Scientific Reports

https://doi.org/10.1038/s41598-026-35197-y

Article in Press

Primacy of feature engineering over architectural complexity for intermittent demand forecasting

Received: 12 December 2025

Accepted: 2 January 2026

Published online: 06 January 2026

Cite this article as: Nathan B.S., Aravinth P.M., Reddy B.V.S. *et al.* Primacy of feature engineering over architectural complexity for intermittent demand forecasting. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-35197-y

B. Sendhil Nathan, P. M. Aravinth, B. Veera Siva Reddy, C. Chandrasekhara Sastry, Sachin Salunkhe & Robert Cep

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Primacy of Feature Engineering over Architectural Complexity for Intermittent Demand Forecasting

B. Sendhil Nathan[a,b], P.M. Aravinth[a], B. Veera Siva Reddy[a]*, C. Chandrasekhara Sastry[a]*, Sachin Salunkhe[c], Robert Cep[d]

[a]Department of Mechanical Engineering, Indian Institute of Information Technology Design and Manufacturing Kurnool (IIITDM Kurnool), Kurnool 518008, Andhra Pradesh, India

[b]Ford Motor Private Limited, Ford Global Technology & Business Center, Chennai 600119, Tamil Nadu, India

[c]Department of Mechanical Engineering, Gazi University, Ankara, Turkey

[d]Department of Machining, Assembly and Engineering Metrology, Faculty of Mechanical Engineering, VSB-Technical University of Ostrava, 70800 Ostrava, Czech Republic

*Corresponding authors: sivareddy.bobbili@gmail.com, chandrasekhar@iiitk.ac.in

ARTICLE IN PRESS

Abstract

Intermittent demand forecasting remains a fundamental challenge in large-scale supply chains due to extreme demand sparsity, irregular occurrence patterns, and highly variable demand magnitudes. While recent studies have increasingly adopted complex multi-stage model architectures to address these challenges, the role of statistically grounded feature engineering has received comparatively less attention. This study proposes the Smoothed Hybrid Occurrence-Size (SHOS) framework, which generates adaptive, series-specific estimates of demand occurrence probability and conditional demand size using sparsity-aware exponential smoothing. These estimates are incorporated as features into supervised machine learning models trained on large-scale, zero-padded

ARTICLE IN PRESS

panel data. The proposed approach is evaluated on an automotive aftermarket dataset comprising approximately 1.4 million monthly observations across 56,000 spare-part time series, using an 11-fold rolling-window cross-validation protocol. Empirical results demonstrate that SHOS-enhanced models achieve substantial performance improvements over baseline feature sets, reducing mean absolute error (MAE) by approximately 50% and weighted mean absolute percentage error (WMAPE) by over 40% in highly intermittent demand segments. Notably, despite their increased architectural complexity, two-stage hurdle-based models do not outperform the proposed single-stage SHOS-enhanced framework. Formal statistical testing using the Wilcoxon signed-rank test confirms that the performance advantage of the single-stage SHOS model is consistent and statistically significant across all validation folds ($p < 0.001$). These findings reveal an unexpected but practically important insight: robust, statistically informed feature engineering can be more effective than increased model complexity for intermittent demand forecasting. The results highlight the value of simple, interpretable, and computationally efficient forecasting frameworks for large-scale operational deployment, while motivating future validation across additional application domains.

Keywords: Feature engineering; intermittent demand forecasting; machine learning for supply chains; sparse time series; statistical-ML hybrid models

1. Introduction

ARTICLE IN PRESS

Intermittent demand forecasting is characterized by long periods of zero demand punctuated by irregular, low-frequency demand events remain a fundamental challenge in supply chain and inventory management. This challenge is particularly pronounced in automotive aftermarket and spare-parts systems, where organizations must manage tens of thousands of slow-moving stock keeping units (SKUs) across distributed dealer networks. Inaccurate forecasts in such settings lead to a costly trade-off between excessive inventory holding and obsolescence on one hand, and service-level failures and lost revenue on the other [1-3]. As supply chains scale in size and complexity, the need for forecasting methods that are both robust to sparsity and scalable across large item populations has become increasingly critical.

Classical intermittent demand forecasting methods, beginning with Croston's decomposition and later extended through the Syntetos-Boylan Approximation (SBA) and the Teunter-Syntetos-Babai (TSB) method, explicitly model demand as the product of a demand occurrence process and a conditional demand size process [4-6]. These methods offer strong interpretability and stability under sparsity, but they are inherently univariate and operate independently on each time series. As a result, they are unable to exploit auxiliary information or cross-sectional structure that is increasingly available in modern enterprise systems.

In contrast, modern machine learning (ML) models such as gradient-boosted decision trees (e.g., LightGBM and XGBoost) can leverage rich, multivariate feature representations and capture complex nonlinear

ARTICLE IN PRESS

relationships across large panels of time series [7,8]. However, when applied directly to intermittent demand data, these models often suffer from degraded performance due to extreme sparsity, weak signal-to-noise ratios, and a tendency to overfit rare demand events [9,10]. This has motivated the use of increasingly complex modeling strategies, including multi-stage or hurdle-based architectures, under the assumption that structural complexity is necessary to handle intermittent demand.

Despite these developments, an important gap remains in the literature. Much of the recent focus has been placed on architectural complexity, while the role of statistically grounded feature representation has received comparatively less systematic attention. Evidence from recent forecasting competitions and empirical studies suggests that hybrid approaches where statistical signals are embedded into machine learning pipelines can yield substantial performance gains without resorting to complex model structures [11,12]. This raises a fundamental question: can carefully designed statistical features capture the essential dynamics of intermittent demand more effectively than increasingly complex model architectures?

To address this question, we propose the Smoothed Hybrid Occurrence-Size (SHOS) framework. SHOS is an adaptive statistical procedure that generates series-specific estimates of two latent components of intermittent demand: (i) the probability of demand occurrence and (ii) the conditional demand size given occurrence. These components are estimated using sparsity-aware exponential smoothing

ARTICLE IN PRESS

that accounts for both prolonged inactivity and demand reactivation. Importantly, SHOS is not used as a standalone forecasting model; instead, its outputs referred to as SHOS features are incorporated directly as input features within standard supervised machine learning models. This design enables the integration of statistically meaningful demand signals into scalable ML pipelines without altering downstream model architectures.

The proposed framework is evaluated on a large-scale automotive aftermarket dataset comprising more than 56,000 dealer-part time series, expanded to approximately 1.4 million monthly observations through zero padding. All models are assessed using an 11-fold rolling-window cross-validation protocol, ensuring temporal integrity and realistic deployment conditions. Comparative experiments include standard machine learning baselines, SHOS-enhanced single-stage models, and more complex two-stage hurdle-based architectures.

ARTICLE IN PRESS

The results demonstrate that SHOS feature augmentation leads to substantial and statistically significant improvements in forecasting accuracy, with mean absolute error reductions of approximately 50% in highly intermittent demand segments. Notably, despite their greater structural sophistication, two-stage hurdle models do not consistently outperform a single-stage regressor augmented with SHOS features. This finding reveals a key and somewhat unexpected insight: in large-scale intermittent demand forecasting, representation quality and statistically informed feature engineering can outweigh architectural complexity. The study thus contributes both a practical forecasting framework and a

ARTICLE IN PRESS

conceptual clarification of where performance gains truly arise in intermittent demand modeling.

2. Literature review

The challenge of forecasting intermittent demand has a rich history, with research evolving from foundational statistical decompositions to complex, data driven machine learning architectures. This review synthesizes the key methodological streams to contextualize the contributions of this paper.

2.1 Foundational paradigms for intermittent demand modelling

The bedrock of intermittent demand forecasting is the principle of decomposition. Recognizing that standard time series models fail when demand is frequently zero, early research focused on methods that separately model the two underlying stochastic processes: the occurrence of a demand event and the size of that demand.

ARTICLE IN PRESS

The seminal contribution in this area is Croston's method, which applies separate exponential smoothing to the inter demand intervals and the positive demand sizes [4]. While foundational, Croston's method was later shown to possess a positive statistical bias, a finding that spurred a series of crucial refinements [13]. The Syntetos Boylan Approximation (SBA) introduced a debiasing factor that has been empirically shown to improve both forecast accuracy [5] and inventory performance [14]. A further significant evolution came with the Teunter Syntetos Babai (TSB) method, which shifted from smoothing the inter demand interval to smoothing the demand probability itself [6]. By updating this probability

ARTICLE IN PRESS

in every period including zeros the TSB method is more responsive to changes in demand frequency. The SHOS algorithm proposed in this work is conceptually related to the TSB method, as it also focuses on the direct smoothing of occurrence probability and conditional size.

The Teunter-Syntetos-Babai (TSB) method represents an important advancement over Croston-style estimators by directly smoothing the probability of demand occurrence rather than inter-demand intervals. By updating the occurrence probability in every period, including zero-demand periods, TSB improves responsiveness to changes in demand frequency. However, the method relies on fixed smoothing parameters and uniform update rules across all series, which can introduce systematic bias under extreme intermittency.

In highly sparse demand environments, TSB may exhibit two limitations. First, prolonged zero-demand periods can cause the smoothed occurrence probability to decay slowly, leading to delayed adaptation when demand resumes. Second, isolated demand spikes can exert disproportionate influence on both the occurrence and size estimates, particularly when the smoothing parameters are not calibrated to series-specific sparsity. These effects can result in biased or unstable forecasts in non-stationary intermittent demand scenarios.

The SHOS algorithm builds upon the conceptual decomposition introduced by TSB while explicitly addressing these limitations through adaptive and recency-aware mechanisms, as detailed in Section 3.6.

ARTICLE IN PRESS

ARTICLE IN PRESS

These classical methods are robust, interpretable, and remain important benchmarks. However, their critical limitation is their univariate nature. They operate solely on the historical demand series, which prevents them from incorporating the rich external feature sets (e.g., price, product attributes) that often drive demand in real world supply chains.

2.2 Machine learning and deep learning approaches

With the proliferation of data, machine learning (ML) and deep learning (DL) models have become powerful tools for forecasting, valued for their ability to learn complex, nonlinear relationships from a wide array of features. Tree based ensembles like LightGBM, XGBoost, and CatBoost are now considered state of the art for many tabular forecasting tasks [7,8,15]. Concurrently, deep learning has introduced sequential architectures like Recurrent Neural Networks (RNNs) to capture complex temporal dependencies [16]. However, the primary failure of these powerful models is their data hungry nature, which makes their performance on highly sparse time series unstable; they are prone to overfitting to the few non zero data points and often struggle in the low signal environment where long periods of zero demand provide weak gradients for learning [9].

2.3 Hybridization and the identified research gap

Recognizing the respective limitations of pure statistical and pure ML models, the frontier of forecasting research has moved towards hybridization. The principle is to combine the strengths of different

ARTICLE IN PRESS

modeling paradigms to achieve performance superior to any single model. This approach has been validated at scale in major forecasting competitions, where the winning methods were overwhelmingly hybrid. The 4th and 5th Makridakis Competitions (M4 and M5) [11,12], which are major benchmarks for time series forecasting, conclusively demonstrated that top performing models often combined statistical methods with ML components, such as the M4 winner's hybrid of exponential smoothing and an LSTM network [17].

Recent work has explored embedding statistical signals directly into ML pipelines not as after ensembles, but as input features that guide learning [18]. However, a critical gap remains: there is limited research on adaptive, series-specific statistical priors that are both theoretically grounded and computationally efficient for large scale deployment. Moreover, it remains an open question whether complex, specialized architectures such as two stage or zero inflated models are necessary when ML regressors are equipped with high quality statistical features that already encode demand occurrence and size dynamics.

Recent forecasting literature increasingly emphasizes the role of feature design over model complexity, particularly in sparse and noisy settings. Feature-based forecasting and hybrid statistical-machine learning approaches have shown that incorporating domain-consistent statistical signals as model inputs can substantially improve generalization and stability. Rather than relying solely on raw lagged values, these

ARTICLE IN PRESS

approaches embed structural knowledge of the demand-generating process into features that guide learning.

In the context of intermittent demand, features that explicitly encode demand occurrence and conditional size are especially important, as they reflect the underlying renewal structure of the process. The SHOS algorithm follows this principle by generating adaptive, series-specific statistical estimates that are subsequently used as input features for machine learning models.

This work addresses these gaps by introducing the SHOS algorithm as a feature generation mechanism and demonstrating that a single stage regressor, when augmented with SHOS derived signals, achieves performance that matches or exceeds more complex alternatives. The results support the hypothesis that, in intermittent demand forecasting, feature engineering can supersede architectural complexity.

ARTICLE IN PRESS

2.3 Feature representation, model complexity, and generalization in modern ML

Recent studies in machine learning increasingly highlight that predictive performance and generalization are governed not only by model architecture, but more critically by the quality and structure of feature representations. Across engineering, measurement science, geoscience, and risk-assessment applications, hybrid learning frameworks that integrate statistically meaningful features, uncertainty modeling, or domain knowledge with machine learning have consistently demonstrated

ARTICLE IN PRESS

improved robustness and interpretability compared to purely data-driven approaches [19-22].

A growing body of work has shown that explicitly engineered or adaptively learned feature representations often combined with ensemble models such as XGBoost or LightGBM can dominate architectural complexity in determining model generalization. Several recent studies employing feature importance and explainability techniques (e.g., SHAP) have demonstrated that informative features derived from physical insight, entropy measures, or statistical priors contribute more significantly to predictive accuracy than increasing model depth or complexity [23-26]. These findings are particularly pronounced in sparse, noisy, or highly nonlinear systems, where over-parameterized models tend to suffer from instability and overfitting.

Parallel developments in uncertainty-aware and physics-informed machine learning further reinforce the importance of structured representations. Hybrid models incorporating entropy theory, cloud models, or physics-guided constraints have been shown to enhance generalization under limited data availability and non-stationary conditions, emphasizing the limitations of black-box learning in such regimes [26-29].

The present study aligns with these emerging directions by prioritizing feature representation over architectural innovation. The proposed SHOS framework generates adaptive statistical representations of intermittent demand by explicitly modeling demand occurrence

ARTICLE IN PRESS

probability and conditional demand size. These SHOS-derived features encode sparsity, recency, and demand reactivation dynamics in a statistically grounded and interpretable manner, enabling improved generalization when used with standard machine learning models. Consistent with recent literature, our results demonstrate that representation-centric design can yield substantial performance gains without increasing model complexity, particularly in highly intermittent demand environments [19-29].

3. Methodology and experimental setup

This section details the SHOS algorithm, the machine learning frameworks under evaluation, and the experimental design used to compare their performance. The overall methodological framework from data preprocessing to final performance analysis is illustrated in Figure 1. The pipeline begins with raw transactional data, which is aggregated to a monthly frequency and zero padded to ensure a consistent temporal grid. A comprehensive feature engineering process then generates a rich set of predictors, including statistically grounded signals from the SHOS algorithm. Finally, a suite of forecasting models is trained using hyperparameters and evaluated via a rolling window cross validation protocol that respects temporal order.

ARTICLE IN PRESS

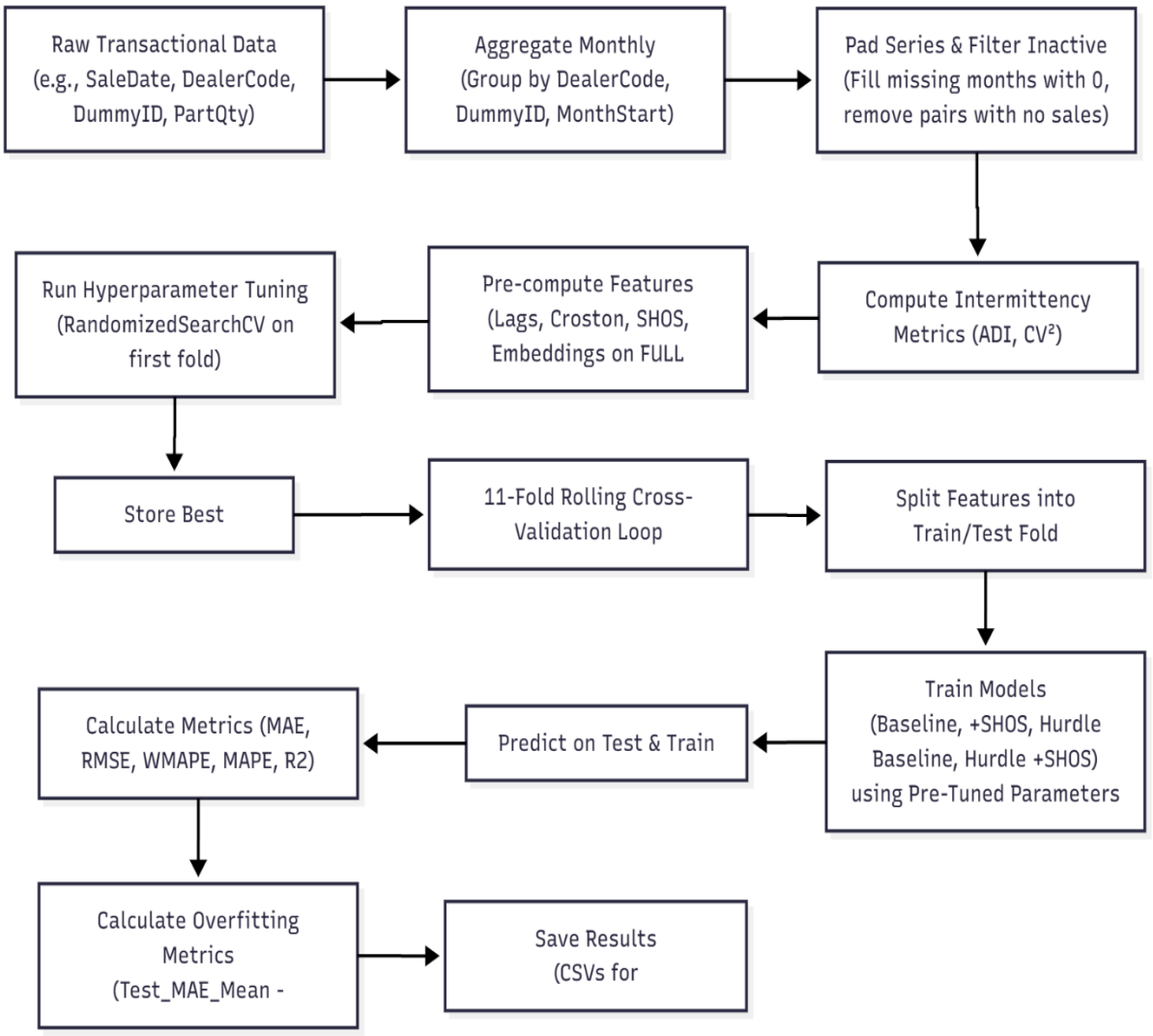


Figure 1. End-to-end experimental workflow for intermittent demand forecasting, illustrating data preprocessing, SHOS feature generation, model training, and rolling-window evaluation

3.1 Dataset description

The experiments use a large-scale intermittent demand dataset from the automotive aftermarket sector, comprising daily dealer part sales transactions from 2022 to 2024. The raw data includes 59 dealers and 7,702 unique parts, yielding approximately 56,000 active dealer part time series. Due to extreme sparsity at the daily level, the data are aggregated

to a monthly frequency. To enable consistent feature engineering across all series, the monthly data are zero padded over the full date range, resulting in a final dataset of approximately 1.4 million observations. This padded dataset preserves the high intermittency of the original data, with a median Average Demand Interval (ADI) of 25 and median coefficient of variation squared cv^2 of 24 as shown in Figure 2. The target variable is the total monthly part quantity sold (PartQty). Covariates include static part attributes such as price, weight, and cube dimensions.

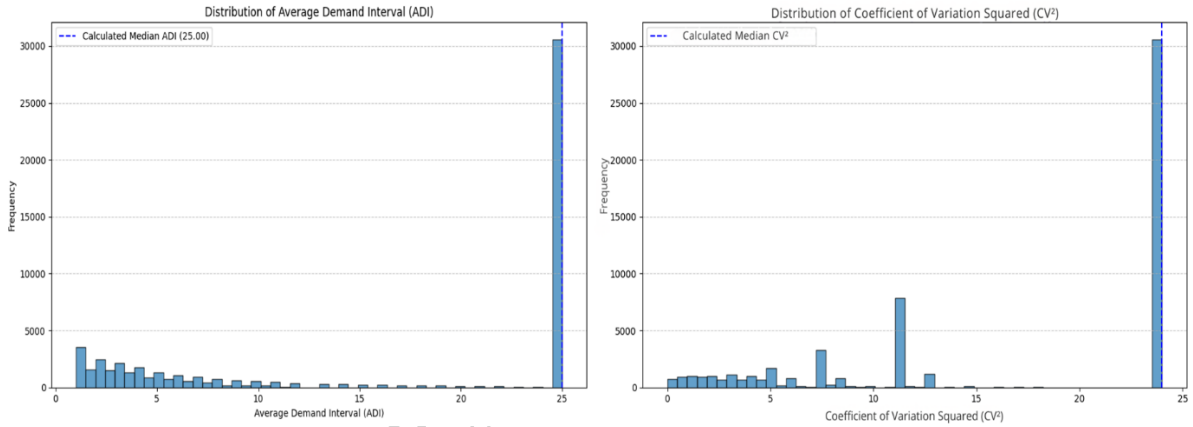


Figure 2. Empirical distribution of intermittency characteristics (ADI and CV²) across 56,000 active dealer-part time series

All experimental analyses in this study are conducted on the zero-padded, monthly-aggregated dataset described above. The zero-padding step establishes a consistent temporal grid across all dealer-part time series, which is a prerequisite for uniform feature engineering, rolling-window cross-validation, and fair comparison across models. No results are reported on raw, unpadded transactional data.

The zero-filling of monthly demand series is a statistically meaningful transformation rather than a purely engineering-driven choice.

ARTICLE IN PRESS

In the context of intermittent demand, zero-valued observations explicitly represent periods of non-occurrence and therefore carry essential information for modeling demand sparsity and renewal dynamics. Treating these periods as missing values would eliminate critical signals required for estimating demand occurrence probabilities and bias feature distributions.

3.2 Preprocessing and feature engineering

A structured feature engineering pipeline transforms the padded monthly data into a high dimensional feature set. The pipeline includes:

- Temporal features: Lagged values of sales quantity up to 12 months, rolling sums and means (3- and 6-month windows), count of non-zero periods in the last 6 months, and time elapsed since the last sale.
- Algorithmic features: Forecasts from Croston's method [4] and the three core outputs of the SHOS algorithm: smoothed occurrence probability (q_{shos}) smoothed conditional size (z_{shos}), and the SHOS point forecast. Crucially, SHOS features are computed using an adaptive smoothing mechanism with sparsity scaling factor k = 0.1, and an improved initialization that uses the first three months of data or falls back to the global mean of positive demands.
- Global and latent features: Dealer and part level historical averages, and low rank embeddings (12 dimensions). To strictly preserve temporal integrity and prevent data leakage, the Truncated SVD factorization is dynamically recomputed within each cross-validation fold. The dealer part activity matrix is constructed exclusively using

ARTICLE IN PRESS

the training data available in that specific fold, ensuring no future information influences the embeddings.

For fairness, baseline ML models are trained on base features only (temporal, global, and latent), while SHOS features are reserved for ablation studies.

3.3 Forecasting models for comparison

The evaluation includes five machine learning models: LightGBM, XGBoost, ElasticNet, Ridge, and Random Forest. Models labeled +SHOS use the full feature set (including SHOS derived features), while baseline variants use only base features. Two stage (hurdle) variants are also evaluated, where a LightGBM classifier predicts demand occurrence and a separate LightGBM regressor (trained on positive demands with a Tweedie objective) predicts conditional size.

All models use fixed, pre optimized hyperparameters, determined through a preliminary hyperparameter search using time series cross validation on a validation fold. The final values were then hard coded for all folds to ensure consistent and efficient evaluation. The specific parameters used are presented in Table 1.

ARTICLE IN PRESS

Table 1. Hyperparameter settings for all evaluated models

Model	Hyperparameter	Value
Ridge Regression	alpha	0.01
XGBoost	n_estimators	500
	max_depth	9
	learning_rate	0.05

ARTICLE IN PRESS

	subsample	0.8
LightGBM	n_estimators	500
	max_depth	9
	learning_rate	0.05
	num_leaves	31
Hurdle Model	base regressor	LightGBM
SHOS	alpha_base	0.2
	beta_base	0.2
	sparsity scaling (k)	0.1

Gradient boosted models (LightGBM and XGBoost) employ a Tweedie objective to better handle the zero inflated and skewed nature of intermittent demand. To independently assess the impact of feature engineering and model architecture, the experimental design follows a controlled ablation framework. All forecasting models are trained on the same zero-padded dataset using identical cross-validation folds, preprocessing steps, and fixed hyperparameters. The only controlled variations are (i) the inclusion or exclusion of SHOS-derived features and (ii) the use of a single-stage versus two-stage (hurdle) architecture.

This design ensures that differences in predictive performance can be attributed specifically to feature engineering or architectural complexity, rather than to confounding factors such as data preprocessing, hyperparameter tuning, or evaluation protocol. Importantly, all evaluation metrics reported in this study are computed on the test windows derived from the processed dataset, ensuring a direct correspondence between the data preparation steps, the experimental protocol, and the reported results.

ARTICLE IN PRESS

Classical intermittent demand forecasting methods such as Croston, SBA, and TSB are not included in the primary benchmark comparison. This is because the present study focuses on feature-based learning within a supervised machine learning framework operating on zero-padded panel data, where models are trained across a large cross-section of series using engineered features. In contrast, classical methods generate direct time-series forecasts for individual series and do not naturally produce feature representations compatible with cross-sectional learning or model-based attribution analysis. TSB, in particular, serves as a conceptual basis for the proposed SHOS algorithm and is therefore discussed in the methodological section rather than treated as a competing baseline. This design choice ensures a fair and internally consistent comparison among models operating under the same learning paradigm.

3.4 Evaluation protocol and metrics

Model performance is evaluated using a rolling-window cross-validation protocol designed to respect temporal ordering and prevent look-ahead bias. In each fold, models are trained on a contiguous historical window and evaluated on the immediately following test window, thereby closely simulating real-world deployment conditions. All data-dependent transformations including feature scaling and the construction of low-rank latent embeddings using truncated singular value decomposition (SVD) are fitted exclusively on the training data of each fold and then applied to the corresponding test data. This ensures strict temporal integrity and avoids information leakage. Model performance is reported as the mean

ARTICLE IN PRESS

and standard deviation of evaluation metrics computed across all eleven rolling validation folds.

Let y_i denote the observed demand and $\hat{y}_i$ the corresponding forecast for observation i, with N representing the total number of observations in the evaluation set and $\grave{y}$ denoting the mean of the observed demand. Forecast accuracy is assessed using multiple complementary metrics that capture both absolute and relative error characteristics relevant to intermittent demand forecasting.

The Root Mean Squared Error (RMSE) is used to quantify sensitivity to large forecast errors and is particularly important for avoiding severe underprediction that can lead to stockouts. It is defined as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}. \quad (1)$$

The Mean Absolute Error (MAE) measures the average magnitude of forecast errors and provides an intuitive measure of typical deviation between predicted and observed demand. It is computed as

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|. \quad (2)$$

The coefficient of determination, R^2, evaluates the proportion of variance in the observed demand explained by the model and serves as a normalized measure of explanatory power. It is given by

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \grave{y})^2}. \quad (3)$$

ARTICLE IN PRESS

To assess relative forecast accuracy in a volume-weighted manner, the Weighted Mean Absolute Percentage Error (WMAPE) is employed. Unlike conventional percentage errors, WMAPE remains well-defined for intermittent demand series with frequent zero values. It is defined as

$$\text{WMAPE} = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{\sum_{i=1}^{N} |y_i|} \times 100\%. \quad (4)$$

In addition to aggregate accuracy metrics, overfitting is diagnosed by comparing training and test errors across validation folds to assess generalization behavior. Model interpretability and feature-level contributions are further examined using SHAP (SHapley Additive exPlanations) analysis, providing insight into the drivers of predictive performance.

To reporting aggregate error metrics, statistical significance testing is performed to assess whether observed performance differences between competing models are meaningful. Given the paired structure of rolling-window cross-validation and the non-normal error distributions typical of intermittent demand, the Wilcoxon signed-rank test is employed to compare per-fold MAE and WMAPE values between models. All tests are conducted at a 5% significance level.

3.5 Problem formulation

Following the foundational work on intermittent demand forecasting [4,6], a time series of demand $\{y_t\}_{t=1}^{T}$ is modeled as the product of two latent processes: an occurrence process and a size process. Let y_t denote the observed demand quantity in period t, with $O_t = I(y_t > 0)$ indicating

ARTICLE IN PRESS

whether a demand event occurred, and $z_t = y_t | y_t > 0$ representing the conditional size. The forecasting objective is to estimate the conditional expectation:

$$E[y_{t+1}|F_t] = P(y_{t+1} > 0|F_t) \cdot E[y_{t+1}|y_{t+1} > 0, F_t], \quad (5)$$

where F_t is the information set available up to time t. This work focuses on generating robust, series-specific estimates of these two components and incorporating them as features into machine learning models. In this formulation, the forecasting objective is to estimate the conditional expectation of demand given the information set F_{t-1}, which includes all historical demand observations and available covariates up to time $t - 1$.

3.6 The SHOS algorithm: adaptive statistical priors

The SHOS (Smoothed Hybrid Occurrence Size) algorithm is an adaptive statistical feature-generation framework designed to operate robustly under extreme demand intermittency, where time series are dominated by long sequences of zero demand punctuated by sporadic, irregular non-zero observations. The algorithm decomposes intermittent demand into two latent components estimated at each time step: the probability of demand occurrence and the conditional demand size given that demand occurs. This formulation allows SHOS to update continuously over time, including during prolonged zero-demand periods, which is essential for maintaining stable estimates in highly sparse demand environments. The overall logical flow of the SHOS algorithm is illustrated in Figure 3. The SHOS algorithm explicitly leverages zero-filled observations as informative indicators of demand non-occurrence,

enabling continuous updating of occurrence probability during inactivity periods and ensuring statistically consistent estimation under extreme sparsity.

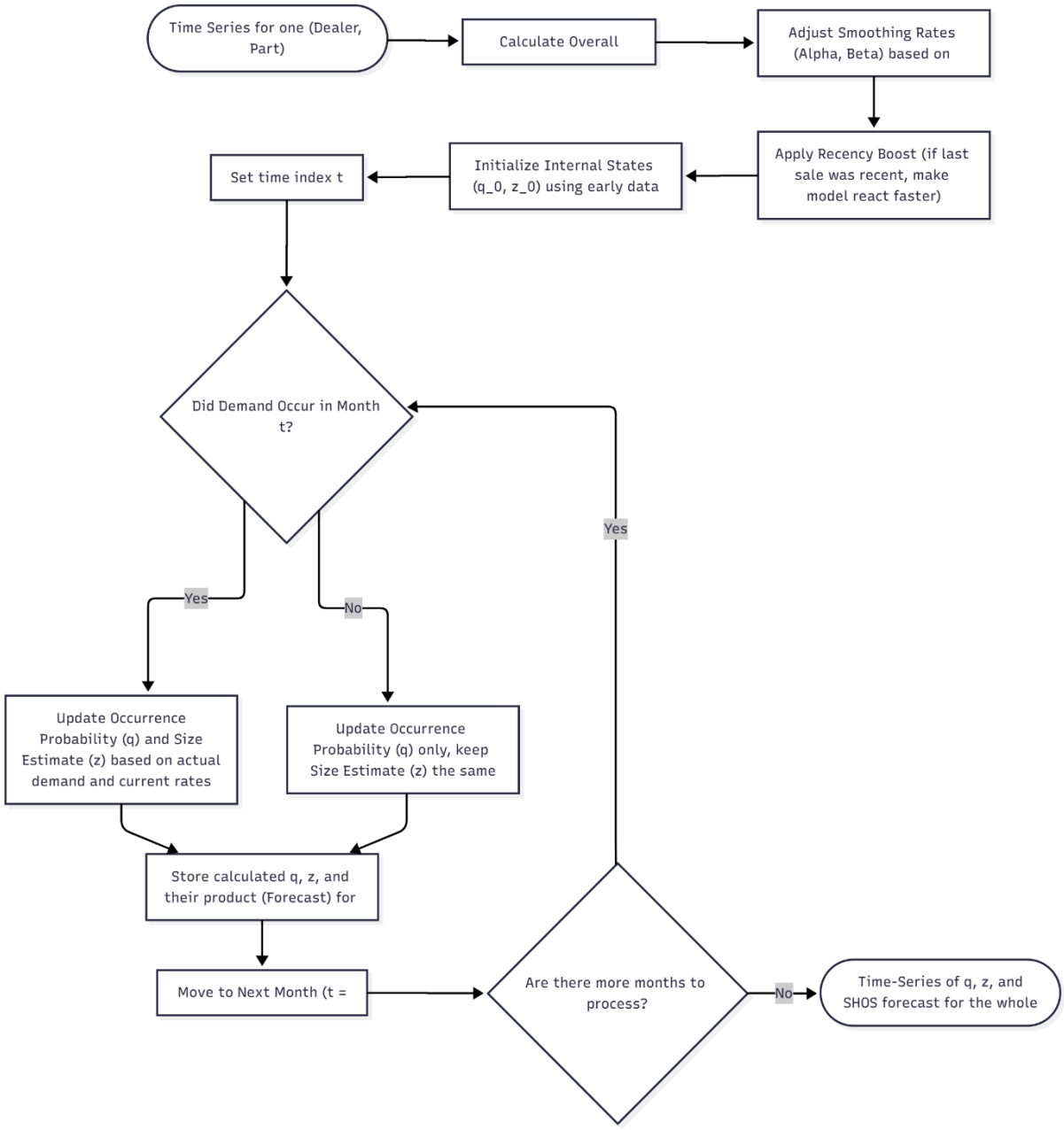


Figure 3. Algorithmic flow of the SHOS framework, illustrating adaptive smoothing, recency adjustment, and occurrence-size updates

ARTICLE IN PRESS

Algorithm. SHOS feature-generation procedure

Input: Demand series $\{y_t\}_{t=1}^{T}$; base smoothing parameters $\alpha_{base},\beta_{base} \in (0,1)$; sparsity scaling factor $k \in [0,1]$; recency window $r \in \mathbb{N}$; cap $\alpha_{max},\beta_{max} \in (0,1]$.

Output: SHOS features $\{\hat{q}_t,\hat{z}_t,\hat{y}_t\}_{t=1}^{T}$.

1. **Compute demand indicator:** for all t, set $I_t \leftarrow 1(y_t > 0)$.
2. **Compute series sparsity:**

$$\text{sparsity} \leftarrow 1 - \frac{1}{T}\sum_{t=1}^{T} I_t .$$

3. **Set sparsity-adaptive smoothing:**

$\alpha \leftarrow \alpha_{base}(1 - k \cdot \text{sparsity}), \beta \leftarrow \beta_{base}(1 - k \cdot \text{sparsity})$. **Initialize** $\hat{q}_0$ **and** $\hat{z}_0$**:** compute from the first three periods if sufficient positive observations exist; otherwise set $\hat{q}_0$ to the empirical occurrence rate in the initialization window and $\hat{z}_0$ to the global mean of positive demands (computed on the training data of the fold).

4. **For** $t = 1,2,\ldots,T$ **do:**
    1. **Recency-aware boost (optional):** let τ_t be the number of periods since the most recent positive demand. If $\tau_t \le r$, set $\alpha_t \leftarrow \min(\alpha_{max},\alpha), \beta_t \leftarrow \min(\beta_{max},\beta)$; otherwise set $\alpha_t \leftarrow \alpha$ and $\beta_t \leftarrow \beta$.
    2. **Update occurrence probability:**

ARTICLE IN PRESS

ARTICLE IN PRESS

$\hat{q}_t \leftarrow \beta_t I_{t-1} + (1 - \beta_t)\hat{q}_{t-1}$, with I_0 defined from the initialization window.

3. **Update conditional size:**

$$\hat{z}_t \leftarrow \begin{cases} \alpha_t y_{t-1} + (1 - \alpha_t)\hat{z}_{t-1}, & \text{if } y_{t-1} > 0, \\ \hat{z}_{t-1}, & \text{if } y_{t-1} = 0, \end{cases}$$ with y_0 defined from the initialization window.

4. **Compute SHOS point forecast:**

$\hat{y}_t \leftarrow \hat{q}_t \cdot \hat{z}_t$. **Return** $\{\hat{q}_t, \hat{z}_t, \hat{y}_t\}_{t=1}^{T}$.

Note: In the rolling-window cross-validation protocol, all statistics required for initialization (e.g., global mean of positive demands) are computed using only the training window of the corresponding fold to avoid leakage.

ARTICLE IN PRESS

Let y_t denote the observed demand quantity at time period t, and let $I_t = 1(y_t > 0)$ be a binary indicator that equals 1 when demand occurs and 0 otherwise. SHOS maintains a smoothed estimate of the probability of demand occurrence, denoted by $\hat{q}_t$, and a smoothed estimate of the conditional demand size, denoted by $\hat{z}_t$. The SHOS point forecast of expected demand is given by $\hat{y}_t = \hat{q}_t \cdot \hat{z}_t$.

3.6.1 Smoothed occurrence probability and conditional size

In highly intermittent demand series, naive estimators of occurrence probability and demand magnitude are often unstable or unresponsive due to the dominance of zero observations. SHOS mitigates this issue by explicitly updating the demand occurrence probability at every time step,

ARTICLE IN PRESS

including periods with zero demand. This ensures that the estimated occurrence probability decays gradually during inactivity rather than remaining fixed, thereby preventing persistent overestimation following isolated demand events. The occurrence probability is updated using exponential smoothing as

$$\hat{q}_t = \beta \cdot I_{t-1} + (1 - \beta) \cdot \hat{q}_{t-1}, \tag{6}$$

where $\beta \in (0,1)$ is the smoothing parameter controlling the responsiveness of the occurrence estimate. The conditional demand size $\hat{z}_t$ is updated only when a positive demand is observed, thereby avoiding contamination of magnitude estimates by zero-demand periods. Specifically, the update rule is defined as

$$\hat{z}_t = \begin{cases} \alpha \cdot y_{t-1} + (1 - \alpha) \cdot \hat{z}_{t-1}, & \text{if } y_{t-1} > 0, \\ \hat{z}_{t-1}, & \text{if } y_{t-1} = 0, \end{cases} \tag{7}$$

where $\alpha \in (0,1)$ is the smoothing parameter for the conditional size process. By separating the update mechanisms for occurrence and size, SHOS ensures that magnitude estimates remain statistically meaningful even when non-zero observations are rare.

3.6.2 Adaptive smoothing, sparsity handling, and initialization

To further stabilize estimation under anomalous intermittent demand patterns, SHOS employs sparsity-adaptive smoothing. Series-specific sparsity is quantified using the proportion of zero-demand periods over the observed horizon T, defined as

ARTICLE IN PRESS

$$\text{sparsity} = 1 - \frac{1}{T}\sum_{t=1}^{T} I(y_t > 0). \quad (8)$$

The effective smoothing parameters are dynamically adjusted as a function of this sparsity measure according to

$$\alpha = \alpha_{base} \cdot (1 - k \cdot \text{sparsity}), \beta = \beta_{base} \cdot (1 - k \cdot \text{sparsity}), \quad (9)$$

where α_{base} and β_{base} are base smoothing values and k is a sparsity scaling factor controlling the strength of adaptation. In this study, $\alpha_{base} = \beta_{base} = 0.2$ and $k = 0.1$, which provides stronger smoothing for highly sparse series while preserving responsiveness for less intermittent ones.

To evaluate the sensitivity of SHOS to the sparsity scaling factor k, a sensitivity analysis was conducted over a wide range $k \in [0.1, 0.9]$. The mean MAE values obtained for representative k values are 0.6647 ($k = 0.1$), 0.6654 ($k = 0.3$), 0.6657 ($k = 0.5$), 0.6611 ($k = 0.7$), and 0.6665 ($k = 0.9$). The maximum relative variation in MAE across this range is only 0.81%, indicating that SHOS performance is highly robust to the choice of k. Based on this analysis, $k = 0.1$ is adopted in this study as it provides stable smoothing while maintaining responsiveness to demand reactivation.

To improve robustness in short or extremely sparse series, SHOS employs a conservative initialization strategy. Initial estimates $\hat{q}_0$ and $\hat{z}_0$ are computed from the first three observed periods when sufficient positive demand exists; otherwise, they default to the global mean of positive demand across the dataset. In addition, a recency-aware adjustment mechanism temporarily increases the effective smoothing weights (subject to an upper bound of 0.9) when demand occurs after an

ARTICLE IN PRESS

extended zero-demand run. This allows SHOS to respond rapidly to genuine demand reactivation without destabilizing long-term estimates.

3.6.3 SHOS versus TSB: bias mitigation and adaptivity

Although SHOS and the Teunter-Syntetos-Babai (TSB) method shares a common conceptual foundation by modeling demand as the product of an occurrence probability and a conditional size, their estimation mechanisms differ fundamentally. In TSB, both components are updated using fixed exponential smoothing parameters, implicitly assuming homogeneous intermittency and stationarity across all series. While effective under moderate intermittency, this assumption can lead to delayed responsiveness and variance-induced bias in highly sparse or non-stationary demand environments.

SHOS addresses these limitations by introducing sparsity-adaptive smoothing, whereby the degree of smoothing is dynamically scaled according to series-specific intermittency. This reduces overreaction to isolated demand events in highly sparse series while maintaining responsiveness in less intermittent ones. Furthermore, the recency-sensitive adjustment mechanism enables SHOS to adapt more rapidly than TSB when demand reappears after prolonged inactivity, mitigating the delayed-response bias commonly observed in fixed-parameter smoothing approaches [30].

Finally, the robust initialization strategy employed by SHOS stabilizes early estimates in short or highly intermittent series, further reducing estimation bias relative to classical methods. Collectively, these

ARTICLE IN PRESS

mechanisms enable SHOS to generate stable, informative, and adaptive estimates of demand occurrence probability and conditional size, making it particularly well suited for feature generation in anomalous intermittent demand scenarios.

3.7 Machine learning frameworks

To systematically examine the relative importance of feature representation versus model architectural complexity, two machine learning frameworks are evaluated. Both frameworks use the raw monthly demand quantity (PartQty) as the prediction target and operate under identical data preprocessing, feature construction, and evaluation protocols. All models are trained using pre-tuned hyperparameters obtained from the initial tuning stage to ensure a fair and controlled comparison.

ARTICLE IN PRESS

Importantly, the outputs of the SHOS algorithm namely the estimated demand occurrence probability q_{shos} and conditional demand size z_{shos}, are incorporated exclusively as input features in both frameworks. No hybrid or transformed target variables are introduced, ensuring that any observed performance differences arise solely from feature augmentation rather than changes to the prediction objective.

Single-stage regressor (+SHOS)

In the single-stage framework, a LightGBM regressor is trained directly on the full feature set, which includes SHOS-derived features, temporal lag features, intermittency indicators, learned embeddings, and

ARTICLE IN PRESS

static part and dealer attributes. The model produces a direct estimate of expected demand for each time period. This framework serves as the primary benchmark for evaluating the effectiveness of SHOS-based feature engineering in a standard supervised learning setting, without introducing additional architectural complexity.

Two-stage hurdle model (+SHOS)

The two-stage hurdle framework adopts a decomposed modeling strategy that explicitly separates demand occurrence and demand magnitude. It consists of two sequential components:

- **Stage 1 (Occurrence model):** A LightGBM classifier estimates the probability of observing non-zero demand, $P(y_t > 0)$, using the full base feature set augmented with the SHOS occurrence prior q_{shos}.
- **Stage 2 (Conditional size model):** A LightGBM regressor with a Tweedie objective predicts the expected demand quantity $y_t | y_t > 0$, using the same base features augmented with the SHOS conditional size estimate z_{shos}.

The final demand forecast is obtained by multiplying the outputs of the two stages, yielding an estimate of expected demand that accounts for both occurrence likelihood and conditional magnitude [31]. The overall modeling pipeline for this two-stage framework is illustrated in Figure 4.

ARTICLE IN PRESS

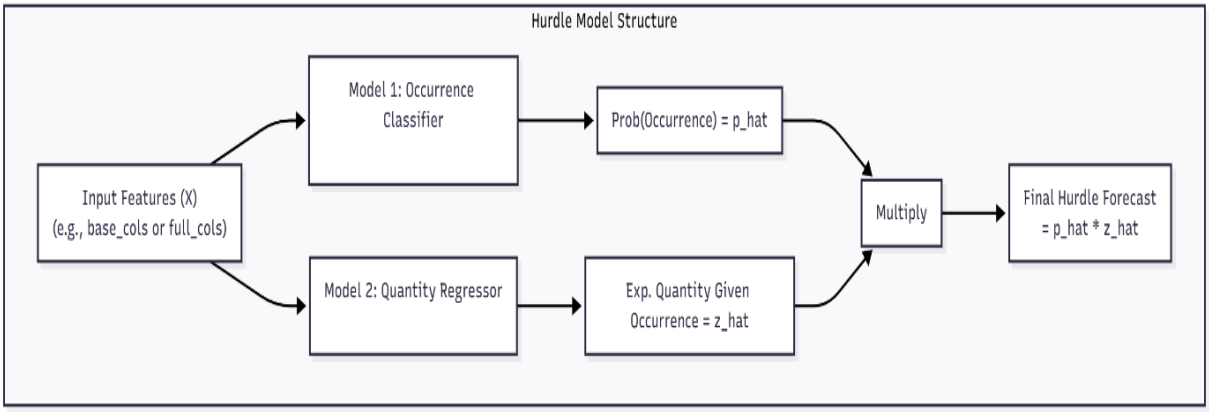


Figure 4. Two-stage hurdle model architecture illustrating separate demand-occurrence classification and conditional-quantity regression with optional SHOS feature integration

4. Results and discussions

The results presented in this section follow directly from the experimental protocol described in Section 3. All models are trained and evaluated on the same zero-padded, monthly-aggregated automotive aftermarket dataset using an identical rolling-window cross-validation framework. This dataset comprises approximately 1.4 million monthly observations derived from 56,000 dealer-part time series. By holding the data selection, preprocessing, feature construction, and evaluation protocol constant across all experiments, any observed differences in forecasting performance can be attributed solely to variations in feature sets or model architectures. Within this controlled setting, we compare the predictive performance of machine-learning models augmented with SHOS-derived features against a suite of baseline machine-learning approaches, thereby providing a rigorous empirical assessment of the

ARTICLE IN PRESS

contribution of statistically grounded feature engineering under extreme demand intermittency.

4.1 Quantitative comparison of all forecasting models

The complete performance results are presented in Table 2. Models augmented with SHOS features achieve substantially lower error than all ML baselines. The single stage +SHOS model achieves a mean R^2 of 0.8430 ± 0.0460, MAE of 0.0712 ± 0.0028, and WMAPE of 20.39% ± 0.57%. The proposed model achieves a substantial reduction in MAE compared to the baseline, computed as (0.1417 - 0.0712)/0.1417 ≈ 49.75 %, which corresponds to an approximately 50% improvement.

The two stage Hurdle_SHOS model achieves comparable performance (MAE = 0.0701, R^2 = 0.8186), confirming that architectural complexity provides no consistent advantage when high quality SHOS features are available.

ARTICLE IN PRESS

To ensure low error rates were not artifacts of sparsity, we benchmarked against a 'Naive Zero baseline. The naive strategy failed to capture variance (R^2 = -0.0176, MAE = 0.3492), whereas the +SHOS model achieved an $R^2$of 0.8430 and an MAE of 0.0712 an 80% error reduction. This confirms the model successfully captures genuine demand signals rather than merely exploiting the target's zero inflation [32].

Table 2. Comparative Performance of Machine Learning Models (Mean ± Std)

ARTICLE IN PRESS

Model	MAE	RMSE	MAPE (%)	WMAPE (%)	R^2 Score
SHOS	0.0712 ± 0.0028	1.0310 ± 0.1416	17.73 ± 0.47	20.39 ± 0.57	0.8430 ± 0.0460
Hurdle_SHOS	0.0701 ± 0.0023	1.1136 ± 0.1102	16.84 ± 0.44	20.10 ± 0.47	0.8186 ± 0.0413
Hurdle_Baseline	0.1341 ± 0.0050	1.2911 ± 0.0934	37.91 ± 1.29	38.41 ± 0.71	0.7577 ± 0.0379
LightGBM	0.1326 ± 0.0046	1.2900 ± 0.1079	38.25 ± 1.20	38.00 ± 0.68	0.7578 ± 0.0418
XGBoost	0.1413 ± 0.0045	1.3960 ± 0.1074	40.41 ± 1.37	40.49 ± 1.26	0.7156 ± 0.0494
Random Forest	0.1364 ± 0.0046	1.3411 ± 0.1057	40.28 ± 1.54	39.08 ± 1.03	0.7377 ± 0.0484
Ridge	0.3439 ± 0.0134	1.5101 ± 0.1161	69.99 ± 0.61	98.57 ± 4.28	0.6685 ± 0.0535
ElasticNet	0.3439 ± 0.0134	1.5101 ± 0.1160	70.00 ± 0.61	98.59 ± 4.26	0.6685 ± 0.0535
Naive Zero	0.3492 ± 0.0137	2.6557 ± 0.0643	-	100.00 ± 0.00	-0.0176 ± 0.0014

Note: MAPE is undefined for the Naive Zero model due to zero actual demand values, rendering percentage error metrics meaningless.

ARTICLE IN PRESS

To assess the statistical significance of the performance difference between the single-stage SHOS model and the two-stage hurdle structure, a Wilcoxon signed-rank test was applied to per-fold MAE and WMAPE values across the 11 rolling validation folds. The test results indicate statistically significant differences for both metrics (MAE: $W = 0.0$, $p = 0.00098$; WMAPE: $W = 0.0$, $p = 0.00098$) are presented in Table 3. This finding confirms that the single-stage SHOS model consistently outperforms the hurdle-based formulation under the evaluated experimental setting.

ARTICLE IN PRESS

Table 3. Statistical significance testing between SHOS and Hurdle models

Metric	Test	Test statistic (W)	p-value	Significance (α = 0.05)
MAE	Wilcoxon signed-rank	0.0	0.00098	Significant
WMAPE	Wilcoxon signed-rank	0.0	0.00098	Significant

4.2 Visual analysis of model performance

To visually synthesize the key findings, a series of plots were generated. Figure 5 shows a grouped bar chart of the absolute error metrics (RMSE and MAE). The +SHOS model achieves the lowest average error. Error bars indicate the standard deviation across rolling validation folds and highlight performance stability. Figure 6 compares percentage-based metrics (WMAPE and MAPE). WMAPE is volume-weighted and is therefore especially relevant for supply chain decision-making. The +SHOS model performs best on this metric.

ARTICLE IN PRESS

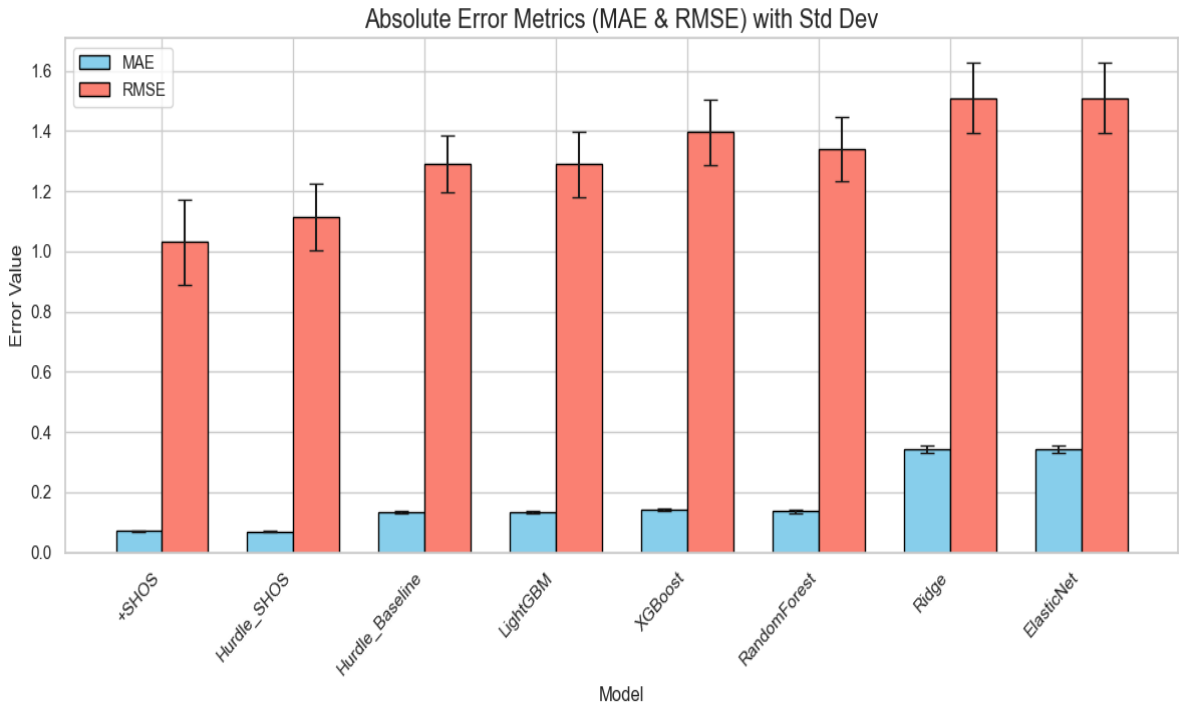


Figure 5. Comparison of absolute forecasting error metrics (MAE and RMSE) across models, reported as mean ± standard deviation over rolling validation folds

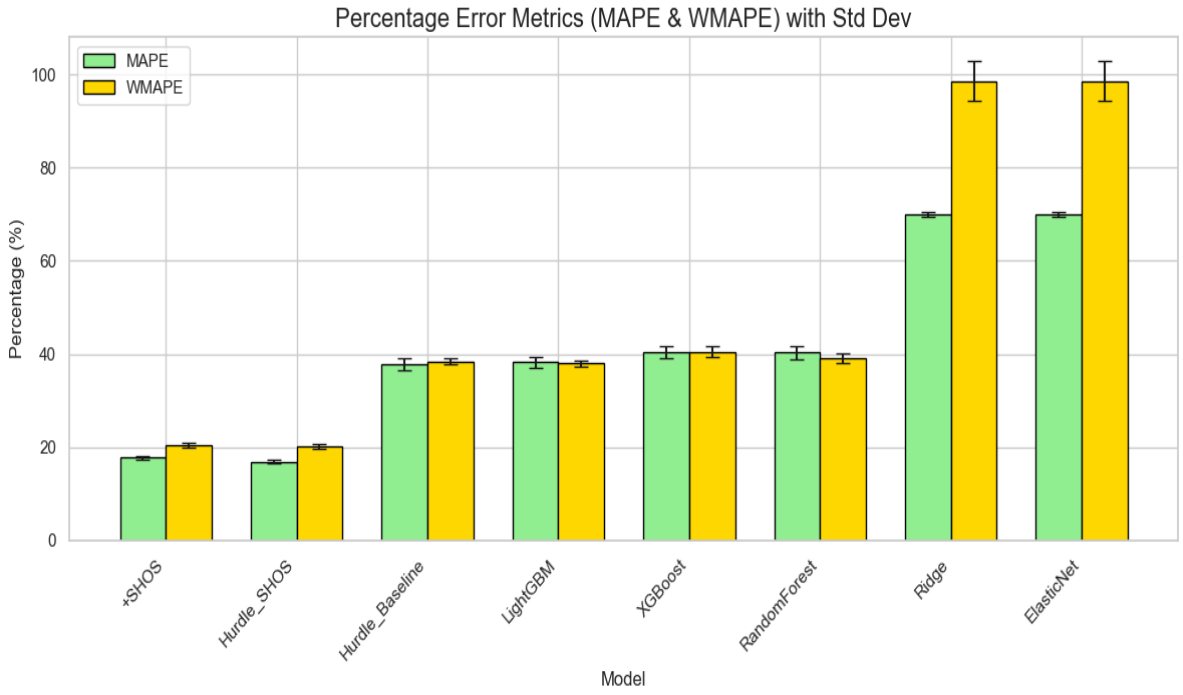


Figure 6. Comparison of percentage-based forecasting error

ARTICLE IN PRESS

metrics (MAPE and WMAPE) across models, reported as mean ± standard deviation over rolling validation folds

4.3 Component analysis

To isolate the sources of performance improvement, we conducted a controlled component (ablation) analysis. We compare models that differ in only one factor at a time. This allows performance differences to be attributed directly to feature design or architecture. Table 4 presents a feature-controlled comparison between a baseline LightGBM regressor and the same model augmented with SHOS-derived features. Since model architecture, hyperparameters, training data, and evaluation protocol are identical, the substantial reduction in MAE (≈50%) and RMSE (>28%) can be directly attributed to the inclusion of SHOS features. This confirms that feature engineering alone accounts for the majority of the observed performance gain.

ARTICLE IN PRESS

Table 4. Component Analysis: Impact of SHOS Features

Model Name	R^2 Score	RMSE	MAE
Baseline	0.7279	1.3657	0.1417
+SHOS	0.843	1.031	0.0712

The Table 5 isolates the effect of architectural complexity by comparing a single-stage regressor and a two-stage hurdle architecture trained on the same SHOS-enhanced feature set. The comparable performance across MAE, WMAPE, and R^2 demonstrates that increasing architectural complexity provides no consistent advantage once statistically grounded SHOS features are available. This result validates

ARTICLE IN PRESS

that the primary driver of performance is feature quality rather than architectural decomposition.

Table 5. Component analysis: impact of architectural complexity

Model Name	R^2 Score	RMSE	MAE
Hurdle_SHOS	0.8186	1.1136	0.0701
+SHOS	0.843	1.031	0.0712

The conclusion regarding architectural complexity is based on a controlled comparison between single-stage and two-stage forecasting frameworks trained on identical data, using the same feature representations, hyperparameter settings, and rolling-window evaluation protocol. Under these controlled conditions, the two-stage hurdle architecture does not provide a performance advantage over the single-stage SHOS-enhanced model. Statistical significance testing using the Wilcoxon signed-rank test further confirms that the performance differences are consistent and not attributable to random variation. These results indicate that, once statistically grounded SHOS features are incorporated, additional architectural complexity does not yield meaningful gains within the evaluated setting.

ARTICLE IN PRESS

4.4 Overfitting diagnosis

To assess generalization, a diagnostic comparison of training and test performance was performed. Figure 7 shows a bar chart comparing the mean MAE on the training set versus the test set for each model. The top-performing models show only a small gap between train and test MAE. For +SHOS the difference is +0.0212, and for Hurdle_SHOS it is +0.0335.

Models without SHOS priors (e.g., XGBoost and Random Forest) show larger gaps (+0.06-0.07), suggesting stronger overfitting. These results indicate that SHOS features regularize learning and improve out-of-sample generalization.

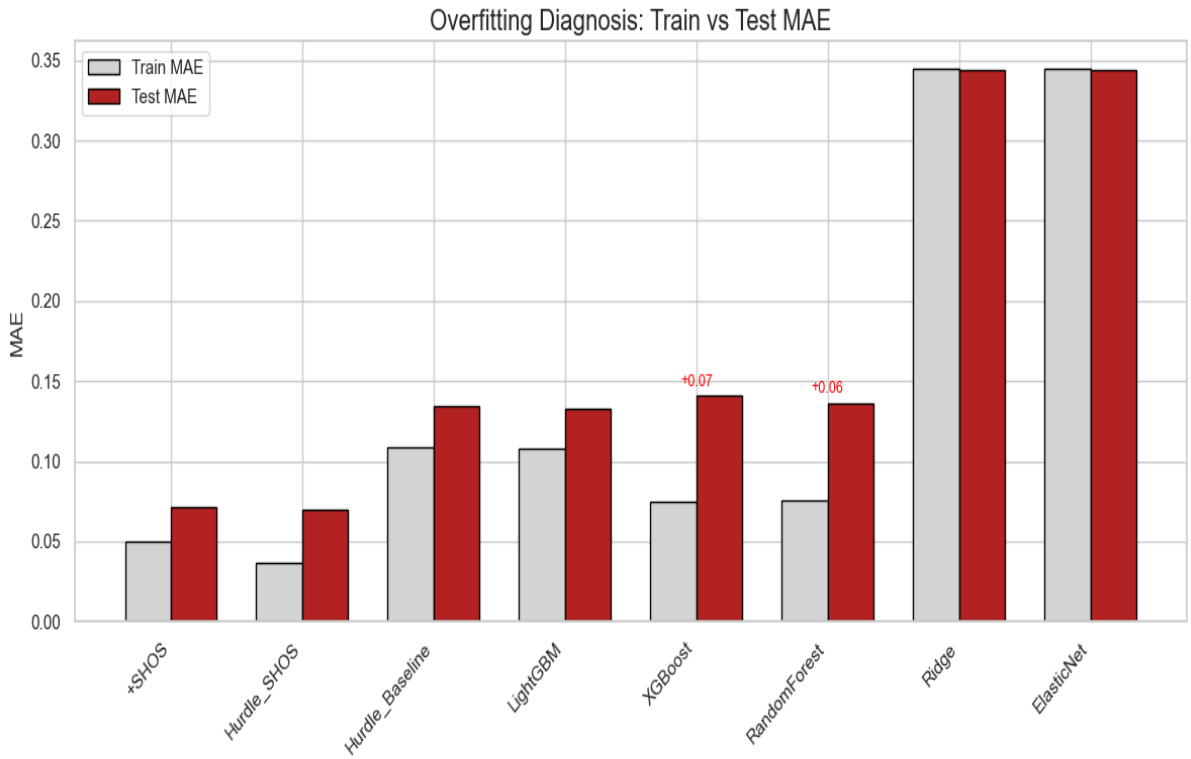


Figure 7. Overfitting diagnosis across models based on train-test MAE comparison, where positive gaps indicate reduced generalization performance

Although the experimental evaluation in this study is conducted using a large-scale automotive aftermarket spare parts dataset, the proposed SHOS framework is not domain-specific. SHOS is designed to model core properties of intermittent demand such as high sparsity, irregular demand occurrence, and variable conditional demand size which are commonly observed in other application domains, including aerospace and defense

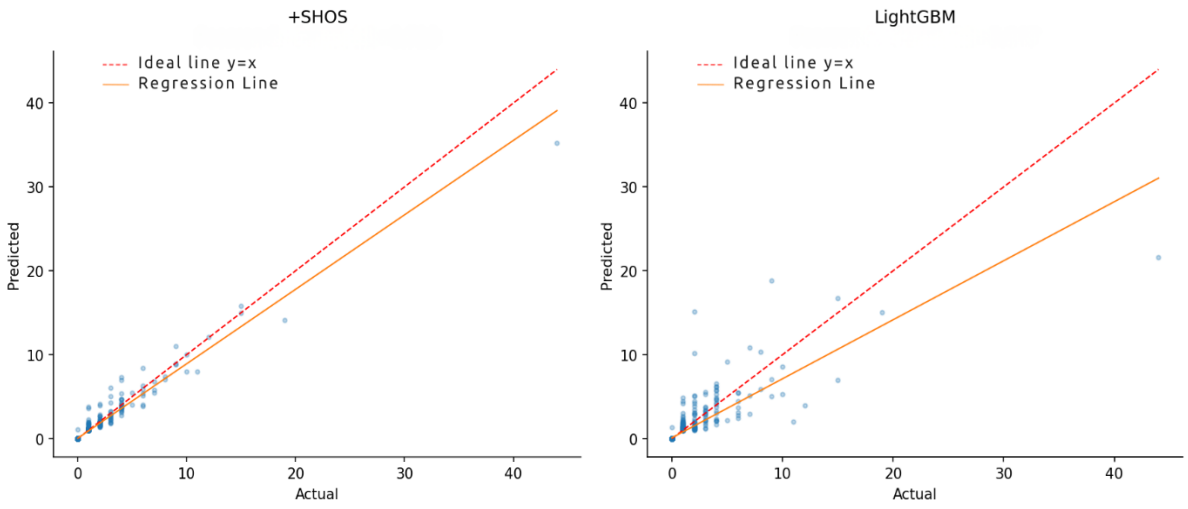
ARTICLE IN PRESS

spare parts, industrial maintenance inventories, healthcare consumables, and slow-moving retail items [33]. As the framework operates on aggregated demand histories and does not rely on domain-specific covariates, it can be readily transferred to other settings where intermittent demand is prevalent. Nevertheless, empirical validation in additional domains and under different temporal aggregation schemes remains an important direction for future work.

4.5 Diagnostic analysis of the top models

A diagnostic analysis of the top performing models was performed to assess their predictive behaviour. Figure 8 show scatter plots of predicted versus actual values for the first rolling validation fold for the baseline LightGBM and the +SHOS model, respectively. The points for both models are tightly clustered around the diagonal y=x line, indicating strong calibration and the absence of significant systematic bias. The +SHOS model demonstrates superior alignment with the true demand, as evidenced by its higher $R^2 = 0.86$ compared to the baseline $R^2 = 0.813$.

ARTICLE IN PRESS



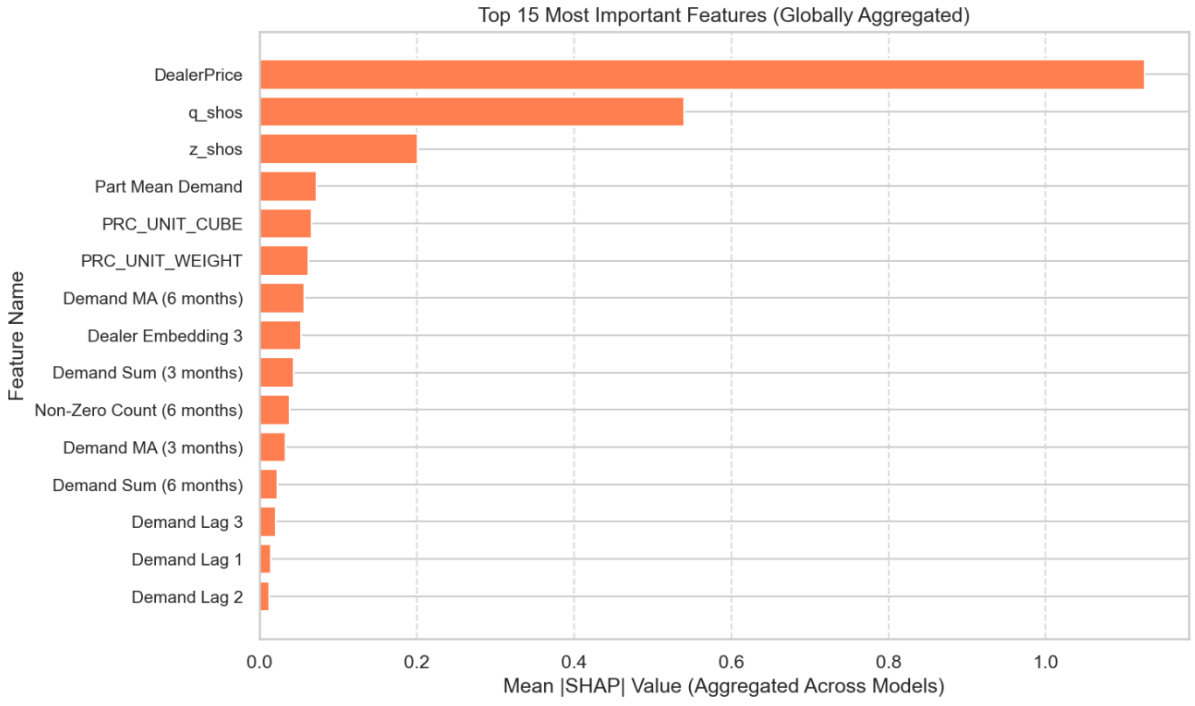
ARTICLE IN PRESS

Figure 8. Predicted versus actual demand values for the SHOS-enhanced single-stage model (+SHOS) and the LightGBM baseline, illustrating calibration and bias under intermittent demand

4.6 SHAP based interpretability analysis

To understand the drivers of predictive performance, SHAP (SHapley Additive exPlanations) analysis was performed on the all models. Figure 9 shows the global feature importance plot, where each point represents a SHAP value for a feature in a single prediction. The analysis reveals that the SHOS derived features q_shos (occurrence probability) and z_shos (conditional size) rank among the top three most important features globally, confirming their critical role in guiding the model's forecasts. This provides direct empirical support for the efficacy of the SHOS algorithm as a feature generator.

ARTICLE IN PRESS



ARTICLE IN PRESS

Figure 9. Global feature importance based on mean absolute SHAP values aggregated across forecasting models

The SHAP-based analysis provides further evidence for this interpretation. While conventional features such as recent lags, rolling averages, and static attributes contribute to model performance, SHOS-derived occurrence probability and conditional size consistently rank among the most influential predictors across models. This indicates that the benefit of SHOS features arises from their alignment with the underlying intermittent demand structure, rather than from arbitrary feature proliferation. Together, the ablation results and interpretability analysis form a systematic comparative evaluation, demonstrating that statistically grounded, domain-consistent features play a dominant role in stabilizing and improving forecasts under extreme intermittency.

ARTICLE IN PRESS

Beyond indicating global feature importance, the SHAP analysis provides direct insight into the model's decision-making process. The SHOS-derived occurrence probability and conditional size features consistently exhibit the largest absolute SHAP values, indicating that they contribute most strongly to individual predictions. This demonstrates that the model does not merely use SHOS features as auxiliary signals, but instead relies on them as primary drivers when forming demand estimates.

Positive SHAP values for the SHOS occurrence probability correspond to increased predicted demand during periods of recent activity, while negative contributions suppress predictions during extended zero-demand regimes. Similarly, the SHOS conditional size

feature governs the magnitude of forecasts once demand is deemed likely, allowing the model to decouple whether demand will occur from how much demand is expected.

The Figure 10 provides a side-by-side comparison of feature importance for the baseline LightGBM model (using base features only) and the +SHOS model (using full features). In the baseline model, static features like dealer price dominate, while in the +SHOS model, the SHOS features q_shos and z_shos become central, highlighting the shift in the model's decision-making process when provided with statistically grounded signals.

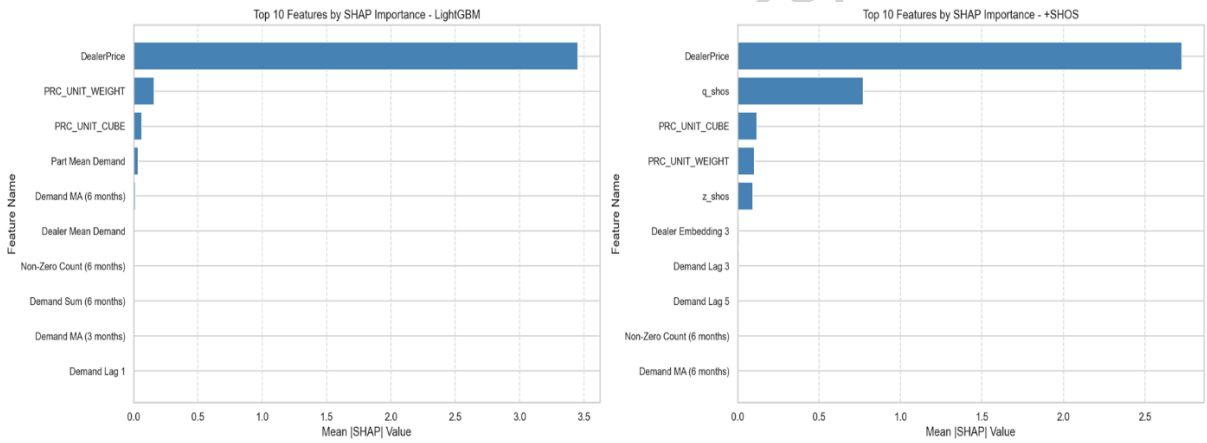


Figure 10. Comparison of SHAP-based feature importance for the baseline LightGBM model and the SHOS-enhanced model (+SHOS)

A representative case illustrates this behavior. For a highly intermittent dealer-part series characterized by long zero-demand runs, the baseline LightGBM model relies primarily on static features and recent lag values, often producing noisy or unstable predictions. In contrast, the

ARTICLE IN PRESS

+SHOS model first anchors its forecast using a low SHOS occurrence probability during inactivity, effectively dampening spurious demand signals. When demand reappears after prolonged inactivity, the recency-adjusted SHOS occurrence probability increases sharply, resulting in a corresponding positive SHAP contribution that elevates the forecast in a controlled manner. The SHOS conditional size feature then determines the forecast magnitude, preventing overreaction to isolated demand spikes.

This behavior confirms that SHOS features act as statistically grounded priors that shape the model's response to sparse and irregular demand patterns, rather than merely improving performance through indirect correlation effects.

The SHAP analysis for the hurdle models further illustrates this effect. As shown in Figure 11, for the occurrence classifier in the Hurdle_SHOS model, `q_shos` is a key input, while for the quantity regressor, `z_shos` is the dominant feature. This confirms that the two-stage architecture leverages the SHOS priors effectively but does not surpass the simpler single stage model in overall performance.

ARTICLE IN PRESS

ARTICLE IN PRESS

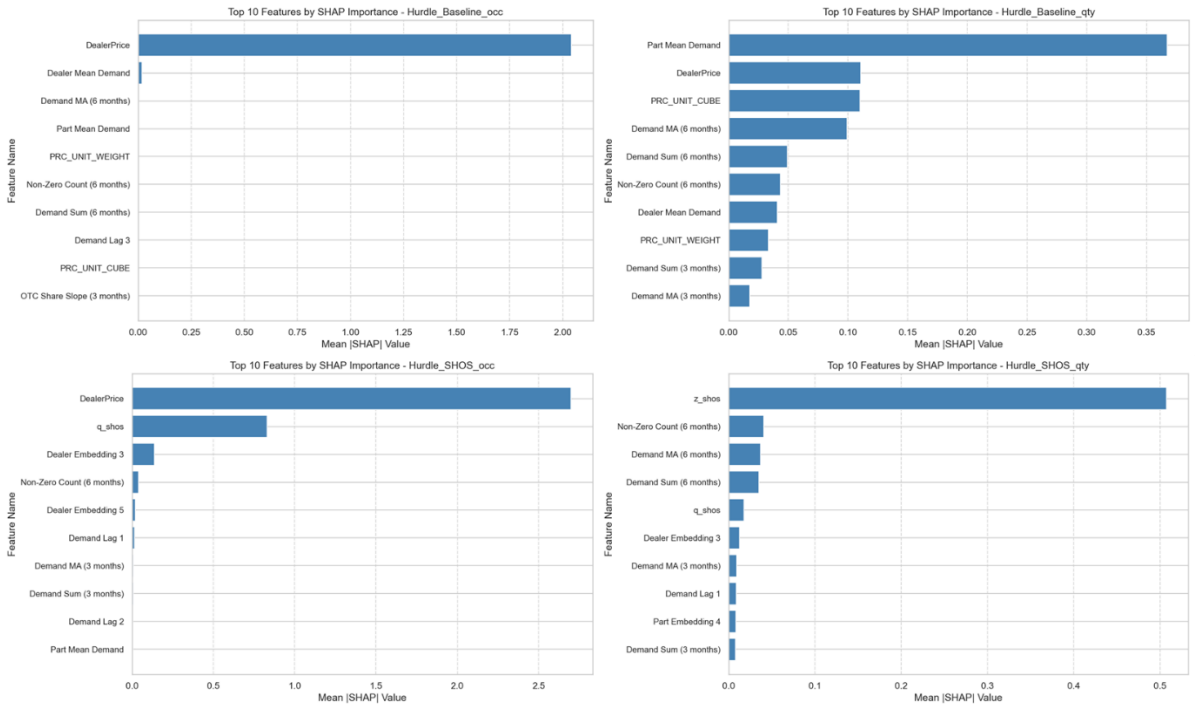


Figure 11. SHAP-based feature importance for two-stage hurdle model components, comparing occurrence classification and conditional quantity regression with and without SHOS features

From a process-level perspective, the effectiveness of the SHOS framework can be explained by the physical nature of intermittent demand generation in supply chains. Demand for slow-moving spare parts is typically driven by discrete operational events such as equipment failure, scheduled maintenance, or delayed replacement cycles, rather than continuous consumption. This results in long periods of inactivity punctuated by irregular demand events with highly variable magnitudes. SHOS explicitly models this mechanism by decoupling demand occurrence from conditional demand size and allowing both components to evolve adaptively over time. By doing so, SHOS-derived features encode latent demand readiness and reactivation dynamics that are not directly

ARTICLE IN PRESS

observable from raw demand histories. These features provide machine learning models with a physically meaningful representation of the underlying demand-generating process, explaining why improved generalization is achieved without increasing architectural complexity [34].

4.7 The Dominance of statistically grounded feature engineering

The performance gains observed with SHOS-enhanced models should be interpreted in the context of systematic feature comparison rather than as an inherent dominance of a specific feature set. The central question addressed in this study is not whether SHOS features are universally superior, but whether features that encode the structural properties of intermittent demand provide more informative signals than generic lag-based or static covariates. This question is evaluated through controlled ablation experiments, where identical models are trained with and without SHOS-derived features, and through interpretability analysis that reveals how different feature groups contribute to predictions.

ARTICLE IN PRESS

The most significant finding of this study is the dramatic performance improvement achieved by integrating SHOS derived priors into a standard machine learning framework. The comparison between the baseline LightGBM Regressor (MAE = 0.1417, RMSE = 1.3657) and the +SHOS model (MAE = 0.0712, RMSE = 1.0310) demonstrates that the core of the predictive power comes not from architectural novelty, but from providing the machine learning model with a stable, statistically

ARTICLE IN PRESS

grounded signal of the underlying demand pattern a concept aligned with recent advances in feature based forecast combination [18].

The SHOS algorithm effectively acts as a denoising and signal processing layer. While a standard ML model struggles to learn from the raw, sparse, and noisy time series [9], the SHOS features provide robust, adaptive estimates of demand occurrence probability q_{shos} and conditional size (z_{shos}) The machine learning model can then learn nuanced, nonlinear adjustments to this statistical prior using rich temporal and static features. This synergy where SHOS supplies a strong baseline and the ML model captures residual signal is the primary driver of the observed performance gains.

This insight is empirically validated by SHAP analysis (Figure 9), which shows that q_{shos} and z_{shos} rank among the top three most important features globally, confirming their central role in the model's decision-making process.

ARTICLE IN PRESS

4.8. The case for simplicity: feature quality over architectural complexity

This study's second critical insight emerges from the comparison between the +SHOS model and the more complex Hurdle_SHOS architecture. The results show that the simpler single stage regressor achieves comparable or slightly better performance on key operational metrics: MAE (0.0712 vs. 0.0701), WMAPE (20.39% vs. 20.10%), and R^2 (0.8430 vs. 0.8186). Although the Hurdle model has a marginally lower MAE, the +SHOS model achieves a higher R^2 and demonstrates superior

ARTICLE IN PRESS

calibration (R = 0.860 vs. lower for Hurdle variants), indicating better variance capture.

This finding challenges the assumption that specialized, multi stage architectures such as zero inflated or hurdle models [35,36] are inherently superior for intermittent demand. A likely explanation is the avoidance of error propagation in the two-stage framework, miscalibration in the occurrence classifier directly compounds in the final forecast. In contrast, the single stage model learns a direct, end to end mapping from features including SHOS priors to expected demand, resulting in greater robustness. Importantly, these conclusions are not based on isolated model comparisons but on controlled ablation experiments that independently vary feature engineering and architectural complexity. This structured evaluation supports the central claim that, for intermittent demand forecasting, high-quality statistically grounded features dominate the choice of model architecture.

ARTICLE IN PRESS

These results strongly support the principle of parsimony: when equipped with high quality statistical features, a simple architecture not only matches but often exceeds the performance of more complex alternatives.

4.9 Practical implications for inventory management

From a practical point of view, the performance of the +SHOS model in error metrics such as RMSE and MAE is critical for supply chain operations. Forecast errors in intermittent demand contexts are highly asymmetric [37]: under prediction causes stockouts and lost sales [3],

ARTICLE IN PRESS

while over prediction leads to obsolescence and holding costs [6]. The +SHOS model's low MAE and tight error distribution confirmed by scatter plots showing strong alignment with actuals (Figure 8) indicate it is both accurate and reliable, providing the statistical foundation necessary to minimize costly deviations.

Furthermore, the model's simplicity enhances deplorability. A single stage LightGBM regressor is easier to train, maintain, monitor, and explain to stakeholders than a multi component pipeline. This makes the proposed methodology not only accuracy but also practical for real world implementation in large scale automotive supply chains. Finally, overfitting diagnostics confirm that the model generalizes well indicating robust out of sample performance essential for dynamic inventory planning under uncertainty [38].

ARTICLE IN PRESS

While the empirical evaluation in this study is conducted on an automotive aftermarket spare-parts dataset, this setting represents a prototypical intermittent demand environment characterized by extreme sparsity, long inactivity periods, and irregular demand reactivation. As such, it provides a rigorous testbed for assessing forecasting methods under challenging intermittency conditions. Nevertheless, the reported results should be interpreted within the scope of this domain, and direct extrapolation to other industries should be made with appropriate caution.

The reliance on fully zero-filled monthly panel data has important statistical implications. Zero-filling establishes a regular temporal structure that allows for consistent feature construction, rolling-window

ARTICLE IN PRESS

validation, and fair model comparison. However, it also encodes a specific assumption regarding the demand-generating process namely that periods without transactions correspond to true demand absence rather than censored or unobserved events. While this assumption is standard in intermittent demand forecasting, alternative representations based on irregular-time or event-driven modeling may lead to different conclusions and warrant further investigation.

The conclusions drawn in this study should be interpreted within the scope of the evaluated dataset, temporal aggregation level, and modelling framework. While the results demonstrate that statistically grounded feature representations dominate architectural complexity in the present setting, alternative feature designs, hyperparameter optimization strategies, or evaluation protocols may influence outcomes under different conditions. In this study, hyperparameter tuning, preprocessing, and evaluation procedures are held constant across models to isolate the effects of feature representation and architectural structure.

ARTICLE IN PRESS

Beyond performance improvements, the present study reveals a notable and study-specific phenomenon in intermittent demand forecasting: once demand sparsity and occurrence dynamics are explicitly captured through statistically grounded features, additional architectural complexity offers limited marginal benefit. Despite the widespread assumption that multi-stage or hurdle-based structures are necessary to model intermittent demand, our results demonstrate that a single-stage learning framework enriched with SHOS-derived features can more

ARTICLE IN PRESS

effectively internalize demand occurrence and magnitude information. This finding highlights the dominant role of representation quality over structural complexity in large-scale intermittent demand forecasting.

5. Conclusions

Forecasting intermittent demand remains a persistent challenge in large-scale supply chains due to extreme sparsity, long zero-demand periods, and irregular demand reactivation. In this study, we investigated whether improvements in forecasting performance arise primarily from increasing model architectural complexity or from incorporating statistically grounded, domain-consistent features. To this end, we introduced the SHOS (Smoothed Hybrid Occurrence Size) algorithm as an adaptive feature-generation mechanism and evaluated its effectiveness within a large-scale automotive aftermarket spare-parts dataset. The key findings of this study can be summarized as follows:

ARTICLE IN PRESS

1. Statistically grounded feature engineering is the dominant driver of performance.

   Incorporating SHOS-derived features specifically smoothed demand occurrence probability and conditional demand size into standard machine learning regressors leads to substantial improvements in forecasting accuracy compared to models relying solely on conventional temporal and static features.

2. Adaptive, series-specific priors improve robustness under extreme intermittency.

   The SHOS algorithm stabilizes learning in highly sparse demand

ARTICLE IN PRESS

series by dynamically adjusting smoothing strength based on series-level intermittency and by incorporating recency-aware updates that enable controlled adaptation when demand reactivates.

3. Architectural complexity provides limited additional benefit once informative features are available. Controlled ablation experiments demonstrate that a single-stage LightGBM regressor augmented with SHOS features achieves performance comparable to, or better than, a more complex two-stage hurdle architecture trained on the same feature set. This indicates that explicit architectural decomposition is not necessary when high-quality statistical features already encode the underlying demand structure.

4. Interpretability analysis confirms the mechanistic role of SHOS features. SHAP-based analysis shows that SHOS-derived occurrence probability and conditional size consistently rank among the most influential predictors, directly shaping model decisions by suppressing spurious predictions during inactivity and guiding magnitude estimation when demand occurs.

5. The conclusions are supported within the evaluated intermittent demand context. All empirical results are based on a large-scale automotive spare-parts dataset aggregated at a monthly level, which represents a canonical and operationally relevant example of extreme intermittent demand.

From a practical deployment perspective, the SHOS algorithm is computationally lightweight and introduces minimal additional cost

ARTICLE IN PRESS

beyond standard exponential smoothing operations. The feature-generation process operates in linear time with respect to the number of time periods and does not require iterative optimization or model retraining, making it suitable for large-scale supply chain systems with tens of thousands of items. SHOS-derived features can be computed offline or incrementally and integrated directly into existing machine learning pipelines without modifying downstream model architectures. Implementation constraints are therefore primarily related to data availability and aggregation choices rather than computational burden. As with other feature-based approaches, the effectiveness of SHOS depends on consistent temporal granularity and reliable historical demand records, which are standard in most enterprise resource planning systems. These characteristics make SHOS a practical and scalable solution for real-world intermittent demand forecasting applications.

ARTICLE IN PRESS

While the results clearly demonstrate the effectiveness of SHOS-enhanced feature engineering within the studied setting, the findings should be interpreted in the context of the evaluated data characteristics and aggregation level. Future work will focus on validating the proposed framework across additional intermittent demand domains, such as high-volatility retail and seasonal consumer goods, as well as at alternative temporal resolutions. Further extensions include integrating exogenous drivers (e.g., promotions or pricing effects) and learning SHOS smoothing parameters in a data-driven or meta-learning framework. Collectively, this study highlights that carefully designed, statistically informed features can

ARTICLE IN PRESS

offer a practical, interpretable, and scalable alternative to increasing architectural complexity in intermittent demand forecasting.

Data Availability

All the data and material used in this study is available in the manuscript, and further details if required, the corresponding author will provide the same, through proper requisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Conflict of Interest

All the authors declare that they have no conflict of interest.

Contributions

S. N. B.: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Investigation (equal); Methodology (equal); Writing-original draft (lead); Writing-review & editing (equal). **A. P. M.**: Conceptualization (equal); Methodology (equal); Formal analysis (lead); Investigation (equal); Writing & editing (equal). **B. V. S. R.**: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Investigation (equal); Methodology (equal); Writing-original draft (lead); Writing & editing (equal). **C. C. S.**: Conceptualization (equal); Methodology (equal); Formal analysis (lead); Funding acquisition (lead); Supervision (lead); Investigation (equal); Writing & editing (lead). **S. S.**: Funding acquisition (lead); Investigation (equal); Writing & editing (equal). **R. C.**: Funding acquisition (lead); Investigation (equal); Writing & editing (equal).

ARTICLE IN PRESS

Funding

This work was co-funded by the European Union under the REFRESH - Research Excellence for Region Sustainability and High-tech Industries project (Project No. CZ.10.03.01/00/22_003/0000048) via the Operational Programme Just Transition. This article was also supported by the Students Grant Competition SP2024/087, Specific Research of Sustainable Manufacturing Technologies, financed by the Ministry of Education, Youth and Sports (MEYS), Czech Republic, and the Faculty of Mechanical Engineering, VŠB-Technical University of Ostrava.

References

1. K. Nikolopoulos. We need to talk about intermittent demand forecasting. *European Journal of Operational Research* **291**, 549-559 (2021). doi.org/10.1016/j.ejor.2019.12.046.
2. N. Kourentzes, G. Athanasopoulos. Elucidate structure in intermittent demand series. *European Journal of Operational Research* **288**, 141-152 (2021). doi.org/10.1016/j.ejor.2020.05.046
3. Ç. . Pinçe, L. Turrini, J. Meissner, Intermittent demand forecasting for spare parts: A critical review, Omega (United Kingdom) 105 (2021) 102513. doi.org/10.1016/j.omega.2021.102513
4. J. D. Croston. Forecasting and Stock Control for Intermittent Demands. *Operational Research Quarterly* **23**, 289-303 (1972).
5. A. A. Syntetos, J. E. Boylan. The accuracy of intermittent demand estimates. *International Journal of Forecasting* **21**, 303-314 (2005). doi.org/10.1016/j.ijforecast.2004.10.001
6. R. H. Teunter, A. A. Syntetos, M. Z. Babai, Intermittent demand: Linking forecasting to inventory obsolescence, European Journal of Operational Research 214 (3) (2011) 606-615 doi.org/10.1016/j.ejor.2011.05.018
7. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. 3146-3154 In *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., 2017.
8. T. Chen, C. Guestrin. XGBoost: A Scalable Tree Boosting System. 785-794 In *Proceedings of the 22nd ACM SIGKDD International*

ARTICLE IN PRESS

Conference on Knowledge Discovery and Data Mining. ACM, 2016. doi.org/10.1145/2939672.2939785

9. M. Z. Babai, M. Arampatzis, M. Hasni, F. Lolli, A. Tsadiras. On the use of machine learning in supply chain management: a systematic review. *IMA Journal of Management Mathematics* **36**, 21-49 (2025). doi.org/10.1093/imaman/dpae029

10. R. S. Gutierrez, A. O. Solis, S. Mukhopadhyay. Lumpy demand forecasting using neural networks. *International Journal of Production Economics* **111**, 409-420 (2008). doi.org/10.1016/j.ijpe.2007.01.007

11. S. Makridakis, E. Spiliotis, V. Assimakopoulos. The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting* **34**, 802-808 (2018). doi.org/10.1016/j.ijforecast.2018.06.001

12. Spyros Makridakis, Evangelos Spiliotis, Vassilios Assimakopoulos. The M5 competition: Background, organization, and implementation. *International Journal of Forecasting* **38**, 1325-1336 (2022). doi.org/10.1016/j.ijforecast.2021.07.007

13. A. A. Syntetos, J. E. Boylan. On the bias of intermittent demand estimates. *International Journal of Production Economics* **71**, 457-466 (2001). doi.org/10.1016/S0925-5273(00)00143-2

14. A. H. C. Eaves, B. G. Kingsman. Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society* **55**, 431-437 (2004). doi.org/10.1057/palgrave.jors.2601697

15. R. H. Teunter, A. A. Syntetos, M. Z. Babai. Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research* **214**, 606-615 (2011). doi.org/10.1016/j.ejor.2011.05.018

16. G. Zhang, B. E. Patuwo, M. Y. Hu, Forecasting with artificial neural networks: The state of the art, International Journal of Forecasting 14 (1) (1998) 35-62. doi.org/10.1016/S0169-2070(97)00044-7

17. S. Smyl. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting* **36**, 75-85 (2020). doi.org/10.1016/j.ijforecast.2019.03.017

18. L. Li, Y. Kang, F. Petropoulos, F. Li. Feature-based intermittent demand forecast combinations: accuracy and inventory implications. *International Journal of Production Research* **61**, 7557-7572 (2023). doi.org/10.1080/00207543.2022.2153941

19. Ma, T., Lin, Y., Zhou, X., & Zhang, M. (2022). Grading evaluation of goaf stability based on entropy and normal cloud

ARTICLE IN PRESS

model. *Advances in Civil Engineering*, *2022*(1), 9600909. https://doi.org/10.1155/2022/9600909

20. Xu, B., Tan, Y., Sun, W., Ma, T., Liu, H., & Wang, D. (2023). Study on the prediction of the uniaxial compressive strength of rock based on the SSA-XGBoost model. *Sustainability*, *15*(6), 5201. https://doi.org/10.3390/su15065201
21. Cheng, Y., He, D., Ma, T., Lin, H., Hu, X., & Liu, H. (2024). Hybrid data-driven model and shapley additive explanations for peak dilation angle of rock discontinuities. *Materials Today Communications*, *40*, 110194. https://doi.org/10.1016/j.mtcomm.2024.110194
22. Ma, T., Hu, X., Liu, H., Peng, K., Lin, Y., Chen, Y., ... & Chen, M. (2025). Elastic modulus prediction for high-temperature treated rock using multi-step hybrid ensemble model combined with coronavirus herd immunity optimizer. *Measurement*, *240*, 115596. https://doi.org/10.1016/j.measurement.2024.115596
23. Ma, T., Chen, C., Shen, L., Luo, K., Jiang, Z., Xie, S., ... & Peng, K. (2025). Physics-informed neural networks for capturing the true relationships between parameters to predict the dynamic triaxial strength of rocks in cold environments. *Measurement*, 118900. https://doi.org/10.1016/j.measurement.2025.118900
24. Shen, L., Ma, T., Chen, H., Ye, B., Luo, R., & Sun, H. (2025). A new CNN-GRU deep learning framework optimized by CHIO for precise prediction of debris flow velocity. *Stochastic Environmental Research and Risk Assessment*, 1-21. https://doi.org/10.1007/s00477-025-02973-7
25. Ma, T., Shen, L., Zhang, M., Luo, K., Hu, X., Jiang, Z., ... & Peng, K. (2025). Hybrid Empirical-Data-Driven Neural Network for Predicting Air-Entry Value in Unsaturated Soils. *Mathematical Geosciences*, 1-30. https://doi.org/10.1007/s11004-025-10230-4
26. Liu, H., Ma, T., Lin, Y., Peng, K., Hu, X., Xie, S., & Luo, K. (2024). Deep learning in rockburst intensity level prediction: performance evaluation and comparison of the NGO-CNN-BiGRU-attention model. *Applied Sciences*, *14*(13), 5719. https://doi.org/10.3390/app14135719
27. Wang, Y., Ma, T., Shen, L., Wang, X., & Luo, R. (2025). Prediction of thermal conductivity of natural rock materials using LLE-transformer-lightGBM model for geothermal energy applications. *Energy Reports*, *13*, 2516-2530. https://doi.org/10.1016/j.egyr.2025.02.003
28. Xie, S., Lin, H., Ma, T., Peng, K., & Sun, Z. (2025). Prediction of joint roughness coefficient via hybrid machine learning model combined with principal components analysis. *Journal of Rock Mechanics and Geotechnical Engineering*, *17*(4), 2291-2306. https://doi.org/10.1016/j.jrmge.2024.05.059
29. Ma, T., Hu, X., Liu, H., Peng, K., Lin, Y., Chen, Y., ... & Chen, M. (2025). Elastic modulus prediction for high-temperature treated rock using multi-step hybrid ensemble model combined with

ARTICLE IN PRESS

coronavirus herd immunity optimizer. *Measurement*, *240*, 115596. https://doi.org/10.1016/j.measurement.2024.115596

30. Nathan, B. S., Siva Reddy, B. V., Sastry, C. C., Krishnaiah, J., & Eswaramoorthy, K. V. (2024). Innovative framework for effective service parts management in the automotive industry. *Frontiers in Mechanical Engineering*, *10*, 1361688. https://doi.org/10.3389/fmech.2024.1361688

31. B, S.N., Reddy, B.V.S., Sujan, V.S.S. *et al.* A machine learning framework for long-term forecasting of spare part demand in end-of-life product scenarios. *Sci Rep* (2025). https://doi.org/10.1038/s41598-025-31171-2

32. Bobbili, V. S. R., Chandrasekhara Sastry, C., Krishnaiah, J., Hafeezur Rahman, A., Surya Kumar, S., & Subba Reddy, N. (2025). Physics-informed neural networks for predicting high-strain-rate energy absorption in additively manufactured lattice materials. *Progress in Additive Manufacturing*, 1-27. https://doi.org/10.1007/s40964-025-01460-3

33. Reddy, B. V. S., Shaik, A. M., Sastry, C. C., Krishnaiah, J., Bhise, C. A., & Ramakrishna, B. (2024). Machine learning approaches for predicting mechanical properties in additive manufactured lattice structures. *Materials Today Communications*, *40*, 109937. https://doi.org/10.1016/j.mtcomm.2024.109937

34. Reddy, B. V. S., Shaik, A. M., Sastry, C. C., Krishnaiah, J., Patil, S., & Nikhare, C. P. (2025). Performance evaluation of machine learning techniques in surface roughness prediction for 3D printed micro-lattice structures. *Journal of Manufacturing Processes*, *137*, 320-341. https://doi.org/10.1016/j.jmapro.2025.01.082

35. M. Lichman, P. Smyth. Prediction of sparse user-item consumption rates with zero-inflated poisson regression. 719-728 In *The Web Conference 2018 - Proceedings of the World Wide Web Conference, WWW 2018*. ACM, 2018. doi.org/10.1145/3178876.3186153

36. P. Wallström, A. Segerstedt. Evaluation of forecasting error measurements and techniques for intermittent demand. *International Journal of Production Economics* **128**, 625-636 (2010). doi.org/10.1016/j.ijpe.2010.07.013

37. A. C. Cameron, P. K. Trivedi. Regression Analysis of Count Data. Cambridge University Press, 2012.

38. N. Kourentzes. On intermittent demand model optimisation and selection. *International Journal of Production Economics* **156**, 180-190 (2014). doi.org/10.1016/j.ijpe.2014.06.007