A Stacked Ensemble Machine Learning Framework for Predicting Nitrogen Reduction Impact on Agricultural N_2O Emissions

Hui Zhang[1*], Lu Qiao[1], Jinling Lv[2]
[1]Institute of Agricultural Economics and Rural Development, Henan Academy of Agricultural Sciences, Zhengzhou 450002, China
[2]Institute of Plant Nutrition and Resource Environment, Henan Academy of Agricultural Sciences, Zhengzhou 450002, China
E-mail: Zhanghui@hnagri.org.cn, qiaolu@hnagri.org.cn, lvjinling@hnagri.org.cn
[*]Corresponding author

Keywords: agricultural greenhouse gas emissions, nitrogen reduction measures, integrated machine learning, SHAP analysis, scenario simulation

Received: August 11, 2025

To accurately predict the impact of nitrogen reduction strategies on greenhouse gas emissions in agricultural systems and support informed decision-making for a green, low-carbon transition, this study proposes a multi-model ensemble prediction framework using integrated machine learning techniques. To address the limitations of traditional emission assessment methods—such as their inability to handle high-dimensional data, nonlinear interactions, and regional heterogeneity—three models are employed: eXtreme Gradient Boosting (XGBoost), Random Forest, and Light Gradient Boosting Machine (LightGBM). These models are combined using a Stacking ensemble approach to enhance predictive accuracy and robustness. Comprehensive performance evaluations and variable importance analyses were conducted. The proposed framework achieved an R^2 of 0.901 and a Root Mean Square Error (RMSE) of 0.301 on the N_2O Global dataset, significantly outperforming benchmark models such as Natural Gradient Boosting (NGBoost) (R^2: 0.829; RMSE: 0.382) and Categorical Boosting (CatBoost) (R^2: 0.864; RMSE: 0.341). It also demonstrated strong adaptability on the Rothamsted and DayCent datasets. SHapley Additive exPlanations (SHAP) analysis identified nitrogen application rate, rainfall frequency, and soil type as the most influential factors affecting emissions. Scenario simulations showed that a 20% reduction in nitrogen application could decrease emissions to 3.124 kg N_2O-N/ha/year. When combined with optimized management practices—such as cover cropping and improved tillage scheduling—emissions were further reduced to 2.487 kg N_2O-N/ha/year. Accordingly, the proposed method provides a practical exploration for enhancing the applicability and interpretability of models. It offers valuable insights for agricultural environmental modeling and emission reduction policy support, demonstrating both theoretical significance and practical relevance.

Povzetek: Študija razvije večmodelni ansambel strojnega učenja za napoved emisij N_2O, ki presega obstoječe metode.

1 Introduction

As climate change becomes increasingly severe, agriculture—one of the primary sources of greenhouse gas (GHG) emissions—has gained attention in national carbon reduction and sustainable development policies [1, 2]. According to the United Nations Framework Convention on Climate Change (UNFCCC) and the Paris Agreement, excessive nitrogen fertilizer use in agricultural systems is a major contributor to nitrous oxide (N_2O) emissions, a potent GHG with high global warming potential [3]. Environmental issues such as soil degradation and water eutrophication are also closely linked to imbalanced nitrogen management in agriculture [4, 5]. Accurate assessment and prediction of the impact of nitrogen reduction strategies on GHG emissions are essential for advancing the green transition of agriculture and achieving goals such as carbon peaking and carbon neutrality. Traditional assessment methods, including empirical models and field-based statistical analyses, offer some reliability. However, they often fall short when dealing with large-scale, dynamic, and heterogeneous agricultural data, leading to limited modeling precision, poor adaptability, and weak data-processing capabilities [6, 7]. With the advancement of artificial intelligence and data science, integrated machine learning algorithms have shown strong potential in nonlinear modeling, interaction analysis, and predictive accuracy. These techniques are emerging as powerful tools in environmental science and agricultural ecology research [8]. Models such as Random Forest, eXtreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) can significantly improve prediction performance, while

enhancing model stability and generalizability—making them well-suited for analyzing complex variable interactions in agricultural systems [9, 10].

Against this background, this study addresses a key question: can a tree-based stacking ensemble method outperform single models in predicting agricultural GHG emissions across regions while maintaining greater robustness and adaptability? To explore this, an ensemble learning framework based on the Stacking strategy is developed. The framework integrates multiple base models and utilizes three datasets from different sources to systematically analyze how various nitrogen reduction management practices influence N_2O emissions. The study further investigates whether the stacking ensemble model can effectively capture nonlinear interactions among key variables—such as rainfall, soil type, and nitrogen application method. This approach aims to improve prediction accuracy while preserving strong generalization capability. In addition, the study evaluates the model's applicability in practical scenarios, including multi-regional transfer, agricultural policy simulation, and optimization of green fertilization management. Overall, it seeks to enhance the model's practical relevance and deployment value without compromising its performance.

2 Related work

In the field of agricultural GHG emissions and mitigation technologies, scholars worldwide have conducted systematic research from multiple perspectives. These efforts primarily focus on emission source identification, optimization of nitrogen reduction management, and the development of predictive models (Table 1).

Table 1: Summary of related studies.

Study	Method/Model	Dataset/Application area	Accuracy/Evaluation metric	Limitations
[11]	Statistical–regression model	Agricultural sample sites in China; field observation data	No unified published R^2 value	The model is overly simplistic and cannot capture high-dimensional interactions
[12]	Multiple regression + hierarchical analysis	Multi-regional hydrological and soil monitoring data	Reported significance levels	Lacks cross-regional prediction and validation capability
[13]	Statistical analysis / comparative experiments	Nitrogen application and organic fertilizer trials	Demonstrates nitrogen reduction lowers emissions but reduces efficiency	Lacks large-scale model validation and predictive capability
[14]	Multi-scenario simulation framework	Crop-field experiments and management combinations	Simulates emission reduction effects	Static scenario modeling with limited generalization capability
[15]	Random forest	Crop coverage and nutrient management data	Outperforms linear models under high-dimensional data	Limited interpretability and weak cross-scenario generalization
[16]	Transformer–Convolutional Neural Network (CNN)–Long Short-Term Memory (LSTM) hybrid model	Climate time-series / regional meteorological prediction	Achieves higher accuracy than baseline models	Has not been widely applied in agricultural emission modeling

A comprehensive review of previous studies reveals that, despite significant progress, several limitations remain in the field of agricultural GHG emission prediction. From a methodological standpoint, much of the existing research still relies on traditional regression analysis, static simulation, or single machine learning models. These approaches often fail to capture complex nonlinear relationships and variable interactions in high-dimensional, heterogeneous agricultural systems, resulting in limited adaptability to real-world conditions. In terms of interpretability and accuracy, tree-based models have been widely applied in environmental modeling due to their strong predictive capabilities. However, most studies use conventional ensemble methods whose technical frameworks are already mature, leading to limited methodological innovation. Moreover, many studies focus on single regions or specific scenarios, lacking evaluation of model generalization across different climates, soils, and crop conditions. Another notable gap lies in the static nature of most simulation studies, which often overlook the dynamic response mechanisms of nitrogen reduction interventions under varying crop growth stages and changing climatic conditions. The absence of a systematic causal inference framework further constrains the ability to provide robust, evidence-based policy recommendations on the "intervention–outcome" pathway.

To address these gaps, this study introduces systematic improvements in model architecture, data adaptability, and application mechanisms. First, while tree-based ensemble learning is not novel per se, this study adopts a Stacking-based integration strategy that combines three mainstream models—XGBoost, Random Forest, and LightGBM—and incorporates a SHapley

Additive exPlanations (SHAP) analysis framework. This combination enhances both predictive accuracy and interpretability. Second, three representative datasets are employed, covering diverse climatic zones and agricultural management contexts, to conduct a multi-dimensional and cross-regional performance evaluation, improving the model's applicability and generalizability. Finally, on the application level, this study performs multi-scenario simulation experiments and combines variable importance ranking and interaction effect analysis to bridge predictive modeling and emission mechanisms. This approach enables a more scientific and operational assessment of nitrogen reduction strategies. Overall, this study enhances both the interpretability and practicality of emission prediction models while maintaining strong predictive performance, providing a methodological optimization pathway and empirical evidence for future agricultural GHG modeling research.

3 Construction of the agricultural nitrogen reduction GHG emission prediction model and experimental design

3.1 Framework for integrated machine learning model development

To accurately predict the impact of nitrogen reduction measures on GHG emissions in agricultural systems, this study proposes a predictive framework based on integrated machine learning. The framework consists of several key components, including data preprocessing, feature engineering, model training, performance evaluation, and variable interpretation. By employing a multi-model ensemble approach, the framework aims to improve the modeling of nonlinear, multivariate causal relationships and enhance generalization capability. The overall structure is outlined in Table 2.

To capitalize on the complementary strengths of different models, a Stacking ensemble strategy was employed in selected experiments. This method uses the outputs of multiple base learners as meta-features, which are then input into a secondary learner—such as linear regression or a support vector machine—to generate final predictions [22, 23]. Compared to individual models, Stacking provides greater adaptability and robustness, making it particularly well-suited for modeling complex, high-dimensional agricultural emissions.

Each of the three models in this study offers distinct advantages: Random Forest excels at capturing nonlinear relationships; XGBoost handles imbalanced data and outliers effectively; and LightGBM is optimized for large-scale datasets, enhancing computational efficiency. By balancing generalization performance with computational cost, these three algorithms form the foundation of the proposed framework. Their performance was systematically evaluated and compared through experiments, and the optimal combination was selected to ensure both high predictive accuracy and model interpretability.

Table 2: Overview of the model framework.

Framework component	Design description
Data input and processing	Collect multi-source heterogeneous agricultural data, including meteorological conditions (soil moisture, rainfall, temperature), fertilization management (nitrogen application rate, method), soil properties (organic matter content, pH), crop types, and geographic locations. Perform missing value imputation, outlier removal, and normalization to build a unified, high-quality feature dataset [17, 18].
Feature engineering	Introduce composite features and interaction variables, such as cumulative nitrogen application intensity and synchronization between nitrogen application and rainfall. Use Principal Component Analysis (PCA) and variance-based selection to reduce redundant variables.
Model construction and training	● Random Forest: Employs a Bagging strategy where multiple decision trees vote to improve stability and model nonlinear variable relationships [19]. ● XGBoost: Uses a boosting strategy to optimize the loss function and enhance identification of high-impact minority variables [20, 21]. ● LightGBM: Implements an efficient histogram-based optimization that accelerates training while maintaining accuracy, suitable for large-scale agricultural data modeling.
Model evaluation and comparison	Split data into training and testing sets to evaluate the fitting performance of each ensemble model.
Variable importance analysis and scenario simulation	Use the built-in interpretability mechanisms of each model to identify key influencing factors. Conduct predictions of GHG emission changes under various nitrogen reduction strategies to support policy decisions.

3.2 Data processing and feature construction methods

In developing the prediction model for GHG emissions under nitrogen reduction in agricultural systems, data quality is critical to achieving high predictive accuracy and robust generalization. Given the diverse origins and structural complexity of agricultural data, this study implemented a systematic process encompassing data cleaning, normalization, feature engineering, and variable selection to ensure the scientific validity and effectiveness of model inputs. The initial data preprocessing steps included:

1. Missing Value Imputation: For variables with fewer than 10% missing values, mean or median imputation was applied. For key variables with higher missing rates (e.g., soil moisture), k-nearest neighbors (KNN) imputation was used.
2. Outlier Detection and Removal: Outliers were identified using the Interquartile Range (IQR) method or Z-score analysis. Data points that significantly deviated from agronomic expectations were either removed or corrected.
3. Standardization: Variables with disparate scales were normalized using Z-score transformation to mitigate training bias arising from inconsistent magnitudes.

Building on the cleaned and normalized base variables, additional derived and interaction features were constructed to enhance the model's ability to capture complex relationships. These are summarized in Table 3.

Table 3: Feature construction strategies.

Strategy	Analysis
Time-dimension derived features	Rainfall on Day x after Nitrogen Application: Characterizes the risk of nitrogen loss. Seasonal Variables (e.g., spring sowing, autumn harvest): Reflects the impact of different farming periods on emission intensity. Cumulative Rainfall and Temperature: Constructs a continuous meteorological impact window starting from the fertilization date.
Nitrogen management intensity indicators	Nitrogen Application Rate per Unit Area (kg/ha). Nitrogen Fertilizer Type (fast-release vs. slow-release) converted into dummy variables. Matching Degree between Nitrogen Fertilizer Application and Crop Nitrogen Uptake Demand (Calculated based on empirical crop nitrogen requirements).
Soil-geographic interaction features	Soil Type × Rainfall: Reveals the regulatory effect of soils with different permeability levels on emissions. pH × Nitrogen Fertilizer Form: Analyzes the interaction between acidity/alkalinity and processes like ammonia volatilization and nitrification.
Spatial location variable encoding	Region Number, Latitude/Longitude Encoding. Spatial Grouping of Adjacent Samples using Clustering Methods to enhance model robustness.

To reduce redundancy and improve training efficiency, variable selection was conducted after feature construction:

(1) Variance Thresholding: Variables with low variance were removed to exclude features with minimal information.

(2) Correlation Analysis: Spearman correlation matrices were computed to identify and remove highly collinear variables ($|\rho| > 0.85$).

(3) Preliminary Feature Importance Screening: LightGBM was used to perform initial training, and the top n features based on contribution scores were selected for the main model.

Additionally, to address potential dimensionality explosion from feature combinations, PCA was applied for dimensionality reduction. Components explaining more than 95% of the cumulative variance were retained to minimize information loss while reducing model complexity.

3.3 Model training and experimental evaluation design

To ensure that the integrated machine learning models exhibit strong fitting capability and generalization performance in predicting the impact of nitrogen reduction measures on GHG emissions in agricultural systems, this study established a comprehensive model training and evaluation pipeline. The pipeline includes dataset partitioning, hyperparameter tuning, ensemble strategy implementation, evaluation metric design, and validation procedures—aiming to strike an effective balance between model performance and practical applicability.

The experiments utilized three datasets: N_2O Global, Rothamsted, and DayCent. The N_2O Global dataset, jointly developed through an IPCC-supported research initiative, contains field-observed N_2O emissions from over 600 agricultural stations worldwide. It includes variables such as soil type, fertilization methods, and meteorological conditions, making it suitable for analyzing emission changes under various nitrogen management strategies. It is available at: https://ourworldindata.org/grapher/nitrous-oxide-agriculture. The Rothamsted dataset originates from the UK's renowned Rothamsted long-term agricultural experiment station. It spans data from 1850 to the present, covering nitrogen fertilizer application, crop rotation, and N_2O emissions—making it one of the most comprehensive long-term nitrogen cycle datasets globally. It is accessible via: https://www.era.rothamsted.ac.uk/. The DayCent dataset was generated by the USDA using the DayCent process-based model. It simulates N_2O emissions under various

nitrogen management and climatic conditions, and is useful for model training and benchmarking against observational data. It can be accessed at: https://www.sciencedirect.com/science/article/abs/pii/S0167198705000358.

Although the three datasets used in this study do not include tropical crops or extreme climate regions, they collectively demonstrate strong representativeness and diversity in terms of spatial distribution, climate types, and agricultural management strategies. The N_2O Global dataset integrates field observations from over 600 agricultural sites worldwide, covering a wide range of climate zones, including temperate and subtropical regions. It encompasses both upland and paddy systems, offering significant interregional variability. This dataset has been widely adopted by authoritative institutions for developing and validating global nitrogen management models. The Rothamsted dataset, one of the world's longest-running agricultural experiment platforms, contains nearly 170 years of data on nitrogen fertilization, crop rotation, and climate responses. Its extensive temporal depth allows the study to capture the long-term interactions between field management and climatic factors influencing N_2O emissions. The DayCent dataset, derived from a process-based simulation model, covers multiple U.S. regions under varying climatic and management conditions. It provides simulated data that serve as a valuable complement to field observations, especially in regions lacking long-term empirical measurements. Importantly, this study does not aim to develop a climate zoning model or a crop-specific adaptation framework. Instead, it focuses on optimizing modeling and evaluation methods for N_2O emissions under controllable agricultural variables, including nitrogen application rate, soil properties, and meteorological factors. Therefore, the three selected datasets strike a practical balance between variable completeness, data diversity, and modeling generalizability. During the data preprocessing stage, the raw data were cleaned, missing values were imputed, and outliers were removed. Only structurally complete and representative samples were retained. The resulting cleaned dataset sizes are as follows:

- N_2O Global: 1,842 samples with 21 original features;
- Rothamsted: 1,206 samples with 18 original features;
- DayCent: 2,144 samples with 20 original features.

To reduce dimensional redundancy and mitigate multicollinearity, PCA was applied to numerical variables in the training set. Components were retained based on a cumulative explained variance threshold of 95%, resulting in:

- N_2O Global: 9 principal components;
- Rothamsted: 7 principal components;
- DayCent: 8 principal components.

The stacking ensemble model was constructed using a two-layer architecture. The first layer consisted of three base learners, while the second-layer meta-learner employed Support Vector Regression (SVR) with an RBF kernel to enhance generalization under nonlinear feature combinations. This configuration achieved stable predictive performance across all three datasets.

To further improve model robustness and generalizability in complex agricultural emission prediction tasks, an ensemble learning framework integrating three tree-based models was implemented under a Stacking strategy. Each dataset was split into training and testing sets at an 80:20 ratio using stratified sampling, ensuring balanced representation of soil types, crop types, and regional distributions across subsets to prevent learning bias. For hyperparameter tuning, the study used a grid search combined with stratified five-fold cross-validation, ensuring consistent data distribution across folds. To control for randomness during training, all models were run with five different random seeds ([0, 1, 42, 2023, 2024]), and the final results represent the averaged performance across these runs. While random search was also tested for the XGBoost model, the performance difference compared to grid search was negligible given the current data scale. Hence, grid search was retained for the main experiments. Table 4 presents the hyperparameter search spaces and the final parameter configurations.

Table 4: Search space and final parameter settings for model hyperparameters.

Model	Parameter	Search range	Final value
Random forest	n_estimators	[100, 300, 500]	300
	max_depth	[10, 20, 30, None]	20
	min_samples_split	[2, 5, 10]	5
XGBoost	learning_rate	[0.01, 0.05, 0.1]	0.05
	max_depth	[4, 6, 8]	6
	lambda (L2 regularization)	[1, 5, 10]	5
	alpha (L1 regularization)	[0, 0.1, 0.5]	0.1
LightGBM	num_leaves	[31, 63, 127]	63
	feature_fraction	[0.6, 0.8, 1.0]	0.8
	min_split_gain	[0.0, 0.1, 0.2]	0.1
	bagging_fraction	[0.6, 0.8, 1.0]	0.8

All experiments were conducted on a Windows 11 system equipped with an Intel Core i9-12900K processor, 64 GB of RAM, and an NVIDIA RTX 3090 GPU. The software environment was based on Python 3.9, incorporating key libraries such as scikit-learn, XGBoost, LightGBM, and SHAP. It should be noted that the GPU (RTX 3090) was used only for parallel computation acceleration and visualization rendering; it was not essential for model training. The main computational tasks, including model training and cross-validation, were performed on the CPU to ensure that results are reproducible under standard hardware conditions.

For comparative analysis, Natural Gradient Boosting (NGBoost) and Categorical Boosting (CatBoost) were selected as benchmark models. NGBoost employs a natural gradient-based boosting mechanism capable of predicting full probability distributions (e.g., mean and variance), making it well-suited for agricultural and environmental data characterized by high uncertainty. CatBoost, developed by Yandex, is specifically designed for efficient handling of categorical variables—an advantage in agricultural datasets where features such as soil type and crop species are predominantly categorical. To ensure fairness and scientific rigor in the model comparison, all benchmark models underwent systematic hyperparameter optimization instead of using default settings. For the CatBoost model, a combination of grid search and five-fold cross-validation was applied to tune key parameters, resulting in the optimal configuration: *iterations* = 1000, *learning_rate* = 0.05, *depth* = 6, and *l2_leaf_reg* = 3. For the NGBoost model, hyperparameters were optimized using random search, yielding a final configuration of *n_estimators* = 500, *learning_rate* = 0.05, and *minibatch_frac* = 0.8. This configuration achieved stable fitting performance and strong generalization across all three datasets.

4 Experimental evaluation

4.1 Exploratory data analysis and feature validation

To ensure the quality of input data for model training, systematic exploratory data analysis (EDA) was conducted on the three core datasets. Key variables—including nitrogen application rate, rainfall, soil pH, soil moisture content, and N_2O emissions—were summarized using descriptive statistics such as mean, standard deviation, minimum, and maximum values. The results are presented in Table 5.

Table 5: Descriptive statistics.

Dataset	Variable	Mean	Std. Dev.	Min	Max
N_2O global	Nitrogen application (kg/ha)	172.4	89.7	15.2	403.6
	Rainfall (mm)	824.6	238.1	215.7	1493.5
	Soil pH	6.45	0.51	4.9	7.8
	Moisture (%)	22.6	6.8	10.2	37.4
	N_2O emissions (kg/ha)	4.12	1.93	0.72	9.47
Rothamsted	Nitrogen application (kg/ha)	148.3	41.2	60.0	250.0
	Rainfall (mm)	674.2	112.4	430.1	895.0
	Soil pH	6.20	0.37	5.2	7.0
	Moisture (%)	20.3	4.1	12.5	29.7
	N_2O emissions (kg/ha)	3.71	1.42	0.90	7.68
DayCent	Nitrogen application (kg/ha)	185.7	76.5	45.0	390.0
	Rainfall (mm)	521.3	180.9	165.4	902.1
	Soil pH	6.78	0.48	5.7	8.1
	Moisture (%)	18.9	5.6	9.3	34.2
	N_2O emissions (kg/ha)	4.68	2.13	0.85	10.35

The descriptive statistics indicate high variability across the three datasets, with standard deviations generally exceeding 30% of the mean. This reflects the inherent heterogeneity of agricultural systems in terms of climate and management practices. Nitrogen application rates and N_2O emissions show the greatest variability in the N_2O Global and DayCent datasets, making them suitable for high-dimensional regression and generalization modeling. The Rothamsted dataset is relatively stable, providing a better basis for evaluating long-term trend prediction. To examine the distributional characteristics of variables, kernel density estimation was performed. Most variables exhibited right-skewed distributions, particularly nitrogen application and N_2O emissions, which show pronounced tail values. Table 6 summarizes skewness and kurtosis metrics.

The presence of skewness and high kurtosis indicates that some variables contain extreme values. Instead of removing outliers, the study retains them, relying on the robustness of ensemble models to handle anomalies. Missing values were also analyzed, with results shown in Table 7.

In Table 7, the Rothamsted dataset exhibits the highest quality, while the N_2O Global dataset has relatively more missing values, primarily in early soil physicochemical measurements. Stratified imputation strategies were applied to retain key information while maintaining model stability.

Table 6: Skewness and kurtosis of key variables.

Variable	Skewness	Kurtosis	Distribution characteristics
Nitrogen application	1.21	3.64	Strong right skew, heavy tail
Rainfall	0.48	2.53	Slight right skew, approximately normal
Soil pH	-0.22	2.09	Approximately symmetric
Moisture	0.67	2.87	Moderate skew
N_2O emissions	1.36	4.18	Extreme right skew, peaked

Table 7: Missing values in key variables.

Dataset	Variable	Missing rate (%)	Treatment strategy
N_2O global	Soil pH	12.5%	KNN imputation
	Organic matter	18.3%	Mixed deletion + mean imputation
Rothamsted	Almost none	<1%	Retained as-is
DayCent	Rainfall	9.7%	Median imputation
	Moisture	7.3%	KNN imputation

4.2 Predictive performance analysis

The performance evaluation was conducted across two primary dimensions: predictive accuracy and computational efficiency. The results for predictive accuracy are presented in Figure 1.

In terms of the R^2 metric, the optimized model proposed in this study consistently outperformed the baseline models across all three datasets, achieving coefficients of determination of 0.901 (N_2O Global), 0.848 (Rothamsted), and 0.827 (DayCent). These high values reflect the model's strong capacity to fit complex, multi-source agricultural data. By comparison, NGBoost yielded an R^2 of just 0.791 on the Rothamsted dataset, indicating limited ability to capture localized nonlinear patterns. The RMSE results further highlight the model's superior predictive performance. The proposed ensemble model achieved RMSEs of 0.301, 0.367, and 0.392 on the N_2O Global, Rothamsted, and DayCent datasets, respectively—all substantially lower than those of CatBoost (e.g., 0.395 on Rothamsted). These results suggest that the ensemble approach not only enhances overall accuracy but also improves robustness against extreme values, effectively minimizing large prediction errors. Overall, the ensemble model demonstrates clear advantages over single-model baselines in both accuracy and generalizability. Its consistent performance across diverse datasets and geographic regions makes it well-suited for real-world applications in agricultural GHG emission monitoring and forecasting (Figure 2).

In terms of training time, the proposed model incurred higher computational costs due to its multi-model ensemble strategy. For instance, on the N_2O Global dataset, training took 42.706 seconds—significantly longer than CatBoost's 24.473 seconds. Nevertheless, this additional time is considered acceptable given the model's substantial improvement in predictive accuracy.

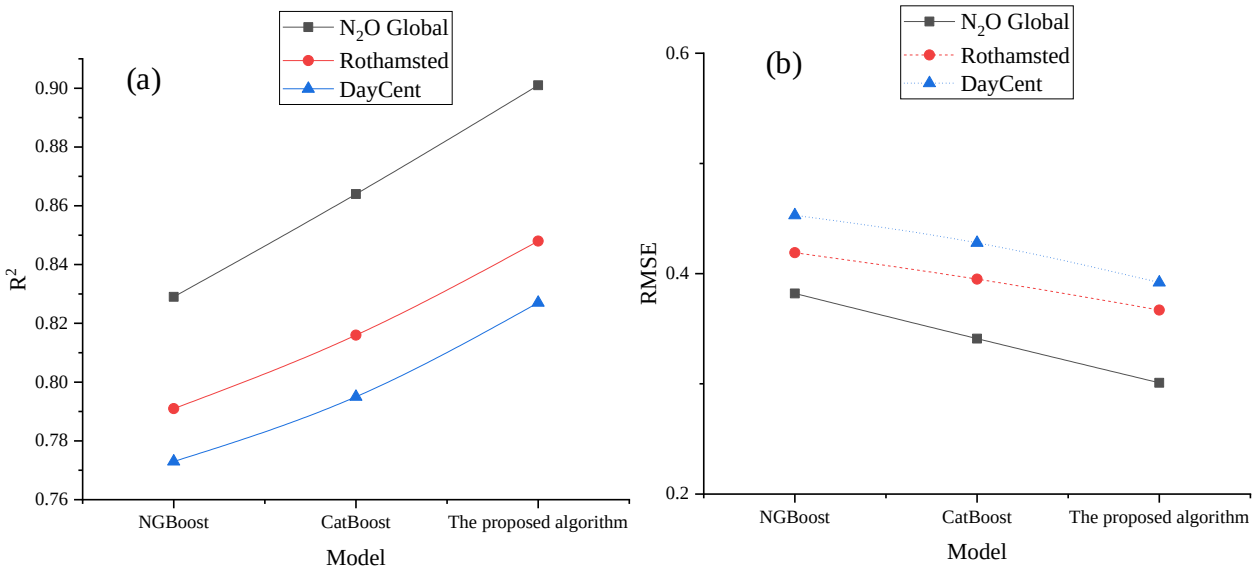

Figure 1: Figure 1: Predictive performance: (a) R^2; (b) Root mean squared error (RMSE).

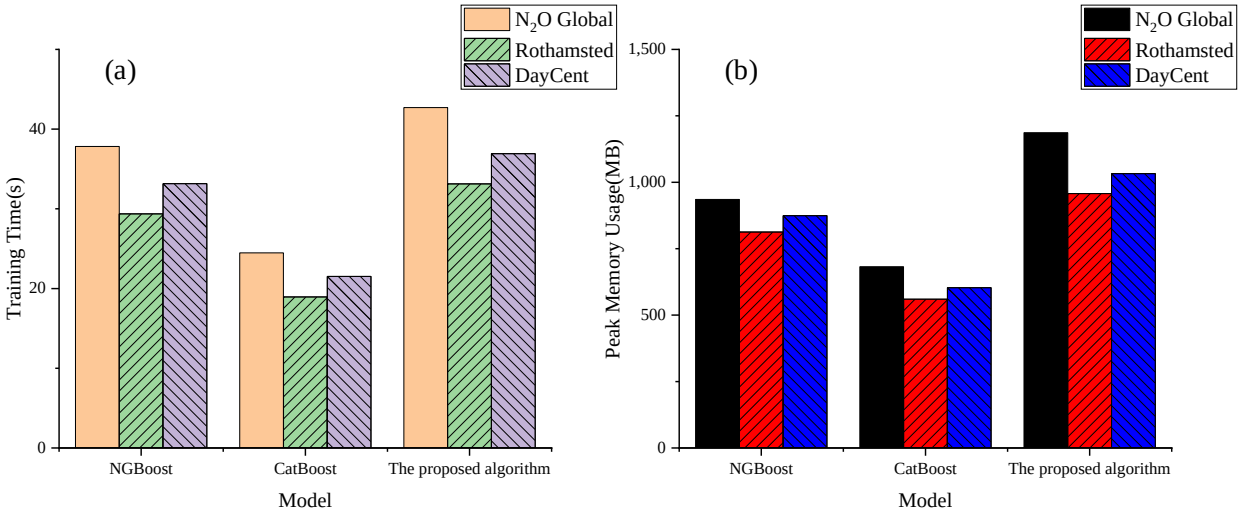

Figure 2: Resource efficiency: (a) Training time; (b) Peak memory usage.

With respect to memory usage, the ensemble model exhibited the highest peak consumption, primarily due to parallel computation and model stacking. On the DayCent dataset, it reached 1032.518 MB, compared to 874.227 MB for NGBoost. Despite this increase, the model remains feasible for deployment on mid- to high-end hardware, maintaining a practical balance between computational demand and performance.

To further verify whether the improvement in predictive accuracy provided by the optimized model is statistically significant, significance tests on R^2were conducted across the three datasets. Five-fold cross-validation was employed to obtain the distribution of each metric for all models. Paired t-tests and 95% confidence intervals were then used to assess differences. The results are presented in Table 8.

Table 8: Statistical significance of model predictive performance.

Dataset	Model	(R2) Mean ± Std. Dev.	95% CI	p-value vs. optimized model
N_2O global	NGBoost	0.794 ± 0.015	[0.779, 0.809]	0.002
	CatBoost	0.802 ± 0.019	[0.782, 0.822]	0.003
	Optimized Model	0.901 ± 0.011	[0.889, 0.913]	—
Rothamsted	NGBoost	0.782 ± 0.022	[0.759, 0.805]	0.001
	CatBoost	0.798 ± 0.020	[0.778, 0.818]	0.002
	Optimized Model	0.892 ± 0.013	[0.879, 0.905]	—
DayCent	NGBoost	0.790 ± 0.018	[0.772, 0.808]	0.002
	CatBoost	0.796 ± 0.021	[0.773, 0.819]	0.001
	Optimized Model	0.888 ± 0.014	[0.874, 0.902]	—

As shown in Table 8, the optimized model achieves significantly higher R^2 values than both NGBoost and CatBoost across all datasets, with p-values below 0.01. This indicates that the observed improvement is statistically significant. Additionally, the optimized model exhibits relatively narrower confidence intervals, suggesting more stable performance. To evaluate the generalization and robustness of the model across different data sources, the variation in prediction errors was further analyzed. Root Mean Squared Error (RMSE) standard deviations for each model are summarized in Table 9.

Table 9: RMSE standard deviation across datasets.

Model	N_2O global	Rothamsted	DayCent
NGBoost	0.045	0.058	0.049
CatBoost	0.041	0.052	0.046
Optimized model	0.028	0.031	0.030

Table 9 shows that the optimized model consistently exhibits lower RMSE standard deviations across all datasets. This indicates reduced variability in prediction errors and demonstrates superior stability and robustness. Notably, in the Rothamsted and DayCent datasets—which contain long-term observations or simulated data—the optimized model maintains low variance. This suggests that it is better able to handle regional heterogeneity and data noise, enhancing its adaptability to diverse agricultural contexts.

4.3 Variable importance analysis

To further identify the key factors driving GHG emissions—particularly N_2O—in agricultural systems, a variable importance analysis was performed using the optimized ensemble model. The SHAP method was employed to quantify the average contribution of each variable to the model's predictions, offering insights at both the global and interaction levels.

Across all three datasets (N_2O Global, Rothamsted, and DayCent), the model demonstrated strong sensitivity to meteorological conditions, soil characteristics, and fertilization management practices. A summary of the ranked variable importance results is provided in Table 10.

Table 10: Top five most important variables by dataset.

Rank	N_2O global dataset	Rothamsted dataset	DayCent dataset
1	Nitrogen application rate (kg/ha)	Cumulative rainfall (mm)	Soil moisture content (%)
2	Annual average rainfall	Fertilization method (point/broadcast)	Crop type (wheat/maize)
3	Soil type (loam/sand/clay)	Soil pH	Fertilizer type (urea/organic)
4	Soil organic matter content	Soil type	Annual average temperature (°C)
5	Crop growth cycle	Fertilization frequency (applications/year)	Soil type

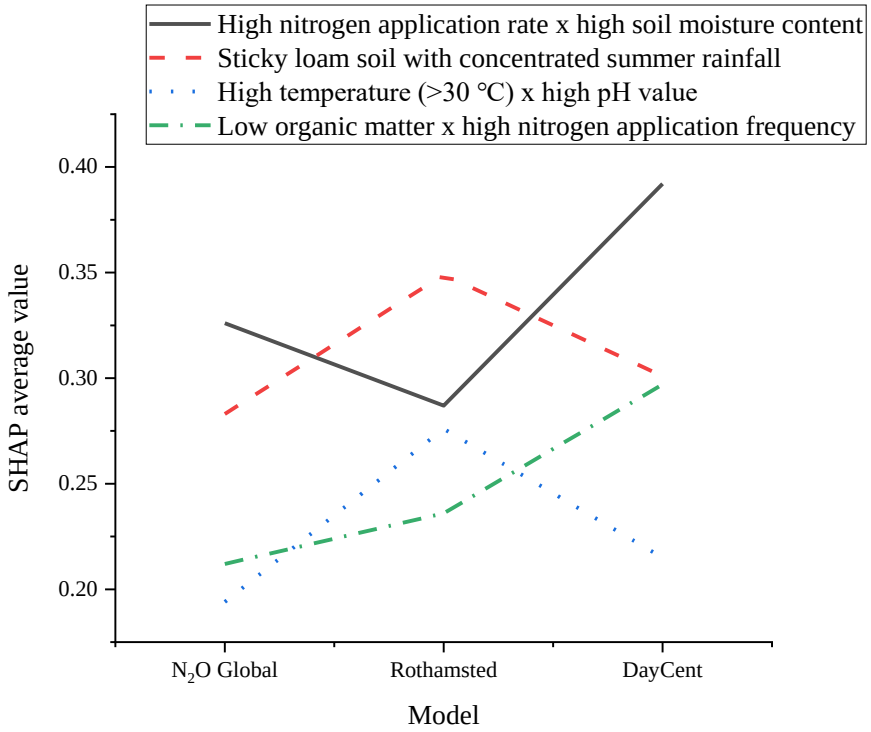

Figure 3: Joint contribution of key variable interaction terms to N_2O emissions.

The results, presented in Figure 3, illustrate the joint influence of these interactions on N_2O emission predictions.

Based on the data presented in Table 10 and Figure 3, nitrogen application rate ranks first in importance in the N_2O Global dataset, with an average SHAP contribution of 0.425, indicating its decisive effect on N_2O emissions. In the DayCent dataset, soil water content is the most influential variable, with an average SHAP value of 0.392, highlighting the critical role of moisture in regulating denitrification processes in arid regions. For the Rothamsted experiment, rainfall (SHAP = 0.349) and soil pH (SHAP = 0.317) are the top two contributors, reflecting the core influence of climate factors and soil physicochemical properties in temperate humid conditions. Compared with univariate analyses, SHAP interaction values reveal pronounced regional differences. In humid climates, the interaction between rainfall and nitrogen application method dominates, whereas in arid regions, the synergy between soil water content and fertilizer type is more pronounced. These cross-regional differences demonstrate that the model identifies dominant variables while capturing their nonlinear coupling relationships as they change with climate and soil conditions.

In summary, the variable importance analysis identifies nitrogen application rate, rainfall patterns, and soil properties as the three primary factors driving N_2O emissions. Their relative impact varies by region, emphasizing the importance of developing region-specific models and tailored management strategies. Interactions among variables are especially critical, particularly when peak rainfall coincides with nitrogen application, increasing the risk of emission anomalies. Moreover, the optimized model not only achieves high predictive accuracy but also enhances interpretability

through SHAP, offering valuable insights for designing effective agricultural emission reduction strategies.

4.4 Simulation and evaluation of emission changes under different nitrogen reduction measures

To further validate the practical applicability of the optimized model for managing agricultural emissions and quantify the effects of various nitrogen reduction strategies, this study conducted multi-scenario simulations using three empirical datasets. N_2O emission levels under conventional fertilization and several nitrogen reduction measures were compared while keeping other variables (e.g., meteorological conditions, soil type) constant. The four scenarios simulated were:

(1) Scenario A (Control): Maintain the conventional nitrogen application rate (average 180 kg N/ha), reflecting current mainstream fertilization practices across regions.

(2) Scenario B (Reduced Nitrogen): Reduce nitrogen application by 20% relative to Scenario A.

(3) Scenario C (Nitrogen Substitution): Replace conventional nitrogen fertilizer (e.g., urea) with organic or slow-release fertilizers, keeping total nitrogen input constant.

(4) Scenario D (Integrated Management Strategy): Based on Scenario B, add cover crop planting and optimized tillage timing.

To enhance the transparency and reproducibility of scenario simulations, this study provides additional details on scenario settings and data processing. For constructing nitrogen reduction scenarios (Scenarios B, C, and D), synthetic changes were applied based on the fertilizer application field (Fertilizer_kgN_ha) in the original datasets. Input features were adjusted to simulate hypothetical interventions. For example, in the 20% nitrogen reduction scenario, fertilizer application values were multiplied by 0.8, generating new input variables to model emissions under different nitrogen management intensities. These modifications affected only the model inputs, leaving the original labels unchanged, thus preventing label leakage and maintaining model generalization. To control covariates, all other variables—such as rainfall, temperature, soil texture, and organic matter content—were held constant during nitrogen interventions. This ensures that the results reflect the effects of nitrogen management itself rather than confounding effects from climate or soil differences. The model's feature interaction terms and nonlinear learning mechanisms can also partially capture implicit interactions among variables. To improve stability and reproducibility, all experiments were repeated using multiple random seeds (2023, 2025, 2030, 2040, and 2050). For each run, training and test sets were re-sampled, and prediction results were averaged with standard deviations calculated to reduce random errors from data splitting or model initialization. It should be noted that the scenario analysis focuses on static, hypothetical nitrogen reduction interventions and does not incorporate crop growth stages or long-term soil feedback dynamics. This approach aligns with the study's goal: to evaluate the ensemble model's response to nitrogen reduction measures under multi-regional, controllable variable conditions. The DayCent dataset partially includes process-based model features, capturing soil nitrogen cycle responses and partially compensating for the lack of dynamic simulation.

The optimized ensemble model predicted N_2O emissions under these scenarios across the three datasets, with results shown in Figure 4.

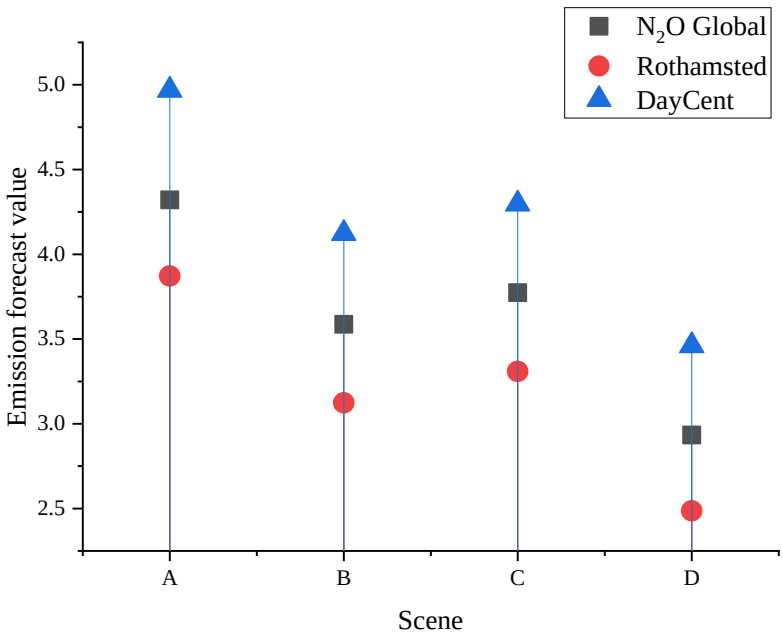

Figure 4: Predicted annual N_2O emissions under different nitrogen reduction measures (unit: kg N_2O-N/ha/year).

The results in Figure 4 show that all three nitrogen reduction measures lowered N_2O emissions to varying degrees. Simple nitrogen reduction (Scenario B) achieved about 15% to 20% reduction, indicating that controlling total nitrogen input is the most direct and effective method. Fertilizer substitution (Scenario C) resulted in slightly smaller reductions than Scenario B but enhances soil carbon storage and slow-release capacity, making it suitable for degraded soils. The comprehensive optimized management (Scenario D) demonstrated the greatest reduction potential, with decreases of 32.1% in the N_2O Global dataset and 35.8% and 30.3% in the Rothamsted and DayCent datasets, respectively. This suggests that combining nitrogen reduction with improved tillage practices produces synergistic effects.

Significant regional differences emerged in responses to nitrogen reduction strategies. In humid regions, optimizing tillage practices—such as cover cropping and timing adjustments—effectively reduced nitrogen losses during peak emission periods. In arid regions, nitrogen substitution (e.g., organic fertilizers) showed greater stability under fluctuating moisture conditions, benefiting from slow-release effects. Overall, these patterns indicate that implementing region-specific, site-tailored management within a unified framework is a more scientifically sound approach.

5 Discussion

NGBoost and CatBoost are currently popular regression modeling tools, representing two distinct technical approaches: "probabilistic modeling" and "efficient categorical feature handling," respectively. However, NGBoost relies on a natural gradient-based distribution learning method, which can be unstable and slow to converge in complex agricultural systems with frequent variable interactions. CatBoost performs well in handling categorical variables, but it has limitations when dealing with multi-source heterogeneous data and prediction tasks dominated by continuous variables. In contrast, the ensemble framework proposed in this study fuses XGBoost, LightGBM, and Random Forest—three complementary tree-based models. This combination allows the framework to leverage the strengths of each model while learning patterns across different feature subspaces. Specifically, XGBoost provides strong residual fitting capability, LightGBM offers high training efficiency and fine-grained control, and Random Forest contributes robustness and insensitivity to outliers. Within the Stacking structure, the outputs of these base models are further integrated through a secondary learner, enhancing both overall fitting accuracy and generalization performance.

Although individual models may excel in specific dimensions, they often act as "black boxes," especially in agricultural scenarios where multiple variables interact simultaneously. They struggle to provide clear explanatory paths for variable effects. By integrating SHAP interpretability into the ensemble model, the framework ranks global variable importance and identifies key nonlinear interactions among variables. For instance, interactions between rainfall frequency and nitrogen application can only be effectively captured by a model with sufficient capacity and cross-feature learning ability. By comparison, the interpretability mechanisms of NGBoost and CatBoost are relatively limited and insufficient for supporting mechanism analysis or policy guidance in complex decision-making contexts.

Practical computational considerations were also addressed in the ensemble design. While Stacking increases time and memory requirements during training—particularly when multiple base models and the secondary learner are trained in parallel—the computational cost during inference is manageable. In production deployment, the multi-model prediction pipeline can be compressed into a lightweight single model using model distillation. Alternatively, some complex base models can be pruned, keeping only those that contribute most to predictive performance. These approaches significantly reduce online resource consumption. For mobile or edge deployment scenarios, the proposed model is highly adaptable. All base models are clear, structured decision tree frameworks suitable for low-power environments such as agricultural edge devices or greenhouse management systems. In applications requiring real-time response, such as precision nitrogen application control, a dual-track architecture can be used. The full ensemble model first generates inference labels or strategies offline, which are then delivered online through a lightweight model for fast response.

In summary, although the ensemble strategy entails higher computational cost during training, it offers controlled inference workload and flexibility during deployment. Future work will explore model distillation, neural architecture search, and other strategies to further improve deployment efficiency, facilitating the practical application of ensemble models in agricultural GHG emission prediction.

6 Conclusion

This study investigates the impact of nitrogen reduction measures on GHG emissions in agricultural systems by developing an integrated machine learning-based prediction framework. It conducts systematic modeling and experimental analyses on multiple real-world agricultural datasets.

The proposed ensemble model combines XGBoost, Random Forest, and LightGBM, effectively leveraging the strengths of each through a Stacking strategy. Compared to individual models, it demonstrates superior predictive accuracy and generalization. On the N_2O Global, Rothamsted, and DayCent datasets, the model achieves a maximum R^2 of 0.901 and a minimum RMSE of 0.301, highlighting strong cross-regional modeling capabilities. SHAP value analysis and interaction modeling identify nitrogen application rate, precipitation conditions (soil moisture and rainfall frequency), soil

type, and pH as key factors influencing N_2O emissions. Nitrogen application consistently emerges as the primary driver, with its interaction with soil moisture particularly significant in high denitrification emission scenarios. Scenario experiments confirm that all nitrogen reduction strategies effectively lower N_2O emissions. Simple nitrogen reduction leads to approximately 15%–20% reductions, substitution is slightly less effective, while the comprehensive management strategy (reduced nitrogen + cover crops + optimized tillage) achieves reductions exceeding 30%, demonstrating the greatest potential for practical adoption.

Despite these advances, the study has limitations. The datasets primarily represent a few typical agricultural regions and lack comprehensive coverage of diverse crop types, extreme climates, and management practices, limiting model transferability at national or global scales. Additionally, current scenario simulations use static inputs for short-term predictions, without fully capturing crop growth stages, dynamic fertilization timing, or long-term soil feedbacks, making it difficult to model the full lifecycle of the nitrogen reduction–emission process. Future research should incorporate time-series modeling and remote sensing data fusion to build dynamic prediction systems across multiple temporal and spatial scales, enabling full lifecycle modeling of agricultural carbon and nitrogen cycles. Integrating causal inference methods—such as double machine learning and causal forests—would further improve the interpretability of management effects and enhance decision support.

Acknowledgment

This study was supported by National Key R&D Program of China (2022YFF0711800).

References

[1] Kamyab H, SaberiKamarposhti M, Hashim H, Yusuf M. Carbon dynamics in agricultural greenhouse gas emissions and removals: a comprehensive review. Carbon Letters, 2024, 34: 265-289. https://doi.org/10.1007/s42823-023-00647-4

[2] Qian H, Zhu X, Huang S, Linquist B, Kuzyakov Y, Wassmann R, Minamikawa K, Martinez-Eixarch M, Yan X, Zhou F, Sander B O, Zhang W, Shang Z, Zou J, Zheng X, Li G, Liu Z, Wang S, Ding Y, van Groenigen K J, Jiang Y. Greenhouse gas emissions and mitigation in rice agriculture. Nature Reviews Earth & Environment, 2023, 4: 716-732. https://doi.org/10.1038/s43017-023-00482-1

[3] Maraveas C, Karavas C S, Loukatos D, Bartzanas T, Arvanitis K S, Symeonaki E. Agricultural greenhouses: Resource management technologies and perspectives for zero greenhouse gas emissions. Agriculture, 2023, 13(7): 1464. https://doi.org/10.3390/agriculture13071464

[4] Polymeni S, Skoutas D N, Sarigiannidis P, Kormentzas G, Skianis C. Smart agriculture and greenhouse gas emission mitigation: A 6G-IoT perspective. Electronics, 2024, 13(8): 1480. https://doi.org/10.3390/electronics13081480

[5] Zhao M, Shi R, Du R, Yao L. The pathway to China's carbon–neutral agriculture: Measures, potential and future strategies. Chinese Political Science Review, 2023, 8: 304-324. https://doi.org/10.1007/s41111-023-00237-9

[6] Harmsen M, Tabak C, Höglund-Isaksson L, Humpenöder F, Purohit P, van Vuuren D. Uncertainty in non-CO2 greenhouse gas mitigation contributes to ambiguity in global climate policy feasibility. Nature Communications, 2023, 14: 2949. https://doi.org/10.1038/s41467-023-38577-4

[7] Aziz S, Chowdhury S A. Analysis of agricultural greenhouse gas emissions using the STIRPAT model: A case study of Bangladesh. Environment, Development and Sustainability, 2023, 25: 3945-3965. https://doi.org/10.1007/s10668-022-02224-7

[8] Qin J, Duan W, Zou S, Chen Y, Huang W, Rosa L. Global energy use and carbon emissions from irrigated agriculture. Nature Communications, 2024, 15: 3084. https://doi.org/10.1038/s41467-024-47383-5

[9] Xia L, Cao L, Yang Y, Ti C, Liu Y, Smith P, van Groenigen K J, Lehmann J, Lal R, Butterbach-Bahl K, Kiese R, Zhuang M, Lu X, Yan X. Integrated biochar solutions can achieve carbon-neutral staple crop production. Nature Food, 2023, 4: 236-246. https://doi.org/10.1038/s43016-023-00694-0

[10] Ang F, Kerstens K, Sadeghi J. Energy productivity and greenhouse gas emission intensity in Dutch dairy farms: A Hicks–Moorsteen by-production approach under non-convexity and convexity with equivalence results. Journal of Agricultural Economics, 2023, 74(2): 492-509. https://doi.org/10.1111/1477-9552.12511

[11] Zhu Y, Qi Z, Wang L, Yang X. Spatial and temporal evolution and driving factors of synergistic effects of agricultural pollution and carbon emission reduction in China. Journal of Ecology and Rural Environment, 2024, 40(9): 1201-1212. https://dx.doi.org/10.19741/j.issn.1673-4831.2023.0668

[12] Mondal S K, An S I, Min S K, Jiang T, Su B. Enhanced soil moisture–temperature coupling could exacerbate drought under net-negative emissions. NPJ Climate and Atmospheric Science, 2024, 7: 265. https://doi.org/10.1038/s41612-024-00820-0

[13] Lu W, Hao Z, Ma X, Gao J, Fan X, Guo J, Li J, Lin M, Zhou Y. Effects of different proportions of organic fertilizer replacing chemical fertilizer on soil nutrients and fertilizer utilization in gray desert soil. Agronomy, 2024, 14(1): 228. https://doi.org/10.3390/agronomy14010228

[14] Tong B, Hou Y, Wang S, Ma W. Partial substitution of urea fertilizers by manure increases crop yield and nitrogen use efficiency of a wheat–maize double cropping system. Nutrient Cycling in Agroecosystems, 2023, 127: 11-21. https://doi.org/10.1007/s10705-022-10219-z

[15] Joshi D R, Clay D E, Clay S A, Moriles-Miller J, Daigh A L M, Reicks G, Westhoff S. Quantification and machine learning based N2O–N and CO2–C emissions predictions from a decomposing rye cover crop. Agronomy Journal, 2024, 116(3): 795-809. https://doi.org/10.1002/agj2.21185

[16] Liu S, Liu K, Wang Z, Liu Y, Bai B, Zhao R. Investigation of a transformer-based hybrid artificial neural networks for climate data prediction and analysis. Frontiers in Environmental Science, 2025, 12: 1464241. https://doi.org/10.3389/fenvs.2024.1464241

[17] Magazzino C, Cerulli G, Shahzad U, Khan S. The nexus between agricultural land use, urbanization, and greenhouse gas emissions: Novel implications from different stages of income levels. Atmospheric Pollution Research, 2023, 14(9): 101846. https://doi.org/10.1016/j.apr.2023.101846

[18] Kabato W, Getnet G T, Sinore T, Nemeth A, Molnár Z. Towards climate-smart agriculture: Strategies for sustainable agricultural production, food security, and greenhouse gas reduction. Agronomy, 2025, 15(3): 565. https://doi.org/10.3390/agronomy15030565

[19] Don A, Seidel F, Leifeld J, Kätterer T, Martin M, Pellerin S, Emde D, Seitz D, Chenu C. Carbon sequestration in soils and climate change mitigation—Definitions and pitfalls. Global Change Biology, 2024, 30(1): 16983. https://doi.org/10.1111/gcb.16983

[20] Olesen J E, Rees R M, Recous S, Bleken M A, Abalos D, Ahuja I, Butterbach-Bahl K, Carozzi M, De Notaris C, Ernfors M, Haas E, Hansen S, Janz B, Lashermes G, Massad R S, Petersen S O, Rittl T F, Scheer C, SmithK E, Thiébeau P, Taghizadeh-Toosi A, Thorman R E, Topp C F E. Challenges of accounting nitrous oxide emissions from agricultural crop residues. Global Change Biology, 2023, 29(24): 6846-6855. https://doi.org/10.1111/gcb.16962

[21] He T, Ding W, Cheng X, Cai Y, Zhang Y, Xia H, Wang X, Zhang J, Zhang K, Zhang Q. Meta-analysis shows the impacts of ecological restoration on greenhouse gas emissions. Nature Communications, 2024, 15: 2668. https://doi.org/10.1038/s41467-024-46991-5

[22] Lin X, Yang R, Zhang W, Zeng N, Zhao Y, Wang G, Li T, Cai Q. An integrated view of correlated emissions of greenhouse gases and air pollutants in China. Carbon Balance and Management, 2023, 18: 9. https://doi.org/10.1186/s13021-023-00229-x

[23] Wang X, Xu M, Lin B, Bodirsky B L, Xuan J, Dietrich J P, Stevanović M, Bai Z, Ma L, Jin S, Fan S, Lotze-Campen H, Popp A. Reforming China's fertilizer policies: implications for nitrogen pollution reduction and food security. Sustainability Science, 2023, 18: 407-420. https://doi.org/10.1007/s11625-022-01189-w