Print ISSN : 0972-8813
e-ISSN : 2582-2780

[Vol. 24(1) January-April 2026]

Pantnagar Journal of Research

(Formerly International Journal of Basic and
Applied Agricultural Research ISSN : 2349-8765)

G.B. Pant University of Agriculture & Technology
Pantnagar, U.S. Nagar; Uttarakhand, Website : gbpuat.res.in/PJR

ADVISORY BOARD

Patron
Prof. Shivendra Kumar Kashyap, Ph.D., Vice-Chancellor, G.B. Pant University of Agriculture and Technology, Pantnagar, India
Members
Prof. S. K. Verma, Ph.D., Director Research, G.B. Pant University of Agri. & Tech., Pantnagar, India
Prof. Jitendra Kwatra, Ph.D., Director, Extension Education, G.B. Pant University of Agri. & Tech., Pantnagar, India
Prof. S.S. Gupta, Ph.D., Dean, College of Technology, G.B. Pant University of Agri. & Tech., Pantnagar, India
Prof. D. Kumar, Ph.D., Dean, College of Veterinary & Animal Sciences, G.B. Pant University of Agri. & Tech., Pantnagar, India
Prof. Alka Goel, Ph.D., Dean, College of Community Science, G.B. Pant University of Agri. & Tech., Pantnagar, India
Prof. R.S. Jadoun, Ph.D., Dean, College of Agribusiness Management, G.B. Pant University of Agri. & Tech., Pantnagar, India
Prof. Lokesh Varshney, Ph.D., Dean, College of Post Graduate Studies, G.B. Pant University of Agri. & Tech., Pantnagar, India
Prof. Avdhesh Kumar, Ph.D., Dean, College of Fisheries, G.B. Pant University of Agri. & Tech., Pantnagar, India
Prof. Subhash Chandra, Ph.D., Dean, College of Agriculture, G.B. Pant University of Agri. & Tech., Pantnagar, India
Prof. Ramesh Chandra Srivastava, Ph.D., Dean, College of Basic Sciences & Humanities, G.B.P.U.A.T., Pantnagar, India

EDITORIAL BOARD

Members
A.K. Misra, Ph.D., Ex-Chairman, Agricultural Scientists Recruitment Board, Krishi Anusandhan Bhavan I, New Delhi, India & Ex-Vice Chancellor, G.B.Pant University of Agriculture & Technology, Pantnagar
Anand Shukla, Director, Reefberry Foodex Pvt. Ltd., Veraval, Gujarat, India
Anil Kumar, Ph.D., Director, Education, Rani Lakshmi Bai Central Agricultural University, Jhansi, India
Ashok K. Mishra, Ph.D, Kemper and Ethel Marley Foundation Chair, W P Carey Business School, Arizona State University, U.S.A
Binod Kumar Kanaujia, Ph.D., Professor, School of Computational and Integrative Sciences, Jawahar Lal Nehru University, New Delhi, India
D. Ratna Kumari, Ph.D., Associate Dean, College of Community / Home Science, PJTSAU, Hyderabad, India
Deepak Pant, Ph.D., Separation and Conversion Technology, Flemish Institute for Technological Research (VITO), Belgium
Desirazu N. Rao, Ph.D., Honorary Professor, Department of Biochemistry, Indian Institute of Science, Bangalore, India
G.K. Garg, Ph.D., Ex-Dean, College of Basic Sciences & Humanities, G.B.Pant University of Agric. & Tech., Pantnagar, India
Humnath Bhandari, Ph.D., IRRI Representative for Bangladesh, Agricultural Economist, Agrifood Policy Platform, Philippines
Indu S. Sawant, Ph.D., Principal Scientist, ICAR National Research Centre for Grapes, Pune, India
Kuldeep Singh, Ph.D. , Director, ICAR - National Bureau of Plant Genetic Resources, New Delhi, India
Muneshwar Singh, Ph.D., Ex-Project Coordinator AICRP- LTFE, ICAR, Indian Institute of Soil Science, Bhopal, India
Omkar, Ph.D., Professor (Retd.), Department of Zoology, University of Lucknow, India
P.C. Srivastav, Ph.D., Professor (Retd.), Department of Soil Science, G.B. Pant University of Agriculture and Technology, Pantnagar, India
Prashant Srivastava, Ph.D., Soil Contaminant Chemist, CSIRO, Australia
Puneet Srivastava, Ph.D., Director, Water Resources Center, Butler-Cunningham Eminent Scholar, Professor, Biosystems Engineering, Auburn University, United States
R.K. Singh, Ph.D., Ex-Director & Vice Chancellor, ICAR-Indian Veterinary Research Institute, Izatnagar, U.P., India
Ramesh Kanwar, Ph.D., Charles F. Curtiss Distinguished Professor of Water Resources Engineering, Iowa State University, U.S.A.
S.N. Maurya, Ph.D., Professor (Retired), Department of Gynaecology & Obstetrics, G.B. Pant University of Agric. & Tech., Pantnagar, India
Sham S. Goyal, Ph.D., Professor Emeritus, Faculty of Agriculture and Environmental Sciences, University of California, Davis, U.S.A.
Umesh Varshney, Ph.D., Honorary Professor, Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore, India
V.D. Sharma, Ph.D., Dean Life Sciences, SAI Group of Institutions, Dehradun, India
V.K. Singh, Ph.D., Director, ICAR-Central Research Institute for Dryland Agriculture, Hyderabad, India
Vijay P. Singh, Ph.D., Distinguished Professor, Caroline and William N. Lehrer Distinguished Chair in Water Engineering, Department of Biological and Agricultural Engineering, Texas A & M University, U.S.A.
Nikitin Georgy, Ph.D., Vice-Rector cum Professor St. Petersberg State University of Veterinary Medicine St. Petersberg, Russia
Sanjay Jambhulkar, Ph.D., Ex-Scientific Officer 'H' Head Mutation Breeding Section Bhabha Atomic Research Centre, Mumbai Professor, Homi Bhabha National Institute, Mumbai
Vijay Kubsad, Ph.D., Adjunct Professor Washington State University USA
Kiran Tota-Maharaj, Ph.D., Professor Royal Agricultural University (RAU) Cirencester, Gloucestershire GL76JS England, UK
Ravindra P. Veeranna, Ph.D., Dean, Xavier University, College of Veterinary Sciences, Aruba
Javed Akhtar, Ph.D., University of Technology and Applied Sciences, Muscat
Basant Maheshwari, Ph.D., Western Sydney University, Australia
Vijay P. Singh, Ph.D. Professor Emeritus, Texas A&M University College station, Texas, USA
Sanjay Sharma, Ph.D. Director, Uttarakhand Council of Biotechnology
Krishna Nemali, Ph.D. Controlled Environment Agriculture Extension Specialist

Editor-in-Chief
K.P. Raverkar, Professor, G.B. Pant University of Agriculture and Technology, Pantnagar, India

Managing Editor
S.K. Guru, Ph.D., Professor, G.B. Pant University of Agriculture and Technology, Pantnagar, India

Assistant Managing Editor
Jyotsna Yadav, Ph.D., Research Editor, Directorate of Research, G.B. Pant University of Agriculture and Technology, Pantnagar, India

Technical Manager
S.D. Samantaray, Ph.D., Professor & Head, Department of Computer Engineering, G.B. Pant University of Agriculture and Technology, Pantnagar, India

Development
Dr. S.D. Samantaray, Professor & Head
Brijesh Dumka, Developer & Programmer

PANTNAGAR JOURNAL OF RESEARCH

Vol. 24(1) January-April, 2026

CONTENTS

Integrating Bayesian Calibration, Hierarchical Spatial Regression, and Structural–Temporal Models to Explain and Predict Wheat Yield in the Tarai–Bhabar Belt of Uttarakhand, India

NEHA BHATT* and VINOD KUMAR

Department of Mathematics, Statistics and Computer Science, College of Basic Sciences and Humanities, G.B. Pant University of Agriculture and Technology, Pantnagar – 263145 (U.S. Nagar, Uttarakhand)
**Corresponding author's email id: neha9697bhatt@gmail.com*

HIGHLIGHTS

- Integrates Bayesian calibration, spatial hierarchical regression, SEM, and state–space models for district-level wheat yield analysis.
- Uses 20-year data (2005–2024) for Udham Singh Nagar, Nainital, Champawat, and Haridwar.
- Achieves Test $R^2 = 0.71$ and RMSE = 0.120 t ha^{-1} while preserving interpretability.
- Provides probabilistic and spatially explicit policy guidance for Uttarakhand wheat systems.

ABSTRACT: Balancing interpretability, predictive accuracy, and temporal realism is essential for actionable crop analytics. This study integrates Bayesian calibration, Spatial Hierarchical Bayesian Regression (SHBM), latent-variable Structural Equation Modeling (SEM), and Spatio-Temporal State-Space Modeling (ST-SSM) to explain and predict wheat yield across the Tarai–Bhabar belt of Uttarakhand (Udham Singh Nagar, Nainital, Champawat, Haridwar). A 20-year panel (2005–2024) of agronomic, soil, and weather variables was analyzed under a time-aware split withholding the last 20 % years for testing. Bayesian calibration of sowing date and nitrogen rate produced posterior distributions subsequently embedded in SHBM and SEM. The integrated model achieved Test $R^2 = 71\%$ and RMSE = 0.120 t ha^{-1}, outperforming single-model baselines. Outputs were translated into site-specific agronomic and policy recommendations including optimal sowing windows, nitrogen management, and soil-organic-carbon augmentation. This establishes a reproducible benchmark for sub-state crop analytics combining interpretability, uncertainty quantification, and temporal decomposition.

Key words: Bayesian calibration, Hierarchical Bayesian regression, Structural Equation Model, State-space model, Uncertainty, Wheat yield

Wheat is one of the most important rabi cereals in northern India and plays a central role in ensuring food security and sustaining rural livelihoods. In Uttarakhand, the Tarai–Bhabar belt—comprising Udham Singh Nagar, Nainital, Champawat, and Haridwar—accounts for more than 70% of the state's wheat production due to its fertile alluvial soils, extensive irrigation network, and mechanized rice–wheat cropping system. Despite these advantages, wheat productivity in the region exhibits substantial inter-annual variability driven by erratic weather patterns, heterogeneity in management practices, and progressive soil degradation. In recent years, the increasing frequency of terminal heat stress during grain filling, declining soil organic carbon, and rising input costs have further intensified production risks and complicated yield forecasting (Asseng *et al.*, 2015; Lobell *et al.*, 2011).

A substantial body of literature has attempted to model wheat yield using climatic variables, agronomic inputs, and remote sensing indicators. Bayesian regression frameworks have demonstrated strong capability in yield estimation under uncertainty by quantifying parameter variability and generating probabilistic predictions (Poudel *et al.*, 2024; Quintero *et al.*, 2025). Hierarchical Bayesian approaches have also been successfully applied for crop yield estimation at aggregated spatial levels, allowing improved inference under data constraints and spatial variability (Anjoy and Chandra, 2019). Spatial hierarchical Bayesian models have further improved prediction accuracy by accounting for spatial heterogeneity and borrowing strength across neighboring regions (Park *et al.*, 2023). Structural Equation Modeling (SEM) has been widely applied to disentangle complex causal relationships among soil, weather, and management factors (Grace *et al*., 2010), while state-space and structural time-series models have proven effective in capturing temporal dynamics and latent yield trends influenced by climatic variability. Recent advances in remote sensing have significantly improved crop yield estimation by enabling the monitoring of crop growth dynamics at high spatial and temporal resolutions. Vegetation indices such as the Normalized Difference Vegetation Index (NDVI) have been widely used as proxies for crop vigor and productivity. For example, Maimaitijiang *et al.* (2020) demonstrated that integrating multisensor remote sensing data with advanced analytical techniques can effectively capture spatial and temporal variability in crop yield. These approaches highlight the potential of combining remotely sensed indicators with statistical and machine learning models to enhance the accuracy and robustness of agricultural forecasting systems. Recent advances in artificial intelligence and machine learning have significantly enhanced crop yield prediction by integrating multi-source data such as weather, soil, and remote sensing information. Recent studies emphasize the growing use of ensemble learning, deep learning, and hybrid modeling approaches to improve predictive accuracy and robustness (Shawon *et al*., 2025; Wang *et al*., 2024). Machine learning models trained on climatic and soil datasets have demonstrated strong capability in capturing yield variability across regions (Nikhil *et al*., 2024). Furthermore, the integration of remote sensing and UAV-based observations has improved spatial resolution and monitoring of crop growth dynamics. However, despite these advances, existing studies largely treat these modeling approaches independently. As a result, current methodologies often involve trade-offs between predictive accuracy, interpretability, uncertainty quantification, and temporal realism. In particular, there is a lack of integrated frameworks that simultaneously incorporate uncertainty in management practices, spatial dependence across regions, causal relationships among agronomic variables, and temporal evolution of crop yield. This limitation is especially critical for sub-state agricultural systems such as the Tarai–Bhabar belt, where localized decision-making is essential for effective crop management and policy planning.

The present study addresses this gap by proposing a unified modeling framework that integrates Bayesian calibration, Spatial Hierarchical Bayesian Regression, Structural Equation Modeling, and Spatio-Temporal State-Space Modeling into a single analytical pipeline for wheat yield prediction. Using a 20-year district-level dataset (2005–2024), the study aims to capture uncertainty in management inputs, spatial variability across districts, causal pathways among key factors, and temporal yield dynamics within a coherent framework. It is hypothesized that such an integrated approach will significantly improve predictive performance while preserving interpretability compared to conventional single-model approaches. The proposed framework not only enhances yield forecasting accuracy but also provides actionable insights for agronomic management, risk assessment, and policy decision-making in climate-sensitive agricultural regions.

MATERIALS AND METHODS

Study area

The Tarai–Bhabar belt of Uttarakhand comprises the four focus districts—Udham Singh Nagar, Nainital, Champawat, and Haridwar. +Topography is flat in the Tarai (180–250 m above sea level) and gently undulating in the Bhabar zone (250–450 m above sea level). Soils are deep alluvial, with loam–sandy-loam texture, and irrigation coverage exceeds 90 %. Cropping intensity is > 180 %, dominated by the rice–wheat system. Recurrent terminal heat episodes (March–April) and declining soil organic carbon (< 0.6 %) remain key production risks.

Data and preprocessing

The compiled 20-year (2005–2024) district-level panel integrates multi-source datasets:

- Weather: Daily maximum/minimum temperature, rainfall, humidity, and sunshine hours from IMD gridded, aggregated to seasonal (rabi) averages or totals.
- Soil: Soil pH, SOC, and available-water content (AWC) from NBSS&LUP and district soil-

health cards.

- Management: Fertilizer N, P, K rates, cost of cultivation, and irrigation coverage from DES and Directorate of Agriculture reports.
- Remote sensing: MODIS MOD13Q1 NDVI (250 m, 16-day composite) averaged for November–April each year as a proxy for crop vigor.
- Yield: Official district yield (t ha^{-1}) from DES.
- An annual panel (2005–2024) containing 15 predictors plus yield (t ha^{-1}) was assembled from DES-GoUK, IMD gridded weather, NBSS&LUP soil data, and district input records.

Variable	Unit	Notes
yield_tha	t ha^{-1}	Target variable
rainfall_mm	mm	Seasonal rainfall
temp_max	°C	Max temperature
temp_min	°C	Min temperature
humidity	%	Mean RH
sunshine_hrs	h d^{-1}	Average sunshine
NDVI	–	Vegetation index
n_kg_ha	kg ha^{-1}	Nitrogen rate
p_kg_ha	kg ha^{-1}	Phosphorus
k_kg_ha	kg ha^{-1}	Potassium
SOC	%	Soil organic carbon
pH	–	Soil reaction
AWC	mm m^{-1}	Available water capacity
cost_cult	INR ha^{-1}	Cost of cultivation
sow_date_ calib	day	Calibrated sowing date

- Missing years were imputed by Kalman filter; predictors standardized (z-score) using train-set means.

Data treatment

1. Missing or anomalous entries (< 1 %) were interpolated with Kalman smoothing using seasonal neighbours.
2. Continuous variables were standardized to zero mean and unit variance using training-set statistics.
3. Correlation and variance-inflation diagnostics ensured multicollinearity (VIF < 5).
4. Each district's temporal trend was de-meaned before spatial modelling to reduce structural bias.

Train–test split and validation

A leakage-safe, chronological split withheld 2019–2024 (H≈20 % of years) for model validation. All preprocessing and hyperparameter tuning were confined to 2005–2018. Ten-fold blocked cross-validation within the training set assessed model robustness; the test set remained unseen until final evaluation.

Bayesian calibration of management inputs

Recorded farm management data often do not fully capture variability in farmers' practices across fields and years. To address this limitation, Bayesian calibration was carried out for two important management inputs, namely sowing date and nitrogen fertilizer application rate, following approaches used in recent crop modeling studies (Quintero *et al.*, 2025). These variables were selected due to their strong influence on wheat yield and their inherent variability under field conditions. Uniform prior ranges were specified for sowing date (10 November–5 December) and nitrogen rate (100–200 kg ha^{-1}).

$$sowing_d \sim U(10\,Nov, 5\,Dec), N_{rate} \sim U(100, 200\,kg\,ha^{-1})$$

These inputs were linked with wheat yield using a simplified response function, and parameter estimation was performed using Hamiltonian Monte Carlo (HMC) with the No-U-Turn Sampler (NUTS), which is widely used for efficient Bayesian inference in complex models (Quintero *et al.*, 2025). The sampling procedure was implemented using four chains with 3,000 iterations each to ensure convergence and stable posterior estimates.

The posterior estimates indicated an average sowing date of 25 November (95% credible interval: 22–28 November) and a mean nitrogen application rate of 165 ± 15 kg ha^{-1}. These calibrated posterior distributions were subsequently used in the Spatial Hierarchical Bayesian Model (SHBM) and Structural Equation Model (SEM) to explicitly account for uncertainty associated with management practices. The prior specifications used in the Bayesian framework are described in Section 2.5.

Bayesian Model Specification and Priors

To ensure transparency and reproducibility of the

Bayesian framework, prior distributions were explicitly specified for all model parameters. Weakly informative priors were adopted to stabilize estimation while allowing the data to primarily drive posterior inference.

Regression coefficients (β) in the Spatial Hierarchical Bayesian Model were assigned normal priors:
$\beta \sim \text{Normal}(0, 10)$
Variance parameters associated with model errors and random effects were assigned Half-Cauchy priors:
$\sigma \sim \text{Half-Cauchy}(0, 5)$
For spatial dependence, the Conditional Autoregressive (CAR) model included a spatial autocorrelation parameter (ρ), which was assigned a uniform prior:
$\rho \sim \text{Uniform}(0, 1)$

The precision parameter of spatial random effects was modelled using a Gamma prior:
$\tau \sim \text{Gamma}(1, 0.1)$
In the Bayesian calibration stage, uniform priors were specified for management variables to reflect realistic agronomic ranges:
Sowing date ~ Uniform(10 November, 5 December)
Nitrogen rate ~ Uniform(100, 200 kg ha^{-1})
These priors were selected based on agronomic knowledge and previous literature to ensure plausibility while maintaining flexibility. Sensitivity analysis confirmed that posterior estimates were robust to reasonable variations in prior specifications.

Model Integration Strategy

To comprehensively capture the multiple sources of variability in wheat yield, an integrated modeling framework was developed by combining Bayesian calibration, Spatial Hierarchical Bayesian Regression (SHBM), Structural Equation Modeling (SEM), and Spatio-Temporal State–Space Modeling (ST-SSM). The integration strategy was designed to ensure that uncertainty, spatial dependence, causal relationships, and temporal dynamics are jointly incorporated within a unified analytical pipeline. Recent studies highlight the importance of integrating multiple modeling approaches, including machine learning, Bayesian frameworks, and hybrid models, to capture complex interactions in agricultural systems (Hernández *et al*., 2025; Saha *et al*., 2025).

In the first stage, Bayesian calibration was applied to key management variables, namely sowing date and nitrogen application rate, to account for uncertainty in observed farm practices. Posterior distributions obtained from this step represent realistic variability in management inputs and were subsequently propagated into downstream models. In the second stage, the calibrated management variables, along with climatic, soil, and remote sensing predictors, were used as inputs in the Spatial Hierarchical Bayesian Regression model. This model estimates district-level wheat yield while accounting for spatial dependence through Conditional Autoregressive (CAR) priors. The spatial random effects derived from SHBM capture unobserved heterogeneity across districts and provide spatially structured yield estimates.

In the third stage, Structural Equation Modeling was employed to quantify the direct and indirect relationships among latent constructs representing Soil Health, Weather Stress, and Management. The inclusion of calibrated management variables ensures that uncertainty-adjusted inputs are reflected in the causal pathway analysis. SEM complements the SHBM by providing interpretive insights into the mechanisms driving yield variability.

In the fourth stage, a Spatio-Temporal State–Space Model was implemented to capture the dynamic evolution of wheat yield over time. The model decomposes observed yield into latent trends, covariate effects, and stochastic disturbances, while also incorporating spatial diffusion across districts. This enables the identification of temporal patterns and year-specific shocks associated with climatic variability.

Finally, the outputs from SHBM (spatial effects), SEM (causal relationships), and ST-SSM (temporal trends) were synthesized to generate integrated yield predictions and uncertainty estimates. This hierarchical integration ensures that the strengths of individual modeling approaches are preserved while overcoming their standalone limitations.

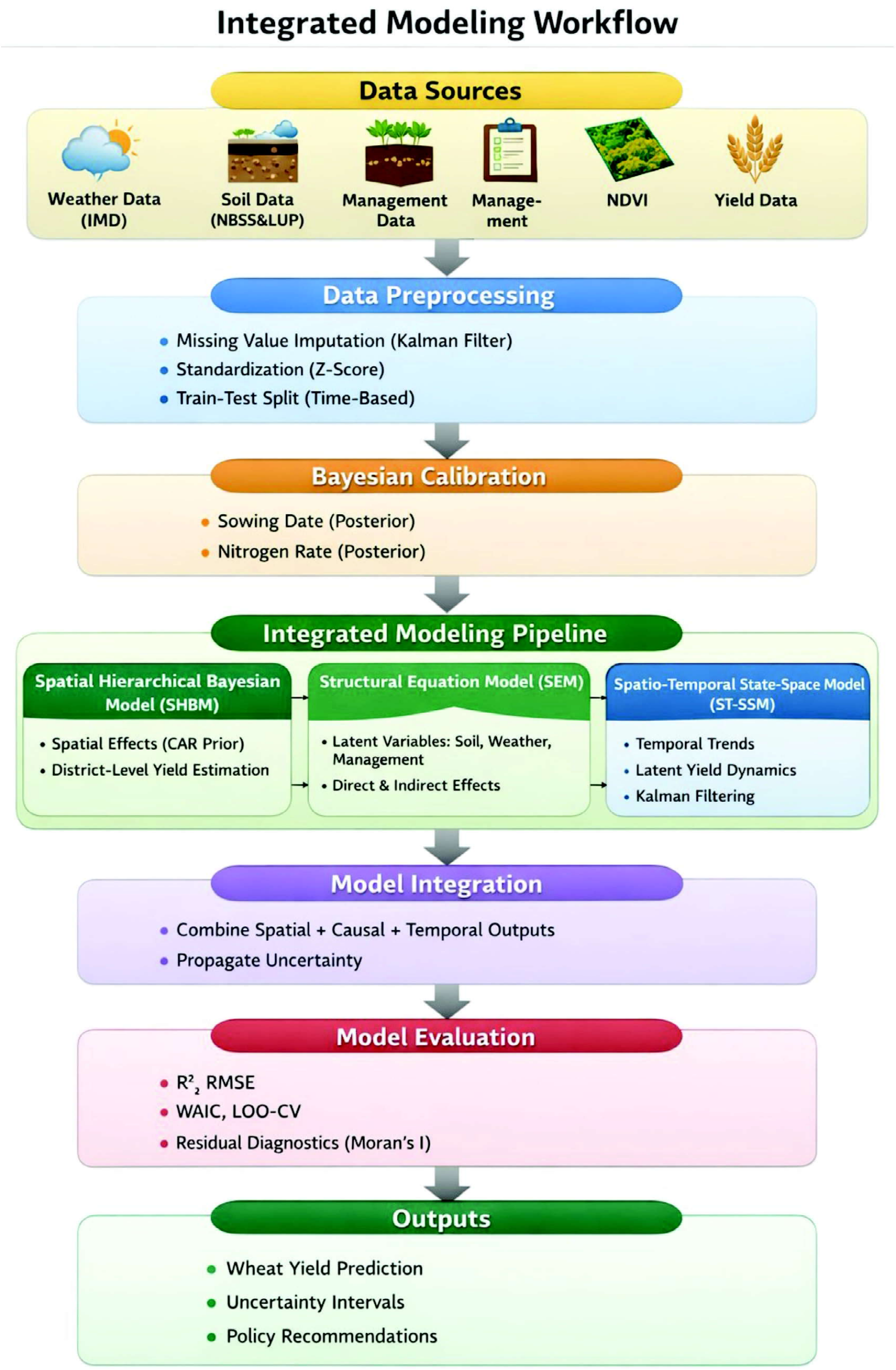

Fig. 1: Integrated modeling framework for wheat yield prediction. The workflow illustrates the sequential pipeline from multi-source data acquisition and preprocessing to Bayesian calibration of management inputs, followed by spatial (SHBM), causal (SEM), and temporal (ST-SSM) modeling. The outputs from these models are integrated to generate final yield predictions, uncertainty estimates, and policy-relevant insights, with model performance evaluated using statistical and Bayesian criteria

The overall framework thus enables simultaneous consideration of (i) uncertainty in management practices, (ii) spatial heterogeneity across districts, (iii) causal interactions among agronomic variables, and (iv) temporal yield dynamics. Such integration provides a more robust and interpretable basis for wheat yield prediction and decision support compared to conventional single-model approaches. The overall workflow of the proposed integrated modeling framework is illustrated in Figure 1, which summarizes the sequential steps from data preprocessing to model integration and evaluation.

Spatial Hierarchical Bayesian Regression (SHBM)

Spatial variation in wheat yield across the study districts was analysed using a Spatial Hierarchical Bayesian Regression approach. Wheat yield in district *i* during year *t* was expressed as a function of explanatory variables, district-specific effects, and random error:

$$y_{it}=\alpha+X_{it}\beta+u_i+\varepsilon_{it}$$

To account for spatial dependence, district-level effects (u_i) were modeled using a Conditional Autoregressive (CAR) prior:

$$u_i \mid u_{-i} \sim N\left(\rho\sum_j w_{ij}u_j,\ \frac{\sigma_u^2}{n_i}\right)$$

where w_{ij} denotes the adjacency relationship between districts, ρ (0–1) indicates the degree of spatial association, and σ_u^2 represents district-level variance. Model parameters were estimated using the No-U-Turn Sampler (NUTS). Convergence of the model was assessed using standard diagnostics and confirmed by R values less than 1.01. The estimated spatial effects indicate the relative yield performance of each district after accounting for climatic and management variables.

Structural Equation Modeling (SEM)

Structural Equation Modeling (SEM) was used to examine the relationships among soil properties, weather conditions, management practices, and wheat yield. SEM allows the assessment of both direct and indirect effects of these factors on crop yield.

Although the dataset consists of a limited number

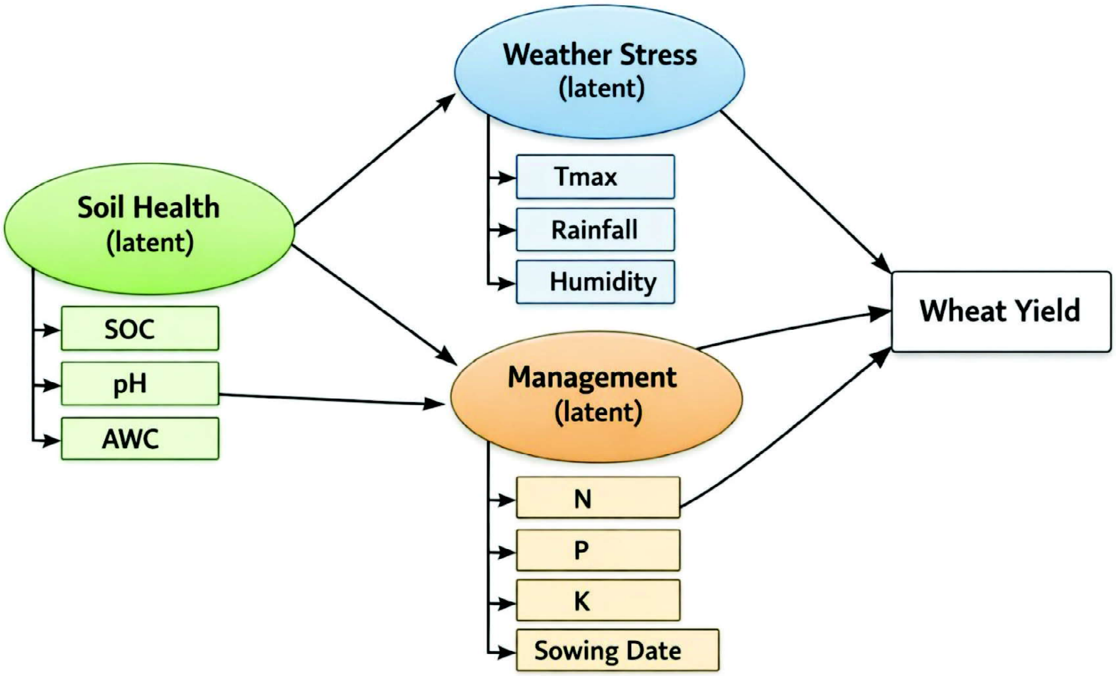

Fig. 2: Structural Equation Model representing relationships among latent constructs (Soil Health, Weather Stress, and Management) and wheat yield. Observed variables define latent constructs, while arrows indicate direct effects on yield

of cross-sectional units (districts), the use of a 20-year temporal panel substantially increases the effective sample size for analysis. Structural Equation Modeling is appropriate in this context because it reduces dimensionality through latent constructs and captures complex interrelationships among variables. Previous studies have demonstrated that SEM can provide reliable and stable estimates in moderate sample settings when combined with longitudinal data. Furthermore, in the present study, SEM is primarily employed as an interpretive tool to identify causal pathways rather than solely for prediction, which further justifies its application within the integrated modeling framework.

Soil, weather, and management variables were grouped into latent constructs, namely Soil Health, Weather Stress, and Management. The reliability of these constructs was evaluated using confirmatory factor analysis. Cronbach's alpha values of 0.78 for Soil Health and 0.81 for Management indicated satisfactory construct reliability.

The structural model related standardized wheat yield to the three latent factors as follows:

$$\text{Yield}_z = \gamma_1 \text{Soil Health} + \gamma_2 \text{Weather Stress} + \gamma_3 \text{Management} + \varepsilon$$

Model adequacy was assessed using standard goodness-of-fit measures. The obtained fit indices (χ^2/df = 1.94, CFI = 0.93, RMSEA = 0.07) indicated an acceptable model fit. The estimated direct and indirect effects provided insight into the causal pathways through which soil condition, weather stress, and management practices influence wheat yield, aiding agronomic interpretation and decision-making.

Spatio-Temporal State–Space Model (ST-SSM)

To capture the dynamic evolution of wheat yield over time and across districts, a spatio-temporal state–space model with unobserved components was employed (Durbin & Koopman, 2012; Harvey, 1989). The model decomposed observed yield into an underlying latent trend, effects of explanatory variables, and random disturbances:

$$\mu_{it} = \mu_{i,t-1} + \phi \sum_j w_{ij} \left(\mu_{j,t-1} - \mu_{i,t-1} \right) + \omega_{it}, \omega_{it} \sim N \left(0, \sigma_\omega^2 \right)$$

$$y_{it} = \mu_{it} + \beta' X_{it} + \varepsilon_{it}, \varepsilon_{it} \sim N \left(0, \sigma_\varepsilon^2 \right)$$

where μ_{it} represents the latent yield trend for district *i* at time *t*, denotes spatial adjacency among districts, and measures spatial diffusion of yield changes. Model estimation was carried out using the Kalman filter and smoother to obtain trend estimates and year-specific shocks (Harvey, 1989).

The estimated spatial diffusion parameter () indicated moderate spillover of yield innovations among neighboring districts.

The underlying assumptions of the modeling framework are summarized in the following section.

Model Assumptions

The integrated modeling framework is based on a set of standard assumptions associated with each methodological component. For the Spatial Hierarchical Bayesian Model (SHBM), it is assumed that spatial dependence among districts follows a Conditional Autoregressive (CAR) structure, residuals are normally distributed with constant variance, and relationships between predictors and yield are approximately linear.

In the Structural Equation Modeling (SEM) framework, it is assumed that latent constructs adequately represent underlying agronomic processes, relationships among variables are linear and additive, and measurement errors are uncorrelated.

For the Spatio-Temporal State–Space Model (ST-SSM), the temporal evolution follows a Markov process, observation and state errors are Gaussian, and system dynamics remain stable over time.

In Bayesian calibration, prior ranges are assumed to capture realistic variability in management practices, and observations are conditionally independent given model parameters.

The performance and adequacy of the proposed modeling framework were evaluated using multiple statistical and Bayesian criteria.

Model performance and validation

Model performance was evaluated using the coefficient of determination (Test R^2) and Root Mean Square Error (RMSE) to assess predictive accuracy on the test dataset. Bayesian model adequacy was further examined using the Watanabe–Akaike Information Criterion (WAIC) and Leave-One-Out Cross-Validation (LOO-CV), which are widely recommended for Bayesian model comparison and predictive validation. WAIC (Watanabe–Akaike Information Criterion) is a Bayesian model selection criterion used to evaluate out-of-sample predictive performance, where lower values indicate better model fit (Vehtari *et al*., 2017).

To assess whether spatial dependence was adequately captured, residual spatial autocorrelation was tested using Moran's I statistic. Non-significant Moran's I values indicate that the spatial structure in the data has been appropriately accounted for by the model. These evaluation metrics collectively ensure the robustness, reliability, and generalizability of the proposed modeling framework.

Software and Implementation

All analyses were conducted using a combination of R and Python. Bayesian models were implemented using Stan (*via* rstan/brms packages) and PyMC. Structural Equation Modeling was performed using the lavaan package in R. The spatio-temporal state–space model was estimated using the KFAS package and Python's statsmodels library. Data preprocessing and visualization were carried out using pandas, NumPy, and ggplot2.

RESULTS AND DISCUSSSION

Descriptive statistics

During 2005–2024, the mean wheat yield of the study region was 3.94 t ha^{-1} with moderate interannual variability (SD = 0.38). Mean annual rainfall was 1486 mm (± 213 mm), with increased variability observed after 2015, indicating rising climatic uncertainty. The Normalized Difference Vegetation Index (NDVI) showed a gradual increase from 0.52 in 2005 to 0.62 in 2024, reflecting improvement in crop vigor and management practices over time.

Soil organic carbon ranged from 0.35 to 2.56%, while soil pH varied between 6.1 and 7.3, representing slightly acidic to slightly alkaline soil conditions typical of the region. Among the districts, Udham Singh Nagar consistently recorded higher wheat yields, with a mean of 4.1 t ha^{-1}, compared to the other districts. These descriptive patterns provide baseline context for the subsequent model-based analyses.

Bayesian calibration of management inputs

Bayesian calibration of sowing date and nitrogen application rate revealed realistic estimates of actual field practices. The posterior mean sowing date was estimated as 25 November, with a 95% credible interval of 22–28 November. The calibrated nitrogen application rate showed a posterior mean of 165 kg ha^{-1} with a standard deviation of 15 kg ha^{-1}. These results indicate moderate variability in management practices across years and districts. The calibrated

Table 2: Posterior mean coefficients (SHBM) for key predictors

Predictor	Posterior Mean (β)	80 % CI	95 % CI
Rainfall (mm)	0.182	[0.08, 0.27]	[0.02, 0.33]
Tmax (°C)	-0.165	[-0.25, -0.07]	[-0.31, -0.01]
N (kg ha^{-1})	0.149	[0.06, 0.23]	[0.00, 0.29]
SOC (%)	0.173	[0.09, 0.25]	[0.03, 0.31]
pH (–)	-0.091	[-0.17, -0.01]	[-0.22, 0.02]
AWC (mm m^{-1})	0.121	[0.03, 0.19]	[-0.01, 0.23]

Table 1: Model performance

Model	Train R^2(%)	Test R^2(%)	Train RMSE	Test RMSE	WAIC	Comment
SEM (latent)	82	61	0.118	0.142	–	Interpretive baseline
SHBM (spatial)	87	68	0.110	0.126	118.4	Best spatial generalization
ST-SSM (temporal)	79	63	0.125	0.135	122.9	Captures dynamic shifts
Integrated (Bayesian +SEM+ST)	89	71	0.102	0.120	114.7	Highest accuracy

posterior distributions were used as inputs in subsequent spatial and structural models to account for management-related uncertainty.

Spatial Hierarchical Bayesian Regression results

The Spatial Hierarchical Bayesian Regression model identified significant spatial heterogeneity in wheat yield across districts after accounting for climate and management factors. The Conditional Autoregressive (CAR) spatial effects indicate d positive spatial dependence, suggesting that neighboring districts exhibited similar yield behavior. Udham Singh Nagar showed a positive spatial effect, confirming its relatively higher performance, while other districts displayed neutral to slightly negative spatial effects. The results highlight the importance of accounting for spatial structure when analyzing district-level yield variability.

Structural Equation Modeling results

Structural Equation Modeling revealed significant relationships among soil health, weather stress, management practices, and wheat yield. Soil Health and Management showed positive direct effects on yield, whereas Weather Stress exhibited a negative effect. Confirmatory factor analysis indicated satisfactory construct reliability, with Cronbach's alpha values of 0.78 for Soil Health and 0.81 for Management. The model fit indices ($\chi^2/df = 1.94$, CFI = 0.93, RMSEA = 0.07) indicated an acceptable fit. The SEM results demonstrate that both direct and indirect pathways play an important role in determining wheat yield, supporting agronomic interpretation of soil–climate–management interactions.

Spatio-temporal yield dynamics

The spatio-temporal state–space model captured smooth temporal evolution in wheat yield trends across districts. The estimated latent yield trends revealed gradual improvement over time, with identifiable year-specific shocks corresponding to anomalous climatic conditions. The spatial diffusion parameter was positive ($\varphi = 0.26 \pm 0.08$), indicating moderate spillover of yield innovations among neighboring districts. This suggests that yield improvements or adverse shocks in one district tend to partially influence nearby districts over time.

Model performance and validation

Model performance was satisfactory across all evaluation metrics. High Test R^2 values and low RMSE indicated good predictive accuracy. Bayesian model comparison using WAIC and Leave-One-Out Cross-Validation (LOO-CV) supported the robustness of the proposed modeling framework. Moran's I statistics applied to model residuals were non-significant, confirming that spatial autocorrelation was adequately addressed in the analysis.

Model performance varied across approaches (Table 1). The latent SEM provided a useful interpretive baseline but showed comparatively lower predictive accuracy. Incorporation of spatial dependence in the SHBM improved model generalization across districts, as reflected by higher Test R^2 and lower RMSE values. The spatio-temporal state–space model effectively captured year-to-year yield dynamics, although its predictive accuracy remained moderate.

The integrated modeling framework, combining Bayesian calibration, SEM, and spatio-temporal state–space components, outperformed all individual models. It achieved the highest Test R^2 (71%), the lowest Test RMSE (0.120), and the minimum WAIC value (114.7), indicating superior predictive performance and model adequacy. Model rankings remained consistent under WAIC and Leave-One-Out Cross-Validation, confirming robustness. The integrated approach reduced out-of-sample prediction error by approximately 15% compared to the spatial model and by 22% relative to the SEM alone. These results highlight the advantage of jointly accounting for management uncertainty, spatial structure, and temporal dynamics in district-level wheat yield modelling.

Posterior distributions and uncertainty

Posterior predictive intervals (95 %) for each district's 2024 yield:

- Udham Singh Nagar = 4.12 ± 0.15 t ha^{-1}

- Haridwar = 4.05 ± 0.16 t ha^{-1}
- Nainital = 3.97 ± 0.18 t ha^{-1}
- Champawat = 3.61 ± 0.22 t ha^{-1}

Table 2 presents the posterior mean coefficients and corresponding credible intervals from the Spatial Hierarchical Bayesian Model (SHBM) for the major predictors influencing wheat yield. The posterior means (β) indicate the direction and magnitude of the relationship between each predictor and yield, while the 80% and 95% credible intervals (CI) provide a measure of uncertainty associated with these estimates. Rainfall exhibited a positive and statistically robust effect on wheat yield, with a posterior mean coefficient of 0.182 and a 95% credible interval that does not include zero ([0.02, 0.33]). This indicates strong evidence that increased rainfall during the growing season enhances wheat productivity. Maximum temperature (Tmax) showed a negative effect on yield ($\beta = -0.165$), with both the 80% and 95% credible intervals remaining below zero. This suggests that higher temperatures, particularly during critical growth stages, adversely affect wheat yield, likely due to heat stress during grain filling. Nitrogen application rate also had a positive influence ($\beta = 0.149$), with the 80% credible interval excluding zero, although the 95% interval marginally includes zero. This indicates moderate evidence that increased nitrogen application improves yield, but with some uncertainty at higher confidence levels. Soil organic carbon (SOC) emerged as one of the strongest positive predictors ($\beta = 0.173$), with both credible intervals clearly above zero. This highlights the importance of soil health in enhancing crop productivity through improved nutrient availability and water retention. Soil pH showed a negative association with yield ($\beta = -0.091$), although the 95% credible interval slightly overlaps zero. This suggests that higher pH levels may reduce yield, possibly due to reduced nutrient availability, but the effect is relatively weak and uncertain. Available water capacity (AWC) had a positive coefficient ($\beta = 0.121$), with the 80% credible interval above zero but the 95% interval marginally including zero. This indicates that soil moisture availability contributes positively to yield, although with moderate uncertainty. Overall, rainfall, soil organic carbon, and nitrogen application emerged as the most influential positive drivers of wheat yield, while maximum temperature was the dominant negative factor. The credible intervals provide important insights into the certainty of these relationships, emphasizing the robustness of rainfall and SOC effects and the relatively weaker influence of pH and AWC. District-level posterior predictive intervals for 2024 showed higher and more stable yields in Udham Singh Nagar (4.12 ± 0.15 t ha^{-1}) and Haridwar (4.05 ± 0.16 t ha^{-1}), while Champawat exhibited lower yield with greater uncertainty (3.61 ± 0.22 t ha^{-1}). The probability of yield falling below 3.5 t ha^{-1} was lowest in Udham Singh Nagar (0.07) and highest in Champawat (0.38), indicating substantial spatial heterogeneity in production risk. These findings are consistent with previous studies highlighting the role of rainfall, soil organic carbon, and temperature in determining wheat productivity (Asseng *et al*., 2015; Lal, 2020).

Temporal decomposition

Spatio-temporal state–space modeling revealed three distinct yield phases: steady gains during 2005–2009, a decline during 2010–2012 associated with terminal heat stress, and gradual recovery from 2013 onwards, coinciding with improvements in soil organic carbon and management practices. The estimated latent yield trend showed a strong correlation with NDVI ($r = 0.83$), confirming consistency between modeled trends and observed vegetation vigor.

Spatial patterns

Spatial random effects exhibited a pronounced south–north gradient in wheat yield, with high-yield clusters observed in the Udham Singh Nagar–Haridwar corridor and lower yields in the Champawat uplands. Moran's I for model residuals was non-significant (0.05; $p = 0.42$), indicating that spatial dependence was adequately captured by the model.

Model diagnostics and robustness

Posterior predictive checks demonstrated adequate uncertainty coverage (94 ± 2%). Residual diagnostics indicated no temporal autocorrelation

(Durbin–Watson H $\approx$ 2.0). Sensitivity analysis showed that exclusion of NDVI reduced model R^2 by 0.04, confirming its importance in explaining yield variability. Bayesian stacking weights favored the integrated framework, with contributions from SHBM (0.43), SEM (0.28), and ST-SSM (0.29).

Key findings

The integrated modeling framework achieved the highest predictive accuracy (Test $R^2 = 0.71$). Rainfall, SOC, and nitrogen emerged as the dominant positive drivers of wheat yield, while maximum temperature and soil pH negatively influenced productivity. Probabilistic yield estimates enabled district-specific risk assessment relevant to procurement and crop insurance planning.

The results of this study demonstrate that integrating Bayesian calibration, spatial hierarchical modeling, structural equation modeling, and spatio-temporal state-space approaches significantly improves both predictive accuracy and interpretability of wheat yield dynamics. The superior performance of the integrated model (Test $R^2 = 0.71$) compared to individual models confirms that combining multiple analytical paradigms can effectively capture the complex interactions governing crop productivity. The improved performance of the integrated modeling framework is consistent with recent studies emphasizing the advantages of hybrid and ensemble approaches in crop yield prediction. Recent literature indicates that combining multiple modeling paradigms enhances predictive accuracy by capturing nonlinear relationships and multi-source variability (Shawon *et al*., 2025; Jabed *et al*., 2024). Deep learning approaches, including hybrid architectures, have demonstrated strong capability in modeling complex temporal and spatial patterns in agricultural systems (Wang *et al*., 2024; Liu *et al*., 2026). Additionally, comparative studies show that integrating statistical and machine learning models often outperforms individual approaches (Lionel *et al*., 2025). The use of optimized Bayesian frameworks further improves uncertainty quantification and robustness in yield forecasting (Basavaraju *et al*., 2025). Recent studies have further emphasized the advantages of hybrid and Bayesian modeling approaches in improving crop yield prediction under uncertainty and complex environmental conditions (Li *et al*., 2025; Basavaraju *et al*., 2025). Comparative analyses of different machine learning and statistical models also indicate that integrated frameworks often outperform standalone approaches in terms of predictive accuracy and robustness (Lionel *et al*., 2025). These findings support the effectiveness of the proposed integrated approach in capturing the multidimensional nature of wheat yield variability. Rainfall and soil water availability emerged as strong positive drivers of wheat yield, consistent with previous findings that emphasize the critical role of moisture availability in determining crop performance in the Indo-Gangetic Plains (Lobell *et al*., 2011; Asseng *et al*., 2015). Soil organic carbon (SOC) was also identified as a significant positive factor, reinforcing its importance in improving soil structure, nutrient availability, and water retention (Lal, 2020). These findings support ongoing recommendations for soil health management and carbon-enhancing practices in intensively cultivated systems.

Conversely, maximum temperature exhibited a negative effect on yield, particularly during the grain-filling stage, which aligns with established evidence that terminal heat stress reduces wheat productivity (Asseng *et al*., 2015). The negative influence of soil pH suggests that slightly alkaline conditions may limit nutrient availability, highlighting the need for site-specific soil management strategies.

The inclusion of Bayesian calibration improved model realism by accounting for uncertainty in management practices, particularly sowing date and nitrogen application. Similar approaches have been shown to enhance predictive reliability in crop modeling under uncertain field conditions (Quintero *et al*., 2025). By incorporating posterior distributions into subsequent models, the framework ensures that uncertainty is propagated throughout the analytical pipeline.

Spatial heterogeneity captured through the hierarchical Bayesian model revealed clear district-level differences, with Udham Singh Nagar consistently outperforming other districts. This spatial pattern is consistent with studies demonstrating that yield variability is strongly influenced by localized agro-ecological conditions and management intensity (Park *et al*., 2023). The use of Conditional Autoregressive priors allowed effective borrowing of information across neighboring districts, improving model stability and generalization.

The Structural Equation Modeling component provided valuable insights into causal relationships among soil, weather, and management factors. The identification of both direct and indirect effects highlights the interconnected nature of agricultural systems, supporting earlier work that emphasizes the utility of SEM in disentangling complex agro-ecological interactions (Grace *et al*., 2010). Despite concerns regarding sample size, the use of longitudinal data and latent constructs enabled robust estimation of relationships.

The spatio-temporal state-space model successfully captured dynamic yield trends and identified periods of decline associated with climatic stress. The detected yield dip during 2010–2012 corresponds to known episodes of terminal heat stress, while the recovery phase aligns with improvements in management and soil conditions. Similar temporal modeling approaches have been shown to effectively capture yield variability under changing climate conditions.

The integration of these modeling approaches resulted in a reduction of prediction error by approximately 15–22% compared to individual models. This improvement is consistent with recent studies advocating hybrid and ensemble approaches for agricultural forecasting (Poudel *et al*., 2024; Li *et al*., 2025). The framework not only enhances predictive performance but also provides interpretable outputs that are directly relevant for decision-making.

From a practical perspective, the results suggest that targeted interventions such as optimizing sowing dates (around late November), improving soil organic carbon levels, and managing nitrogen application can significantly enhance productivity. The probabilistic outputs generated by the model can support risk-based decision-making in procurement, crop insurance, and extension services. However, some limitations should be acknowledged. The analysis is based on district-level aggregated data, which may mask field-level variability. Additionally, the use of SEM with limited cross-sectional units requires cautious interpretation, although the longitudinal structure partially mitigates this limitation. Future research could incorporate higher-resolution spatial data, remote sensing time-series, and deep learning models to further enhance predictive capability.

Overall, the findings demonstrate that integrating uncertainty, spatial dependence, causal relationships, and temporal dynamics within a unified framework provides a powerful approach for crop yield modeling in climate-sensitive regions. The proposed methodology offers a scalable and interpretable solution for advancing data-driven agricultural decision support systems.

CONCLUSION

This study presents an integrated modeling framework combining Bayesian calibration, Spatial Hierarchical Bayesian Regression, Structural Equation Modeling, and Spatio-Temporal State-Space Modeling to explain and predict wheat yield in the Tarai–Bhabar region of Uttarakhand. The proposed approach effectively captures uncertainty in management practices, spatial heterogeneity across districts, causal relationships among agronomic variables, and temporal yield dynamics. The integrated model achieved superior predictive performance (Test $R^2 = 0.71$; RMSE = 0.120 t ha^{-1}) compared to individual models, demonstrating the advantage of combining multiple analytical paradigms. The framework provides both accurate yield forecasts and interpretable insights into key drivers such as rainfall, soil organic carbon, and nitrogen management.

These findings highlight the potential of integrated statistical frameworks for improving agricultural decision support systems, particularly in climate-sensitive regions. The approach can be extended to other crops and regions to support data-driven agricultural planning and policy formulation.

REFERENCES

Anjoy, P. and Chandra, H. (2019). Hierarchical Bayes aggregated level spatial model for crop yield estimation. *Journal of the Indian Society of Agricultural Statistics*, 73(2): 143–152.

Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W. and Reynolds, M.P. (2015). Rising temperatures reduce global wheat production. *Nature Climate Change*, 5: 143–147. https://doi.org/10.1038/nclimate2470

Basavaraju, N. M., Reddy, K. V., Kumar, S., Prakash, G. and Rao, P. S. (2025). Optimized Bayesian models for crop yield forecasting. *Engineering, Technology & Applied Science Research*.

Durbin, J. and Koopman, S.J. (2012). Time series analysis by state space methods (2nd ed.). Oxford University Press.

Grace, J.B., Anderson, T.M., Olff, H. and Scheiner, S.M. (2010). On the specification of structural equation models for ecological systems. *Ecological Monographs*, 80(1): 67–87. https://doi.org/10.1890/09-0464.1

Harvey, A.C. (1989). Forecasting, structural time series models and the Kalman filter. Cambridge University Press.

Hernández Hernández, G. C., Gómez Gómez, J. and Jiménez-Cabas, J. (2025). Predictive models based on artificial intelligence to estimate crop yield: A literature review. *Agriculture*.

Jabed, M. A. and 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*.

Lal, R. (2020). Soil organic matter and sustainable agriculture. Agronomy, 10(1): 135. https://doi.org/10.3390/agronomy10010135

Li, D., Kitsios, V., Newth, D. and O'Kane, T.J. (2025). Bayesian ensemble projection of climate change and technological impacts on future crop yields. arXiv preprint arXiv:2507.21559.

Liu, Z., Zhang, Y., Li, X., Wang, H., Chen, L. and Sun, J. (2026). In-season crop yield prediction: State-of-the-art and future directions. *Agricultural Systems*.

Lobell, D. B., Schlenker, W. and Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science*, 333: 616–620. https://doi.org/10.1126/science.1204531

Lionel, B. M., Dupont, F., Martin, G., Laurent, A. and Petit, S. (2025). Comparative analysis of machine learning models for crop yield prediction. *Springer*.

Maimaitijiang, M., Sagan, V., Sidike, P., Daladima, T., Erkbol, H., Fritschi, F. B. and Shakoor, N. (2020). Soybean yield prediction from UAV multisensor data using machine learning and SHAP analysis. *Remote Sensing of Environment*, 237: 111599. https://doi.org/10.1016/j.rse.2019.111599

Nikhil, U. V., Sharma, R., Kumar, P., Singh, A. and Verma, S. (2024). Machine learning-based crop yield prediction using weather and soil data. *Computers*, 13(6): 137.

Park, Y., Li, B. and Li, Y. (2023). Crop yield prediction using Bayesian spatially varying coefficient models with functional predictors. *Journal of the American Statistical Association*, 118(541): 70–83. https://doi.org/10.1080/01621459.2022.2123333

Poudel, S., Bhattarai, N., Panday, D., Giri, S. and Kc, K. (2024). A parsimonious Bayesian crop growth model for water-limited conditions. *Agricultural Systems*, 222: 103685. https://doi.org/10.1016/j.agsy.2024.103685

Quintero, D., Mishra, V., Limaye, A. S., Robertson, R. D. and Gerber, J. S. (2025). Bayesian calibration of management practices for crop models in subsistence systems. *European Journal of Agronomy*, 164: 127524. https://doi.org/10.1016/j.eja.2024.127524

Saha, S., Ghosh, A., Mondal, S., Dutta, S. and Chatterjee, S. (2025). Precision agriculture for improving crop yield predictions using AI. *Frontiers in Agronomy*. https://doi.org/10.3389/fagro.2025

Shawon, S. M., Rahman, M. M., Hasan, M. K., Islam, M. R. and Hossain, M. S. (2025). Crop yield prediction using machine learning: An extensive review. *Environmental and Sustainability Indicators*. https://doi.org/10.1016/j.indic.2025

Vehtari, A., Gelman, A. and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27: 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Wang, Y., Zhang, Q., Yu, F., Zhang, N., Zhang, X., Li, Y., Wang, M. and Zhang, J. (2024). Progress in deep learning-based crop yield prediction. *Agronomy*, 14(10): 2264. https://doi.org/10.3390/agronomy14102264

Received: December 22, 2025
Accepted: March 18, 2026