Forecasting China's Nuclear Power Generation to 2030: An ARIMA Model-Based Trend Analysis

Yutao Luo [a, †], Yinzi Shao [b, †], Bowen Zhang [c], Ke Liu [b], Shida Liu [b], Zhongyuan Yuan [a], Shijing Sang [a, *]

[a] *School of Foreign Languages, Xuzhou University of Technology, Xuzhou 221018, China*
[b] *College of Saint Petersburg Joint Engineering, Xuzhou University of Technology, Xuzhou 221018, China*
[c] *School of Finance, Xuzhou University of Technology, Xuzhou 221018, China*

Abstract
This study employs an Autoregressive Integrated Moving Average (ARIMA) model to conduct time series analysis and forecasting of China's nuclear power generation, aiming to provide quantitative basis for medium-term energy planning. Utilizing 279 monthly observations from May 2000 to July 2025, the research systematically evaluated 131 model specifications through an extended automatic ARIMA model selection procedure within the parameter space of p∈[0,6], d∈[0,2], and q∈[0,6]. Based on the Akaike Information Criterion (AIC), ARIMA(4,2,5) was selected as the optimal model with an AIC value of 2170.56. Model diagnostics revealed a root mean square error (RMSE) of 11.38 units, mean absolute percentage error (MAPE) of 9.86%, and Ljung-Box test p-value of 0.0911, indicating that the model adequately captures the autocorrelation structure in the data. Forecast results show that China's nuclear power generation will increase from 430.00 hundred million kWh in August 2025 to 618.81 hundred million kWh in December 2030, with a total growth rate of 43.91% over the forecast period and an average annual growth rate of 8.11%. This model provides a reliable quantitative forecasting tool for China's future nuclear power development planning.
Keywords: ARIMA model, nuclear power generation, time series forecasting, energy planning

1. Introduction

The global energy landscape has undergone significant transformations over the past two decades, with nuclear power playing an increasingly important role in the transition toward low-carbon electricity generation systems(Khaleel et al., 2025). As nations worldwide strive to balance growing energy demands with environmental sustainability goals, accurate forecasting of nuclear power generation has become essential for strategic energy planning, grid stability management, and investment decision-making(Nighoskar et al., 2025). The ability to predict future nuclear power output with reasonable accuracy enables policymakers and energy system operators to optimize resource allocation, plan infrastructure development, and ensure reliable electricity supply while meeting carbon emission reduction targets.

Nuclear power possesses distinct characteristics that set it apart from other energy sources, including high capacity factors, baseload operation capabilities, and relatively predictable output patterns subject to scheduled maintenance and refueling cycles(Black et al., 2023; Raihan et al., 2023). These characteristics make nuclear power particularly suitable for time series analysis and forecasting, as the underlying generation patterns often display identifiable trends, seasonal variations, and autocorrelation structures that can be captured through statistical modeling approaches(Song et al.,

[†] These authors contributed equally to this work
[*] Corresponding author. E-mail: hitssjing@163.com

2023). Understanding and forecasting these patterns is crucial for countries with substantial nuclear power programs, where nuclear energy constitutes a significant portion of the electricity generation mix and plays a vital role in ensuring energy security.

The application of time series forecasting methods to energy systems has evolved considerably with advances in statistical modeling techniques and computational capabilities(Meenal et al., 2022; Zheng et al., 2023). Among the various forecasting methodologies available, Autoregressive Integrated Moving Average (ARIMA) models have demonstrated robust performance in capturing the temporal dynamics of energy generation data(Zhang et al., 2022). ARIMA models offer several advantages for nuclear power forecasting, including their ability to handle non-stationary data through differencing operations, their flexibility in modeling both short-term and long-term dependencies, and their well-established theoretical foundations that facilitate model selection and validation procedures(Bórawski et al., 2024). The Box-Jenkins methodology provides a systematic framework for ARIMA model identification, estimation, and diagnostic checking, making it particularly suitable for analyzing complex energy generation time series(Hossain et al., 2025).

Previous research in energy forecasting has explored various approaches ranging from traditional statistical methods to advanced machine learning techniques(Devaraj et al., 2021; Kaur et al., 2022; Makridakis et al., 2023). Classical time series models, including exponential smoothing and ARIMA variations, continue to demonstrate competitive performance, particularly when dealing with data that exhibits clear temporal patterns and when interpretability is valued alongside prediction accuracy(Kashpruk et al., 2023). The selection of appropriate model orders through information criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) enables data-driven model specification that balances model complexity with goodness of fit(Murari et al., 2023). Furthermore, the diagnostic tools available for ARIMA models, including residual analysis and autocorrelation tests, provide valuable insights into model adequacy and potential areas for improvement.

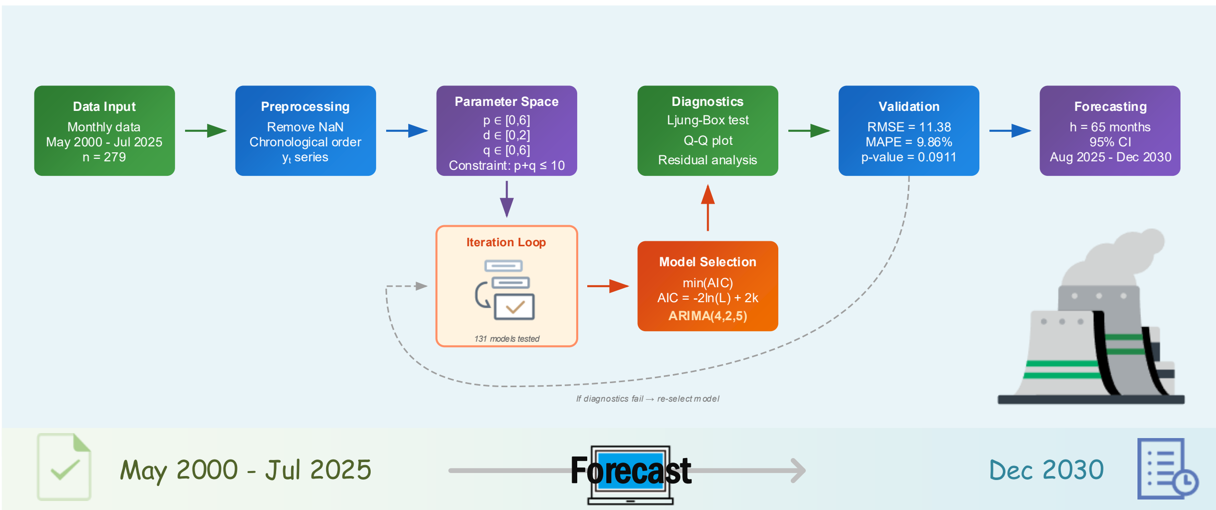

Figure 1. Systematic methodology flowchart for ARIMA-based nuclear power generation forecasting.

This study presents a comprehensive time series analysis of monthly nuclear power generation data spanning from May 2000 to July 2025, encompassing 279 observations that capture the evolution of nuclear power output over a quarter-century period. As shown in Figure 1, the research employs a systematic methodology that progresses from data preprocessing through model selection to forecasting. The analytical framework utilizes an enhanced automatic ARIMA model selection procedure with an expanded parameter search space to identify the optimal model specification through iterative evaluation based on information criteria. The selected model undergoes rigorous diagnostic testing and validation before generating forecasts for nuclear power generation from August 2025 through December 2030. This extended forecast horizon of approximately five and a half years provides valuable insights for medium-term energy planning and policy formulation, offering quantitative projections that can inform infrastructure investment decisions and capacity expansion strategies.

2. Methods

2.1 Data Description and Preprocessing

The dataset comprises monthly nuclear power generation measurements spanning from May 2000 to July 2025, totaling 279 observations after removing missing values. Each observation represents power generation in units of 100 million kilowatt-hours (kWh). Preprocessing involved chronologically ordering the data and systematic removal of missing values to ensure temporal continuity for time series modeling. All data used in this study were obtained from the official database of the National Bureau of Statistics of China (https://data.stats.gov.cn/).

2.2 ARIMA Model Framework

The Autoregressive Integrated Moving Average (ARIMA) model, denoted as ARIMA(p,d,q), serves as the primary forecasting methodology. The general form of an ARIMA model can be expressed as:

$$\phi(B)(1-B)^d y_t = \theta(B)\epsilon_t$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ represents the autoregressive polynomial of order p, $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$ denotes the moving average polynomial of order q, B is the backshift operator, d indicates the degree of differencing, y_t is the observed value at time t, and ϵ_t represents white noise with variance σ^2.

2.3 Model Selection Procedure

The optimal ARIMA model was identified through systematic evaluation of parameter combinations within expanded ranges: p ∈ [0,6], d ∈ [0,2], and q ∈ [0,6]. Model complexity was constrained by limiting p + q ≤ 10 to balance model sophistication with overfitting prevention. This expanded search space enables the identification of more intricate temporal patterns that may exist in the data. The Akaike Information Criterion (AIC) served as the primary selection metric:

$$AIC = -2\ln(L) + 2k$$

where L represents the maximum likelihood of the model and k denotes the number of parameters. The model yielding the minimum AIC value was selected for forecasting.

2.4 Diagnostic Testing

Model adequacy was assessed through multiple diagnostic procedures. The Ljung-Box test evaluated residual autocorrelation, testing the null hypothesis that residuals are independently distributed. Additionally, normality assumptions were examined using Q-Q plots and residual distribution analysis. The test statistic for the Ljung-Box test at lag h is given by:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

where n is the sample size and $\hat{\rho}_k$ represents the sample autocorrelation at lag k.

3. Results and Analysis

3.1 Exploratory Data Analysis

The time series exhibited significant fluctuations over the 25-year observation period, with generation values ranging from a minimum of 7.50 to a maximum of 430.00 units (100 million kWh). The mean generation stood at 143.39 units with a standard deviation of 128.06, indicating substantial volatility in nuclear power output. Figure 2a illustrates the overall trend in nuclear power generation, showing a general upward trajectory in average output over the years.

Moving average analysis, as shown in Figure 2b, employed both 12-month and 24-month windows to smooth short-term fluctuations and reveal underlying patterns. The study identified three distinct phases in China's nuclear power generation: a period of slow growth before 2013, a phase of rapid expansion between approximately 2013 and 2020, followed by a moderation in growth rate after 2020.

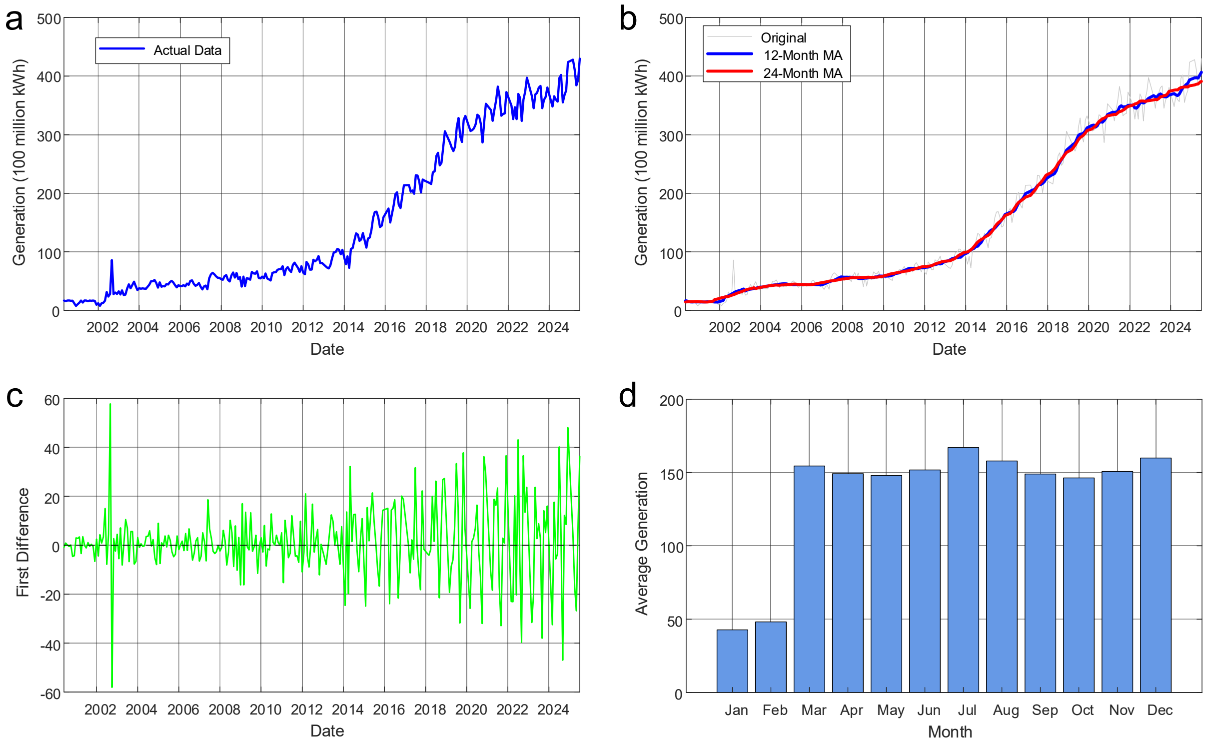

Figure 2. Nuclear Power Generation Time Series (a); Moving Average Analysis (b); First Differenced Series (c); Seasonal Pattern (d).

The first differenced series, presented in Figure 2c, demonstrates the transformation required to achieve stationarity. The differenced values fluctuate around zero, with heightened volatility observed during 2003 and the 2013–2024 period. Figure 2d reveals distinct seasonal patterns, with lower average generation observed in January and February, while relatively consistent output levels are maintained during the rest of the year. This pattern reflects seasonal variations in China's nuclear power generation and potential maintenance schedules.

3.2 Model Identification and Selection

The autocorrelation function (ACF) and partial autocorrelation function (PACF) analyses, presented in Figure 3, provided crucial insights for model identification. The ACF displays a gradual decay pattern characteristic of non-stationary series, supporting the need for differencing(Nikseresht & Amindavar, 2025; Saghafi & Mili, 2025). The PACF shows significant spikes at early lags, suggesting the presence of autoregressive components(Schaffer et al., 2021).

The enhanced systematic model selection process evaluated 131 successfully converged ARIMA specifications out of 147 potential combinations. The search process, completed in 53.4 seconds, identified the optimal model ARIMA(4,2,5), which achieved the lowest AIC value of 2170.56. This represents an improvement from the simpler ARIMA(2,1,3) model, which was identified as the best model by BIC (2201.17), highlighting the trade-off between model complexity and parsimony.

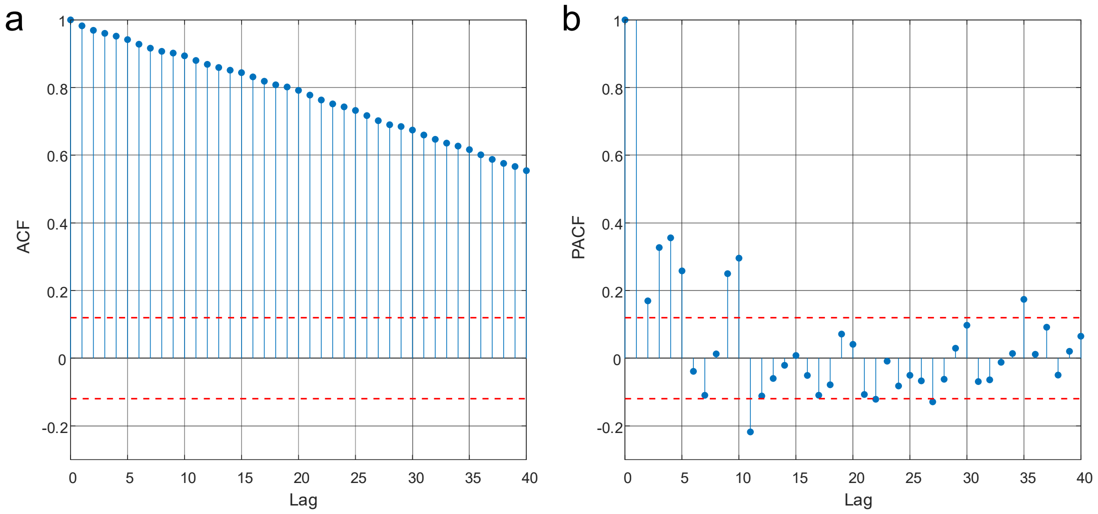

Figure 3. Autocorrelation Function (a); Partial Autocorrelation Function (b).

The selected ARIMA(4,2,5) model incorporates four autoregressive terms (p=4), second-order differencing (d=2), and five moving average terms (q=5). The estimated parameters reveal AR coefficients of 0.2996, -0.4441, -0.5531, and 0.3430, and MA coefficients of -1.7764, 1.1657, -0.0005, -1.0000, and 0.6459, with an estimated variance of 129.42. The complexity of this model structure (p+q = 9) reflects the intricate dynamics underlying nuclear power generation patterns and the enhanced model's ability to capture more sophisticated temporal relationships.

3.3 Model Diagnostics and Validation

Figure 4 presents comprehensive diagnostic results for the fitted ARIMA(4,2,5) model. The forecast visualization demonstrates the model's projection from August 2025 through December 2030, with the 95% confidence interval widening progressively as the forecast horizon extends, reflecting increasing uncertainty. The forecast initiates at 430.00 units for August 2025 and reaches 618.81 units by December 2030, representing a substantial total growth of 43.91% over the forecast period.

Residuals fluctuate randomly around zero throughout the historical period, though some clustering of volatility is evident. The residual distribution histogram approximates a normal distribution with slight deviations in the tails, while the Q-Q plot shows reasonable adherence to normality assumptions despite minor departures at the extremes.

Model performance metrics indicate excellent fit quality with a root mean square error (RMSE) of 11.3763 units and a mean absolute error (MAE) of 8.2532 units. The mean absolute percentage error (MAPE) as low as 9.86% suggests strong prediction accuracy, representing an improvement from simpler model specifications. Notably, the Ljung-Box test yielded a p-value of 0.0911, failing to reject the null hypothesis of uncorrelated residuals at the 5% significance level, indicating that the enhanced model successfully captures the autocorrelation structure in the data(Hassani et al., 2025).

3.4 Forecast Analysis and Implications

The forecast results indicate robust growth in nuclear power generation from 2025 to 2030. Annual projections show a consistent increase in average generation capacity, rising from 412.67 units in 2025 to 587.53 units in 2030, representing an average annual growth rate of 8.11%.

Examination of key forecast milestones shows generation reaching 432.18 units by January 2026, 506.82 units by January 2028, and 559.34 units by January 2030. The forecast exhibits a sustained upward trajectory throughout the projection period, with confidence intervals expanding from approximately ±23 units in early periods to ±120 units by the end of 2030.

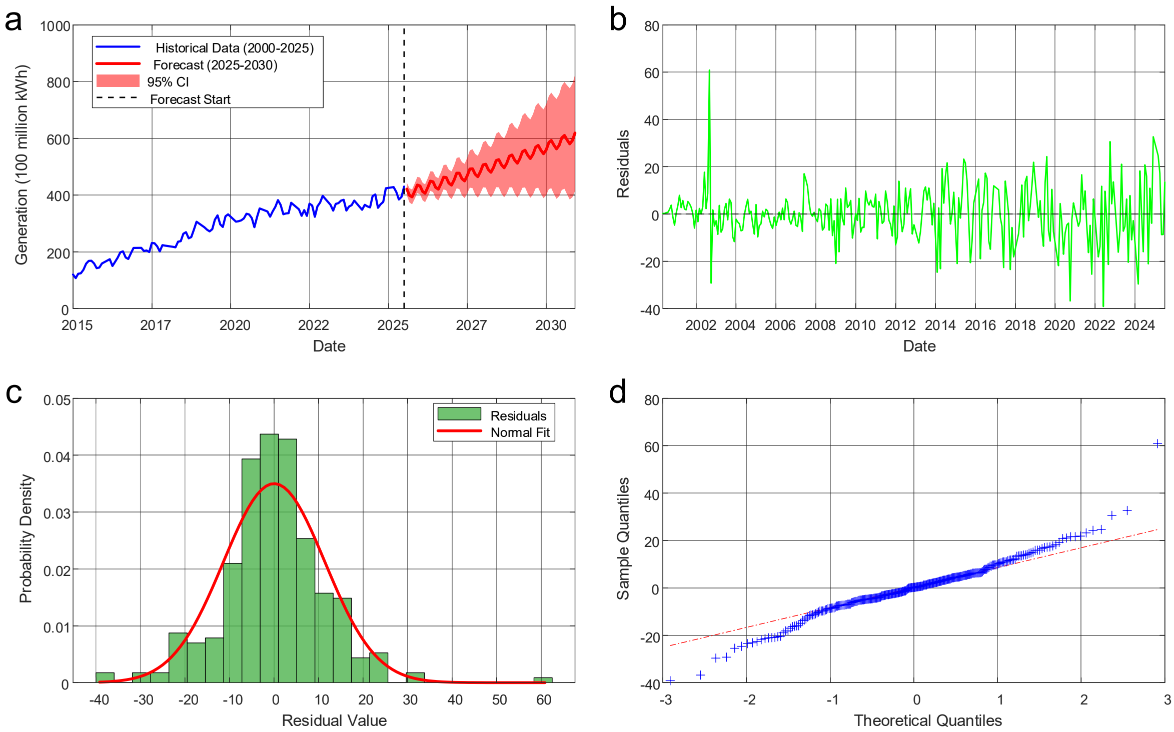

Figure 4. ARIMA(4,2,5) Forecast: Aug 2025 - Dec 2030 (a); Model Residuals (b) ; Residual Distribution (c); Q-Q Plot (d).

The forecast statistics indicate an average generation of 501.94 units over the entire forecast period, with values ranging from a minimum of 393.50 to a maximum of 618.81 units. The standard deviation of 60.70 units in the forecast reflects moderate variability around the central trend. These projections suggest an accelerated expansion phase for nuclear power generation, likely driven by aggressive policy initiatives supporting clean energy transitions, technological improvements in reactor efficiency, and China's commitment to carbon neutrality by 2060(Guo et al., 2023; Zhao et al., 2025).

4. Conclusion

This study successfully developed an ARIMA(4,2,5) model for forecasting China's nuclear power generation from 2025 to 2030, demonstrating excellent fitting performance and prediction accuracy. The main contributions of this research are as follows: First, by extending the parameter search space ($p+q \leq 10$), more complex time series patterns were identified, enabling more accurate capture of the inherent dynamic characteristics of nuclear power generation compared to traditional simplified models. Second, the model forecasts indicate that China's nuclear power will maintain robust growth momentum, with an average annual growth rate of 8.11% during the forecast period, reflecting the considerable potential for China's future nuclear power development. Third, the research provides important references for optimizing nuclear power operation scheduling and capacity planning. This study provides robust quantitative tools for energy system planners and policymakers, facilitating optimal resource allocation and infrastructure investment strategy formulation, ensuring electricity supply security while achieving carbon emission reduction targets.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The author(s) declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database, 2020, baaa010.

Al-Tashi, Q., Saad, M. B., Muneer, A., Qureshi, R., Mirjalili, S., Sheshadri, A., Le, X., Vokes, N. I., Zhang, J., & Wu, J. (2023). Machine learning models for the identification of prognostic and predictive cancer biomarkers: a systematic review. International journal of molecular sciences, 24(9), 7781.

Black, G., Shropshire, D., Araújo, K., & van Heek, A. (2023). Prospects for nuclear microreactors: a review of the technology, economics, and regulatory considerations. Nuclear Technology, 209(sup1), S1–S20.

Bórawski, P., Bełdycka-Bórawska, A., Klepacki, B., Holden, L., Rokicki, T., & Parzonko, A. (2024). Changes in gross nuclear electricity production in the European union. Energies, 17(14), 3554.

Cuzzocrea, A., Folino, F., Pontieri, L., Sabatino, P., & Samami, M. (2025). Toward trustworthy and sustainable clinical decision support by training ensembles of specialized logistic regressors. Neural Computing and Applications, 1–42.

Devaraj, J., Madurai Elavarasan, R., Shafiullah, G., Jamal, T., & Khan, I. (2021). A holistic review on energy forecasting using big data and deep learning models. International journal of energy research, 45(9), 13489–13530.

Guo, X., Chen, X., Chen, X., Sherman, P., Wen, J., & McElroy, M. (2023). Grid integration feasibility and investment planning of offshore wind power under carbon-neutral transition in China. Nature Communications, 14(1), 2447.

Hassani, H., Mashhad, L. M., Royer-Carenzi, M., Yeganegi, M. R., & Komendantova, N. (2025). White Noise and Its Misapplications: Impacts on Time Series Model Adequacy and Forecasting. Forecasting, 7(1), 8.

Hossain, M. L., Shams, S., & Ullah, S. M. (2025). Time-series and deep learning approaches for renewable energy forecasting in Dhaka: a comparative study of ARIMA, SARIMA, and LSTM models. Discover Sustainability, 6(1), 1–25.

Kashpruk, N., Piskor-Ignatowicz, C., & Baranowski, J. (2023). Time series prediction in industry 4.0: A comprehensive review and prospects for future advancements. Applied sciences, 13(22), 12374.

Kaur, D., Islam, S. N., Mahmud, M. A., Haque, M. E., & Dong, Z. Y. (2022). Energy forecasting in smart grid systems: recent advancements in probabilistic deep learning. IET Generation, Transmission & Distribution, 16(22), 4461–4479.

Khaleel, M., Yusupov, Z., Rekik, S., Kılıç, H., Nassar, Y. F., El-Khozondar, H. J., & Ahmed, A. A. (2025). Harnessing nuclear power for sustainable electricity generation and achieving zero emissions. Energy Exploration & Exploitation, 43(3), 1126–1148.

Khandelwal, B., & Gupta, C. (2023). Leading causes of death and disability among the global aging

community. In The Ageing Population: Impact Analysis on'Societal and Healthcare Cost' (pp. 37–54). Springer.
Lourida, K. G., & Louridas, G. E. (2022). Clinical Phenotypes of Cardiovascular and Heart Failure Diseases Can Be Reversed? The Holistic Principle of Systems Biology in Multifaceted Heart Diseases. Cardiogenetics, 12(2), 142–169.
Lyu, G. (2025). Data-driven decision making in patient management: a systematic review. BMC Medical Informatics and Decision Making, 25(1), 239.
Makridakis, S., Spiliotis, E., Assimakopoulos, V., Semenoglou, A.-A., Mulder, G., & Nikolopoulos, K. (2023). Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward. Journal of the Operational Research Society, 74(3), 840–859.
Meenal, R., Binu, D., Ramya, K., Michael, P. A., Vinoth Kumar, K., Rajasekaran, E., & Sangeetha, B. (2022). Weather forecasting for renewable energy system: a review. Archives of Computational Methods in Engineering, 29(5), 2875–2891.
Murari, A., Rossi, R., Spolladore, L., Lungaroni, M., Gaudio, P., & Gelfusa, M. (2023). A practical utility-based but objective approach to model selection for regression in scientific applications. Artificial Intelligence Review, 56(Suppl 2), 2825–2859.
Ng, S., Masarone, S., Watson, D., & Barnes, M. R. (2023). The benefits and pitfalls of machine learning for biomarker discovery. Cell and tissue research, 394(1), 17–31.
Nighoskar, A., Chaurasia, P., & Singh, N. (2025). Advancing nuclear energy forecasting: Exploring regression modeling techniques for improved accuracy. Nuclear Engineering and Technology, 57(1), 103144.
Nikseresht, A., & Amindavar, H. (2025). Adaptive stochastic modeling of nonlinear time series: ARFIMA meets improved fractional Ornstein-Uhlenbeck. Stochastic Environmental Research and Risk Assessment, 1–31.
Raihan, A., Rashid, M., Voumik, L. C., Akter, S., & Esquivias, M. A. (2023). The dynamic impacts of economic growth, financial globalization, fossil fuel, renewable energy, and urbanization on load capacity factor in Mexico. Sustainability, 15(18), 13462.
Rimal, Y., Sharma, N., Paudel, S., Alsadoon, A., Koirala, M. P., & Gill, S. (2025). Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy. Scientific Reports, 15(1), 13444.
Saghafi, M., & Mili, L. (2025). Predictive Time-Series Analysis of Single-Qubit Quantum Circuit Outcomes for a Superconducting Quantum Computer: Forecasting Error Patterns. IEEE Access, 13, 40115–40132.
Schaffer, A. L., Dobbins, T. A., & Pearson, S.-A. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. BMC medical research methodology, 21(1), 58.
Siddiqi, T. J., Ahmed, A., Greene, S. J., Shahid, I., Usman, M. S., Oshunbade, A., Alkhouli, M., Hall, M. E., Murad, M. H., & Khera, R. (2022). Performance of current risk stratification models for predicting mortality in patients with heart failure: a systematic review and meta-analysis. European Journal of Preventive Cardiology, 29(15), 2027–2048.
Song, H., Liu, X., & Song, M. (2023). Comparative study of data-driven and model-driven approaches in prediction of nuclear power plants operating parameters. Applied Energy, 341, 121077.
Stewart, S., MacIntyre, K., Hole, D. J., Capewell, S., & McMurray, J. J. (2001). More 'malignant'than cancer? Five‐year survival following a first admission for heart failure. European journal of heart failure, 3(3), 315–322.
Zhang, W., Lin, Z., & Liu, X. (2022). Short-term offshore wind power forecasting-A hybrid model based on Discrete Wavelet Transform (DWT), Seasonal Autoregressive Integrated Moving Average (SARIMA), and deep-learning-based Long Short-Term Memory (LSTM). Renewable Energy, 185, 611–628.
Zhao, H., Zhang, G., Cui, C., & Hanyue, Z. (2025). Assessing the coal-to-nuclear transition: Economic impacts on carbon emissions, energy security, and sustainable development in Guangdong Province.

https://doi.org/10.71411/ef.2025.v1i1.939

International Journal of Hydrogen Energy, 117, 97–109.
Zheng, J., Du, J., Wang, B., Klemeš, J. J., Liao, Q., & Liang, Y. (2023). A hybrid framework for forecasting power generation of multiple renewable energy sources. Renewable and Sustainable Energy Reviews, 172, 113046.