



AGRI MAGAZINE

(International E-Magazine for Agricultural Articles)

Volume: 03, Issue: 10 (October, 2026)

Available online at <http://www.agrimagazine.in>

©Agri Magazine, ISSN: 3048-8656

Statistical Models for Crop Yield Prediction: Turning Weather, Soil and Crop Data into Better Agricultural Decisions

*Dr. Sachin S. Chinchorkar

Associate Professor, Polytechnic in Agricultural Engineering, Anand Agricultural University, Dahod, Gujarat, India

*Corresponding Author's email: csachin.chinchorkar@gmail.com

Crop yield is influenced by a complex interaction of weather, soil, crop variety, management practices, pests, diseases and environmental conditions. For farmers, researchers and policymakers, knowing the likely crop yield before harvest is extremely valuable for planning inputs, storage, marketing, food security and agricultural policies. Statistical models provide a scientific bridge between historical agricultural observations and future yield expectations. By analysing relationships between yield and variables such as rainfall, temperature, sunshine hours, humidity, soil characteristics and crop-growing conditions, these models can generate useful and location-specific yield forecasts. Regression models, correlation analysis, time-series models, weather-index models and multivariate statistical techniques have traditionally played an important role in agricultural forecasting. More recently, statistical approaches are increasingly being combined with remote sensing, geographic information systems, satellite-derived vegetation indices and machine-learning techniques. Recent research emphasizes that local and crop-specific models can be particularly valuable because the relationship between weather and yield may vary considerably between locations. Modern studies are therefore moving towards integrated, hybrid and data-driven approaches that combine the interpretability of statistical models with the predictive power of advanced computational methods. Statistical modelling should not be viewed merely as a mathematical exercise; it is a practical decision-support tool capable of converting agricultural data into meaningful information. When properly developed, validated and communicated, yield prediction models can help farmers reduce uncertainty, improve resource-use efficiency and build more climate-resilient agricultural systems.

Keywords: Crop yield prediction, statistical modelling, regression analysis, weather indices, agricultural forecasting, rainfall, climate variability, yield forecasting, precision agriculture, decision support system

Introduction

Agriculture is an activity of uncertainty. A farmer may sow a crop with great care, but the final harvest depends on many factors that cannot be completely controlled. Rainfall may arrive late, temperatures may rise unexpectedly, a dry spell may occur during flowering, or excessive rainfall may damage the crop near maturity. Soil fertility, irrigation, crop variety and farm management further influence the final yield. Crop yield prediction is not simply about guessing the future. It involves analysing historical relationships between crop yield and the factors that influence crop growth. Statistical models allow researchers to quantify these relationships and use them to estimate future production. Weather is particularly important. Rainfall and temperature influence germination, vegetative growth, flowering, grain formation and maturity. A recent systematic review of statistical and machine-learning approaches found that incorporating meteorological information is important for yield prediction and emphasized the value of location- and crop-specific models because climate-yield relationships can differ considerably between regions.

What is a Statistical Yield Prediction Model?

A statistical model is a mathematical representation of the relationship between crop yield and one or more influencing factors. For example, researchers may develop a model relating wheat yield to seasonal rainfall and temperature:

$$\text{Yield} = a + b_1(\text{Rainfall}) + b_2(\text{Temperature})$$

Here, yield is the dependent variable, while rainfall and temperature are explanatory variables. The coefficients indicate how changes in these variables are associated with changes in expected yield. The simplicity of such models is one of their greatest strengths. Farmers and agricultural planners can understand the relationship between the input variables and predicted yield without needing to understand highly complex computer algorithms.

Regression: The Foundation of Yield Prediction

Regression analysis has traditionally been one of the most widely used statistical approaches in agricultural forecasting. In simple regression, yield may be related to one major factor such as rainfall. In multiple regression, several variables can be considered simultaneously, such as rainfall, maximum temperature, minimum temperature, relative humidity and sunshine duration.

For example:

$$\text{Crop Yield} = f(\text{Rainfall} + \text{Temperature} + \text{Humidity} + \text{Soil} + \text{Management})$$

Multiple regression is particularly useful because crop yield rarely depends on a single factor. Statistical models can also include weather indices, which summarize rainfall or temperature during important crop-growth periods. Instead of treating all rainfall received during a season equally, researchers can distinguish rainfall during sowing, vegetative growth, flowering and grain-filling stages.

This can provide a more biologically meaningful prediction.

Time-Series Models: Learning from the Past

Agricultural production also has a strong historical pattern. Yield in one year is often related to technological progress, variety changes, irrigation development and previous production trends. Time-series models analyse these historical patterns to estimate future values. Such models can be useful when long-term yield records are available. They can identify trends, cycles and fluctuations and provide forecasts before the crop season or during its early stages. However, yield is affected by changing climate and management conditions. Therefore, relying only on historical yield trends may not always be sufficient. Combining historical yield with weather and environmental information can improve the usefulness of forecasting systems.

Weather-Based Yield Forecasting

Weather-based statistical forecasting is particularly important in rainfed agriculture. Rainfall distribution may be more important than seasonal rainfall alone. Two seasons may receive similar total rainfall but produce very different yields if rainfall occurs at different crop-growth stages. For example, adequate rainfall during germination and flowering may support good crop development, while a prolonged dry spell during flowering can cause substantial yield reduction.

Statistical models can therefore incorporate:

- Total seasonal rainfall
- Weekly or monthly rainfall
- Number of rainy days
- Dry-spell duration
- Maximum temperature
- Minimum temperature
- Relative humidity
- Sunshine duration
- Evaporation
- Growing degree days

Such information can help transform weather observations into practical yield forecasts.

From Statistical Models to Modern Data-Driven Agriculture

The agricultural data environment has changed dramatically. Earlier, yield prediction often depended on crop-cutting experiments, historical production data and weather observations.

Today, researchers can combine these traditional datasets with:

Satellite imagery + weather data + soil information + crop observations + farm management data

Satellite-derived vegetation indices such as NDVI and EVI provide information about crop growth and vegetation condition. Recent reviews show that environmental variables and vegetation indices are among the important inputs used in modern crop-yield prediction systems. This development does not make statistical models obsolete. Instead, it creates opportunities to develop hybrid models in which statistical methods provide interpretability while machine-learning methods identify complex patterns.

Statistical Models and Machine Learning: Competitors or Partners?

The agricultural forecasting community is increasingly moving away from the idea that one model must be universally superior. A linear regression model may be highly useful when relationships are relatively stable and understandable. Random Forest, Support Vector Machine, Gradient Boosting or neural networks may perform better when relationships are highly nonlinear. A 2025 systematic review of crop-yield prediction research found that Linear Regression, Random Forest and Gradient Boosting were among commonly applied machine-learning approaches, while CNN and LSTM were important deep-learning approaches.

The key lesson is simple:

The best model is not necessarily the most complicated model; it is the model that provides reliable predictions for the intended location, crop and decision.

Importance of Local Models

Agriculture is highly location-specific. A rainfall amount that is beneficial for one crop or region may be inadequate or excessive for another. Similarly, the response of a crop to temperature depends on variety, soil, irrigation and growth stage. Therefore, yield prediction models should ideally be calibrated and validated using local or regionally representative data. This is especially important in India, where agricultural conditions can change dramatically between districts and agro-climatic zones. Recent research specifically examining statistical and machine-learning yield prediction emphasizes the importance of location-specific modelling because climate–yield relationships can change at local scales.

How Accurate is a Yield Prediction?

No statistical model can predict crop yield with absolute certainty.

The objective is to reduce uncertainty sufficiently to support better decisions.

Researchers commonly evaluate prediction models using statistical indicators such as:

- R^2 (coefficient of determination) – indicates how much variation in yield is explained by the model.
- RMSE (Root Mean Square Error) – measures the magnitude of prediction errors.
- MAE (Mean Absolute Error) – represents the average absolute prediction error.
- MAPE (Mean Absolute Percentage Error) – expresses error in percentage terms.

Recent systematic reviews identify RMSE, R^2 and MAE among the most frequently used evaluation measures in crop-yield prediction research. However, statistical accuracy alone is not enough. A model should also be tested using independent data and under different weather conditions.

Benefits for Farmers and Policymakers

Reliable yield forecasts can provide benefits at several levels. For farmers, early yield information can support decisions about harvesting, storage, marketing and input management. For agricultural departments, forecasts can support estimates of regional

production. For food-security planning, early estimates can help governments anticipate potential shortages or surpluses. For markets, yield forecasts can provide useful information about expected supply. For crop insurance, reliable production estimates can potentially improve risk assessment. Thus, yield prediction is not only an academic exercise. It has direct economic and social importance.

The Future: Intelligent and Climate-Resilient Yield Forecasting

The future of agricultural forecasting will likely involve the integration of multiple data sources rather than dependence on a single variable. A future prediction system may combine:

Weather + soil + crop variety + remote sensing + farm management + historical yield + climate forecasts

Statistical models can provide the foundation for interpreting these relationships, while machine learning and artificial intelligence can help identify complex nonlinear patterns. A 2026 systematic review comparing historical and contemporary crop-yield prediction models reported that modern approaches integrating machine learning, deep learning and remote sensing can achieve higher predictive performance than many traditional approaches, while also emphasizing the continuing importance of contextual calibration, data integration and expert validation. The challenge now is to make these technologies affordable and accessible, especially for small and marginal farmers.

Conclusion

Crop yield prediction represents one of the most important applications of agricultural data science. Statistical models provide a scientifically sound way to transform historical yield, weather, soil and crop information into meaningful forecasts. Their greatest strength is not simply mathematical sophistication but interpretability and practical usefulness. The future will not necessarily belong to statistical models alone, nor to artificial intelligence alone. It will belong to integrated systems that combine statistics, agronomic knowledge, remote sensing, climate information and modern computing. In an increasingly uncertain climate, farmers cannot control every factor affecting their crops. But better prediction can help them prepare for uncertainty. Ultimately, the purpose of yield prediction is not to predict the harvest perfectly—it is to make agricultural decisions smarter before the harvest arrives.

References

1. Shawon, S.M., Ema, F.B., Mahi, A.K., Niha, F.L. & Zubair, H.T. (2025). Crop yield prediction using machine learning: An extensive and systematic literature review. *Smart Agricultural Technology*, 10, 100718. DOI
2. De Clercq, D. & Mahdi, A. (2025). Modern computational approaches for rice yield prediction: A systematic review of statistical and machine learning-based methods. *Computers and Electronics in Agriculture*, 109852. ScienceDirect
3. Javed, M.A. & Murad, M.A.A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon*, 10(24), e40836. DOI
4. Ajith, S., Debnath, M.K. & Karthik, R. (2024). Statistical and machine learning models for location-specific crop yield prediction using weather indices. *International Journal of Biometeorology*, 68, 2453–2475. PubMed
5. Hernández Hernández, G.C., Gómez Gómez, J. & Jiménez-Cabas, J. (2025). Predictive Models Based on Artificial Intelligence to Estimate Crop Yield: A Literature Review. *Agriculture*, 15(23), 2438. mdpi.com
6. Historical and contemporary crop yield prediction models: Key lessons and innovations (2026). *Smart Agricultural Technology*, 13, 101672.