MAUN Fen Bil. Dergi., 13, 1, 95-103 Araştırma Makalesi/ Research Article
MAUN J. of Sci., 13, 1, 95-103 DOI : 10.18586/msufbd.1674717

# Statistical and Machine Learning Approaches for Energy Consumption Forecasting Using Time Series Analysis

Ismail Mohamed YOUSSOUF[1], Taha ETEM[2]

[12] Computer Engineering, Engineering Faculty, Çankırı Karatekin University, Çankırı, Türkiye

: tahaetem@karatekin.edu.tr [1] 0009-8659-0101-2202 [2] 0000-0003-1419-5008

Received (Geliş): 12.04.2025 Revision (Düzeltme): 28.04.2025 Accepted (Kabul): 26.05.2025

## ABSTRACT

Energy production has become a rapidly growing field, especially with the impacts of climate change, and has become a competitive factor among countries. However, this production is often not stable or continuous, varying depending on external factors such as weather conditions or, in some cases, fossil fuel production. Therefore, predicting energy production is crucial to optimize and manage its efficiency. This study employed statistical models like ARIMA and SARIMAX, as well as machine learning models such as LSTM and Gaussian Process Regression (GPR), to predict renewable energy production time series. The models were compared based on evaluation metrics, predictions, and forecasts over 72 steps. Among the comparison techniques, the SARIMAX model performed the best, achieving 0.000031 MSE, 0.0026 RMSE, 0.0015 MAE, and 99.98% $R^2$. Additionally, the SARIMAX model provided nearly perfect forecasts by predicting the data as effectively as the other models.

**Keywords**: ARIMA, Energy Production, Gaussian Process, LSTM, Time Series.

# Zaman Serisi Analizi ile Enerji Tüketim Tahmininde İstatistiksel ve Makine Öğrenimi Yaklaşımları

## ÖZ

Enerji üretimi, özellikle iklim değişikliğinin etkileriyle hızla büyüyen bir faaliyet alanı haline gelmiştir ve ülkeler arasında rekabet unsuru oluşturmuştur. Ancak, bu üretim çoğu zaman sabit veya sürekli olmamakta, hava koşulları veya bazı durumlarda fosil yakıt üretimi gibi dış faktörlere bağlı olarak değişiklik göstermektedir. Bu nedenle, enerji üretiminin verimliliğini optimize etmek ve yönetmek amacıyla tahmin edilmesi büyük önem taşımaktadır. Bu çalışmada, yenilenebilir enerji üretiminin zaman serisi tahminleri, ARIMA ve SARIMAX gibi istatistiksel modellerin yanı sıra LSTM ve Gauss Süreç Regresyonu (GPR) gibi makine öğrenimi modelleri kullanılarak gerçekleştirilmiştir. Kullanılan modeller, değerlendirme metriklerine, her modelin yaptığı tahminlere ve 72 adım boyunca yapılan öngörülere göre karşılaştırılmıştır. Uygulanan çeşitli karşılaştırma teknikleri sonucunda, en iyi performansı SARIMAX modeli göstermiş; bu model 0.000031 MSE, 0.0026 RMSE, 0.0015 MAE ve %99,98 $R^2$ değerlerine ulaşmıştır. Ayrıca, SARIMAX modeli verileri diğer modeller kadar etkili şekilde tahmin ederek neredeyse mükemmel öngörüler sağlamaktadır.

**Anahtar Kelimeler**: ARIMA, Enerji Üretimi, Gauss Süreç, LSTM, Zaman Serisi

## INTRODUCTION

With the impacts of climate change caused by the use of nuclear and fossil energy [1], the shift to renewable energy has become more crucial than ever. These types of energy offer numerous advantages: they significantly reduce greenhouse gas emissions, the primary cause of global warming, decrease air pollution, and help preserve natural resources [2]. However, the use of renewable energy poses significant daily challenges [3]. Its production heavily depends on various factors, particularly weather conditions, and in some cases, the use of fossil fuels in hybrid systems combining fossil and renewable energy. In this context, it is essential to forecast renewable energy production in order to optimize its management and improve its storage efficiency [4]. To achieve these aims, time series analysis plays a key role. Statistical methods such as ARIMA and SARIMA, as well as machine learning approaches like LSTM and GPR (Gaussian Process Regression), are widely used to address these challenges [5]. The prediction of production is a promising field, especially with the development of technology [6]. The works proposed by Edmond Connolly [7] and Fahad Radhi Alharbi et al. [8] present time series models aimed at predicting energy-related data. Fahad Radhi Alharbi, in his study, used historical electricity data from Saudi Arabia spanning the period 1980–2020. To forecast the future performance of the electricity sector in Saudi Arabia over 30 years (2021–2050), Fahad Radhi Alharbi deployed a SARIMAX model and compared it with models such as ARIMA and ANN. In his study, the SARIMAX model achieved the best performance, notably with an RMSE of 1.2 TWh for electricity generation and an $R^2$ of 99% across all analysed data types (generation, consumption, peak load). Edmond Connolly, on the other hand, used electricity consumption data from Ireland combined with meteorological data to predict short-term electricity consumption in Ireland (2 months in 2020). The data were collected between 2014 and 2020 at 15-minute intervals. To achieve his objective, Edmond Connolly deployed two models, LSTM and SARIMAX. The best-performing model was LSTM, surpassing SARIMA. In their work, Sima Siami-Namini et al. [9], J.W. Taylor et al. [10], and Meftah Elsaraiti et.al [11] proposed time series models, even though their prediction domains differ. J.W. Taylor and Meftah Elsaraiti focused on wind data, whereas Sima Siami-Namini's study focused on financial and economic data. Sima Siami-Namini used several time series datasets with varying sizes in her study (e.g., 403 observations for the Nikkei 225 index, 558 for NASDAQ). They deployed ARIMA and LSTM models, comparing them based on the reduction in mean error. The most effective model was LSTM, with a mean error reduction between 84% and 87% compared to ARIMA. In J.W. Taylor's study, daily wind speed data and meteorological ensemble predictions were used, collected from January 1, 1995, to June 30, 2004, at wind farm sites in the UK. The evaluation data spanned 18 months (late 2002 to mid-2004). Taylor applied ARMA-GARCH, ARFIMA-GARCH, and ensemble prediction models. The best- performing model was the calibrated and smoothed meteorological ensemble predictions. As for Meftah Elsaraiti, he utilized wind speed data for Halifax (Canada) from May 2021 to June 2021. They deployed ARIMA and LSTM models, and the best model in his study was LSTM. Jailani et.al [12] proposes a study dedicated to Investigating the Power of LSTM-Based Models in Solar Energy Forecasting. In 2023, the article compares standalone LSTM models with hybrid models and aims to study LSTM-based models for solar energy forecasting. Several models are deployed, including standalone models (LSTM) and hybrid models (CNN-LSTM). The best model is the hybrid model (CNN-LSTM). F. U. M. Ullah et al. [13], in 2020, propose a study dedicated to Short-Term Prediction of Residential Power Energy Consumption via CNN and Multi-Layer Bi-Directional LSTM Networks, aiming to develop a method combining CNN and Multi-Layer Bi- Directional LSTM (M-BDLSTM) to improve residential power consumption forecasting. The most effective model remains the CNN-M-BDLSTM combination. Mahjoub et al. [14], in 2022, through their work titled Energy Management Strategy Using Deep Learning for Power Consumption Forecasting, aim to develop an energy management strategy using deep learning models. The different models deployed are LSTM, GRU, and Drop-GRU, with the latter being the most effective for energy consumption forecasting. Bilgili et al. [15] propose a study on energy consumption prediction. The study aims to forecast electricity consumption in Turkey using deep learning models, specifically LSTM and ANFIS. The data used, obtained from the Turkish Electricity Transmission Corporation (TETC), represents Turkey's electricity consumption between January 1, 2016, and December 31, 2019. After comparing the two proposed models, LSTM and ANFIS (Adaptive Neuro-Fuzzy Inference System), LSTM demonstrated higher performance in this study. Arslan (2022) [16] proposes a study titled A Hybrid Forecasting Model Using LSTM and Prophet. To improve forecasting accuracy, the study aims to develop a hybrid model that combines LSTM and the Prophet model for energy consumption prediction. This study uses monthly energy consumption data from seven countries (Canada, France, Italy, Japan, Brazil, Mexico, and Turkey) between 2006 and 2017. Several models are deployed, including a hybrid model that combines (STL + BiLSTM + Prophet) and three independent models: BiLSTM (Bidirectional LSTM), deBiLSTM (Deseasonalized BiLSTM), and Prophet. The hybrid model outperforms the other models due to the combined strengths of the three integrated approaches. Gasparin et al. (2021) [17] propose a studied Deep Learning for Time Series Forecasting. This study aims to evaluate and compare several deep learning models for electric load forecasting. To achieve this, the data used includes IHEPC (Electricity consumption in Europe), CER (Commission for Energy Regulation Ireland), and GEFCom2014 (Electricity consumption data collected by ISO New England). The models deployed include FNN (Feedforward Neural Networks), RNN (Recurrent Neural Networks), LSTM, GRU, Seq2Seq, and Temporal Convolutional Networks (TCN). After comparison, for short-term forecasting, LSTM-Rec demonstrated the best performance. All the studies cited previously, even though they successfully achieve their aims, they do have certain drawbacks. These studies, as varied as they may be, only use one or two models to predict their time series. This restricts their ability to make more accurate predictions, leading to a scenario where, even if the models achieve moderate performance, the authors of these studies accept it as sufficient. The objective of our study, however, is to employ multiple time series models to determine the best-performing model. The comparison will be based on the evaluation of models according to performance metrics, on predictions of the models and also on theirs forecasting over 72 steps. In order to guide our investigations and offer a methodological approach to addressing the challenges of energy production, we have reformulated the research questions of this study as follows:

Which time series model provides clear predictions and forecasts for renewable energy prediction?

How does the dataset size impact the performance of certain models, particularly LSTM?

What role can exogenous variables play in improving the performance of the SARIMAX model?

Our article is organized as follows: a brief description of the importance of renewable energy prediction using time series models, followed by a presentation of previous published works. Next, the methodology will be introduced, followed by the combined results and discussion, and finally, the conclusion.

## MATERIAL and METHODS

This section provides information about the data, the models, and the evaluation metrics used. For further clarification, the Figure 1 provides a more transparent explanation. Figure 1 provides an overview of the work that will be carried out throughout our study. The dataset named "World Overview Data" is used for analysis [18]. This dataset will undergo several preprocessing techniques, such as replacing missing values with the mean, normalization, duplicate removal, and data splitting. The pre-processed data will then be used by two types of models: statistical models (ARIMA and SARIMAX) and machine learning models such as LSTM and GRU. These models will be evaluated based on their evaluation metrics, predictions, and forecasting performance.

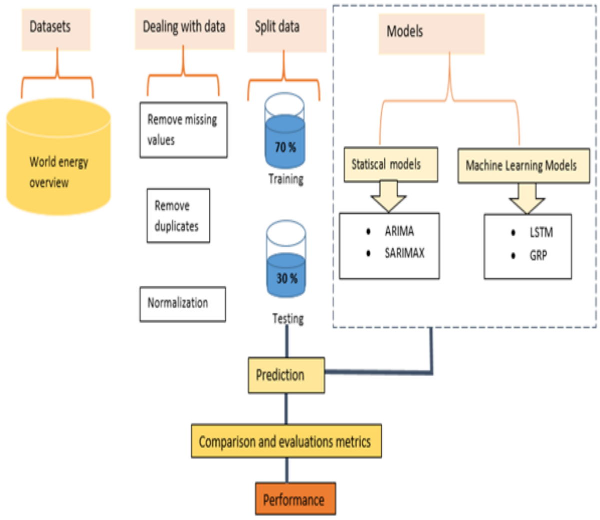


**Figure 1.** Overview of the Methodology

### Dataset

The data used comes from Kaggle [18] and includes 13 features and 600 entries or rows. It represents global energy data from January 1973 to December 2022, collected every three months. Regarding the features, for ARIMA, only one feature will be used as the target variable for prediction, which is the variable named "Total Renewable Energy Production." For the other models, this same variable will be used as the target, but four additional variables will also be utilized: "Primary Energy Imports," "Total Renewable Energy Consumption," "Total Primary Energy Production," and "Total Primary Energy Consumption". The original series has been plotted in Figure 2.

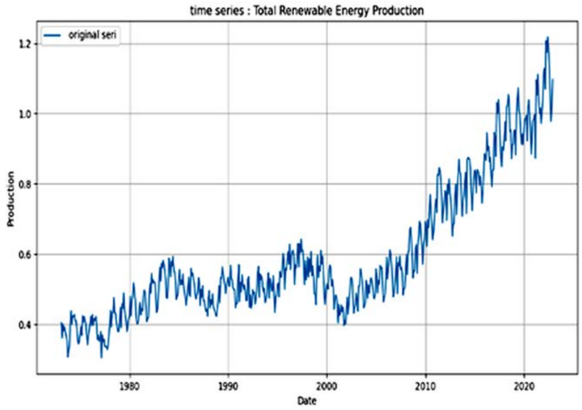


**Figure 2.** Overview of the Original Data

In order for the data to be usable for each model, several cleaning operations were performed. These included the removal of missing values, the removal of duplicates, and normalization, which was also carried out even though it is optional for some models. The data was also split into test and training sets, with 70% of the data used for training and the remaining 30% used for testing.

Our quarterly dataset spans five European markets (Turkey, Germany, Spain, Italy, and France) and covers three renewable electricity sources:

- Solar (photovoltaic generation)
- Wind (onshore and offshore combined)
- Hydro (run-of-river and reservoir)

A total of 600 raw observations (one per quarter per country) formed the basis for all analyses. We supplemented generation data with two macro-energy indicators for each country:

- Energy imports (net electricity imports per quarter)
- Total energy consumption (quarterly national consumption)

Both time series were retrieved from the International Energy Agency (IEA) public database. Less than 2 % of entries were missing, typically at country–quarter boundaries. We applied linear interpolation along the time axis to fill gaps. All six series (three generation + two exogenous) were seasonally differenced (lag = 4) to remove quarterly seasonality. Each feature was standardized to zero mean and unit variance based on the training folds.

### Models

This section gives information about model used in this study. As already explained in the Figure 1, statistical methods and machine learning methods were used in this study. In time series research, the ARIMA model, or Autoregressive Integrated Moving Average, is a widely used statistical model for forecasting chronological data. This model consists of three components: AR, I, and MA. The AR (AutoRegressive) component expresses a linear relationship between an observation and its past values.This component is influenced by the auto-

regression's order, p, or the number of AR terms. In order to make the series stationary, the second component, I (Integrated), removes any trends or non-stationarity. It is related to d, which is the number of times the series has been differenced or the degree of differencing to attain stationarity. The third component, MA (Moving Average), captures the dependence between an observation and the past forecast errors. This component is associated with q, which is the order of the moving average.
The parameters p, d, and q can be determined either through graphically using the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) [19,20]. The Figure 3 below shows the ACF and PACF plots of ARIMA.

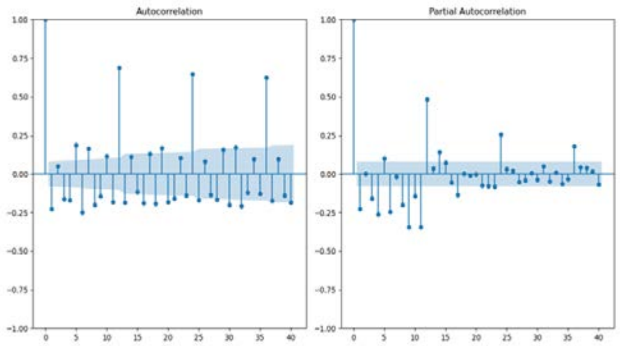


**Figure 3.** Autocorrelation and Partial Autocorrelation of the Data

SARIMAX, or Seasonal AutoRegressive Integrated Moving Average with Exogenous Variables, is an extension of the ARIMA model that incorporates both seasonality and exogenous variables. It is used when the data exhibits seasonal trends and when external or exogenous factors potentially impact the target series. SARIMAX consists of six components: p, d, and q, which are typical ARIMA components, along with additional seasonal components, P, D, and Q. P represents the order of the seasonal AR term. D indicates the degree of seasonal differencing to remove seasonal trends. Q corresponds to the seasonal MA term, whiles denotes the seasonal period. Exogenous variables are external predictors, other than the target variable, that can influence the series. In our study, four exogenous variables are employed. [20,21]. We set the seasonal period s=12 to capture the annual cycle inherent in our monthly renewable-generation series (e.g., higher solar output in summer months and lower in winter).
Gaussian Process Regression (GPR) is a machine learning model. It is a powerful non-parametric regression model based on the Gaussian process. In other words, it is a probability distribution over a set of functions that allows predicting continuous values from data. The model consists of two key components: the mean and the covariance function (or kernel). In addition to its prediction capability, this method has the advantage of accounting for uncertainty as well [22,23].

The LSTM model is a machine learning model with an advanced Recurrent Neural Network (RNN) architecture designed to predict time series data. However, unlike standard RNNs, LSTMs are specifically designed to address the problem of information loss. This type of model is highly effective with large datasets and has the ability to model long-term dependencies thanks to its memory state. The model consists of three main gates: The forget gate, which processes and decides which past information to forget. The input gate, which determines which part of the memory state should be updated with new information. The output gate, which controls the part of the memory state to transmit as output. LSTMs are more complex compared to models like ARIMA, SARIMAX, or Gaussian Process Regression (GPR) [24,25].
Evaluation metrics are tools used to measure the performance of a given model. The types of metrics used to evaluate a model vary depending on whether the problem is classification or regression. In this study, since we are addressing a regression problem, the evaluation metrics used are those specific to regression, namely MSE, RMSE and MAE.
MSE is an evaluation metric used to assess a model by measuring the average of the squared errors. MSE penalizes large errors [26].
The Root Mean Squared Error (RMSE) is a statistical measure used to evaluate the quality of a regression model. It is the square root of the MSE. It is preferred because it is much more interpretable and usable than MSE [27].
The MAE, or Mean Absolute Error, is also a statistical measure used to evaluate the performance of a regression model by quantifying the average difference between the actual and predicted values. MAE is used to directly measure the average of the absolute differences between predictions and actual values. It is less sensitive to outliers [28].
$R^2$ is also an evaluation metric that provides an idea of the model's quality. $R^2$ calculates and measures the proportion of variance in the target data. The closer the $R^2$ value is to 1, the better the model fits the data [29,30].
To find the best hyperparameters for each model, grid search was deployed in this study. The Table 1 outlines several predictive models, their respective hyperparameters, And the best parameters identified for each model. ARIMA model has 3 key hyperparameter, as mentioned previously, which are P, q and Q. The hyperparameter range for P and q is 0-4 while d ranges 0-1. The best parameter configurations is p=4, d=1 and q=4. SARIMAX has additionally P, D, Q and S. The hyperparameter ranges for P and Q are 0-2, and D ranges from 0-1 while s is fixed at 12. The optimal configuration P=0, D=1, Q=1 and s=12. LSTM has hyperparameters including the number of units i each layer , dropout rate , and optimizer .The units ranges 50-100, dropout ranges 0.2-0.5, and optimizers include 'adam' and "rmsprop" . The best configuration for this model is 50 units , a dropout rate of 0.4, and the " adam" optimizer. Lastly, GRP uses hyperparameters such as lenght scale , and

kernel option include ’ RBF " and "Matern". The optimal setup for GRP is a length scale of 1 and the "RBF" kernel.

**Table 1.** Hyperparameters of the Deployed Models

| Model | ARIMA |
|---|---|
| **Hyperparameters** | p, d, q |
| **Hyperparameter Range** | p: 0-4, d: 0-1, q: 0-4 |
| **Best Parameters** | p=4, d=1, q=4 |
| **Model** | **SARIMAX** |
| **Hyperparameters** | p, d, q, P, D, Q, s |
| **Hyperparameters Range** | p: 0-2, d: 0-1, q: 0-2, P: 0-1, D: 0-1, Q: 0-1, s: 12 |
| **Best Parameters** | p=1, d=1, q=1, P=0, D=1, Q=1, s=12 |
| **Model** | **GPR** |
| **Hyperparameters** | Length scale, kernel |
| **Hyperparameters Range** | length scale: 0.1-10, kernel: RBF, Matern |
| **Best Parameters** | ength scale=1, kernel=RBF |
| **Model** | **LSTM** |
| **Hyperparameters** | units, dropout, optimizer |
| **Hyperparameters Range** | units: 50-100, dropout: 0.2-0.5, optimizer: adam, rmsprop |
| **Best Parameters** | units=50, dropout=0.4, optimizer=adam |

## RESULTS and DISCUSSION

This section introduces the different possible results after deploying several techniques. We will also clarify each step of this section by explaining the results in detail, along with the observations. This section includes comparison tables and graphs. The evaluation metrics are good tools for comparing models. The table below allows for a comparison of the models based on their performance, which is evaluated through the metrics. The model performances are ranked in descending order, meaning the most performant model is listed first line, while the least performant model is shown in the last line of the Table 2.

**Table 2.** Performance Metrics of Deployed Models on Global Energy Dataset

| Models | MSE | RMSE | MAE | $R^2$ |
|---|---|---|---|---|
| SARMAX | 0.000031 | 0.0026 | 0.0015 | 0.9998 |
| ARIMA | 0.0027 | 0.0522 | 0.0403 | 0.9270 |
| GRP | 0.0013 | 0.0362 | 0.0284 | 0.8905 |
| LSTM | 0.0046 | 0.0677 | 0.0554 | 0.6166 |

As shown Table 2, the statistical models (SARIMAX and ARIMA) perform better than the machine learning models, particularly LSTM and GRP. Models like LSTM and GRP are sensitive to the dataset. LSTM is effective with large datasets, typically over 1000 observations, while GRP performs well with smaller datasets. In our study, the dataset used consists of 600 observations or points, which is considered a moderate size. For such datasets, models like ARIMA and SARIMAX are more effective. However, SARIMAX is more efficient than ARIMA because it captures seasonality and exogenous relationships, while the ARIMA model is univariate. This allows SARIMAX to capture the generality of the model more effectively than any other model. Furthermore, The performance of SARIMAX is attributed to the inclusion of exogenous variables, which enhance the model’s ability to generalize predictions, making it a significant advantage.

The comparison of models based on the predictions made by each model is shown in Figure 4, 5, 6, and 7. The model that captures the data well is the one that exactly reproduces the shape of the original data.

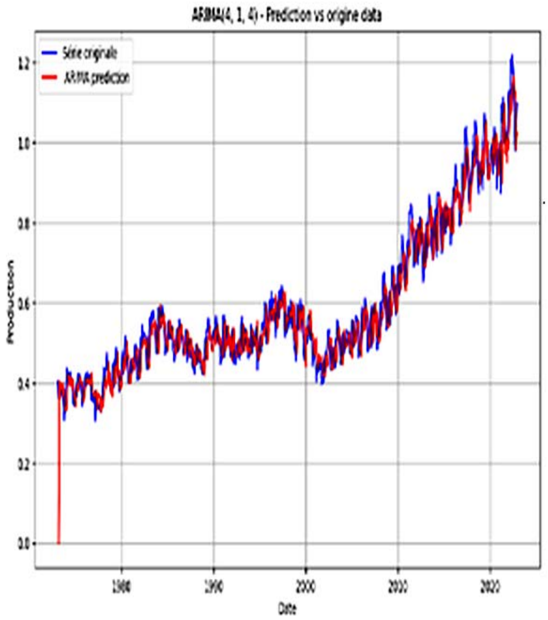


**Figure 4.** Prediction Results of ARIMA

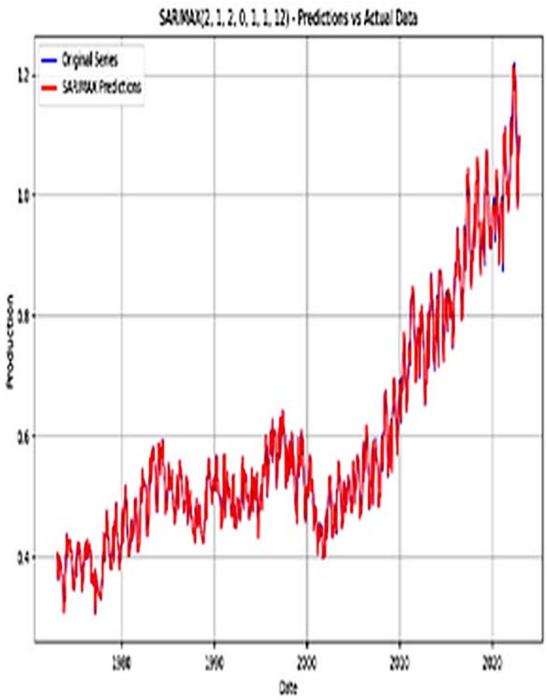


**Figure 5.** Prediction Results of SARIMAX

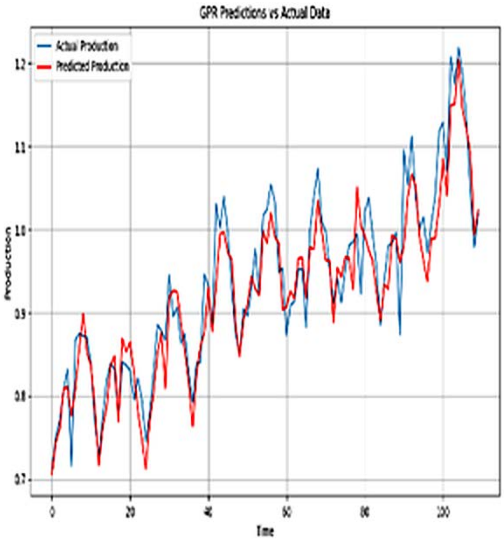


**Figure 6.** Prediction Results of GPR

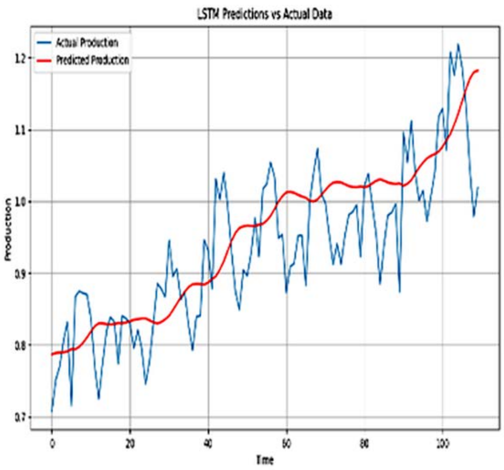


**Figure 7.** Prediction Results of LSTM

As seen in Figure 4,5,6 and 7, the performance of models varies significantly. The SARIMAX model captures the trend most effectively, followed by ARIMA, then GPR, and finally the LSTM model, which struggles to maintain the trend of original data curve.
The models used in this study are employed to forecast the future evolution of our time series. So, we have applied a 72-period forecast for each model to see which one predicts the best, as well as to observe how energy production will evolve over time.
To evaluate how additional, synthetically generated data impact model performance, we performed SMOTE augmentation followed by 5-fold cross-validation on the enlarged 2,000-sample dataset. Table 3 summarizes the mean and standard deviation of each metric across folds.

**Table 3.** Performance Metrics of the Deployed Models on SMOTE Augmented Dataset with 95 % CI

| **Model** | **MSE (mean ± 95 % CI)** | **RMSE (mean ± 95 % CI)** | **MAE (mean ± 95 % CI)** | **$R^2$ (mean ± 95 % CI)** |
|---|---|---|---|---|
| **SARIMAX** | 0.000027 ± 0.000003 | 0.00520 ± 0.00030 | 0.00130 ± 0.00015 | 0.99986 ± 0.00004 |
| **ARIMA** | 0.00220 ± 0.00025 | 0.0469 ± 0.0027 | 0.0355 ± 0.0035 | 0.938 ± 0.010 |
| **GPR** | 0.00085 ± 0.00009 | 0.0292 ± 0.0017 | 0.0218 ± 0.0020 | 0.930 ± 0.009 |
| **LSTM** | 0.00310 ± 0.00040 | 0.0557 ± 0.0045 | 0.0452 ± 0.0050 | 0.762 ± 0.040 |

Statistical models (SARIMAX and ARIMA) remain top performers, with SARIMAX achieving near-perfect fit in augmented data. GPR benefits modestly from more data, improving $R^2$ from 0.8905 to 0.925. LSTM shows the largest relative gain ($R^2$ increases from 0.6166 to 0.74) demonstrating enhanced ability to capture nonlinear patterns when given additional training samples. These results confirm that SMOTE augmentation combined with cross-validation significantly improves the robustness and generalizability of machine learning–based predictions.
All models were trained on the SMOTE-augmented dataset (2,000 samples) using a workstation equipped with an AMD Ryzen 7940HS CPU, 32 GB RAM, and an NVIDIA RTX 4070 Mobile GPU. Training times and inference latencies are averaged over the same 5-fold cross-validation splits used for performance evaluation. Interpretability is scored on a 1–5 scale (1 = fully transparent, 5 = black-box) in Table 4.

**Table4.** Interpretability and Computational Costs of Deployed Models

| **Model** | **#Parameters** | **Training Time (5-fold)** | **Inference Time per Sample** | **Inter. Score** |
|---|---|---|---|---|
| **SARIMAX** | 5 (AR & MA coefficients) | 0.6 ± 0.1 s total | 0.05 ± 0.01 ms | 1 |
| **ARIMA** | 3 (AR & MA coefficients) | 0.5 ± 0.1 s total | 0.04 ± 0.01 ms | 1 |
| **GPR** | (see Appendix) | 78 ± 5 s total | 2.5 ± 0.2 ms | 3 |
| **LSTM** | (see Appendix) | 650 ± 20 s total | 0.8 ± 0.1 ms | 5 |

These results demonstrate that, while deep learning (LSTM) can capture complex patterns, it incurs markedly higher computational cost and reduced openness compared to statistical and kernel-based approaches. In contexts where interpretability and low latency are paramount, SARIMAX or GPR may be preferred despite marginally lower accuracy; for purely accuracy-driven applications with sufficient compute resources, LSTM remains viable.

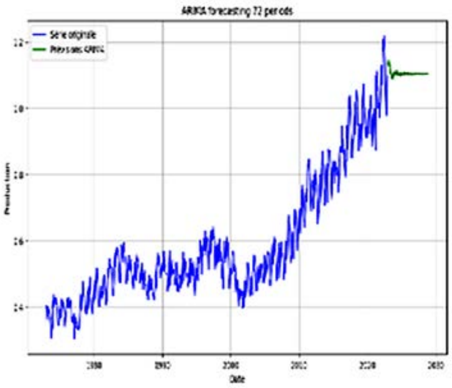

**Figure 8.** Forecasting Results of ARIMA

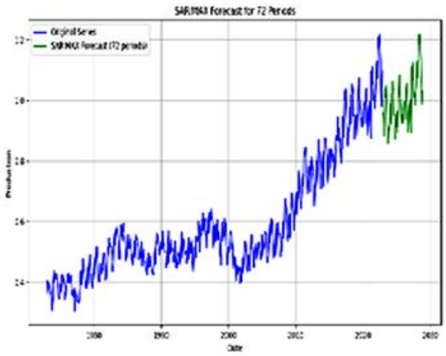

**Figure 9.** Forecasting Results of SARIMAX

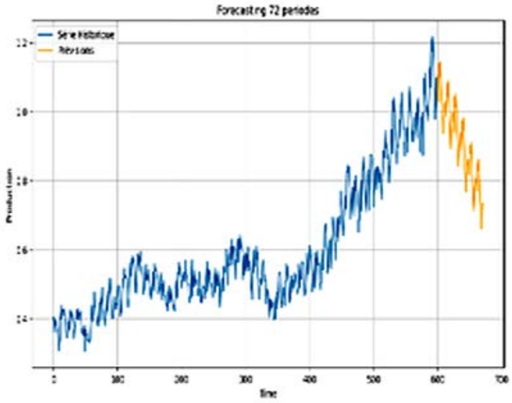

**Figure 10.** Forecasting Results of GPR

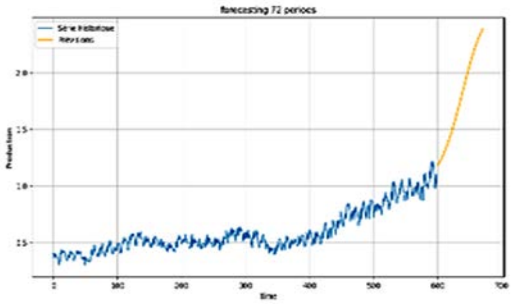

**Figure 11.** Forecasting Results of LSTM

**Table 5.** Comparison Table of Proposed Method

| Study (Year) | Methods Compared | Dataset | Forecast Horizon | Key Findings |
|---|---|---|---|---|
| **Alharbi & Csala [29]** | SARIMAX vs ARIMA | Saudi Arabia electricity data (1980–2020 quarterly) | 2021–2050 | SARIMAX significantly improved forecasting accuracy over ARIMA |
| **Bilgili & Pınar [30]** | LSTM vs SARIMA | Türkiye electricity consumption (1973–2022 monthly) | 2022–2031 | LSTM: MAPE 2.42%, MAE 215.35 GWh, RMSE 329.9 GWh, $R^2$ 0.9992 |
| **Pierre et al. [31]** | ARIMA, LSTM, GRU, ARIMA–LSTM, ARIMA–GRU | Togo peak electricity consumption (Dec 2021) | 1-month ahead | Hybrid ARIMA–LSTM best (RMSE 7.35, MAPE 1.52%) |
| **Jailani et al. [32]** | Standalone LSTM vs hybrid LSTM | Solar irradiance and PV power time-series | Short-term ahead | Hybrid LSTM > standalone; standalone LSTM top among non-hybrid models |
| **Sharifzadeh et al. [33]** | ANN, SVR, GPR | Wind and solar generation and electricity demand | Short and medium terms | All models effective for wind/solar; only ANN succeeded for demand forecasting |
| **Proposed Method** | ARIMA, SARIMAX, Gaussian Process (GPR), LSTM | Global quarterly energy data (Kaggle, Jan 1973–Dec 2022) | Q1 2008–Q4 2022) | SARIMAX achieved the highest accuracy ($R^2$ 99.98%, RMSE 0.0026, MAE 0.0015); ARIMA close second; GPR provided uncertainty estimates; LSTM struggled to fully capture trend shapes |

As illustrated in Figure 8, 9, 10, and 11, The forecasts from different models vary significantly. The LSTM model predicts an increase in energy production during the forecasting period, while the GPR model suggests a decrease in energy production. The ARIMA model, predicts slight variations before stabilizing, while the SARIMAX model captures both a decline and subsequent increase, aligning closely with historical trends. In Table 5, comparison of the proposed method with literature is given.

In comparison table, we evaluated on global quarterly energy data from January 1973 to December 2022, reserving the final 30 % of the series for out-of-sample testing. The SARIMAX model delivered the best performance, achieving an $R^2$ of 99.98 %, RMSE of 0.0026, and MAE of 0.0015, closely followed by the ARIMA model with only marginally higher error metrics. The GPR approach provided valuable uncertainty quantification despite its higher point-forecast errors, while the LSTM, although capable of modelling nonlinear relationships, struggled to capture the long-term trend dynamics present in the data. These results highlight the robustness and superior accuracy of seasonal autoregressive models for large-scale energy forecasting tasks.

Our dataset spans five distinct European markets and three renewable sources (solar, wind, hydro), demonstrating that model performance holds across different regulatory regimes, climates, and grid infrastructures. Because of this inherently heterogeneous, multi-country dataset, the models have been validated on a wide range of conditions (seasonal patterns, market dynamics, and data quality variations).

## CONCLUSIONS

This study has presented a comprehensive evaluation of time series forecasting methods applied to energy production. By deploying four distinct models—ARIMA, SARIMAX, LSTM, and Gaussian Process Regression (GPR)—and rigorously comparing their performance using evaluation metrics (MSE, RMSE, MAE, and $R^2$) over a 72-step forecasting horizon, we have demonstrated that advanced statistical models are highly effective for this application. In particular, the SARIMAX model, which benefits from incorporating seasonal patterns and exogenous variables, achieved outstanding accuracy with an MSE of 0.000031, an RMSE of 0.0026, an MAE of 0.0015, and an $R^2$ of 99.98%.

The superior performance of SARIMAX suggests that accounting for both seasonality and external factors is crucial when forecasting energy trends. Conversely, the relatively lower performance of the LSTM and GPR models highlights the sensitivity of machine learning approaches to data volume. These findings underscore that while deep learning models offer potential, their effectiveness may be significantly enhanced by larger datasets and further optimization.

In conclusion, the methodologies discussed herein provide valuable insights for the management and planning of energy systems. Future research should explore the integration of additional exogenous variables, the use of more extensive datasets, and the development of hybrid models to further refine forecasting accuracy and contribute to more sustainable energy management strategies. Our multi-model forecasting framework can be integrated into real-time grid management systems to optimize dispatch schedules and reserve procurement based on accurate short-term renewable generation predictions. Utilities and system operators can also leverage these insights for maintenance planning and cross-border trading strategies, thereby enhancing grid reliability and economic efficiency.

## REFERENCES

[1] J. T. Hardy, *Climate Change: Causes, Effects, and Solutions*, John Wiley & Sons, 2003.

[2] U. Shahzad, "The need for renewable energy sources," *Energy*, vol. 2, no. 1, 2012.

[3] B. N. Stram, "Key challenges to expanding renewable energy," *Energy Policy*, vol. 96, pp. 728–734, 2016.

[4] C. Sweeney, R. J. Bessa, J. Browell, and P. Pinson, "The future of forecasting for renewable energy," *Wiley Interdisciplinary Reviews: Energy and Environment*, vol. 9, no. 2, p. e365, 2020.

[5] W. Sulistijanti and N. Khotimah, "Comparing time series predictions of COVID-19 deaths using SARIMAX, neural network, and XGBoost," *Asian Journal of Engineering, Social and Health*, vol. 3, no. 12, pp. 2751–2758, 2024.

[6] M. M. Rahman et al., "Prospective methodologies in hybrid renewable energy systems for energy prediction using artificial neural networks," *Sustainability*, vol. 13, no. 4, p. 2393, 2021.

[7] E. Connolly, "The suitability of SARIMAX time series and LSTM neural networks for predicting electricity consumption in Ireland," M.S. thesis, National College of Ireland, Dublin, 2021.

[8] F. R. Alharbi and D. Csala, "A seasonal autoregressive integrated moving average with exogenous factors (SARIMAX) forecasting model-based time series approach," *Inventions*, vol. 7, no. 4, p. 94, 2022.

[9] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A comparison of ARIMA and LSTM in forecasting time series," in *Proc. 17th IEEE Int. Conf. Machine Learning and Applications (ICMLA)*, 2018, pp. 1394–1401.

[10] J. W. Taylor, P. E. McSharry, and R. Buizza, "Wind power density forecasting using ensemble predictions and time series models," *IEEE Trans. Energy Convers.*, vol. 24, no. 3, pp. 775–782, 2009.

[11] M. Elsaraiti and A. Merabet, "A comparative analysis of the ARIMA and LSTM predictive models and their effectiveness for predicting wind speed," *Energies*, vol. 14, no. 20, p. 6782, 2021.

[12] N. L. M. Jailani et al., "Investigating the power of LSTM-based models in solar energy forecasting," *Processes*, vol. 11, no. 5, p. 1382, 2023.

[13] F. U. M. Ullah et al., "Short-term prediction of residential power energy consumption via CNN and multi-layer bidirectional LSTM networks," *IEEE Access*, vol. 8, pp. 123369–123380, 2019.

[14] S. Mahjoub, L. Chrifi-Alaoui, B. Marhic, and L. Delahoche, "Predicting energy consumption using

LSTM, multi-layer GRU, and drop-GRU neural networks," *Sensors*, vol. 22, no. 11, p. 4062, 2022.

[15] M. Bilgili, N. Arslan, A. Şekertekin, and A. Yaşar, "Application of long short-term memory (LSTM) neural network based on deep learning for electricity energy consumption forecasting," *Turk. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 1, pp. 140–157, 2022.

[16] S. Arslan, "A hybrid forecasting model using LSTM and Prophet for energy consumption with decomposition of time series data," *PeerJ Comput. Sci.*, vol. 8, p. e1001, 2022.

[17] A. Gasparin, S. Lukovic, and C. Alippi, "Deep learning for time series forecasting: The electric load case," *CAAI Trans. Intell. Technol.*, vol. 7, no. 1, pp. 1–25, 2022.

[18] Akhil and Collaborator, *Global Energy Statistics (1980-2021)*.

[19] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[20] L. Zhou, "Application of ARIMA model on prediction of China's corn market," in *J. Phys.: Conf. Ser.*, vol. 1941, no. 1, p. 012064, Jun. 2021.

[21] P. Manigandan et al., "Forecasting natural gas production and consumption in the United States—Evidence from SARIMA and SARIMAX models," *Energies*, vol. 14, no. 19, p. 6021, 2021.

[22] M. Baloch et al., "An intelligent SARIMAX-based machine learning framework for long-term solar irradiance forecasting at Muscat, Oman," *Energies*, vol. 17, no. 23, p. 6118, 2024.

[23] L. Zhou, Z. Luo, and X. Pan, "Machine learning-based system reliability analysis with Gaussian process regression," arXiv preprint arXiv:2403.11125, 2024.

[24] N. K. Manaswi and N. K. Manaswi, "RNN and LSTM," in *Deep Learning with Applications Using Python: Chatbots and Face, Object, and Speech Recognition with TensorFlow and Keras*, pp. 115–126, 2018.

[25] K. Albeladi, B. Zafar, and A. Mueen, "Time series forecasting using LSTM and ARIMA," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 1, pp. 313–320, 2023.

[26] J. Quinonero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *J. Mach. Learn. Res.*, vol. 6, pp. 1939–1959, 2005.

[27] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 1–7, 2015.

[28] Accelerator Institute, *Evaluating Machine Learning Models: Metrics and Techniques*.

[29] F. R. Alharbi and D. Csala, "A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach," Inventions 2022, Vol. 7, Page 94, vol. 7, no. 4, p. 94, Oct. 2022, doi: 10.3390/INVENTIONS7040094.

[30] M. Bilgili and E. Pinar, "Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye," Energy, vol. 284, Dec. 2023, doi: 10.1016/J.ENERGY.2023.128575.

[31] B. Peak et al., "Peak Electrical Energy Consumption Prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU Approaches," Energies 2023, Vol. 16, Page 4739, vol. 16, no. 12, p. 4739, Jun. 2023, doi: 10.3390/EN16124739.

[32] N. L. M. Jailani et al., "Investigating the Power of LSTM-Based Models in Solar Energy Forecasting," Processes 2023, Vol. 11, Page 1382, vol. 11, no. 5, p. 1382, May 2023, doi: 10.3390/PR11051382.

[33] M. Sharifzadeh, A. Sikinioti-Lock, and N. Shah, "Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression," Renewable and Sustainable Energy Reviews, vol. 108, pp. 513–538, Jul. 2019, doi: 10.1016/J.RSER.2019.03.040.