Using Machine Learning Algorithms to Improve Crop Modeling Predictions for Monitoring and Mitigating Climate Change

Ankita Chauhan[1]* and Narender K. Sankhyan[2]

[1]*Ph. D. (Soil science), Department of Soil Science, CSK HPKV, Palampur, HP, India-176062*
[2]*Head cum Principal Scientist, Department of Soil Science, CSK HPKV, Palampur, HP, India-176062*

Corresponding Author
Ankita Chauhan
Email: Ankitachauhan94185@gmail.com

OPEN ACCESS

Keywords

Machine learning, Crop modelling, Climate change, Crop yield, Machine learning algorithms, Climate mitigation.

How to cite this article:

Chauhan, A. and. Sankhyan, N. K., 2024. Using Machine Learning Algorithms to Improve Crop Modeling Predictions for Monitoring and Mitigating Climate Change. *Vigyan Varta* 5(10): 65-69.

ABSTRACT

Integrating machine learning (ML) algorithms into crop modeling offers a transformative approach to enhancing agricultural systems' predictive accuracy and adaptability amid climate change pressures. This article examines recent advancements in ML applications within crop modeling, emphasizing improvements in real-time monitoring, scenario analysis and climate-resilient strategy development. Key ML methods—including supervised, unsupervised, and deep learning—are explored for analyzing extensive datasets such as remote sensing imagery, meteorological data, and soil properties. These methods enable precise predictions of crop growth and yield, support early detection of climate-induced stressors and allow for simulation of future climate impacts. Additionally, ML-driven optimization of crop management, including planting schedules and resource allocation, fosters effective adaptation strategies, enhancing crop resilience and mitigating greenhouse gas emissions. Despite these advancements, challenges persist related to data quality, model interpretability and the scalability of ML applications across diverse agricultural settings.

INTRODUCTION

Climate change has significantly impacted global temperatures, with a rise of 1.1°C observed between 2011 and 2020 compared to the pre-industrial period (1850-1900) (IPCC AR6 Synthesis Report). This warming, more pronounced over land than oceans, poses a critical threat to global food security by increasing the frequency and severity of extreme weather events such as droughts, floods and heatwaves, which in turn cause substantial crop yield losses. Soil, an essential natural resource, is not immune to the impacts of climate change. Both direct effects, such as changes in rainfall patterns, elevated temperatures and increased CO_2 levels and indirect effects through altered ecosystem functions, significantly influence soil properties and processes.

Accurate prediction of crop productivity and response to climate variability is essential for shaping agricultural policies, setting food aid priorities and developing effective climate change adaptation strategies. Two main types of models are used for crop yield prediction: crop simulation models and data-driven models like Machine Learning (ML) and Neural Networks (NN). Each has its own strengths and limitations, contributing to the ongoing development of more robust agricultural forecasting tools.

Modeling in the age of big data: The rise of machine learning

Machine learning (ML), unlike process-based modeling that relies on human expertise, allows computers to learn from data. Originating from Alan Turing's question, "Can machines think?", ML has evolved significantly over the last two decades, finding widespread application in fields like agriculture. Early use of ML in agriculture dates back to the 1990s, with artificial neural networks becoming prevalent, likely spurred by the development of the back-propagation algorithm.

ML excels in handling complex, non-linear relationships, making it effective for tasks such as yield prediction, remote sensing analysis, disease and weed detection and phenotyping. However, ML models are often criticized as black-box models due to their opaque internal workings, which complicates interpretation and limits their ability to generalize beyond the training data. Additionally, ML models typically require large datasets for effective training.

Types of machine learning techniques in crop modeling

1. **Supervised learning** involves training models on labeled data to map inputs like weather conditions to outputs such as crop yield. Techniques include regression models (e.g., Linear Regression, Random Forests, SVMs) for predicting continuous outcomes and classification models (e.g., Decision Trees, Logistic Regression) for tasks like disease prediction. These methods are used for yield forecasting, disease monitoring and optimizing planting schedules.

2. **Unsupervised learning** identifies patterns in unlabeled data. Clustering algorithms like K-Means and DBSCAN group similar data, such as regions with analogous soil properties, while dimensionality reduction methods like PCA simplify datasets by focusing on key variables. These techniques help identify climate zones, segment crop types and uncover hidden data patterns.

3. **Deep learning** employs neural networks with multiple layers for complex tasks. Convolutional Neural Networks (CNNs) analyze spatial data, such as satellite

images, to assess vegetation health, while Recurrent Neural Networks (RNNs) and LSTM Networks are used for time-series analysis, predicting crop growth stages based on weather trends. These approaches are crucial in precision agriculture for crop classification, phenological stage prediction and early detection of drought stress.

4. **Data sources and preprocessing**: Data for crop modeling is sourced from remote sensing (satellite and drone imagery), meteorological data (weather patterns), soil data (soil properties) and historical yield records. Preprocessing involves cleaning and normalizing this data by removing noise, filling gaps and standardizing it. Feature engineering selects and transforms relevant variables, enhancing model performance, while dimensionality reduction techniques like PCA simplify analysis by focusing on key variables.

5. **Feature engineering**: Effective feature selection is crucial for identifying variables with the most significant impact on crop outcomes and capturing complex interactions, such as between temperature and soil moisture. Automated feature engineering, particularly through deep learning, can extract features directly from raw data, while optimization techniques like Genetic algorithms refine feature selection, improving model efficiency and accuracy.

Collaboration between humans and machines: Allying PBMs with ML

Both process-based models (PBMs) and machine learning (ML) models have inherent limitations in model simulations. To achieve high prediction accuracy and interpretability, integrating human knowledge with ML has been a key focus, leading to the emergence of interpretable ML. In a hybrid modeling approach, known as the knowledge- and data-driven modeling (KDDM) approach, PBMs guide and constrain ML in several ways:

1. **Task allocation:** ML algorithms are chosen and adapted based on the knowledge embedded in PBMs. For instance, LSTM networks are selected for their ability to model long-term climate effects on crop growth, though they have yet to be widely applied in plant growth prediction.

2. **Direct use of PBM equations:** In some cases, PBM equations can directly serve as ML models, where ML learns parameter values like intercepts and slopes from data.

3. **Embedding PBM constraints:** PBM-derived variable boundaries can be incorporated into ML models through constrained loss functions or specific activation functions, enhancing interpretability.

4. **Simulation data from PBMs:** PBMs can generate simulation data to reduce the need for extensive experimental data, a strategy used in autonomous control research for training ML in smart decision-making.

This hybrid approach leverages the strengths of both PBMs and ML, aiming to create models that are both accurate and interpretable.

Approaches to combine PBMs and ML

Three key structures are used to combine process-based models (PBMs) with machine learning (ML) in knowledge- and data-driven modeling (KDDM) approaches, each offering unique advantages for agricultural modeling:

1. **Parallel structure**: In this approach, PBM and ML models share the same input variables. The final output is derived by combining the outputs of both models, either through addition or multiplication.

For example, ML can be trained on the residuals between experimental data and PBM predictions, improving prediction accuracy, as demonstrated by Fan et al. (2015) with the GreenLab model for predicting tomato dry weights.

2. **Serial structure**: This structure involves running either the PBM or ML model first, with the output of the first model serving as input for the second. For instance, ML can process sensor data to generate inputs for PBMs, or PBMs can provide preliminary predictions for ML refinement. Kaneko et al. (2022) illustrated this with a hybrid model that first calculated leaf photosynthetic rates using a PBM, then used an artificial neural network to predict canopy photosynthesis, achieving higher accuracy than either method alone.

3. **Modular structure**: In this flexible approach, specific PBM modules can be replaced or supplemented with ML models. For example, Feng et al. (2019) improved wheat yield predictions under extreme climate events by substituting a PBM module in the APSIM model with ML, enhancing accuracy by 33%. This structure allows for the integration of additional data sources, such as remote sensing, where ML processes and interprets unstructured data, further improving model accuracy.

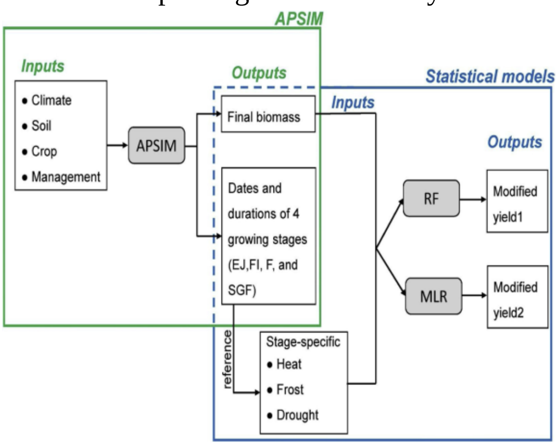

Diagram of the input and output per model for the APSIM+RF (or MLR) hybrid model applied in this study. EJ: end of juvenile; FI: floral initiation; F: flowering; SGF: start of grain filling; RF: random forest; MLR: multiple linear regression. [Feng et al. (2019)]

These KDDM structures effectively leverage the strengths of both PBMs and ML, enhancing the predictive power and flexibility of agricultural models.

Application in climate change monitoring and mitigation

1. **Real-time monitoring**: Integrating machine learning (ML) with satellite imagery, weather data and IoT sensors enables real-time crop health monitoring. ML algorithms analyze this data continuously to detect issues like drought stress or pest outbreaks early. For instance, NASA's deep learning models process multispectral satellite images, tracking vegetation indices and soil moisture to allow timely interventions, such as adjusting irrigation or fertilization.

2. **Scenario analysis**: ML simulates climate scenarios to predict their impact on crop productivity, guiding adaptive strategies. In India, ensemble ML models have been used to predict how climate variations might affect crops like wheat and rice. These models, which integrate climate projections with soil and management data, suggest strategies such as shifting planting dates and adopting drought-resistant varieties to mitigate yield declines.

3. **Adaptation strategies**: ML-enhanced crop models optimize adaptation strategies for climate change. In the U.S. Midwest, ML models analyze historical climate and soil data to recommend optimal planting dates for crops like maize, helping farmers maximize yields despite erratic weather patterns.

4. **Carbon sequestration and emission reduction**: ML supports carbon sequestration and emission reduction by optimizing land use and crop management. In Australia, precision agriculture techniques powered by ML refine fertilizer application, reducing nitrous oxide emissions while maintaining crop yields and promoting soil health.

CONCLUSION

Climate change presents a global challenge as it significantly affecting agriculture and food security. Traditional crop models often oversimplify complex interactions, whereas machine learning models require extensive data and can suffer from issues of interpretability. Hybrid-modeling approach has potential for accurate predictions that integrates a theory driven crop growth model into a data driven machine learning method. So, advances in machine learning and simulation crop modeling have created new opportunities to improve predictions in agriculture while considering the dynamic effects of climate change.

REFERENCES

Intergovernmental Panel on Climate Change (IPCC). 2023. Climate Change 2023: Synthesis report

Fan XR, Kang MZ, Heuvelink E, de Reffye P and Hu BG. 2015. A knowledge- and data-driven modeling approach for simulating plant growth: A case study on tomato growth. *Ecological Modelling* 312: 363–373

Feng P, Wang B, Liu DL, Waters C and Yu G. 2019. Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia. *Agricultural and Forest Meteorology* 275: 100-113

Kaneko T, Nomura K, Yasutake D, Iwao T, Okayasu T, Ozaki Y, Mori M, Hirota T and Kitano M. 2022. A canopy photosynthesis model based on a highly generalizable artificial neural network incorporated with a mechanistic understanding of single-leaf photosynthesis. *Agricultural and Forest Meteorology* 323: 109036