

Mathematical Modelling of Inventory Systems for Optimal Production Planning under Cost–Service Trade-offs

Taruna Sharma^{1*}, R.K. Srivastava²

¹Research Scholar, Agra College, Affiliated to Dr. B.R. Ambedkar University, Agra (India)

²Professor, Agra College, Affiliated to Dr. B.R. Ambedkar University, Agra (India)

*Corresponding Author

Abstract: This study develops a mathematical framework for inventory-based production planning under explicit cost–service trade-offs by formulating a continuous-time nonlinear dynamic system in which inventory level, backlog, service responsiveness, and production rate evolve simultaneously. The proposed model captures service-sensitive demand, backlog recovery, linear and nonlinear inventory loss, and the effect of production responsiveness on service performance, thereby providing a more realistic representation than classical linear inventory models. A discounted cost functional is constructed to balance production, holding, shortage, service-support, and deterioration costs subject to service-level requirements. To obtain analytical insight, the full model is reduced under practically meaningful assumptions to a Riccati-type nonlinear differential equation, from which a closed-form inventory trajectory, equilibrium solution, and stability condition are derived. A numerical case study demonstrates monotone convergence of inventory to a stable equilibrium and reveals a clear cost–service frontier: higher service targets improve fill rate and customer fulfillment, but they also require larger equilibrium inventory, greater production effort, and higher total cost. The results show that nonlinear deterioration plays a critical role in limiting feasible inventory expansion and highlight the importance of balancing customer service ambitions with operational cost efficiency in production planning.

Keywords: Inventory systems, optimal production planning, nonlinear differential equations, cost–service trade-offs, service-sensitive demand, Riccati equation, equilibrium analysis, fill rate, nonlinear deterioration, production control.

I. INTRODUCTION

In modern production and supply systems, firms must maintain an efficient balance between cost control and customer service performance, since inadequate inventory causes shortages and poor fulfillment, while excessive production and stockholding increase carrying, deterioration, and operating costs. This paper addresses that problem by developing a nonlinear inventory-production model in which production rate, inventory level, shortage accumulation, and service responsiveness interact dynamically over time. The model assumes that demand is positively influenced by service intensity and negatively affected by inventory congestion, while inventory is subject to both linear and nonlinear loss, and shortages are only partially recoverable. These assumptions reflect realistic operating environments in which service improvement attracts demand but also requires stronger replenishment support and more careful control of stock-related inefficiencies. To analyze this system, the study formulates a discounted optimization problem, derives optimality conditions, and then obtains a closed-form analytical solution for a reduced Riccati-type model. The numerical results further show that higher service targets raise effective demand, equilibrium inventory, production rate, and total cost, while fill rate improves with diminishing marginal gains. Therefore, the study is motivated by the need for a mathematically rigorous yet managerially interpretable framework that connects inventory control, production responsiveness, and service-level design in a unified decision model.

Albey *et al.* (2015) examined how forecast evolution can be modeled more realistically for production planning environments in which demand information changes over time rather than remaining fixed. Their contribution is important because they showed that planning systems perform better when they recognize that forecasts are revised gradually as new market information arrives. The study highlighted the operational value of linking demand uncertainty with dynamic planning decisions, especially in industries where production commitments must be made before final demand is fully known. By emphasizing the pattern and structure of forecast updates, their work strengthened the idea that production plans should not be based only on point forecasts, but on the way information matures across planning periods. Bilginer and Erhun (2015) focused on new product introduction under capacity limitations and explained that

production and sales decisions must be coordinated carefully when firms launch products into uncertain markets. Their discussion showed that launch planning is not merely a manufacturing problem, because marketing, demand timing, and capacity usage interact strongly. They demonstrated that improper coordination can create shortages, lost opportunities, or inefficient use of resources, whereas integrated planning improves both responsiveness and economic performance during the early stages of a product's life cycle. **De Kok (2015)** discussed uncertainty buffering in high-tech supply chains and emphasized that advanced manufacturing systems face a combination of volatile demand, long lead times, and expensive components. The work clarified that uncertainty cannot be absorbed effectively through a single rule or a single stock point; rather, managers must balance responsiveness, inventory investment, and structural flexibility throughout the network. **Glock and Grosse (2015)** reviewed production ramp-up decision support models and showed that the transition from development to full-scale production is one of the most sensitive phases in operations management. Their synthesis made clear that ramp-up performance depends on coordinated decisions about capacity, labor, process stability, inventory, and timing. They also revealed that production ramp-up research had become fragmented across several subtopics, and their review helped organize the field by identifying major modeling streams and practical gaps. **Afrouzy et al. (2016a)** proposed an approach based on evolutionary optimization for supply chain configuration in the presence of new product development considerations. Their work underlined that network design becomes more complex when firms must simultaneously decide facility structure, material flow, and product-development-related choices. The importance of the study lies in showing that classical exact approaches may struggle with such combinatorial complexity, while metaheuristic procedures can provide strong near-optimal solutions in large and realistic settings. **Afrouzy et al. (2016b)** extended this line of thinking by incorporating fuzziness, stochastic behavior, and multiple objectives into supply chain configuration. Their contribution is especially meaningful because real supply chains rarely optimize only one target; instead, cost, responsiveness, risk, and strategic development concerns often coexist. By integrating uncertainty and multi-objective reasoning, they provided a more realistic basis for decision-making and demonstrated how managers can evaluate trade-offs rather than rely on a single performance measure. **Eruguz et al. (2016)** provided a broad synthesis of guaranteed-service approaches for multi-echelon inventory optimization and clarified the main assumptions, structures, and extensions of that literature. Their work is valuable because it organized a large body of research into a coherent framework, enabling scholars to understand where guaranteed-service models perform well and where their assumptions become restrictive. **Graves and Schoenmeyr (2016)** examined safety stock placement under capacity constraints and showed that capacity fundamentally changes the logic of inventory positioning in supply chains. Their study demonstrated that when production resources are limited, inventory decisions cannot be separated from capacity usage, because stock buffers also serve as protection against congestion and delay. **Huh et al. (2016)** investigated capacitated multiechelon inventory systems and developed policy insights together with analytical bounds. Their work contributed to the understanding of how structured control rules behave in systems where replenishment opportunities are restricted by upstream limitations. **Ji et al. (2016)** analyzed a two-echelon serial inventory setting with limited capacities and explained how optimal policies must account for both echelon interaction and finite replenishment ability. Collectively, the 2016 studies deepened the field's understanding of how capacity, uncertainty, and system structure jointly shape inventory and supply chain performance. **Angelus and Zhu (2017)** studied multi-stage capacitated inventory systems from an upstream perspective and showed that decisions made earlier in the chain have major consequences for downstream availability and overall system efficiency. Their work is significant because it shifted attention toward the strategic importance of upstream control and the need to view multistage systems as tightly connected decision environments rather than isolated stocking points. **Meistering and Stadler (2017)** explored capacitated lot sizing with multiple products in rolling planning settings and highlighted the challenge of maintaining service performance while repeatedly revising schedules over time. Their findings reinforced the idea that short-term schedule stability and customer fill-rate targets must be treated jointly, especially when several products compete for common resources. These 2017 contributions together emphasized dynamic coordination, multistage dependence, and the need for planning rules that remain effective under repeated revision and capacity competition. **De Kok (2018)** presented a broad treatment of inventory management with emphasis on real-life supply chains and empirical validity. His discussion is especially important because it moved beyond purely theoretical modeling and argued that inventory research should be judged by how well it explains and improves actual practice. He underscored that real systems are shaped by behavioral rules, structural complexities, and data imperfections that are often simplified away in stylized models. **De Kok et al. (2018)** developed a typology and review of stochastic multi-echelon inventory models, offering a structured classification of the field. Their contribution helped clarify the diversity of assumptions used in the literature regarding uncertainty, network form, control logic, and optimization goals. By organizing the domain systematically, they made it easier to identify both mature areas and unresolved challenges, especially for complex supply networks. **Woerner et al. (2018)** examined simulation-based optimization for capacitated assembly systems under service constraints and demonstrated the usefulness of combining simulation with optimization when analytical models become difficult to solve exactly. Their study is important because assembly networks often involve nonlinear interactions, service requirements, and capacity restrictions that resist simple closed-form treatment. **Aouam et al. (2021)** addressed production capacity and safety stock placement when inventory budgets are finite, thereby bringing financial

realism into guaranteed-service reasoning. Their work showed that even when inventory is strategically desirable, firms operate under budget limits that force selective allocation of buffers across the network. This contribution is useful because it connects service-oriented planning with capital discipline and makes safety stock placement more relevant to managerial reality. **Tavaghof Gigloo and Minner (2021)** studied stochastic capacitated lot sizing under service-level constraints and compared planning approaches for handling uncertainty when resources are limited. Their findings highlighted the importance of integrating service targets directly into lot-sizing logic rather than treating them as an afterthought. The study contributed to the growing recognition that dynamic production planning must explicitly balance uncertainty, capacity, and customer-service promises. **Voas et al. (2021)** discussed the semiconductor shortage and its wider implications for global systems, drawing attention to the fragility of technologically advanced supply networks. Their analysis showed that disruptions in one critical sector can quickly produce economic and strategic consequences across industries and countries. In the broader context of production and inventory research, this work reinforced why resilience, lead time awareness, and supply risk have become central concerns in modern high-tech operations. **Bhuniya et al. (2021)** developed a supply chain perspective in which service level requirements are treated as a central performance condition rather than a secondary outcome. Their work is important because it explains how uncertainty affects both operational cost and the ability of a system to satisfy customers on time and in the desired quantity. By incorporating service restrictions directly into the decision structure, they showed that production, replenishment, and distribution policies must be designed in a balanced way so that cost efficiency does not come at the expense of customer responsiveness. Their contribution also strengthened the literature on uncertainty-aware planning by demonstrating that the optimal policy in a supply chain environment depends not only on inventory and production parameters, but also on the quality of service promised to the market. This makes the study highly relevant for mathematical modelling of inventory systems under cost-service trade-offs. **Hrabec et al. (2022)** examined integrated decision-making across production, inventory, and routing and showed that these areas should not be optimized in isolation. Their contribution is especially valuable because they synthesized and evaluated a broad body of prior work, making clear that fragmented planning often leads to hidden inefficiencies. When production timing, stock levels, and transportation choices are coordinated jointly, firms can reduce duplicated effort, lower total costs, and improve service performance at the same time. Their analysis reinforces the idea that inventory systems should be modelled as part of a wider operational network in which decisions interact across stages and functions. For research on optimal production planning, this study provides strong support for integrated models that explicitly capture trade-offs among capacity usage, stock positioning, and delivery quality. **Alkahtani et al. (2022)** addressed the relationship between inventory modelling and process outsourcing in supply chain optimization. Their study is significant because it broadened the traditional inventory control problem by including outsourcing as a strategic decision variable. This means that stock management is no longer only about how much to order or produce, but also about where work should be performed and how internal and external processes should be coordinated. By introducing outsourcing into the mathematical structure, the authors highlighted that cost minimization, continuity of supply, and operational flexibility are strongly connected. Their work is useful for modern production planning because many firms rely on hybrid supply structures, and the inventory consequences of outsourcing can substantially affect service reliability and total system performance. **Ghadimi et al. (2023)** investigated safety stock placement when market selection and load-dependent lead times are present. Their work is particularly important because it recognizes that lead times are not always fixed or exogenous; instead, they can increase as the production system becomes more loaded. This makes the inventory decision much more realistic and more difficult. The study showed that safety stock placement should not be treated as a static buffering exercise, because upstream congestion, market priorities, and system loading all influence the best positioning of inventory. Their contribution advances the literature by linking strategic inventory placement with operational conditions in capacity-sensitive environments. In the context of cost-service trade-offs, the study demonstrates that good service performance often depends on protecting the system from delay variability through carefully targeted stock decisions. **Alejo-Reyes et al. (2023)** studied an inventory management and order allocation problem in which both nonlinear quantity discounts and nonlinear price-dependent demand play important roles. Their contribution is meaningful because it moves beyond simpler assumptions of constant demand and linear purchasing structures. In real markets, customer demand often reacts to price, and suppliers often offer discounts that make order sizing more complex. The authors showed that the optimal inventory policy must therefore account for both procurement incentives and market response simultaneously. This creates a richer and more realistic model for production and replenishment planning, especially in environments where firms need to balance purchasing economy, selling strategy, and stock availability. Their study is a useful example of how inventory modelling can incorporate economic behavior directly into operational decision-making. **León et al. (2024)** focused on service-level optimization in a multi-warehouse supply chain network under stochastic demand. Their work is highly relevant because it addresses one of the key challenges of multi-location systems: how to position inventory across warehouses so that customer service can be maintained despite uncertain demand patterns. The study emphasizes that service level should be treated as an optimization target that interacts with stock allocation, replenishment planning, and network structure. By considering multiple warehouses together, the authors highlighted the importance of coordination across locations rather than relying on isolated local rules. Their contribution supports the view that optimal production and inventory planning must be

network-aware, especially when firms seek to improve service without incurring excessive holding and transportation costs. **Jabeur et al. (2024)** examined integrated lot sizing and production scheduling in flexible flow line systems while also incorporating energy-related considerations and a two-level learning-based solution logic. Their work is significant because it shows that modern production planning is no longer limited to traditional cost components such as setup, holding, and shortage costs. Energy use has become an important operational concern, and integrating it into lot-sizing and scheduling changes the planning problem substantially. The study illustrates that production quantities, sequencing decisions, and operational efficiency must be coordinated in a unified framework. By doing so, the authors expanded the meaning of cost–service trade-offs to include resource and energy efficiency, which is increasingly relevant in sustainable manufacturing contexts. Their work also suggests that advanced computational methods can help solve large integrated planning problems that are difficult to manage with classical approaches alone. **Kohlmann and Sahling (2024)** developed a flexible approach for integrated lot sizing and rework planning in the presence of defective products. Their contribution is important because it acknowledges that imperfect production is a common feature of real manufacturing systems, and rework decisions can significantly influence inventory dynamics. Instead of assuming that all produced items are usable, they included the possibility that some units require correction, which creates additional interactions among production planning, inventory control, and quality management. The study demonstrates that ignoring rework can lead to distorted lot-sizing decisions and poor service performance, particularly when defect proportions are uncertain. Their work strengthens the inventory-planning literature by showing that production quality variability must be incorporated into mathematical models if firms want realistic and robust decision support.

II. ASSUMPTIONS OF THE MODEL

The model is based on the assumptions that the system deals with a single product, production occurs continuously over time, demand is influenced by service quality, inventory experiences both linear and nonlinear losses, shortages are partially recoverable, and service performance changes dynamically depending on production responsiveness. The planning horizon is finite, and the decision-maker has direct control over the production rate subject to capacity limitations.

Table 1: Nomenclature table for the model

Symbol	Description
$I(t)$	Inventory level at time (t)
$B(t)$	Backlog or shortage level at time (t)
$s(t)$	Service responsiveness index
$u(t)$	Production rate control
d_0	Base demand rate
κ	Service elasticity of demand
ν	Inventory saturation parameter
ρ	Backlog recovery rate
θ	Linear deterioration coefficient
γ	Nonlinear inventory-loss coefficient
η	Backlog-clearing efficiency
α	Service gain from production responsiveness
β	Service decay rate
χ	Nonlinear service fatigue coefficient
c_p	Production cost parameter
h	Holding cost parameter
p	Shortage cost parameter
c_s	Service support cost parameter
c_d	Deterioration cost parameter
r	Discount rate
$FR(t)$	Fill rate or service fulfillment ratio
\underline{FR}	Minimum target service level
u_{max}	Maximum production capacity

III. NONLINEAR MATHEMATICAL MODEL

In order to study optimal production planning under cost–service trade-offs, a continuous-time nonlinear inventory system is developed in which production rate, inventory level, shortage accumulation, and service responsiveness evolve simultaneously over time. The model is designed to capture the realistic interaction between production decisions and customer service requirements, since a higher service level generally requires faster replenishment and higher inventory support, while excessive production leads to increased holding, deterioration, and operating costs. Let $I(t)$ denote the on-hand inventory level at time t , $B(t)$ denote the backlog or shortage level, $s(t)$ denote the service responsiveness index, and $u(t)u(t)u(t)$ denote the controllable production rate. The system is assumed to operate over a finite planning horizon $0 \leq t \leq T$, with initial conditions

$$I(0) = I_0, B(0) = B_0 \text{ and } s(0) = s_0 \quad (1)$$

The demand rate is assumed to depend nonlinearly on the service level and inversely on the available inventory congestion state. A suitable form of service-sensitive demand is taken as

$$D(t) = \frac{d_0[1+\kappa s(t)]}{1+\nu I(t)} \quad (2)$$

where $d_0 > 0$ is the base demand rate, $\kappa > 0$ measures the sensitivity of demand to service responsiveness, and $\nu > 0$ represents the inventory saturation effect. This expression implies that improved service quality attracts demand, whereas a large inventory level may slow turnover because of market saturation, storage inefficiency, or mismatch between production and actual sales velocity.

The inventory dynamics are governed by the balance between production inflow, customer demand, backlog recovery, and inventory loss. Therefore, the inventory state equation is written as

$$\frac{dI}{dt} = u(t) - D(t) + \rho B(t) - \theta I(t) - \gamma I^2(t) \quad (3)$$

where $\rho > 0$ is the backlog recovery rate, $\theta > 0$ is the linear deterioration coefficient, and $\gamma > 0$ is the nonlinear inventory-loss coefficient. The term $\theta I(t)$ captures ordinary spoilage, obsolescence, or carrying inefficiency proportional to the stock level, while the quadratic term $\gamma I^2(t)$ reflects nonlinear losses that become more significant when inventory is high, such as congestion cost, damage risk, overstock burden, or accelerated obsolescence.

The shortage or backlog evolves according to unmet demand and recovery through production adjustment. The backlog equation is defined as

$$\frac{dB}{dt} = D(t) - \eta u(t) \frac{B(t)}{1+B(t)} - \rho B(t) \quad (4)$$

where $\eta > 0$ is the shortage-clearing efficiency parameter. The term $\eta u(t) \frac{B(t)}{1+B(t)}$ introduces nonlinearity in backlog clearance and reflects the practical situation in which an increase in production helps reduce backorders, but the marginal effectiveness of additional production declines when the shortage level becomes very large. The term $\rho B(t)$ represents the fraction of backlog that is gradually recovered and converted into fulfilled demand.

Since customer service is an important strategic factor in production planning, a service state variable $s(t)$ is incorporated into the model. Its dynamics are described by

$$\frac{ds}{dt} = \alpha u(t) - \beta s(t) - \chi s^2(t) \quad (5)$$

where $\alpha > 0$ measures the positive influence of production responsiveness on service improvement, $\beta > 0$ is the natural decay rate of service performance, and $\chi > 0$ is the nonlinear fatigue or diminishing-return coefficient. This equation implies that faster and more adaptive production improves service, but sustaining a high service level becomes increasingly difficult and costly over time because of managerial fatigue, capacity pressure, or service-system saturation.

This system is suitable because it captures the following:

- (i) production replenishes inventory,
- (ii) demand depletes inventory,
- (iii) unsatisfied demand creates backlog,

- (iv) backlog can be recovered later,
- (v) inventory experiences linear and nonlinear loss,
- (vi) service quality rises with production responsiveness but weakens over time.

3.1. Cost functional and optimization problem: The decision variable of the model is the production rate $u(t)$, which is assumed to satisfy the operational constraint

$$0 \leq u(t) \leq u_{max} \tag{6}$$

A suitable total discounted cost functional over planning horizon $[0, T]$ is

$$J(u, s) = \int_0^T e^{-rt} [c_p u^2(t) + hI(t) + pB(t) + c_s s^2(t) + c_d \gamma I^2(t)] dt \tag{7}$$

subject to equations (2)–(4), initial conditions in equations (1) and service requirement

$$FR(t) \geq \underline{FR} \tag{8}$$

where:

c_p : production adjustment cost

h : inventory holding cost

p : shortage penalty

c_s : service-support cost

c_d : nonlinear deterioration cost

r : discount rate

$FR(t)$: fill rate or service fulfillment ratio

\underline{FR} : Minimum target service level

A smooth proxy for fill rate can be expressed as

$$FR(t) = 1 - e^{-\tau I(t)} \tag{9}$$

with $\tau > 0$. This makes the service-performance relation differentiable and analytically convenient.

3.2. Optimality conditions: The decision variable of the model is the production rate $u(t)$, which is assumed to satisfy the operational constraint

$$\mathcal{H} = e^{-rt} [c_p u^2 + hI + pB + c_s s^2 + c_d \gamma I^2] + \lambda_1 (u - D + \rho B - \theta I - \gamma I^2) + \lambda_2 \left(D - \eta u \frac{B}{1+B} - \rho B \right) + \lambda_3 (\alpha u - \beta s - \chi s^2) \tag{10}$$

Then the adjoint equations are

$$\dot{\lambda}_1 = -\frac{\partial \mathcal{H}}{\partial I}, \dot{\lambda}_2 = -\frac{\partial \mathcal{H}}{\partial B}, \dot{\lambda}_3 = -\frac{\partial \mathcal{H}}{\partial s} \tag{11}$$

and the stationary condition for optimal production is

$$\frac{\partial \mathcal{H}}{\partial u} = 0 \tag{12}$$

$$2c_p e^{-rt} u + \lambda_1 - \lambda_2 \eta \frac{B}{1+B} + \alpha \lambda_3 = 0 \tag{13}$$

which yields

$$u^* = - \frac{\lambda_1 - \lambda_2 \eta \frac{B}{1+B} + \alpha \lambda_3}{2c_p e^{-rt}} \tag{14}$$

IV. ANALYTICAL SOLUTION OF THE REDUCED MODEL

To obtain a closed-form solution, impose the following practically meaningful approximations:

- (i) service adjusts faster than inventory, so $s(t) = s^*$ over the medium horizon
- (ii) backlog is small under good control, so $B(t) \approx 0$
- (iii) the production policy is linear state feedback $u(t) = u_0 + u_1 I(t)$
- (iv) For moderate $\nu I(t)$, expand demand using first-order approximation

$$\frac{1}{1+\nu I(t)} \approx 1 - \nu I(t) \tag{15}$$

Then demand becomes

$$D(t) \approx d_0 [1 + \kappa s^*] [1 - \nu I(t)] \tag{16}$$

Substituting (15) and (16) into (2), we obtain

$$\frac{dI}{dt} = u_0 + u_1 I - d_0 [1 + \kappa s^*] [1 - \nu I] - \theta I(t) - \gamma I^2$$

$$\text{Rearranging, } \frac{dI}{dt} = a_0 + a_1 I - a_2 I^2 \tag{17}$$

$$\text{Where } a_0 = u_0 - d_0(1 + \kappa s^*) \tag{18}$$

$$a_1 = u_1 + \nu d_0(1 + \kappa s^*) - \theta \tag{19}$$

$$a_2 = \gamma \tag{20}$$

Equation (17) is a Riccati-type nonlinear differential equation.

4.1. Closed-form solution: Let the roots of

$$a_2 I^2 - a_1 I - a_0 = 0 \tag{21}$$

$$I_{\pm} = \frac{a_1 \pm \sqrt{a_1^2 + 4a_0 a_2}}{2a_2}$$

$$\frac{dI}{dt} = -a_2 (I - I_+) (I - I_-) \tag{22}$$

Separating variables gives

$$\int \frac{dI}{(I-I_+)(I-I_-)} = -a_2 t + C$$

Hence the exact solution is

$$\frac{I(t)-I_+}{I(t)-I_-} = \frac{I_0-I_+}{I_0-I_-} e^{-a_2(I_+-I_-)t}$$

$$\text{or equivalently, } I(t) = \frac{I_+ - C I_- e^{-\Delta t}}{1 - C e^{-\Delta t}} \tag{23}$$

$$\text{Where } C = \frac{I_0 - I_+}{I_0 - I_-}, \Delta = a_2(I_+ - I_-)$$

This gives the inventory trajectory explicitly.

4.2. Equilibrium and stability: The positive equilibrium is $I^* = I_+$. It is asymptotically stable whenever $a_2 > 0$ and $I_+ > I^-$, since the solution converges to I_+ as $t \rightarrow \infty$. This is economically meaningful because nonlinear loss prevents unlimited inventory buildup.

4.3 Propositions: We include results such as:

Proposition 1. If $a_0 > 0, a_2 > 0$ and $a_1^2 + 4a_0a_2 > 0$, then the reduced inventory system admits a unique positive equilibrium $I^* = I^+$

Proposition 2. The equilibrium $I^* = I^+$ is locally asymptotically stable for the reduced model.

Proposition 3. An increase in service intensity s^* increases effective demand and therefore shifts the equilibrium inventory upward unless production parameters are adjusted downward.

Proposition 4. The marginal gain in service level diminishes as equilibrium inventory increases, while total operating cost grows more than proportionally because of nonlinear production and inventory costs.

V. NUMERICAL CASE STUDY

A medium service target is first selected with $s^* = 0.80$. Substituting this value into the effective demand expression gives

$$d_0(1 + \kappa s^*) = 90(1 + 0.8 \times 0.80) = 147.6 \tag{24}$$

Table 2: Parameter values for the numerical case study		
Parameter	Description	Value
d_0	base demand rate	90
κ	service elasticity of demand	0.8
ν	inventory saturation parameter	0.01
θ	linear deterioration rate	0.12
γ	nonlinear inventory-loss coefficient	0.004
u_0	baseline production rate	170
u_1	inventory feedback coefficient	0.1
τ	fill-rate sensitivity parameter	0.008
c_p	production cost coefficient	2.2
h	holding cost coefficient	0.9
c_s	service-support coefficient	180
I_0	initial inventory level	40

Hence the reduced-model coefficients are

$$a_0 = 170 - 147.6 = 22.4 \tag{25}$$

$$a_1 = 0.10 + 0.01(147.6) - 0.12 = 1.456 \tag{26}$$

$$a_2 = 0.004 \tag{27}$$

$$I_{\pm} = \frac{a_1 \pm \sqrt{a_1^2 + 4a_0a_2}}{2a_2} \Rightarrow I_+ = 378.784, I_- = -14.784 \tag{28}$$

Since the negative root has no physical interpretation, the economically relevant equilibrium inventory is $I^* = 378.784$ (29)

This value represents the long-run inventory level that stabilizes the nonlinear production-inventory system under the chosen service target. The corresponding steady-state production rate is obtained from the feedback rule

$$u^* = u_0 + u_1I^* = 170 + 0.10(378.784) = 207.878 \tag{30}$$

The steady-state fill rate is then calculated from

$$FR^* = FR(t) = 1 - e^{-\tau I^*} \approx 0.9517 \tag{31}$$

Therefore, under the chosen parameter values, the system achieves a fill rate of approximately 95.17 percent. The exact time path of inventory is obtained from the analytical solution

$$I(t) = \frac{I_+ - CI_-e^{-\Delta t}}{1 - Ce^{-\Delta t}} \tag{32}$$

$$C = \frac{I_0 - I_+}{I_0 - I_-} = \frac{40 - 378.784}{40 + 14.784} = -6.184 \tag{33}$$

$$\text{And } \Delta = a_2(I_+ - I_-) = 1.5743 \tag{34}$$

Substituting these values into equation (32), the inventory trajectory can be evaluated for different times. The computed values are reported in Table 2.

Table 3: Time path of inventory level for $s^* = 0.80$	
Time (t)	Inventory (I(t))
0	40
1	137.694
2	296.243
3	355.654
5	377.858
10	378.784

The results show that inventory increases rapidly from its initial value and converges smoothly to the stable equilibrium level. This confirms the theoretical stability result established earlier. The nonlinear loss term prevents unlimited inventory growth, while the production rule ensures that stock is increased sufficiently to meet the higher service-induced demand.

To measure the operating burden associated with this equilibrium, the total steady-state weighted cost is approximated by

$$TC = c_p u^* + hI^* + c_s (s^*)^2$$

Substituting the computed values gives

$$TC = 2.2(207.878) + (0.9)(378.784) + 180(0.8)^2 = 913.438$$

Thus, the total weighted cost associated with the medium service policy is approximately 913.44. In order to reveal the cost–service trade-off more clearly, the same procedure is repeated for three different service targets, namely $s^* = 0.40, s^* = 0.80, s^* = 1.20$. For each service level, the equilibrium inventory, corresponding production rate, fill rate, and total cost are computed. The results are summarized in Table 3.

Table 4: Cost–service trade-off under different service targets

Service level (s^*)	Effective demand	Equilibrium inventory (I^*)	Production rate (u^*)	Fill rate (FR^*)	Total cost (TC)
0.4	118.8	309.144	200.914	0.9157	784.84
0.8	147.6	378.784	207.878	0.9517	913.44
1.2	176.4	449.136	214.914	0.9725	1086.48

Table 4 clearly indicates that as the service target increases, effective demand also rises because customers respond positively to better service. To support this higher demand and avoid shortages, the system requires a larger equilibrium inventory and a higher production rate. As a result, the fill rate improves steadily, rising from 91.57 percent at $s^* = 0.40$ to 97.25 percent at $s^* = 1.20$. However, this service improvement is achieved at the expense of a substantial increase in operating cost. The total weighted cost rises from 784.84 to 1086.48 across the three scenarios. This demonstrates the central managerial insight of the study: higher service commitments enhance customer fulfillment but impose increasing production and inventory-related costs.

A further sensitivity analysis is performed on the nonlinear inventory-loss parameter γ , which captures congestion, spoilage, or overstock burden. Since this parameter plays a major stabilizing role in the nonlinear model, its effect on equilibrium inventory and fill rate is of particular interest. Keeping the service target fixed at $s^* = 0.80$, three values of γ are examined. The results are shown in Table 5.

Table 5: Sensitivity analysis with respect to nonlinear inventory-loss parameter γ

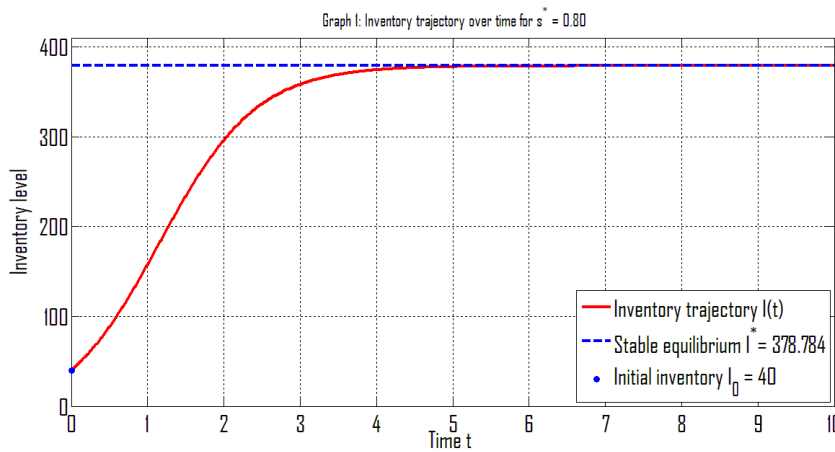
γ	Equilibrium inventory (I^*)	Production rate (u^*)	Fill rate (TC)
0.0032	469.897	216.99	0.9767
0.004	378.784	207.878	0.9517
0.0048	318.008	201.801	0.9215

The sensitivity results show that increasing the nonlinear loss parameter reduces the stable inventory level. This is consistent with the model structure, because stronger nonlinear loss penalizes large stock accumulation more heavily. Consequently, as γ rises, the fill rate declines, indicating that excessive congestion or deterioration makes it more difficult to sustain high service performance. These findings suggest that in industries with perishable products, storage congestion, or rapid obsolescence, managers must either accept lower service targets or adopt more aggressive production and replenishment policies.

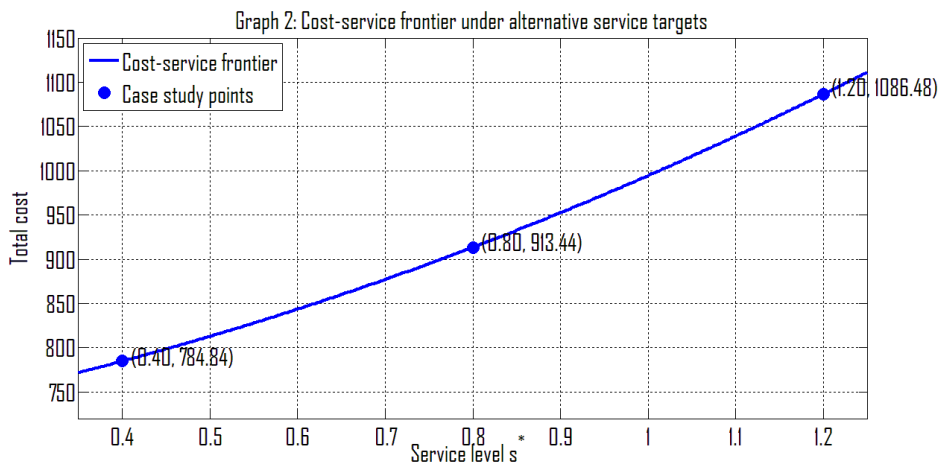
From a managerial perspective, the numerical case study confirms three important conclusions. First, the nonlinear inventory system converges to a stable equilibrium under the feedback production policy, which validates the usefulness

of the proposed modelling approach for operational planning. Second, service enhancement is not free: higher service goals directly raise inventory requirements, production effort, and total cost. Third, the nonlinear loss effect is crucial in determining how far inventory can be expanded to support customer fulfillment. Ignoring such nonlinearities may cause classical linear models to underestimate the real cost of maintaining high service levels.

The case study therefore demonstrates that the proposed model is capable of supporting production decisions in environments where customer service and inventory cost are tightly linked. By combining explicit differential-equation analysis with equilibrium computations and sensitivity analysis, the model provides a mathematically rigorous yet managerially interpretable framework for optimal production planning.

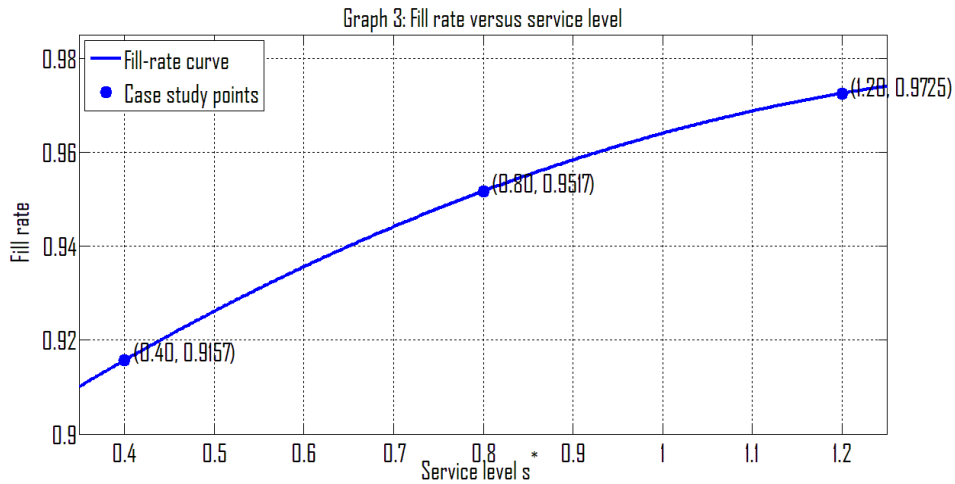


The graph (1) shows that the inventory level starts from a low initial value of 40 units and then rises rapidly during the early time periods as production exceeds depletion, indicating a strong adjustment response in the system. After this sharp increase, the curve gradually flattens and approaches the horizontal dashed line representing the stable equilibrium inventory level $I^* = 378.784$. This monotone convergence means that the inventory does not oscillate or overshoot, but instead moves smoothly toward its long-run target level, confirming the stability of the nonlinear inventory model under the chosen production policy. Economically, the figure suggests that the system is capable of restoring inventory efficiently from an initially understocked condition while avoiding excessive accumulation, so the equilibrium level acts as a sustainable balance between replenishment, demand, and inventory-related losses.

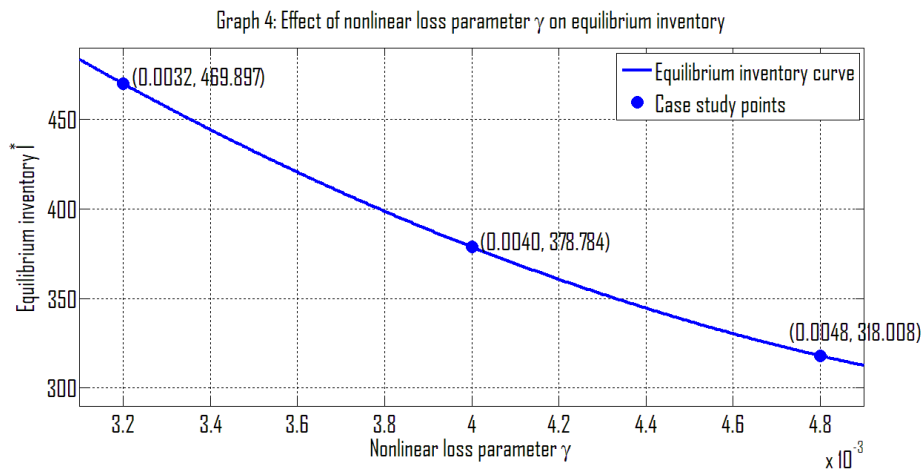


The graph (2) illustrates a clear positive relationship between service level and total cost, showing that as the service target increases from 0.40 to 1.20, the total cost also rises from about 784.84 to 1086.48. The upward-sloping shape of the curve indicates that achieving better customer service requires greater production effort, larger inventory support, and higher operating expenditure. The curve also appears convex, which suggests that the cost of improving service becomes steeper at higher service levels, meaning that moderate improvements are comparatively less expensive while very high service targets demand disproportionately larger costs. This reflects the core cost-service trade-off in production

planning, where firms must balance the benefit of improved service performance against the increasing financial burden needed to sustain it.

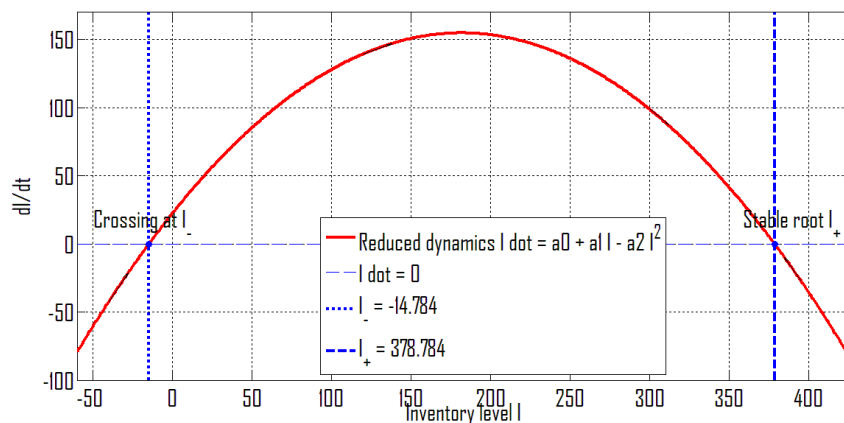


The graph (3) shows that the fill rate increases as the service level rises, which means that better service intensity improves the firm’s ability to meet customer demand. Specifically, when the service level increases from 0.40 to 1.20, the fill rate improves from about 0.9157 to 0.9725, indicating a higher proportion of demand being satisfied without shortage. However, the curve becomes flatter as service level grows, revealing diminishing marginal gains: early improvements in service produce a noticeable increase in fill rate, whereas further increases in service generate smaller additional benefits. This pattern suggests that although higher service targets enhance customer fulfillment, the incremental performance gain becomes limited at advanced service levels, which is important for managers when deciding whether the extra cost of raising service intensity is justified by the relatively smaller improvement in service performance.



The graph (4) shows a clear downward relationship between the nonlinear loss parameter γ and the equilibrium inventory level I^* . As γ increases from 0.0032 to 0.0048, the equilibrium inventory decreases from about 469.897 to 318.008, indicating that stronger nonlinear loss effects reduce the amount of inventory the system can sustain in the long run. This happens because higher values of γ represent greater deterioration, congestion, spoilage, or stock-related burden, which penalize inventory accumulation more heavily. As a result, the system settles at a lower stable inventory level. The graph therefore highlights the important managerial implication that when nonlinear losses are high, firms cannot rely on large inventory stocks to maintain service performance and must instead improve production efficiency or adopts more careful replenishment policies.

Graph 5: Phase portrait for the reduced inventory dynamics



The graph (5) represents the phase portrait of the reduced inventory dynamics by plotting the rate of change of inventory, dI/dt , against the inventory level I . The curve crosses the horizontal axis at two equilibrium points, $I_- = -14.784$ and $I_+ = 378.784$, where the inventory change becomes zero. Since the curve is above the horizontal axis between these two roots, dI/dt is positive in that region, meaning inventory increases over time; outside this region the curve is below the axis, so dI/dt is negative and inventory decreases. This behavior shows that the positive equilibrium I_+ is stable, because inventory levels near it move back toward it, while the negative root I_- is not economically meaningful in practice since negative inventory cannot occur in a real system. Thus, the graph confirms that the reduced nonlinear model has a stable long-run inventory target at $I_+ = 378.784$, toward which the system naturally converges under the given production policy.

VI. CONCLUSION AND FUTURE RESEARCH

The study concludes that nonlinear mathematical modelling provides an effective framework for understanding optimal production planning under cost–service trade-offs. By integrating service-sensitive demand, backlog recovery, nonlinear deterioration, and production feedback into a continuous-time inventory system, the model demonstrates that inventory policy cannot be designed solely on the basis of cost minimization or service maximization in isolation. The analytical solution of the reduced Riccati-type model establishes the existence of a stable positive equilibrium, while the numerical case study confirms that inventory converges smoothly to this long-run target under the chosen policy. The results further show that increasing service commitments improves fill rate and customer satisfaction but also leads to higher equilibrium inventory, stronger production requirements, and greater total operating cost. In addition, the sensitivity analysis highlights that stronger nonlinear loss effects reduce the sustainable inventory level and make high service targets more difficult to maintain. Overall, the paper shows that the most effective production policy is one that achieves a sustainable compromise between customer fulfillment and operational efficiency, and it provides a useful foundation for future extensions involving stochastic demand, multi-item systems, preservation investment, machine disruptions, or other real-world complexities.

REFERENCES

- [1]. Afrouzy Z.A., Nasser S.H., Mahdavi I. (2016a): “A genetic algorithm for supply chain configuration with new product development”, *Computers & Industrial Engineering*, 101:440–454.
- [2]. Afrouzy Z.A., Nasser S.H., Mahdavi I., Paydar M.M. (2016b): “A fuzzy stochastic multi-objective optimization model to configure a supply chain considering new product development”, *Applied Mathematical Modelling*, 40(17–18):7545–7570.
- [3]. Albey E., Norouzi A., Kempf K.G., Uzsoy R. (2015): “Demand modeling with forecast evolution: an application to production planning”, *IEEE Transactions on Semiconductor Manufacturing*, 28(3):374–384.
- [4]. Alejo-Reyes A., Sánchez-Partida D., Cano-Olivos P., Martínez-Flores J.L. (2023): “A Mathematical Model for an Inventory Management and Order Quantity Allocation Problem with Nonlinear Quantity Discounts and Nonlinear Price-Dependent Demand”, *Axioms*, 12(6):547.
- [5]. Alkahtani M., Ghosh S.K., Khan M.A.A., et al. (2022): “Mathematical Modelling of Inventory and Process Outsourcing for Optimization of Supply Chain Management”, *Mathematics*, 10(7):1142.
- [6]. Angelus A., Zhu W. (2017): “Looking upstream: Optimal policies for a class of capacitated multi-stage inventory systems”, *Production and Operations Management*, 26(11):2071–2088.

- [7]. Aouam T., Ghadimi F., Vanhoucke M. (2021): “Finite inventory budgets in production capacity and safety stock placement under the guaranteed service approach”, *Computers & Operations Research*, 131:105266.
- [8]. Bhuniya S., Pareek S., Sarkar B. (2021): “A supply chain model with service level constraints and strategies under uncertainty”, *Alexandria Engineering Journal*, 60(6):6035–6052.
- [9]. Bilginer Ö., Erhun F. (2015): “Production and sales planning in capacitated new product introductions”, *Production and Operations Management*, 24(1):42–53.
- [10]. de Kok T. (2015): “Buffering against uncertainty in high-tech supply chains”, in *2015 Winter Simulation Conference (WSC)*, 1–10.
- [11]. De Kok T. (2018): “Inventory management: Modeling real-life supply chains and empirical validity”, *Foundations and Trends in Technology, Information and Operations Management*, 11(4):343–437.
- [12]. De Kok T., Grob C., Laumanns M., Minner S., Rambau J., Schade K. (2018): “A typology and literature review on stochastic multi-echelon inventory models”, *European Journal of Operational Research*, 269(3):955–983.
- [13]. Eruguz A.S., Sahin E., Jemai Z., Dallery Y. (2016): “A comprehensive survey of guaranteed-service models for multi-echelon inventory optimization”, *International Journal of Production Economics*, 172:110–125.
- [14]. Ghadimi F., Aouam T., Uzsoy R. (2023): “Safety stock placement with market selection under load-dependent lead times”, *IISE Transactions*, 55(3):314–328.
- [15]. Glock C.H., Grosse E.H. (2015): “Decision support models for production ramp-up: a systematic literature review”, *International Journal of Production Research*, 53(21):6637–6651.
- [16]. Graves S.C., Schoenmeyr T. (2016): “Strategic safety-stock placement in supply chains with capacity constraints”, *Manufacturing & Service Operations Management*, 18(3):445–460.
- [17]. Hrabec D., Hvattum L.M., Hoff A. (2022): “The value of integrated planning for production, inventory, and routing decisions: A systematic review and meta-analysis”, *International Journal of Production Economics*, 248:108468.
- [18]. Huh W.T., Janakiraman G., Nagarajan M. (2016): “Capacitated multiechelon inventory systems: Policies and bounds”, *Manufacturing & Service Operations Management*, 18(4):570–584.
- [19]. Jabeur M.H., Mahjoub S., Toublanc C., Cariou V. (2024): “Optimizing integrated lot sizing and production scheduling in flexible flow line systems with energy scheme: A two level approach based on reinforcement learning”, *Computers & Industrial Engineering*, 190:110095.
- [20]. Ji Q., Sun L., Hu X., Hou J. (2016): “Optimal policies of a two-echelon serial inventory system with general limited capacities”, *International Journal of Production Research*, 54(20):6142–6155.
- [21]. Kohlmann P., Sahling F. (2024): “A flexible planning approach for integrated lot sizing and rework planning with random proportion of defective products”, *International Journal of Production Research*, 62(19):6961–6978.
- [22]. León R., Miranda P.A., Cabrera G., et al. (2024): “An Inventory Service-Level Optimization Problem for a Multi-Warehouse Supply Chain Network with Stochastic Demands”, *Mathematics*, 12(16):2544.
- [23]. Meistering M., Stadtler H. (2017): “Stabilized-cycle strategy for capacitated lot sizing with multiple products: Fill-rate constraints in rolling schedules”, *Production and Operations Management*, 26(12):2247–2265.
- [24]. Tavaghof Gigloo D., Minner S. (2021): “Planning approaches for stochastic capacitated lotsizing with service level constraints”, *International Journal of Production Research*, 59(17):5087–5107.
- [25]. Voas J., Kshetri N., DeFranco J.F. (2021): “Scarcity and global insecurity: the semiconductor shortage”, *IT Professional*, 23(5):78–82.
- [26]. Woerner S., Laumanns M., Wagner S.M. (2018): “Simulation-based optimization of capacitated assembly systems under beta-service level constraints”, *Decision Sciences*, 49(1):180–217.