Analytical and Numerical Solutions for Nonlinear Equations
©Available online at https://ansne.du.ac.ir/
Online ISSN: 3060–785X
2025, Volume 10, Issue 2, pp. 162–190
Research article

An Improved Time-Delay Grey Verhulst Model Optimized by Multi-Agent Reinforcement Learning for Electricity Market Forecasting

Sajedeh Hedayatollahi Pour[1], Seyed Ali Alavizadeh [2], Omid Solaymani Fard[1,*]

[1] Department of Applied Mathematics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

[2] Alma Mater Studiorum Università di Bologna, Bologna, Italy

* **Corresponding author(s):** soleimani@um.ac.ir; omidsfard@gmail.com

Received: 08/10/2025 **Revised:** 29/11/2025 **Accepted:** 11/12/2025 **Published:** 30/12/2025 10.22128/ansne.2025.3087.1160

Abstract

Accurate forecasting of electricity prices is still a difficult task because of the volatile and nonlinear nature of energy markets, as well as the limited availability of reliable data. Grey forecasting models are often used for this purpose, but they usually lack enough flexibility to capture delayed effects and complex interactions among several variables. To address these issues, this study aims to present an Improved Time-Delay Grey Multivariable Verhulst Model (ITGMVM), a new grey model developed for short-term electricity price prediction. The model introduces time-delay parameters that represent lagged relationships between variables, helping it respond better to dynamic market behavior. Two optimization methods are designed for parameter calibration: Partial Parameter Estimation (PPE) and Full Parameter Estimation (FPE), where the latter adjusts all parameters at the same time. These methods are supported by a new hybrid optimization framework called MARL-WOA, which combines Multi-Agent Reinforcement Learning (MARL) with Whale Optimization Algorithm (WOA). This combination improves the search process, leading to faster convergence and higher accuracy. The model is evaluated using real-world data from Australia's National Electricity Market (NEM), specifically focusing on Sundays and Wednesdays between December 2023 and March 2024. Results show that ITGMVM, when optimized with FPE-MARL-WOA, outperforms six existing grey and hybrid models across multiple statistical metrics, achieving exceptional forecasting accuracy and robustness. The obtained results demonstrate the strength of integrating adaptive AI techniques with grey modeling to support decision-making in data-constrained energy environments.

Keywords: Grey model, Time series, Multi-Agent reinforcement learning, Metaheuristic algorithms, Electricity market

Mathematics Subject Classification (2020): 65K10, 68T20, 90C59

1 Introduction

Accurate Electricity Price Forecasting (EPF) is vital for the efficient operation of electricity markets and for the wider energy sector. It lets market participants – generators, retailers and large consumers – make informed choices about bidding, contracting and investment planning [12]. As an example, generators use price forecasts to organize maintenance and unit dispatch, while retailers set tariffs and hedge procurement risks on that basis. Policymakers and regulators also rely on sound forecasts when they design energy policies, support

renewable integration and work to keep markets stable [13].

The need for EPF has grown with the rapid spread of wind and solar power. These sources are periodic and depend on weather, so they add extra volatility to electricity prices and make the price path harder to predict [14]. Because of this, good short-term forecasts are essential for balancing supply and demand, omitting price spikes and protecting grid reliability.

However, price forecasting is difficult. Prices react sharply to fuel-cost swings, sudden load changes and network bottlenecks, all of which raise volatility and complicate modeling [15]. The situation is made harder by the growing use of distributed resources and demand-side programs, which change the way prices form in the market. Data problems add another layer: Real-time records of consumption, generation, and grid status can be unorganized, inconsistent, or protected by privacy rules; these specifics limit the model quality and slow decision-making down [14].

To meet these issues, researchers have turned to advance tools such as machine learning algorithms, hybrid statistical models and physics-inspired methods, that can learn non-linear patterns and adapt to fast-changing markets. Some examples of recent studies which have used the advanced tools and mechanism can be mentioned,

- In [29], Ghimire et al. combined wavelet-based noise filtering with a CNN feature extractor and a lightweight Random-Vector-Functional-Link regressor, resulting in sharper half-hourly predictions than any stand alone network could achieve.
- In [30], Xu et al. decomposed the raw price signal into cleaner sub-bands using variational-mode decomposition, processed these bands through an attention-enhanced LSTM, and then used the Grey Wolf Optimizer to fine-tune every hyper parameter seamlessly integrating signal decomposition, deep learning, and nature-inspired optimization into a single adaptive loop.
- In [31], Meng et al. tackled the volatility of a wind- and solar-heavy market by training an attention-based LSTM, whose weights and learning rates were guided by the Criss-Cross Optimization algorithm keeping day-ahead errors in check even during steep renewable ramps.

These studies show that hybrid deep learning architectures, metaheuristic tuning, and attention mechanisms have become a key focus for capturing the tangled, fast moving dynamics of electricity prices. Even so, these approaches still depend on high quality data, underscoring how crucial better data collection and open data sharing will be for future progress in EPF. Taken together, these studies suggest that hybrid deep learning pipelines paired with metaheuristic algorithms have set a new benchmark for predictive accuracy in data rich power markets.

The catch, however, is that each component of these pipelines from attention modules and deep networks to evolutionary hyper parameter tuning relies on abundant, high quality data. Yet in many regional or newly deregulated markets, such data remain fragmented, noisy, or locked behind commercial fire walls. When data are scarce, even the most sophisticated tech stack struggles to deliver accurate results.

This ongoing lack of data has sparked new interest in forecasting methods that can work with limited and unreliable information. One of the most important of these methods is called grey system theory.

Grey system theory, originally proposed by Deng in the 1980s, is a powerful forecasting framework designed to model uncertain systems with incomplete or sparse data. Unlike traditional statistical models that require large data sets and clear probability distributions, grey models are particularly effective in scenarios where data is limited, noisy, or only partially known making them ideal for energy market forecasting where such conditions often arise.

The classical model GM(1,1) is a first-order univariate grey differential model that uses data transformation and accumulation to uncover underlying trends. It has since been extended into several key variants to capture more complex system dynamics. For example, GM(1,N) introduces multiple influencing variables, enhancing the model's ability to represent multivariable relationships. The grey Verhulst model incorporates nonlinear saturation effects and is suitable for forecasting in systems that demonstrate logistic growth behavior, such as energy demand. More recent developments include the Time-Delay Grey Model (TD-GM), which accounts for lagged effects by integrating past values with time-lag weight parameters. This is particularly relevant in electricity price forecasting, where the influence of variables like load, temperature, and generation may exhibit delayed impact.

Recent research has increasingly explored the use of these grey model variants for electricity price prediction. For instance, *Li et al.* [16] proposed a hybrid GM(1,N)-ARIMA model to capture both short-term trends and stochastic residuals in market prices. *Zhao et al.* [17] introduced an improved time-delay grey model combined with a genetic algorithm for forecasting peak load prices. Moreover, *Chen et al.* [18] developed a multivariable grey Verhulst model optimized via particle swarm optimization (PSO) to improve forecasting accuracy under volatile market conditions. These studies demonstrate that grey models, especially when hybridized with intelligent optimization

or machine learning techniques, can outperform conventional models like ARIMA and support vector regression (SVR) in capturing both linear and nonlinear price dynamics.

The growing popularity of grey forecasting methods in energy systems is largely due to their robustness in handling uncertain, small-sample environments precisely the type of scenario encountered in regional electricity price data sets. As a result, grey models and their variants continue to be widely adopted and enhanced, offering a flexible foundation for accurate electricity price forecasting in complex, data-constrained environments.

The remainder of this paper is organized as follows. Section 2 reviews the main concepts and recent developments related to grey forecasting models, metaheuristic optimization algorithms, and MARL. Section 3 introduces two optimization approaches, including the formulation of the Grey Multivariable Verhulst Model and its time-delay extension. The proposed hybrid optimization strategy is presented in Section 4, where the MARL-WOA framework is described and analyzed. Section 5 provides an empirical investigation, including the evaluation framework, performance comparison, and forecasting experiments for Sunday and Wednesday electricity price predictions. Finally, Section 6 summarizes the main findings and conclusions of this study.

2 Preliminaries

In this section, we review the essential background and theoretical foundations relevant to the proposed model. First, the recent research on grey forecasting models for electricity price prediction is discussed to highlight the advantages and current challenges of existing approaches. Next, the role of metaheuristic algorithms in improving grey model performance is examined, emphasizing their capability in parameter tuning and global optimization. Finally, the concept of MARL is introduced, providing the basis for the hybrid optimization strategy developed later in this paper.

2.1 Recent Research on Grey Models for Electricity Price Forecasting

Recent advances in grey system theory have significantly extended its application in electricity price forecasting, particularly through the development of hybrid models that integrate grey forecasting with optimization and machine learning techniques. Table 1 summarizes key studies from recent years, highlighting the diversity of grey model variants and their integration with other algorithms for enhanced prediction accuracy.

Grey forecasting models have evolved into a rich taxonomy, as illustrated in Figure 1. The classical GM(1,1) remains a foundational tool for single-variable forecasting. Its multivariable extension, GM(1,N), accounts for multiple influencing factors, while the Verhulst model incorporates logistic-type nonlinearity. Time-delay grey models (TD-GM) represent an important innovation, enabling the integration of lagged effects crucial in electricity markets where system inputs affect prices over time. Recent advances further include fractional-order grey models, such as C-FGM, which enhance adaptability by capturing memory effects in complex load dynamics. Additionally, nonlinear mixed-frequency grey panel models have expanded the framework's capability to handle heterogeneous temporal resolutions, improving forecasting accuracy in modern electricity price environments.

These advances reflect an overarching trend: while grey models are effective in small-sample, uncertain environments, their predictive accuracy improves significantly when coupled with intelligent optimization. Hybrid approaches often use metaheuristic algorithms such as genetic algorithms, particle swarm optimization, or firefly algorithms to fine-tune grey model parameters. This synergy enables grey models to better adapt to complex, dynamic price signals and makes them more suitable for real-world deployment in deregulated electricity markets.

Thus, the integration of grey models with metaheuristic algorithms forms a promising direction for developing robust forecasting tools.

2.2 Literature Gaps

Time-delay grey models like TD-GM(1,N) are useful because they can include past information, but they still have important limitations. Their delay structures are usually fixed or too simple, and they treat all delayed inputs the same way. As a result, they cannot fully describe the complex and nonlinear behavior of electricity prices, especially in markets with strong and changing lag effects. The proposed ITGMVM model solves these problems by using a nonlinear Verhulst structure together with flexible weighted time delays, which allows it to capture different strengths of lag and saturation effects more accurately. In addition, instead of estimating parameters separately like older models, ITGMVM uses a full-parameter learning method supported by the MARL-WOA algorithm, which helps it find better parameter

Table 1. Recent Applications of Grey Models in Electricity Price (and Load) Forecasting (20212025)

Study	Grey Model Variant	Hybrid Method	Application / Contribution
Li et al. (2021) [16]	GM(1,N)	ARIMA	Captures short-term trends and residuals in power market prices.
Zhao et al. (2022) [17]	TD-GM(1,1)	Genetic Algorithm (GA)	Improves peak price forecast accuracy with delay parameter tuning.
Chen et al. (2023) [18]	GM-Verhulst	Particle Swarm Optimization (PSO)	Models nonlinear saturation effects in volatile market conditions.
Wang et al. (2022) [19]	GM(1,1)	Support Vector Regression (SVR)	Combines GM trend fitting with SVR residual learning.
Liu et al. (2020) [20]	GM(1,N) with delay	Firefly Algorithm	Forecasts electricity spot prices using metaheuristic tuning of grey parameters.
Yu et al. (2025) [21]	C-FGM	Global Optimization	Short-term electrical load forecasting using a clustering fractional-order grey model.
Gou et al. (2025) [22]	Nonlinear Mixed-Frequency Grey Panel Model	Spatio-Temporal Hybrid Prediction	Short-term electricity price forecasting using a nonlinear mixed-frequency grey model.

combinations and avoid early convergence. Compared with common algorithms such as GA or PSO, MARL-WOA can explore the search space more intelligently and adjust its strategy during learning, leading to more stable and reliable optimization. This gives the model a stronger ability to detect complex parameter interactions and achieve higher forecasting accuracy. Overall, ITGMVM is more than a minor enhancement; it represents a clearly stronger and more reliable model that provides higher accuracy and stability in electricity price forecasting.

2.3 Metaheuristic Algorithms and Their Role in Enhancing Grey Forecasting Models

Metaheuristic algorithms are general problem solving methods that help simpler rules search large, complicated solution spaces more efficiently. Unlike traditional optimization techniques, they cannot guarantee the absolute best answer, but they usually find very good ones in a reasonable amount of time. They are especially useful for problems that are nonlinear, have many peaks and valleys, or are NP-hard cases where standard methods often get stuck or take far too long to run [23].

In general, we can divide metaheuristic methods into four main groups:

- **Evolutionary Algorithms (EAs):** Genetic Algorithm (GA), Differential Evolution (DE), Evolutionary Strategies (ES). Inspired by natural selection, EAs evolve a population of candidate solutions through operators such as selection, crossover, and mutation.
- **Swarm Intelligence Algorithms (SIAs):** Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Firefly Algorithm (FA). These simulate the collective behaviour of social organisms where agents cooperate or compete to search the solution space.
- **Physics-Inspired Algorithms:** Simulated Annealing (SA), Gravitational Search Algorithm (GSA), Whale Optimization Algorithm (WOA). These methods are inspired by physical processes, natural phenomena, or mathematical analogies.
- **Human-Based Algorithms:** TeachingLearning-Based Optimization (TLBO), Brain Storm Optimization (BSO), Cultural Algorithm (CA). These methods model elements of human learning, social interaction, brainstorming, or cultural evolution.

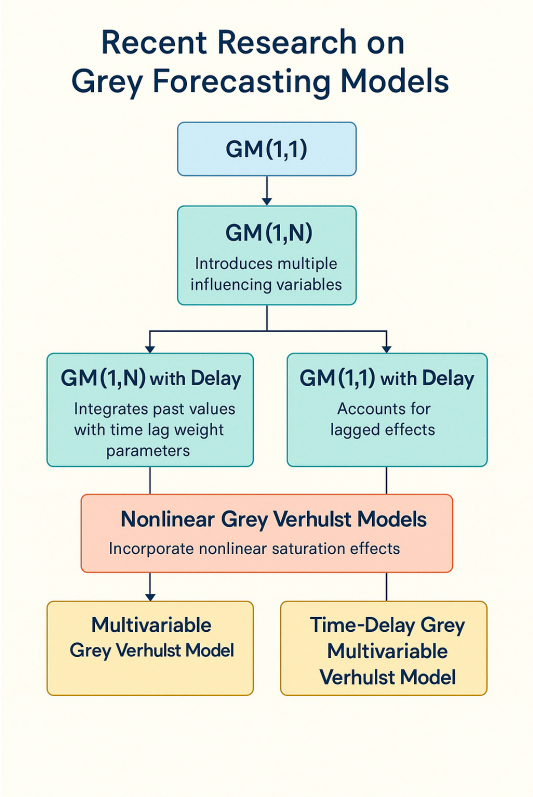


Figure 1. Taxonomy of Grey Models Used in Energy Forecasting

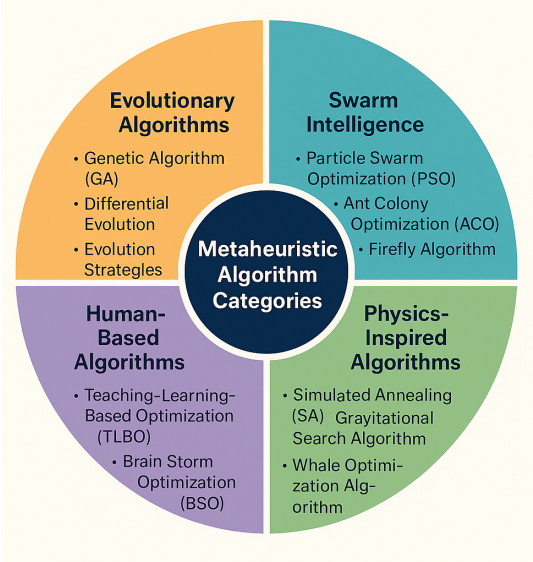


Figure 2. Classification of Metaheuristic Algorithms

Figure 2 illustrates the taxonomy of widely used metaheuristic algorithms.

Recently, metaheuristic algorithms have become popular in energy system forecasting because they are both strong and flexible. In grey system models, they help tune important settings such as development factors, time delay weights, and control values. Older methods for choosing these numbers like least squares fitting often fail when the search space is uneven, noisy, or when the settings affect one another.

These hybrid models combine the interpretability and small-sample efficiency of grey models with the adaptive search power of

metaheuristic methods, resulting in systems that are both accurate and robust under uncertainty. This integration forms the foundation for our proposed MARL-WOA method, which further leverages the adaptability of multi-agent reinforcement learning to guide the optimization process.

2.4 Multi-Agent Reinforcement Learning

As discussed in the previous section, integrating metaheuristic optimization techniques with grey forecasting models significantly improves prediction performance by enabling more adaptive and robust parameter tuning. While classical metaheuristic algorithms like PSO or GA operate based on fixed search strategies, recent research has explored the use of learning-based methods particularly Reinforcement Learning (RL) to dynamically guide the search process.

Recent advancements in optimization techniques have seen the emergence of hybrid models that combine metaheuristic algorithms with RL to tackle complex optimization problems more effectively. These hybrid approaches aim to leverage the global search capabilities of metaheuristic methods and the adaptive learning features of RL.

One notable example is the Reinforced Hybrid Genetic Algorithm (RHGA) (Figure 3) proposed by Zheng et al. [24], which integrates Q-learning with a genetic algorithm to solve the Traveling Salesman Problem (TSP). The Q-learning component guides the selection of genetic operators, enhancing the algorithm's ability to escape local optima and improving convergence speed.

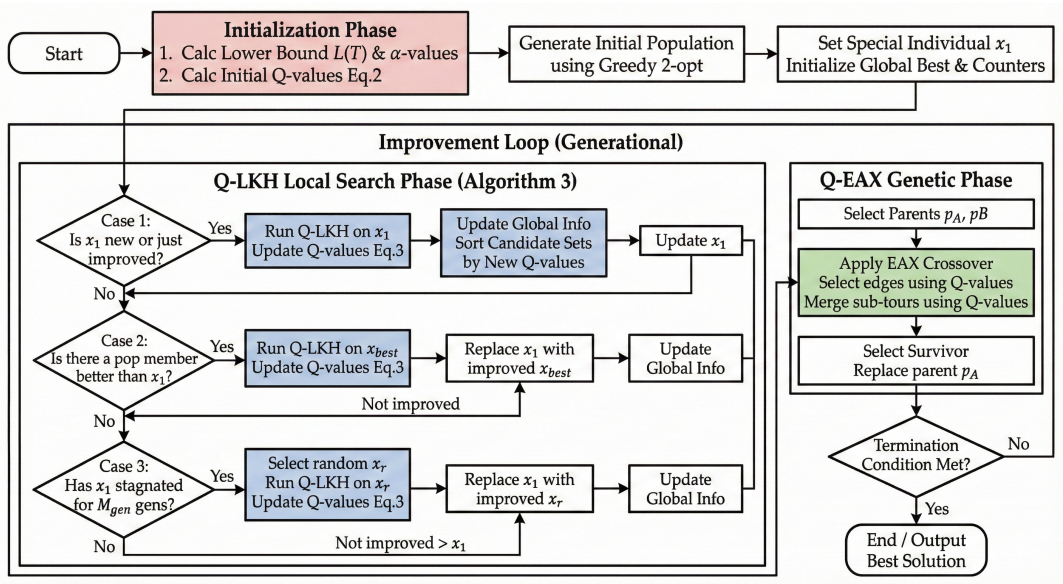


Figure 3. Reinforced Hybrid Genetic Algorithm (RHGA)

Similarly, the Tribal Intelligent Evolution Optimization (TIEO) algorithm (Figure 4) introduced by Kiani et al. [25] combines reinforcement learning with metaheuristic strategies to solve global optimization problems. TIEO employs a tribal-based structure where each tribe represents a potential solution, and RL is used to adaptively adjust the search strategies based on performance feedback. Recent studies have explored reinforcement-learningdriven ensemble frameworks that dynamically adapt model selection to fluctuating wind and solar conditions, substantially improving renewable energy forecasting performance [27]. More broadly, reinforcement learning has also been applied to general time-series ensemble forecasting, providing an adaptive mechanism for determining model combination weights and achieving notable accuracy gains across diverse forecasting tasks [28].

These hybrid models demonstrate the potential of combining metaheuristic algorithms with reinforcement learning to enhance the performance of optimization tasks. By incorporating adaptive learning mechanisms, these approaches can dynamically adjust their search strategies, leading to more efficient and accurate solutions.

Building upon this, MARL extends RL by involving multiple interacting agents, each capable of learning and adapting based on its environment and the behaviors of other agents. In the context of optimization, MARL can be leveraged to distribute the search process among agents, each exploring different regions of the solution space and sharing knowledge through coordinated policies. This

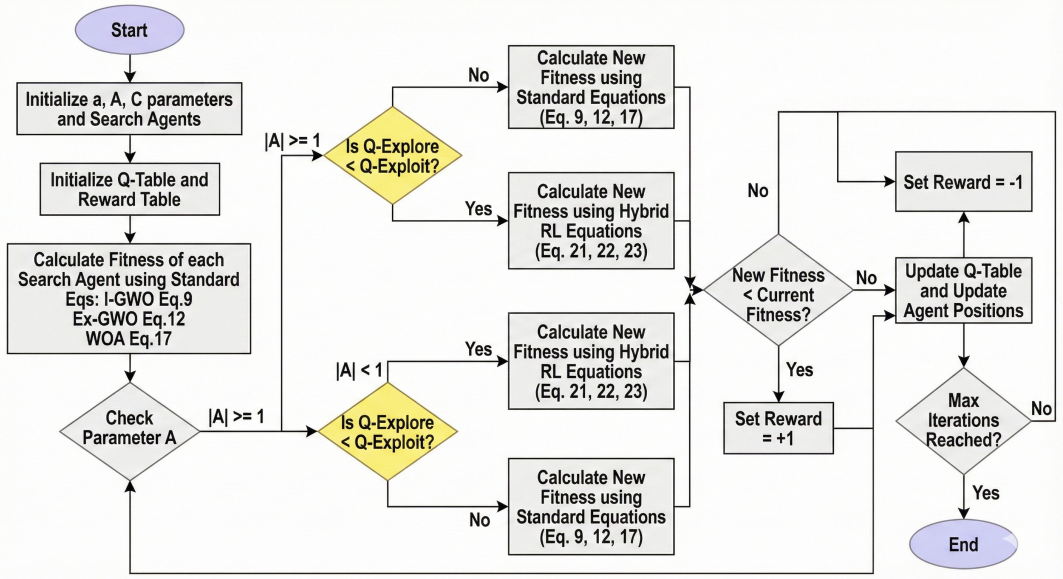


Figure 4. Tribal Intelligent Evolution Optimization (TIEO) algorithm

decentralized learning structure promotes exploration diversity and adaptability two features especially valuable in high-dimensional or dynamic optimization landscapes [5].

3 Two Grey Model Optimization Approaches

3.1 Grey Multivariable Verhulst Model

Definition 1. *Assume that* $X_1^{(0)} = (x_1^{(0)}(1), x_1^{(0)}(2), \ldots, x_1^{(0)}(n))$ *is the sequence of dependent variables and independent sequences of variables are* $X_i^{(0)} = (x_i^{(0)}(1), x_i^{(0)}(2), \ldots, x_i^{(0)}(n))$, *where* $i = 2, 3, \ldots, m$. *The definition of the first-order accumulation of the observed data is as in Eq.* (5)

$$X^{(1)} = (X_1^{(1)}, X_2^{(1)}, \ldots, X_m^{(1)}(n)) \tag{1}$$

where,

$$X_i^{(1)} = (x_i^{(1)}(1), x_i^{(1)}(2), \ldots, x_i^{(1)}(n)), \qquad i = 1, 2, \ldots, m. \tag{2}$$

and

$$x_i^{(1)}(k) = \sum_{i=0}^{k} x_i^{(0)}(j), \qquad k = 1, 2, \ldots, n. \tag{3}$$

Definition 2. *Let* $X_1^{(0)}, X_i^{(0)}$ *and* $X^{(1)}$ *be expressed as Definition 1, The Grey Multi variable Verhulst Model, denoted as* $(GMVM(1,N))$,*is then described as follows:*

$$x_1^{(0)}(k) + \beta_1 z_1^{(1)}(k) = \sum_{i=2}^{m} \beta_i (x_i^{(1)}(k))^2 + u. \tag{4}$$

Here u is a grey control parameter and $z_1^{(1)}(k)$ *is the adjoining mean generated sequence of* $x_1^{(1)}(k)$

$$Z_1^{(1)} = (z_1^{(1)}(1), z_1^{(1)}(2), \ldots, z_1^{(1)}(n)) \tag{5}$$

where

$$z_1^{(1)}(k) = 0.5(x_1^{(1)}(k) + x_1^{(1)}(k-1)), \qquad (k = 2, 3, \ldots, n) \tag{6}$$

and $z_1^{(1)}(1) = x_1^{(1)}(1)$.

The expression for the differential equation of the grey multivariable Verhulst (GMVM$(1,N)$*) model is as in Eq. (7)*

$$\frac{dx_1^{(1)}(k)}{dk} + \beta_1 x_1^{(1)}(k) = \sum_{i=2}^{m} \beta_i (x_i^{(1)}(k))^2 + u, \qquad (i = 2,3,\ldots,n) \tag{7}$$

while β_1, β_i, u *are real numbers. In Eq. (7), the parameter* β_1 *is a development coefficient,* $\beta_i (x_i^{(1)}(k))^2$ *and* β_i *are, respectively, the driving terms and driving coefficients; besides u is the grey control parameter.*

Theorem 1. *[1] The parameter vector* $(\beta_1, \beta_i, u)^T, (i = 2,3,\ldots,n)$ *in (GMVM*$(1,N)$*) can be estimated by applying the least squares method (LSM) provided that the matrix* $B^T B$ *is invertible Eq. (8). That is,*

$$(\beta_1, \beta_i, u)^T = (B^T B)^{-1} B^T Y \tag{8}$$

Where the matrices B and Y are defined as in Eqs.(9) and (10) respectively;

$$B = \begin{pmatrix} -z_1^{(1)}(2) & \left(x_2^{(1)}(2)\right)^2 & \cdots & \left(x_m^{(1)}(2)\right)^2 & 1 \\ -z_1^{(1)}(3) & \left(x_2^{(1)}(3)\right)^2 & \cdots & \left(x_m^{(1)}(3)\right)^2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -z_1^{(1)}(n) & \left(x_2^{(1)}(n)\right)^2 & \cdots & \left(x_m^{(1)}(n)\right)^2 & 1 \end{pmatrix}, \tag{9}$$

$$Y = \begin{pmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(n) \end{pmatrix}. \tag{10}$$

3.2 Time-Delay Grey Multivariable Verhulst Model

In this subsection, we introduce our proposed time-lag weight parameter multivariable Verhulst model, which is a novel extension of the grey multivariable Verhulst model. This approach incorporates time-lag weight parameters to delayed interactions among variables and offers a fresh perspective for modeling complex systems.

Definition 3. *Assume* $X_1^{(0)}, X_i^{(0)}$ *and* $Z_1^{(1)}(k)$ *are defined in Definition 1. The proposed model can be formulated using Eqs. (11) and (12). The improved time-delay grey multivariable Verhulst model with one order and N variables, shorted for (ITGM*$(1,N)$*), can be presented as*

$$x_1^{(0)}(k) + \beta_1 z_1^{(1)}(k) = \sum_{i=2}^{m} \sum_{j=1}^{k} \beta_i \lambda_i^{(k-j)} (x_i^{(1)}(j))^2 + u, \qquad k = 2,\ldots,n. \tag{11}$$

The corresponding whitening differential equation is presented below:

$$\frac{dx_1^{(1)}(k)}{dk} + \beta_1 x_1^{(1)}(k) = \left(\sum_{i=2}^{m} \int_1^k \beta_i \lambda_i^{(k-s)} (x_i^{(1)}(s))^2 ds\right) + u \tag{12}$$

Theorem 2. *Suppose* $X_1^{(0)}, X_i^{(0)}$ *and* $Z_1^{(1)}(k)$ *are the same as defined in Definition 1 and 2. Once* λ_i *are given, the procedure to find the parameter vector* $\theta = (\beta_1, \beta_i, u)^T, i = 2,3,\ldots,m$ *in ITGMVM*$(1,N)$ *by LSM method is as follows:*

1. *When* $m+1 = n-1$, $|B| \neq 0$, *then* $\theta = B^{-1}Y$.
2. *When* $m+1 > n-1$, $|BB^T| \neq 0$, *then* $\theta = B^T(BB^T)^{-1}Y$.
3. *When* $m+1 < n-1$, $|B^T B| \neq 0$, *then* $\theta = (B^T B)^{-1} B^T Y$.

where

$$B=\begin{pmatrix} -z_1^{(1)}(2) & \sum_{j=1}^{2}\lambda_2^{2-j}(x_2^{(1)}(j))^2 & \cdots & \sum_{j=1}^{2}\lambda_m^{2-j}(x_m^{(1)}(j))^2 & 1 \\ -z_1^{(1)}(3) & \sum_{j=1}^{3}\lambda_2^{3-j}(x_2^{(1)}(j))^2 & \cdots & \sum_{j=1}^{3}\lambda_m^{3-j}(x_m^{(1)}(j))^2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -z_1^{(1)}(n) & \sum_{j=1}^{n}\lambda_2^{n-j}(x_2^{(1)}(j))^2 & \cdots & \sum_{j=1}^{n}\lambda_m^{n-j}(x_m^{(1)}(j))^2 & 1 \end{pmatrix} \tag{13}$$

and

$$Y=\begin{pmatrix} x_1^{(1)}(2) \\ x_1^{(1)}(3) \\ \vdots \\ x_1^{(1)}(n) \end{pmatrix}.$$

Proof. By substituting $k=2,3,\ldots,n$ into Eq. (11), we obtain

$$\begin{cases} x_1^{(0)}(2)=-\beta_1 z_1^{(1)}(2)+\sum\limits_{i=2}^{m}\sum\limits_{j=1}^{2}\beta_i\lambda_i^{2-j}(x_i^{(1)}(j))^2+u \\ x_1^{(0)}(3)=-\beta_1 z_1^{(1)}(3)+\sum\limits_{i=2}^{m}\sum\limits_{j=1}^{3}\beta_i\lambda_i^{3-j}(x_i^{(1)}(j))^2+u \\ \quad\vdots \\ x_1^{(0)}(n)=-\beta_1 z_1^{(1)}(n)+\sum\limits_{i=2}^{m}\sum\limits_{j=1}^{n}\beta_i\lambda_i^{n-j}(x_i^{(1)}(j))^2+u \end{cases} \tag{14}$$

In terms of $\hat{\theta}=(\hat{\beta}_1,\hat{\beta}_i,\hat{u})^T, i=2,3,\ldots,m$, fitted value can be expressed by

$$\hat{Y}=B\hat{\theta}. \tag{15}$$

When $m+1<n-1$. Accordingly, the error sequence can be elaborated as

$$\varepsilon=Y-\hat{Y}=Y-B\hat{\theta} \tag{16}$$

The Sum of Squares for Error (SSE) is employed in this Section. Denote

$$SSE=\varepsilon^T\varepsilon=(Y-B\hat{\theta})^T(Y-B\hat{\theta})=\sum_{k=2}^{n}[x_1^{(0)}(k)+\hat{\beta}_1 z_1^{(1)}(k)-\sum_{i=2}^{m}\sum_{j=1}^{k}\hat{\beta}_i\lambda_i^{(k-j)}(x_i^{(1)}(j))^2-\hat{u}]^2 \tag{17}$$

To minimize SSE, we take partial derivation with respect to each parameter. Thus, $\hat{\theta}=(\hat{\beta}_1,\hat{\beta}_2,\ldots,\hat{\beta_m},\hat{u})^T$ should satisfy

$$\begin{cases} \frac{\partial SSE}{\partial\hat{\beta}_1}=2\sum_{k=2}^{n}\left[x_1^{(0)}(k)+\hat{\beta}_1 z_1^{(1)}(k)-\sum_{i=2}^{m}\sum_{j=1}^{k}\hat{\beta}_i\lambda_i^{(k-j)}(x_i^{(1)}(j))^2-\hat{u}\right]z_1^{(1)}(k)=0 \\ \frac{\partial SSE}{\partial\hat{\beta}_2}=-2\sum_{k=2}^{n}\left[x_1^{(0)}(k)+\hat{\beta}_1 z_1^{(1)}(k)-\sum_{i=2}^{m}\sum_{j=1}^{k}\hat{\beta}_i\lambda_i^{(k-j)}(x_i^{(1)}(j))^2-\hat{u}\right]\sum_{j=1}^{k}\lambda_2^{k-j}(x_2^{(1)}(j))^2=0 \\ \vdots \\ \frac{\partial SSE}{\partial\hat{u}}=-2\sum_{k=2}^{n}\left[x_1^{(0)}(k)+\hat{\beta}_1 z_1^{(1)}(k)-\sum_{i=2}^{m}\sum_{j=1}^{k}\hat{\beta}_i\lambda_i^{(k-j)}(x_i^{(1)}(j))^2-\hat{u}\right]=0 \end{cases} \tag{18}$$

Namely,

$$\begin{cases} \sum_{k=2}^{n}\left[x_1^{(0)}(k)+\hat{\beta}_1 z_1^{(1)}(k)-\sum_{i=2}^{m}\sum_{j=1}^{k}\hat{\beta}_i\lambda_i^{(k-j)}(x_i^{(1)}(j))^2-\hat{u}\right]z_1^{(1)}(k)=0 \\ \sum_{k=2}^{n}\left[x_1^{(0)}(k)+\hat{\beta}_1 z_1^{(1)}(k)-\sum_{i=2}^{m}\sum_{j=1}^{k}\hat{\beta}_i\lambda_i^{(k-j)}(x_i^{(1)}(j))^2-\hat{u}\right]\sum_{j=1}^{k}\lambda_2^{k-j}(x_2^{(1)}(j))^2=0 \\ \vdots \\ \sum_{k=2}^{n}\left[x_1^{(0)}(k)+\hat{\beta}_1 z_1^{(1)}(k)-\sum_{i=2}^{m}\sum_{j=1}^{k}\hat{\beta}_i\lambda_i^{(k-j)}(x_i^{(1)}(j))^2-\hat{u}\right]=0 \end{cases} \tag{19}$$

Eq. (19) can be expressed in matrix form as

$$B^T\varepsilon = B^T(Y-B\hat{\theta})=0 \tag{20}$$

This implies that

$$B^TY = B^TB\hat{\theta}. \tag{21}$$

Clearly, by solving the Eq. (21), we obtain

$$\hat{\theta}=(B^TB)^{-1}B^TY, \tag{22}$$

where the matrix B^TB is non-singular, meaning that $|B^TB|\neq 0$.

In this case, the rank of B is m, indicating that the column vectors of matrix B are linearly independent.

When $m+1>n-1$, the full-rank matrix B can be decomposed as

$$B=DC. \tag{23}$$

Assuming B^+ is the generalized inverse of B, it can be computed as

$$B^+=C^T(CC^T)^{-1}(D^TD)^{-1}D^T. \tag{24}$$

Thus,

$$\hat{\theta}=B^+Y=C^T(CC^T)^{-1}(D^TD)^{-1}D^TY. \tag{25}$$

Since $D=I_{n-1}C$ (where I_{n-1} is the identity matrix), we have $B=C$, leading to

$$\hat{\theta}=B^T(BB^T)^{-1}Y. \tag{26}$$

Again, B^TB is non-singular, meaning $|B^TB|\neq 0$.

When $m+1=n-1$.

According to the above, we can expand Eq. (20) or Eq. (21). Taking Eq. (21) as an example for explanation, we have

$$\hat{\theta}=(B^TB)^{-1}B^TY=B^{-1}(B^T)^{-1}B^TY=B^{-1}Y. \tag{27}$$

Here, the matrix B is non-singular, meaning that $|B|\neq 0$.

In summary, the Ordinary Least Squares (OLS) method can be used to obtain the parameter estimates. By substituting these estimated parameters, we can leverage the convolution integral to solve the model. □

The parameter vector $\hat{\theta}=(\beta_1,\beta_i,u)^T$ is provided in Theorem 2 which also outlines the procedure for estimating the parameters of $ITGMVM(1,N)$. Once all parameter estimates are determined, the corresponding solution to the $ITGMVM(1,N)$ is derived, as presented in Theorem 3.

Theorem 3. *Assuming that B, Y, and $\hat{\theta}$ are as defined in Theorem 2, then the time response function for the forecasting sequence can be expressed as*

$$\begin{cases} \hat{x}_1^{(1)}(k)=x_1^{(0)}(1)e^{-\beta_1(k-1)}+\sum_{\eta=2}^{k}\left\{e^{-\beta_1\left(k-\eta+\frac{1}{2}\right)}\frac{1}{2}\left[g(\eta)+g(\eta-1)\right]\right\}+\frac{1}{\beta_1}u(1-e^{-\beta_1(k-1)}), & \beta_1\neq 0,\quad k=2,3,\ldots,n. \\ \hat{x}_1^{(1)}(1)=x_1^{(0)}(1), & \text{as the initial value.} \end{cases} \tag{28}$$

where the function $g(\eta)$ is defined as

$$g(\eta)=\frac{1}{2}\sum_{i=2}^{m}\sum_{\varpi=2}^{\eta}\beta_i\left\{\lambda_i^{\eta-\varpi+\frac{1}{2}}\left[(x_i^{(1)}(\varpi))^2+(x_i^{(1)}(\varpi-1))^2\right]\right\}, \tag{29}$$

for $\eta\geq 2$. Additionally, for $\eta=1$ we have $g(1)=0$.

Proof. To solve the Eq. (12), we set the right side of the equation equal to 0, then we have

$$\frac{dx_1^{(1)}(k)}{dk}+\beta_1 x_1^{(1)}(k)=0. \tag{30}$$

After solving the equation, we have

$$x_1^{(1)}(k)=e^{-\beta_1 k}. \tag{31}$$

On the other hand, if we consider $k\geq 2$ based on the Gaussian rule [2], the integral part of the Eq. (12) can be transformed into

$$\int_1^k \beta_i\lambda_i^{(k-s)}(x_i^{(1)}(s))^2 ds)=\frac{1}{2}\sum_{\varpi=2}^{k}\beta_i\left\{\lambda_i^{k-\varpi+\frac{1}{2}}\left[(x_i^{(1)}(\varpi))^2+(x_i^{(1)}(\varpi-1))^2\right]\right\}, \tag{32}$$

Thus, we can express the Eq. (12) as follow

$$\frac{dx_1^{(1)}(k)}{dk}+\beta_1 x_1^{(1)}(k)=\frac{1}{2}\sum_{i=2}^{m}\sum_{\varpi=2}^{k}\beta_i\left\{\lambda_i^{\eta-\varpi+\frac{1}{2}}\left[(x_i^{(1)}(\varpi))^2+(x_i^{(1)}(\varpi-1))^2\right]\right\}+u \tag{33}$$

Obviously, the response function will be solved by Eq. (33), which will be represented as

$$\hat{x}_1^{(1)}(k)=x_1^{(0)}(1)e^{-\beta_1(k-1)}+\int_1^k g(\eta)e^{-\beta_1(k-1)}+u \tag{34}$$

where

$$g(\eta)=\frac{1}{2}\sum_{i=2}^{m}\sum_{\varpi=2}^{\eta}\beta_i\left\{\lambda_i^{\eta-\varpi+\frac{1}{2}}\left[(x_i^{(1)}(\varpi))^2+(x_i^{(1)}(\varpi-1))^2\right]\right\}, \tag{35}$$

The Gaussian rule is originally employed for discretizing the convolution integral in Eq. (34). Subsequently, the time response function is determined by the formula below:

$$\begin{cases}\hat{x}_1^{(1)}(k)=x_1^{(0)}(1)e^{-\beta_1(k-1)}+\sum_{\eta=2}^{k}\left\{e^{-\beta_1\left(k-\eta+\frac{1}{2}\right)}\frac{1}{2}\left[g(\eta)+g(\eta-1)\right]\right\}+\frac{1}{\beta_1}u(1-e^{-\beta_1(k-1)}), & \beta_1\neq 0,\quad k=2,3,\ldots,n.\\ \hat{x}_1^{(1)}(1)=x_1^{(0)}(1), & \text{as the initial value.}\end{cases} \tag{36}$$

□

Fainally, using the first-degree inverse accumulating generation operation (1-IAGO), reduced values can be derived. In general, the predictive formula with unknown coefficients λ_j (where $j=2,3,\ldots,m$) can be obtained from the previous expressions. For more accurate predictions, it is necessary to estimate these unknown coefficients λ_j (where $j=2,3,\ldots,m$).

3.3 Two Optimization Problems

In this subsection, two distinct strategies are explored for parameter estimation in the proposed $ITGMVM(1,N)$ model. The first adopts a sequential procedure: the delay parameters λ_j are initially estimated using an optimization algorithm, after which the remaining parameters are determined via LSE. The second strategy formulates parameter estimation as a constrained nonlinear optimization problem, where all parameters including delay factors, model coefficients, and control variables are estimated simultaneously by minimizing a MAPE-based objective function. This integrated approach enables the model to capture parameter interdependencies more effectively, leading to improved accuracy and robustness. It is particularly suitable for nonlinear systems and scenarios involving uncertain data or limited prior information.

i. **Partial Parameter Estimation (PPE):** This approach concentrates on estimating the delay parameters λ_j (for $j = 2,3,\ldots,m$) using an advanced optimization algorithm. Once the optimal delays are identified, the remaining model parameters including the coefficients β_i and the control parameter u are computed via the LSE method. The estimation process is guided by a well-defined objective function and subject to a set of problem-specific constraints.

$$\min g(\lambda_2,\lambda_3,\ldots,\lambda_m) = \frac{1}{n-1}\sum_{k=2}^{n}\left|\frac{\hat{x}^{(0)}(k)-x^{(0)}(k)}{x^{(0)}(k)}\right| \times 100\% \tag{37}$$

s.t.

$$\begin{cases}
(\beta_2,\beta_3,\ldots,\beta_m,u)^T = (B^TB)^{-1}B^TY \\
\lambda = [\lambda_2,\lambda_3,\ldots,\lambda_m], \quad 0<\lambda_i<1 \\
\beta_i \in (a_i,b_i) \\
u \in U \\
\beta_1 \neq 0 \\
\hat{x}_1^{(1)}(k) = x_1^{(0)}(1)e^{-\beta_1(k-1)} + \sum_{\eta=2}^{k}\left\{e^{-\beta_1\left(k-\eta+\frac{1}{2}\right)}\frac{1}{2}\left[g(\eta)+g(\eta-1)\right]\right\} + \frac{1}{\beta_1}u(1-e^{-\beta_1(k-1)}), \\
g(\eta) = \frac{1}{2}\sum_{i=2}^{m}\sum_{\varpi=2}^{\eta}\beta_i\left\{\lambda_i^{\eta-\varpi+\frac{1}{2}}\left[(x_i^{(1)}(\varpi))^2+(x_i^{(1)}(\varpi-1))^2\right]\right\}, \\
\hat{x}_1^{(1)}(1) = x_1^{(0)}(1), \\
x_i^{(0)}(k) = x_i^{(1)}(k) - x_i^{(1)}(k-1), \\
k = 2,3,\ldots,n, \quad i = 2,3,\ldots,m.
\end{cases}$$

ii. **Full Parameter Estimation (FPE):** The second strategy aims to estimate all parameters of the $ITGMVM(1,N)$ model simultaneously, including the delay factors λ_i, the coefficients β_1 and β_i, and the control parameter u. In this integrated framework, all parameters are treated as optimization variables, and a MAPE-based objective function is minimized within a fully constrained nonlinear optimization setting. By employing the proposed optimization algorithm in this context, the approach effectively captures parameter interdependencies and yields a more robust and flexible modeling framework. This is particularly advantageous in situations where prior information about parameter values is unavailable. The complete formulation of the parameter estimation problem is presented as follows:

$$\min g(\lambda_i,\beta_1,\beta_i,u) = \frac{1}{n-1}\sum_{k=2}^{n}\left|\frac{\hat{x}^{(0)}(k)-x^{(0)}(k)}{x^{(0)}(k)}\right| \times 100\% \tag{38}$$

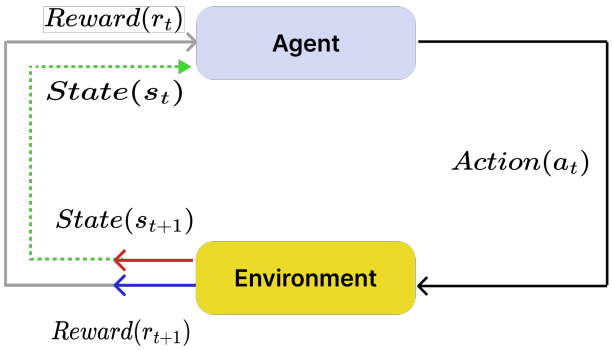
s.t

$$\begin{cases}
\lambda = [\lambda_2,\lambda_3,\ldots,\lambda_m], \quad 0<\lambda_i<1 \\
\beta_1 \neq 0 \\
\beta_i \in (a_i,b_i) \\
u \in U \\
\hat{x}_1^{(1)}(k) = x_1^{(0)}(1)e^{-\beta_1(k-1)} + \sum_{\eta=2}^{k}\left\{e^{-\beta_1\left(k-\eta+\frac{1}{2}\right)}\frac{1}{2}\left[g(\eta)+g(\eta-1)\right]\right\} + \frac{1}{\beta_1}u(1-e^{-\beta_1(k-1)}), \\
g(\eta) = \frac{1}{2}\sum_{i=2}^{m}\sum_{\varpi=2}^{\eta}\beta_i\left\{\lambda_i^{\eta-\varpi+\frac{1}{2}}\left[(x_i^{(1)}(\varpi))^2+(x_i^{(1)}(\varpi-1))^2\right]\right\}, \\
\hat{x}_1^{(1)}(1) = x_1^{(0)}(1), \\
x_i^{(0)}(k) = x_i^{(1)}(k) - x_i^{(1)}(k-1), \\
k = 2,3,\ldots,n, \quad i = 2,3,\ldots,m.
\end{cases}$$

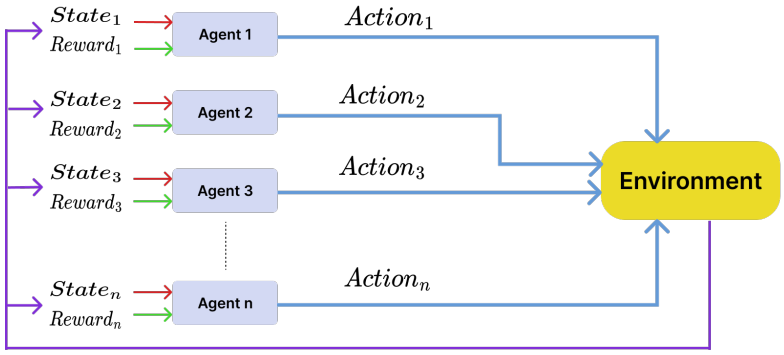
4 Methodology

In this section, we focus on improving a metaheuristic algorithm based on Multi-Agent Reinforcement Learning to solve two optimization problems (37) and (38).

4.1 Multi-Agent Reinforcement Learning



(a) Reinforcement Learning



(b) Multi-Agent Reinforcement Learning

Figure 5. Reinforcement Learning, Multi-Agent Reinforcement Learning

While traditional RL (see Fig. 5.a) focuses on training a single agent to interact with its environment, many real-world problems involve multiple interacting entities or agents [4]. MARL extends RL to scenarios involving multiple agents, each of which operates within a shared environment (Fig. 5.b). These agents can collaborate, compete, or coexist, making MARL a powerful paradigm for solving complex, decentralized, and multi-objective problems [5].

MARL methods can be broadly classified into cooperative, competitive, and mixed settings:

- **Cooperative MARL**: In cooperative settings, agents work together to achieve a common goal, often optimizing a shared reward signal. Algorithms such as *Centralized Training with Decentralized Execution (CTDE)* have gained prominence. CTDE trains agents collectively using global information but allows them to act independently during execution.

- **Competitive MARL**: In competitive environments, agents aim to maximize their own rewards, often at the expense of others. Strategies like *self-play* and *opponent modeling* enable agents to improve by interacting with and predicting the actions of adversaries.

In this work, we utilize a cooperative MARL approach, where agents collaborate to achieve the shared objective of optimizing the search process.

4.1.1 Multi-Agent Q-Learning in Meta-Heuristic Algorithms

One of the prominent methodologies in MARL is Multi-Agent Q-Learning (MAQL), an extension of the traditional Q-Learning algorithm [4]. In MAQL, multiple agents collaboratively update their Q-values based on their interactions with the shared environment, working together toward a common optimization objective. This cooperative framework is particularly well-suited for integration with meta-heuristic algorithms, which aim to solve complex optimization problems through iterative improvement strategies. By incorporating Multi-Agent Q-Learning into meta-heuristics, the adaptive learning mechanisms driven by agent-environment interactions can significantly enhance exploration and exploitation capabilities, ensuring robust and efficient optimization.

4.2 Proposed Method: MARL-WOA

MARL-WOA is a hybrid optimization algorithm that integrates **MARL** with the **WOA**(see Appendix A). MARL consists of three essential elements, described as

- **State** The state represents the current position of each whale in the search space, along with the ongoing iteration or phase (exploration or exploitation) of the optimization process.
- **Action** The action space includes two key actions: (1) exploration, where agents perform random position updates to discover new regions in the search space, and (2) exploitation, which involves refining known solutions using encircling or spiral movement strategies.
- **Reward** The reward function evaluates each agent's performance based on fitness improvement and solution diversity. Agents receive higher rewards for improving fitness or maintaining diversity, while penalties are applied for worsening solutions or clustering around suboptimal areas.

The main objective is to combine the global exploration abilities of WOA with the adaptive decision-making capabilities of MARL, improving the search process and ensuring more efficient local optimization.
The proposed framework utilizes **Multi-Agent Reinforcement Learning** to improve decision-making in aspects of the WOA algorithm that usually rely on randomness, such as choosing between exploration and exploitation, as well as determining agents' encircling or spiral movements. Each agent in the WOA swarm uses Q-learning [4], an RL-based technique, to learn optimal actions for exploration and exploitation over the course of iterations. The agents adapt their decision-making based on their interactions with the environment. WOA governs the movement behavior of the agents inspired by the bubble-net hunting strategy of humpback whales. The decision-making regarding the movement patterns (e.g., encircling or spiral) is guided by the learned Q-values.

4.2.1 Reward System

The reward system in the MARL-WOA framework is designed to guide agents toward improving fitness while maintaining diversity in the population. After each action, the reward for an agent is calculated based on two key factors: fitness improvement and diversity score. These factors ensure that agents are encouraged to explore new areas of the solution space while also refining existing solutions.

Diversity Score: The diversity score, D_i, for agent i is computed as the average distance between the agent's current position $\vec{X}_i$ and the positions of all other agents in the population:

$$D_i = \frac{1}{N-1} \sum_{j=1, j\neq i}^{N} \left\| \vec{X}_i - \vec{X}_j \right\|$$

This score measures how well the population is distributed across the search space, encouraging agents to avoid clustering too early in the optimization process.

Reward Calculation: The reward is determined by the agent's fitness improvement and its diversity score. Fitness improvement is calculated as:

$$\text{improvement} = f(\vec{X}_i(t)) - f(\vec{X}_i(t+1))$$

where $f(\vec{X}_i(t))$ and $f(\vec{X}_i(t+1))$ are the fitness values of the agent i in consecutive time steps.

The reward for agent i is then assigned based on the following rules:

1. When the agent achieves a positive fitness improvement, it is rewarded with the improvement value plus a small bonus proportional to its diversity score.

$$\text{improvement} + 0.1 \times D_i, \text{if improvement} > 0$$

2. When there is no improvement or a negative change in fitness, a penalty proportional to the magnitude of the loss is applied.

$$-0.5 \times |\text{improvement}|, \text{if improvement} \leq 0$$

So, the reward system is summarized as below

$$\text{reward} = \begin{cases} \text{improvement} + 0.1 \times D_i, & \text{if improvement} > 0 \\ -0.5 \times |\text{improvement}|, & \text{if improvement} \leq 0 \end{cases}$$

4.2.2 Epsilon-Greedy Strategy for Action Selection

The **epsilon-greedy strategy** is employed in MARL to balance exploration and exploitation (encircling and bubble-net attack) effectively during action selection. The strategy works as follows:

- With a probability of ε, the agent explores the action space by selecting a random action. This allows the algorithm to discover potentially better solutions and avoids premature convergence.
- With a probability of $1-\varepsilon$, the agent exploits (or choose bubble-net attack) the learned Q-values by choosing the action that maximizes its expected reward based on the current Q-table.
- Over time, the value of ε is decayed gradually to shift the focus from exploration to exploitation (from encircling to bubble-net attack) as the learning process progresses.

The action a at state s is selected as

$$a = \begin{cases} \text{random action} & \text{with probability } \varepsilon, \\ \arg\max_{a'} Q(s,a') & \text{with probability } 1-\varepsilon. \end{cases} \tag{39}$$

To manage the exploration-exploitation trade-off over time, ε is updated using the following decay formula:

$$\varepsilon_t = \varepsilon_{\min} + (\varepsilon_{\max} - \varepsilon_{\min}) \cdot e^{-\lambda t}, \tag{40}$$

where

- ε_t: The value of ε at time step t,
- $\varepsilon_{\min}$: The minimum value of ε, which ensures exploration never completely stops,
- $\varepsilon_{\max}$: The initial value of ε at the start of the learning process,
- λ: The decay rate, controlling how quickly ε decreases over time,
- t: The current time step or iteration.

This adaptive mechanism ensures a robust trade-off between exploration and exploitation, enhancing the overall efficiency of the MARL-WOA framework. The Q-learning update rule used for each agent is as follows:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[R + \gamma \max_a Q(s',a) - Q(s,a) \right], \tag{41}$$

where s represents the current state, a is the action taken, R is the reward, and γ is the discount factor that encourages long-term gains.

The integration of the WOA with MARL enhances WOA's adaptability and decision-making capabilities by leveraging learning-based mechanisms. In this approach, each whale in the population acts as an agent equipped with a Q-table, which serves as a memory to guide decisions during the optimization process. Initially, the whales are randomly distributed within the solution space, and their movements are driven by a balance between exploration (searching new areas) and exploitation (refining known solutions). This balance is managed through an epsilon-greedy strategy, where decisions to explore or exploit are influenced by the Q-values stored in the table, enabling whales to learn and adapt their behavior dynamically based on past interactions.

As the whales move, they perform either encircling or spiral movements, guided by Q-values that prioritize actions contributing to fitness improvement or search diversity. After each action, the reward is computed by evaluating changes in solution quality and diversity, providing feedback to update the Q-table using the Q-learning algorithm. Over time, this process allows the whales to refine their strategies, promoting more effective exploration and exploitation of the search space, see Algorithm **??** and Fig.6

Algorithm 1: Hybrid MARL-WOA

Input: Objective function f, population size (agent size) N, max iterations T, learning rate α, discount factor γ, exploration rate ε

Initialize population $\vec{X}$ randomly;

Initialize two Q-tables for each agent:

- $Q_1(s,a)$ for explore/exploit decision-making
- $Q_2(s,a)$ for encircling/bubble-net attack during exploitation

Set the global best position $\vec{X}^*$;

for $t = 1$ *to* T **do**

Calculate adaptive coefficient $a = 2 - 2t/T$;

for *each agent i in the agents* **do**

Observe current state s_i;

Select action a_{1i} from Q_1 using ε-greedy strategy (Equation 40):

- **Explore**: Perform exploration update (Equation 46)
- **Exploit**: Select action a_{2i} from Q_2 using ε-greedy strategy ((Equation 40)):
  - **Encircle prey**: Update position using encircling equation (Equation 42)
  - **Bubble-net attack**: Update position using bubble-net equation (Equation 45)

Evaluate fitness $f(\vec{X}_i)$;

$$D_i = \frac{1}{N-1} \sum_{j=1, j\neq i}^{N} \left\| \vec{X}_i - \vec{X}_j \right\|$$

$$\text{improvement} = f(\vec{X}_i(t)) - f(\vec{X}_i(t+1))$$

$$\text{reward} = \begin{cases} \text{improvement} + 0.1 \times D_i, & \text{if improvement} > 0 \\ -0.5 \times |\text{improvement}|, & \text{if improvement} \leq 0 \end{cases}$$

Update Q_1 and Q_2 using Q-learning update rules (Equation 41);

end for

Update the global best position $\vec{X}^*$ if a better solution is found;

end for

Output: Best solution $\vec{X}^*$ and its fitness $f(\vec{X}^*)$.

4.3 Result Analysis

The performance of the proposed MARL-WOA method is evaluated on a set of standard benchmark functions listed in Table 2. These functions are widely used in the metaheuristic literature for testing exploration and exploitation capabilities [8–11]. Each benchmark function represents a unique optimization challenge with varying levels of complexity, dimensionality, and search space characteristics.

Experimental Setup: The experiments were conducted with a population size of 50 agents and a maximum of 500 epochs for each run. To ensure statistical robustness, each benchmark function was tested 100 times independently. This approach is in line with standard practices for comparing optimization algorithms [8]. The optimization process aimed to minimize the objective value of each function, starting from randomly initialized positions within the defined search range. The global best solution and its corresponding fitness value were recorded for each run.

Evaluation Metrics: The key evaluation metrics for each benchmark function included **Mean Fitness**, **Standard Deviation**, and **Convergence Speed**. Such metrics are frequently employed to analyze both the *accuracy* and the *consistency* of metaheuristic algorithms [9]. Furthermore, visualizing the convergence curves helps illustrate how quickly the population approaches the optimum over time [10].

Table 2. Benchmark Functions

Func.	Benchmark Function	Formula	Dim	Range	Optimal
F_1	Sphere	$f(x) = \sum_{i=1}^{n} x_i^2$	30	$[-100, 100]$	0
F_2	Rosenbrock	$f(x) = \sum_{i=1}^{n-1} \left[100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2\right]$	30	$[-100, 100]$	0
F_3	Rastrigin	$f(x) = \sum_{i=1}^{n} \left[x_i^2 - 10\cos(2\pi x_i) + 10\right]$	30	$[-5.12, 5.12]$	0
F_4	Ackley	$f(x) = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + \exp(1)$	30	$[-32, 32]$	0

where N is the number of agents, D is the dimensionality of the problem, T is the number of iterations, and A denotes the number of actions in the Q-learning phase.

4.3.1 Advantages of the MARL-WOA Method

The proposed MARL-WOA method integrates reinforcement learning into the decision-making process of WOA, replacing random value selection with a learning-driven approach. By employing Q-learning, agents dynamically balance exploration and exploitation, progressively refining their actions based on accumulated experience. Parameters such as the exploration rate and the encircling control parameter are adaptively adjusted, ensuring stability and flexibility throughout the optimization process. Furthermore, the reward mechanism combines fitness improvement with diversity preservation, maintaining a well-distributed population of solutions, and significantly reducing the risk of stagnation in local optima. According to the results illustrated in Table 3 and Fig. 7.

This method is highly scalable and parallelizable, leveraging its MARL framework for efficient parallel processing, making it suitable for large-scale optimization problems. The collaborative dynamics of MARL allow agents to share learning experiences indirectly, enhancing the robustness and efficiency of the optimization process. By combining reinforcement learning's adaptive capabilities with WOA's population-based search, the algorithm achieves superior performance, demonstrating faster convergence rates, higher-quality solutions, and the ability to tackle complex, high-dimensional optimization challenges effectively.

5 Empirical Study

Australia's electricity market is intricate and highly organized, consisting of several interconnected regions, each with distinct market dynamics. The core of this system is the National Electricity Market (NEM), which covers the eastern and southern regions of the country, including Queensland, New South Wales (NSW), Victoria, and Tasmania. This study utilizes real market data from the NSW regional market within the NEM, specifically focusing on data from Sundays and Wednesdays. These days were chosen to capture both weekday

Table 3. Performance Comparison of Algorithms on Benchmark Functions

Func.	Metric	PSO	WOA	MARL-WOA
F_1	Mean	1.17e-23	2.49e-203	2.10e-250
	Std	0	0.0	0.0
	Execution Time(s)	0.19	0.21	0.17
F_2	Mean	33.41	26.29	3.04e-11
	Std	1.58	0.67	2.34e-11
	Execution Time(s)	0.22	0.35	0.29
F_3	Mean	48.26	49.53	5.68e-16
	Std	14.65	77.86	5.65e-15
	Execution Time(s)	0.19	0.33	0.29
F_4	Mean	7.63e-6	4.81e-15	4.44e-16
	Std	4.85e-4	5.02e-15	1.23e-15
	Execution Time(s)	0.22	0.32	0.31

Table 4. Computational Complexity Comparison of Optimization Algorithms

Algorithm	Relative Complexity (O-notation)
PSO	$O(N \times D \times T)$
WOA	$O(N \times D \times T)$
MARL-WOA (proposed)	$O(N \times D \times T) + O(N \times A)$

and weekend consumption patterns, reflecting the variations in electricity demand throughout the week, and offering valuable insights into the region's market dynamics.(see figure 8)

5.1 Performance Evaluation Framework

The proposed FPE-MARL-WOA model is rigorously evaluated against six advanced benchmarks: (1) the conventional $GM(1,N)$ [40] as a baseline multivariate grey model; (2) its enhanced variant $ITGM(1,N)$ [41] with time-varying adaptations; (3) the metaheuristic-driven $GVBM-FHO$ [42](Grey Vector-Based Model with Fire Hawk Optimizer); and four hybrid derivatives of the proposed framework itself *PPE-WOA*, *FPE-WOA*, *PPE-MARL-WOA*, and *FPE-MARL-WOA* designed to isolate the contributions of LSE, WOA, and MARL components. This comprehensive comparison validates the incremental advantages of the full hybrid architecture.

5.2 Performance Forecasting Assessment

The performance of the proposed model is evaluated using seven statistical indicators listed in Table 5. A model is considered highly accurate when the values of APE, MAPE, RMSPE, MSE, and MAE are low. In contrast, higher values of IA and R suggest better model performance. The Index of Agreement (IA) measures how closely the testing data align with the training data, with values ranging from 0 (complete disagreement) to 1 (perfect agreement). The R value, which ranges from -1 to $+1$, reflects the strength and direction of a linear relationship values near $+1$ indicate a strong positive correlation, values near -1 show a strong negative correlation, and values around 0 suggest no linear correlation. To enhance clarity, the highest-performing result for each metric is highlighted in bold.

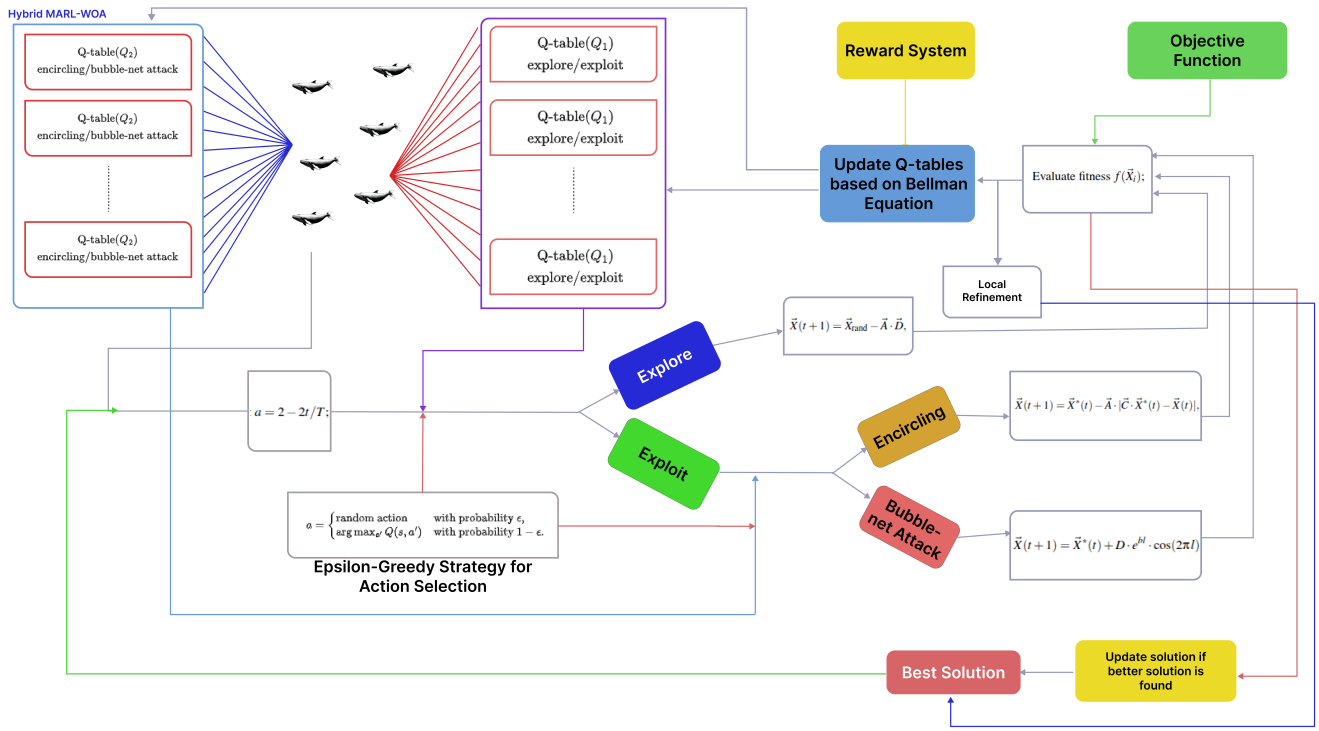


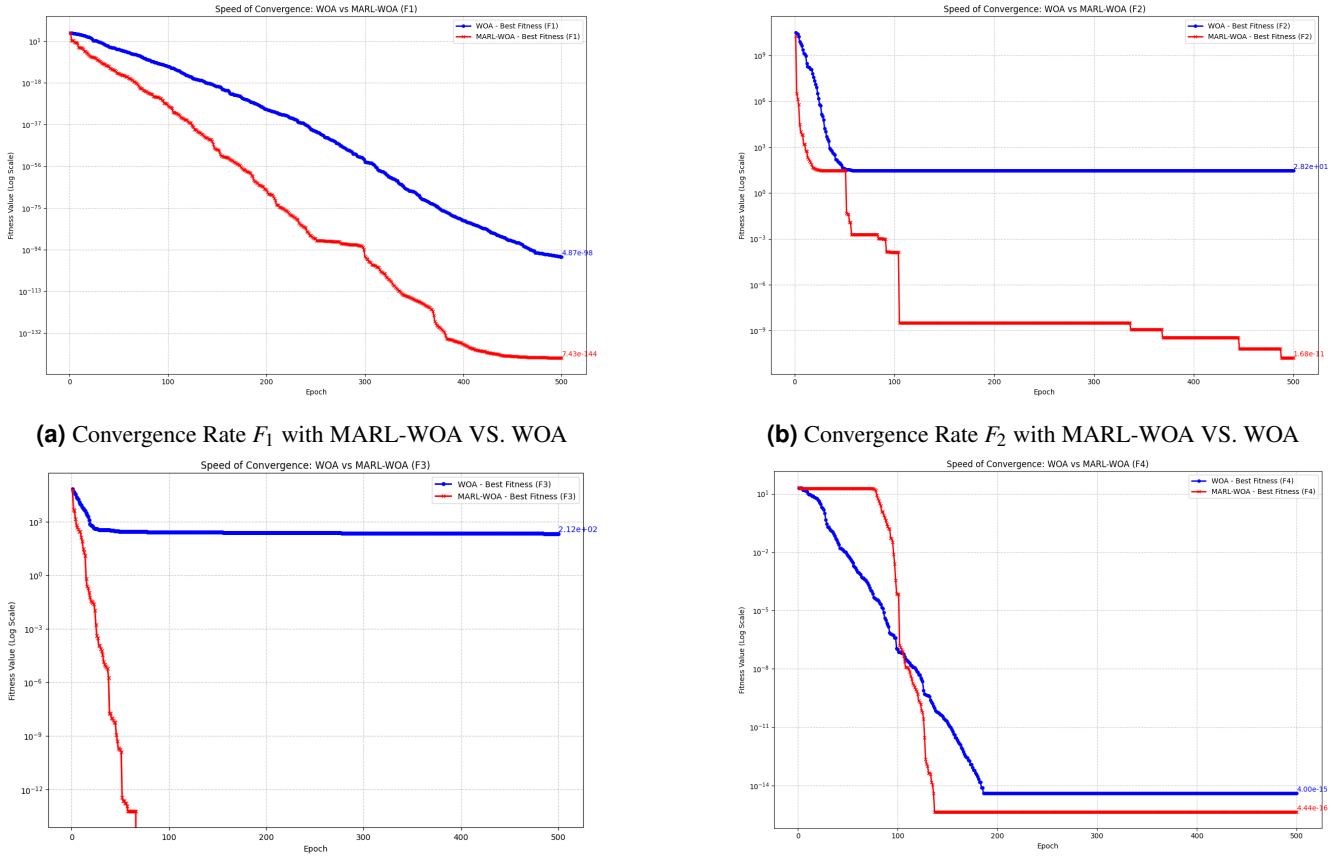
Figure 6. Hybrid MARL-WOA

Table 5. Equations to evaluate model.

Index	Formula
Absolute Percentage Error (APE)	$APE = \frac{\lvert\hat{x}^{(0)}(k) - x^{(0)}(k)\rvert}{x^{(0)}(k)} \times 100\%$
Mean Absolute Percentage Error (MAPE)	$MAPE = \frac{1}{n}\sum_{k=1}^{n} \frac{\lvert\hat{x}^{(0)}(k) - x^{(0)}(k)\rvert}{x^{(0)}(k)} \times 100\%$
Mean Absolute Error (MAE)	$MAE = \frac{1}{n}\sum_{k=1}^{n} \lvert\hat{x}^{(0)}(k) - x^{(0)}(k)\rvert$
Mean Squared Error (MSE)	$MSE = \frac{1}{n}\sum_{k=1}^{n} (\hat{x}^{(0)}(k) - x^{(0)}(k))^2$
Root Mean Square Percentage Error (RMSPE)	$RMSPE = \sqrt{\frac{1}{n}\sum_{k=1}^{n} (\frac{\hat{x}^{(0)}(k) - x^{(0)}(k)}{x^{(0)}(k)})^2} \times 100\%$
Index of Agreement (IA)	$IA = 1 - \frac{\sum_{k=1}^{n} (\hat{x}^{(0)}(k) - x^{(0)}(k))^2}{\sum_{k=1}^{n} (\lvert\hat{x}^{(0)}(k) - \bar{x}\rvert + \lvert\bar{x} - x^{(0)}(k)\rvert)^2}$
Correlation coefficient (R)	$R = \frac{Cov(x^{(0)}(k), \hat{x}^{(0)}(k))}{\sqrt{Var(x^{(0)}(k)) Var(\hat{x}^{(0)}(k))}}$

5.3 Experiment I: Sunday Forecast Comparison

The data set from the NSW electricity market includes weekly electricity prices on Sundays, covering the period from December 22, 2023, to March 30, 2024. To evaluate the forecasting models, the data is divided into training and testing sets. The training data includes the first 12 Sundays, from December 22, 2023, to March 9, 2024. This part is used to train each forecasting model and optimize parameters. The



(a) Convergence Rate F_1 with MARL-WOA VS. WOA

(b) Convergence Rate F_2 with MARL-WOA VS. WOA

(c) Convergence Rate F_3 with MARL-WOA VS. WOA

(d) Convergence Rate F_4 with MARL-WOA VS. WOA

Figure 7. Convergence Rates of MARL-WOA VS. WOA

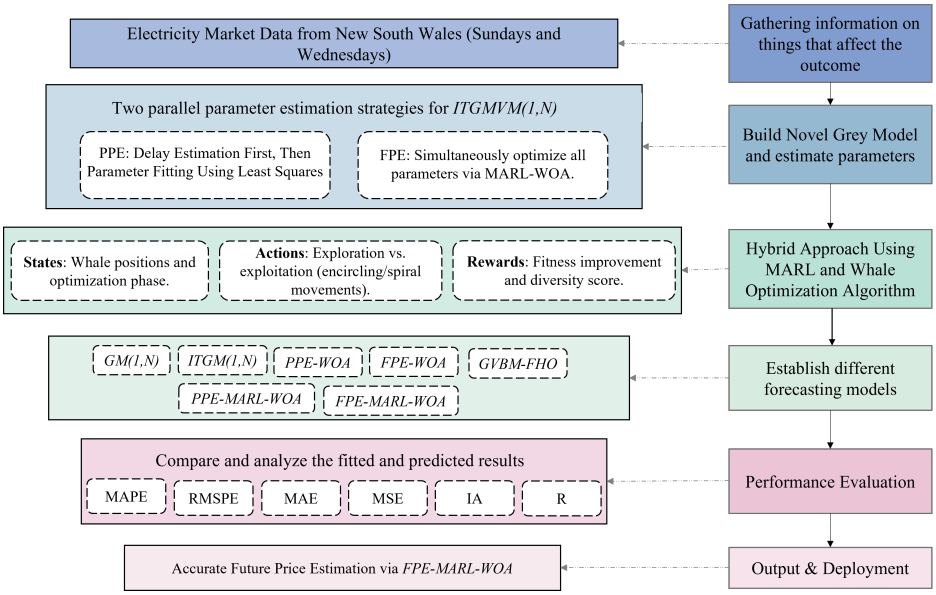


Figure 8. Overall framework of the proposed ITGMVM(1,N) electricity forecasting model, including data preparation, parameter estimation using PPE and FPE strategies, hybrid MARLWOA optimization, model construction, and performance evaluation.

remaining three Sundays March 16, March 23, and March 30, 2024 serve as the test data, used to evaluate how well each model generalizes to unseen data.

Among all the models, *FPE-MARL-WOA* achieves the best results in both the training and test phases. During training, it produces the lowest MAPE (1.66%), lowest RMSPE (0.231), and the highest correlation (R = 0.9983), indicating strong fitting ability. It also maintains superior accuracy in the test data, with the lowest error metrics compared to other models, such as $GM(1,N)$ and $ITGM(1,N)$, which show significantly higher errors (e.g., MAPE $>$ 46% in the test phase). This confirms that *FPE-MARL-WOA* not only fits well to historical data but also generalizes effectively to new data, making it the most reliable model for electricity price prediction in this case. The detailed numerical results are presented in Table 6, and the performance comparison across models is illustrated in Figures 9, 10, 11 and 12.

Remark 1. The analysis of the foregoing leads to the conclusion that the designed *FPE-MARL-WOA* model can achieve excellent electricity price forecasting performance by effectively integrating the strengths of the three member models grey modeling, metaheuristic optimization, and reinforcement learning. In particular, the model demonstrates outstanding accuracy in predicting weekend electricity prices, as evidenced by its superior performance on Sunday data. This suggests its suitability for applications where understanding weekend consumption dynamics is critical.

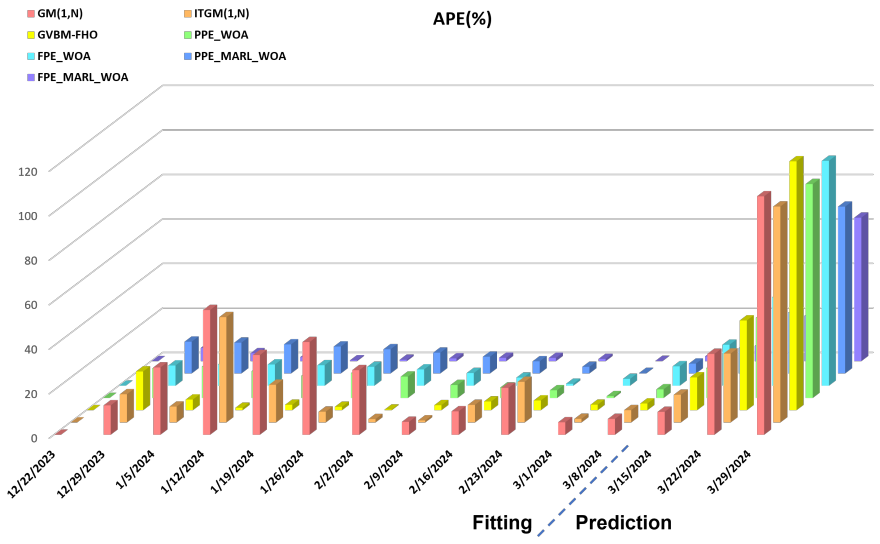


Figure 9. The Absolute Percentage Error (APE) calculated for seven grey forecasting models used to predict Experiment I, Table 6

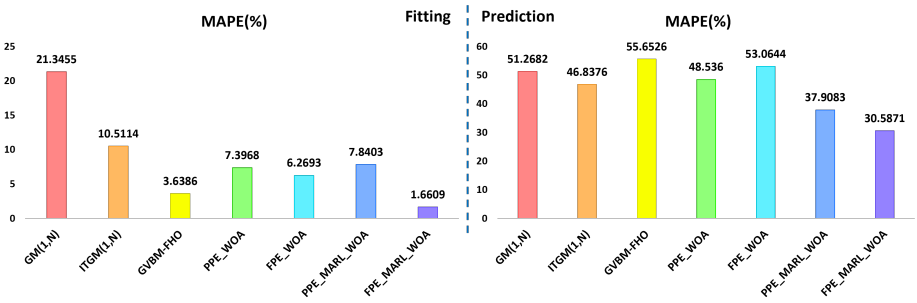


Figure 10. The Mean Absolute Percentage Error (MAPE) calculated for seven grey forecasting models used to predict Experiment I, Table 6

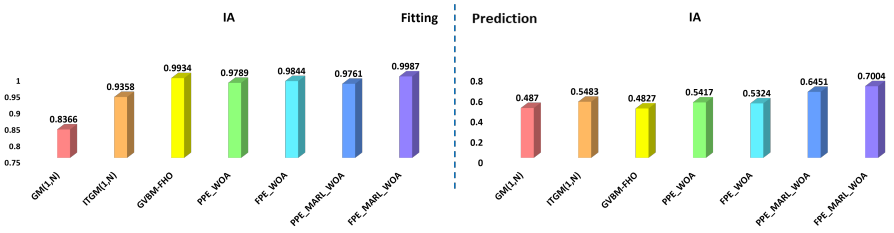


Figure 11. The Index of Agreement (IA) calculated for six grey forecasting models used to predict Experiment I, Table 6

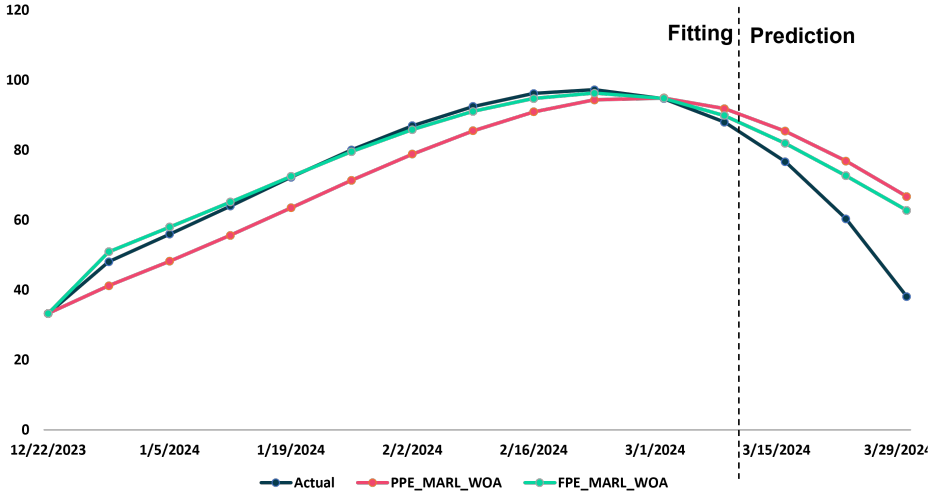


Figure 12. The simulated and predicted values of the novel models for Experiment I (Sunday Forecast Comparison)

5.4 Experiment II: Wednesday Forecast Comparison

The dataset from the NSW electricity market includes weekly electricity prices on Wednesdays, spanning from December 18, 2023, to March 26, 2024. To evaluate the forecasting models, the data is divided into training and testing sets. The training data includes the first 12 Wednesdays, from December 18, 2023, to March 19, 2024. This portion is used to train each forecasting model and optimize parameters. The remaining Wednesday March 26, 2024 serves as the test data, used to evaluate how well each model generalizes to unseen data.

Among all the models, *FPE-MARL-WOA* performs the best in the test phase. Although it shows relatively weaker results in the training phase, with higher MAPE and other error metrics, it demonstrates strong predictive power in the test data, achieving the lowest error metrics such as MAPE (5.2481%), low RMSPE (0.6291%), and high correlation (R = 0.9896). This indicates that while the model does not fit the training data as perfectly as some others, it is able to generalize effectively to unseen data. On the other hand, *GVBM-FHO* exhibits good performance during the training phase, achieving low MAPE and high correlation values. However, it struggles in the test data, with significantly higher errors, making it less reliable for predicting future prices. This comparison highlights that *FPE-MARL-WOA*, despite its weaker fit during training, is better suited for forecasting electricity prices in this case because of its superior generalization to unseen data. The detailed numerical results are presented in Table 7, and the performance comparison across models is illustrated in Figures 13, 14, 15 and 16

Remark 2. The results indicate that the *FPE-MARL-WOA* model also performs well for weekday electricity price forecasting, as shown by its strong generalization ability and low error metrics on Wednesday data.

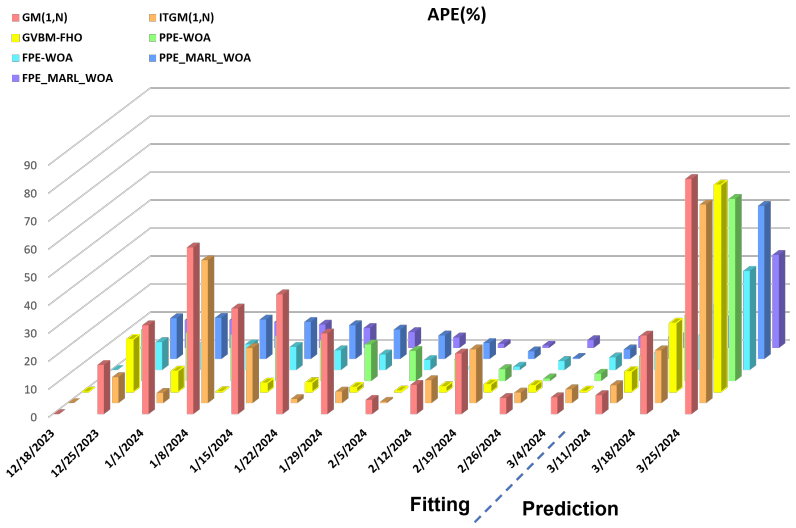


Figure 13. The Absolute Percentage Error (APE) calculated for seven grey forecasting models used to predict Experiment II, Table 7

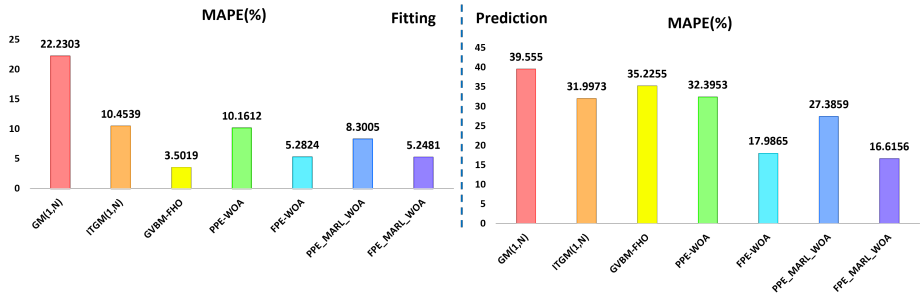


Figure 14. The Mean Absolute Percentage Error (MAPE) calculated for seven grey forecasting models used to predict Experiment II, Table 7

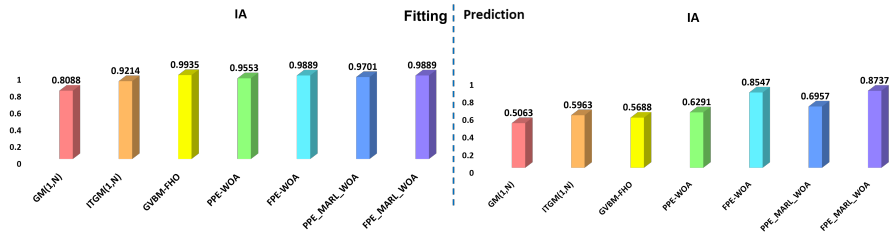


Figure 15. The Index of Agreement (IA) calculated for six grey forecasting models used to predict Experiment II, Table 7

6 Conclusion

In the current paper, we developed an advanced forecasting framework called the Improved Time-Delay Grey Multivariable Verhulst Model (ITGMVM) to address the challenges of predicting electricity prices in dynamic and data-scarce energy markets. By introducing time-delay parameters into the traditional grey Verhulst model, the proposed method effectively captured delayed interactions among multiple variables and better represented the nonlinear behavior observed in real electricity markets. To overcome the complexity of parameter estimation, two optimization strategies PPE and FPE were applied. Both were enhanced through a new hybrid algorithm that integrates Multi-Agent Reinforcement Learning with the Whale Optimization Algorithm (MARL-WOA). This integration improved the

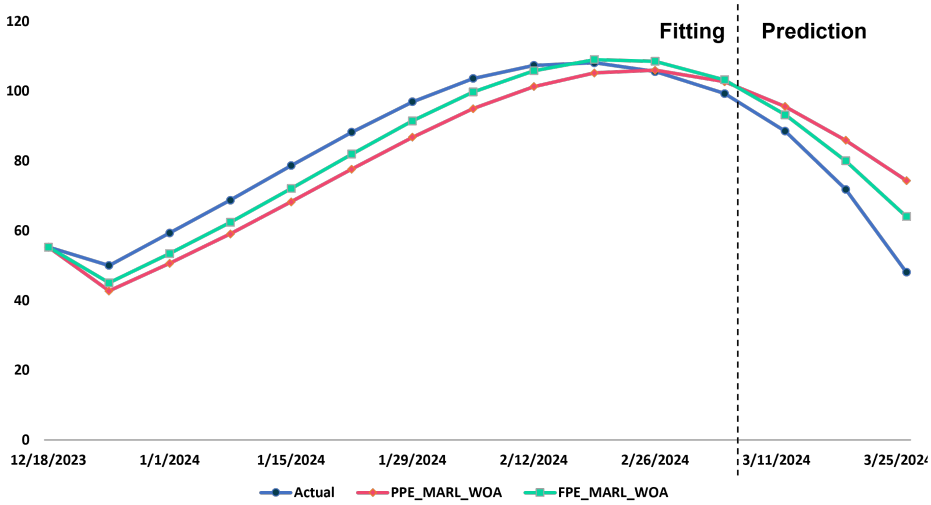


Figure 16. The simulated and predicted values of the novel models for Experiment II(Wednesday forecast Comparison)

Table 6. Sunday Comparative results of seven grey models

Date	Price	GM(1,N)		ITGM(1,N)		GVBM-FHO		PPE-WOA		FPE-WOA		PPE-MARL-WOA		FPE-MARL-WOA	
		Val	APE	Val	APE	Val	APE	Val	APE	Val	APE	Val	APE	Val	APE
12/22/2023	33.2876	33.2876	0	33.7318	0	33.2876	0	33.2876	0	33.2876	0	33.2876	0	33.2876	0
12/29/2023	48.0765	41.7256	13.21	41.9769	12.69	39.7064	17.41	41.7078	13.25	43.7063	9.09	41.2111	14.28	**50.9452**	5.97
1/5/2024	55.9675	72.8974	30.25	51.9717	7.14	53.2842	4.79	48.0805	14.09	50.6088	9.57	48.2188	13.85	**57.9687**	3.58
1/12/2024	63.9880	99.8579	56.06	94.2868	47.35	**64.694**	1.10	56.2797	12.05	57.8851	9.54	55.6321	13.06	65.1471	1.81
1/19/2024	72.2303	98.0989	35.81	84.4805	16.96	74.005	2.46	65.0194	9.98	65.5721	9.22	63.4963	12.09	**72.4993**	0.37
1/26/2024	80.0806	113.4416	41.66	76.1922	4.86	81.312	1.54	71.1481	11.15	73.2769	8.50	71.3823	10.86	**79.5586**	0.65
2/2/2024	86.9748	112.2188	29.02	88.3117	1.54	**86.7252**	0.29	78.6214	9.60	80.6067	7.32	78.8556	9.34	85.8771	1.26
2/9/2024	92.4672	97.837	5.81	**91.6094**	0.93	90.3753	2.26	87.0863	5.82	87.1968	5.70	85.5071	7.53	91.0635	1.52
2/16/2024	96.1963	106.3331	10.54	103.8064	7.91	92.4142	3.93	91.7052	4.67	92.6977	3.64	90.9378	5.47	**94.7663**	1.49
2/23/2024	97.2740	117.7635	21.06	115.0478	18.27	93.0122	4.38	93.91	3.46	**96.4338**	0.86	94.3794	2.98	96.2819	1.02
3/2/2024	94.6979	100.009	5.61	96.2442	1.63	92.3517	2.48	95.4565	0.80	97.6066	3.07	94.914	0.23	**94.8098**	0.12
3/9/2024	87.9631	94.2216	7.11	92.8268	5.53	90.6207	3.02	91.3817	3.89	95.6352	8.72	91.8451	4.41	**89.8523**	2.15
MAPE			21.35		10.51		3.64		7.40		6.27		7.84		**1.66**
RMSPE			2.70		1.63		0.57		0.88		0.71		0.93		**0.23**
MAE			15.93		7.58		2.51		5.32		4.65		5.64		**1.15**
MSE			387.65		128.25		10.99		36.37		27.02		41.20		**1.98**
IA			0.84		0.94		0.99		0.98		0.98		0.98		**1.00**
R			0.88		0.91		0.99		0.98		0.98		0.98		**1.00**
3/16/2024	76.6947	84.5835	10.29	86.2572	12.47	88.0061	14.75	86.9559	13.38	90.8096	18.40	85.4216	11.38	**81.9517**	6.85
3/23/2024	60.3479	82.2837	36.35	79.043	30.98	84.6878	40.33	82.1148	36.07	84.3131	39.71	76.8728	27.38	**72.6616**	20.40
3/30/2024	38.1513	79.0379	107.17	75.1831	97.07	80.8336	111.88	74.8375	96.16	76.7137	101.08	66.7508	74.96	**62.7598**	64.50
MAPE			51.27		46.84		55.65		48.54		53.06		37.91		**30.59**
RMSPE			6.56		5.93		6.92		5.98		6.36		4.65		**3.93**
MAE			23.57		21.76		26.11		22.90		25.55		17.95		**14.06**
MSE			738.38		604.10		847.39		641.66		753.54		389.05		**261.61**
IA			0.49		0.55		0.48		0.54		0.53		0.65		**0.70**
R			**1.00**		0.97		1.00		1.00		1.00		1.00		1.00

models adaptive learning capability, resulting in faster convergence, greater stability, and higher prediction accuracy. Furthermore, the results showed that MARL-WOA achieved better optimization performance, faster convergence, and higher consistency compared with the original WOA and other popular metaheuristic algorithms. This confirmed its strong search efficiency and robustness in solving complex nonlinear optimization problems.

Empirical experiments using real electricity market data from New South Wales, Australia, further demonstrated that the FPE-MARL-WOA model consistently outperformed conventional grey models such as GM(1,N) and ITGM(1,N), as well as other hybrid

Table 7. Wednesday Comparative results of six grey models

Date	Price	GM(1,N)		ITGM(1,N)		GVBM-FHO		PPE-WOA		FPE-WOA		PPE-MARL-WOA		FPE-MARL-WOA	
		Val	APE	Val	APE	Val	APE	Val	APE	Val	APE	Val	APE	Val	APE
12/18/2023	55.2466	55.2466	0	55.2466	0	55.2466	0	55.2466	0	55.2466	0	55.2466	0	55.2466	0
12/25/2023	49.9743	41.2097	17.54	**45.3642**	9.23	40.6063	18.75	40.8134	18.33	45.0108	9.93	42.7021	14.55	45.0043	9.95
1/1/2024	59.3231	78.1235	31.69	**57.1718**	3.63	54.9061	7.45	49.1017	17.23	53.4353	9.92	50.6025	14.70	53.4151	9.96
1/8/2024	68.7068	109.5581	59.46	103.7313	50.98	**68.7676**	0.09	57.3403	16.54	62.4192	9.15	59.0522	14.05	62.3805	9.21
1/15/2024	78.6120	108.232	37.68	94.0402	19.63	**81.0942**	3.16	65.9995	16.04	72.1332	8.24	68.2117	13.23	72.0697	8.32
1/22/2024	88.2000	125.9005	42.74	86.9178	1.45	**91.1444**	3.34	75.379	14.54	82.0165	7.01	77.5886	12.03	81.9206	7.12
1/29/2024	96.9063	124.798	28.78	100.8862	4.11	**98.5414**	1.69	84.2684	13.04	91.5656	5.51	86.7447	10.49	91.4283	5.65
2/5/2024	103.5956	108.8559	5.08	104.07	0.46	**103.2159**	0.37	92.4739	10.74	99.8973	3.57	94.906	8.39	99.7077	3.75
2/12/2024	107.3508	118.5372	10.42	116.1237	8.17	105.3248	1.89	99.0442	7.74	**106.0607**	1.20	101.2442	5.69	105.8062	1.44
2/19/2024	108.0896	131.4592	21.62	128.7767	19.14	105.1684	2.70	103.5342	4.21	109.3561	1.17	105.1628	2.71	**109.022**	0.86
2/26/2024	105.5592	111.6428	5.76	109.5242	3.76	103.1208	2.31	104.6097	0.90	108.9323	3.20	**105.9384**	0.36	108.502	2.79
3/5/2024	99.2849	105.2309	5.99	104.1572	4.91	**99.5766**	0.29	101.8861	2.62	103.7312	4.48	102.672	3.41	103.1853	3.93
MAPE			22.23		10.45		**3.50**		10.16		5.28		8.30		5.25
RMSPE			2.83		1.73		**0.61**		1.21		0.63		0.99		0.63
MAE			17.96		8.44		**2.41**		8.03		4.10		6.53		4.06
MSE			495.59		171.06		**11.97**		85.16		21.31		56.60		21.18
IA			0.81		0.92		**0.99**		0.96		0.99		0.97		0.99
R			0.84		0.90		**0.99**		0.97		0.99		0.98		0.99
3/12/2024	88.5088	94.4938	6.76	94.1284	6.35	94.9155	7.24	96.4356	8.96	93.8848	6.07	95.5438	7.95	**93.2015**	5.30
3/19/2024	71.8328	91.9672	28.03	85.3389	18.80	89.4798	24.57	88.5115	23.22	80.8199	12.51	85.905	19.59	**79.9736**	11.33
3/26/2024	48.0608	88.371	83.87	82.1073	70.84	83.564	73.87	79.3057	65.01	65.062	35.37	74.3112	54.62	**64.0227**	33.21
MAPE			39.56		32.00		35.23		32.40		17.99		27.39		**16.62**
RMSPE			5.12		4.25		4.51		4.02		2.19		3.38		**2.05**
MAE			22.14		17.72		19.85		18.62		10.45		15.79		**9.60**
MSE			688.71		457.72		537.65		439.09		132.90		312.20		**114.36**
IA			0.51		0.60		0.57		0.63		0.85		0.70		**0.87**
R			**1.00**		0.94		1.00		1.00		1.00		1.00		1.00

approaches like GVBM-FHO, PPE-WOA, and PPE-MARL-WOA. The model achieved superior results across several evaluation metrics, including MAPE, RMSPE, MAE, MSE, IA, and R. Its strong generalization abilityconfirmed on both weekday (Wednesday) and weekend (Sunday) datasetshighlights the advantage of combining interpretable grey modeling with intelligent, learning-based optimization. The FPE-MARL-WOA model achieved very low prediction errors and near-perfect correlation with observed data, confirming its reliability for forecasting under uncertainty and limited data conditions.

By integrating grey theory, metaheuristic optimization, and reinforcement learning, this study presents a significant advancement over traditional forecasting approaches. Our proposed model combines interpretability and adaptability, making it a reliable and practical tool for modern electricity markets characterized by volatility and incomplete information.

Future work may extend the ITGMVM by incorporating additional exogenous factors such as economic indicators, weather-driven renewable generation, and policy or market structural changes to better capture real-world variability. Enhancing the time-delay mechanism with adaptive or data-driven kernels could further improve the representation of heterogeneous lag effects. Moreover, integrating more advanced MARL-based optimization strategies or deep RL architectures may strengthen convergence and robustness. Finally, expanding the framework to multi-step and probabilistic forecasting would increase its applicability in operational decision-making within electricity markets.

Authors' Contributions

All authors have the same contribution.

Data Availability

The manuscript has no associated data or the data will not be deposited.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Ethical Considerations

The authors have diligently addressed ethical concerns, such as informed consent, plagiarism, data fabrication, misconduct, falsification, double publication, redundancy, submission, and other related matters.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or nonprofit sectors.

A Appendix

A.1 Whale Optimization Algorithm (WOA)

The Whale Optimization Algorithm (WOA) is a bio-inspired algorithm based on the hunting and foraging behaviors of humpback whales [3]. This algorithm mimics two key strategies used by humpback whales to hunt their prey: **encircling prey** and **bubble-net attacking**. These behaviors are used to drive the search agents towards the optimal solution in the search space (see Algorithm 2).

A.2 Encircling Prey (Exploitation Phase)

In the WOA, the position of each search agent is updated based on its current position and the position of the best solution (prey). The position update rule is as follows:

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot |\vec{C} \cdot \vec{X}^*(t) - \vec{X}(t)|, \tag{42}$$

where:

- $\vec{X}^*(t)$ is the best solution at time t
- $\vec{X}(t)$ is the current position of a search agent
- $\vec{A}$ and $\vec{C}$ are coefficient vectors that control the movement

The vectors $\vec{A}$ and $\vec{C}$ are computed as:

$$\vec{A} = 2 \cdot a \cdot \vec{r} - a, \tag{43}$$

$$\vec{C} = 2 \cdot \vec{r}, \tag{44}$$

where $\vec{a}$ decreases linearly from 2 to 0, and $\vec{r}$ is a random vector in $[0, 1]$.

A.3 Bubble-Net Attack (Exploitation Phase)

There are two ways to simulate the bubble-net attack:

- **Shrinking Encircling Mechanism:** The shrinking of the search space as the whales get closer to the prey
- **Spiral Position Update:** A helix-shaped movement towards the prey.

In the spiral model, the position update of a whale is given by:

$$\vec{X}(t+1) = \vec{X}^*(t) + D \cdot e^{bl} \cdot \cos(2\pi l), \tag{45}$$

where:

- $D = |\vec{X}^*(t) - \vec{X}(t)|$ is the distance between the whale and the prey
- b controls the shape of the spiral
- l is a random number in the range $[-1, 1]$

A.4 Exploration Phase

The position update during exploration is as follows:

$$\vec{X}(t+1) = \vec{X}_{\text{rand}} - \vec{A} \cdot \vec{D}, \tag{46}$$

where $\vec{X}_{\text{rand}}$ is a randomly chosen agent from the population, and $\vec{D} = |\vec{C} \cdot \vec{X}_{\text{rand}} - \vec{X}|$ is the distance between the agent and the randomly chosen one.

The vector $\vec{A}$ is modified during exploration to encourage global search, meaning that if $|\vec{A}| > 1$, the agent is allowed to explore a wider area in the search space.

Algorithm 2: Whale Optimization Algorithm (WOA) [3]

Input: Population size n, maximum number of iterations $t_{\max}$
Output: Optimal solution X^*

```
Initialize the population of whales X_i, for i = 1, 2, ..., n ;
Evaluate the fitness of each whale agent and identify the best solution X* ;
t ← 0 ;
while t < t_max do
    for each whale agent X_i do
        Update the control parameters a, A, C, l, and p ;
        if p < 0.5 then
            if |A| < 1 then
                Update position using the "encircling prey" behavior ;
            else if |A| ≥ 1 then
                Select a random whale agent X_rand ;
                Update position using exploration dynamics ;
        else if p ≥ 0.5 then
            Update position using the "bubble-net foraging" mechanism ;
        end if
    end for
    Check and correct whale positions that violate the search boundaries ;
    Re-evaluate the fitness of each whale agent ;
    Update X* if a better solution is found ;
    t ← t + 1 ;
end while
return X* ;
```

References

[1] Jiang, H., Kong, P., Hu, Y. C., and Jiang, P. Forecasting China's CO_2 emissions by considering interaction of bilateral FDI using the improved grey multivariable Verhulst model. International Journal of Computer Mathematics, 23, 225–240, (2021).

[2] Ding, S. and Li, R. A new multivariable grey convolution model based on Simpson's rule and its applications. Complexity, 2020(1), 4564653, (2020).

[3] Mirjalili, S. and Lewis, A. The Whale Optimization Algorithm. Advances in Engineering Software, 95, 51–67, (2016).

[4] Watkins, C. J. and Dayan, P. Q-learning. Machine Learning, 8(3-4), 279–292, (1992).

[5] Buoniu, L., Bbuka, R., and De Schutter, B. A comprehensive survey of multiagent reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 38(2), 1–18, (2008).

[6] Hu, Z. and Yu, X. Reinforcement learning-based comprehensive learning grey wolf optimizer for feature selection. Applied Soft Computing, 149, 110959, (2023).

[7] Seyyedabbasi, A., Aliyev, R., Kiani, F., Gulle, M. U., Basyildiz, H., and Shah, M. A. Hybrid algorithms based on combining reinforcement learning and metaheuristic methods to solve global optimization problems. Knowledge-Based Systems, 223, 107044, (2021).

[8] Digalakis, J. and Margaritis, K. On benchmarking functions for genetic algorithms. International Journal of Computer Mathematics, 77(4), 481–506, (2001).

[9] Molga, M. and Smutnicki, C. Test functions for optimization needs. Technical Report, (2005).

[10] Yang, X. S. Firefly algorithm, stochastic test functions and design optimisation. International Journal of Bio-Inspired Computation, 2(2), 78–84, (2010).

[11] Yao, X., Liu, Y., and Lin, G. Evolutionary programming made faster. IEEE Transactions on Evolutionary Computation, 3(2), 82–102, (1999).

[12] Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting, 30(4), 1030–1081, (2014).

[13] He, Y., Wang, C., and Chen, Z. State-of-the-art electricity load and price forecasting for wholesale markets: A comprehensive review. Energies, 17(22), 5797, (2024).

[14] Li, J., Zhao, S., and Wang, L. Impact of renewable energy on extreme volatility in wholesale electricity prices. Journal of Cleaner Production, 442, 140256, (2024).

[15] Tang, R., Zhou, K., and Yang, Q. Challenges and trends in high-volatility electricity price forecasting. Energies, 17(22), 5802, (2024).

[16] Li, X., Zhou, H., and Liu, J. A hybrid GM(1,N)-ARIMA model for electricity price forecasting. Energy Reports, 7, 2874–2882, (2021).

[17] Zhao, Y., He, Z., and Xu, M. Improved time-delay grey model optimized by GA for power price peak forecasting. Energy, 254, 124363, (2022).

[18] Chen, L., Wang, X., and Zhang, H. A PSO-optimized multivariable grey Verhulst model for energy market forecasting. Expert Systems with Applications, 211, 118635, (2023).

[19] Wang, B., Lin, Z., and Wu, T. Electricity price forecasting using hybrid GM(1,1)-SVR with residual correction. Applied Energy, 309, 118460, (2022).

[20] Liu, M., Zhang, Y., and Xu, D. Grey-based hybrid prediction model for spot electricity price using Firefly Algorithm. Journal of Cleaner Production, 258, 120743, (2020).

[21] Yu, X., Lu, L., Qi, J., Qian, Y., Zhao, L., Tan, C., Chen, Y., and Han, Z. A clustering fractional-order grey model in short-term electrical load forecasting. Scientific Reports, 15(1), 6207, (2025).

[22] Gou, X., Mi, C., Yang, Y., and Zeng, B. A nonlinear mixed-frequency grey prediction model with two-stage lag parameter optimization and its application. Applied Mathematical Modelling, 116360, (2025).

[23] Talbi, E. G. Metaheuristics: From Design to Implementation. Wiley Series on Parallel and Distributed Computing, (2009).

[24] Zheng, J., Zhong, J., Chen, M., and He, K. Reinforced Hybrid Genetic Algorithm for the Traveling Salesman Problem. arXiv preprint arXiv:2107.06870, (2021).

[25] Kiani, F., Seyyedabbasi, A., Aliyev, R., et al. Hybrid algorithms based on combining reinforcement learning and metaheuristic methods to solve global optimization problems. Knowledge-Based Systems, 223, 107044, (2021).

[26] Nguyen, H. T., Nguyen, H. V., and Nguyen, T. T. Proposing a hybrid metaheuristic optimization algorithm for energy consumption prediction in buildings. Scientific Reports, 12, 4923, (2022).

[27] Fu, Y., Wu, D., Boulet, B., and Zinflou, A. Time series forecasting via reinforcement learning-based model combination. IEEE Internet of Things Journal, (2025).

[28] Wu, Z., Fang, G., Ye, J., Zhu, D. Z., and Huang, X. A reinforcement learning-based ensemble forecasting framework for renewable energy forecasting. Renewable Energy, 244, 122692, (2025).

[29] Ghimire, S., Deo, R. C., Casillas-Perez, D., Sharma, E., Salcedo-Sanz, S., Barua, P. D., and Acharya, U. R. Half-hourly electricity price prediction with a hybrid convolution neural network–random vector functional link deep learning approach. Applied Energy, 374, 123920, (2024).

[30] Xu, Y., Huang, X., Zheng, X., Zeng, Z., and Jin, T. VMD-ATT-LSTM electricity price prediction based on grey wolf optimization algorithm in electricity markets considering renewable energy. Renewable Energy, 236, 121408, (2024).

[31] Meng, A., Wang, P., Zhai, G., Zeng, C., Chen, S., Yang, X., and Yin, H. Electricity price forecasting with high penetration of renewable energy using attention-based LSTM network trained by criss-cross optimization. Energy, 254, 124212, (2022).

[32] Forrest, S. Genetic algorithms: Principles of natural selection applied to computation. Science, 261(5123), 872–878, (1993).

[33] Storn, R. and Price, K. V. Differential Evolution—A simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization, 11(4), 341–359, (1997).

[34] Beyer, H.-G. and Schwefel, H.-P. Evolution strategies: A comprehensive introduction. Natural Computing, 1(1), 3–52, (2002).

[35] Poli, R., Kennedy, J., and Blackwell, T. Particle swarm optimization—An overview. Swarm Intelligence, 1(1), 33–57, (2007).

[36] Dorigo, M., Maniezzo, V., and Colorni, A. The Ant System: Optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man, and Cybernetics—Part B, 26(1), 29–41, (1996).

[37] Yang, X.-S. Firefly algorithm, stochastic test functions and design optimisation. International Journal of Bio-Inspired Computation, 2(2), 78–84, (2010).

[38] Kirkpatrick, S., Gelatt, C. D., Jr., and Vecchi, M. P. Optimization by simulated annealing. Science, 220(4598), 671–680, (1983).

[39] Rashedi, E., Nezamabadi-pour, H., and Saryazdi, S. GSA: A gravitational search algorithm. Information Sciences, 179(13), 2232–2248, (2009).

[40] Ma, X. and Liu, Z. The kernel-based nonlinear multivariate grey model. Applied Mathematical Modelling, 56, 217–238, (2018).

[41] Ding, S., Hu, J., and Lin, Q. Accurate forecasts and comparative analysis of Chinese CO_2 emissions using a superior time-delay grey model. Energy Economics, 126, 107013, (2023).

[42] Pour, S. H., Fard, O. S., and Zeng, B. A grey prediction model based on Von Bertalanffy equation and its application in energy prediction. Engineering Applications of Artificial Intelligence, 143, 110012, (2025).