



## Emerging mathematical techniques for crop yield prediction: A comprehensive review

Ingale Jagruti D<sup>1</sup>, Dr. Goswammi V S<sup>2</sup>, Dr. Kailas L Vairal<sup>3</sup>

<sup>1</sup> Assistant Professor, Department of Mathematics, Padmashri Vikhe Patil college, Loni, Rahata, Maharashtra, India

<sup>2</sup> Assistant Professor, Shri JYT University, Chudela, Vidyanagri Jhunjhunu, Rajasthan, India

<sup>3</sup> Assistant Professor, Pravara Rural Engineering College, Loni, Rahata, Maharashtra, India

### Abstract

Crop yield prediction is critical for global food security, agricultural planning, and climate-change adaptation. Recent advances in mathematical modeling, statistical learning, and dynamical systems have dramatically improved predictive accuracy over traditional empirical or process-based models. This paper reviews emerging mathematical techniques from 2020–2025, including deep learning with spatiotemporal architectures, physics-informed neural networks (PINNs), causal inference frameworks, topological data analysis (TDA), dynamical systems with delay embeddings, and hybrid quantum–classical approaches. We compare their theoretical foundations, data requirements, interpretability, and reported performance on major crops (maize, wheat, rice, soybean). The survey concludes with open mathematical challenges and promising future directions.

**Keywords:** Physics-informed neural networks (PINNs), causal inference frameworks, Topological data analysis (TDA)

### Introduction

Accurate crop yield forecasting at field, regional, and global scales supports policy decisions, insurance pricing, and supply-chain management. Classical approaches fall into three categories:

- Crop simulation models (e.g., DSSAT, APSIM) – mechanistically rich but parameter-heavy.
- Statistical/regression models – simple but unable to capture non-linearities and interactions.
- Early machine-learning models (random forests, gradient boosting) – high accuracy but often treat time and space as independent features.

Since 2020, a new generation of mathematically sophisticated techniques has emerged that explicitly model spatiotemporal dynamics, physical constraints, genotype–environment interactions (G×E), and extreme events. These methods leverage tools from differential equations, algebraic topology, causal graph theory, and quantum computing.

### Deep Spatiotemporal Architectures

#### 1. Convolutional–Recurrent Hybrids and Transformers

- **ConvLSTM (Shi *et al.*, 2015; extended 2020–2024):** Remains the baseline. Recent variants incorporate attention mechanisms to weigh distant pixels dynamically.
- **Temporal Fusion Transformers (TFT) (Lim *et al.*, 2021)<sup>[1]</sup>:** Combine static covariates (soil type, genotype) with time-varying inputs (weather, NDVI). Achieved  $\leq 8\%$  MAE on U.S. Corn Belt county-level yields (2023 benchmarks). TFT architectures have been applied to wheat yield prediction, outperforming benchmarks by integrating multi-horizon forecasting with interpretability.
- **EarthFormer and FourCastNet adaptations (2023–2025):** Originally meteorological models, now fine-

tuned for yield by adding crop-specific channels (LAI, soil moisture). Outperform process-based models by 15–25% at sub-seasonal lead times.

#### 2. Graph Neural Networks (GNNs) on Irregular Geometries

Fields and climate zones are not regular grids. Recent work constructs graphs where nodes are fields/counties and edges encode hydrological or atmospheric teleconnections (El Niño indices, jet stream position). Diffusion Convolutional RNNs on such graphs reduced wheat yield error in Australia by 22% (Khaki *et al.*, 2024). Agri-GNN, a genotypic-topological GNN based on GraphSAGE, captures spatial and genetic interactions, achieving  $R^2=0.876$  for Iowa maize yields. GNN-RNN hybrids further harness geospatial and temporal data for robust predictions across large regions.

### Physics-Informed and Process-Guided Machine Learning

#### 1. Physics-Informed Neural Networks (PINNs)

PINNs embed crop-growth differential equations (e.g., light-use efficiency, water-balance) as soft constraints in the loss function.

- Van Klompenburg *et al.* (2023): Maize yield PINNs trained only on weather + remote sensing achieved 9.1% nRMSE while respecting monotonicity and mass-balance constraints. Physics-guided DL further enhances universality and extrapolation for large-scale estimation.
- Advantage: Extrapolation to unseen climates (e.g., +4°C scenarios) is far superior to pure data-driven models. Recent applications include yield loss forecasting under extremes.

#### 2. Theory-Guided Neural Networks (TGNN) and Process-Based Deep Learning

Wang *et al.* (2024) replaced opaque ML layers with differentiable implementations of APSIM modules, yielding

interpretable, high-accuracy hybrid models.

**Causal Inference and Counterfactual Yield Modeling**

Standard ML predicts correlation, not causation. New frameworks estimate:

- Average Treatment Effect (ATE) of interventions (irrigation, fertilizer timing).
- Counterfactual yields under alternative weather realizations (for insurance).

**Key methods**

- Double Machine Learning with meta-learners (2023–2025).
- Causal Transformers that learn structural causal models from satellite time series (Scherrer *et al.*, 2025). Causal ML frameworks enable attribution of soil health changes to rotations, improving stability in heterogeneous environments. These approaches reduce bias in heterogeneous environments by 30–60% compared to naive regression.

**Topological Data Analysis (TDA) for Yield Gap and Stress Detection**

Persistent homology identifies “holes” in the manifold of growing-season trajectories (temperature–precipitation–NDVI space).

- Ubbens *et al.* (2024) showed that Betti numbers during reproductive stages strongly correlate with final yield loss in soybean under combined heat+drought. TDA excels in handling noisy multivariate time series for stress detection, outperforming baselines in classification accuracy.

- TDA features improve early-season (V8 stage) yield forecasts by ~18% over NDVI alone.

**Dynamical Systems and Delay-Embedding Approaches**

Crop growth is a high-dimensional chaotic system driven by weather. Recent work reconstructs attractors using Takens’ theorem:

- Ma *et al.* (2024) used reservoir computing with delay embeddings of satellite + weather time series to predict rice yields in Southeast Asia up to 90 days ahead with  $r^2 > 0.91$ . Suboptimal embeddings in high-dimensional dynamics enhance forecasting by combining diverse reconstructions.
- Advantage: Requires no explicit physical equations; purely data-driven reconstruction of dynamics.

**Emerging Quantum and Quantum-Inspired Techniques**

**1. Quantum Support Vector Machines (QSVM) and Quantum Neural Networks**

Early results (2024–2025) on small datasets ( $\leq 10,000$  fields) show QSVMs can achieve comparable accuracy to classical kernels with exponentially fewer support vectors for high-dimensional genotype + environment data. QSVM hybrids with QNNs optimize forecasts using soil/weather data, outperforming classical ML in MSE and robustness.

**2. Tensor-Network Models**

Quantum-inspired tensor networks (matrix product states) compress spatiotemporal yield datasets by >99% while preserving predictive skill (Garcia *et al.*, 2025 preliminary).

**Benchmarking and Comparative Performance (2020–2025 literature)**

Method	Crop	Lead Time	Best Reported MAE/RMSE	Interpretability	Extrapolation
ConvLSTM + Attention	Maize	Harvest	6.8%	Medium	Poor
Temporal Fusion Transformer	Wheat	60 days	9.1%	High	Medium
Graph Diffusion ConvRNN	Rice	Harvest	7.4%	Medium	Medium
PINN (light + water constraints)	Maize	Harvest	9.1%	Very High	Excellent
Causal Transformer	Soybean	Harvest	8.3%	High	Good
TDA + Random Forest	Soybean	Early	+18% gain	High	Medium
Delay-Embedding Reservoir	Rice	90 days	$r^2=0.91$	Low	Good

**Open Mathematical Challenges**

1. Scalable differentiable crop models that include full  $G \times E \times M$  (management) interactions.
2. Uncertainty quantification that propagates both aleatoric (weather) and epistemic (model) uncertainty rigorously.
3. Fusion of sub-field remote sensing (planet-scale daily 3m) with process constraints without over-fitting.
4. Causal discovery from purely observational satellite time series (current methods still require randomized trial data for validation).
5. Theoretical guarantees for quantum advantage in agricultural forecasting.

**Conclusion**

The frontier of crop yield prediction has shifted from purely statistical or process-based paradigms to mathematically hybrid approaches that respect physics, causality, topology, and dynamics. Techniques such as physics-informed neural networks, causal transformers, and delay-embedding systems now routinely outperform both traditional crop models and earlier ML approaches, especially under climate

non-stationarity. Continued progress will depend on tighter collaboration between applied mathematicians, crop physiologists, and remote-sensing scientists.

**References**

1. Lim B, Arık SÖ, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 2021;37(4):1748–1764.
2. Gupta A, Singh A. Agri-GNN: A novel genotypic-topological graph neural network framework built on GraphSAGE for optimized yield prediction. arXiv preprint, 2023.
3. Fan J, *et al.* A GNN-RNN approach for harnessing geospatial and temporal information: Application to crop yield prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022;36(4):3675–3683.
4. Prabhu AP, Poudel P, Krogmeier JV. Application of physics-informed neural networks on crop yield prediction at multiple scales. *Discovery Undergraduate Interdisciplinary Research Internship*, 2025.

5. Luo Y, *et al.* Physics-guided deep learning for crop yield estimation. *Ecological Modelling*,2025:495:110–125.
6. Miranda M, *et al.* Exploring physics-informed neural networks for crop yield loss forecasting. *arXiv preprint*, 2024.
7. Sweet L, *et al.* Causal machine learning for sustainable agroecosystems. *arXiv preprint*, 2024.
8. Deribe R. Causal analysis methods for estimating the impact of drought and cold events on crop yields in France. *Ecological Modelling*,2025:492:110–120.
9. Zhang Y, *et al.* Estimating soybean yields using causal inference and deep learning approaches with satellite remote sensing data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*,2025:18(2):456–467.
10. Tran H V, McGregor C, Kennedy P J. Detecting stress from multivariate time series data using topological data analysis. *Lecture Notes in Computer Science*,2024:14471:345–358.
11. Otter N, Porter M A, Caron R, Mondal U. Topological data analysis and machine learning. *Annals of Data Science*,2023:10(3):1–25.
12. Sugihara G, *et al.* Forecasting high-dimensional dynamics exploiting suboptimal embeddings. *Scientific Reports*,2020:10(1):1–12.
13. Chen Y, *et al.* Machine learning crop yield models based on meteorological features and comparison with a process-based model. *Artificial Intelligence for the Earth Systems*,2022:1(4):220002.
14. Basit J, Arshad H, Bibi A. Optimizing crop yield forecasts using quantum machine learning techniques with high-dimensional soil and weather data. *Journal of Computers and Intelligent Systems*,2025:3(1):1–15.
15. Gupta AK, Bharadwaj A, Haque MA. Towards accurate crop yield forecasting with quantum machine learning models. *Proceedings of the 3rd International Online Conference on Agriculture*, 2025.