World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

A novel machine learning-driven approach for predicting nitrous oxide flux in precision managed agricultural systems

Srikanth Samy [1, *, #], Krishiv Jaini [1, #] and Sarah Preheim [2]

[1] *Dublin High School, Dublin, CA 94568.*
[2] *Johns Hopkins University, Whiting School of Engineering, Baltimore, MD 21218.*
[#] *These authors contributed equally.*

World Journal of Advanced Research and Reviews, 2024, 24(01), 679–685

Publication history: Received on 27 August 2024; revised on 04 October 2024; accepted on 06 October 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.24.1.3041

Abstract

N_2O, also known as nitrous oxide, is a greenhouse gas that is roughly 300 times more potent than CO_2 and destroys the the stratospheric ozone layer causing climate change. One of the primary causes of the rapid increase of N2O in our ecosystem is the application of nitrogen fertilizer to agricultural land. This stimulates N_2O emissions and accounts for approximately 5% of the global greenhouse gases, forcing harm to the environment and atmosphere (Aronson and Allison, 2012). Previous models severely underestimated N2O flux in various crops, causing inaccurate predictions to form. In our study, we utilized data from automated flux chambers to train and evaluate different machine-learning models to predict the field-level flux of N_2O which assist farmers to predict fertilizer amounts to use. The best machine learning model, Random Forest, performed considerably better than the standard empirical and biophysical models by roughly 15%, and show promise in improving predictive accuracy and guiding sustainable agricultural practices.

Keywords: Machine learning-driven approach; Nitrous oxide flux; Agricultural systems; N_2O

1. Summary

Nitrous oxide (N2O), is a greenhouse gas that is roughly 300 times more potent than carbon dioxide causing climate change and destroys the stratospheric ozone layer. One of the primary causes of the rapid increase of N2O in the atmosphere is the application of nitrogen fertilizer to agricultural land. This stimulates N2O emissions, which account for a significant percentage of global greenhouse gasses. Previous models severely underestimated N2O flux in various crops, causing inaccurate predictions. Implementing Machine Learning methods to predict nitrous oxide flux in precision-managed agricultural systems trained on data split by crop and experiment type can accurately predict N2O flux and also shows promise in outperforming empirical and biophysical models. In our study, we utilized data from automated flux chambers to train and evaluate different machine-learning models to predict the field-level flux of N2O. After testing various Machine Learning techniques on the Random Forest, LSTM, and XGBoost models, we gathered results on which model would be most accurate, in addition to what code should be implemented on the model. The Machine Learning model, Random Forest, proved to be the best and performed considerably better than the standard empirical and biophysical models by roughly 15. This new method of using Machine Learning to predict N2O flux showed promise in improving predictive accuracy and guiding sustainable agricultural practices using data segmentation techniques that we had designed and implemented.

* Corresponding author: Srikanth Samy

2. Introduction

2.1. Problem

The agricultural sector is a major contributor to global greenhouse gas emissions, with nitrous oxide (N2O) being a significant component (Rees et al., 2020). N2O is a potent, long-lived greenhouse gas with a global warming potential 265-298 times greater than carbon dioxide (3). N2O emissions account for 5% of the global greenhouse gasses in the atmosphere (Aronson and Allison, 2012) and over the past 50 years, N2O concentrations in the atmosphere have increased by 20%, largely due to the increased use of nitrogen fertilizers and manure application in agriculture to meet this growing global food demand (Sapkota et al., 2021; Rees et al., 2020). Additionally, the demand for food has increased with global population rise, which has led to intensified agricultural practices, including the extensive use of nitrogen-based fertilizers. While these practices are critical for enhancing crop yields and food security, they have inadvertently contributed significantly to global greenhouse gas emissions. These agricultural inputs lead to increased N2O emissions, a greenhouse gas that has a far greater impact on global warming than CO2. Projected climate change will further increase emissions if current practices continue (Rees et al., 2020). This is due to mechanisms such as increased temperatures, which lead to higher energy demands for cooling, and the release of nitrous oxide from agricultural soils. Warmer temperatures can also alter natural carbon sinks and increase the decomposition rates in soils, releasing more greenhouse gasses (Rees et al., 2020). Furthermore, the feedback loop between agriculture and climate change needs to be examined. As the climate warms, it affects agricultural productivity, leading to potential changes in farming practices that could either mitigate or exacerbate greenhouse gas emissions. For instance, higher temperatures and altered precipitation patterns can influence the rate of N2O emissions from soil. Potential mitigation strategies, such as the adoption of precision agriculture techniques, improved nitrogen management practices, and the use of alternative fertilizers, are essential for reducing N2O emissions while maintaining agricultural productivity.

2.2. Current Model

Current models to predict N2O flux in agricultural systems include process-based models like ecosys, which incorporate mechanistic representations of biophysical and biochemical processes in agroecosystems (Yang et al., 2022). When validated against field measurements, ecosys demonstrated R^2 values of 0.64 for cumulative N2O emissions (RMSE=0.89 kg N/ha/y) and 0.83 for corn yield (RMSE=1.91 Mg/ha) across multiple sites in the US Midwest. The model was validated using a variety of data sources, including AmeriFlux network sites for CO2 flux, soil temperature, and soil moisture; sites with N2O observations during growing seasons and winters using static and automated chambers; and a total of 108 site-year estimates of cumulative N2O emissions across eight Midwestern states. Additionally, snow depth and soil temperature data were used for winter validations at one site. The model relies on inputs such as air temperature, precipitation, relative humidity, wind speed, net solar radiation, and soil properties from the gSSURGO database. These performances are comparable to similar studies in North American cropping systems using other process-based models. However, the authors acknowledge uncertainties in the regional estimates due to model limitations, estimation of observations, and lack of validations on non-growing season N2O emissions in some areas, N leaching, and soil organic carbon changes (Yang et al., 2022). While machine learning (ML) models also face challenges in scaling to global predictions, they offer a promising approach by learning complex patterns from large datasets. ML models can integrate diverse data sources, including soil properties, weather data, and management practices, to improve prediction accuracy. Unlike traditional models, ML models can potentially adapt to new data, making them more flexible and capable of capturing the spatial and temporal variability of N2O emissions. However, they still require extensive, high-quality data to train effectively and perform well across different scales. Current models have severely underestimated N2O flux, which leads to inaccurate predictions of emissions from fertilizer usage and other practices. Specifically, predicting N2O flux peaks has been an imperative issue. To address the limitations of current models, it is crucial to explore advanced machine learning (ML) techniques that offer both improved accuracy and computational efficiency. By identifying and implementing ML models that excel in predictive power while maintaining rapid processing times, we can significantly reduce errors in the existing approaches.

2.3. ML Model

To address these limitations, researchers have proposed the use of advanced ML models to enhance the predictive power of N2O emission models (Xu and Zeng, 2022)(Dorich et al., 2020). Recent studies have demonstrated the potential of advanced ML models like random forests, gradient boosting, and neural networks to outperform traditional empirical and biophysical models in predicting N2O flux from agricultural soils due to their ability to capture the non-linear relationship between N2O flux and its underlying processes (Szelgg et al. 2023)(Saha et al., 2021). Our study, like those by Szelgg et al. (2023) and Saha et al. (2021), leverages advanced ML models to improve N2O emission predictions and highlight the limitations of traditional models. However, we focus on agricultural systems and use data segmentation by crop and experiment type, whereas Szelgg et al. (2023) focuses on wastewater treatment plants, and

Saha et al. (2021) couples ML with cropping systems models. Additionally, we test three different types of ML models for this situation: XGBoost, LSTM, and Random Forest to find the best performing model. Saha found that using a Random Forest model along while training based on different experiments they were able to achieve a maximum R2 of 0.89 on the Arlington site. We were able to achieve similar results as we found the best option was using the Random Forest model, and our highest R2 value was 0.95 on the Arlington site as well. Szlegg on the other hand used a different dataset due to the fact that the experiment was based on finding N2O in wastewater treatment plants, however, his results showed that he had the lowest error using the XGBoost model, which is similar to our study as we got the second lowest error using the XGBoost model as well. Szlegg did not use the Random Forest model in his testing so we do not know if using it would have yielded higher accuracy similar to ours. Additionally, the functionality of using a Machine Learning model is now significantly more plausible due to the availability of long- term high-frequency observations from automated flux chambers that improve predictability and allow us to better understand the factors controlling daily N2O flux variation (Saha et al., 2021). In summary, understanding the contribution of agriculture to greenhouse gas emissions, the subsequent impact on climate change, and the potential pathways for mitigating these effects is crucial for ensuring sustainable agricultural practices in the face of a changing climate. We hypothesized that training different Machine Learning models on data split by experiment and vegetation type would perform better than standard biophysical and empirical models in predicting N2O while utilizing the optimal model. After testing our hypothesis and implementing different code, we found that the Random Forest Model performed best, although it is important to implement proper techniques such as data segmentation on specific features included.

2.4. Data

The dataset for this study was obtained from three long-term experiments conducted at the W.K. Kellogg Biological Station in Michigan ("BCSE_KBS" and "MCSE-T2") and the Arlington Agricultural Research Station in Wisconsin ("Arlington WI") (Saha et al., 2020). These experiments included continuous no-till corn and corn-soybean-wheat rotations, with BCSE_KBS and Arlington WI having only corn but MCSE-T2 having all 3 types of vegetation. N2O flux measurements were taken using automated flux chambers, which provided daily average N2O flux data, resulting in a dataset with 2246 data points across 15 non-consecutive years. Additionally, the dataset included various predictive variables such as soil moisture, air temperature, cumulative precipitation, and nitrogen fertilization rates. These variables are crucial for understanding the environmental factors influencing N2O emissions. Data preprocessing involved iterative imputation to handle missing values and robust scaler to normalize the features, ensuring accurate and reliable inputs for the machine learning models. The preprocessing steps aimed to enhance the stability and performance of the models to make a model that would be able to predict N2O flux efficiently.

3. Methodology

3.1. Model Building

We used the Python programming language with the pandas, numpy, scikit-learn, matplotlib, tensor-flow, XGBoost, and LightGBM libraries to build and train the XGBoost, Random Forest, and LSTM machine learning models. XGBoost is a gradient-boosting decision tree algorithm ideal for tabular data, training trees sequentially with Gradient Descent Optimization and L1 and L2 regularization to prevent overfitting (Dhillon et al., 2023)(Shahhosseini et al., 2019)(Wang et al., 2020). Random Forest is a sequential technique fitting multiple decision tree models on subsets of the training data, averaging their predictions to increase accuracy, and is known for its simplicity and efficiency. The Long Short-Term Memory (LSTM) Recurrent Neural Network was chosen for modeling temporal dynamics of the N2O flux time series, using specific cells to regulate the flow of information and capture temporal data effectively. We trained it to predict daily N2O flux from the numeric features provided in the data.

3.2. Data Manipulation

To process the data, we had to go through many steps, including data imputation, feature scaling, binning, hyperparameter optimization, and proper train test split to be compatible with the binning. For data imputation, we used scikit- learn's iterative imputer, which iteratively modeled each feature with missing values as a function of other features, allowing for more accurate estimates of the missing data. To handle feature scaling, which allows all features to contribute to the training process equally, we used robust scaler, which scales according to the median and interquartile range. This makes it less sensitive to outliers compared to other scaling techniques. For hyperparameter tuning, we used GridSearchCV from scikit-learn to perform an exhaustive search and find the best hyperparameters using K-Fold cross-validation. For training and testing data split, we split the data up using a 70% training and 30% evaluation/testing split. The most important data manipulation technique we did was Quantile binning which handled the uneven distribution of our target variable, N2O . This divided the target variable into bins of equal size allowing for a uniform distribution of the target variable across the bins. This binning strategy improved the stability and

performance of the model compared to training on the raw continuous target variables. The model was evaluated using the standard regression evaluation methods R^2, RMSE, and MAE.

4. Results

4.1. Model Performance

We employed three different types of Machine Learning models (XGBoost, Random Forest, and LSTM) to predict N2O flux based on N2O flux data and associated environmental observations from three different experiments (BCSE_KBS, MCSE_T2, Arlington, WI). N2O flux data was collected from automated flux chambers and the feature data we trained on included a range of environmental and management variables. These variables encompassed nitrogen management (e.g., N fertilization rate, days after fertilizer application), precipitation, air temperature, soil properties (e.g., water-filled pore space, NH4-N and NO3-N content, clay concentration, and soil organic matter in the top 25-cm soil layer), and temporal information. The experiments covered different vegetation types: continuous no-till corn in BCSE_KBS and Arlington, WI, and a three-year no-till rotation of corn, soybean, and wheat in MCSE_T2. We then tested the different models based on a split of vegetation type and experiment location after training them on this data and evaluated the different results. The experiments BCSE_KBS and Arlington, WI were primarily implemented in similar ways, particularly regarding the vegetation type under study. Both BCSE_KBS and Arlington, WI focused on the vegetation type of corn, ensuring a consistent approach in these two experiments. On the other hand, the experiment MCSE_T2 diverged significantly from the other two in terms of the variety of vegetation types it included. Unlike BCSE_KBS and Arlington, WI, which were limited to corn, MCSE_T2 was a rotation experiment, covering Corn, Soybean (GLYMYX), and Wheat (TRIAE).

Of the three types of ML models, the Random Forest model (Figure 1) performed best overall, explaining 95.26% of the variability in the Arlington WI experiment located at a station in Wisconsin, USA, 99.38% in the BCSE_KBS experiment, and 96.07% in the MCSE-T2 experiment which were both experiments at the W.K. Kellog Biological Station in southwest Michigan, USA. Specifically, the high percentage of variability explained in BCSE_KBS indicates that the model captured nearly all the significant factors affecting N2O flux in this region. The evaluation R Mean Square Error (RMSE) for the Random Forest model was 11.817, 1.177, and 1.462, and the Mean Absolute Error (MAE) was 2.2632, 0.2886, and 0.2384 for Corn, Wheat (TRIAE), and Soybean (GLYMX), respectively, reflecting the model's high precision and accuracy across different experimental sites (Table 1). Additionally, the model's performance by vegetation type showed 92.17% variability explained for Corn, 95.03% for Soybean (GLYMX), and 97.16% for Wheat (TRIAE).

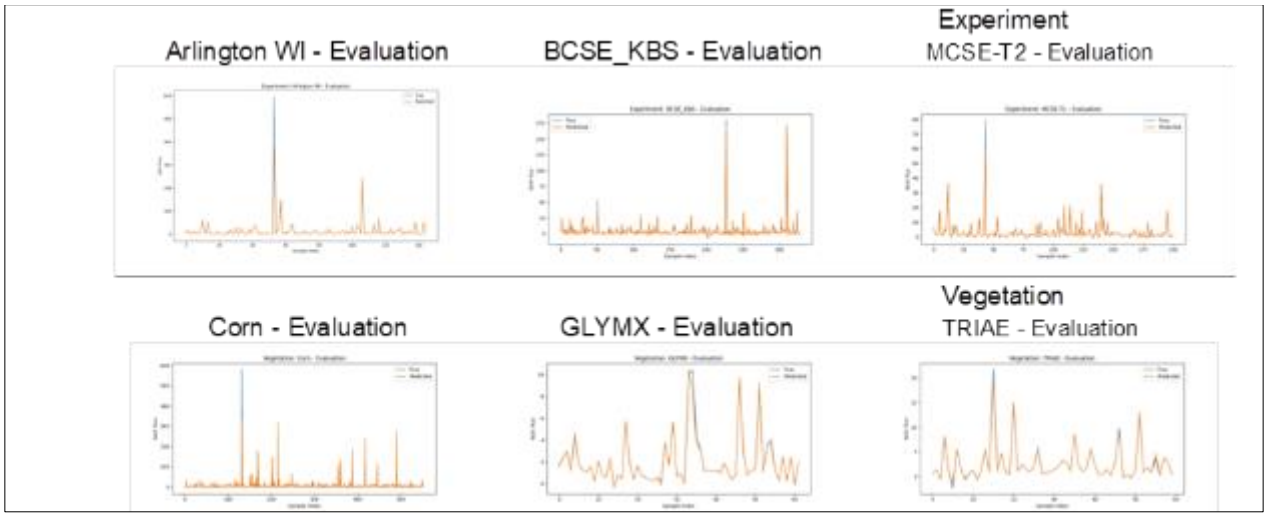

Figure 1 Evaluation of N2O flux prediction with Random Forest on testing split of data. The X-axis represents the Sample Index and the Y-axis represents N2O flux in g N2O-N ha-1d-1

These results highlight the model's versatility in handling diverse crop types, with the highest accuracy observed for Wheat, however, this was likely due to the smaller nature of the dataset with 197 samples leading to less variability. This is likely also the reason that Corn had the lowest accuracy due to the large sample size of 1844, leading to more variability. The superior performance of the Random Forest model can be attributed to its ability to handle complex,

non-linear relationships and interactions among variables, making it particularly effective in capturing the variability in N2O flux across different experimental conditions and vegetation types.

4.2. Critical Predictor Variables

Water-Filled Pore Space at 25 centimeters (WFPS25cm), Nitrogen Fertilization Rate (N_rate), and Air Temperature (AirT) were identified as key variables in predicting N2O flux, capturing essential aspects of soil moisture, nitrogen application, and temperature. The other variables that were used in training included the Cumulative precipitation in the last two days before gas sampling (PP2), Cumulative precipitation in the last week before gas sampling (PP7), Daily average air temperature (AirT), Days after top-dressed N fertilizer application (DAF_TD), Days after side-dressed N fertilizer application (DAF_SD), NH4-N content in the top 25-cm soil layer in kg ha-1 (NH4), NO3-N content in the top 25-cm soil layer in kg ha-1 (NO3), Clay concentration in the top 25-cmsoil layer in g kg-1 (Clay), soil organic matter concentration (SOM). In Arlington WI, WFPS25cm's mean partial dependence ranged from 13.3157 to 14.2298, suggesting a strong influence of soil moisture on N2O flux. Similarly, the N_rate's range from 15.3309 to 14.0046 highlights the significant impact of nitrogen fertilization, while AirT ranged from 13.9678 to 14.0152, underscoring temperature's role in emissions. In BCSE_KBS, WFPS25cm ranged from 4.5953 to 4.6638, and AirT from 4.6239 to 4.6713, showing how these factors vary regionally. For MCSE-T2, WFPS25cm ranged from 3.0932 to 3.1506, N_rate from 3.1360 to 3.1514, and AirT from 3.1256 to 3.1476, indicating consistent influences across different sites. These results demonstrate that WFPS25cm, N_rate, and AirT are critical in driving N2O flux across various conditions. Grouping data points by experiment and vegetation type with quantile binning significantly improved model performance by capturing unique patterns within each subgroup, leading to more accurate and reliable predictions. The variability and influence of these factors in different regions and conditions emphasize the importance of considering both environmental and experimental contexts in modeling N2O emissions.

5. Discussion

The Random Forest model demonstrated superior performance compared to XGBoost and LSTM in predicting N2O flux in agricultural systems. It effectively captured variability across different experimental sites, with R2 values of 95.26% in Arlington WI, 99.38% in BCSE_KBS, and 96.07%.

Table 1 Model Evaluation Results

Category	Train MSE	Train MAE	Train R^2	Eval MSE	Eval MAE	Eval R^2
Arlington WI	Random Forest					
	104.46	1.96	96.50%	139.70	2.26	95.26%
BCSE_KBS	7.20	0.35	97.31%	1.38	0.29	99.38%
MCSE-T2	1.71	0.18	96.99%	2.14	0.24	96.07%
Corn	12.57	0.83	98.56%	94.47	1.26	92.17%
GLYMX	0.11	0.25	99.63%	0.28	0.38	95.03%
TRIAE	2.31	0.51	97.12%	0.44	0.41	97.16%
Arlington WI		XGBoost				
	0.02	0.09	99.99%	280.95	1.93	90.46%
BCSE_KBS	0.00	0.04	99.99%	4.14	0.34	98.13%
MCSE-T2	0.00	0.02	99.99%	4.35	0.24	92.01%
Corn	0.00	0.03	99.99%	140.65	1.05	88.35%
GLYMX	0.00	0.00	99.99%	0.07	0.12	98.69%
TRIAE	0.00	0.00	99.99%	0.17	0.13	98.89%
Arlington WI		LSTM				
	365.32	2.97	87.75%	731.20	3.41	75.17%

BCSE_KBS	5.68	1.06	97.88%	2.62	1.08	98.82%
MCSE-T2	0.53	0.40	99.08%	0.54	0.41	99.01%
Corn	71.19	2.17	91.84%	83.80	2.25	93.06%
GLYMX	0.14	0.25	99.55%	0.07	0.21	98.83%
TRIAE	2.94	0.95	96.34%	0.59	0.66	96.15%

in MCSE-T2. This outperformance was also consistent across various vegetation types, achieving accuracies of 92.17% for Corn, 95.03% for Soybean (GLYMX), and 97.16% for Wheat (TRIAE).

Key predictor variables identified by the Random Forest model such as Water-Filled Pore Space at 25 centimeters (WFPS25cm), Nitrogen Fertilization Rate (N_rate), and Air Temperature (AirT) were found to be crucial in explaining N2O flux. In contrast, XGBoost and LSTM models placed different levels of emphasis on these predictors, leading to less accurate predictions. The grouping of data by experiment and vegetation type, along with the use of quantile binning, significantly enhanced the Random Forest model's performance by enabling it to capture unique patterns within the data.

While the Random Forest model achieved an almost perfect fit (R2 = 0.999994) within the training region, its generalization ability was challenged when applied to data from different regions, where the R2 value decreased to 0.904599. This discrepancy underscores the importance of contextual factors such as environmental conditions and experimental design that influence model performance. By integrating these contextual elements we are able to highlight the necessity of considering both environmental and experimental contexts for accurate N2O emissions predictions, which was less effectively addressed by the alternative models.

This discrepancy suggests that while the model performs exceptionally well within the confines of the training data, its generalizability to new regions with varying climates, soil types, and agricultural practices is limited. Several factors could have influenced the model's performance. Firstly, the model was trained on data from only three experiments involving three crop types. The limited variety of crops and experimental conditions may have constrained the model's ability to generalize to other agricultural systems. Additionally, wheat and soybeans were only represented in one of the experiments, limiting the model's training diversity. More replicates and a broader variety of crops and conditions would likely improve the model's robustness.

Despite these limitations, the findings highlight the potential and constraints of using Random Forest models for predicting N2O flux in precision agriculture. The model's high accuracy within the training region indicates that it effectively captures complex, non-linear interactions in controlled settings. However, the drop in performance in different regions underscores the need for caution when applying the model to diverse agricultural systems. While the model shows promise, its current limitations suggest that it should not be solely relied upon for N2O emission predictions in regions significantly different from the training environment. The research contributes to the understanding of machine learning applications in agriculture, emphasizing the importance of diverse and high-quality data.

Improving the model's generalizability across different regions and conditions requires adaptations and enhancements. Additional data and experimental designs are necessary to bolster the model's efficiency. Long-term climate changes impact N2O emissions, and adjusting the model to account for these variations is essential. Future experiments should focus on longitudinal studies, conducting experiments over 10-20 years to provide insights into long-term N2O emissions and improve prediction reliability. Implementing techniques like SMOTE can address data imbalances and improve model accuracy, while tools like SHapley Additive exPlanations (SHAP) can help interpret model predictions and understand the influence of different variables on N2O flux. Expanding the dataset to include a broader range of crops, soil types, and climatic conditions will also enhance the model's applicability.

6. Conclusion

In summary, the Random Forest model coupled with data segmentation shows significant potential for predicting nitrous oxide flux in precision-managed agricultural systems. The high accuracy within the training region highlights the model's capability to capture complex interactions within N2O flux and its underlying factors. However, the reduced performance in new regions indicates the need for broader, more diverse datasets and improved experimental designs.

Future research should focus on addressing these limitations to enhance the model's generalizability and reliability, ultimately contributing to better-informed agricultural management

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Saha, D., Basso, B., & Robertson, G. P. (2020). Data from: Machine learning improves predictions of agricultural nitrous oxide (N2O) emissions from intensively managed cropping systems [Dataset]. Dryad. https://doi.org/10.5061/dryad.bnzs7h493

[2] Bartosz, Szelag¸., Ewa, Zaborowska., Jacek, Makinia. (2023). An algorithm for selecting a machine learning method for predicting nitrous oxide emissions in municipal wastewater treatment plants. Journal of water process engineering. https://doi.org/10.1016/j.jwpe.2023.103939

[3] Xu, X., & Zeng, B. (2023). Application of a novel second-order differential equation grey model to forecast NOx emissions in China. Environmental Science and Pollution Research, 30, 24441-24453. https://doi.org/10.1007/s11356-022-23662-w

[4] Dorich, C. D., De Rosa, D., Barton, L., et al. (2020). Global Research Alliance N2O chamber methodology guidelines: Guidelines for gap-filling missing measurements. Journal of Environmental Quality, 49, 1186-1202. https://doi.org/10.1002/jeq2.20138

[5] Shahhosseini, M., Martinez-Feria, R. A., Hu, G., & Archontoulis, S. (2019). Maize yield and nitrate loss prediction with machine learning algorithms. Environmental Research Letters, 14(12), 124026. https://doi.org/10.1088/1748-9326/ab5268

[6] Wang, Y., Zhang, Z., Feng, L., Du, Q., & Runge, T. (2020). Combining multi-source data and machine learning approaches to predict winter wheat yield in the conterminous United States. Remote Sensing, 12(8), 1232. https://doi.org/10.3390/rs12081232

[7] Dhillon, M. S., Dahms, T., Kuebert-Flock, C., Rummler, T., Arnault, J., Steffan-Dewenter, I., & Ullmann, T. (2023). Integrating random forest and crop modeling improves the crop yield prediction of winter wheat and oil seed rape. Frontiers in Remote Sensing, 3. https://www.frontiersin.org/articles/10.3389/frsen.2022.1010978

[8] Sapkota, T. B., Jat, M. L., & Rana, D. S. (2021). Crop nutrient management using Nutrient Expert improves yield, increases farmers' income and reduces greenhouse gas emissions. Scientific Reports, 11, 1564. https://doi.org/10.1038/s41598-020-79883-x

[9] Rees, R. M., Maire, J., Florence, A., Cowan, N., Skiba, U., van der Weerden, T., & Ju, X. (2020). Mitigating nitrous oxide emissions from agricultural soils by precision management. Frontiers of Agricultural Science and Engineering, 7(1), 75-80. https://doi.org/10.15302/J-FASE-20192947

[10] Udvardi, M., Brodie, E. L., Riley, W., Kaeppler, S., & Lynch, J. (2015). Impacts of agricultural nitrogen on the environment and strategies to reduce these impacts. Procedia Environmental Sciences, 29, 303. https://doi.org/10.1016/j.proenv.2015.07.275

[11] Aronson, E. L., & Allison, S. D. (2012). Meta-analysis of environmental impacts on nitrous oxide release in response to N amendment. Frontiers in Microbiology, 3, 272. https://doi.org/10.3389/fmicb.2012.00272

[12] Pärn, J., Verhoeven, J.T.A., & Butterbach-Bahl, K. (2018). Nitrogen-rich organic soils under warm well-drained conditions are global nitrous oxide emission hotspots. Nature Communications, 9, 1135. https://doi.org/10.1038/s41467-018-03540-1

[13] Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ, 6, e5518. https://doi.org/10.7717/peerj.5518

[14] Liu, F., Wu, H., Zhao, Y., Li, D., Yang, J.-L., Song, X., Shi, Z., Zhu, A.-X., & Zhang, G.-L. (2021). Mapping high resolution National Soil Information Grids of China. Science Bulletin. https://doi.org/10.1016/j.scib.2021.10.013