Optimizing Gas Consumption Predictions: A Comprehensive Study of Individual and Hybrid Modeling Approaches with Practical Implications for Energy Policies

Changhao Zhang* and Mengyu Ren

Henan University of Animal Husbandry and Economy; Zhengzhou, Henan, 450000, China
E-mail: qq22630186@163.com
**Corresponding Author*

Received 13 April 2025; Accepted 28 November 2025

Abstract

Natural gas, the cleanest fossil fuel, is increasingly important due to its abundance and lower carbon emissions. However, accurately forecasting gas demand remains challenging. To forecast gas usage, this study uses sophisticated machine learning (ML) techniques, including CatBoost, XGBoost, and MLP. Six prediction models and hyperparameter optimization are created and assessed. Hybrid XGBoost models, particularly XGBoost-SSA and XGBoost-SMA, demonstrate superior convergence and accuracy. Visual aids like correlation matrices and scatter plots provide insights into model performance. The research contributes to enhancing the efficiency of gas distribution operations, ensuring energy security, economic stability, and

Strategic Planning for Energy and the Environment, Vol. 45_2, 531–560.
doi: 10.13052/spee1048-5236.45210

environmental sustainability. By integrating renewable energy and leveraging real-time analytics, the study addresses the evolving dynamics of gas consumption forecasting, offering valuable implications for energy policies and investment strategies.

Keywords: Gas consumption forecasting, individual modeling, hybrid modeling, CatBoost, MLP, XGBoost, SMA, SSA, predictive accuracy, energy management, resource allocation.

1 Introduction

Natural gas, recognized as the cleanest fossil fuel, has gained increasing prominence owing to its abundant resources, rising investments, and lower carbon emissions compared to other fossil fuels [1]. Its role has expanded globally across residential, industrial, and power generation sectors, while policy efforts in Europe and elsewhere have further reinforced its importance as a transitional energy source in the shift toward low-carbon systems [2, 3]. Despite these advantages, accurately forecasting natural gas demand remains challenging due to its sensitivity to environmental, economic, and geopolitical factors.

The reliability of gas demand forecasts is strategically significant, as it underpins energy policy formulation, infrastructure investment, and market stability. Precise predictions help avoid supply–demand imbalances, reduce risks of price volatility, and support the integration of natural gas with renewable energy systems [4]. In countries such as Italy, where regional suppliers manage procurement, distribution, and pipeline operations, demand forecasting directly influences operational efficiency, service reliability, and security [5]. This reflects a broader global challenge: ensuring that forecasting methods are robust enough to manage volatility while enabling sustainable, secure, and cost-effective energy systems.

Conventional forecasting methods – statistical and econometric models – offer useful insights but are typically inadequate to account for high variability of gas consumption patterns when conditions are rapidly changing [6]. Several recent developments in machine learning (ML) and optimization methods have provided more flexible and accurate tools for forecasting, which are able to capture nonlinear relationships and adapt to continuously changing environments [7, 8]. On the other hand, there are clear gaps in these areas of literature: many studies rely on single-model methods, offering limited examination of hybrid or optimized frameworks that have the potential to dramatically improve prediction accuracy.

This study addresses these gaps by applying state-of-the-art ML algorithms – CatBoost, MLP, and XGBoost – together with metaheuristic optimization methods (SMA and SSA) to forecast gas consumption. By systematically comparing both individual and hybrid models, the research not only benchmarks their predictive performance but also highlights the strategic value of optimized hybrid approaches in enhancing forecasting accuracy. The methodological and empirical contributions of this study thus extend beyond technical performance, offering implications for energy policy, system planning, and the integration of gas into future low-carbon energy landscapes.

The literature on natural gas consumption forecasting spans statistical modelling, econometric analysis, machine learning approaches, and hybrid methods, each offering unique strengths and limitations [9]. Research has evolved through four main stages: the Initial phase (1950–1970), the Conventional era (1970–1992), the integration of AI techniques (1992–2006), and the All-round phase (2006–present) [10]. Early studies, such as Verhulst's (1950) work on French gas plants, applied basic statistical models to link gas consumption with factors like income and prices, laying the groundwork for later advancements [11]. This trajectory highlights the gradual shift from structural statistical models to modern AI-driven approaches, revealing both methodological progress and emerging research gaps. In more recent studies Svoboda et al. (2023) [12] present a novel framework for multistep ahead natural gas consumption forecasting that integrates change point detection with continual learning in a data stream processing environment. The study employs Hoeffding tree predictors as forecasting models and leverages the Pruned Exact Linear Time (PELT) algorithm to identify structural changes in consumption trends. By defining and testing three different model collection selection procedures – both with and without error feedback loops – the authors demonstrate how adapting model sets to detected change points can enhance forecasting robustness. Their results show that fewer change points reduce forecasting error and that simpler selection procedures, which omit error feedback, often yield more reliable models for real-world continual learning applications. Zhang et al. (2022) [13] propose a hybrid forecasting framework for natural gas consumption (NGC) by integrating Extreme Gradient Boosting (XGBoost) with the Salp Swarm Algorithm (SSA). In this approach, XGBoost serves as the core forecasting model within a nonlinear autoregression structure, while SSA optimizes its most sensitive hyperparameters. Case studies on the UK and Netherlands demonstrate that the hybrid XGBoost–SSA model consistently outperforms 15 alternative hybrid schemes, achieving superior accuracy with mean absolute percentage errors

(MAPE) of 4.98% for the UK and 9.05% for the Netherlands. The results highlight the model's strong potential for reliable multi-step ahead NGC forecasting and its applicability in broader clean energy planning contexts. Khajavi & Rastgoo (2023) [14] introduce a hybrid machine learning framework to forecast heating energy consumption in residential buildings, addressing the complexity of multiple influencing variables. Their approach uses Support Vector Regression (SVR) as the core predictor, enhanced through hyperparameter optimization with six metaheuristic algorithms. Case study evaluations show that the SVR–Battle Royale Optimization (BRO) model achieves the most accurate results, with exceptionally high performance ($R^2 = 0.999386$ for training and 0.998898 for testing). The findings confirm the effectiveness of metaheuristic–SVR hybrids for highly precise energy demand prediction in the building sector. Zhou et al. (2024) [15] developed the first foundation model for natural gas demand forecasting, addressing the limitations of traditional methods that lack generalization across industries. Their approach integrates contrastive learning with advanced noise filtering to improve representation quality and prediction accuracy in noisy real-world datasets. Using large-scale data from over 10,000 customers across multiple sectors, the model significantly outperformed state-of-the-art benchmarks, achieving a 3.68% reduction in MSE and a 6.15% improvement in MASE, thereby proving its robustness and adaptability for energy forecasting applications. Zhou et al. (2024) [16] present a hybrid forecasting framework for electricity consumption that integrates LightGBM and CatBoost models optimized via Arithmetic Optimization and Fruit Fly Optimization algorithms. The study employs sensitivity analysis to evaluate the influence of variables such as customer behavior and temperature on consumption patterns. Results demonstrate exceptional predictive accuracy ($R^2 = 0.9996$ for training and 0.9968 for testing), highlighting the framework's effectiveness in generating reliable electricity demand forecasts and supporting informed energy management decisions.

1.1 Main Contribution & Novelties

This study applied state-of-the-art ML approaches to forecast gas consumption and focused more on improving forecasting accuracy. The research integrated advanced methods such as CatBoost, MLP, and XGBoost, combining them with different optimization algorithms like SMA and SSA. A wide comparison analysis was also performed for these ML methods, singly and in combination. The primary objective was to determine whether the techniques

applied would optimize gas consumption prediction accuracy. This research is important in terms of enhancing the current understanding of the predictive capacity of ML in forecasting gas consumption. It provides rich insights into possible avenues for enhancements with both single and hybrid models to improve forecast accuracy. This research paper's latter portions adhere to a formal academic format. The prediction methods used are thoroughly examined in Section 2, which also provides an overview of the suggested models (CatBoost, MLP, and XGBoost) and explains the features of the optimization strategies. The findings and additional analyses of these models are presented in Section 3, along with a number of charts and tables to aid in a more thorough analysis. The main conclusions and implications drawn from this study are summarized in Section 4, which brings the work to a close. The contribution made by this systematic approach within the research study adds to the academic discourse on ML in the application of gas consumption forecasting and thus lays a foundation for further research studies in this area.

The novelty of this study comes from the combination of state-of-the-art machine learning models (CatBoost, XGBoost, MLP) with innovative metaheuristic optimization techniques (SMA and SSA) to predict natural gas consumption. This approach differs from the traditional statistical or single-model approaches used in prediction, as the framework developed in this study is a hybrid that allows prediction accuracy, convergence speed, and model robustness to improve simultaneously, even with limited or noisy datasets. The implications of these decisions are great, as energy suppliers and policymakers can rely on this method for high-quality and reliable forecasts, which would allow for optimal procurement decisions, optimized infrastructure and supply, and optimized supply-demand balance for natural gas and support for energy transition strategies. Additionally, the contributions to current practice through this study are significant, as we reveal the usefulness of combining metaheuristic optimization with ensemble learning, while also providing a scalable, adaptable, and data-driven solution for energy management, or informed decision-making, for natural gas distributions in real time under uncertainty.

2 Methodology

The employment of cutting-edge ML techniques is the main topic of this research project, with a focus on the CatBoost, MLP, and XGBoost algorithms for gas consumption level prediction. The research has placed

considerable emphasis on improving the predictability of the models through an elaborate process of hyperparameter optimization. This is achieved by tuning and optimizing the key parameters that govern the models' prediction performance using sophisticated optimization methods like SMA and SSA. This paper aims to develop and evaluate six different predictive models for forecasting gas consumption levels. A rigorous comparison between these models is carried out utilizing a strong statistical approach for performance evaluation in order to determine which of them is the best and most dependable in providing effective forecasts. The adopted methodological framework for the research consists of a really careful process of data collection, analysis, and high-stringency validation processes. Data partitioning comes after data pre-processing, and a well-planned approach divides the dataset into a training set and a test set, respectively, with an $80\% : 20\%$ split. It is very important for further work to do this to make the models strong and accurate because then only thorough testing will be possible across two different sets of data. Two steps make up the predictive modeling process: first, the performance of the CatBoost, MLP, and XGBoost algorithms is assessed separately to determine their ability to make predictions on their own; second, the study concentrates on hybrid models that combine the advantages of MLP, XGBoost, and CatBoost to improve prediction accuracy. Precise and effective tuning of hyperparameters is carried out to ensure the best possible performance of these models. Different optimization techniques, such as SMA and SSA, play the most important role in searching for the best hyperparameter settings of CatBoost, MLP, and XGBoost models. The models' predictive capability and dependability in gas consumption forecasts are enhanced by this thorough optimization process, which also yields good results. The results of this study will be presented through extensive discussions in the following sections, supported by informative graphical representations, charts, and tables. The academic community benefits from richer knowledge and interpretation because of the thoughtful integration of these visual aids, which offer a thorough and understandable explanation of the study findings. Additionally, Figure 1 presents a schematic illustration outlining the comprehensive structure of the employed model, along with the methodology applied in this research, depicted as a flowchart. This scholarly inquiry primarily focuses on investigating the geographical area of a campus-distributed energy system in Japan, spanning a temporal range in 2022.

Besides, the Algorithm 1 in the following provides the pseudocode of the overall process of the study.

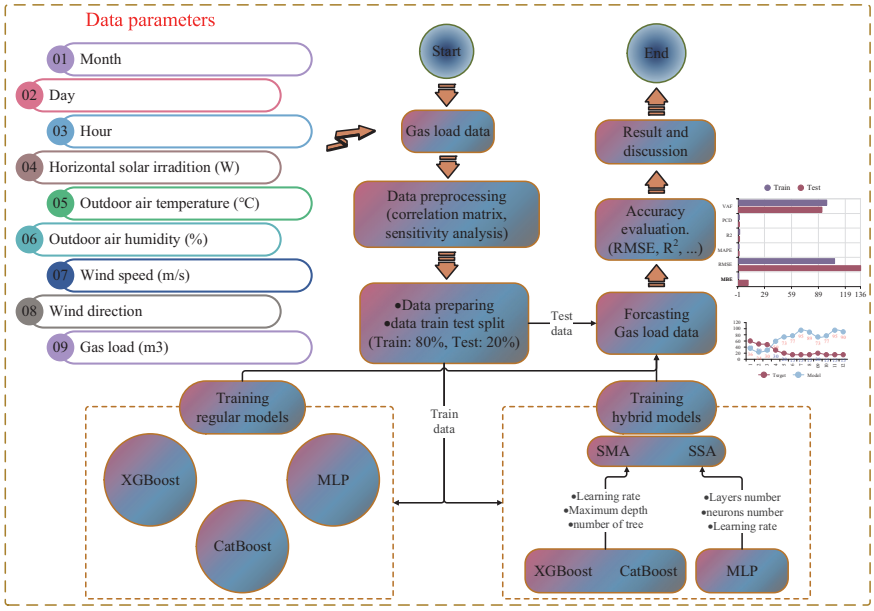

Figure 1 Diagram of the present investigation's flowchart.

Algorithm 1 Overall procedure for gas consumption forecasting.

Require: Raw Data D_{raw} for a campus energy system in Japan (2022).
Ensure: Performance comparison of all models; Identification of the optimal model.

– STAGE 1: Data Preparation –
Step 1: Load and Preprocess Data
$D_{processed} \leftarrow LoadAndClean(\mathrm{D}_{raw})$
Features: Month, Day, Hour, Solar Irradiation, Temperature, etc.
Target: Gas Load (m^*)
Step 2: Analyze Data Characteristics
$\mathrm{M}_{corr} \leftarrow CalculateCorrelationMatrix(\mathrm{D}_{processed})$
$\mathrm{S}_{sens} \leftarrow PerformSensitivityAnalysis(\mathrm{D}_{processed})$
Step 3: Partition Data
$\mathrm{D}_{train}, \mathrm{D}_{test} \leftarrow SplitData(\mathrm{D}_{processed}, train\_ratio = 0.8)$

– STAGE 2: Individual Model Training and Evaluation –
Step 4: Process Standalone Models
for each model M in {CatBoost, MLP, XGBoost} **do**
    $\mathrm{M}_{trained} \leftarrow TrainModel(\mathrm{M}, \mathrm{D}_{train})$
    $\mathrm{P}_{train} \leftarrow EvaluatePerformance(\mathrm{M}_{trained}, \mathrm{D}_{train})$
    $\mathrm{P}_{test} \leftarrow EvaluatePerformance(\mathrm{M}_{trained}, \mathrm{D}_{test})$
    Store results for M in the results repository
end for

– STAGE 3: Hybrid Model Optimization, Training, and Evaluation –
Step 5: Process Hybrid Models
for each ML Model M in {CatBoost, MLP, XGBoost} **do**
  for each Optimizer O in {SMA, SSA} **do**
    ▷ Define the hyperparameter search space for the current ML Model
    $\mathrm{S}_{hyper} \leftarrow DefineSearchSpace(\mathrm{M})$
    ▷ Use the optimizer to find the best hyperparameters
    $\mathrm{H}_{optimal} \leftarrow OptimizeHyperparameters(\mathrm{O}, \mathrm{M}, \mathrm{D}_{train}, \mathrm{S}_{hyper})$
    ▷ Train the final hybrid model with the optimal hyperparameters
    $\mathrm{M}_{hybrid} \leftarrow InitializeModel(\mathrm{M}, \mathrm{H}_{optimal})$
    $\mathrm{M}_{hybrid\_trained} \leftarrow TrainModel(\mathrm{M}_{hybrid}, \mathrm{D}_{train})$
    ▷ Evaluate the final hybrid model
    $\mathrm{P}_{hybrid\_train} \leftarrow EvaluatePerformance(\mathrm{M}_{hybrid\_trained}, \mathrm{D}_{train})$
    $\mathrm{P}_{hybrid\_test} \leftarrow EvaluatePerformance(\mathrm{M}_{hybrid\_trained}, \mathrm{D}_{test})$
    Store results for hybrid model (M-O) in the results repository.
  end for
end for

– STAGE 4: Comprehensive Analysis and Comparison –
Step 6: Collate and Compare Results
Gather all stored performance metrics (RMSE, $\mathrm{R}^{\wedge}$, MAPE, VAF, PCD) from the results repository.
Generate comparative plots and tables (e.g., Figures 4, 6, 7, and Tables 3, 4).
Analyze convergence, stability, and runtime performance (e.g., Figures 9, 11).
Step 7: Identify Optimal Model
$\mathrm{M}_{best} \leftarrow IdentifyBestModel(\text{all results})$ based on superior predictive accuracy, stability, and lower error on the test set.
return The comprehensive analysis and the identified best-performing model, M_{best}.

2.1 Data

The data used in this investigation comes from earlier studies by JIN, Xiaoyu, and Beeravalli [17]. This scholarly investigation primarily centers on exploring the geographical scope of a campus-based distributed energy system in Japan, spanning the temporal extent of 2022. It is crucial to acknowledge that the estimation of gas usage is influenced by a multitude of significant factors. This target, the prediction of gas consumption, encompasses several parameters: Boiler (m3), Gas engine (m3), Fuel Cell (m3), and three Absorption chillers (m3). A complete statistical study of these relevant variables is painstakingly identified and clarified in Table 1. This tabular representation facilitates a thorough understanding of the numerous elements that exert considerable influence on gas consumption forecasting. For the analysis and conclusions that follow, it provides an empirical foundation for this research. This research, therefore, builds a strong foundation for understanding the

Table 1 The statistical details that correspond to the input parameters

Variables	Count	Mean	Std	Min	25%	50%	75%	Max
Month	7296	5.526316	2.871212	1	3	6	8	10
Day	7296	15.71382	8.794684	1	8	16	23	31
Hour	7296	11.5	6.922661	0	5.75	11.5	17.25	23
Horizontal solar irradiation (W)	7296	188.0566	281.1532	0	0	10	302	1193
Outdoor air temperature (°C)	7296	18.42057	8.79052	−0.3	10.8	19.3	26.2	36.4
Outdoor air humidity (%)	7296	67.79866	14.21803	20	58	68	78	99
Wind speed (m/s)	7296	2.053673	1.641432	0	0.9	1.7	2.8	15.3
Wind direction	7296	269.156	121.4241	1	175	251	380	539
Gas load (m3)	7296	750.5373	747.9824	49	232	438	963.25	3559
Month	7296	5.526316	2.871212	1	3	6	8	10
Day	7296	15.71382	8.794684	1	8	16	23	31

complex interaction of factors through a systematic delineation in an influencing factor structured table that affects the predictive modeling of gas consumption.

2.2 Machine Learning Methods

An academic summary of the methods used to anticipate gas consumption is provided in this section. The study uses advanced models like CatBoost, MLP, and XGBoost to achieve this objective. Optimization methods like SMA and SSA are incorporated into the study framework for these prediction models to be more accurate and efficient. These are the methods that have been carefully selected since choosing the right approach is crucial for enhancing the accuracy and durability of gas consumption forecasts. This aspect will be discussed further in the following sections of this thorough investigation. The methodological approach followed here not only underlines the sophistication of the models used but also gives full signification to the deliberate efforts made in order to optimize their performance, thus contributing to the overall reliability and effectiveness of the gas consumption forecasting process.

2.2.1 Categorical gradient boosting (CatBoost)

CatBoost [18] is a substantial development in the area of gradient boosting and decision tree methods in academic publications. The underlying concept of boosting is to integrate a series of weak models in order to construct a strong predictive model that performs better than random guessing. By generating decision trees one after the other, gradient boosting reduces errors by allowing the new trees to learn from the faults of the older ones. This

repeated procedure of adding a new function continues until the selected loss function reaches its lowest value. The innovative way CatBoost works in constructing decision trees makes it stand out from classical gradient-boosting models. Rather than using conventional gradient-boosting trees, CatBoost adopts the method of "oblivious trees." These trees possess a distinctive feature such that every node at a given level assesses the identical predictor under uniform conditions. This arrangement increases the process's overall efficiency by making it easier to determine a leaf's index only through bitwise operations. Also, CatBoost uses a new way of selecting data to construct the ht+1 decision tree by random permutation of the elements in D and an arbitrary sequence. A more sensitive method is to create a new set $Dk = \{x_1, x_2, \ldots, x_{k-1}\}$, where $x_1, x_2, \ldots, x_{k-1}$ are the items of D stated in the order that the random permutation σ determines, and (k) is the kth element of D under the permutation σ. CatBoost employs a modified technique to determine the encoded value X_k for the ith categorical variable while fitting the decision tree h^{t+1}, in contrast to the strict adherence to Equation (1):

$$\hat{X}_k^i = \frac{\sum x_j \in D_k 1 x_j^i = x_k^i . y_j + ap}{\sum x_j \in D_k 1 x_j^i = x_k^i + a} \tag{1}$$

2.2.2 Multi-layer perceptron (MLP)

The structural models and neurons of the human brain served as the inspiration for the 1943 development of Artificial Neural Networks ($ANNs$) [19]. The ability to recognize intricate relationships between input and output vectors is a fundamental feature of these networks. Since then, ANN variants have appeared, such as radial basis function networks, feed-forward, recurrent, spiking, and Kohonen self-organizing networks. An ANN subtype called feed-forward neural networks ($FNNs$) is made up of parallel input, hidden, and output layers. An FNN with only one hidden layer is called an MLP. Each layer's number of neurons varies, and only input vectors affect neurons in the input layer; outputs from the layer above inform layers below. Stochastic optimization methods are frequently used to determine the hidden layer's neuron count. ANNs employ learning mechanisms to adjust weights and biases of network connections, aligning outputs with desired targets. This adjustment occurs iteratively, guided by input-output comparisons until a satisfactory level of output alignment is reached. The function of the network is expressed mathematically, defining the connection between input and output, and a variety of stochastic and deterministic learning techniques have been proposed in previous literature. In this study, tangent and purelin

functions were used in the hidden and output layers of the MLP, respectively. The calculation of MLP output involves three steps, determined by the weights and biases [20].

In the first step, the weighted sums of inputs are initially computed as outlined in Equation (2):

$$r_j = \sum_{i=1}^{l} (W_{ij}, Y_i) + B_j \tag{2}$$

Here, l stands for the number of input nodes, W_{ij} for the connection weight between the ith input node and the jth hidden node, and h for the number of hidden nodes, where j ranges from 1 to h. B_j is the bias of the jth hidden node, and Y_i is the ith input.

The second step is to use Equation (3) to calculate each concealed node's output:

$$R_j = tansig(r_j) = \frac{2}{1 + exp(-2 \times r_j)} - 1 \tag{3}$$

Step 3 involves computing the ultimate result based on Equations (4) and (5):

$$f = \sum_{j=1}^{h} (Wj, Ri) + B' \tag{4}$$

$$F = purelin(f) = f \tag{5}$$

The connection weight (Wj) between the jth hidden node and the output node and the output node's bias (B') are essential elements in this scholarly investigation. The MLP parameters were trained using the ALO and PSO algorithms.

2.2.3 eXtreme Gradient Boosting (XGBoost)

According to scholarly discourse, XGBoost represents a substantial improvement over the gradient-boosting system that Chen and Guestrin first presented in 2016 [21]. This tree-boosting method, distinguished for its scalability and comprehensiveness, has gained widespread acceptance and has demonstrated remarkable efficacy in both classification and regression tasks. A key advantage lies in its ability to address overfitting while enabling the parallel construction of decision trees, thereby substantially accelerating the computational process. Equation (6) illustrates how $XGBoost$, which operates as a group of regression trees, or $CART$, combines the results from each tree to

provide prediction scores.

$$\hat{Y} = \sum_{m=1}^{M} f_m(X) \tag{6}$$

In this particular scenario, M represents the total count of trees, where each individual tree within the ensemble is represented by f_m. In contrast to Friedman's 2001 original gradient boosting system, $XGBoost$ adds a regularization component to the loss function. As seen in Equation (7), the goal of this regularization is to enhance the mth iteration's performance.

$$G^m = \sum_{i=1}^{n} l(y_i, \hat{y}_i^m) + \sum_{j=1}^{m} \Omega(f_j) \tag{7}$$

The number n in this sentence denotes the total number of data samples. To measure the difference between the actual target, y_i, and the expected output, $\hat{y}_i^m$, the differential loss function, $l()$, is employed. Furthermore, the term $\Omega(\text{fi})$ represents the regularization term aimed at alleviating overfitting, as detailed in Equation (8).

$$\hat{y}_i^m = \hat{y}_i^{m-1} + f_m(x_i) \tag{8}$$

$$G^m = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{m-1} + f_m(x_i)) + \Omega(f_m) + const \tag{9}$$

In this particular context, the symbol 'const' denotes the summation of the numerical values $\sum_{j=1}^{m-1}(\text{fi})$. Known as $(m-1)$, this number was already established by the tree in the iteration that came right before the present one and is regarded as a fixed amount. The objective function may be reformulated in a different way by using the second-order Taylor expansion on the loss function.

$$G^m = \sum_{i=1}^{n} [l(y_i, \hat{y}_i^{m-1}) + g_i f_m(x_i) + 0.5 h_i f_m^2(x_i)] + \Omega(f_m) \tag{10}$$

In this particular scenario, the symbols g_i and h_i denote the primary and secondary differentials of the loss function correspondingly. Compared to previous gradient-boosting frameworks, $XGBoost$ enables faster and more accurate gradient convergence by using the second-order Taylor expansion approach for the loss function. Additionally, this technology streamlines the

loss function derivation procedure. Furthermore, by adding a regularization component to the objective function, $XGBoost$ facilitates a balanced trade-off between minimizing the objective function, reducing model complexity, and effectively addressing overfitting problems [22].

2.2.4 Slime mould algorithm (SMA)

In academia, the SMA was initially presented by [23] as a novel optimization technique aimed at global optimization. It emulates the natural oscillatory behavior observed in slime molds. The mathematical representation of SMA is delineated below:

1. Food Approach Phase: This stage simulates the movement of the slime mold towards food sources. The subsequent equation delineates this phase:

$$Z = \begin{cases} Z_b + v_b.(W.Z_A - Z_B) r < p \\ v_c.Zr \geq p \end{cases} \tag{11}$$

While 'v_c' falls linearly from 1 to 0, 'v_b' is restricted to the interval $[-a, a]$. 'Z_b' denotes the optimal solutions identified throughout the optimization procedure. Furthermore, 'Z_A' and 'Z_B' represent two solutions chosen randomly from the available set. The parameter 'W' signifies the weight of the slime mold, while 'p' is calculated as follows:

$$p = tanh|S(i) - DF| \tag{12}$$

$S(i)$ indicates the 'Z' solution's fitness values. DF stands for the best fitness value out of all the options. The parameter 'a,' which determines v_b in Equation (1), is calculated in the subsequent manner:

$$a = arctan\ h\left(-\left(\frac{t}{MAX_t}\right) + 1\right) \tag{13}$$

During the optimization process, 't' symbolizes the ongoing iteration, while 'MAX_t' denotes the predetermined maximum iteration number for the algorithm. The determination of the 'W' value in the SMA is accomplished through the following procedure:

$$\overrightarrow{w(smellIndex(i))} = \begin{cases} 1 + r.log\left(\dfrac{bF - S(i)}{bF - wF} + 1\right), & condition \\ 1 - r.log\left(\dfrac{bF - S(i)}{bF - wF} + 1\right), & other \end{cases}$$

$$smellIndex = sort(S) \tag{14}$$

The rank of '$S(i)$' is inside the first half of the population, as shown by the variable '$Cond$'. Furthermore, 'r' indicates a value that is created at random from the interval $[0, 1]$. The variables 'bF' and 'wF' represent the best and worst fitness values within the population solutions, respectively. Moreover, 'smellIndex' stores the fitness values sorted in ascending order, calculated using the 'sort' function applied to the vector 'S.' As such, the sorted fitness values will be in '$SInd$'.

Considering food packaging: In this stage, the SMA simulates the slime mold's position update. Using the following formula, the update is calculated:

$$Z = \begin{cases} rand(UB - LB) + LBrand < z \\ Z_b + v_b.(W.Z_A - Z_B)r < p \\ v_c.Zr \geq p \end{cases} \tag{15}$$

The lower and upper bounds of the exploration space are represented by LB and UB in this equation, respectively. 'r' and 'rand' refer to random variables drawn from a uniform distribution ranging between 0 and 1. Oscillation: During this stage, the variable vb undergoes adjustment within the interval [−a, a], while 'vc' is modified within the range of -1 to 1.

2.2.5 Sparrow search algorithm (SSA)

Xue Jiankai presented the SSA for the first time [24], taking cues from sparrow populations' feeding and anti-predation habits. Being a broad and varied group of birds, sparrows may be found all over the world. They exhibit a notable ability to adapt to human settlements, often coexisting in areas populated by humans. Their diet primarily comprises seeds derived from grains or herbs. Sparrows demonstrate remarkable cognitive abilities and possess a robust memory. The sparrow population exhibits two different behavioral types: scroungers and producers. Whereas scavengers rely on the efforts of producers to survive, producers actively seek food sources. Producers play a crucial part in discovering places rich in food supplies by acting cautiously and selecting foraging sites or routes that are advantageous to all scavengers. Even while a sparrow might become a producer by aggressively looking for the best food sources, the population's proportion of producers to scavengers stays the same. Producers are responsible for identifying and supplying the best food sources for scavengers. The following process is used to update the

producer's position:

$$\begin{cases} X_{i,j}^{t+1} = X_{i,j}^{t}.\exp\left(-\dfrac{1}{\alpha.iter_{\max}}\right) & if\ R_2 < ST\ or \\ X_{i,j}^{t+1} = X_{i,j}^{t} + Q.L & if\ R_2 \geq ST \end{cases} \tag{16}$$

The revised value of the ith sparrow's jth dimension in the following iteration is represented in this formulation as $X_{i.j}^{t+1}$. The word $iter_{max}$ refers to the maximum number of iterations. The parameter α is a uniformly distributed random variable ranging from 0 to 1. The variables R_2 (lying within the range of 0 to 1) and ST (ranging between 0.5 and 1.0) correspond to the values associated with fap (false alarm probability) and the threshold for solidity, respectively. Q is an additional random variable that is used to guarantee a normal distribution. Each member of the one-dimensional matrix L, which has a size of $1 \times d$, is initialized to 1. The sparrow uses a broad search approach when the R_2 value is less than ST, which indicates that there are no predators nearby. On the other hand, all sparrows are required to move quickly (at a velocity indicated by fy) to safer areas if R_2 is greater than or equal to ST, indicating the presence of predators recognized by specific sparrows [25].

2.3 Model Verification and Evaluation

A range of analytical methods and performance measures are used to guarantee the validity of the suggested models. These metrics are designed to identify differences between observed and predicted values by assessing residual errors. The employed metrics encompass MBE, RMSE, MAPE, R^2, VAF, and PCD [26]. The specific mathematical formulations for these statistical measures are provided in Table 2.

3 Results and Discussion

The outcomes of using both individual and hybrid modeling approaches to anticipate gas consumption are shown in the following section. The standalone models examined in this investigation encompass CatBoost, MLP, and XGBoost techniques. Meanwhile, these separate methods are combined in

Table 2 Indicators for statistical validation

Statistics	Criteria	Equation
VAF	Variance Accounted For	$100\% \times \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\mathrm{y_i} - \overline{\mathrm{y}}\right) \left(\mathrm{f_i} - \overline{\mathrm{f}}\right)}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\mathrm{y_i} - \overline{\mathrm{y}}\right)^2}$
RMSE	Root Mean Square Error	$\sqrt{\frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i}\right)^2}{\mathrm{n}}}$
MBE	Mean Bias Error	$\frac{1}{n} \sum_{i=1}^{n} \left(f_i - y_i\right)$
MAPE	Mean Absolute Percentage Error	$\frac{100\%}{N} \sum_{i=0}^{N-1} \frac{\lvert y_i - \hat{y}_i \rvert}{\lvert y_i \rvert}$
R^2	Coefficient of Determination	$1 - \frac{\sum_{i=1}^{n} \left(y_i - \widehat{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}$
PCD	Prediction of Change in Direction	$\frac{1}{n-1} \sum_{i=2}^{n} \mathrm{I}\left(\left(f_i - f_{i-1}\right)\left(y_i - y_{i-1}\right) > 0\right)$

the composite models, which are improved by combining three different optimization techniques: SMA and SSA. The empirical findings arising from these experiments are systematically showcased utilizing various charts, visual representations, and meticulously structured tables. As shown in Figure 2, a thorough construction of a correlation matrix has been carried out, specifically encompassing both the input and output variables within this model.

Detailed discussion and analytical evaluation of the findings will be carried out in later sections of this manuscript, enabling a thorough exploration of the research results. Month, Day, Hour, Outdoor Air Temperature, Outdoor Air Humidity, Horizontal Solar Irradiation, Wind Speed, and Wind Direction are the input factors taken into consideration. Simultaneously, the designated outcome or target variable for this study is gas load (m3). The color gradient represented on the chart spans a range from -0.305 to $+1$. It is imperative to emphasize that negative values indicate an inverse correlation, whereas positive values signify a direct correlation. Upon examination of the visual representation, it becomes apparent that certain parameters, notably Month, Day, and Outdoor air humidity, demonstrate a negative correlation with the target variable. Conversely, the parameter Horizontal solar irradiation exhibits a notable positive association with the target variable. This suggests that increases in these metrics correspond to an increase in the output energy, as inferred from the interpretation of the visual depiction.

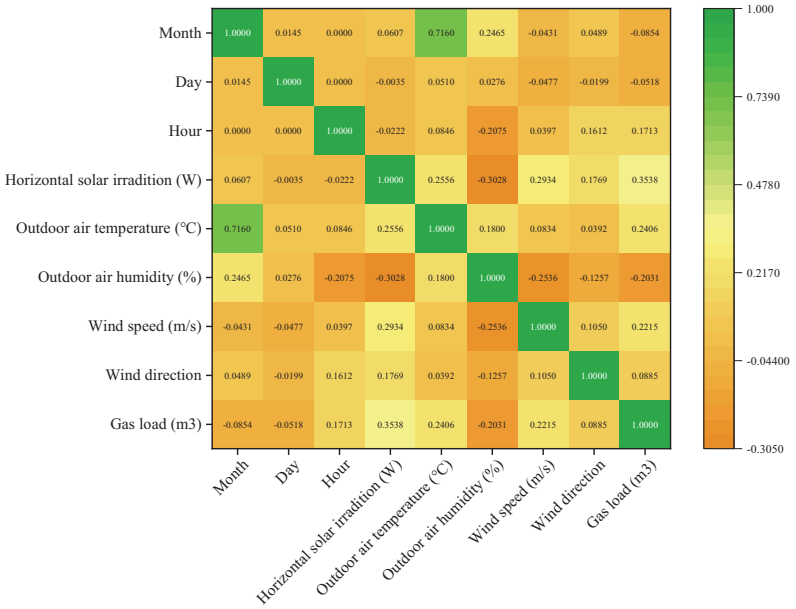

Figure 2 Correlation analysis of dataset attributes.

Figure 3 illustrates the evaluation of input significance and sensitivity regarding model outputs of gas consumption. These graphical representations provide valuable insights into the comparative influence of different model parameters on the overall output. Notably, the parameter "Outdoor air temperature" emerges as the most pivotal factor, displaying the highest sensitivity and exerting a substantial impact. Conversely, parameters such as "Day, Wind direction, Wind speed, and Outdoor air humidity" demonstrate minimal sensitivity, indicating their relatively negligible influence on the model output when compared to other input variables.

Figure 4 offers a thorough overview of the outcomes from each model, including a scatterplot, a chronological data point sequence, and performance indicators for the training and testing stages. The exceptional performance of both models during the training phase is highlighted by the temporal representation in the time series graph, which clearly shows the error rate in red. However, after being exposed to the test dataset, the CatBoost model outperforms the alternative model in terms of performance. A detailed examination of the scatterplot and key statistical metrics, with specific attention to the R^2 depicted in the graphical representation, reveals that the CatBoost algorithm exhibits superior performance, achieving an R^2 value of 0.8431, in contrast to

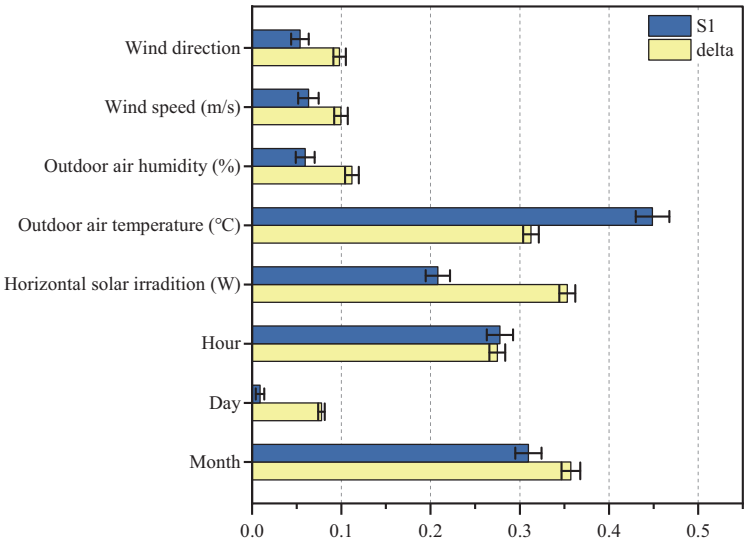

Figure 3 Sensitivity analysis of variables.

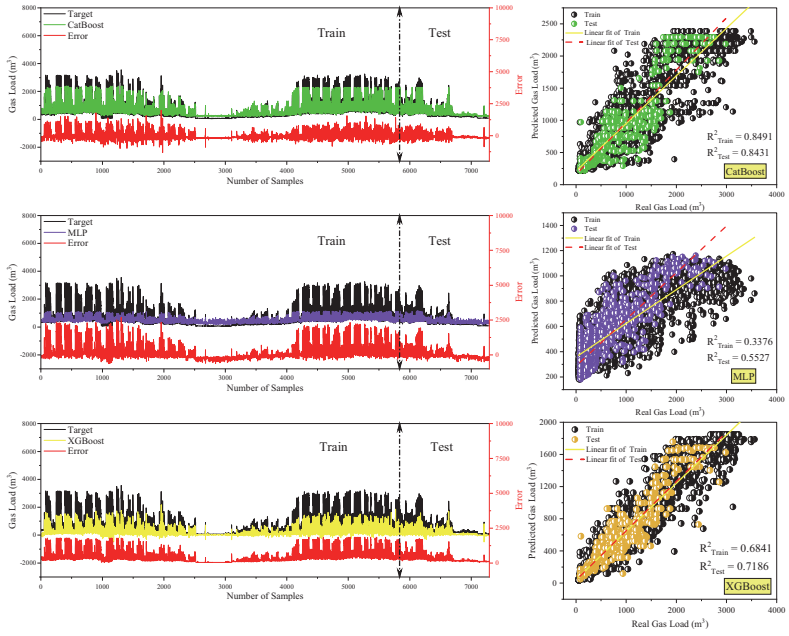

Figure 4 Performance comparison across $CatBoost$, MLP, and $XGBoost$ models.

Table 3 Performance indicators of *CatBoost*, *MLP*, and *XGBoost*

Optimizer	CatBoost	MLP	XGBoost
		Train	
MBE	−0.53152	−239.667	−282.93
RMSE	304.8561	638.6818	441.1456
MAPE	0.556652	0.746783	0.323675
R^2	0.849094	0.337654	0.684006
PCD	0.397326	0.521598	0.480802
VAF	84.90948	43.09222	81.39844
		Test	
MBE	63.04174	−25.8139	−154.865
RMSE	199.1512	336.2808	266.7039
MAPE	0.534423	0.649314	0.284539
R^2	0.84312	0.552693	0.718641
PCD	0.352538	0.509602	0.40535
VAF	85.88401	55.53285	81.35066

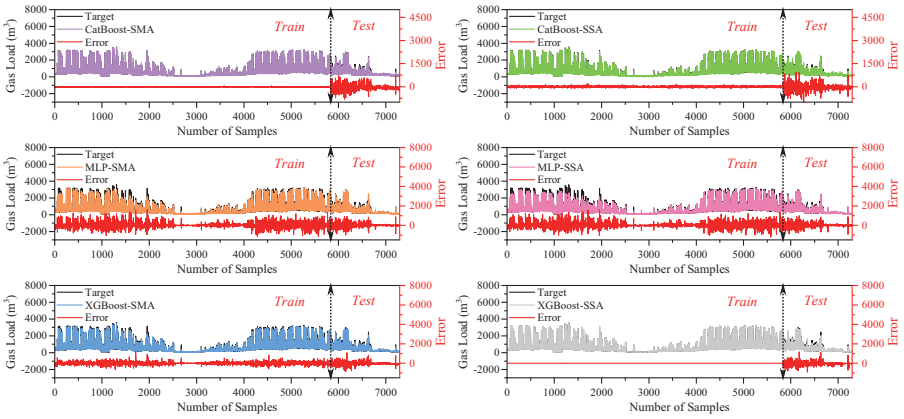

Figure 5 Temporal sequences illustrating the actual and predicted data derived from hybrid models employing CatBoost, MLP, and XGBoost methodologies.

the other models. These results highlight the *CatBoost* algorithm's predictive accuracy and resilience in academic modeling and forecasting contexts. For a detailed overview, Table 3 provides a systematic enumeration of these indices for the models.

SMA and SSA optimizers are added to the time series data of the training and testing sets for the *CatBoost*, MLP, and XGBoost models, as shown in Figure 5. The error rates associated with each hybrid model are depicted

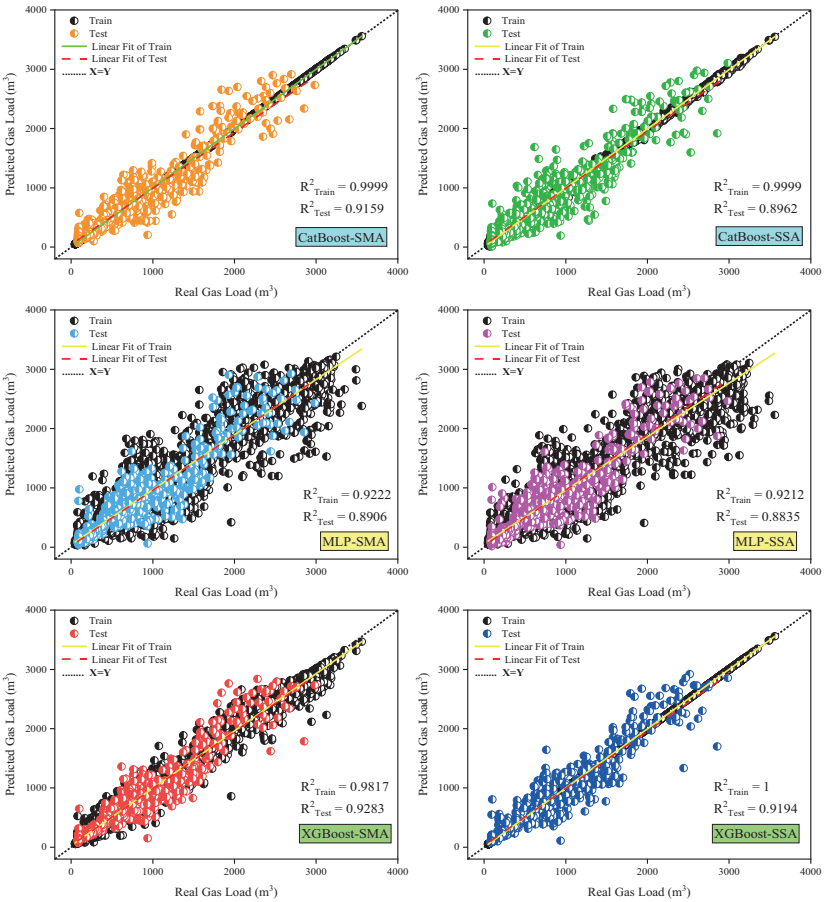

Figure 6 Actual vs. predicted value distribution for hybrid models leveraging CatBoost, MLP, and XGBoost algorithms.

in red. It is clear from the analysis that the hybrid $XGBoost$ models perform better in terms of prediction than the other two models since their error rate on the testing set is significantly lower. Among these, the XGBoost-SMA model stands out for its minimal error. This finding underscores the effectiveness of hybridization, especially when combined with the XGBoost algorithm and the SMA optimization technique, in improving predictive accuracy.

Figure 6 shows scatter plots of hybrid models with the statistical R^2 index for a more thorough analysis of these models. An examination of the graphical displays shows that the test data in the hybrid $XGBoost$ exhibits less

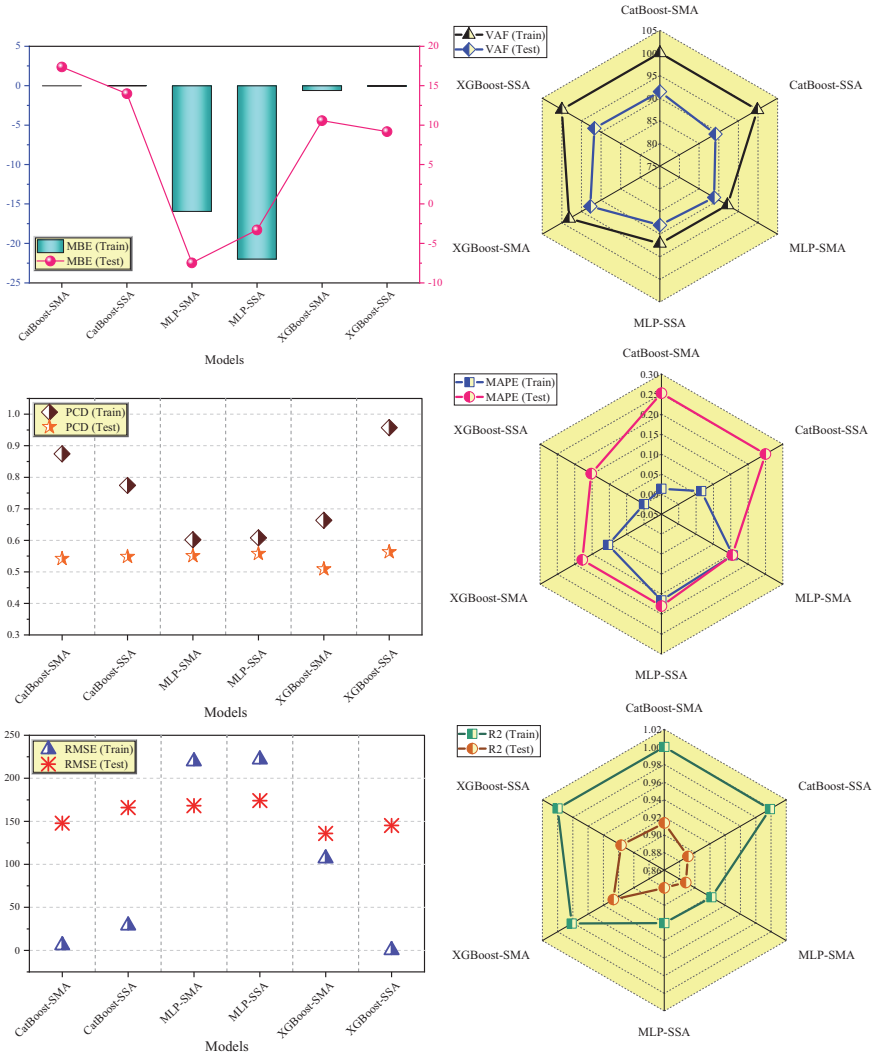

Figure 7 Graphical illustration of performance errors for *CatBoost*-, *MLP*-, and *XGBoost*-based hybrids.

dispersion and greater clustering around the unity line ($x = y$) plot. Notable is the XGBoost-SMA, which has a remarkable R^2 of 0.9283, demonstrating its remarkable prediction accuracy. The importance of parameter tuning for *XGBoost* models' effectiveness cannot be overstated. The XGBoost model's initial R^2 was 0.7186, as was previously mentioned. However, as depicted in

Table 4 Metrics assessing the effectiveness of hybrid models built on $CatBoost$, MLP, and $XGBoost$

Optimizer	CatBoost-SMA	CatBoost-SSA	MLP-SMA	MLP-SSA	XGBoost-SMA	XGBoost-SSA
			Train			
MBE	0.000625	0.014148	−15.9311	−21.9864	−0.60132	−0.05265
RMSE	5.833202	28.94623	219.5785	221.7854	106.8733	0.249309
MAPE	0.013559	0.064912	0.155857	0.165288	0.103647	6.92E-05
R^2	0.999945	0.99864	0.921712	0.92013	0.981454	1
PCD	0.874357	0.774254	0.601988	0.607988	0.663524	0.957491
VAF	99.99448	99.86395	92.21241	92.09153	98.14544	99.99999
			Test			
MBE	17.37555	13.99032	−7.46731	−3.29036	10.5656	9.174911
RMSE	147.787	165.9527	168.1806	174.0652	135.8976	145.0815
MAPE	0.251982	0.250237	0.154578	0.179624	0.178111	0.152188
R^2	0.913608	0.891064	0.88812	0.880153	0.926949	0.916742
PCD	0.541838	0.548011	0.550754	0.557613	0.508916	0.5631
VAF	91.48021	89.18385	88.83402	88.01962	92.73908	91.7075

Figure 6, the refinement of model parameters, along with the incorporation of optimizers, has substantially improved the performance of the XGBoost model. In contrast, while there has been some enhancement in the performance of two other models, the degree of improvement remains relatively moderate. The higher performance of the hybrid XGBoost models in this study may be explained by their capacity to produce excellent results even with fewer datasets.

Metrics of inaccuracy associated with the hybrid models are presented in Figure 7. Analyzing the R^2 and VAF metrics reveal that the hybrid XGBoost models have performed admirably on average during the prediction phase. The MLP-SSA model performs poorly, displaying the lowest R^2 value, while the hybrid XGBoost-SMA model performs better than the others, getting the highest value. The XGBoost-SMA model continuously demonstrates better performance, and this performance trend is consistent across a number of additional statistical indices. Table 4 provides a systematic presentation of the values corresponding to these indices for the hybrid models for a more thorough and in-depth view.

Plots of the results of hybrid models using $CatBoost$, MLP, and $XGBoost$ in the training and testing datasets are shown in Figure 8; in the training dataset, the hybrid CatBoost models have a smaller dispersion and a lower error than the other two hybrid models, with their median line close to zero. Among these hybrid models, the CatBoost-SMA model performs better, but in the testing dataset, the efficiency of the CatBoost models declines, and the hybrid XGBoost models perform better, with the hybrid XGBoost-SMA

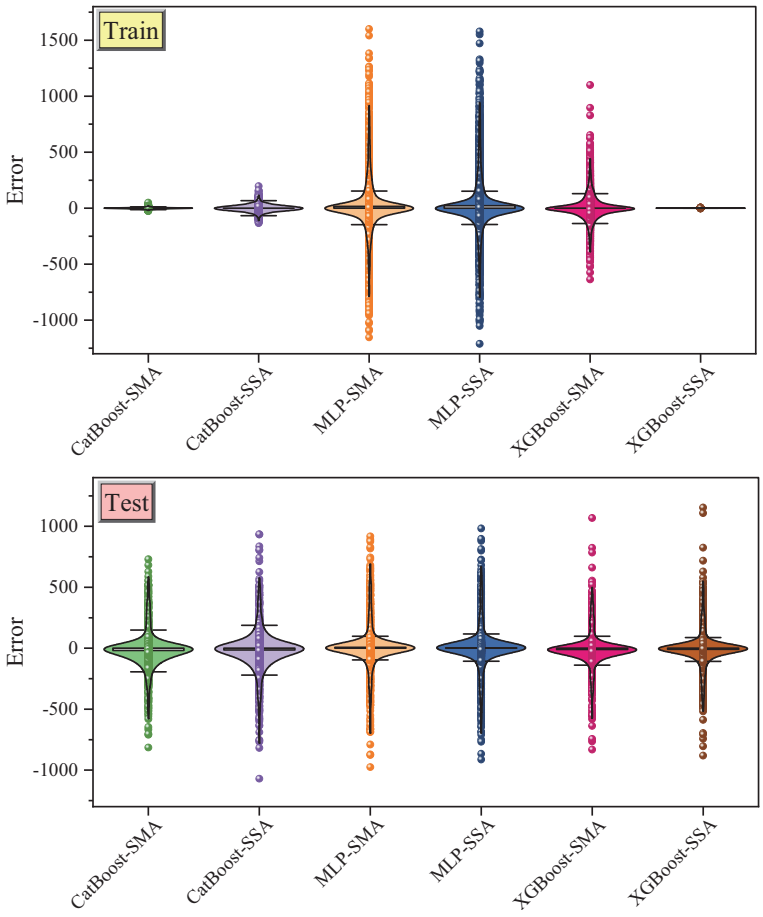

Figure 8 Distribution of error values in hybrid models across both training and testing datasets.

model standing out by showing less dispersion, with its median line near zero. These findings imply a lower error margin, indicating a respectable degree of predictive performance.

In Figure 9, the graphical depiction illustrates the temporal duration of execution for each algorithm across successive iterations. The fluctuations observed in execution time imply that algorithms initially showing temporal variations tend to have lower stability compared to those exhibiting oscillations. However, as multiple iterations progress, these algorithms tend to converge towards a state of relative stability. It is important to note that

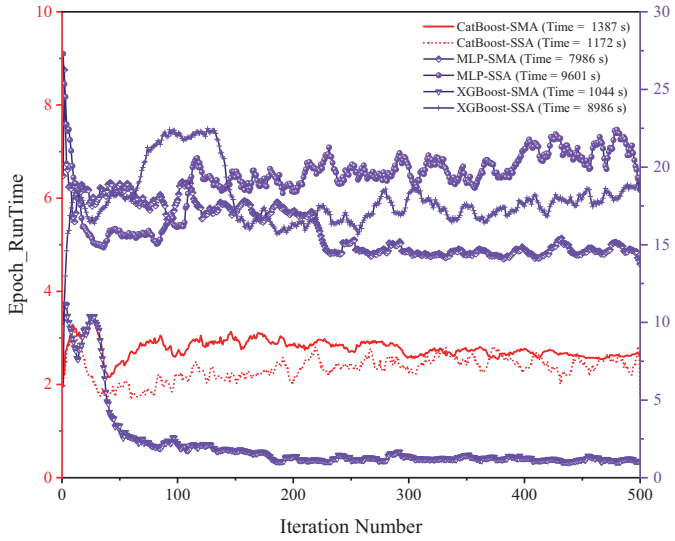

Figure 9 Runtime performance of hybrid models under study.

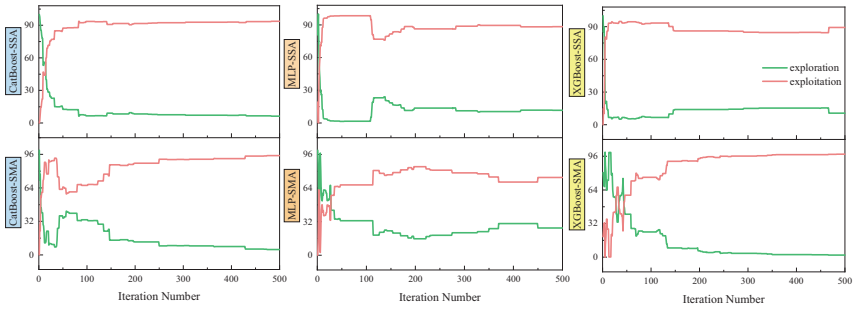

Figure 10 Exploration versus exploitation in hybrid models: An analysis.

the XGBoost-SMA and CatBoost-SSA algorithms adhere to this observed pattern, manifesting greater stability in contrast to other algorithms. This stability attribute serves as an indicative gauge of their proximity to an optimal state.

Figure 10 shows how two metrics – exploitation and exploration – relate to one another and their respective proportions throughout 500 iterations during the hybrid models' exploration phase. The graphical depiction in Figure 10 emphasizes the swift convergence of the hybrid MLP models toward an optimal state. Of the models that were examined, the MLP model that was improved using the SMA approach showed the fastest and most effective

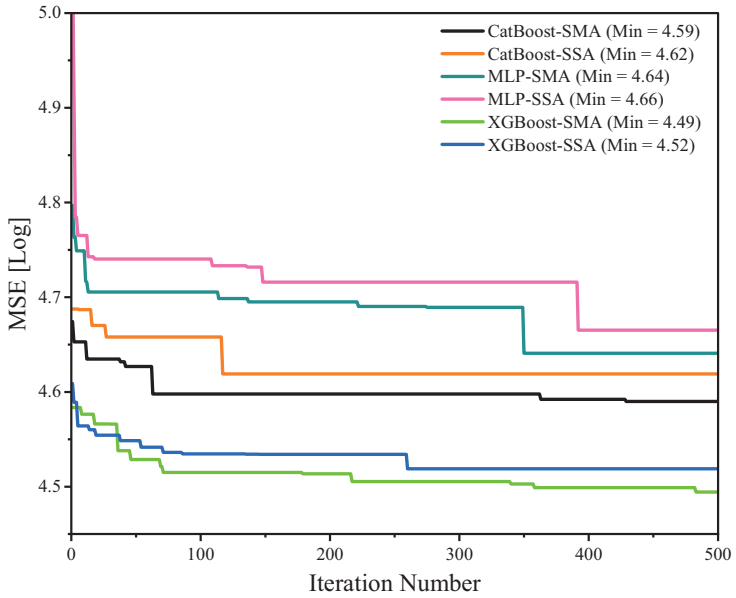

Figure 11 Analysis of convergence behavior in the $CatBoost$, MLP, and $XGBoost$ hybrid models.

progress toward reaching optimality. Additionally, the XGBoost model with SMA optimization also exhibited significant progress.

The convergence graph in Figure 11 shows the state of hybrid model convergence, with special attention to the $CatBoost$, MLP, and $XGBoost$ models. The MSE index is used to assess the degree of convergence. Notably, hybrid XGBoost models exhibit superior convergence performance compared to the other models, as evidenced by notably lower MSE indices. Among these, the XGBoost-SSA and XGBoost-SMA models stand out, displaying the most favorable convergence with the lowest MSE values in this comparison.

4 Conclusion

This study advances the state of gas consumption forecasting by systematically integrating advanced machine learning algorithms with metaheuristic-based hyperparameter optimization. While previous research has predominantly applied single predictive models or relied on conventional statistical approaches, this work demonstrates that hybrid frameworks – most

notably XGBoost combined with SMA and SSA – substantially enhance predictive accuracy and convergence stability. The XGBoost-SMA, with an R^2 value of 0.9283, demonstrates how targeted tuning of hyper-parameters and the integration of an optimizer can increase the performance of the model beyond its baseline (0.7186), thus reducing the gap in accuracy found in prior studies. The same can be said for CatBoost's strong stand-alone performance. Despite the strong performance of CatBoost as a stand-alone model, the comparison analysis does indicate that combining models using optimizers gives us a tangible and strategically significant improvement.

In addition to performance benchmarking, the study provides a useful consideration of methodological trade-offs. The performance stability of the hybrid models, especially the XGBoost-SMA and CatBoost-SSA, reinforces the critical relationship of metaheuristic optimization between exploration and exploitation. Furthermore, the results illustrate that improvements arrived at by such hybridization are not homogeneous – Some hybrids, such as the MLP-SSA hybrid produced only marginal improvements – evidencing that improvements from optimization is model-dependent. Thus, the study contributes to the continued dialogue in the energy forecasting literature around hybridization, which is presumptively assumed to generate improved performance.

In a strategic sense, even beyond accuracy in prediction, better forecasting for gas demand can lead to improved energy distribution, foster good policy development, and might further support the integration of natural gas with renewable energy schemes. In that respect, this will help address global energy transition imperatives by providing tools that enable not only more efficient resource allocation, but also reinforce energy security and lessen environmental impacts.

Nonetheless, the research faces limitations. The dataset, constrained to a single geographical context and one year of operation, restricts generalizability across broader temporal or regional variations. Furthermore, while SSA and SMA proved effective, the study does not exhaustively compare all available metaheuristic optimizers, leaving open the possibility that alternative strategies may yield further gains.

Future studies should take steps to address these deficiencies by incorporating data across multiple regions and time horizons, including exogenous variables such as market dynamics or variability in renewable generation, and evaluating different forms of optimization methods. Study designs including real-time adaptive models could also be examined in order to adjust forecasts dynamically in real time in situations with rapidly changing conditions.

Acknowledgments

2023 Henan Provincial Philosophy and Social Sciences Planning Annual: "Research on the Mechanism and Path of Collaborative Development of Digital and Green Agriculture in Henan Province". (Project Number: 2023BJJ036).

References

[1] J. Essandoh-Yeddu, "Natural gas market," in *Advances in Natural Gas Technology*, IntechOpen, 2012.

[2] J.-J. Her and H.-J. Lim, "An analysis of growth factors on the city-gas industry by input-output structural decomposition analysis," *Journal of Energy Engineering*, vol. 21, no. 2, pp. 158–167, 2012.

[3] S. I. and T. D. R. Center, M. Ishwaran, W. King, M. Haigh, T. Lee, and S. Nie, "The Global Natural Gas Market," *China's Gas Development Strategies*, pp. 345–378, 2017.

[4] L. Baldacci, M. Golfarelli, D. Lombardi, and F. Sami, "Natural gas consumption forecasting for anomaly detection," *Expert Syst Appl*, vol. 62, pp. 190–201, 2016, doi: 10.1016/j.eswa.2016.06.013.

[5] W. Liu, X. Feng, Y. Xu, and X. Gu, "Framework Design and Implementation of Real-Time Monitoring and Dynamic Analyzing System of Urban Gas Network," in *ICPTT 2011: Sustainable Solutions For Water, Sewer, Gas, And Oil Pipelines*, 2011, pp. 295–303.

[6] T. N. Beloglazova and T. N. Romanova, "Effective Implementation of Digital Technologies in the Field of Gas Supply," *Problemele Energeticii Regionale*, vol. 53, no. 1, pp. 141–151, 2022.

[7] B. Soldo, "Forecasting natural gas consumption," *Appl Energy*, vol. 92, pp. 26–37, 2012.

[8] M. Mohebbi and B. Sobhani, "Enhancing Residential Electricity Consumption Forecasting with Meta-Heuristic Algorithms," *Advances in Engineering and Intelligence Systems*, vol. 003, no. 02, pp. 143–175, 2024, doi: 10.22034/aeis.2024.458696.1197.

[9] D. Bajatović, A. S. AnĐelković, I. Ćosić, and R. Maksimović, "Application of predictive models for natural gas needs-current state and future trends review," *Tehnički vjesnik*, vol. 27, no. 2, pp. 648–655, 2020.

[10] J. Liu, S. Wang, N. Wei, X. Chen, H. Xie, and J. Wang, "Natural gas consumption forecasting: A discussion on forecasting history and future challenges," *J Nat Gas Sci Eng*, vol. 90, no. February, p. 103930, 2021, doi: 10.1016/j.jngse.2021.103930.

[11] M. J. Verhulst, "The theory of demand applied to the French gas industry," *Econometrica*, pp. 45–55, 1950.
[12] R. Svoboda, S. Basterrech, J. Kozal, J. Platos, and M. Wozniak, "A Natural Gas Consumption Forecasting System for Continual Learning Scenarios based on Hoeffding Trees with Change Point Detection Mechanism," Aug. 2024.
[13] L. Zhang, X. Ma, H. Zhang, G. Zhang, and P. Zhang, "Multi-Step Ahead Natural Gas Consumption Forecasting Based on a Hybrid Model: Case Studies in The Netherlands and the United Kingdom," *Energies (Basel)*, vol. 15, no. 19, p. 7437, Oct. 2022, doi: 10.3390/en15197437.
[14] H. Khajavi and A. Rastgoo, "Improving the prediction of heating energy consumed at residential buildings using a combination of support vector regression and meta-heuristic algorithms," *Energy*, vol. 272, p. 127069, Jun. 2023, doi: 10.1016/j.energy.2023.127069.
[15] X. Zhou et al., "Towards Universal Large-Scale Foundational Model for Natural Gas Demand Forecasting," Sep. 2024.
[16] J. Zhou, Q. Wang, H. Khajavi, and A. Rastgoo, "Sensitivity analysis and comparative assessment of novel hybridized boosting method for forecasting the power consumption," *Expert Syst Appl*, vol. 249, p. 123631, Sep. 2024, doi: 10.1016/j.eswa.2024.123631.
[17] X. Jin, "Energy Station Data," 2024, doi: 10.6084/m9.figshare.24978645.v1.
[18] Y. Zhang, Z. Zhao, and J. Zheng, "CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China," *J Hydrol (Amst)*, vol. 588, p. 125087, 2020.
[19] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull Math Biophys*, vol. 5, pp. 115–133, 1943.
[20] M. Askari and F. Keynia, "Mid-term electricity load forecasting by a new composite method based on optimal learning MLP algorithm," *IET Generation, Transmission and Distribution*, vol. 14, no. 5, pp. 845–852, 2020, doi: 10.1049/iet-gtd.2019.0797.
[21] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
[22] J. Ma, J. C. P. Cheng, Z. Xu, K. Chen, C. Lin, and F. Jiang, "Identification of the most influential areas for air pollution control using XGBoost and Grid Importance Rank," *J Clean Prod*, vol. 274, p. 122835, 2020.

[23] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: A new method for stochastic optimization," *Future Generation Computer Systems*, vol. 111, pp. 300–323, 2020.
[24] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: sparrow search algorithm," *Systems science & control engineering*, vol. 8, no. 1, pp. 22–34, 2020.
[25] L. Zhang, Y. Chen, and Z. Yan, "Predicting the short-term electricity demand based on the weather variables using a hybrid CatBoost-PPSO model," *Journal of Building Engineering*, vol. 71, p. 106432, Jul. 2023, doi: 10.1016/J.JOBE.2023.106432.
[26] A. Rastgoo and H. Khajavi, "A novel study on forecasting the Airfoil self-noise, using a hybrid model based on the combination of CatBoost and Arithmetic Optimization Algorithm," *Expert Syst Appl*, p. 120576, 2023.

Biographies

Changhao Zhang, male, native place: Luohe, Henan, born in March 1981. He graduated from Zhongnan University of Economics and Law in 2004. He is currently an associate professor in the School of Business Administration of Henan University of Animal Husbandry and Economy. He has long undertaken the teaching tasks of "Marketing", "Marketing Case Analysis" and other courses. He has many years of enterprise practice experience and has rich professional skills and professional training guidance experience for students. His research focuses on digital marketing, digital economy, and green economy.

Mengyu Ren, female, native place: Xuchang, Henan, born in September 1981. She graduated from Henan University of Technology in 2004. She is currently a lecturer in the School of Business Administration of Henan University of Animal Husbandry and Economy. She has long undertaken the teaching tasks of " Business etiquette", "Marketing Case Analysis" and other courses. She has many years of enterprise practice experience and has rich professional skills and professional training guidance experience for students. Her research focuses on digital marketing and Marketing management, etc.