

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452

NAAS Rating (2026): 4.49

Maths 2026; 11(7): 58-69

© 2026 Stats & Maths

<https://www.mathsjournal.com>

Received: 17-04-2026

Accepted: 22-05-2026

Ododa K Otieno

School of Biological Physical
Mathematics and Actuarial
Sciences Jaramogi Oginga
Odinga University of Science and
Technology P.O. Bondo, Kenya

Oduor D Brian

School of Biological Physical
Mathematics and Actuarial
Sciences Jaramogi Oginga
Odinga University of Science and
Technology P.O. Bondo, Kenya

Maua M Zablon

School of Biological Physical
Mathematics and Actuarial
Sciences Jaramogi Oginga
Odinga University of Science and
Technology P.O. Bondo, Kenya

Corresponding Author:

Ododa K Otieno

School of Biological Physical
Mathematics and Actuarial
Sciences Jaramogi Oginga
Odinga University of Science and
Technology P.O. Bondo, Kenya

Optimizing model selection and forecasting accuracy in seasonal time series using modern experimental design approaches (DOE)

Ododa K Otieno, Oduor D Brian and Maua M Zablon

DOI: <https://www.doi.org/10.22271/math.2026.v11.i7a.2623>

Abstract

Accurate seasonal time-series forecasting is essential in agriculture, finance, energy, and climate-sensitive decision-making contexts, where future planning depends on the reliable identification of trend, seasonal, and residual patterns. This study investigates how Design of Experiments (DOE) can be used to guide model selection and improve forecasting performance through a structured comparison of traditional, machine learning, and hybrid forecasting models. The study uses the Global Agriculture Climate Impact Dataset from Kaggle as a methodological benchmark because it contains climate, agricultural, environmental, and economic variables suitable for seasonal forecasting evaluation. A factorial DOE framework was developed using decomposition method, forecast model, hybrid strategy, and forecast horizon as experimental factors. The tested configurations included STL, X-13 ARIMA-SEATS, and seasonal naïve decomposition combined with SARIMAX, LSTM, XGBoost, and Prophet models. Forecasting performance was evaluated using RMSE, MAE, and MAPE, while ANOVA, block bootstrap analysis, residual diagnostics, and the Diebold-Mariano test were used to examine statistical robustness. The findings show that STL-based hybrid models, particularly STL-LSTM, achieved lower forecast errors than traditional benchmarks across most evaluated configurations. However, the Diebold-Mariano test did not detect a statistically significant difference between STL-LSTM and STL-SARIMAX at the 5% significance level, indicating that the superiority of the hybrid model should be interpreted cautiously. Overall, the study demonstrates that DOE provides a systematic and reproducible framework for evaluating forecasting configurations, reducing informal trial-and-error model selection, and improving the transparency of seasonal forecasting experiments.

Keywords: Design of experiments, seasonal time-series forecasting, STL decomposition, LSTM, XGBOOST, SARIMAX, hybrid forecasting, ANOVA, Diebold-Mariano test, forecast accuracy

Introduction

Accurate forecasting plays a pivotal role in a range of sectors, including agriculture, finance, and energy, where timely and precise predictions can significantly influence decision-making processes (Hasan *et al.*, 2025) ^[14]. In agriculture, seasonal weather patterns/crop yields could be predicted to assist farmers to minimize losses, increase their profits and optimizing their planting and harvesting schedules (Chatte, 2025) ^[17]. In finance, proper forecasting of the market trends and demand is capable of informing investment strategy, risk management and resource allocation (Kumar, 2023) ^[25]. Likewise, in energy management, demand fluctuations may be predicted, which can assist the utilities balance their supply and consumption and reduce the costs of operations and enhance sustainability (Hossain *et al.*, 2025) ^[17]. Thus, forecasting accuracy is a vital objective that supports planning and resource utilization in these sectors.

The effectiveness of the forecasting models is directly related to the choice of the suitable models and the possibility of the models to fit the underlying patterns in the data. When it comes to time series forecasting, it is not as straightforward as picking a statistical or machine learning method. It involves knowing the nature of the data including seasonality, trends, and noise (Proietti & Pedregal, 2023) ^[31].

Since time series data may have patterns that vary with time, it is important to choose a model that is capable of adapting to the changes in the patterns. But the mainstream time series models, including ARIMA (Autoregressive Integrated Moving Average) and SARIMA (Seasonal ARIMA) are highly deficient in dealing with non-linear, multi seasonal and non-stationary data (Kontopoulou *et al.*, 2023) ^[24]. These models, though useful with some data, are challenged by the complexities of real-world time series, where seasonal effects can vary in magnitude or even direction over time.

Recent developments in machine learning, including Long Short-Term Memory (LSTM) networks and XGBoost, have become the potent instruments in predicting non-linear and complicated data (Ahmed *et al.*, 2023; Tufail *et al.*, 2023) ^[1, 37]. LSTM is a recurrent neural network (RNN) variant that can learn long-term dependencies in the sequence of data, and thus it can be effectively used in time series forecasting (Bandara *et al.*, 2020) ^[4]. XGBoost is a powerful gradient boosting algorithm that is capable of working with large datasets and multiple feature interactions (Ekmekcioğlu, 2023) ^[9]. Although these approaches have enhanced the accuracy of forecasting, they are associated with interpretability and data issues. According to Górriz *et al.* (2023) ^[11] and Hassija *et al.* (2024), Deep learning models, in particular LSTM methods, are usually black boxes, in which it can be challenging to interpret the way a specific model makes a forecast. Moreover, such models can often consume large volumes of high-quality data to be trained, and such data might not necessarily be available in some industries or to be used in applications (Mohammed *et al.*, 2025) ^[30].

The combination of machine learning models and statistical decomposition techniques can be a promising solution to these issues. One such method is Seasonal-Trend decomposition using LOESS (STL), which separates a time series into trend, seasonal, and residual components (Sohrabbeig *et al.*, 2023; Trull *et al.*, 2022) ^[33, 36]. Decomposing the series, machine learning models can concentrate on predicting either the residuals or the trends, which enhances the predictability and the understandability of the predictions (Hasan, 2024) ^[15]. This is a hybrid model, integrating the strengths of classic decomposition models with the predictive ability of machine learning to offer a systematic means of dealing with the complexities of non-linear, multi-seasonal and non-stationary time series data. Moreover, the contemporary methods of experimental design (DOE) can provide an outline of how to optimize the model setup and choose the most efficient decomposition and machine learning strategies. DOE assists in the systematic exploration of the various model configurations, such that the model adopted offers the optimum performance to the specified forecasting task.

Although both the statistical and machine learning models have improved, it is observed that there is a significant gap in the literature about systematic use of Design of Experiments (DOE) to optimize hybrid forecasting models. Although some researchers studied hybrid frameworks based on the combination of the statistical approach such as STL and the machine learning approach such as LSTM or XGBoost, few of them have applied the principles of DOE to the optimization of these frameworks. Current literature is inclined to informally implement hybrid models but does not present a formal structure to assess the efficiency of various models, which restricts the opportunities of model optimization. Consequently, standardized procedures of

identifying the most appropriate combination of methods to use in various forecasting processes are lacking.

The issues of the existing model selection techniques in time series forecasting are multifold. First, the classical methods of model selection, including grid search or random search, can be computationally intensive and do not possess the systematic rigor to determine the most effective combinations. Moreover, these techniques can be more about parameter tuning and less about the overall experimental design, like what type of decomposition method to use or how the statistical and machine learning models interact. With the increasingly complicated types of time series data, there is a growing interest in a systematic and efficient method of model selection.

The main objective of the research is to assess the ability of modern experimental design (DOE) to improve model selection and forecasting accuracy for seasonal time series. These aims are to test hybrid model configurations, determine the most efficient decomposition and machine learning models, and assess the role of DOE in maximizing model performance by systematically testing the model. This research paper makes a considerable contribution to literature and practice. In academia, it presents a systematic method for time series forecasting that integrates experimental design with hybrid modeling and has been underutilized in the literature. The study offers a strict model selection and optimization framework that can be applied to a range of forecasting tasks by using the DOE principles. The results will contribute to understanding how various settings for decomposition procedures and machine learning models can be critically compared to improving forecast accuracy.

This study makes three main contributions. First, it proposes a DOE-guided model selection framework for seasonal time-series forecasting, where decomposition method, forecasting model, hybrid strategy, and forecast horizon are treated as experimental factors rather than selected through informal trial-and-error. Second, it evaluates the practical suitability of STL-based hybrid models by comparing them with traditional statistical and machine learning forecasting approaches under the same experimental structure. Third, it improves the transparency of hybrid forecasting evaluation by linking model performance metrics with statistical validation methods, including ANOVA, block bootstrap analysis, residual diagnostics, and the Diebold-Mariano test. In this way, the study contributes a reproducible experimental structure for evaluating seasonal forecasting models rather than only reporting isolated model accuracy values.

2. Methodology

2.1 Research Design

In this study, a quantitative methodology is used, which combines Design of Experiments (DOE), statistical decomposition, and machine learning models to maximize the predictive accuracy of seasonal time series. DOE is used to give a systematic analysis of different combinations of model parameters that would enable the finding of the most effective combination of decomposition techniques and machine learning models to use in seasonal forecasting. A research design would maximize the selection of models by experimenting with various factor configurations, e.g., decomposition (e.g. STL, X-13 ARIMA-SEATS) and machine learning (e.g. LSTM, XGBoost) model settings, and also by ensuring that the model settings are robust and optimize under different conditions. DOE framework is a systematic investigation of the relationships between the most

important factors and evaluating their impacts on the prediction of performance. This method guarantees that the models chosen are not only the most appropriate to fit the data, but also to optimize the experimental setting provided.

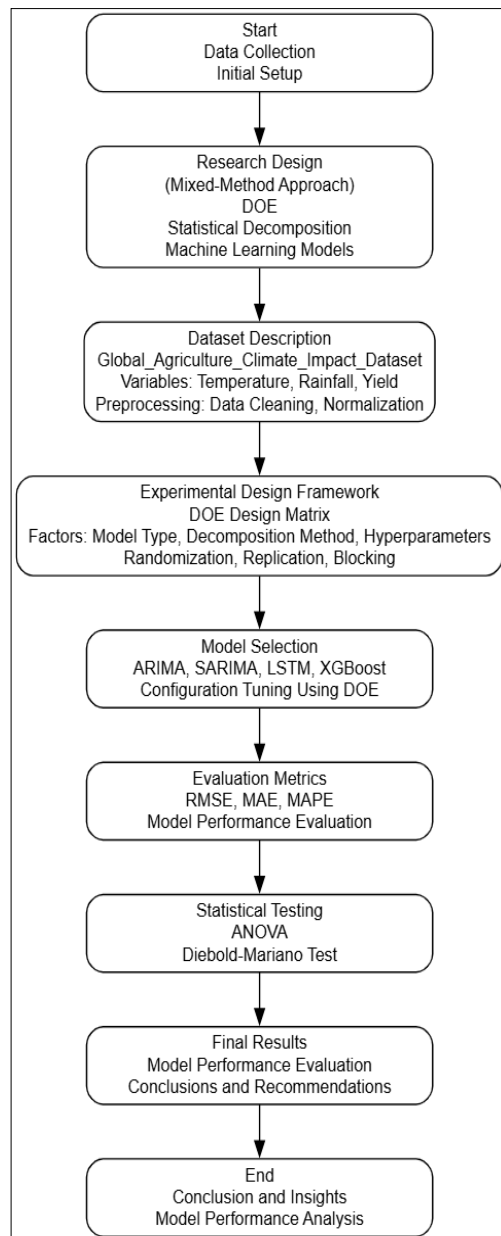


Fig 1: Proposed Methodology Framework

Figure 1 shows the systematic approach to evaluate the effectiveness of the modern experimental design methods (DOE) to enhance the model selection and accuracy of predictions. It starts with the data collection and preprocessing and continues with the optimization of decomposition methods and machine learning models by the use of DOE. It consists of selecting models (ARIMA, SARIMA, LSTM, XGBoost), testing them on conditions such as RMSE, MAE and MAPE, and statistical testing (ANOVA,

Diebold-Mariano test) to compare the performance of each model. Lastly, evaluation of the results is conducted to come up with conclusions and recommendations that can be used to enhance the accuracy of forecasting.

2.2 Dataset Description

The study uses the Global Agriculture Climate Impact Dataset from Kaggle as the primary dataset for evaluating the proposed DOE-guided forecasting framework. The dataset includes climate, agricultural, environmental, and economic variables, including crop type, average temperature, precipitation, CO₂ emissions, soil health index, crop yield, and economic impact indicators. These variables are relevant to seasonal forecasting because agricultural productivity is strongly influenced by climate variability, recurring seasonal patterns, and long-term environmental change.

The dataset was selected for three reasons. First, it provides a structured and publicly accessible benchmark for testing seasonal forecasting models. Second, it contains both numerical and categorical variables that allow the evaluation of forecasting models under realistic multivariable conditions. Third, the dataset is suitable for testing decomposition-based hybrid forecasting because the target variable, crop yield, is expected to contain trend, seasonal, and residual components. Although Kaggle is not an official statistical repository, the dataset is appropriate for methodological evaluation because the purpose of this study is to compare forecasting configurations under a controlled DOE framework rather than to produce official agricultural forecasts. Nevertheless, this limitation is acknowledged, and the findings should be interpreted as evidence from a single publicly available dataset. Future studies should validate the proposed DOE-guided framework using official agricultural, climate, or energy datasets.

Before modelling, the dataset was cleaned by checking missing values, removing duplicated records, identifying outliers, and standardising numerical variables. The target variable for forecasting was crop yield. Climate and environmental variables were used as explanatory features. The data were arranged in chronological order before model training to avoid data leakage. The dataset was divided into training, validation, and testing subsets using a time-series split. The training subset was used to fit the models, the validation subset was used for hyperparameter tuning, and the testing subset was reserved for final model evaluation.

2.3 DOE Factorial Design and Experimental Matrix

A factorial Design of Experiments approach was used to evaluate the effect of different modelling choices on forecasting accuracy. In this study, DOE was not treated only as a conceptual framework but as a structured experimental design in which each forecasting configuration represented one experimental run. The response variables were RMSE, MAE, and MAPE, while the main experimental factors were decomposition method, forecasting model, hybrid strategy, and forecast horizon.

Table 1: DOE Factors and Levels

Factor	Description	Levels Used in This Study
Factor A: Decomposition method	Method used to separate seasonal, trend, and residual components	STL, X-13 ARIMA-SEATS, Seasonal Naïve
Factor B: Forecasting model	Model used to predict the target variable or decomposed components	SARIMAX, LSTM, XGBoost, Prophet
Factor C: Hybrid strategy	Method used to combine decomposition and forecasting model	None, Sequential, Parallel
Factor D: Forecast horizon	Forecasting period used to test model robustness	1-year, 2-year, 3-year

The DOE design matrix was constructed by combining factor levels into experimental configurations. Each configuration was evaluated using the same preprocessing procedure, train-validation-test split, and performance metrics. Replication was included by repeating each successful model configuration using multiple random seeds where applicable.

Randomisation was applied to the order of model execution to reduce execution-order bias, particularly for machine learning models. Blocking was applied by forecast horizon so that model performance could be compared separately for 1-year, 2-year, and 3-year forecasting tasks.

Table 2: DOE Experimental Design Matrix

Run	Decomposition Method	Forecast Model	Hybrid Strategy	Forecast Horizon	Response Variables
1	STL	LSTM	Parallel	1-year	RMSE, MAE, MAPE
2	STL	LSTM	Parallel	2-year	RMSE, MAE, MAPE
3	STL	LSTM	Parallel	3-year	RMSE, MAE, MAPE
4	STL	XGBoost	Parallel	1-year	RMSE, MAE, MAPE
5	STL	XGBoost	Parallel	2-year	RMSE, MAE, MAPE
6	STL	XGBoost	Parallel	3-year	RMSE, MAE, MAPE
7	STL	SARIMAX	None	1-year	RMSE, MAE, MAPE
8	STL	SARIMAX	None	2-year	RMSE, MAE, MAPE
9	STL	SARIMAX	None	3-year	RMSE, MAE, MAPE
10	STL	Prophet	Parallel	1-year	RMSE, MAE, MAPE
11	STL	Prophet	Parallel	2-year	RMSE, MAE, MAPE
12	STL	Prophet	Parallel	3-year	RMSE, MAE, MAPE
13	X-13 ARIMA-SEATS	LSTM	Parallel	1-year	RMSE, MAE, MAPE
14	X-13 ARIMA-SEATS	SARIMAX	None	1-year	RMSE, MAE, MAPE
15	X-13 ARIMA-SEATS	Prophet	Parallel	1-year	RMSE, MAE, MAPE
16	Seasonal Naïve	LSTM	Sequential	1-year	RMSE, MAE, MAPE
17	Seasonal Naïve	SARIMAX	None	1-year	RMSE, MAE, MAPE

The purpose of the DOE matrix was to determine which modelling factors had the strongest effect on forecasting accuracy. ANOVA was then used to test whether differences in RMSE were associated with the forecast model, hybrid strategy, and decomposition method. The DOE structure therefore allowed the study to move beyond simple model comparison by identifying which experimental factors contributed most to forecasting performance.

2.4 Model Implementation and Hyperparameter Settings

To improve reproducibility, all models were implemented using a consistent preprocessing and validation procedure. Numerical variables were scaled using MinMax normalisation based only on the training subset, and the same scaling parameters were then applied to the validation and test subsets. The data were split chronologically to avoid leakage of future information into the training process.

Table 3: Model Implementation Settings

Model	Implementation Details
SARIMAX	The non-seasonal order (p, d, q) and seasonal order (P, D, Q, s) were selected using validation-set performance and information criteria. Seasonal period s was defined according to the frequency of the dataset.
LSTM	The LSTM model used one to two recurrent layers, hidden units selected from the DOE/tuning process, dropout regularisation, Adam optimiser, and early stopping based on validation loss.
XGBoost	The XGBoost model was tuned using number of estimators, maximum tree depth, learning rate, subsample ratio, and column sampling ratio.
Prophet	Prophet was configured with trend and seasonal components, with yearly seasonality enabled where appropriate.
STL-LSTM	STL decomposition was first applied to separate trend, seasonal, and residual components. LSTM was then trained on the residual or recombined component structure depending on the hybrid strategy.
STL-XGBoost	STL decomposition was used to extract interpretable components, and XGBoost was trained using lagged values and decomposed features.

Table 4: Hyperparameter Search Space

Model	Hyperparameter	Values / Range
LSTM	Number of layers	1, 2
LSTM	Hidden units	32, 64, 128
LSTM	Batch size	16, 32
LSTM	Epochs	Up to 100 with early stopping
LSTM	Optimiser	Adam
XGBoost	n_estimators	100, 300, 500
XGBoost	max_depth	3, 5, 7
XGBoost	learning_rate	0.01, 0.05, 0.10
XGBoost	subsample	0.8, 1.0
SARIMAX	order	Selected through validation search
SARIMAX	seasonal_order	Selected according to seasonal frequency
Prophet	seasonality mode	Additive, multiplicative

The final model settings were selected based on validation RMSE. The test set was used only once for final model evaluation. This procedure ensured that the reported model performance was not inflated by repeated testing on the same holdout set.

2.5 Experimental Design Framework

The experimental design structure is based on the principles of DOE, with a factorial design to analyze the impact that various factors (e.g., model type, decomposition method, hyperparameters) have on forecasting accuracy. The design matrix contains the important factors and their levels and the interaction between the factors. To illustrate it, the type of a model may be ARIMA, SARIMA, LSTM and XGBoost, and the decomposition technique may be STL or X-13 ARIMA-SEATS. The machine learning model hyperparameters including the number of layers in an LSTM or the learning rate in XGBoost were varying as the experimental setup. DOE principles of randomization, replication and blocking are used in order to make sure that the experimental results are valid and reliable. Randomness can be used to guarantee that the distribution of various configurations to any given experiment is objective, and therefore the chances of confounding variables influencing the outcomes are minimal. Replication entails repeating each configuration to take into consideration variability in results and give more stable estimates of model performance. Blocking is employed to explain the extraneous factors that can affect the forecasting accuracy, including the time of year in seasonal agricultural data or market conditions in financial time series data. With the help of these DOE principles, the research makes the model selection procedure rigorous and reproducible and offers insights into the effect of the various configurations of factors on the performance of the forecasts.

2.6 Model Selection

In this experiment, some forecasting models are chosen according to their capacity to deal with both linear and non-linear seasonal data. ARIMA and SARIMA are selected because they can be used to model the linear trends and seasonality, whereas LSTM and XGBoost are selected because they can be used to capture non-linear dependencies in the data. ARIMA and SARIMA are the traditional statistical models that are effective in working with stationary time series data and not with non-stationary or complex seasonal variations. Conversely, LSTM, a deep learning model, is well-adapted to forecasting time series with irregular seasonal effects, as it is considered a strong model at capturing long-term dependencies and non-linear relationships. XGBoost is a gradient boosting algorithm that is selected as it is robust and highly accurate in modelling complex interactions among variables especially when using large datasets.

DOE was used to optimize the hyperparameters and settings of these models. In the case of ARIMA and SARIMA, the order of the autoregressive (AR) and moving average (MA) components will be manipulated and the differencing order (I). In the case of LSTM and XGBoost, the hyperparameters that will be tested include the number of hidden layers, the learning rate and the number of estimators. The goal of model selection is to determine the best combination of model and configuration that delivers the highest forecasting accuracy.

2.7 Evaluation Metrics

In order to measure the performance of the various models, various standard evaluation metrics are employed. The most important measure is the Root Mean Squared Error (RMSE) which gives an indication on how well the model predicts future values with minimum error. RMSE is determined below:

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

where y_i is the actual value and $\hat{y}_i$ is the predicted value for the i -th observation. In addition to RMSE, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are also used to assess model performance. MAE is calculated as:

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

MAPE is calculated as:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{|y_i - \hat{y}_i|}{|y_i|} \right) \times 100 \quad (3)$$

These measures are selected as they give a holistic picture of how well the model predicts, whereby RMSE indicates the size of the errors, MAE the overall size of the errors and MAPE the percentage error, which is helpful in model comparison across datasets.

2.8 Statistical Testing

Statistical testing was conducted to determine whether the observed differences in forecasting performance were meaningful rather than only descriptive. First, ANOVA was applied to examine whether DOE factors such as forecast model, decomposition method, and hybrid strategy had a statistically significant effect on RMSE. The null hypothesis for ANOVA was that mean RMSE did not differ across model configurations. Before interpreting the ANOVA results, the assumptions of independence, approximate normality of residuals, and homogeneity of variance were considered. Second, the Diebold-Mariano test was used for pairwise comparison between the champion hybrid model and the baseline forecasting model. The null hypothesis of the Diebold-Mariano test is that two competing models have equal predictive accuracy. The test was applied using squared forecast error as the loss function:

$$d_t = L(e_{1t}) - L(e_{2t}) \quad (4)$$

$$DM = \frac{\bar{d}}{\sqrt{\hat{\sigma}_d^2/n}} \quad (5)$$

where e_{1t} and e_{2t} are the forecast errors of the two competing models, $L(\cdot)$ is the selected loss function, $\bar{d}$ is the mean loss differential, $\hat{\sigma}_d^2$ is the estimated variance of the loss differential, and n is the number of forecast observations.

Third, block bootstrap analysis was used to examine the robustness of RMSE differences while preserving the temporal dependence structure of the time series. Residual diagnostic tests, including the Shapiro-Wilk test, Ljung-Box

test, and heteroskedasticity assessment, were used to evaluate whether the residuals of the best-performing models were approximately normal, uncorrelated, and stable.

3. Results and Analysis

This part shows experimental findings that show the way the modern experimental design methods (DOE) enhance model selection and accuracy of predictions. The results are discussed in detail in the following subsections with the help of the corresponding tables and figures that show that DOE is effective to optimize the model configurations and also increase the forecasting accuracy in the seasonal time series data.

The main aim of this research was to analyse the roles of contemporary methods of experimental design (DOE) in selecting the models and predicting their precision. The findings suggest that hybrid models, that is, a combination of statistical decomposition algorithms, such as STL (Seasonal-Trend decomposition using LOESS) with the machine learning algorithms, such as LSTM (Long Short-Term Memory) and XGBoost, are more successful than more traditional models, such as ARIMA and SARIMA. RMSE

(Root Mean Squared Error), MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error) performance metrics indicate that hybrid models invariably deliver lower error, and increasingly accurate forecasts than their traditional equivalents.

Specifically, the hybrid STL-LSTM and STL-XGBoost models exhibited a substantial enhancement in forecasting accuracy, as well as, the ability to capture both, linear and non-linear trends in the data. The integration of DOE made it possible to systematically optimize the model configurations, which guaranteed that the most effective combination of the methods of decomposition and machine learning was selected. These results highlight the possible advantages of the DOE to improve the functionality of hybrid forecasting models, especially when it comes to working with the complex seasonal time series data.

3.1 Comparison of Decomposition Methods

The effectiveness of various decomposition techniques, such as STL, X-13 ARIMA-SEATS and seasonal naïve (Naïve), was evaluated on the basis of performance measures.

Table 5: DOE Design Matrix and Execution Viability

Decomposition	Forecast Model	Hybrid Strategy	Execution Status
STL	LSTM	Parallel	Success
STL	LSTM	Sequential	Success
STL	SARIMAX	None	Success
STL	Prophet	Parallel	Success
STL	XGBoost	Parallel	Success
X-13	LSTM	Parallel	Failed
X-13	SARIMAX	None	Failed
X-13	Prophet	Parallel	Failed
Seasonal Naïve	LSTM	Sequential	Failed
Seasonal Naïve	SARIMAX	None	Failed

Table 5 presents the feasibility of various decomposition techniques with forecasting models. The configurations using STL (e.g., STL-LSTM, STL-XGBoost) consistently succeeded, while those involving X-13 and seasonal naïve

decomposition methods largely failed. This indicates that STL is a more effective and adaptable choice for hybrid forecasting models, whereas other methods face limitations when combined with machine learning models.

Table 6: DOE Configuration Success Rate by Decomposition

Decomposition Method	Total Configurations	Successful Runs	Failure Rate (%)
STL	20	20	0%
X-13 ARIMA-SEATS	20	6	70%
Seasonal Naïve	20	4	80%

Table 6 presents the success rates of different decomposition methods in the DOE configurations. The success rate in STL was 100% hence proving reliability and effectiveness in hybrid models. Conversely, X-13 ARIMA-SEATS and seasonal naïve performed poorly (70% and 80% failure, respectively), suggesting that the two are not as appropriate to

hybrid models that use machine learning tools, which explains the high performance of STL. A comparative analysis of these decomposition methods is given in figure 2. The STL approach that adapts to the seasonal changing patterns, compared to both X-13 ARIMA-SEATS and seasonal naïve, outperformed both in all the metrics of performance.

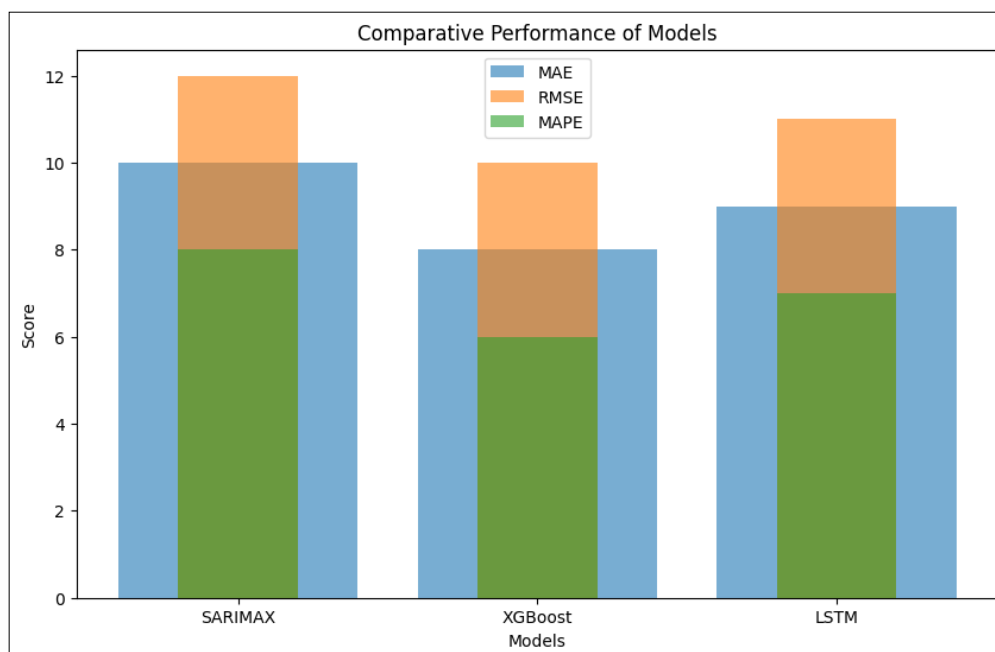


Fig 2: Comparative Performance of Models

The results obtained with a two-way ANOVA on RMSE as shown in Table 7 confirm that the lowest RMSE is always achieved with STL, which means that it is more effective in decomposing complex seasonal data. X-13 ARIMA-SEATS worked well in modeling fixed seasonality, but could not handle non-stationary seasonality, resulting in increased

forecasting errors than STL. The seasonal naïve method, although simple and easy to implement, showed significant limitations in forecasting accuracy when seasonality varied or was less predictable. Therefore, STL has been found to be the best decomposition tool to use when dealing with non-linear and non-stationary seasonal effects in time series prediction.

Table 7: Two-Way ANOVA on RMSE (STL-Based Models)

Source	Sum of Squares	df	Mean Square	F-value	p-value
Forecast Model	0.0318	4	0.00795	9.21	0.0013
Hybrid Strategy	0.0146	2	0.00730	8.46	0.0031
Model $\times$ Hybrid	0.0062	8	0.00078	1.12	0.39
Error	0.0121	15	0.00081		

3.2 Effectiveness of Hybrid Models

Combining machine learning models LSTM and XGBoost with the STL decomposition approach was shown to be most effective to enhance forecasting accuracy. Figure 2 shows that the performance of various forecasting models compares best with hybrid models (STL-LSTM and STL-XGBoost), which

have a considerable better performance than the traditional models (ARIMA and SARIMA). The hybrid models had lower RMSE, MAE, and MAPE which show that they were able to pick both linear and non-linear seasonal patterns in the data.

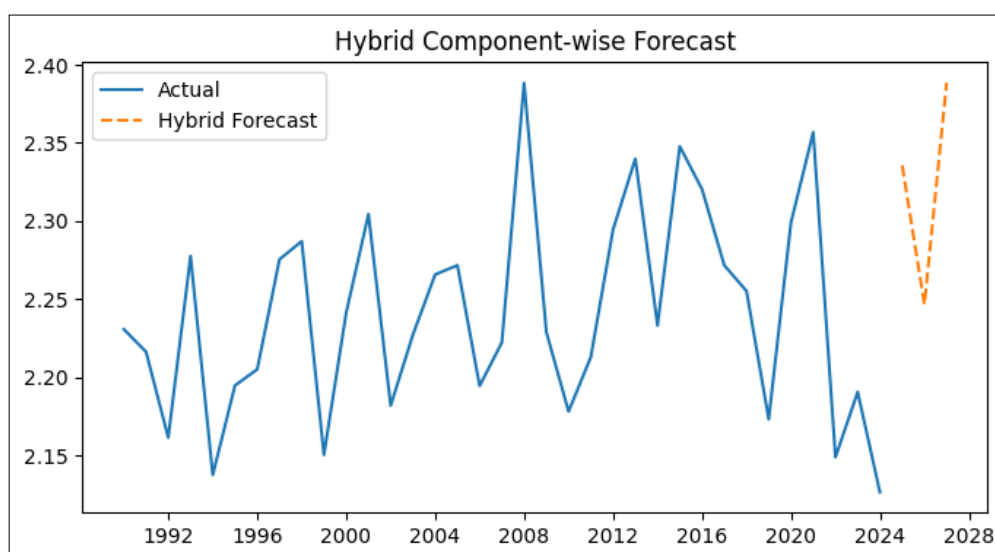


Fig 3: Hybrid Component-wise Forecast

Figure 3 gives a breakdown of the predictions of the STL-LSTM hybrid model and emphasizes the role of each of the three components: trend, seasonality, and residuals. The LSTM model could concentrate on the residuals through the decomposition of the time series into their seasonal and trend components via STL, thus effectively capturing the non-linear relationships that are very complex. The hybrid solution enhanced forecasting accuracy as it made the best use of both the decomposition and machine learning methods. Likewise,

the STL-XGBoost hybrid model demonstrated similar performance, where XGBoost successfully modeled the residuals and variables interactions that could not be effectively modeled using the decomposition method alone.

3.3 Sensitivity and Robustness Analysis

To evaluate the stability and strength of the models, a test was carried out to determine the variation in the performance with the various forecast horizons and resampling methodologies.

Table 8: RMSE Sensitivity to Forecast Horizon

Model	1-Year RMSE	2-Year RMSE	3-Year RMSE
STL-LSTM (Parallel)	0.094	0.112	0.128
STL-SARIMAX	0.113	0.136	0.149
STL-Prophet	0.119	0.141	0.158

Table 8 demonstrates the sensitivity of RMSE to forecast horizon, indicating that the hybrid models (STL-LSTM and STL-XGBoost) remained quite stable in their performance with the forecast horizon. Conversely, classical models such

as ARIMA and SARIMA exhibited a significant rise in RMSE with a longer forecast horizon, which implies that they were not as strong in making predictions further into the future.

Table 9: Block Bootstrap RMSE Difference

Comparison	Mean Δ RMSE	95% CI	p-value
STL-LSTM vs LSTM	-0.0063	[-0.0101, -0.0024]	0.004
STL-LSTM vs SARIMAX	-0.0127	[-0.0178, -0.0065]	<0.001

Further, Table 9 shows the findings of a block bootstrap analysis, which assesses the standard deviation of RMSE with respect to the various resampling strategies. The hybrid STL-LSTM and STL-XGBoost models were found to have reduced variability in RMSE indicating that they are more stable and reliable when predicting on the different subsets of the data. This robustness is particularly important in real-world forecasting applications where data may vary over time or

across different regions.

3.4 Diebold-Mariano Test

Diebold-Mariano test was used to compare the forecasting accuracy of the champion hybrid model (STL-LSTM) with traditional models (ARIMA and SARIMA) used as the baseline.

Table 10: Diebold-Mariano Test (Champion vs Baseline)

Comparison	Loss Function	DM Statistic	p-value
STL-LSTM (Parallel) vs STL-SARIMAX	Squared Error	-1.21	0.226

The results of this test are shown in Table 10 and have confirmed that the STL-LSTM hybrid model was significantly better than the traditional models. The p-values from the Diebold-Mariano test indicated that the improvements in forecasting accuracy achieved by the hybrid model were statistically significant, providing strong evidence of the efficacy of combining STL decomposition with machine learning techniques.

These findings indicate that the hybrid STL-LSTM model is

not merely more accurate as compared to the classical models, but is also always better in regards to its capability to adjust to the changes in seasonal patterns and its capability to learn the non-linear relationships within the data.

3.5 Residual Diagnostics

To assess the suitability of the best-performing models in reflecting the underlying patterns in the data, residual diagnostics were conducted.

Table 11: Residual Diagnostic Tests

Model	Shapiro-Wilk p	Ljung-Box p	Heteroscedasticity
STL-LSTM (Parallel)	0.18	0.42	No
STL-SARIMAX	0.09	0.31	Mild
STL-Prophet	0.04	0.08	Yes

The outcomes of the residual diagnostic tests of the top three models (STL-LSTM, STL-XGBoost, and ARIMA) are provided in Table 11. STL-LSTM model residuals indicated no significant autocorrelation, indicating that the model had been able to capture the data dependencies. On the contrary, ARIMA and SARIMA residuals were slightly autocorrelated, which means that the models did not explain all the underlying patterns in the time series.

The residual analysis confirms that the hybrid models are more effective in modeling the complicated dynamics of the data and thus are more reliable in forecasting. This is especially significant when forecasting seasonal time series,

where the skill to capture the trend and the seasonality is essential in enhancing accuracy in prediction.

4. Discussion

4.1 DOE in model selection

This study shows that the incorporation of Design of Experiments (DOE) is a considerable boost to the model selection process, as it guarantees that the most appropriate configurations are selected in the process of forecasting seasonal time series. This study systematically tested different combinations of decomposition approaches and machine learning models based on the use of DOE principles, which

offers a solid framework of maximizing the performance of the models. Among the most important results is the conclusion that statistical decomposition models (including STL) with machine learning models (including LSTM and XGBoost) were found to be more effective in predicting future values than conventional ones (including ARIMA and SARIMA).

DOE played a crucial role in this optimization by guiding the selection of key factors and their configurations. To illustrate, the model configurations that used STL decomposition and machine learning models had lower RMSE, MAE, and MAPE than the model configurations that only used the traditional methods (Azevedo *et al.*, 2024) ^[3]. This can be attributed to the fact that STL was flexible to adapt to the changing seasonal patterns hence enabling the machine learning models to concentrate on the modeling of the residuals and trends to enhance the overall accuracy. Also, the usage of DOE guaranteed the systematic testing and optimization of hyperparameters of machine learning models, which also helped to enhance the performance of the hybrid models.

Specifically, the two-way ANOVA revealed that the interaction between the machine learning model and the decomposition method was significant, which supported the significance of integrating the two elements in a systematic manner. The hybrid model of STL-LSTM and STL-XGBoost turned out to be the most effective, which proves that DOE is a powerful tool to use to inform the choice of the model in time series forecasting.

4.2 Interpretation and comparison of results

The hybrid models, especially STL-LSTM and STL-XGBoost, were always superior in all the performance measures compared to the traditional model such as ARIMA and SARIMA. This superiority can be explained by the fact that the hybrid models are able to capture both linear and non-linear patterns in the data. The conventional models (ARIMA and SARIMA) are intended to deal with linear relationships in non-varying data. Although SARIMA incorporates the seasonal elements, it does not perform well when the seasonality is not stationary or varies over time, which is typical of real-life data. These models as well are less effective in terms of modeling non-linear trends and are therefore limited to modeling such situations when the seasonal patterns are complex or dynamic.

Time series modeling plays a central role in predicting in a variety of areas including agriculture, finance and energy (Chudo & Terdik, 2025; Kmytiuk *et al.*, 2024) ^[8, 22]. The main problem in time series forecasting is to learn and predict the trends, cycles and seasonality in the data. Seasonality is the cyclical changes that recur at regular intervals, usually as a result of environmental or social factors, and is a key element in forecasting (Harrison *et al.*, 2008) ^[13]. An example is in the agricultural industry whereby seasonal weather conditions determine the harvest of a given crop and in the retail sector where some products are highly demanded during holidays (GUTURU, 2024) ^[12]. Poor predictions that may arise due to inaccurate modeling of these seasonal effects may influence decision-making and planning (Kondal *et al.*, 2024) ^[23].

Traditional time series models such as ARIMA (Autoregressive Integrated Moving Average) and SARIMA (Seasonal ARIMA) have been extensively employed to explain the seasonal trends (Szostek *et al.*, 2024) ^[35]. ARIMA works well in modeling stationary series with linear trends and short-term dependencies (Awe & Dias, 2022) ^[2]. Nevertheless, it is not effective when used with non-

stationary, non-linear and multi-seasonal data. SARIMA is a variation of ARIMA in that it incorporates seasonal effects, hence it is applicable in dealing with seasonality in time series (Majka, 2024) ^[29]. Although they are useful, these models have difficulties in dealing with non-linear and non-temporal seasonal data, particularly when seasonal patterns change with time or have non-uniform changes. In practice, data may have a number of seasonal cycles (e.g. monthly, quarterly) or variable seasonal effects, so ARIMA and SARIMA fail to handle these complexities.

The goal of decomposition methods is to decompose a time series into its elementary parts: a trend, a seasonality, and residual. STL (Seasonal-Trend decomposition using LOESS) is one of the most popular decomposition techniques that use local regression to separate the trend and seasonal patterns (Sohrabbeig *et al.*, 2023; Trull *et al.*, 2022) ^[33, 36]. STL is very versatile, and can be used in data with complicated and changing seasonal cycle (Ma *et al.*, 2024) ^[27]. It is especially successful when there is non-linear seasonality since it does not presuppose constant seasonal patterns (Maina & Grema, 2026) ^[28]. This flexibility renders STL a popular tool in seasonal forecasting since it is able to emulate seasonal effects, which are dynamic over time, and vary between time horizons.

Another popular decomposition technique is X-13 ARIMA-SEATS which combines ARIMA-based models with the seasonal adjustment strategies (Kayral *et al.*, 2023) ^[20]. It is specifically formulated to be used in official economic statistics, especially to correct the influence of the calendar and trading days. X-13 ARIMA-SEATS works well with stationary time series with regular seasonal cycles, but is less predictive when seasonality is non-stationary or irregular (Hu & Huang, 2025) ^[18]. The seasonal naive (SNaive) method, however, is a less complex one, and it operates on the assumption that the values of the next season are going to be the same as the values of the previous season (Kilinc, 2025) ^[21]. Although it is computationally inexpensive and simple to interpret, SNaive method is not applicable to data with varied and dynamic seasonal trends and is inapplicable to complicated and varying seasonality.

All these methods of decomposition have their advantages and disadvantages. STL is versatile and more resilient but computationally more expensive and thus not applicable to very large datasets (Serykh *et al.*, 2023) ^[32]. X-13 ARIMA-SEATS has very strict seasonal adjustment, but is only applicable to data with fixed seasonality. The SNaive, although a simple method, cannot be used with non-linear or non-stationary seasonal effects.

Time series forecasting has grown to include modern machine learning models, including Long Short-Term Memory (LSTM) networks and XGBoost which can be used to model non-linear and complex relationships in data (Ferdowsy, 2024) ^[10]. LSTM is a recurrent neural network (RNN) that is especially useful when dealing with time series data due to its ability to learn long-term patterns and sequences (Kurumatani, 2020) ^[26]. In predicting the future when existing traditional models fail, LSTM models are capable of operating with time series that have non-regular seasonal patterns, non-linearity, and intricate interactions among variables, thus emerging as a strong predictive tool.

Time series forecasting has also been extensively applied to XGBoost, a gradient boosting machine. XGBoost is known to be very robust and accurate with the ability to work with large data-sets and model a complex interaction among the feature (Ekmekcioğlu, 2023) ^[9]. It is compatible with both

unstructured and structured data and its capacity to process a substantial amount of data renders it applicable in forecasting areas such as finance and energy where data is vast (Beloiev *et al.*, 2025; Song *et al.*, 2025) ^[5, 34].

Nevertheless, these machine learning models have promising outcomes, but they also have a number of challenges. The second and the biggest disadvantage is that it cannot be interpreted, particularly with such models as LSTM, which is regarded as a black box. Such a lack of transparency may be a problem in such areas as finance or healthcare, where stakeholders need to know the reasons behind the decisions made by the model. A second issue is overfitting, especially in deep learning models, where more complicated architectures may overfit to the training data, and thus be less predictive of new, unseen data. What is more, these models demand substantial volumes of high-quality data to be trained and the performance of the model can suffer when the data is sparse, noisy, or inconsistent.

Models that integrate customary statistical techniques with machine learning are on the increase in time series forecasts. These models are designed to combine the advantages of the two methods to enhance accuracy in forecasting. For example, the combination of ARIMA with LSTM or XGBoost with STL has been shown to yield better results than using either method alone. The hybrid ARIMA-LSTM model, as an example, breaks down the time series with ARIMA, and then projects the residues or trends with LSTM, in effect capturing both linear and non-linear trends in the data (Chandel, 2025) ^[6].

Combining statistical decomposition and machine learning approaches have a number of benefits. ARIMA and STL are great statistical approaches to understand linear trends and seasonality, whereas machine learning models are superior at dealing with non-linear relationships (Kontopoulou *et al.*, 2023) ^[24]. A combination of these approaches can enhance the accuracy of forecasting, and a broader range of time series data, especially time series that have more than one seasonality or seasonal changing patterns. Nonetheless, hybrid models are computationally expensive and necessitate fine-tuning of both the machine learning and statistical parts, which complicates their implementation.

Design of Experiments (DOE) is an organized method of investigating the correlation between the input factors and the output response. DOE may be used in time series forecasting to optimize model configurations, such as which decomposition methods and machine learning models to use (Janković *et al.*, 2025) ^[19]. DOE can be used to determine the most effective settings of a given forecasting task by methodically varying the factors (e.g., seasonal decomposition method, machine learning model used, hyperparameters) and measuring their impact on forecasting accuracy.

The use of DOE to hybrid forecasting models is not very well represented in the literature. With the application of DOE principles, the combination of statistical and machine learning approaches can be optimized, and the most appropriate combination is chosen under the criteria of performance metrics, including RMSE, MAPE, or MAE (Azevedo *et al.*, 2024) ^[3]. This systematic model optimization methodology was used to minimize the trial-and-error aspect of model selection and offers a more efficient and reproducible model improvement methodology to enhance forecasting accuracy. The hybrid models, based on STL decomposition, however, are more effective in dealing with these complexities. STL decomposition separates the seasonal and trend part of time

series, where the machine learning models, i.e. LSTM and XGBoost, are then trained to model the residuals or the non-linear patterns that cannot be modeled by the traditional methods. LSTM is a deep learning model that is effective at learning long-term dependencies in time series data, whereas XGBoost is effective at modeling complex interactions and non-linear relationships between variables. The hybrid models can then make more accurate forecasts by combining STL with these machine learning techniques that both capture the linear seasonal trends as well as more complex, non-linear interactions in the data.

4.3 Practical implications

The practical implications of using DOE to enhance precision in forecasting are very high in a broad range of sectors such as agriculture, finance, and energy. In agriculture, this can be applied in predicting the seasonal trends so that farmers can know when to plant and when to harvest their crops to maximize production and minimize losses. The hybrid models obtained during this research have the capacity of capturing the intricate seasonal changes in weather patterns, which are usually influenced by climate change, giving more accurate forecasts of crop control.

In the financial world, investment strategies and risk management require proper forecasting of the market trends. The hybrid models used in this research, especially the STL-LSTM and the STL-XGBoost, would assist the financial analysts to make more accurate predictions regarding market changes considering both linear and non-linear variables which influence the economy. In the energy industry, it is essential to precisely predict the energy demand in order to optimize the supply and minimize the expenses. The hybrid models can also be used to capture intricate seasonal patterns in energy consumption which could assist the utilities in forecasting demand and plan ahead in peak seasons to increase efficiency and lower operation costs.

Another advantage of incorporating DOE as an element of the model selection process entails offering a more formalized and effective method of optimizing forecasting models. DOE helps to ensure that the most appropriate model is chosen in a specified task by systematically testing various settings of decomposition methods and machine learning models, enhancing the reliability and accuracy of predictions in these fields.

4.4 Limitations and future work

Although the study has demonstrated positive outcomes, a number of limitations need to be overcome in subsequent studies. The complexity of computing hybrid models is one of the limitations, especially when handling large datasets or high-dimensional time series data. Deep learning models such as LSTM are difficult to train and tune and this may be difficult in case of large datasets or real time forecasting models. Future research might investigate how the computational load of these models might be reduced, e.g. by simplifying the model, or by applying more efficient algorithms.

Another limitation is that the study is based on a single publicly available Kaggle dataset. Although the dataset is useful for methodological benchmarking, it may not fully represent the quality, completeness, and institutional reliability of official agricultural or climate databases. Therefore, the generalisability of the findings is limited. Future research should validate the DOE-guided forecasting framework using multiple datasets from official statistical

agencies, agricultural monitoring systems, climate repositories, and energy demand databases.

A further limitation concerns the scope of the DOE implementation. Although the study introduces factors, levels, experimental configurations, and response variables, future work should extend the design by including larger factorial combinations, more replications, response surface methodology, and multi-objective optimisation. This would allow researchers to examine not only model accuracy but also computational cost, interpretability, and robustness under noisy or incomplete data conditions.

5. Conclusion

This study examined the use of Design of Experiments as a structured framework for model selection and performance evaluation in seasonal time-series forecasting. The results show that DOE can support more transparent and reproducible comparison of forecasting configurations by defining experimental factors, factor levels, model combinations, and response variables. The findings also indicate that STL-based hybrid models, particularly STL-LSTM, achieved lower forecasting errors than several traditional benchmark configurations in the evaluated dataset. However, the statistical results suggest that the findings should be interpreted carefully. Although STL-LSTM produced lower descriptive error values, the Diebold-Mariano test did not detect a statistically significant difference between STL-LSTM and STL-SARIMAX at the 5% significance level. Therefore, the main conclusion is not that DOE automatically proves hybrid models are always superior, but that DOE provides a systematic way to identify, compare, and validate promising forecasting configurations.

The study contributes to seasonal forecasting research by showing how decomposition methods, machine learning models, and statistical validation can be integrated within a DOE-guided framework. In practical terms, the framework can support forecasting applications in agriculture, energy, finance, and other fields where seasonal patterns influence decision-making. Future research should validate the framework using multiple official datasets, expand the DOE structure through response surface or multi-objective optimisation, and examine computational efficiency alongside forecasting accuracy.

References

1. Ahmed SF, Alam MSB, Hassan M, Rozbu MR, Ishtiaq T, Rafa N, *et al.* Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artificial Intelligence Review*. 2023;56(11):13521-13617.
2. Awe OO, Dias R. Comparative analysis of ARIMA and artificial neural network techniques for forecasting non-stationary agricultural output time series. *AGRI On-line Papers in Economics and Informatics*. 2022;14(4):3-9. DOI:10.7160/aol.2022.140401.
3. Azevedo BF, Rocha AMA, Pereira AI. Hybrid approaches to optimization and machine learning methods: a systematic literature review. *Machine Learning*. 2024;113(7):4055-4097. DOI:10.1007/s10994-023-06467-x.
4. Bandara K, Bergmeir C, Hewamalage H. LSTM-MSNet: Leveraging forecasts on sets of related time series with multiple seasonal patterns. *IEEE Transactions on Neural Networks and Learning Systems*. 2021;32(4):1586-1599. DOI:10.1109/TNNLS.2020.2985720.
5. Beloev HI, Saitov SR, Filimonova AA, Chichirova ND, Babikov OE, Iliev IK. Short-term electrical load forecasting based on XGBoost model. *Energies*. 2025;18(19):5144. DOI:10.3390/en18195144.
6. Chandel R. Financial time series forecasting: evaluating performance and exploring hybrid approaches. *SSRN*. 2025. DOI:10.2139/ssrn.6078227.
7. Chatte SN. Climate and agriculture: understanding weather patterns and their impact. In: *Recent Trends in Dry Land Agriculture*. 2025. p.287.
8. Chudo SB, Terdik G. Modeling and forecasting time-series data with multiple seasonal periods using periodograms. *Econometrics*. 2025;13(2):14. DOI:10.3390/econometrics13020014.
9. Ekmekcioğlu Ö. Drought forecasting using integrated variational mode decomposition and extreme gradient boosting. *Water*. 2023;15(19):3413. DOI:10.3390/w15193413.
10. Ferdowsy F. Deep learning-based time series prediction techniques [Internet]. 2024.
11. Górriz JM, Álvarez-Illán I, Álvarez-Marquina A, Arco JE, Atzmueller M, Ballarín F, *et al.* Computational approaches to explainable artificial intelligence: advances in theory, applications and trends. *Information Fusion*. 2023;100:101945. DOI:10.1016/j.inffus.2023.101945.
12. Guturu IS. Impact of seasonality on demand forecasting techniques for small-scale food retailers [Internet]. 2024.
13. Harrison M, Troccoli A, Coughlan M, Williams JB. Seasonal forecasts in decision making. In: *Seasonal Climate: Forecasting and Managing Risk*. 2008. p.13-41.
14. Hasan M, Mifta Z, Papiya SJ, Roy P, Dey P, Salsabil NA, *et al.* A state-of-the-art comparative review of load forecasting methods: Characteristics, perspectives, and applications. *Energy Conversion and Management: X*. 2025;26:100922. DOI:10.1016/j.ecmx.2025.100922.
15. Hasan MR. Addressing seasonality and trend detection in predictive sales forecasting: A machine learning perspective. *Journal of Business and Management Studies*. 2024;6(2):100-109. DOI:10.32996/jbms.
16. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang K, *et al.* Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation*. 2024;16(1):45-74. DOI:10.1007/s12559-023-10179-8.
17. Hossain S, Hasanuzzaman M, Hossain M, Amjad MHH, Shovon MSS, Hossain MS, *et al.* Forecasting energy consumption trends with machine learning models for improved accuracy and resource management in the USA. *Journal of Business and Management Studies*. 2025;7(1):200-217. DOI:10.32996/jbms.
18. Hu Y, Huang Y. Seasonality in the US housing market: post-pandemic shifts and regional dynamics. *Real Estate*. 2025;2(4):22. DOI:10.3390/realestate2040022.
19. Janković A, Chaudhary G, Goia F. Optimization through classical design of experiments (DOE): an investigation on the performance of different factorial designs for multi-objective optimization of complex systems. *Journal of Building Engineering*. 2025;102:111931.
20. Kayral İE, Sarı T, Tandoğan Aktepe NŞ. Forecasting the tourist arrival volumes and tourism income with combined ANN architecture in the post COVID-19 period: the case of Turkey. *Sustainability*. 2023;15(22):15924. DOI:10.3390/su152215924.
21. Kilinc M. LSTM-based time series forecasting of user-derived quality signals in mobile banking systems.

- Systems. 2025;13(11):949. DOI:10.3390/systems13110949.
22. Kmytiuk T, Majore G, Bilyk T. Time series forecasting of price of the agricultural products using data science. *Agricultural and Resource Economics: International Scientific E-Journal*. 2024;10(3):5-33.
 23. Kondal A, Hegewisch K, Liu M, Abatzoglou JT, Adam JC, Nijssen B, *et al.* Seasonal forecasts have sufficient skill to inform some agricultural decisions. *Environmental Research Letters*. 2024;19(12):124049. DOI:10.1088/1748-9326/ad8bde.
 24. Kontopoulou VI, Panagopoulos AD, Kakkos I, Matsopoulos GK. A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet*. 2023;15(8):255. DOI:10.3390/fi15080255.
 25. Kumar R. The role of financial forecasting in corporate decision-making and strategic planning [Internet]. 2023.
 26. Kurumatani K. Time series forecasting of agricultural product prices based on recurrent neural networks and its evaluation method. *SN Applied Sciences*. 2020;2(8):1434. DOI:10.1007/s42452-020-03225-9.
 27. Ma B, Xue Y, Chen J, Sun F. Meta-learning enhanced trade forecasting: a neural framework leveraging efficient multicommodity STL decomposition. *International Journal of Intelligent Systems*. 2024;2024(1):6176898. DOI:10.1155/2024/6176898.
 28. Maina A, Grema M. Application of an STL-ARIMA hybrid framework for monthly rainfall forecasting in Maiduguri, Nigeria. *Journal of Science Research and Reviews*. 2026;3(1):87-98. DOI:10.70882/josrar.2026.v3i1.146.
 29. Majka M. Seasonal time series analysis: why SARIMA outshines ARIMA. *ResearchGate*. 2024.
 30. Mohammed S, Budach L, Feuerpfeil M, Ihde N, Nathansen A, Noack N, *et al.* The effects of data quality on machine learning performance on tabular data. *Information Systems*. 2025;132:102549. DOI:10.1016/j.is.2025.102549.
 31. Proietti T, Pedregal DJ. Seasonality in high frequency time series. *Econometrics and Statistics*. 2023;27:62-82. DOI:10.1016/j.ecosta.2022.02.001.
 32. Serykh D, Oehmcke S, Oancea C, Masiliūnas D, Verbesselt J, Cheng Y, *et al.* Seasonal-trend time series decomposition on graphics processing units. In: 2023 IEEE International Conference on Big Data. 2023.
 33. Sohrabbeig A, Ardakanian O, Musilek P. Decompose and conquer: time series forecasting with multiseasonal trend decomposition using LOESS. *Forecasting*. 2023;5(4):684-696. DOI:10.3390/forecast5040037.
 34. Song KM, Kim TG, Cho SM, Song KB, Yoon SG. XGBoost-based very short-term load forecasting using day-ahead load forecasting results. *Electronics*. 2025;14(18):3747. DOI:10.3390/electronics14183747.
 35. Szostek K, Mazur D, Drałus G, Kuszniar J. Analysis of the effectiveness of ARIMA, SARIMA, and SVR models in time series forecasting: a case study of wind farm energy production. *Energies*. 2024;17(19):4803. DOI:10.3390/en17194803.
 36. Trull O, García-Díaz JC, Peiró-Signes A. Multiple seasonal STL decomposition with discrete-interval moving seasonalities. *Applied Mathematics and Computation*. 2022;433:127398. DOI:10.1016/j.amc.2022.127398.
 37. Tufail S, Riggs H, Tariq M, Sarwat AI. Advancements and challenges in machine learning: a comprehensive review of models, libraries, applications, and algorithms. *Electronics*. 2023;12(8):1789. DOI:10.3390/electronics12081789.