand counts per feature;

b) logic: calculate annual volume and coefficient of variation (CV) for each component;

c) decision rule:

– Tier 1: high volume, low CV → pre-positioned;

– Tier 2: moderate volume/CV → staged based on short-term signal;

– Tier 3: low volume/high CV → ordered after demand;

d) output: CRT plan = assignment of features to Tier 1, 2, or 3;

e) supplier strategy linked:

– Tier 1 = vendor-managed inventory (VMI) or stock-based procurement;

– Tier 2 = flexible contracts;

– Tier 3 = pull-based.

Step 3. Segment layer: mix guardrails and placement:

a) Input: Segment-specific choice maps (% features per segment);

b) Logic: Identify typical share ranges per component family in each segment;

c) Decision rule:

– define guardrails: e.g., [min%, max%] for each component per segment:

– use guardrails to guide:

1) upstream staging proportions;

2) regional placement;

3) finishing capacity reservation;

d) output: segment-specific mix constraints and placement guidance.

Step 4. Configuration layer: Top-K Partial Pre-Kitting:

a) input: frequency-ranked partial configurations (CPU+GPU+RAM+Storage);

b) logic: rank partial-kit signatures by frequency;

c) decision rule:

– choose a value of K (e.g., 100);

– select top-K most common partial kits;

– book short-notice capacity for those kits;

d) output: Top-K kit list with cumulative coverage curve;

e) postponement applied: final differentiation steps (screen, OS, battery) are deferred.

Step 5. Cross-layer coherence and reconciliation:

a) input: CRTs, Guardrails, Top-K list;

b) logic: ensure consistency between layers:

c) project Top-K kits into feature space;

d) verify coverage within Tier-1/2 limits and segment guardrails;

e) if violations → adjust:

– Top-K kit composition (swap marginal signatures);

– Tier-2 readiness quantities;

f) output: coherent, executable plan across layers.

This diagram is illustrated in the Fig. 1 below.

Fig. 1. summarizes the sequence of steps used in the proposed lead-time-first, multi-level planning approach. It shows how historical demand are processed and progressively transformed into planning artifacts, starting from demand signal construction and continuing through component readiness tiering, segment-level mix definition, partial pre-kitting, and cross-layer coherence checks. The figure also indicates the main inputs and outputs at each step and how information is passed between layers to form an executable planning plan.

The approach produces four executable planning artifacts:

– component readiness tiering (CRT);

– segment mix guardrails and placement rules;

– Top-K partial pre-kitting lists;

– a reconciliation-consistent planning cadence.

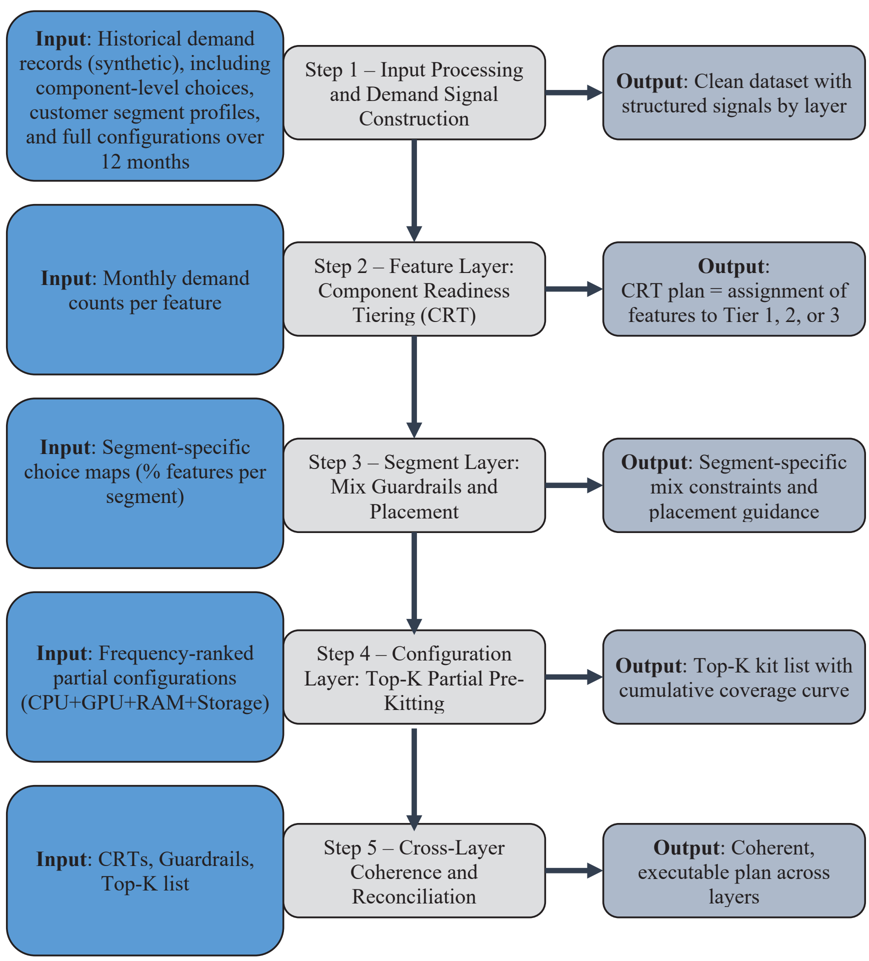

Fig. 1. Step-by-step implementation of the proposed lead-time-first, multi-level planning approach from demand inputs to executable planning decisions

The demand signals used in this study derive from a synthetic dataset with historical structure and serve as a proxy for forecasted inputs. The planning logic relies on their statistical profiles, for example: volume, variability, preference distribution, in order to construct decision rules across the three layers. The overall planning logic is summarized as "prepare early, finish late": pre-position critical components upstream while postponing final differentiation to preserve flexibility [7, 12].

This conceptual formalization establishes the structural foundation for the experimental evaluation presented in the following subsections.

The planning approach was evaluated under a fixed lead-time-first policy "prepare early, finish late", using four lead-time-oriented indicators, where the instant-start effect is captured through the fill rate from on-hand components: P95 customer lead time, fill rate from on-hand components, Top-K partial coverage, and expedites. The synthetic dataset provided 12 full months of orders in 2024, balanced across five customer profiles with stable segment shares. This stability makes the evaluation conservative: any speed gains arise from planning structure rather than easy seasonal shifts.

Table 1 reports the average, minimum, and maximum shares of each customer segment over the 2024 evaluation horizon.

Table 1

Customer segment shares in 2024 (average, minimum, and maximum)

Segment	Avg share	Min	Max
Business executive	20.04%	19.61%	20.26%
Gaming enthusiast	20.04%	19.68%	20.31%
General consumer	19.98%	19.56%	20.30%
Student	19.99%	19.73%	20.21%
Tech professional	19.95%	19.63%	20.30%

The steadiness of shares justifies stable guardrails and highlights that any lead-time gains observed later are due to planning logic, not lucky seasonality.

Fig. 2 presents the monthly distribution of customer segment shares over the 2024 evaluation horizon.

Fig. 2 shows the monthly share of each customer segment over 12 months. Shares remain close to 20%, with only small variations across the year. No segment shows strong growth or decline.

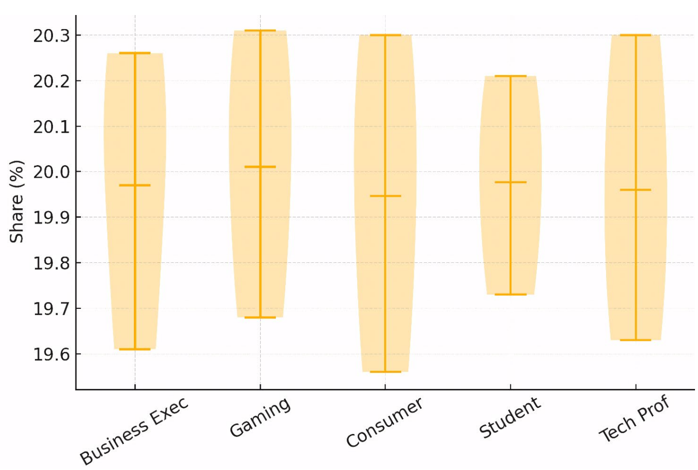

Fig. 2. Monthly distribution of customer segment shares across the evaluation horizon

This confirms that the dataset preserves the intended segment balance. Because demand remains stable, segment-level tools such as Mix Guardrails can be assessed without structural bias. The stable distribution also supports the claim that performance gains result from the planning approach, not from favorable demand conditions.

The results further show that feature-level, segment-level and configuration-level views were applied under the same lead-time-first policy. The same cadence and indicators guided all levels, so readiness decisions, mix limits, and pre-kitting rules operated within a single, coherent planning framework.

5. 2. Feature-level readiness: construction and application of component readiness tiering (CRT)

This subsection reports the results of applying the component readiness tiering rules to the feature-level demand data.

Applying the tiering rule revealed a small but reliable set of Tier-1 components that can be pre-positioned upstream. At the CPU level, three models dominated with high annual volumes and low variability; at the GPU level, integrated graphics and two mainstream discrete families formed the Tier-1 core.

Tables 2, 3 list the Tier-1 components identified through the CRT rules for CPUs and GPUs, respectively. Table 4 then summarizes the resulting instant-start performance.

Table 2

Tier-1 CPU components identified by annual demand volume and coefficient of variation

CPU	Total units	CV
Intel Core i5-1135G7	98,074	0.301
Intel Core i7-1185G7	90,579	0.295
Intel Core i9-12900K	68,869	0.298

Table 3

Tier-1 GPU components identified by annual demand volume and coefficient of variation

GPU	Total units	CV
Integrated graphics (Intel UHD)	75,394	0.294
NVIDIA RTX 3060	46,152	0.298
NVIDIA GTX 1650	44,687	0.301

Table 4

Proportion of orders starting immediately from on-hand components (average, minimum, and maximum)

Metric	Value
Average across months	19.73%
Range (min → max)	19.48% → 20.17%

Staging only these Tier-1 items already provides a tangible benefit: about 20% of orders can start instantly from on-hand CPUs and GPUs.

Fig. 3 presents the classification of selected components into readiness tiers.

Fig. 3 presents the full classification of components into Tier-1, Tier-2, and Tier-3 categories based on predefined volume and variability thresholds.

The visualization confirms that only a limited subset of components satisfies Tier-1 criteria, while the majority remain conditional or finish-to-order items.

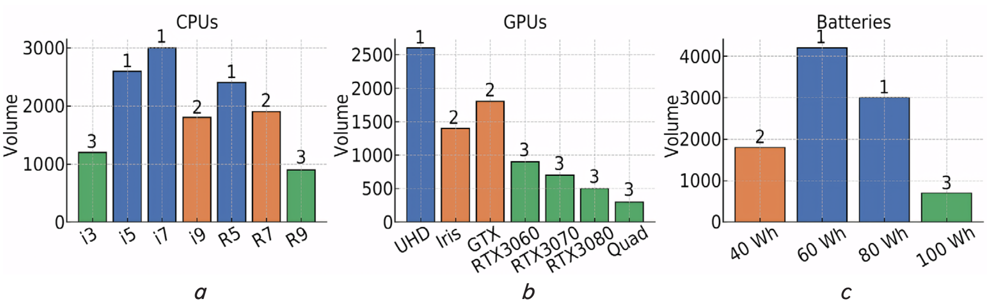

Fig. 3. Classification of selected components into readiness tiers:
a — central processing unit components; *b* — graphics processing unit components; *c* — battery components;
— Tier-1: Pre-position (high volume, low variability); — Tier-2: Conditional (moderate demand / variability);
— Tier-3: Finish-to-order (low volume, high variability)

This distribution reflects the structural concentration of demand within selected component families.

This effect is predictable, with a narrow month-to-month range (≈ 19.5% – 20.2%), providing a stable baseline of instant-start capability, as illustrated in Fig. 4.

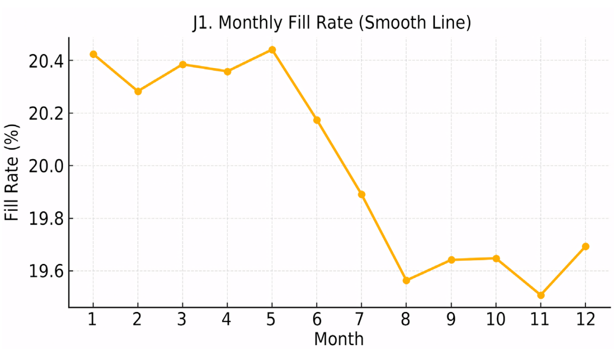

Fig. 4. Month-to-month stability of the fill rate achievement from Tier-1 component readiness

Fig. 4 illustrates the month-to-month stability of the fill rate achieved from pre-positioned Tier-1 components. The curve remains almost flat throughout the year, fluctuating within a very narrow band around 20%. This stability confirms that the Component Readiness Tiering (CRTs) were correctly calibrated: upstream staging consistently covers a predictable portion of demand, regardless of minor monthly variations. The absence of abrupt drops indicates that no structural shortages occurred, while the small oscillations reflect normal demand noise rather than planning failures. Overall, the figure demonstrates that Tier-1 component readiness provides a reliable and repeatable baseline of instant availability across the entire horizon, which is a prerequisite for lead-time reduction in CTO/ATO environments.

These tier assignments also define the sourcing logic. Tier-1 components follow stock-based or vendor-managed arrangements. Tier-2 components rely on flexible supply contracts. Tier-3 components are procured strictly on an order-driven basis.

5. 3. Segment-level mix control: development of segment mix guardrails and placement rules

This subsection reports the results of applying segment-level choice analysis to derive mix guardrails and placement guidance.

The second layer used choice-maps to differentiate profiles. Table 5 illustrates the contrast: Gamers leaned heavily toward RTX-class GPUs, Students concentrated in integrated and entry-level options, while Executives and Tech Professionals favored workstation-grade parts.

Because segment shares were steady, these guardrails could be applied consistently month after month, allowing pre-booking of finishing capacity without risk.

While Table 5 highlights dominant GPU options, Fig. 5 presents the full distribution of GPU selections across segments in absolute terms.

Fig. 6 shows how graphics processing unit choices differ across customer segments. The figure highlights clear and consistent preference patterns between business executives, gaming enthusiasts, general consumers, students, and tech professionals, which form the basis for defining segment-level mix guardrails.

Fig. 6 reveals clear and lasting differences in GPU preferences across customer segments. These patterns remain stable, which supports the use of segment-specific mix guardrails.

By setting guardrails at the segment level, upstream staging proportions and midstream placement decisions can be managed with greater accuracy. This structure also allows the reservation of finishing capacity more reliably and lowers the risk of producing the wrong product mix.

Table 5

Top three GPU choices by customer segment with corresponding demand shares

Segment	#1 GPU (share)	#2 GPU (share)	#3 GPU (share)
Business Executive	Integrated UHD (22.9%)	Intel Iris Plus (16.8%)	AMD Radeon Pro (15.6%)
Gaming Enthusiast	NVIDIA RTX 3060 (18.5%)	RTX 3070 (15.1%)	RTX 3080 (14.8%)
General Consumer	Integrated UHD (25.6%)	GTX 1650 (16.9%)	Intel Iris Plus (13.8%)
Student	Integrated UHD (36.9%)	Intel Iris Plus (23.7%)	GTX 1650 (9.3%)
Tech Professional	NVIDIA Quadro (17.4%)	RTX 3070 (14.0%)	RTX 3060 (10.5%)

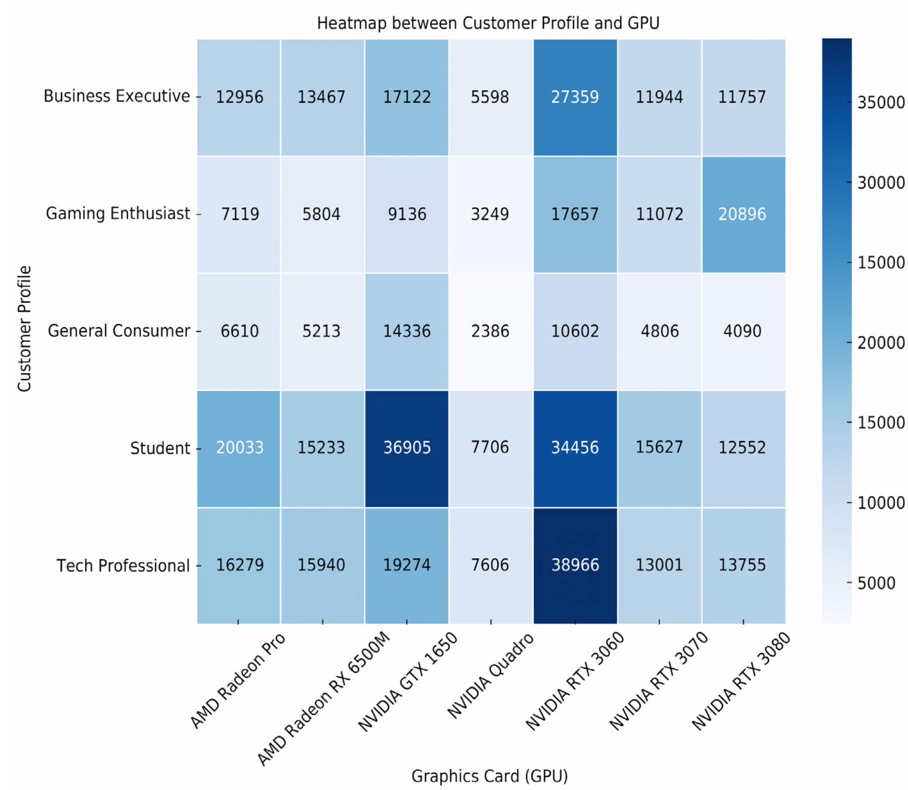

Fig. 5. Segment-specific Graphics Processing Unit choice patterns used to derive percentile-based mix guardrails

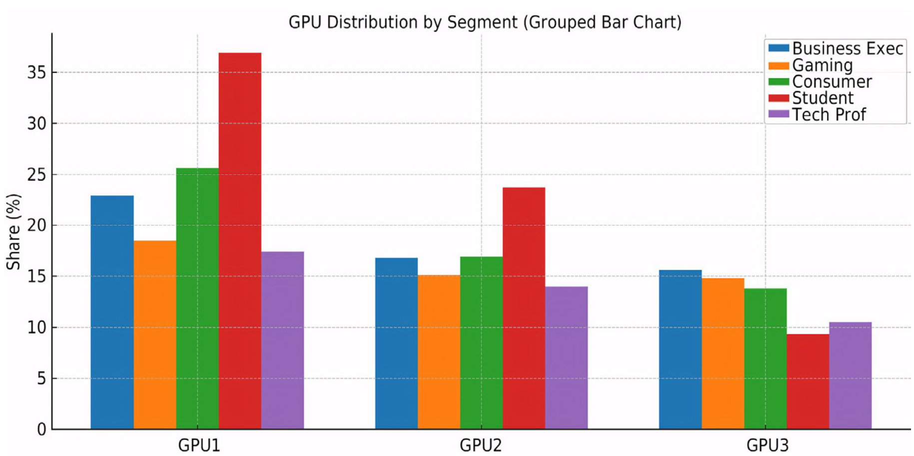

Fig. 6. Segment-specific distribution of Graphics processing unit families illustrating structural differences in component preferences

Table 6

Coverage of customer orders achieved by Top-K partial pre-kitted configurations

K (ready partial kits)	Average	Range (min $\to$ max)
50	20.73%	20.47% $\to$ 21.11%
100	33.20%	32.82% $\to$ 33.65%
200	49.40%	49.08% $\to$ 49.91%

The more than tenfold gap between partial and strict readiness, about 33 percent versus 2 to 3 percent, shows the impact of postponement.

5. 4. Configuration-level readiness: definition of Top-K partial pre-kits and finishing capacity allocation

At the configuration level, the planning approach focused on pre-kitting partial builds of CPU+GPU+RAM+Storage. Ranking signatures by frequency produced coverage curves that rise quickly and then flatten.

Table 6 reports the coverage of customer orders achieved by Top-K partial pre-kitted configurations for different values of K.

Holding 100 partial kits ($K = 100$), selected according to the predefined coverage policy, covers about one-third of orders at modest effort; doubling to 200 kits pushes coverage close to half but increases workload.

Table 7 reports the coverage of customer orders under strict full-configuration readiness.

Table 7

Coverage of customer orders under strict full-configuration readiness

Metric	Value
Average across months	2.34%
Range (min $\to$ max)	2.24% $\to$ 2.53%

This result reflects the operational logic of partial pre-kitting. Core components (CPU, GPU, RAM, Storage) are assembled in advance, while late-differentiating features (screen, battery, operating system, keyboard, and color) remain order-driven.

Short-notice finishing capacity is reserved in proportion to Top-K coverage. This approach allows to complete the final configuration after order confirmation, without committing too much capacity in advance.

5. 5. Cross-layer coherence and operational behavior assessment under the multi-horizon planning cadence

Projecting Top-K plans back into feature space showed no shortages under Tier-1 CRTs. When mixes leaned toward a single GPU, substituting a few marginal kits restored balance without reducing coverage. Segment guardrails were respected, with deviations in gamer-heavy months remaining within ±2 percentage points.

The monthly/weekly/daily rhythm proved sufficient: monthly CRTs and placement defined upstream bounds; weekly refreshes adjusted guardrails and Top-K lists; and daily execution released ready kits while non-Top-K orders started from staged components. No emergency replanning was required over the evaluation horizon.

These coherence results indicate that the planning approach's speed gains are operationally sustainable under routine execution, not merely theoretical.

Fig. 7 below compares the P95 customer lead time under the baseline planning policy and the proposed planning approach using identical lead-time parameters.

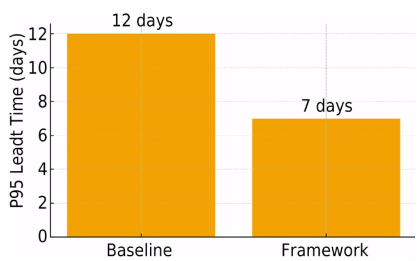

Fig. 7. Comparison of the 95th percentile customer lead time under the baseline Configure-to-Order and Assemble-to-Order policy and the proposed planning approach, using identical lead-time parameters

The comparison shows a clear drop in P95 lead time, from the 12 days under the baseline approach to 7 days with the proposed planning approach. This five-day reduction reflects the combined effect of upstream component readiness, segment-aligned mix discipline, and partial pre-kitting of Top-K configurations.

Together, these planning elements remove the main sources of long delays and allow faster, more consistent order fulfillment. This reduction in worst-case customer lead time is achieved without relying on finished-goods inventory.

These findings also confirm that the multi-horizon planning cadence works coherently across layers. When applied to the synthetic dataset, its operational logic translates into measurable lead-time gains.

6. Discussion of multi-level planning results in CTO/ATO environments

The first result obtained in this study is the formalization of a lead-time-first, multi-level planning approach, which has been specifically designed to address the requirements of mass customization in the context of CTO/ATO. Its distinguishing feature of this approach is the explicit reconfiguration of the planning priorities to achieve the lead-time dominance of procurement and the concentration of structural demand across three coordinated layers: feature readiness, segment mix discipline, and configuration concentration. Contrary to the conventional CTO/ATO planning strategies, which are based on full postponement or aggregated demand treatment, the proposed approach reconfigures the decision priorities around the structural variability and empirical stability of the demand. Its defining features lie in the combination of the following components within a unified decision structure: of variability-based component readiness, percentile-bounded segment mix discipline, and coverage-based configuration pre-kitting within a single coherent decision architecture. Although the approach does not exclude postponement, it redefines it in the sense of introducing upstream readiness for stable components while maintaining the typical postponement strategy for volatile components. The combination of these characteristics differentiates the proposed approach from both traditional make-to-order logic and forecasting-based planning methods.

The results demonstrate a quantifiable and structural improvement in responsiveness under equivalent lead time conditions. The reduction in 95th percentile customer lead time from 12 days under the baseline CTO/ATO policy to approximately 7 days under the proposed approach, as shown in Fig. 7, is a deliberate improvement. This improvement is a direct result of the translation of the multi-level demand structure into operational readiness decisions through the mechanisms of Component Readiness Tiering, Segment Mix Guardrails, and Top-K partial pre-kitting, which have been previously defined.

Firstly, the feature-level analysis in Tables 2–4 and Fig. 5 suggests that a small proportion of high-volume, low-variability components, identified using predefined percentile and coefficient-of-variation thresholds, enables approximately 20% of orders to be dispatched immediately at the point of order release. Given the procurement lead time of 10 days comprises a dominant fraction of the total lead time in the experimental configuration, the elimination of the procurement process for Tier-1 components eliminates the main contributor to the worst-case delay. Moreover, the limited month-to-month variation in the fill rate, between ≈ 19.5% and 20.2%, confirms that the observed effects are not due to seasonal effects but to underlying patterns in the demand concentration. These observations provide empirical support to the conceptual proposition that the role of variability measures such as the annual demand volume V_f and coefficient of variation CV_f is not limited to the domain of statistics but extends to the domain of operations.

Second, insights from the segment-level results shown in Table 5 and Fig. 6 provide an understanding of why mix guardrails are needed even though aggregate segment shares appear stable from the results shown in Table 1 and Fig. 4. Although each segment has an equal 20% of the total demand distribution, the GPU preference for each segment varies significantly. For example, Gaming Enthusiasts prefer RTX-class products exclusively, whereas Students prefer integrated graphics. Without any guardrails based on percentiles defined a priori, upstream planning might result in an aggregate solution that averages across all segments but does not necessarily represent the mix composition of those segments, potentially resulting in delays related to suboptimal mix decisions. The ability of the planning approach to constrain feature proportions within the empirical range of P10 to P90 ensures that such systematic mismatches between inventory positioning and segment-level demand behavior are avoided. The fact that such mismatches are limited to within ±2 percentage points during cross-layer reconciliation serves to reinforce that guardrails are used as stabilizing constraints rather than limiting assumptions.

Third, configuration-level results provided in Tables 6, 7 help to clarify the importance of partial readiness within a postponement logic. As seen, a selection of $K = 100$ partial kits, based on a predetermined cumulative coverage rule, results in a 33% order coverage, whereas a rigorous level of full-configuration readiness accounts for only 2–3% of orders. This greater than ten-fold difference in coverage underscores the idea that deferring late-differentiating features such as screen, OS, battery, keyboard, and color, in favor of pre-assembling key components such as CPU, GPU, RAM, and storage, capitalizes on demand concentration without creating unnecessary SKU proliferation. The diminishing returns experienced by increasing K from 100 to 200 help to corroborate that readiness benefits are nonlinear, and a coverage-based selection of pre-kitted configurations is, in fact, more beneficial than a rigorous and exhaustive configuration preparation approach. The practical implication is that readiness must be focused where structural frequency dictates, a notion that is conceptually aligned with postponement theory but is quantified at a relatively rare level of execution.

Collectively, these three layers explain how a five-day decrease in P95 lead time is achieved. Tier-1 readiness eliminates exposure to procurement variability for stable components, guardrails avoid mix-related shortages, and Top-K partial pre-kitting encodes concentrated configuration demand. As procurement lead time is the most significant element in the baseline formula $LT = LT_{proc} + LT_{assm} + LT_{fin}$, any systematic avoidance of LT_{proc} directly impacts the upper percentile of the lead time distribution. Hence, the results validate the a priori hypothesis that multi-level demand analysis can decrease responsiveness risk independently from finished goods inventory.

Compared to previous literature, this work makes a methodological, not conceptual, contribution. Research on postponement and CODP positioning [12–15] is foundational but does not provide quantified rules for readiness thresholds and configuration coverage. This work, in turn, extends previous work by quantitatively relating variability metrics (V_f, CV_f) and choice distributions to predefined readiness tiers and percentile-based rules. Similarly, intermittent demand forecasting [16–18] is concerned with improving accuracy, not with quantifying rules that influence execution-level outcomes. Hence, unlike previous approaches, which operate at a strategic and forecasting level, the proposed method quantitatively relates variability, heterogeneity, and frequency to a planning artifact with lead time impact.

These findings fill the already mentioned research gap since they provide an approach at the execution level to effectively transform multi-level demand information into a cohesive CTO/ATO planning decisions. The CRT rules reflect the encoding mechanism for upstream readiness; guardrails reflect the mechanism in place for midstream allocation; the Top-K mechanism represents downstream preparation; and finally, there is cross-layer reconciliation through configuration-to-feature incidence mapping. The findings from the empirical analysis provide a sense of coherence in the already mentioned mechanisms under a relatively simple monthly/weekly/daily cadence without emergency replanning. This study thus demonstrates the ability to transform multi-level demand information into a cohesive plan rather than an abstract concept.

In conclusion, the study demonstrates the ability to effectively utilize a multi-level lead-time-first approach in order to realize a significant portion of the responsiveness inherent in make-to-stock systems while maintaining the flexibility of make-to-order systems. This approach thus represents a practical, scalable, and demanding solution to the already mentioned unresolved issues in mass customization.

While there are a number of strengths in this study, there are a few areas that need to be pointed out. For instance, the study only utilized a single synthetic dataset in which the proportion of the segments remains relatively stable. In addition, supplier lead times are assumed to remain fixed. While the dataset was created through a series of structured plausibility constraints [20], there are a number of issues that could have been experienced in a real-world scenario. In addition, the readiness thresholds as well as the coverage parameter α are predefined rather than optimized. This could have a negative impact in terms of flexibility in the event of extreme volatility. Finally, there are no cost considerations in terms of inventory holding versus lead time reduction.

These limitations also suggest the direction for further work. Looking forward, the planning approach can be developed in several ways. First, incorporating dynamic forecasting using statistics or machine learning can enable to periodically adjust readiness tiers and coverage thresholds under evolving demand. Second, incorporating cost-based criteria can enable to simultaneously optimize speed and inventory risk, transforming fixed thresholds into balanced decision boundaries. Third, extending the model to multi-echelon supply networks can require to align readiness across suppliers, regional distribution centers, and final assembly sites. Fourth, testing robustness can validate how well cross-layer alignment can handle changing conditions. In summary, these extensions can develop the current framework from a tightly controlled execution model into a scalable decision support model for the industrial sector.

7. Conclusion

1. The study provides a unified operational planning structure that combines the demand signals of feature, segment, and configuration layers of components into a coherent multi-layer planning process. It demonstrates how multi-level demand data can be combined to form a consistent operational policy for CTO/ATO environments.

2. The study also demonstrates how a small set of components, which have high volume and low variability, can be pre-positioned upstream for reliable results. The Component Readiness Tiering approach identified Tier-1 components such as CPUs and GPUs, which can be used for immediate start of production for approximately 20% of customer orders. This component readiness is consistent across the months, with small variations, which indicate structural gains. This is unlike the traditional CTO/ATO approach of delaying procurement of components until orders are received. In this approach, lead-time risks of components are eliminated without increasing the inventory risks of finished goods. As the tiers are based on volume and variability of components, they naturally lead to differentiated approaches for components such as stock-based components or vendor-managed

arrangements for Tier-1 components, flexible supply contracts for Tier-2 components, and order-driven components for Tier-3 components.

3. On the segment level, the analysis shows that customer profile patterns are evident and sustained for component selection, even if overall demand shares appear to be steady. This provides the basis for mix guardrails for each segment. This provides guidance for how to stage and place inventory by considering the specific segment-level choice distributions. This approach helps to avoid wrong-mix situations, which are cases where inventory exists but does not match actual demand. By considering segment-level differences rather than relying on aggregate demand alone, the linkage between staging and how each segment actually behaves is improving. The deviations in segment-level mix over the evaluation period remained within a tight band of ± 2 percentage points, which confirms the operational stability of the guardrails.

4. At the configuration level, the Top-K Partial Pre-Kitting strategy demonstrates a non-linear payoff. While making arrangements for 100 partial configurations in the pre-kitted state covers about one-third of the orders, striving for full configuration readiness covers merely 2–3%. This indicates the effectiveness of the postponement strategy. By using the method, it is possible to make good progress on the configuration readiness front without increasing the number of SKUs or making commitments. Increasing the configurations covered by the pre-kitting process also shows diminishing returns, which indicates the effectiveness of the approach. However, the finishing capacity also matches the Top-K configurations, so the final customization can be done quickly.

5. Based on the results, it is evident that there is excellent coherence between the different layers. Feature readiness targets, segment guardrails, and Top-K configuration plans remained consistent within a simple monthly, weekly, and daily planning cadence. The deviations stayed within a range of ± 2 percentage points, and there were no emergency replanning. Based on these findings, it appears that the proposed planning approach is operationally effective and does not require constant re-optimization or emergency interventions, in contrast of many heuristic-based approaches used in highly varied environments. When evaluated against a baseline CTO and ATO planning policy, the integrated effect of these planning elements yields a substantial reduction in worst-case customer lead time. The 95^{th} percentile lead time decreases from 12 days under the baseline to approximately 7 days under the proposed approach. This improvement is achieved without relying on finished-goods inventory and explained by the systematic removal of upstream supplier delays, improved mix alignment, and selective partial readiness.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this study, whether financial, personal, authorship or otherwise, that could affect the study and its results presented in this paper.

Financing

The study was performed without financial support.

Data availability

Data will be made available on reasonable request.

Use of artificial intelligence

The authors confirm that no artificial intelligence tools were used to generate or write any substantive part of this manuscript.

Artificial Intelligence tools were used only for basic grammar and spelling checks without modifying the scientific content or structure of the text.

Artificial intelligence tools were not used to generate or modify the study design, methodology, data, results, discussion, or conclusions. All data, analyses, figures, and interpretations were produced and verified by the authors.

All AI-assisted outputs were carefully reviewed and validated by the authors, who take full responsibility for the content of the manuscript. The use of AI tools did not influence the scientific conclusions of the study.

Authors' contributions

Nouhaila El Assad: Conceptualization, Methodology, Formal analysis, Data Curation, Validation, Writing – original draft, Visualization; **Salah-eddine Mokhlis**: Software, Formal analysis, Data curation; **Kawtar El Haouti:** Writing – review & editing, Validation; **Najat Messaoudi**: Supervision, Project administration, Writing – review & editing.

References

1. Wang, N. (2021). Retracted on February 24, 2022: Mass Customization Capabilities: Literature Review. The Sixth International Conference on Information Management and Technology, 1–5. https://doi.org/10.1145/3465631.3465728
2. Oliveira, J. M., Ramos, P. (2019). Assessing the Performance of Hierarchical Forecasting Methods on the Retail Sector. Entropy, 21 (4), 436. https://doi.org/10.3390/e21040436
3. Salvador, F., De Holan, P. M., Salvador, B. Y. F., Holan, P. M. D. E., Piller, F. (2009). Cracking the Code of Mass Customization. MIT Sloan Management Review, 50 (3). Available at: https://www.researchgate.net/publication/265498057_Cracking_the_Code_of_Mass_Customization
4. Fogliatto, F. S., da Silveira, G. J. C., Borenstein, D. (2012). The mass customization decade: An updated review of the literature. International Journal of Production Economics, 138 (1), 14–25. https://doi.org/10.1016/j.ijpe.2012.03.002
5. Gunasekaran, A. (2007). Build-to-order supply chain management. International Journal of Operations & Production Management, 27 (11). https://doi.org/10.1108/ijopm.2007.02427kaa.001

6. Kiefer, D., Grimm, F., Bauer, M., Van, D. (2021). Demand Forecasting Intermittent and Lumpy Time Series: Comparing Statistical, Machine Learning and Deep Learning Methods. Proceedings of the 54th Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2021.172
7. Aviv, Y., Federgruen, A. (2001). Design for Postponement: A Comprehensive Characterization of Its Benefits Under Unknown Demand Distributions. Operations Research, 49 (4), 578–598. https://doi.org/10.1287/opre.49.4.578.11229
8. El Assad, N., Dachry, A., Fouraiji, H., Dachry, W., Messaoudi, N. (2025). Smart manufacturing paradigms in the context of industry 4.0: Bibliometric analysis. E3S Web of Conferences, 601, 00035. https://doi.org/10.1051/e3sconf/202560100035
9. Raza, A., Haouari, L., Pero, M., Absi, N. (2018). Impacts of Industry 4.0 on the Specific Case of Mass Customization Through Modeling and Simulation Approach. Customization 4.0, 217–234. https://doi.org/10.1007/978-3-319-77556-2_14
10. Scholz-Reiter, B., Kück, M., Lappe, D. (2014). Prediction of customer demands for production planning – Automated selection and configuration of suitable prediction methods. CIRP Annals, 63 (1), 417–420. https://doi.org/10.1016/j.cirp.2014.03.106
11. Vithitsoontorn, C., Chongstitvatana, P. (2022). Demand Forecasting in Production Planning for Dairy Products Using Machine Learning and Statistical Method. 2022 International Electrical Engineering Congress (IEECON), 1–4. https://doi.org/10.1109/ieecon53204.2022.9741683
12. van Hoek, R. I. (2001). The rediscovery of postponement a literature review and directions for research. Journal of Operations Management, 19 (2), 161–184. https://doi.org/10.1016/s0272-6963(00)00057-7
13. Olhager, J. (2003). Strategic positioning of the order penetration point. International Journal of Production Economics, 85 (3), 319–329. https://doi.org/10.1016/s0925-5273(03)00119-1
14. Wikner, J., Rudberg, M. (2005). Integrating production and engineering perspectives on the customer order decoupling point. International Journal of Operations & Production Management, 25 (7), 623–641. https://doi.org/10.1108/01443570510605072
15. Yang, B., Burns, N. D., Backhouse, C. J. (2004). Postponement: a review and an integrated framework. International Journal of Operations & Production Management, 24 (5), 468–487. https://doi.org/10.1108/01443570410532542
16. Syntetos, A. A., Boylan, J. E. (2005). The accuracy of intermittent demand estimates. International Journal of Forecasting, 21 (2), 303–314. https://doi.org/10.1016/j.ijforecast.2004.10.001
17. Nikolopoulos, K. (2021). We need to talk about intermittent demand forecasting. European Journal of Operational Research, 291 (2), 549–559. https://doi.org/10.1016/j.ejor.2019.12.046
18. Kourentzes, N., Athanasopoulos, G. (2021). Elucidate structure in intermittent demand series. European Journal of Operational Research, 288 (1), 141–152. https://doi.org/10.1016/j.ejor.2020.05.046
19. Teunter, R. H., Syntetos, A. A., Zied Babai, M. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. European Journal of Operational Research, 214 (3), 606–615. https://doi.org/10.1016/j.ejor.2011.05.018
20. El Assad, N., Mokhlis, S.-E., Messaoudi, N. (2025). Generation and Validation of a Synthetic Dataset for Demand Modelling in Mass Customization Using Artificial Intelligence Tools. Connected Objects, Artificial Intelligence, Telecommunications and Electronics Engineering, 529–535. https://doi.org/10.1007/978-3-032-01536-5_80