International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Optimizing AI-Driven Replenishment Policies Using Response Surface Methodology: A Multi-Echelon Simulation Study

Ibrahim H. Aldali

Abstract

AI doesn't always translate to better inventory economics, even if the demand forecasts it generates are better, if the rules engineered behind replenishment models are still "legacy" and based on legacy planning conditions. This paper proposes a sequential response surface methodology (RSM) protocol to provide an approach for policy tuning after forecasting in a four-echelon Saudi fast-moving consumer goods (FMCG) simulation. FMCG Co. is a generic placeholder rather than a registered company, and all case data are synthetic, generated using realistic assumptions for the Saudi FMCG sector. The analysis is based on 1,248 weekly demand observations for the three regions and the four SKU families, a static LightGBM forecasting engine, and a periodic-review replenishment policy. These five controllable factors were screened using a Resolution V fractional factorial design, steepest descent, a three-factor central composite design, quadratic response modelling, constrained optimization, Derringer-Suich desirability, and held-back confirmation. For the cost-led response-surface search, safety-stock multiplier, review period and demand-signal smoothing were kept. Total cost showed a significant center-point curvature (F = 80.49, p = .0029) and the fitted quadratic models accounted for over 99.9% of the variance in total cost, fill rate, and bullwhip with no significant lack-of-fit tests. The strict model-based constrained solution (k = 1.775, R = 5.095 days, alpha = 0.343) predicted an annual cost of SAR 46.146 million, a 96.000% fill rate, and a bullwhip ratio of 1.595. The desirability solution with a balance of the three factors (k = 2.253, R = 3.520 days, alpha = 0.3635) predicted better service and stability at higher cost. A previous candidate, based on cost, was rejected as infeasible, while the only directly confirmed feasible candidate was able to deliver 97.605% fill and 1.546 bullwhip at SAR 49.706 million. The results indicate that forecast modernization and replenishment-policy modernization are two different interventions and that the policy re-tuning is important to translate the forecasting capability into system-level performance while maintaining interpretability and awareness of constraints.

Keywords: artificial intelligence; replenishment policy; response surface methodology; multi-echelon inventory; bullwhip effect; Saudi Arabia; FMCG; simulation optimization

1. Introduction

AI is now becoming a part of supply chain planning, using machine learning (ML) to enhance demand sensing, forecasting, process optimization, and decision support. Recent reviews indicate that AI applications have evolved to cover various stages of the supply chain and have evolved from stand-alone predictive applications to operational decision systems integrated into operations (Teixeira et al., 2025). This shift raises a more subtle but important issue, which is not always apparent, other than in the model selection phase, that being able to provide more information for prediction improves replenishment decisions following the forecast.

Demand forecasting is an obvious place to start with AI, given that demand in the retail and FMCG sectors has nonlinearities, event effects, different product behaviors, and price, promotion, calendar, and exogenous factor interaction. Generally, studies using machine learning and deep learning show promising levels of predictive ability, with the performance of specific classes of models depending on the characteristics of the data, feature engineering, horizon, and design of the evaluation. Additionally, empirical evidence in retailing demonstrates that ensemble models like trees and recurrent deep learning architectures can effectively leverage vast amounts of high-frequency demand data, highlighting the complementary role of advanced forecasting in operations when ample data exists (Nasseri et al., 2023).

Forecast accuracy is not an operational goal, however. Demand plans take into account service expectations, replenishment constraints, and commercial priorities, which may not always be explicitly defined, along with the statistical forecast, and practitioners can easily mix them all up when making planning judgments (Fahimnia et al., 2023). The second layer of decision-making therefore has to be that which is necessary to translate the forecast information into safety stock, review period, order-up-to point, smoothing of signals and information flows between echelons.

This is particularly significant in multi-echelon networks. An aggressive policy with respect to new information may decrease the amount of local forecast error and increase the variability of upstream orders, and a conservative policy may stabilize orders at the cost of service or inventory. Multi-echelon inventory optimization is naturally a compromise between holding cost, shortage risk, and service; and more recent research considers the network as a whole and not a single stock point as the decision unit (Geevers et al., 2024). This shows that service-constrained inventory optimization is managerially unacceptable, even if it is a low-cost solution, if it fails to meet minimum customer-service requirements (Liu and Nishi, 2024).

This is a beneficial stress test for this issue, as the Saudi FMCG environment is subject to the influence of events in the Islamic calendar, heavy pilgrim traffic, regional variations, and temperature influences on chilled products. Data-intensive decision support has a much broader relevance in event-driven national operations, as Saudi Arabia has already implemented AI-based systems to handle the complexity of the Hajj season (Abalkhail and Al Amri, 2022). These features can generate non-stationary demand peaks, which can interact with replenishment settings, even though the forecast engine is not changed.

This study aims to find the optimal value of the controllable parameters for a given AI demand-forecasting engine that will minimize the total supply chain cost with a weighted retail fill rate of 96% or higher, and a bullwhip effect between plant and retail to be less than 1.60. However, instead of comparing to other AI algorithms, the following study considers the forecasting engine as already a proven information mechanism and optimizes the replenishment policy that uses the output of the forecasting engine. This option emphasizes the following analysis on a typical implementation challenge: an organization can upgrade its forecasting system, but still use policy parameters carried over from a previous planning process.

The methodology used is a combination of discrete-event simulation, sequential design of experiments, and RSM. The protocol includes five controllable factors, a steeper descent direction towards a low-cost region, second-order response surfaces of cost, service, and bullwhip, and explicit constraints and multi-response desirability. It's not as if AI forecasting on its own is an improvement for supply chains, though. It's a transferable experimental setup that can be used to determine if the policy of operation of a fixed AI forecast is calibrated to the new information environment.

The empirical context is highly artificial. The FMCG Co. is a methodological composite Saudi manufacturing company with a distribution network, and the numerical data is simulated using realistic operating assumptions, not from a registered company. The key takeaway with respect to the interpretation is the following: The numerical findings should be interpreted as describing the behavior of the proposed tuning protocol in a realistic case design, not as performance auditing of a specific Saudi firm.

2. Literature Review and Research Gap

2.1 AI-supported demand forecasting and the post-forecast decision layer

Research on supply chain forecasting shifted from the traditional univariate models to machine learning and deep learning approaches that can take advantage of large feature spaces and nonlinear relationships. Review of the literature reveals that ensemble learners, neural network models, and hybrid approaches are increasingly being used for demand forecasting, and the success of these models is influenced by the data-generating environment and the specific application of the forecast (Douaioui et al., 2024). External factors like cannibalization, weather, price, and holiday effects can have a significant contribution to the accuracy of the forecast when compared with a historical sales-based model in the case of FMCG products (Farizal et al., 2021).

The value of improved forecasts only really becomes apparent after they are turned into inventory decisions. Theodorou et al. (2023) present a data-driven inventory-control framework that exhibits a joint influence of the demand patterns and replenishment-policy parameters on the cost and service results, which is an illustration of how machine learning can be used to support the optimization of inventory-control policies in addition to forecasting. This is critical, since the same point forecast may lead to widely divergent levels of inventories with varying safety-stock multipliers, review periods, or shortage-cost structures.

There is also a behavioral implication of the difference between a demand forecast and a demand plan. Experimental findings from FMCG-informed planning tasks indicate that service-level information can influence forecasters' judgments even when the task itself is not to create an operational plan, but to forecast the demand (Fahimnia et al., 2023). Thus, a technically sophisticated forecast is no substitute for the need to articulate explicit policy design; if anything it makes the connection between the forecast, service levels, and replenishment rules even more critical.

2.2 Multi-echelon inventory control, service constraints, and bullwhip

The multi-echelon inventory problem is challenging due to the alteration of the state and uncertainty of the other echelons of the network as a result of decisions made at one echelon. Geevers et al. (2024) demonstrate that integrated multi-echelon policies perform better than benchmark policies, both with linear versus divergent network structure, and with more general network structure, and highlight the computational challenge posed by the stochastic demand, lead times, and network dimensionality. The managerial implication is that local optimization can lead to an incorrect solution if inventories are dynamically linked upstream and downstream.

Some of the low-cost policies are not feasible due to service constraints. Liu and Nishi (2024) develop a multi-echelon inventory problem to minimize cost and maintain a desired service level with the help of data-driven surrogate modeling and evolutionary optimization. The results they obtained confirm a principle applied in the present study: that a minimal cost is desirable within a technically achievable service area, not to optimize first and then to determine if service failure results.

Closed-form models are often limited in their applicability due to the interaction between networks and other non-linear operating rules, and simulation can be very useful in such situations. Sbai and Berrado (2023) claim that simulation is the only approach that allows decision makers to investigate multi-echelon inventory strategies in terms of product availability, customer service, and inventory costs without having to go through the same analytical simplifying assumptions that are imposed by many mathematical formulations. This makes it a good experimental lab for simulation-based tuning of the policy where the network structure and replenishment logic are explicitly modeled.

The bullwhip effect brings a stability criterion that is not covered by fill rate or cost. Control-theoretic supply chain models have revealed that order variability can be amplified between supply chain stages and that a proper design of information and control mechanisms can dampen the effect of order variability amplification (Papanagnou, 2022). Simulation results also show that both the structure and characteristics of the network, as well as the variability parameters, can have a significant impact on bullwhip dynamics, and thus, the dynamic stability should be directly studied rather than inferred through inventory costs (Fussone et al., 2024).

2.3 Response surface methodology as an interpretable simulation-optimization framework

Response surface methodology provides an orderly approach to progress from factor screening to local optimization by means of statistically estimable polynomial models. The use of response surfaces is still being explored in modern research on composite designs, with an emphasis on the need for designs that efficiently estimate response surfaces, interactions, and curvature without extensive enumeration of the factor space (Athanasaki et al., 2024). This is appealing for simulation experiments as there might be many stochastic replications required for each design point.

One important methodological constraint is that, before interpreting the stationary point or predicted optimum of the fitted response model, the response model must be adequate. RSM research focuses on a test for the significance of the model, the adjusted explanatory power, and the non-significant lack of fit for a second-order model, which can be used for optimization (Rheem, 2023). Even though R-squared is high, it should not be used as a criterion for the adequacy of variance assumptions or error behavior, and thus should be seen as complementary.

Supply chain policy tuning is a "multi-response" situation; costs can increase while service and stability increase, or costs can go down, but service and stability go up. Desirability-function approaches transform each response into a common scale of 0-1 and sum them up to get a composite objective, which allows for a transparent balancing of competing engineering responses (Madić and Marinković, 2024). In a replenishment context, this enables decision makers to compare a cost minimization solution with one that strictly respects service and bullwhip thresholds, with a more balanced compromise that knowingly trades off some cost for better service and stability.

2.4 Research gap and analytical contribution

Although there is literature on AI forecasting, multi-echelon inventory optimization, simulation, and supply chain dynamics, they are mostly considered as isolated interventions. Forecasting studies are usually oriented towards forecasting accuracy, while inventory studies focus on optimizing the inventory policy given the assumed demand information. The problem of practical implementation discussed here lies at the intersection of these two extremes: Once the firm has deployed an AI forecasting engine, are the old replenishment parameters still valid for the new signal quality and responsiveness?

This study fills this gap by considering the policy parameters as experimental factors and the AI engine as fixed. The resulting design is deliberately made to be open to interpretation. It estimates factor effects, interactions, curvature, stationary-point geometry, feasibility boundaries, and confirmation performance, rather than returning a black-box optimum. This structure enables the optimization to be auditable by planners, who should be able to understand how and why a recommended policy works rather than just take a parameter vector offered to them by a machine.

3. Context and Methodology

3.1 Synthetic Saudi FMCG case and data provenance

FMCG Co. is a composite case organization representing a Saudi manufacturer and distributor of ambient and chilled fast-moving consumer goods. The synthetic network consists of one plant in Jeddah Industrial City 2, three regional distribution centers (RDCs) located in Riyadh, Jeddah, and Dammam, and retail clusters of some 4,100 outlets in the Central, Western, and Eastern provinces. The modelled network is serial-divergent with four echelons and nine nodes, and it processes around 3.1 million cases per year with the base-year operating cost of the supply chain being approximately SAR 51.4 million.

There are 1,248 weekly observations in the demand dataset, which is created as 104 weeks of demand for three regions for four SKU families. The four families are dairy, bottled water, dates, confectionery, and chilled juice. The actual demand and an intentionally incomplete legacy forecast are included. The legacy series is the series used to represent the forecasting problem and case-specific error; the fixed AI engine is an exogenous case condition and is not re-trained or benchmarked during the analysis.

The simulation is based on Saudi-specific assumptions that introduce non-stationarity to the system: Ramadan and Eid move through the Gregorian calendar, the demand surge in the Western Province due to the Hajj, chilled demand and holding cost vary by season, and the lead time for ingredients imports is longer and more variable. These are not statistics from a specific company, but rather assumptions used in the model for the synthetic case. An additional benefit of the design is distinguishing between the economics of stockout in the modern trade and traditional trade, thereby avoiding treating the service performance as economically homogeneous.

Table 1: Case structure and fixed simulation settings

Item	Specification
Entity	FMCG Co. (generic placeholder; synthetic case)
Network	4 echelons; 9 nodes; serial-divergent topology
Manufacturing/distribution	1 Jeddah plant; 3 RDCs: Riyadh, Jeddah, Dammam
Retail coverage	Approximately 4,100 outlets across Central, Western, Eastern provinces
Demand history	104 weeks x 3 regions x 4 SKU families = 1,248 observations
Fixed forecasting engine	LightGBM gradient boosting; 168 features; 12-week horizon
Replenishment policy	Periodic review (R, S) at stocking nodes
Simulation horizon	156 weeks; first 26 weeks discarded as warm-up
Replications	30 per design point using common random numbers

Item	Specification
Demand input	Empirical bootstrap from 104-week history
Platform	Python 3.11 / SimPy 4.1
Transport capacity	Uncapped (relaxed assumption)

3.2 Experimental factors and response definitions

There were five candidate factors, with each being a continuous, controllable policy setting. These five factors were: A=the safety-stock multiplier k, B=the forecast-retraining interval, C=the periodic review interval R, D=the demand-signal smoothing parameter alpha, and E=the information-sharing lag. Model architecture was intentionally kept the same, as changing the type of model would transform the study from a continuous response-surface optimization to a model comparison experiment, which is not of interest here.

Optimization of three responses. Y1 represented the annual total supply chain costs, including the costs of holding, ordering, transport, stockout, expediting, and model-retraining compute cost (in SAR million). Retail fill rate, volume-weighted, at least 96.0% was volume-weighted retail fill rate (Y2). The plant-versus-retail bullwhip ratio, which is based on relative variability of order and demand, was capped at 1.60 or less. A third response can be used to avoid the optimization from suggesting a cheap policy as a good policy if its service or dynamic stability is not acceptable.

Table 2: Candidate factors and optimization responses

Code	Variable	Range/unit	Role or constraint
A	Safety-stock multiplier (k)	0.50-3.50	Direct service-cost lever
B	Forecast retraining interval	1-30 days	Adaptivity / compute-cost lever
C	Review period (R)	1-14 days	Replenishment frequency
D	Demand-signal smoothing (alpha)	0.10-0.90	Signal responsiveness/bullwhip lever
E	Information-sharing lag	0-7 days	Inter-echelon coupling
Y1	Total supply chain cost	SAR million/year	Minimize
Y2	Weighted retail fill rate	Percent	Maximize; >= 96.0%
Y3	Plant-versus-retail bullwhip	Ratio	Minimize; <= 1.60

A-E are candidate factors. Y1-Y3 are responses. Factor levels were coded for DOE/RSM estimation.

3.3 Sequential experimental design

The analysis was done in four phases. A 2^(5-1) Resolution V fractional factorial screening design was adopted for Phase 1, consisting of 16 factorial points and 4 center points. The main effects on all three responses were estimated; a cost-led screening rule was applied to determine factors to be sought. The center points also gave a clear test for curvature prior to moving on to a second-order design.

The direction of steepest descent from the 1st order Y1 cost model was followed in Phase 2. The path had 9 points, and there were 10 replications of each point in the experiment provided. This phase was not to state an ultimate "optimum" but simply to move the design to a cheaper neighborhood before a quadratic surface is fit.

The three-factor central composite design (CCD) with axial distance 1.682 was used in Phase 3. The design consisted of 20 design points comprising 8 factorial points, 6 axial points, and 6 center points. The composite designs are seen today as efficient designs for the estimation of the second-order terms and interactions required to describe the local curvature (Athanasaki et al., 2024). The mean of 30 simulation replications under common random numbers was used for both Phase 1 and Phase 3 designs, and the replications' standard deviations were stored as uncertainty assessments.

In Phase 4, there were six confirmation points held back, with three of these being replications of a prior nominated cost-oriented candidate, one interior check, an incumbent baseline policy, and an alternate candidate. These points have not been used to fit the Phase 3 surfaces. They were responsible for testing the prediction intervals, auditing feasibility, and estimating the differences from the incumbent prediction interval.

3.4 Statistical analysis and optimization

The coefficient of variation (CV = standard deviation/mean) was used to provide a summary of the demand variability by region and SKU. The mean absolute error, root mean squared error, bias, weighted absolute percentage error (WAPE), and the MAPE excluding zero actual demand were used to evaluate the performance of legacy. WAPE was highlighted due to the presence of one observation with zero demand, and aggregate proportional error is more stable than ordinary MAPE in that case.

The results of screening were based on ordinary least squares main-effect models. All factors were assessed both statistically and for economic effect size for the cost-led search, and the selected factors were then taken forward into the quadratic model. Centre-point curvature was performed to compare the mean of the factorial points with the mean of the center runs with pure error based on center-point replications. The formal criteria used to fit a second-order surface was the presence of a significant curve in the region-seeking response.

In the CCD, three full quadratic models were estimated for Y1, Y2 and Y3, one for each of the three coded factors x1 (safety-stock multiplier), x2 (review period) and x3 (smoothing alpha). Model adequacy was determined using R-squared, overall F tests, lack-of-fit tests using pure error, Shapiro-Wilk test for normality of the model residuals, Breusch-Pagan test for variance of the model, and Durbin-Watson statistics. It is important to note that RSM model adequacy should be assessed beyond just the high explanatory power, as lack of fit and residual behavior should be checked before the interpretation of the optimized model (Rheem, 2023).

Classification of each stationary point based on the eigenvalues of the quadratic form was done by Canonical analysis. The CCD bounds were then used to find a strictly constrained solution by minimizing the predicted Y1 with the constraints that $Y2 \geq 96.0$, $Y3 \leq 1.60$. Individual desirability functions were specified separately for cost, fill rate, and bullwhip, which were to be fully desirable at or below SAR 45.5 million, 96%, and 1.45, respectively, and unacceptable at SAR 51.4 million, 99%, and 1.60 or more, respectively. The composite desirability was computed as a geometric mean of the three individual desirabilities, which is done when considering competing multi-response objectives (Madić and Marinković, 2024).

The experimental workbook provided was used for all analyses, which were performed in Python. The notebook analyzes the means and standard deviations of the generated design points, but does not reconstruct the simulation engine that generated the events because the raw event stream is not stored in the workbook. Thus, model optima that cannot be referred to a held-back run are listed as predictions and will not be said to be directly confirmed unless a new simulation is performed.

4. Results

4.1 Demand variability and legacy forecast performance

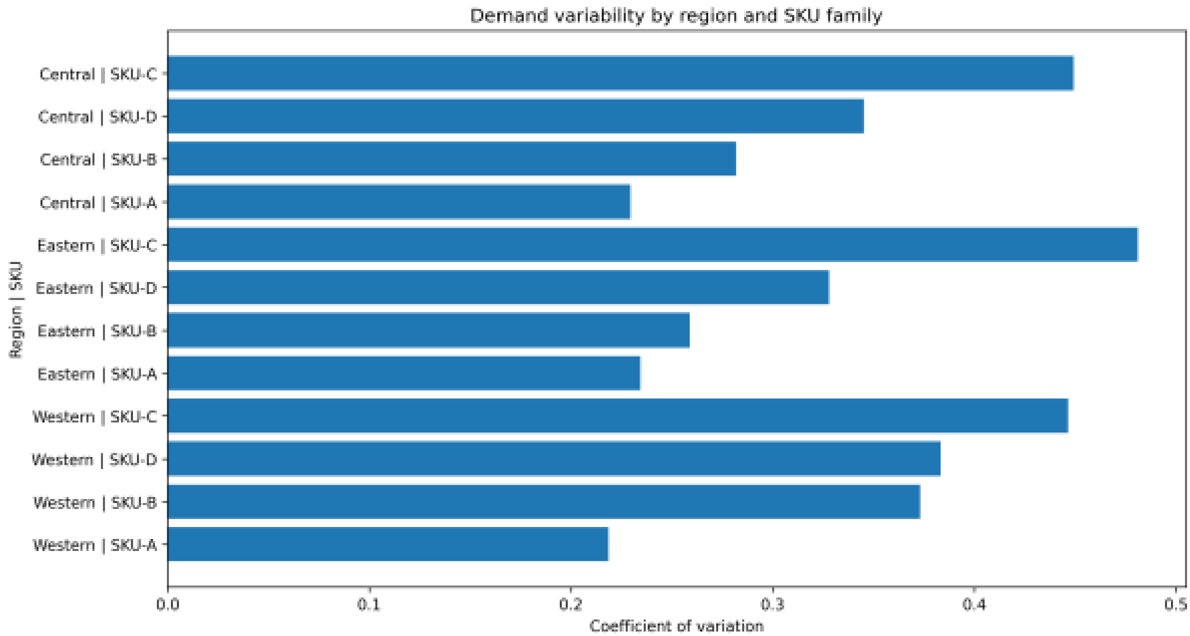
The data set was found to be basically complete, with all 1248 observations of demand having actual and legacy forecasts and one zero demand observation. There were significant variations in demand variability by SKU family. The family with the most variability in all three regions was dates and confectionery (SKU-C) with CV values ranging between 0.447 and 0.481. CV values ranged from 0.219 to 0.234 for long-life dairy (SKU-A), which was the most stable. Western demand was the most volatile regional portfolio on average, across SKU families, compared to Central (0.327) and Eastern (0.326).

Table 3: Demand variability and event-specific legacy forecast error

Region	SKU family	CV	Event	WAPE (%)	Forecast bias (cases)
Central	SKU-A dairy	0.230	Baseline	22.58	+39
Central	SKU-B water	0.282	Ramadan	28.33	-3,261
Central	SKU-C dates/confectionery	0.449	Eid al-Fitr	27.21	-3,585
Central	SKU-D chilled juice	0.345	Hajj season	24.97	-1,758
Eastern	SKU-A dairy	0.234	Summer peak	22.53	+247
Eastern	SKU-B water	0.259	Back-to-school	21.42	+165
Eastern	SKU-C dates/confectionery	0.481	National Day	19.41	-879
Eastern	SKU-D chilled juice	0.328			
Western	SKU-A dairy	0.219			
Western	SKU-B water	0.373			
Western	SKU-C dates/confectionery	0.447			
Western	SKU-D chilled juice	0.383			

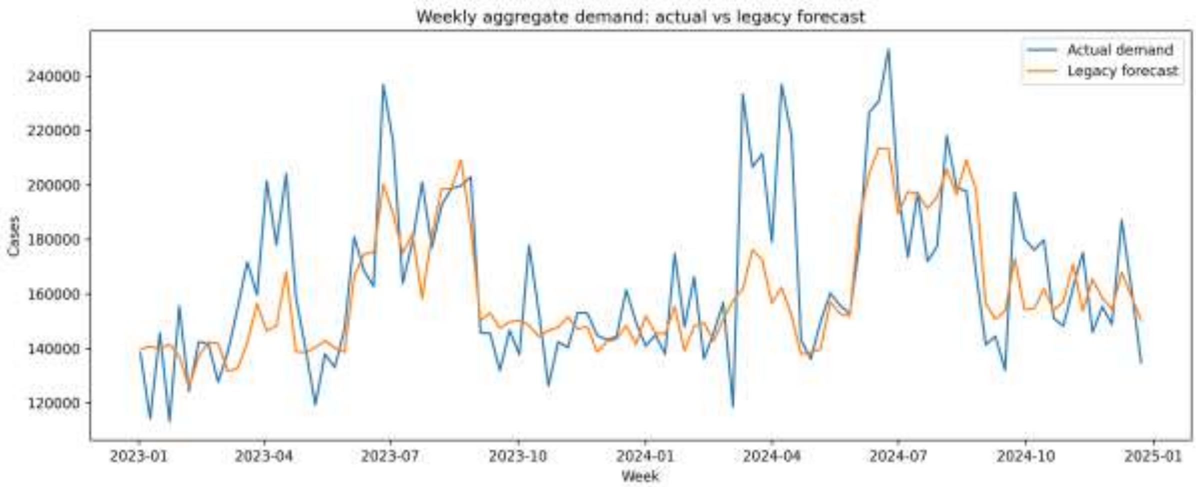
Bias is forecast minus actual demand; negative values indicate under-forecasting. The event error columns summarize pooled observations for the named event.

Figure 1: Demand variability by region and SKU family



As can be seen from Figure 1, there are products which are volatile, and there are products which are volatile regionally. The one SKU with the highest single CV was Eastern SKU-C (0.481); other SKUs with high CVs were also Western SKU-B and SKU-D. This heterogeneity enables the tuning of the policy level in the system, instead of relying on a single generic variability profile to represent the network.

Figure 2: Weekly aggregate actual demand versus legacy forecast



In the aggregate time series in Figure 2, the legacy method has under-forecasted on several occasions. The visual trend is confirmed by the results at the level of events (WAPE): During Ramadan, WAPE increased to 28.33%, during Eid al-Fitr to 27.21%, and during the Hajj season to 24.97% compared to the baseline of 22.58%. The biases during Ramadan, Eid and Hajj are all negative, suggesting systematic under-prediction in the times when service biases are most important.

4.2 Phase 1 screening and centre-point curvature

The safety-stock multiplier effect (effect = 5.979 SAR million, $p < .001$), followed by review period (3.930, $p = .0013$) and smoothing alpha (1.910, $p = .0580$) had the largest screening effects on total cost. Forecast retraining interval and information sharing lag were not statistically significant, having small

Y1 effects of 0.539 SAR million and 0.496 SAR million, respectively. The region search then kept A, C, and D because they were more affordable.

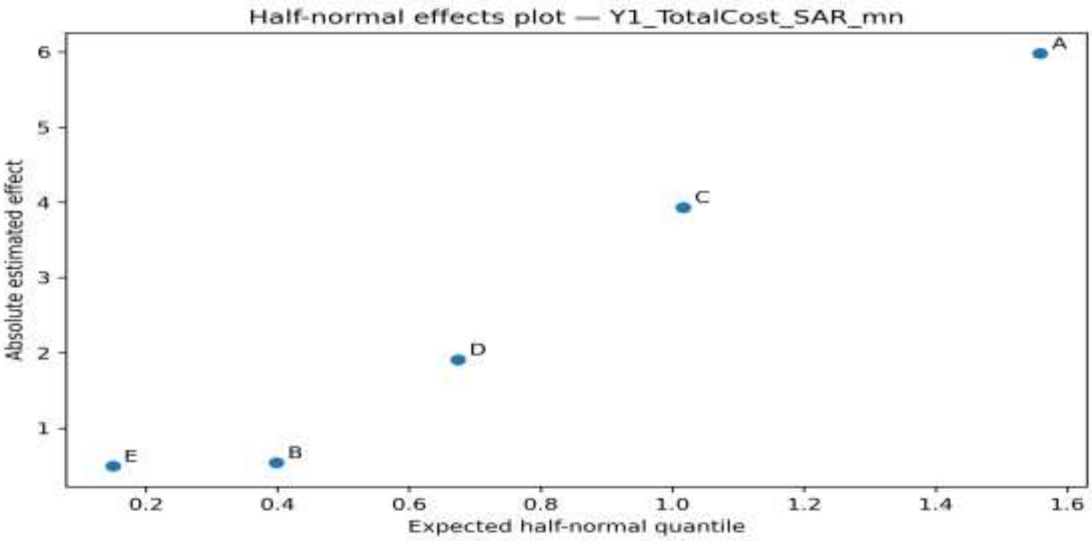
A qualification is needed for the multi-response screening results. Information-sharing lag E was economic for Y1, but was statistically significant when negatively affecting Y3 bullwhip (effect = -0.382, $p < .001$) and marginally affecting Y2 fill rate ($p = .058$). So don't think of E as inconsequential in the global arena. This was not included in the Phase 3 surface as the sequential region search was focused on cost and only a CCD with A, C, and D was provided. The following bullwhip optimization is thus based on E being fixed.

Table 4: Phase 1 main effects across the three responses

Factor	Y1 effect	p	Y2 effect	p	Y3 effect	p	Phase 3 status
A: safety stock	+5.979	<.001	+4.666	<.001	-0.219	.0059	Retain
B: retraining interval	+0.539	.559	-0.340	.257	+0.062	.347	Drop
C: review period	+3.930	.0013	-3.230	<.001	+0.535	<.001	Retain
D: smoothing alpha	+1.910	.0580	-1.660	<.001	+0.668	<.001	Retain
E: information lag	+0.496	.591	-0.605	.0582	-0.382	<.001	Hold constant*

Y1 is SAR million/year, Y2 is percentage points, and Y3 is a ratio effect. *E was small and non-significant for cost but significant for bullwhip; its exclusion is a cost-led sequential-design decision, not evidence of global irrelevance.

Figure 3: Half-normal screening plot for total supply chain cost

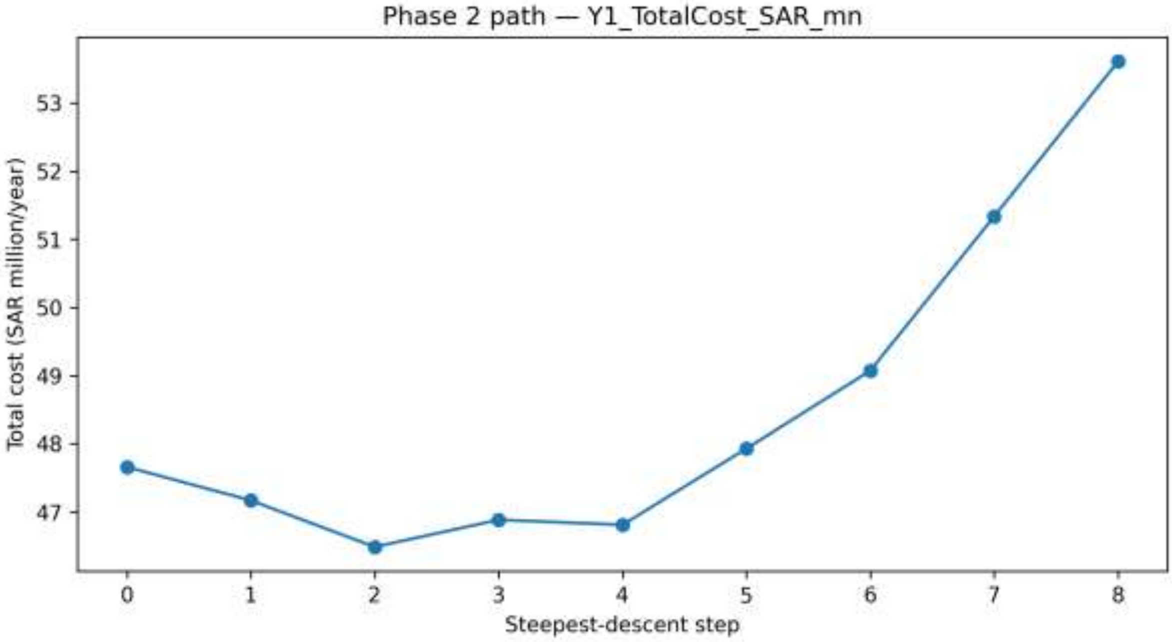


The half-normal pattern in Figure 3 confirms the screening decision regarding the cost response based on the magnitude: A, C, and D are the dominant group. Factor point testing subsequently demonstrated good curvature, as the mean cost for Y1 (SAR 49.385 million) was significantly different from the center (SAR 47.878 million), with $F = 80.49$ and $p = .0029$. Curvature was not significant for Y2 ($p = .7785$) or Y3 ($p = .9435$). Region seeking based on the cost objective led to the use of a second-order response surface as there was a clear statistical basis.

4.3 Steepest descent and movement into an infeasible low-cost region

The direction of steepest descent (DOSD) of the first-order cost model was approximately -0.25 in coded safety stock, -0.164 in review period, and -0.080 in smoothing alpha per principal step. The experimental path obtained in this observation was very close to these proportions. The cost started at SAR 47.659 million at the starting point, and reduced to SAR 46.486 million at Step 2, but increased gradually until SAR 53.608 million at Step 8.

Figure 4: Phase 2 steepest-descent path for annual total supply chain cost



The main conclusion is that the cost valley was not a viable service area. Step 2 results showed a lack of both managerial constraints: fill rate 94.835% and bullwhip 1.7695. The bullwhip was continued to be reduced to around 1.60, but the fill rate decreased significantly to 87.851% in Step 8. The Phase 2 path thus illustrates why it is not possible to take a single-response cost search as the definitive policy: the system can become cheaper locally by washing out service performance more rapidly than can be achieved by making the system more stable.

4.4 Quadratic response models and diagnostics

The explanatory power of all three models obtained by the CCDs was very high. The R-square values for cost, fill rate, and bullwhip were 0.99965, 0.99945, and 0.99977, respectively, while the adjusted R-square values were greater than 0.9989 for each of these. Overall model F tests were highly significant (all $p < .001$), and formal lack-of-fit tests were non-significant for Y1 ($p = .9455$), Y2 ($p = .9115$), and Y3 ($p = .1183$). Results show that the second-order surfaces are able to closely approximate the means of the 20-point CCD within the range sampled.

Table 5: Quadratic model adequacy and residual diagnostics

Response	R^2	Adj. R^2	RMSE	LOF p	Shapiro p	BP F p	DW
Y1 cost	0.99965	0.99934	0.1064	.9455	.0388	.8601	2.358
Y2 fill	0.99945	0.99896	0.0901	.9115	.7068	.5333	2.462
Y3 bullwhip	0.99977	0.99957	0.00890	.1183	.9755	.0057	1.656

LOF = lack of fit; BP = Breusch-Pagan; DW = Durbin-Watson. Y1 shows a mild residual-normality warning, whereas Y3 shows evidence of non-constant variance.

Not all the residual diagnostics were perfect, and so were not all thrown away in interpretation. Residuals from Y1 had Shapiro-Wilk p = .0388, which is a mild violation of normality, but there was no indication of heteroskedasticity in the Breusch-Pagan test (p = .8601). The residuals for Y3 were fairly normally distributed (p = .9755), but the Breusch-Pagan F test suggested non-constant variance (p = .0057). There was no such warning for Y2. The high fit and non-significant lack-of-fit results confirm the usefulness of the model for prediction within the CCD, while the variance warning for Y3 suggests that inferential precision of the coefficients should be viewed with care, and that held-back confirmation is even more important to place greater emphasis on the usefulness of the model for prediction within the CCD.

In coded units, the fitted response equations were:

$$\begin{aligned}\mathrm{Y1} &= 47.496 + 3.172x1 + 1.803x2 + 0.903x3 + 2.662x1^2 + 1.178x2^2 + 0.653x3^2 \\ &\quad + 1.271x1x2 - 0.626x1x3 + 0.491x2x3\end{aligned}$$

$$\begin{aligned}\mathrm{Y2} &= 95.408 + 2.363x1 - 1.548x2 - 0.842x3 - 1.318x1^2 - 0.452x2^2 - 0.117x3^2 \\ &\quad + 0.508x1x2 - 0.169x1x3 - 0.051x2x3\end{aligned}$$

$$\begin{aligned}\mathrm{Y3} &= 1.831 - 0.123x1 + 0.286x2 + 0.341x3 + 0.003x1^2 + 0.076x2^2 + 0.160x3^2 \\ &\quad - 0.068x1x2 + 0.000x1x3 + 0.119x2x3\end{aligned}$$

The cost equation also contains the interaction between material A and C, meaning in this case that the cost of changing the safety stock depends on how frequently it is done, and is strongly curved in safety stock. The fill-rate surface is an upward-sloping surface but decreases with increasing smoothing, increasing the review period, and as safety stock increases, and the negative quadratic terms suggest diminishing returns. The positive C-by-D interaction indicates that the two dimensions of policy in this fitted region interact, with longer review periods and higher smoothing having a greater effect on bullwhip.

4.5 Canonical analysis and response-surface geometry

The Y1 stationary point was proven to be an interior minimum by canonical analysis by virtue of all three of the eigenvalues of its quadratic form being positive. This cost-only stationary point was recorded in engineering units around k = 1.512, R = 6.332 days, and alpha = 0.316, with the predicted cost of SAR 45.849 million. It was not feasible: The predicted fill rate was 94.274%, and the bullwhip was 1.689.

The Y2 stationary point was a maximum but was outside the CCD limits, mainly due to the stationary value of the smoothing alpha being far out of the engineering range. The stationary point of the Y3 surface was a saddle with two positive and one negative eigenvalue and also outside the range of CCD. Such geometries preclude a simple approach, which would be to optimize each of these responses at its stationary point in the absence of constraints and choose whichever seems to be the better result.

Table 6: Comparison of principal candidate policies

Candidate	k	R (days)	alpha	Cost (SAR m)	Fill (%)	Bullwhip	Feasibility
Cost stationary	1.512	6.332	0.316	45.849	94.274	1.689	No
Strict constrained minimum	1.775	5.095	0.343	46.146	96.000	1.595	Model-feasible
Balanced desirability	2.253	3.520	0.3635	48.087	97.563	1.523	Model-feasible
Incumbent centre	2.000	7.000	0.500	47.496	95.408	1.831	No
Confirmed alternative	2.413	5.800	0.280	49.706	97.605	1.546	Observed feasible

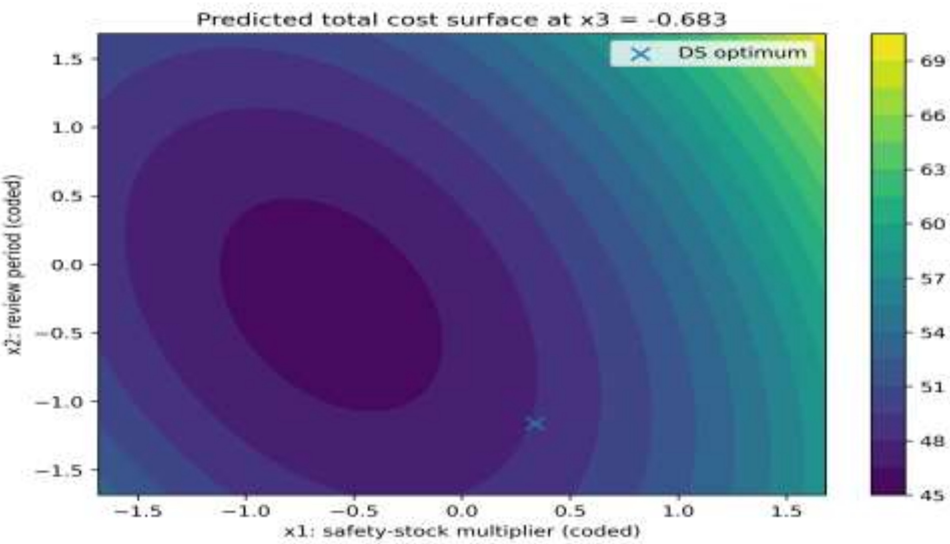
The strict and desirability solutions are predictions from the fitted CCD. The confirmed alternative is a held-back simulation point. The strict solution fill prediction is 95.9999999999992% numerically and is treated as 96.000% using the optimization tolerance.

4.6 Constrained and multi-response optimization

The constrained optimization was a very tight one that directly addressed the management question. The lowest predicted cost inside the CCD subject to fill >= 96.0% and bullwhip <= 1.60 occurred at k = 1.7747, R = 5.0946 days, and alpha = 0.3434. Its predicted annual cost was SAR 46.1465 million, its predicted fill rate was 96.0000%, and its predicted bullwhip was 1.5947. This is a predicted cost savings of approximately SAR 1.350 million (2.84%) with an increase of 0.592 percentage points in fill rate and a reduction of 12.90% in bullwhip, relative to the fitted incumbent center.

The balanced Derringer-Suich solution was further into the feasible service-stability region. At k = 2.2534, R = 3.5204 days, and alpha = 0.3635, predicted cost was SAR 48.0874 million, fill rate was 97.5634%, and bullwhip was 1.5234, with composite desirability 0.5306. The fitted incumbent center is approximately SAR 0.591 million more expensive than the solution, but the solution provides a 2.155 percentage point increase in the fill rate and a 16.80% decrease in bullwhip. Hence, it can be a viable policy if robustness and stability of the service are more important than the next notch of cost minimization.

Figure 5: Predicted total-cost surface at the balanced desirability smoothing level



International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Figure 6: Predicted fill-rate surface at the balanced desirability smoothing level

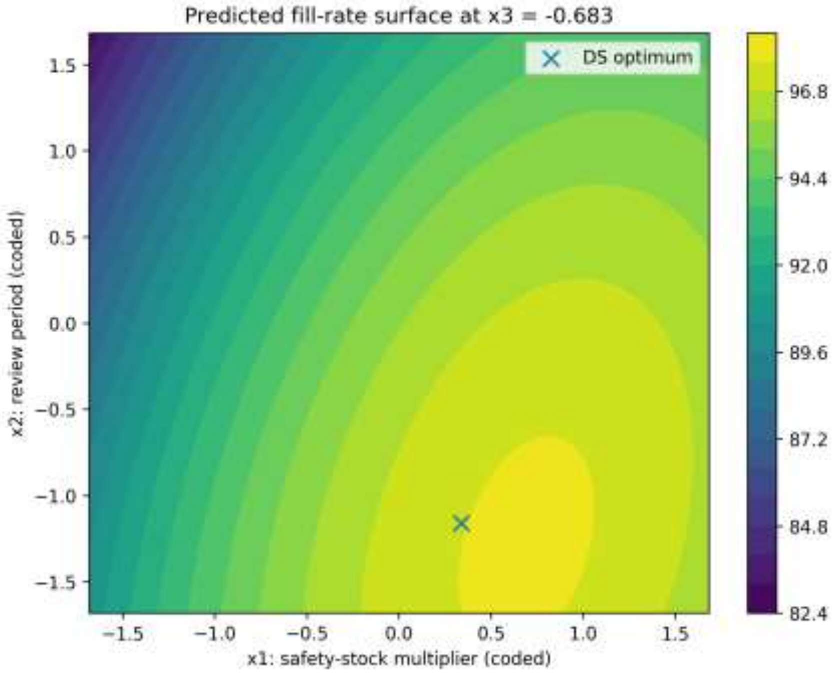
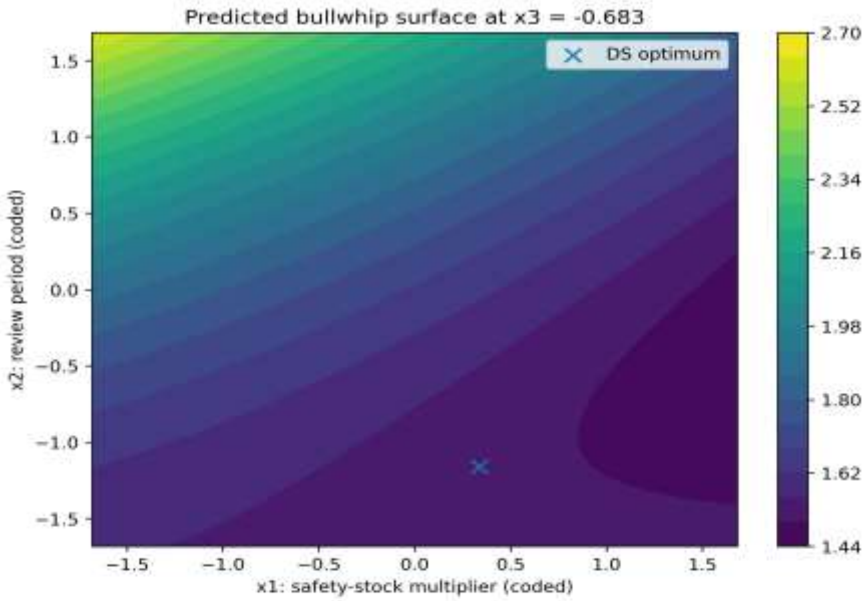


Figure 7: Predicted bullwhip surface at the balanced desirability smoothing level



The trade-off is illustrated in Figures 5, 6 and 7. The lowest-cost basin is situated towards lower levels of safety stock and moderate levels of review periods, while the highest fill-rate region is towards the higher levels of safety stock and lower review periods. Still other combinations are preferred on the bullwhip surface. The balanced solution is chosen from the intersection of acceptable cost, service, and stability, which is not on either of the extreme response points of the absolute cost basin.

4.7 Held-back confirmation and feasibility audit

The most significant validation finding of the study was obtained during the confirmation phase. The provided experiment had three runs marked as a 'Predicted optimum' at k = 1.535, R = 5.95 days, and alpha = 0.32. They had a mean cost of SAR 45.797 million, mean fill rate of 94.575%, and mean bullwhip of 1.6678. The observed responses were within the quadratic model prediction intervals, but the candidate didn't meet both of the managerial constraints. This can therefore be considered a proper confirmation of model prediction, though not a proper confirmation of policy feasibility.

The optimum obtained by fitting the full RSM was not one of the six settings that were held back. It should therefore be presented as an optimum found by the model and not an optimum that has been verified by experiment. Before a manuscript can make a claim to direct confirmation of that particular vector of parameters, a new simulation with exactly these values of k, R, and alpha needs to be performed.

The only possibility was the alternative held-back candidate. At k = 2.413, R = 5.8 days, and alpha = 0.28, it produced a cost of SAR 49.706 million, a fill rate of 97.605%, and bullwhip of 1.5464. Relative to the incumbent baseline run, the candidate increased cost by SAR 2.199 million (95% CI [1.689, 2.709]) but increased fill rate by 2.267 percentage points (95% CI [1.988, 2.546]) and reduced bullwhip by 0.2843 (95% CI [-0.3080, -0.2606]). It is a service-stability premium that is observed, not a cost-saving policy.

Table 7: Held-back confirmation and feasibility audit

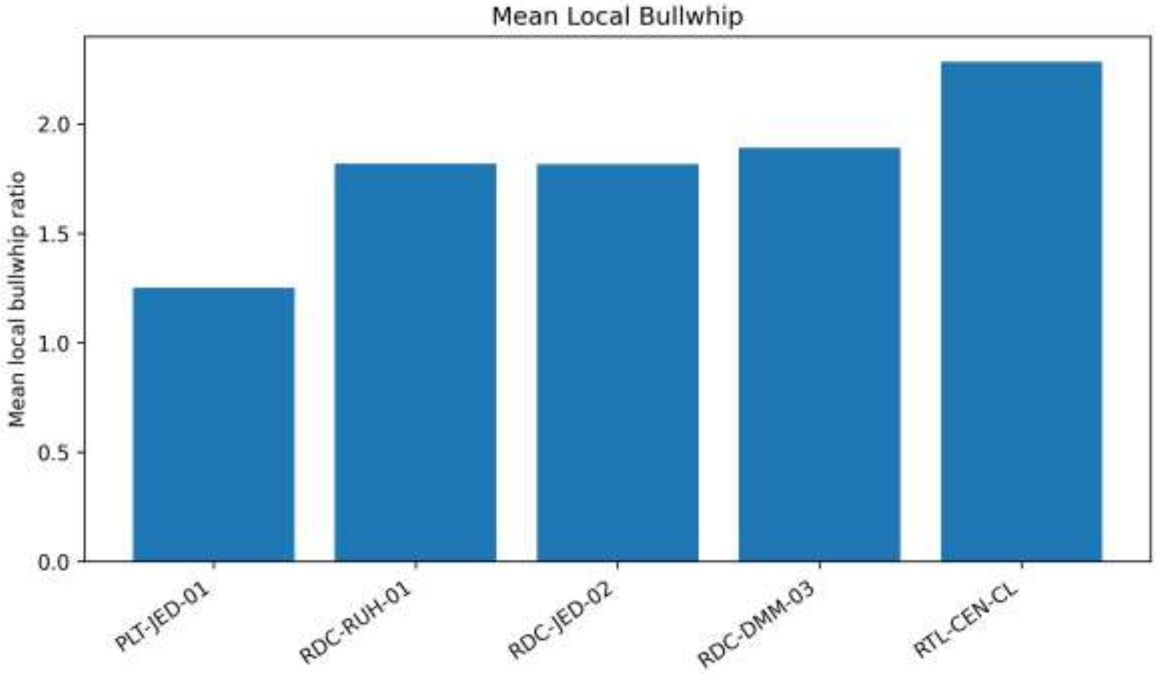
Confirmation setting	k	R	alpha	Cost (SAR m)	Fill (%)	Bullwhip	Feasible
Nominated cost-oriented candidate (mean of 3)	1.535	5.95	0.32	45.797	94.575	1.6678	No
Interior check	1.775	6.40	0.36	46.200	95.379	1.6694	No
Incumbent baseline	2.000	7.00	0.50	47.507	95.338	1.8307	No
Alternative candidate	2.413	5.80	0.28	49.706	97.605	1.5464	Yes

The first row is the mean of three replicate confirmation runs. Feasibility requires fill >= 96% and bullwhip <= 1.60.

4.8 Echelon diagnostics

The reported node subset had an increasing local bullwhip downstream at the center-point diagnostic setting. The local bullwhip at the plant was around 1.250, at the three RDCs it was about 1.817-1.890, and at the Central retail cluster it was about 2.284. The biggest mean cost share was in the plant (SAR 16.257 million), followed by Riyadh RDC (SAR 11.448 million) and Jeddah RDC (SAR 8.924 million). These diagnostics reveal that the overall plant–retail bullwhip response is not just a statistical abstraction; there is material variation across nodes, and it is important to take local variation into account when implementing a Bullwhip policy across the network.

Figure 8: Mean local bullwhip by reported node at the center-point diagnostic setting



5. Discussion

5.1 Forecast modernization does not automatically modernize inventory policy

The key takeaway is that it is not the same intervention to have a strong forecasting capability as it is to have a strong replenishment policy. The cost-only stationary point was attractive on Y1 but did not satisfy the service and bullwhip requirements, and the incumbent center did not satisfy both requirements. This is a validation of the study's idea that the post-forecast decision layer can reverse the impact of a forecasting investment into the desired (financial and service) consequences.

This is in line with the general difference between demand forecasts and operational demand plans. Accordingly, the results of Fahimnia et al. (2023) indicate that information about the service level impacts forecasting judgments in practice; and that forecast generation and operational planning are mutually dependent but conceptually different. In the present simulation, this insight into human judgment carries over to the parameters of the policies: When the generation of the forecast is predetermined, the targets for the services and replenishment rules influence the impact of the forecast on the inventory and on customer results.

The discovery also fits research on the data-driven inventory-control problem that aims to include replenishment parameters in the optimization problem, not demand prediction. Theodorou et al. (2023) learn the mapping between the characteristics of the demand and policy parameters and their corresponding performance, to improve the inventory cost and service. In this context, RSM delivers a more interpretable and lower-dimensional form of that principle, where its key interaction and curvature estimates are around three key policy controls for action.

5.2 Cost, service, and bullwhip define different optima

The Phase 2 path is an example of the pitfalls of cost optimization alone. Rapidly decreasing safety stock and review period resulted in a decrease in cost first, but a loss in fill rate before the cost minimum was achieved. Bullwhip did not improve as rapidly or as much as it could have, and was still over 1.60 even at the lowest cost sampled step. There is thus a real three-way compromise in the system, and no one policy dominates.

This is in line with multi-echelon studies where it has been demonstrated that decisions on inventory have to be based on trading off the holding and shortage costs across the coupled locations and not on optimization of each node independently. Geevers et al. (2024) find that integrated policies can perform above benchmarks for a variety of network configurations, with the performance improvement depending on the method by which states and replenishment actions are coordinated. The current study reinforces the previous result: even if AI planning is done as a signal, it can lead to bad dynamics if the policy is too reactive or miscalibrated.

Bullwhip is important because it can deteriorate, as customer-facing actions may seem fine. Control-based supply chain research shows that the information and replenishment feedback can exacerbate the variability of the upstream supply chain, and that upstream activity is not guaranteed to be stable even if the information is accurate (Papanagnou, 2022). The fitted Y3 surface confirms this interpretation; it shows that in the latter case, responsiveness is useful only if the temporal parameters of the policy help to avoid disproportionate plant orders in response to local demand changes.

5.3 Interpreting the strict and balanced solutions

The management would choose the most cost-effective policy within the feasible set; thus, the strict constrained solution is the correct answer if they consider 96% fill and 1.60 bullwhip to be hard constraints. It is placed very near the fill-rate boundary and near the bullwhip boundary (operationally fragile and economically efficient). This could result in realized performance being outside of one of the limits due to small model error, parameter drift, and/or unmodeled capacity constraints. Therefore, it is important to verify the solution by conducting more stochastic replications before implementation.

A more conservative approach in terms of service and stability is the balanced desirability solution. It reduces the shortness of the review period, increases the safety stock, and does not allow smoothing to fall below the incumbent level, causing a 1.56 percentage-point fill-rate buffer above the contract level and a 0.077 bullwhip buffer below 1.60. The extra cost prediction is the cost of robustness. This is the scenario for the application of multi-response desirability (MRD) methods, which are those methods that are developed for exactly this situation, where none of the responses can be made better at its optimum without compromising another response (Madić and Marinković, 2024).

This distinction is further supported by some observed simulation evidence of the held-back alternative candidate. It was able to meet both constraints without a problem, but it was more expensive than the incumbent. This is consistent with the idea that feasibility can be bought by making changes to the inventory and responsiveness, but it may not be consistent with the idea that the alternative is economically efficient. Thus, the model-based strict optimum is promising in predicting feasibility with less cost, but no exact confirmation run is available so as to make the stronger assertion that it has been validated.

5.4 Implications of the Saudi event-driven context

The analysis of the demand reveals the reason that a single seasonal stationary pattern is not sufficient in this situation. The highest legacy under-forecast biases were in the weeks of Ramadan, Eid al-Fitr, and Hajj, and the most volatile SKU family was dates and confectionery. FMCG forecasting studies have also demonstrated that certain contextual factors such as weather, price, and holidays can significantly improve the results of forecasting models that do not consider such factors, and models are based only

on historical sales (Farizal et al., 2021). This suggests that policy tuning should be tested under the same event regimes that make forecasting challenging, and not only tested under average demand weeks.
Hajj is also a context that is significant in the Saudi context as it is already a high-complexity operational environment with a concentrated focus that uses AI-supported coordination for large-scale management problems (Abalkhail and Al Amri, 2022). In this particular instance, the stress caused by replenishment varies with time of year and the amount of demand for pilgrims from the West, as well as the seasons of the Islamic calendar. It is therefore recommended that the next production implementation should be able to encode Hijri-aligned features of events and also ensure the robustness of the policy over several lunar-calendar years.

5.5 Why RSM is useful despite the availability of black-box optimizers

A genetic algorithm or a Bayesian optimizer or a reinforcement learning policy would be able to explore a wider parameter space and may be able to find an optimum with fewer assumptions of a quadratic local structure. Recent multi-echelon work shows that both deep reinforcement learning and evolutionary algorithms can be powerful approaches to high-dimensional inventory problems (Liu and Nishi, 2024). This study's argument for RSM is not about computational supremacy; it's about diagnostic transparency.
The RSM sequence revealed the significance of the factors, as well as the presence or absence of curvature and the direction(s) in which the 1st-order response curve was inadequate, the location of the stationary points within or outside the experimental domain, and the binding constraint(s). The composite response-surface designs (RSDs) are particularly designed for efficient local estimation of this second-order structure (Athanasaki et al., 2024). An explicit surface can be challenged and audited for managerial adoption of an optimized solution much more readily than a parameter vector provided by a purely black-box optimizer.
The model diagnostics also reveal the weaknesses. The Y3 heteroskedasticity signal and the Y1 normality warning are still displayed, and the confirmation phase, which was held back, shows that even if the predicted values are correct, a cost-attractive candidate can be rejected due to heteroskedasticity. This is a good feature of the sequential protocol: confirmation is not a ceremony of repetition of the fitted optimum, but is a separate inferential stage.

6. Managerial Implications

First, companies should recognize that implementing AI forecast and replenishment-policy calibration should be considered two interdependent workstreams. Information that goes into inventory decisions is impacted by a new forecasting engine, and legacy safety stock, review, and smoothing parameters need to be revalidated, not assumed to be optimal. In practice, any changes to a material forecasting model, data feed, horizon, or feature set are good candidates to kick off a policy-tuning experiment.
Second, it is important that policy tuning is constraint-first, not cost-first. The Phase 2 results indicate that a seemingly successful cost descent can happen quickly over into a low-service area. Hardness of service and stability should be specified beforehand by the manager, and the feasible region should be explicitly reported. In a service-constrained multi-echelon environment, this would be more aligned with the requirements of operational decisions than to pick the cheapest location and see service failures as a low-frequency second problem (Liu and Nishi, 2024).
Third, the strict optimum should be interpreted as a policy of efficiency, and the balanced desirability solution as a policy of robustness. The strict point results in less modeled cost near the feasibility limits,

while the balanced point trades off some of the fill rate and bullwhip for larger buffers. It will depend on the contractual penalties, risk tolerance, exposure to the event season, and cost of emergency replenishment. A seasonal change from the efficiency policy to a more stringent set of parameters may be better for optimal Saudi periods than applying one annual policy throughout the year.

Finally, although information-sharing lag was not included in the three-factor CCD, it is crucial and should not be overlooked. The screening resulted in a statistically significant impact with respect to bullwhip. The organization should revisit E in the context of A, C, and D in a subsequent experiment if the latency of the point of sale, supplier integration, or the frequency of planning data refresh changes. This is an illustration of the importance of applying factor elimination for multi-response studies with all responses, not just the response that is region-seeking.

7. Limitations and Future Research

The first constraint is that of data provenance. The case and its company, FMCG Co., are synthetic and not a real registered company. The demand history, cost structure, lead times, event uplifts and simulation responses are not necessarily audited observations from a named enterprise, but were generated to reflect what would be a realistic FMCG situation in Saudi Arabia. This numerical optimum should not therefore be made a Saudi industry standard. It is the experimental protocol and the behavior it displays under the given assumptions that is offered as a contribution.

Secondly, the forecasting engine is fixed, and its accuracy is an input condition. The data set includes the actual demand and a legacy forecast for baseline characterization, but not a full series of AI forecasts to be able to be independently reproduced to assess forecast improvement. Future research should consolidate all forecast results, as well as the inventory-state transitions and simulation logs of events, to follow the forecast value all the way to cost and service.

Thirdly, the network topology is fixed, the transport capacity is unlimited, there is no product substitution, and the supplier disruption is modeled only by the variation in lead time. In high-inventory and/or high-frequency scenarios, real lane and vehicle constraints would probably make the cost higher in those scenarios and might shift and/or level out the response-surface optimum. The lack of lateral transshipment also precludes the generalization to networks in which RDCs tranship amongst themselves.

Fourth, Phase 3 only optimizes A, C and D. Defensible for the pre-specified cost-led region search, but screening indicated that E had an impact on bullwhip. A multi-response screening rule/augmented CCD based on information lag would be a better foundation for a network-wide stability optimum. Thus, the current Y3 surface should be read as conditional on the information lag that is fixed in these experiments.

Fifth, some mild non-normality in the cost model and heteroskedasticity in the bullwhip model were detected using the residual diagnostics. Results from the non-significant lack-of-fit tests and held-back prediction checks suggest that the surfaces are valid for use within the design region; however, future work should be conducted to investigate the sensitivity of the surfaces recommended both through ordinary least squares and heteroskedasticity-robust inference, through weighted response-surface regression, and through Gaussian-process surrogates.

Lastly, there was a lack of the exact, strictly constrained optimum in the held-back confirmation design. The next experimental action is clear: repeat the simulation at $k = 1.7747$, $R = 5.0946$ days, and $alpha = 0.3434$ with the original common random number protocol of 30 replications and with other random numbers out of sample, and event-specific stress conditions. The strict point should only be considered confirmed if those runs are within prediction intervals and continue to meet both conditions.

8. Conclusion

This paper discussed a scenario that lies between AI forecasting and inventory control: how to tune the replenishment policy once the forecasting engine has been changed. Sequential RSM identified safety stock, review period, and signal smoothing as the fundamental factors that influence the local cost region; the method also found that the total cost exhibits strong curvature, and that the second-order models for cost, fill rate, and bullwhip were highly predictive in this case study of Saudi FMCG.
Results indicate that the unconstrained cost optimum is not operationally acceptable as it does not meeting service and stability requirements. In contrast, constraint-aware optimization identified a model-based policy of k = 1.775, review period = 5.095 days, and alpha = 0.343 that resulted in a forecasted annual cost of SAR 46.146 million, fill = 96.000%, and bullwhip = 1.595. A desirability policy that is balanced takes some cost for more robust service and less bullwhip. Held back confirmation was further evidence of the possibility of a previously nominated cost-oriented candidate being able to predict accurately and yet be managerially infeasible.
The general message is that error rates in forecasts cannot be the sole indicator of the value of AI in supply chains. Forecasts can be used to generate information; replenishment policies can be used to generate inventory, service, as well as information about variability flowing upstream. Therefore, it is important for organizations looking to gain financial value from AI demand planning to re-tune the decision rules around the model, validate multiple responses at once, and finalize the parameter vector before deployment under stochastic simulation.

References

1. Abalkhail, A. A. A., and Al Amri, S. M. A. (2022). Saudi Arabia's management of the Hajj season through artificial intelligence and sustainability. Sustainability, 14(21), 14142. https://www.mdpi.com/2071-1050/14/21/14142
2. Athanasaki, D. E., Georgiou, S. D., and Stylianou, S. (2024). New approaches on composite designs for response surface methodology. PLOS ONE, 19(4), e0301049. https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0301049
3. Douaioui, K., Oucheikh, R., Benmoussa, O., and Mabrouki, C. (2024). Machine learning and deep learning models for demand forecasting in supply chain management: A critical review. Applied System Innovation, 7(5), 93. https://www.mdpi.com/2571-5577/7/5/93
4. Fahimnia, B., Arvan, M., Tan, T., and Siemsen, E. (2023). A hidden anchor: The influence of service levels on demand forecasts. Journal of Operations Management, 69(5), 856–871. https://onlinelibrary.wiley.com/doi/full/10.1002/joom.1229
5. Farizal, F., Dachyar, M., Taurina, Z., and Qaradhawi, Y. (2021). Disclosing fast moving consumer goods demand forecasting predictor using multi linear regression. Engineering and Applied Science Research, 48(5), 627–636. https://ph01.tci-thaijo.org/index.php/easr/article/view/242407
6. Fussone, R., Dominguez, R., Cannella, S., and Framinan, J. M. (2024). Bullwhip effect in closed-loop supply chains with multiple reverse flows: A simulation study. Flexible Services and Manufacturing Journal, 36, 250–278. https://link.springer.com/article/10.1007/s10696-023-09486-x

7. Geevers, K., van Hezewijk, L., and Mes, M. R. K. (2024). Multi-echelon inventory optimization using deep reinforcement learning. Central European Journal of Operations Research, 32, 653–683. https://link.springer.com/article/10.1007/s10100-023-00872-2
8. Liu, Z., and Nishi, T. (2024). Data-driven evolutionary computation for service constrained inventory optimization in multi-echelon supply chains. Complex and Intelligent Systems, 10(1), 825–846. https://link.springer.com/article/10.1007/s40747-023-01179-0
9. Madić, M., and Marinković, V. (2024). Multi-response optimization in laser processing technologies by applying desirability function approach and response surface methodology based on grey relation analysis. Advances in Mechanical Engineering, 16(7), 1–18. https://journals.sagepub.com/doi/10.1177/16878132241259029
10. Nasseri, M., Falatouri, T., Brandtner, P., and Darbanian, F. (2023). Applying machine learning in retail demand prediction—A comparison of tree-based ensembles and long short-term memory-based deep learning. Applied Sciences, 13(19), 11112. https://www.mdpi.com/2076-3417/13/19/11112
11. Papanagnou, C. I. (2022). Measuring and eliminating the bullwhip in closed loop supply chains using control theory and Internet of Things. Annals of Operations Research, 310, 153–170. https://link.springer.com/article/10.1007/s10479-021-04136-7
12. Rheem, S. (2023). Optimizing food processing through a new approach to response surface methodology. Food Science of Animal Resources, 43(2), 374–381. https://link.springer.com/article/10.5851/kosfa.2023.e7
13. Sbai, N., and Berrado, A. (2023). Simulation-based approach for multi-echelon inventory system selection: Case of distribution systems. Processes, 11(3), 796. https://www.mdpi.com/2227-9717/11/3/796/html
14. Teixeira, A. R., Ferreira, J. V., and Ramos, A. L. (2025). Intelligent supply chain management: A systematic literature review on artificial intelligence contributions. Information, 16(5), 399. https://www.mdpi.com/2078-2489/16/5/399/html
15. Theodorou, E., Spiliotis, E., and Assimakopoulos, V. (2023). Optimizing inventory control through a data-driven and model-independent framework. EURO Journal on Transportation and Logistics, 12, 100103. https://www.sciencedirect.com/science/article/pii/S2192437622000280