

Review Article

Building Scalable ERP Modules for Inventory Forecasting and Material Planning

Nisarg Dilipkumar Patel

Northeastern University, Boston, MA

Received Date: 10 June 2026

Revised Date: 15 June 2026

Accepted Date: 25 June 2026

Abstract: ERP systems continue to play a central role in coordinating operations, but inventory forecasting and material-planning capabilities are often implemented more rigidly than modern analytical demands require. This has become more important as demand volatility, product proliferation, shorter planning cycles, and exposure to disruption have increased the need for high-frequency forecasting and closed-loop planning. To clarify how scalable ERP modules for inventory forecasting and material planning can be designed and governed, this review examines the current journal literature on forecasting methods, data-driven planning, reconfigurable supply networks and data-intensive operations management. The review identifies five recurring themes: modular separation of transaction processing and analytics, event-driven orchestration of data, multi-horizon forecasting, feedback-rich planning loops, and the enduring significance of data quality and model interpretability. Forecast accuracy alone does not guarantee planning value unless forecasts are structurally reconciled with bills of material, replenishment logic, lead-time constraints, and exception-processing rules. Significant omissions are a shortage of first-hand evidence of ERP-native applications, a paucity of coverage of computational scalability, little coverage of master-data governance, and disjointed assessment metrics among forecasting and planning research. This area is important because scalable ERP modules increasingly determine whether firms can convert analytical insight into timely material decisions under operational uncertainty.

Keywords: Demand Forecasting, Enterprise Resource Planning, Material Requirements Planning, Microservices Scalability, Inventory Analytics.

I. INTRODUCTION

Enterprise resource planning (ERP) systems have long served as the transactional backbone of procurement, production, inventory management, and order fulfilment, yet recent operations literature suggests that the analytical layer surrounding ERP systems has evolved much faster than the transactional core [1]. Modern planning environments have heightened the need for fine-grained demand sensing, multi-horizon forecasting, exception-based replenishment, and a high frequency of scenario revision. Conventional ERP systems, especially monolithic deployments, are poorly suited to these demands because forecasting and material-planning functions were originally designed for relatively stable demand, slower update rates, and narrower product portfolios. In contemporary manufacturing and distribution, the problem is no longer simply whether material requirements planning (MRP) can be made correct. The difficulty is how to construct ERP modules that are computationally scalable, capable of absorbing volatile signals, are understandable by planners, and govern bills of material, lead times, safety stock and supplier constraints.

Broader supply-chain vulnerability also makes scalable ERP modules more relevant. The recent supply chain viability research has shown that continuity in operations is not only related to redundancy and resilience but also to the virtual capability to re-organize planning decisions in a fast and multi-node, multi-horizon way [2]. These results are of immediate importance to ERP design since the modules where demand changes, supply shocks, and policy changes become tangible purchase, production and allocation decisions are inventory forecasting and material planning. When modules are poorly designed, volatile input data can be translated into unstable planning outputs, increasing nervousness in procurement and shop-floor scheduling. With scalable and well-managed module design, the same platform can be used to provide segmented planning logic, differentiated service targets, as well as quicker recovery from disruptions.

The forecasting literature has supported that fact by demonstrating that turbulent environments are very severe on slow and solely backward-facing planning processes. Pandemic-era studies showed that forecasting and planning cannot be treated as independent organizational activities under rapidly changing environmental conditions, and structural breaks can be observed in the historical demand patterns [3]. When it comes to ERP systems, this implies that forecast modules cannot be modeled as a set of discrete statistical applications but in the form of continuous integrated decision services in conjunction with inventory policies, reorder intervals, capacity limits and communication with suppliers. Material-planning quality therefore does not depend on forecast accuracy alone. It is based on how the forecast results are translated into the



parameters of the planning, the uncertainty and the representation of the uncertainty, the override rules control and the escalation of the exception in the procurement, production and distribution functions.

The second force of change involves the transformation of local and single series forecasting into global and multi-horizon forecasting models that are able to learn on large product portfolios. Recent forecasting studies show that attention-based neural networks and related sequence models can capture complex temporal interactions, as well as, make multi-step predictions that can be utilized in rolling planning loops [4]. These techniques are appealing to the ERP-based forecasting since enterprise systems have thousands of stock-keeping units, non-stationary patterns of demand, and hierarchical products-location systems. However, more advanced models introduce substantial implementation burdens, including training cost, limited interpretability, dependence on feature engineering, and the need for high-quality monitoring. A simple ERP module, which merely loads a high-quality model and does not respond to latency, versioning, fallback logic, and parameter traceability is not likely to be stable in production.

A further architectural implication of digital-twin research is that planning should be treated as a closed-loop interaction among physical processes, data representations, and virtual models [5]. Although much of the digital twin literature does not address the ERP modules individually, a very strong implication is possible; scalable inventory forecasting and material planning are explicitly decoupled and tightly coupled layered structures with transaction capture, analytical processing, scenario evaluation and execution monitoring processes. Practical literature on ERP-native implementations is thinner as compared to related literature on supply chain analytics, forecasting techniques, and digital manufacturing. The consequent disaggregation creates a need for a review focused on scalable ERP modules rather than generic digital transformation. This review addresses that need by synthesizing recent journal literature on forecasting methods, planning systems, digital-twin-enabled control, and reconfigurable supply chains. The following sections present the thematic literature review, a topic-specific conceptual framework, comparison of reported methods, results, finding of contradictions and limitations, and a research agenda of the future, to scale ERP module design to inventory forecasting and material planning.

II. LITERATURE REVIEW

Recent journal literature relevant to scalable ERP modules for inventory forecasting and material planning can be organized into four related streams: large-scale forecasting algorithms, forecast aggregation and portfolio-level learning, digital twin and closed-loop planning, and reconfigurable or digitally integrated supply architectures. There are not many direct studies of the engineering of ERP modules. As a result, the evidence base is disseminated to the operations management, forecasting, manufacturing systems, and the digital supply chain journals. This pattern shows that the field has evolved through adjacent problem domains rather than through a stable body of ERP-focused design research. A review oriented toward scalable ERP modules must therefore translate adjacent evidence into ERP-specific architectural and methodological requirements.

Forecasting is the first major stream. Since 2020, forecasting has evolved rapidly, particularly in the cases where there are large clusters of related time series. The competition-based retail forecasting evidence demonstrated that hybrid and machine-learning-oriented forecasting models have the potential to boost accuracy in forecasting of large-scale SKU portfolios, especially with the use of cross-series learning, and feature-rich demand patterns [6]. A hybrid approach developed by Smyl put more emphasis on the importance of exponential smoothing structure and recurrent learning applied to large collections of forecasts [7], and feature-based forecasting model averaging demonstrated the possibility of scaling method selection via meta-learning instead of by planner selection of methods [8]. Another, more significant review of recurrent neural networks further explained that the current sequence models have significant flexibility in nonlinear and multi-horizon demand signals, but still can be susceptible to overfitting, lack of data on the item level, and lack of interpretability [9]. A key implication for ERP module design is that no single forecasting family is universally superior. The more significant implication is that scalable modules need model governance that is able to accommodate the heterogeneous demand regimes on a single enterprise platform.

The studies on digital-twin have taken the reverse perspective of planning; that is, it starts with the operating system and not the forecasting algorithm. Digital twin surveys in manufacturing suggest the value of real-time or near-real-time interaction between operational states and virtual decision models [10]. Models of digital-twin-driven manufacturing have taken that argument a step further by formalizing bidirectional flows between sensing, data integration, simulation, analytics as well as control [11]. This type of work is highly beneficial when it comes to material planning because ERP modules are increasingly becoming interoperable between the dynamic reality on the shop-floor and planning records. Having a computationally elegant material plan and not associated with determining work-in-process visibility, supplier reliability, or equipment status, a high degree of noise in the replanning process is likely to be typical. Digital-twin research therefore supports understanding scalable ERP modules as mediating structures between operational reality and planning

logic: forecast services can revise the planning logic but that the planning logic can only be adjusted to fit operational feedback loops but not to historical data only.

The disruption-aware planning and supply reconfigurations are a third stream. The most important in the continuity and maintenance of the services through the shock is the ripple effect of dynamic replanning aptitude during pandemics, as revealed by operations research [12]. Conceptual Industry 4.0 and Industry 5.0 work has also represented digital manufacturing architectures as being more data intensive, more interconnected and adaptive and having a more human oriented control compared to the earlier accounts of automation indicated [13]. Empirical studies have also linked Industry 4.0 adoption with improved supply-chain performance through digital integration [14]. Reconfigurable supply chain study went a step further by proposing modular network structures that have the potential of reconfiguring their structure when faced with uncertainty [15]. Although much of this literature addresses the supply chain rather than the ERP module directly, an important implication follows: scalable forecasting and material planning modules need to be able to be reconfigured instead of planning along a set path. A hardened module architecture can be useful in keeping the performance in a steady state but may become a weakness to a sourcing structure, routing decision or fulfilment logic restructuring.

Table 1 summarizes representative studies from 2020–2025 that inform the design of scalable ERP modules. There is a significant pattern that can be seen in the table. Methodological gains are often reported at the model level in forecasting studies and coordination, visibility, and feedback are the focus of digital twin and reconfiguration studies. Few articles combine these two perspective to a single ERP-native design reasoning. It is that gap that the title of the present review focuses on since scalable ERP modules have to work at the intersection of predictive quality, computational efficiency, discipline of planning, and transactional traceability.

Table 1: Summary of Key Findings

ef	Focus	Key Findings
6]	Large-scale retail demand forecasting across extensive SKU portfolios	Benchmark evidence showed that advanced and hybrid methods can outperform simpler baselines in high-dimensional product collections, yet business translation into replenishment outcomes was not automatic.
7]	Hybrid exponential smoothing and recurrent learning for many related time series	A hybrid global approach improved nonlinear demand capture while retaining some structural discipline from classical forecasting, making large-portfolio deployment more feasible than isolated item-level modeling.
8]	Feature-based model averaging for heterogeneous forecasting tasks	Meta-learning improved model selection across diverse series, indicating that scalable ERP forecasting layers may benefit from automated method allocation rather than universal single-model deployment.
9]	Recurrent neural network approaches to time-series forecasting	Sequence models offered strong flexibility for multi-step prediction, but computational burden, data hunger, and limited explainability remained barriers for operational planning environments.
10]	Digital twin applications in manufacturing operations	Virtual-operational coupling improved coordination and monitoring, but implementation success depended heavily on data fidelity, integration discipline, and update latency.
11]	Digital-twin-driven smart manufacturing reference architecture	Closed-loop data flows among sensing, analytics, simulation, and control created a stronger basis for adaptive planning, although ERP coupling was not specified in module-level detail.
12]	Disruption-oriented replanning and ripple-effect management	Dynamic optimization and rapid replanning reduced continuity risk, suggesting that material planning modules need scenario capability rather than only periodic MRP execution.
13]	Digital manufacturing evolution from Industry 4.0 toward human-centric control	Adaptive architecture and human-supervised analytics emerged as critical design principles, implying that scalable ERP modules require transparent override and escalation logic.
14]	Performance effects of digital integration in supply chains	Visibility and responsiveness improved under stronger digital integration, but much evidence remained cross-sectional and offered limited direct proof on ERP module scalability.
15]	Reconfigurable supply chain structures under uncertainty	Modular network designs enhanced adaptability, indicating that ERP planning modules should support segmented policies, structural change, and selective redeployment of planning logic.

Another revelation is made when literature forecasting is juxtaposed to literature planning on a more critical basis. Statistical loss functions of weighted error, scale-free accuracy are commonly optimized in forecasting studies, and continuity, recovery time, resource coordination or schedule stability are commonly optimized in planning studies. The relationship between forecast quality and planning value remains underdeveloped in the literature. This disconnect can be especially consequential in ERP environments. A forecast that improves mean error but does not preserve consistency across hierarchy, lead-time exposure, or substitution logic may still destabilize material plans. On the other hand, a small correct prediction into disciplined planning processes can yield improved service and inventory results since the propagation of parameters is coherent. Recent studies on twins of digital supply chains reinforce that reading by highlighting how resilience and disruption management on a system level instead of an isolated model level are supported [16].

The literature also reveals recurring limitations that hinder direct translation into ERP module design. First, relatively few studies report computational scalability in enterprise-relevant terms such as SKU counts, update frequency, memory cost, or integration latency. Second, most forecasting studies do not model master-data issues, although incorrect master units of measure, bill-of-material, and lead-time bias, and location distortions can negate algorithmic benefits. Third, digital twin and reconfiguration studies often tend to be conceptual or architecture-based, and provide less insight into how uncertainty related to forecasts ought to pass through MRP and procurement logic. Fourth, lumpy and intermittent demand, especially for spare parts and low-turn materials, remains challenging even for advanced models, with state-space methods and other studies demonstrating similar results [18]. Combined, the literature is indicative of significant advancements in technique and design, yet a lack of final convergence on the ERP module as the key implementation unit.

III. METHODOLOGY

A scalable ERP-module framework can be derived by connecting forecasting-method literature with systems-oriented digital-planning research. The reviewed literature suggests that scalable modules should not be understood as isolated algorithms embedded in ERP screens. A more plausible explanation is that the module can be considered as a stratified capability, comprising of transactional data capture, canonical data management, event-based orchestration, prediction services, planning optimization, human exception handling and execution feedback. This model fits the study on digital twins that places analytic and operational layers in a continuous exchange [5], and the manufacturing digitization study that conceptualizes the current production systems as inter-connected, modular, and adaptive [11], [13]. Its implications to inventory forecasting and material planning are direct, namely scaling is attained not so much by increasing the size of a monolithic MRP run but by decoupling responsibilities and interoperable services which can be adjusted independently.

The role of each approach in the ERP decision cycle can also be used to categorise the literature methodological approaches. Base-demand patterns are still appealing to classical statistical techniques and structured hybrids since the behavior of the parameters is relatively traceable and the overhead cost of deployment is less [7]. The larger the enterprise portfolio, rich in features, and cross-series learning can be utilized, the more valuable global machine learning models can be [4], [6], [9]. The logic of forecast combination and allocation of methods, such as feature-based model averaging, can be of particular interest to scalable modules since a heterogeneous SKU base is unlikely to favor one universal method [8]. But prediction is but a stratum. The use of scenario analysis, simulation, and optimization techniques are imperative in converting demand expectations to procurement timing, safety stock, order releases, and capacity-constrained material plans [12], [15]. The literature thus leans towards methodological pluralism which is controlled by modular orchestration as opposed to one paradigm of forecasting.

The same trend is reflected in architectural studies. Digital supply chain twin work and digital twin research demonstrate that closed-loop planning should entail a continuity of relationship between data refresh, state estimation, simulation, and action evaluation [10], [16], [17]. The pattern applies to ERP module design in case forecasting services are considered to be state-estimation engines and material planning services to be decision-propagation engines. In this framework, the authoritative transactional records include inventory positions, purchase orders, work orders, and bills of material, the analytical ERP services include the generation of forecasts, the estimation of uncertainty, the testing of scenarios and the recommendation of parameters, and orchestration services include the version control, event triggers, fallback logic and auditability. An interface discipline (and data contracts) are also significant to a scalable module as much as forecasting accuracy.

Figure 1 is a conceptual architecture that is relying on the literature that has been reviewed. It is deliberately domain-specific: forecasting and material planning is represented as a couple of related, but distinct decision services anchored to ERP transaction data, master data, and exception management. The architecture also emphasizes observability, model registry, and policy controls because recent literature suggests that scalable deployments fail when analytical logic cannot be monitored, interpreted, or linked to planning outcomes.

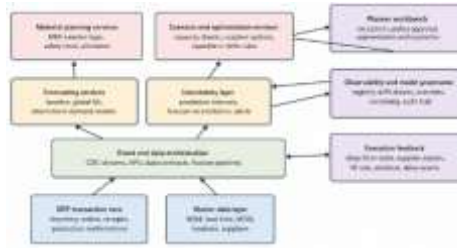


Figure 1: Layered Architecture for Scalable ERP Modules in Inventory Forecasting and Material Planning

This framework suggests a number of design implications which are supported by the literature. Forecasting should not be treated as a single planning task, but as a set of model families delivered through modular forecasting services. Point forecasts and uncertainty estimations, hierarchy-based aggregation and scenario metadata must also be considered in material planning. Planner intervention should be transparent, and should be controlled, especially in the environment where the demand shock or substitution or supplier disruption cannot be handled with algorithmic automation. Finally, it ought to be measured on four dimensions at the same time with respect to portfolio breadth, frequency of updates, latency of computation and organizational traceability. It is weak in ERP terms to have a scaling module that only scales numerically, leaving the interpretability and auditability intact.

IV. RESULTS AND DISCUSSION

Table 2: Method Comparison

ef	Method	Strengths	Limitations
7]	Hybrid exponential smoothing plus recurrent learning	Balances structural trend-seasonality handling with nonlinear pattern capture across many related series; suitable for large SKU portfolios with shared learning.	Retraining pipelines, model drift control, and operational explainability are more demanding than in classical ERP forecasting jobs.
8]	Feature-based forecast model averaging	Automates method allocation across heterogeneous demand patterns, reducing manual planner dependence in broad item portfolios.	Requires reliable meta-features and stable historical descriptors; weaker under abrupt structural breaks and sparse cold-start items.
4]	Temporal fusion transformer and related attention-based multi-horizon models	Handles covariates, multi-step horizons, and some interpretability through attention-based variable importance; useful for rolling planning cycles.	High computational cost, substantial data engineering burden, and limited transparency for parameter-level material planning decisions.
18]	State-space intermittent-demand forecasting	Better suited to lumpy, sparse, and low-turn material demand than dense global models; relevant for spare parts and service inventories.	Less effective for rich feature environments and weaker for integrated portfolio learning across thousands of active SKUs.
10]	Digital-twin-enabled planning models	Links virtual planning logic to operational state updates, improving plan realism and response speed.	Strong dependence on sensor quality, integration maturity, and synchronized master data across operational systems and ERP.
12]	Disruption-oriented optimization and ripple-effect replanning	Supports scenario evaluation, recovery logic, and continuity-aware material decisions under shocks.	Often modeled at network level rather than at deployable ERP module level; integration with transactional workflows remains thinly reported.
15]	Reconfigurable supply network design logic	Enables modular planning rules, segmented policies, and structural adaptation across supply configurations.	Design complexity is high, and empirical guidance on ERP implementation patterns remains limited.

The reviewed literature shows that scalable ERP-module performance depends on a series of interrelated design decisions rather than on any single methodological breakthrough. The first concerns the logic of forecast generation. The literature on competition and benchmarking proves that higher-order methods can indeed lead to a reduction of forecast

error in high-volume retail and multi-series problems [6], [7], [8]. However, it is suggested in the literature on planning that these gains do not necessarily spill over into improved material decisions. In ERP environments, forecasts become operationally useful only when they are mapped to item-location hierarchies, replenishment schedules, order multiples, supplier lead times, and bill-of-material structures. That translation layer is the least standardized part of recent literature. As a result, it is often reported that there are better predictive measures when published research does not have equally strong data on the stability of the service level, alleviation of shortages, or planning nervousness.

A second design question is whether forecasting should operate in an advisory, semi-automated, or closed-loop mode. Table 2 is the comparison of the most important methodological families, which are discussed in the literature. The comparison reveals that strengths and limitations are very contingent on the conditions of operation of ERP. Hybrid statistical-neural models can be scaled to a large number of series, and these models might need more stringent model monitoring and retraining control than classical methods. Planning with a digital-twin enhances responsiveness and realism of scenarios, but the cost of implementation and data dependency are high. Intermittent-demand methods are also valuable when there is low frequency items because high capacity global specifications are not effective in solving lumpy demand automatically.

Table 3 extends the comparison by focusing on reported outcomes and evaluation criteria. A consistent pattern emerges. Forecasting studies are concerned with reducing error, digital twin studies are concerned with planning fidelity and speed of response and supply chain reconfiguration studies are concerned with continuity and adaptability. Such a deviation can help to understand why the implementation of enterprise projects tends to be unsuccessful. Technical teams may evaluate prediction quality, whereas planners may evaluate stockout avoidance, schedule stability, or exception workload by the technical team. Scalable ERP module should thus mediate incompatible measures of success as opposed to supposing that one measure of system value exists.

Table 3: Results comparison

ef	System / Context	Metric / Basis of Comparison	Outcome
6]	Large retail forecasting environment with many related demand series	Competition accuracy metrics such as weighted RMSSE and operational ranking	Advanced and hybrid methods improved portfolio-level forecast accuracy, but operational translation into replenishment value remained context dependent.
7]	Large collections of time series in demand forecasting	Relative forecast accuracy versus benchmark statistical methods	Hybrid global modeling improved accuracy across broad series collections and demonstrated better scalability than isolated local models.
8]	Heterogeneous forecasting portfolios with differing demand signatures	Comparative forecast loss under automated method selection	Feature-based method allocation improved robustness across mixed series classes and reduced dependence on manual model choice.
4]	Multi-horizon forecasting settings with covariates	Quantile loss and interpretability-oriented evaluation	Strong multi-horizon performance and richer driver visibility were reported, although enterprise deployment cost remained high.
10]	Manufacturing digital twin applications	Planning fidelity, responsiveness, and virtual-physical consistency	Closed-loop planning environments improved decision realism, but data synchronization and maintenance burden were substantial.
12]	Disruption-affected supply planning	Recovery speed, continuity, and replanning responsiveness	Dynamic optimization and rapid replanning improved continuity management under shocks.
14]	Digitally integrated supply chain settings	Visibility, responsiveness, and performance association	Digital integration was positively associated with performance, but causal evidence on module-level ERP scalability was limited.
18]	Intermittent and sparse demand environments	Bias, service risk exposure, and demand-pattern suitability	State-space methods remained valuable for lumpy demand where dense global models may mis-handle zero-inflated structure.

The reviewed studies were coded into an evidence map (Figure 2). Forecast horizon and model complexity are shown on the x-axis and the depth of the operating system planning coupling is shown on the y-axis. Bubble size denotes the extent to which each study addresses scalability in terms of portfolio breadth, update frequency, or system-level flexibility. The chart suggests that forecasting studies tend to emphasize predictive sophistication while offering less coupling to planning execution, whereas digital-twin and disruption-planning studies are more tightly coupled to planning but provide less detailed forecasting analysis

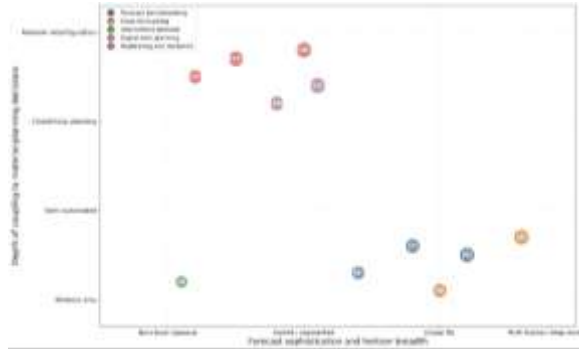


Figure 2: Evidence Map of Reviewed Studies by Forecast Sophistication and Planning Coupling

Figure 2 shows an emerging pattern that helps to critically interpret the literature. Advanced forecasting and deep-learning research are more mature than ERP-coupled planning research and digital twin and resilience research are more mature in architecture and in control than comparative forecasting evaluation. This directly influences that in most enterprise implementations there is a burden of translation between the model innovation and the planning usability. The literature thus prefers the modular ERP designs permitting the sophistication of forecasting to enhance without pressurizing immediate complete automation of material choices.

The other scholarly question that is answered in Figure 3 is: what methodological families are suitable to what ERP module requirements. Rather than evaluating methods in the abstract, the figure compares fit of six criteria of deployment that are of specific concern to inventory forecasting and material planning. The outcome is that there is no method which prevails across all the requirements. The global models and transformers are excellent with respect to horizon breadth and nonlinear learning; and intermittent-demand and structured statistical methods remain superior within the narrow operationally significant realms, such as sparse demand traceability and low-turn inventory control.

Figure 3 makes a comparative argument that adds to an explanation of why the design of ERP modules should be method-agnostic at the orchestration level. A platform that hard-codes a single forecasting family for all item classes is unlikely to be as responsive to fast movers, intermittent spares, promotion-responsive and disruptive item classes. The studied literature thus justifies segmented architecture, policy-based assignment of the methods and fallback rules that maintain the continuity of planning in case of advanced model drift and failure.

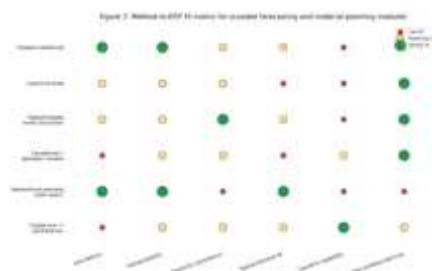


Figure 3: Method-To-ERP Fit Matrix for Scalable Forecasting and Material-Planning Modules

Another group of the reported results is related to the governance and operational feedback. Digital-twin studies repeatedly show that planning value increases when virtual models remain consistent with current operational conditions [10], [11], [17]. The disruption-oriented work also indicates that the ability to dynamically re-plan is more important in the presence of volatility than optimization under a set of assumptions which have become obsolete [12], [16]. In ERP terms, this means that scalable material planning depends not only on computational strength, but also on event-driven refresh, exception thresholds, and traceable override rules. These features tend to be considered as implementation details, but the literature indicates that such features are design features since the unstable feedback loops can nullify methodological benefits.

Figure 4 models that argument into an application map which correlates the forecast uncertainty and lead-time exposure with the module design priorities. The figure does not prioritize specific firms or industries; instead, it identifies planning regimes that require different ERP capabilities. Simpler modules can be utilized in low-uncertainty and short-lead environments, whilst probabilistic forecasting, supplier-collaboration logic, scenario simulation and more powerful exception governance are necessary in high-uncertainty and long-lead environments.

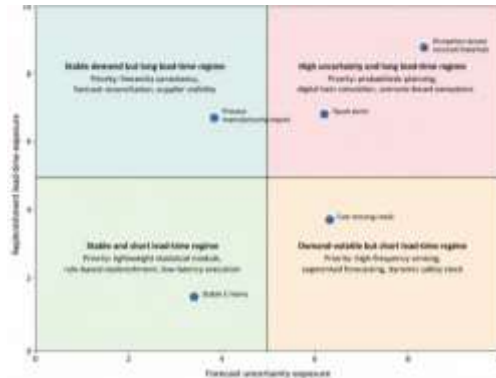


Figure 4: Planning Regime Map for Prioritizing ERP Module Capabilities

Another general inconsistency in recent literature, which is also explained in Figure 4, is grounded on the logic of regime. There are those that are inclined towards more powerful global models and others that put more emphasis on operationally based planning controls and digital twins. The two positions can be justified but to different exposure regimes, each of the positions is appropriate. The more probabilistic and scenario richness of high-uncertainty and long-lead environments; the data refresh and low-latency performance discipline, but not necessarily the maximum model complexity performance, of short-lead environments. Planning regimes should therefore be implemented through configurable ERP modules rather than through a single standardized technical template.

In the literature reviewed, it is possible to single out four more general findings. One, modular separation between transaction processing and analytical services is becoming more and more a necessity due to scale, maintainability and controlled experimentation. Second, the quality of forecasts needs to be measured in combination with planning implications like inventory stability, risk of shortage and the workload of the planner. Third, the master-data quality is a little-recognized bottleneck; inaccurate bills of material, calendar logic, or supplier parameters can erase algorithmic benefits quicker than a model choice can recuperate. Fourth, it lacks any empirical research, especially in terms of ERP-native module implementation. Most of the evidence published has been on either predictive methods external to ERP, or ERP based planning structures. The gap is a high research potential of the fact that the promise of analytics is realized in the enterprise module as operational value or failure.

V. ARCHITECTURAL CASE STUDY: DECOUPLING ANALYTICS VIA MICROSERVICES

The preceding sections established the need to separate transactional ERP processing from high-volume analytical workloads. However, the practical implementation of this separation at industrial scale remains insufficiently discussed in the literature. This section presents an architecture implemented in a high-volume manufacturing environment, where the planning lifecycle is organized as a set of cooperating, long-lived services rather than as stored procedures within a single monolithic database.

The system is organized into approximately a dozen separately deployable processes, each responsible for a defined segment of the planning domain. The application service exposes more than forty REST-facing controller endpoints for the planner interface and external integrations, with authorization enforced at the controller level. The core planning executable functions as the central planning engine and is hosted as a long-running service through systemd or Windows Service integration to support operational telemetry and service management. A work-distribution service polls a Redis-backed queue, retrieves allocation messages, and routes each message to one of three pluggable allocation stages: a deterministic classic allocator, a mathematical optimization solver, or a release-line generator. The selected stage depends on the stage indicator contained in the allocation message. Surrounding this core are specialized services for ERP synchronization, periodic min-max policy execution, purchase-order ingestion, transportation-management integration, and downstream portal publication. Each service is independently deployable, and services interact only through well-defined synchronous REST contracts or asynchronous messaging channels, including RabbitMQ for run-trigger events and Redis for high-volume work distribution.

One of the key design decisions in the system design was the explicit separation of the transactional and analytical data planes. The purchase order, inventory position, consolidated demand and consolidated supply tables, the pegging ledger and the part-master catalog remain in the relational database optimized for write integrity and small, hot working sets. The high-volume analytical state, however, namely the complete pre-consolidation snapshot of all the demand and supply rows for each planning run is staged into a columnar analytical store, and ingested using a streaming bulk-copy interface that can handle tens of thousands of rows of demand and supply data per warehouse snapshot. During planning, the engine loads working sets from the analytical store, consolidates them in memory, and writes only the consolidated results back to the relational database using the same analytical key as the working set. This pattern removes the long-running transactional schema scans which prevented operational writes in the past and enables the planning engine to work over time-series demand on a daily, per-part, per-warehouse granularity, which was previously not possible.

Each planning-run phase is represented as an enumerated type embedded within the work message rather than being managed by a centralized state machine. The message carries the state required for preprocessing, queuing for allocation, classic allocation, optional solver dispatch, solver execution, and post-processing. This message-resident state model removes the need for an additional workflow engine and allows any worker in the cluster to continue the next stage of processing without centralized coordination. Parallelism is achieved across three dimensions: warehouse-cluster tiers, part-warehouse partitions, and independent solver routes.

The architecture has been deployed in a multi-datacenter environment across five geographically dispersed regions, where planning runs reconcile millions of part-warehouse combinations against high-frequency demand and supply signals. The design is not limited to a specific ERP vendor. Its core principles separating transactional and analytical data planes, decomposing the planning lifecycle into independently deployable services, using queues and REST contracts for service interaction, adopting message-resident orchestration, and parallelizing work along the natural structure of the planning universe are applicable to large-scale manufacturing ERP systems that must manage high-velocity demand signals against constrained supply.

VI. FUTURE DIRECTIONS

The most immediate research gap is the lack of empirical study of ERP-native implementations, as distinct from isolated methodological comparisons. Recent literature provides strong evidence on forecasting models and growing evidence on digital planning architectures, yet there remains a lack of direct journal studies on the implementation of such capabilities in the scaled ERP modules. Future research should report implementation anatomy more explicitly in a more explicit form with service boundaries, update frequency, event-driven, model fallback logic, and control of planner overrides. This level of detail is necessary since the architecture of a system and the discipline of operation are as much drivers of scalable material planning as is predictive performance.

Second priority is related to the propagation of uncertainty. Forecasting research has moved beyond point prediction and the material planning research has under-reported the flow of interval forecasts, distributional forecasts or scenario trees through the reorder logic, MRP explosions and constrained supplier planning [4], [6], [12]. Probabilistic ERP modules which are capable of converting uncertainty to differentiated stocking policies, procurement timing, as well as exception thresholds require more work. The concurrent assessment of planning nervousness, risk of stockout and the cost-to-serve must also be taken into account in this work not on the basis of forecast error.

A third research priority concerns data governance and digital-twin calibration. The literature has noted the virtual-operational consistency as having significance but more insight is needed to demonstrate how the flaws of master-data, slow confirmations, or inconsistent updates distort the quality of planning [10], [11], [16], [17]. Future research ought to focus on canonical data models of item-location hierarchies, bill-of-material version control, supplier parameter drift and reliability of event timestamps. Controlled comparisons of batch-based ERP planning and event-driven architectures would be valuable.

One of the last opportunities is in interdisciplinary evaluation design. However, the journals of forecasting, production and enterprise systems continue to evaluate performance using disconnected performance metrics to some extent. Moreover through the integration of measurements of predictive accuracy, service level, inventory turns, planner workload, computational latency, and governance traceability, stronger advances would be achieved. This assessment should be separated into regime plans as the literature studied has shown that different exposure profiles have different amounts of forecasting sophistication and planning control. Future studies should therefore shift toward regime-based design, richer implementation reporting, and multifunctional outcome measurement for scalable ERP modules in inventory forecasting and material planning.

VII. CONCLUSION

Scalable ERP modules for inventory forecasting and material planning occupy a strategically important but underdeveloped position in the current literature. Recent efforts have seen substantial innovations in massive forecasting, digital twins, replanning when faced with disruptions, and reconfigurable supply networks. However, direct evidence on how such advances should be incorporated into ERP-native modules remains limited. The main value of the present review is that the forecasting excellence, planning discipline, system scalability and governance traceability are to be considered a single design problem rather than a collection of technical domains.

This interpretation supports several core conclusions. First, modularity and layered architecture are more appropriate than monolithic planning logic in environments with expanding portfolios, higher uncertainty, and more frequent replanning. Second, forecasting techniques must be chosen and controlled by the item type, planning horizon, and operational regime and not uniformly across the whole enterprise. Third, digital feedback loops, master-data quality, and exception governance are not peripheral implementation details; they are core determinants of whether analytical capability improves actual material decisions. Fourth, the discrepancy between model-level outputs and module-level value is the biggest unaddressed problem in this area.

The perspective is thus positive yet challenging. Probabilistic planning logic, implementation-oriented studies, regime-based module configuration and evaluation measures that link prediction quality to operational consequences will be required to offer better evidence. Innovations in this area could reposition ERP from a transactional backbone to a dynamic decision platform capable of supporting resilient, transparent, and scalable material planning under uncertainty.

- **Interest Conflicts:** The author declares that there is no conflict of interest concerning the publishing of this paper.

VIII. REFERENCES

- [1] Choi, T. M., Wallace, S. W., & Wang, Y. (2020). Big data analytics in operations management. *Production and Operations Management*, 29(8), 1868–1883.
- [2] Ivanov, D., & Dolgui, A. (2020). Viability of intertwined supply networks: Extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak. *International Journal of Production Research*, 58(10), 2904–2915.
- [3] Nikolopoulos, K., Punia, S., Schäfers, A., Tsinopoulos, C., & Vasilakis, C. (2021). Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. *International Journal of Forecasting*, 37(2), 739–766.
- [4] Lim, B., Arik, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748–1764.
- [5] Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the digital twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, 29, 36–52.
- [6] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4), 1346–1364.
- [7] Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, 36(1), 75–85.
- [8] Montero-Manso, P., Athanasopoulos, G., Hyndman, R. J., & Talagala, T. S. (2020). FFORMA: Feature-based forecast model averaging. *International Journal of Forecasting*, 36(1), 86–92.
- [9] Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1), 388–427.
- [10] Cimino, C., Negri, E., & Fumagalli, L. (2020). A review of digital twin applications in manufacturing. *Computers in Industry*, 113, 103130.
- [11] Lu, Y., Liu, C., Wang, K. I.-K., Huang, H., & Xu, X. (2020). Digital twin-driven smart manufacturing: Connotation, reference model, applications, and research issues. *Robotics and Computer-Integrated Manufacturing*, 61, 101837.
- [12] Ivanov, D., & Dolgui, A. (2021). OR-methods for coping with the ripple effect in supply chains during COVID-19 pandemic: Managerial insights and research implications. *International Journal of Production Economics*, 232, 107921.
- [13] Xu, X., Lu, Y., Vogel-Heuser, B., & Wang, L. (2021). Industry 4.0 and Industry 5.0—Inception, conception and perception. *Journal of Manufacturing Systems*, 61, 530–535.
- [14] Fatorachian, H., & Kazemi, H. (2021). Impact of Industry 4.0 on supply chain performance. *Production Planning & Control*, 32(1), 63–81.
- [15] Dolgui, A., Ivanov, D., & Sokolov, B. (2020). Reconfigurable supply chain: The X-network. *International Journal of Production Research*, 58(13), 4138–4163.
- [16] Ivanov, D., & Dolgui, A. (2021). A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. *Production Planning & Control*, 32(9), 775–788.
- [17] Rasheed, A., San, O., & Kvamsdal, T. (2020). Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access*, 8, 21980–22012.
- [18] Svetunkov, I., & Boylan, J. E. (2020). State-space model for intermittent demand forecasting. *European Journal of Operational Research*, 286(1), 321–339.