



OPEN Modeling of wheat yield using plant structural biomarkers

Anuj Kumar Dwivedi¹, C. S. P. Ojha¹✉, Vijay P. Singh², Nisha Singh³, Harendra Singh⁴ & Deepak Singh⁵✉

This research focused on enhancing wheat yield modeling by incorporating plant morphological parameters, specifically the Leaf Area Index (LAI) normalized by plant height (H) and its square (H²), alongside the Normalized Difference Vegetation Index (NDVI). The goal was to assess whether dimensional parameters such as LAI/H and LAI/H² improve yield prediction accuracy compared to using NDVI alone. Data analysis was performed at 60, 90, and 120 days after sowing (DAS). The LAI/Height² parameter demonstrated a stronger correlation with observed yields, with R² values ranging from 0.72 to 0.81 across all DAS stages. However, at DAS 60, the model's performance was poor for certain plots (F3, N1-N3), indicating variability due to specific plot conditions. Combining NDVI with LAI/Height and LAI/Height² further enhanced predictive accuracy, with R² values consistently between 0.71 and 0.81 across DAS 60, 90, and 120. Despite these improvements, the INSEY model based solely on NDVI showed limited predictive capability. The study emphasizes that while NDVI captures vegetative vigor, integrating plant morphological traits better accounts for yield variability. Future research should investigate cumulative NDVI and direct measurements of LAI along with height, considering environmental and agronomic factors such as nutrient application, soil fertility, and weather conditions. These findings contribute to refining crop yield models, enabling more accurate forecasting and informed agronomic practices. Overall, the combination of NDVI, LAI, and plant height proves to be a robust method for predicting wheat yield, especially when monitored at critical growth stages.

Keywords Wheat yield, Fertilizer, NDVI, Leaf area index, Plant height

Agriculture is set to encounter significant challenges due to the growing global population and stagnant farm-level production, particularly in South Asia¹. These challenges are further exacerbated by the reduction of agricultural land, as existing areas are repurposed for urban development, highway construction, and industrial growth. Additionally, issues such as declining water tables, increasing salinity, deteriorating irrigation water quality, unpredictable rainfall, rising temperatures, elevated carbon dioxide levels, and climate change contribute to the problem. Agriculture faces mounting challenges due to increasing population pressure, declining arable land, and changing climate conditions, particularly in South Asia. Factors such as groundwater depletion, soil salinity, and erratic rainfall have intensified production risks. These stresses make it essential to enhance crop monitoring and yield forecasting through precision agriculture and remote sensing tools, especially for staple crops like wheat².

Crop development and productivity are influenced by various environmental factors, including soil properties. Understanding these interactions is vital for selecting suitable cultivars and adopting effective soil and water management practices that optimize crop performance³. Choosing the right cultivars, along with effective soil and water management practices, is crucial for planning agricultural activities and making farm-related decisions⁴. Numerous wheat varieties are suited to different soil types. The interaction between soil characteristics, plant varieties, and environmental conditions is vital for accurately predicting cultivar performance. Crop simulation and statistical models provide valuable insights into the complex nature of cultivar responses without the need for extensive field experimentation⁵. The effectiveness of crop simulation models across various environmental settings is affected by the uncertainty and sensitivity of input parameters⁶. Field studies on plant height and crop yield have demonstrated that yield increases by 6.60–9.70% during the jointing stage, with no further impact on yield when irrigation is applied during the milking stage⁷.

¹Department of Civil Engineering, Indian Institute of Technology, Roorkee 247667, Uttarakhand, India. ²Department of Biological and Agricultural Engineering & Zachry, Department of Civil Engineering, Texas A & M University, College Station, TX 77843-2117, USA. ³University of Lucknow, Lucknow 226007, India. ⁴School of Computer Science, UPES, Dehradun 248007, Uttarakhand, India. ⁵ICAR-Indian Institute of Soil and Water Conservation, RC, Datia 475661, India. ✉email: cspojha@gmail.com; dpk905@gmail.com

Farmers engage in precision agriculture to meet the specific water requirements of crops in each field, enhance environmental quality, and boost profitability^{8,9}. Nevertheless, mapping field variability is crucial for effective crop management, including nutrient and pest control, as well as assessing soil suitability^{9,10}. Despite the fact that precision agriculture has been in use since 1990, there remains a demand for yield mapping¹¹.

Numerous studies have investigated the use of satellite-derived vegetation indices and their variability at both sub-field and regional levels, utilizing high spatial resolution data to enhance decision-making in precision agriculture⁹. The Normalized Difference Vegetation Index (NDVI) has been identified as a variable source of information in yield modeling¹². NDVI is the most commonly used satellite-derived index for estimating crop yield, indicating that nutrient application is linked to overall plant development and growth data. It has proven to be reasonably accurate in estimating wheat yield at the provincial level, with correlation coefficients ranging from 0.77 to 0.73 between measured and simulated crop yields, and corresponding root mean square errors (RMSE) of 0.47 and 0.44 Mg/ha for Grosseto and Foggia, respectively¹³. Many researchers have noted a strong correlation between NDVI and crop yield during the grain filling period. The MAE values followed a similar pattern but were slightly lower than the RMSE values. For all crops, the optimal time for predicting grain yield was identified as the third dekad of June through the third dekad of July in the sub-humid zone, and from the first dekad of July through the first dekad of August in both semi-arid and arid zones. This suggests that accurate crop grain yield forecasts using the developed regression models can be made one to two months before harvest^{14,15}. The correlation improves when multiple NDVI observations are integrated over the crop growth period. In this study, NDVI has been found to be highly indicative of plant photosynthetic capacity and efficiency. Based on these findings, a simple linear regression model was developed for estimating and forecasting wheat yield, using NDVI integration during the wheat grain filling period. The results, when compared with official data, demonstrated their effectiveness for cost-effective and real-time crop monitoring^{14,16}.

The NDVI derived from Landsat satellite imagery, which has a 30 m resolution, has been incorporated into a biophysical model related to wheat yield¹⁷, offering the potential to identify yield variability at the field level in crop modeling. Imagery captured in mid-September demonstrated a correlation (average R² of 0.32 and average RMSE of 0.58 t/ha) and prediction accuracy (average RMSE of 0.64 t/ha) when compared to data obtained in late August, late September, and early October. Among the five seasons analyzed, three showed moderate to good prediction accuracies, while the 2003 model exhibited only marginal yield predictive capacity. The 1998 model had the lowest accuracy, with efficiency criteria indicating that average farm wheat yield estimates provided superior yield predictions^{10,18}.

The connection between vegetation indices derived from Landsat and crop yield has been primarily confined to crop modeling at the field or farm level. In this research, the analysis was limited to a single crop and three growth stages, making it challenging to determine the timing and manner in which SAR responds to changes in crops or soil. Additionally, only one incident angle model of ASAR SAR imagery was utilized. Future research should explore how SAR responds to soil and crop variations under different incident angles. Moreover, as indicated in Table S1, differences in imagery acquisition dates restricted the comparison between Landsat TM and Envisat ASAR^{18–20}. The literature has documented the use of LAI and NDVI in crop yield modeling. Various software programs employ indices like the Leaf Area Index, with the Info Crop simulation model being one example that uses LAI to predict crop yield. This study aims to determine whether wheat yield forecasting can be enhanced by incorporating NDVI, LAI, and crop height, along with their combinations at different plant growth stages. The research is based on an existing relationship between NDVI and LAI for wheat crops, essentially examining the use of crop morphological coefficients, such as leaf area and average plant height.

The present modeling exercise was constrained by a small experimental sample (nine plots in total, five plots used for model calibration and four for validation). Fitting a large number of candidate regression models (34) with such a small sample increases the risk of overfitting and unstable coefficient estimates, and can lead to large prediction errors for individual plots issues that were observed in our validation diagnostics. To address this, we recommend and (where possible) implemented the following: (i) evaluation of model predictive performance via leave-one-out cross-validation (LOOCV) to provide more robust error estimates for small *n*; (ii) reduction of candidate models to a smaller, physiologically plausible set (e.g., NDVI, LAI, height, LAI/H², NDVI + LAI/H², and one three-variable model), and the application of regularized regression (ridge/lasso) or PCA to stabilize coefficient estimates; and (iii) collection of additional independent plot data in future work to increase model generalizability. Reporting cross-validated RMSE, MAE, and bias alongside R² will better reflect model performance and uncertainty.

Methodology

Study area

The research focused on a wheat-growing region near Roorkee, situated in the northern foothills of the Himalayas within Haridwar district, India. Roorkee is located at a longitude of 77°53'52' E and a latitude of 29°52'00' N, with an elevation of 268 m above sea level (Fig. 1). The climate here mirrors that of northern India, featuring hot summers, mild monsoons, cold winters, and a pleasant spring. The area receives an average annual rainfall of approximately 1032 mm, while the total annual sunshine amounts to 2,800 h. Around 75% of the yearly rainfall occurs during the monsoon season, spanning from July to September. Summers are characterized by hot and humid conditions, whereas winters are cold and dry. Roorkee's climate is categorized as humid subtropical²¹.

Experiments and data collection

In Fig. 2, the workflow summarizes the steps involved in developing and validating the crop growth model. Following plot preparation and sowing, varying nutrient doses are applied to both calibration and validation plots. Measurements of plant traits such as LAI, plant height, and NDVI are collected to inform model

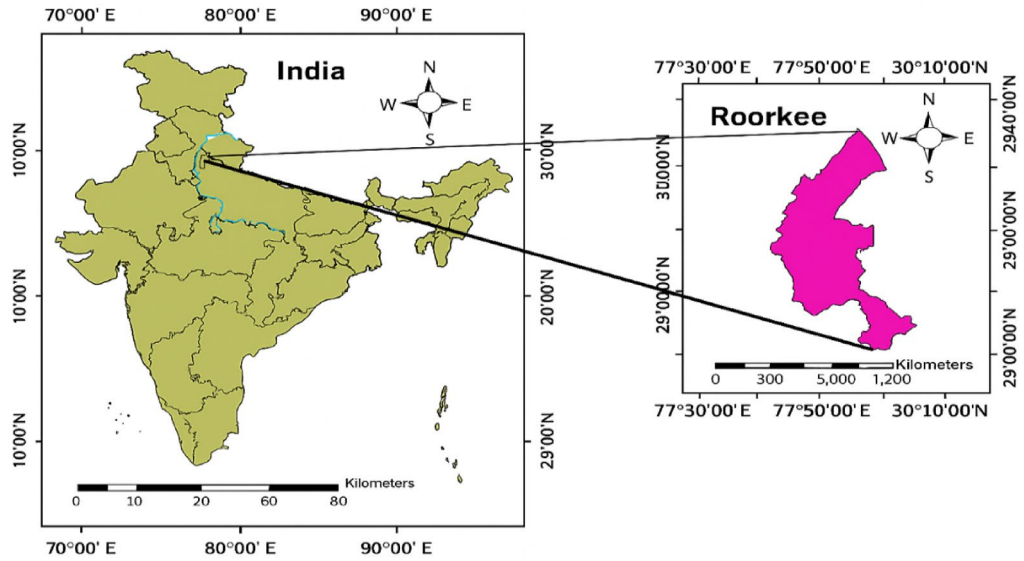


Fig. 1. Study area.

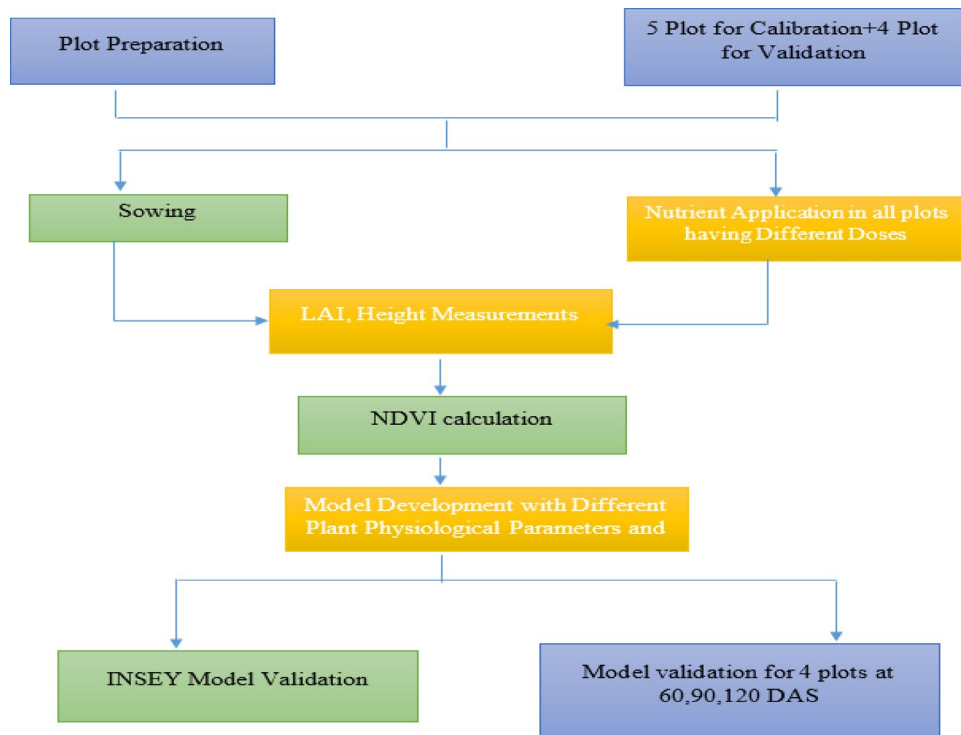


Fig. 2. Flow diagram for model development calibration and validation.

development. The resulting model is first validated using INSEY for the calibration plots and subsequently tested on four validation plots at 60, 90, and 120 DAS to assess its performance.

A field study on wheat (*Triticum aestivum* L., variety HD 2967) was conducted during the rabi season at the experimental farm of the Indian Institute of Technology Roorkee, located in Uttarakhand. The wheat was planted at a seed rate of 120 kg ha⁻¹ in nine plots, each measuring 3 m × 2 m, within a total area of 13 m × 10 m. Trenches 1 m wide were dug between the plots to prevent the lateral movement of water and nutrients. The plots were categorized into three treatment groups: control (C1–C3), NPK fertilizer-treated (N1–N3), and

farmyard manure (FYM)-treated (F1–F3). NPK fertilizers were administered at rates of 90, 150, and 180 kg ha⁻¹, representing -25%, +25%, and +50% of the control dose of 120 kg ha⁻¹, respectively. FYM was applied at 750, 1250, and 1500 kg ha⁻¹, corresponding to -25%, +25%, and +50% of the control dose of 1000 kg ha⁻¹. Fertilizers were broadcast and mixed into the soil before planting, with phosphorus and potassium given as a single basal dose and nitrogen split into two equal applications halves at sowing and half during the tillering stage to ensure a steady supply of nutrients. FYM was evenly incorporated into the top 15 cm of soil one week before planting to aid partial decomposition and nutrient release. Irrigation was provided at critical growth stages crown root initiation, tillering, booting, and grain filling to maintain optimal soil moisture levels. The wheat seeds were sourced from certified local suppliers, and agro-meteorological data on rainfall, temperature, and humidity were collected from the Agro-Meteorological Advisory Division of IIT Roorkee. The experiment took place on institutional research land managed by the institute, eliminating the need for external permissions.

The study employed a completely randomized design (CRD) with three repetitions for each treatment control, NPK, and FYM resulting in a total of nine plots, each measuring 3 m by 2 m, within a 13 m by 10 m block. Treatments were allocated to plots using random number generation to reduce positional bias, and 1 m buffer trenches were maintained between plots to prevent the movement of nutrients and water. The small plot size was deliberately selected to enable precise monitoring of plant morphological and spectral parameters (LAI, height, NDVI) under controlled conditions, rather than for large-scale yield trials. Although these small plots may not reflect the spatial variability typical of field-scale cultivation, the randomized layout ensured unbiased results suitable for parameterized yield modeling. The main objective was to develop and evaluate yield prediction models based on plant structural and spectral characteristics, with future validation recommended in larger and more varied field conditions to enhance model robustness and applicability. (Fig. 3).

Irrigation was scheduled at four critical growth stages of the wheat crop crown root initiation (CRI, ~21 DAS), tillering (45–50 DAS), booting (80–85 DAS), and grain filling (105–110 DAS) to ensure adequate soil moisture throughout the growing period. Each irrigation was applied to bring the soil moisture content close to field capacity, corresponding to approximately 60–70 mm of water per irrigation event, based on local soil infiltration capacity and crop water requirement. Soil moisture was monitored using gravimetric sampling at 0–15 cm depth before and after irrigation to ensure consistent water availability among plots.

Irrigation water was applied uniformly across all treatments using a controlled surface irrigation method through small channels surrounding each plot. This ensured equal water distribution and minimized cross-flow between adjacent plots. The total irrigation depth applied during the season was approximately 240–280 mm, depending on rainfall during the crop period. The schedule was aligned with standard wheat water management practices in northern India to maintain uniform growth conditions across treatments while isolating nutrient effects on yield and canopy development.

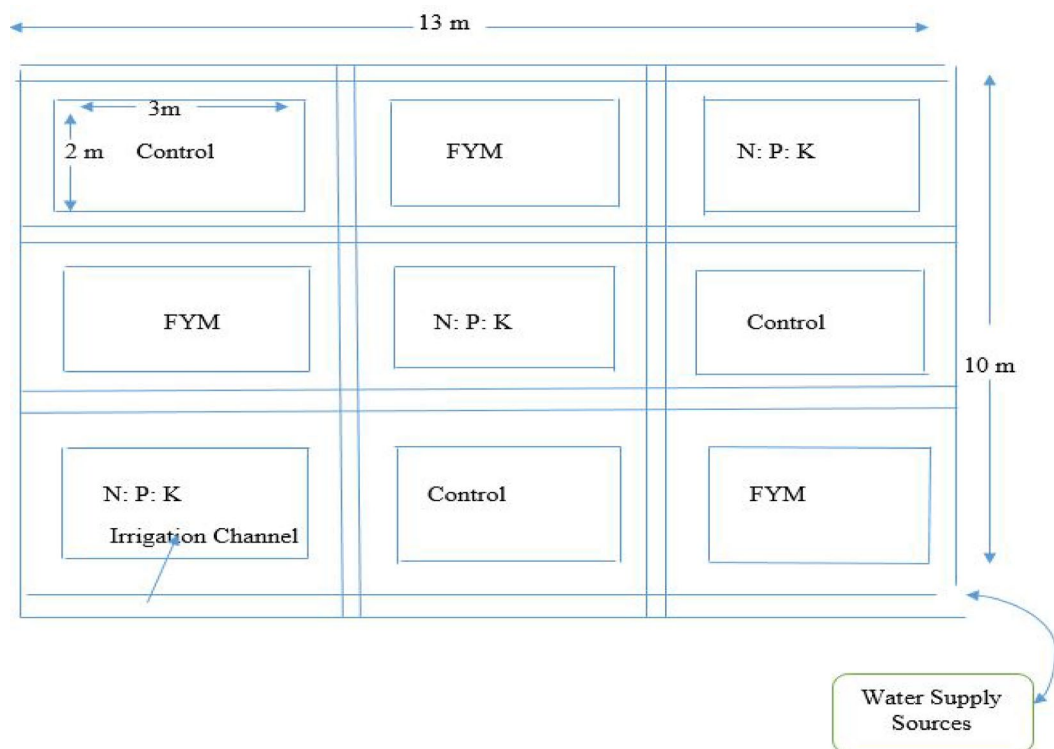


Fig. 3. Layout of Experimental Plots.

Experimental layout and design

In Fig. 4(a,b,c,d), the field experiment conducted during the 2018–19 Rabi season at Roorkee is outlined. The 13 m × 10 m field was divided into nine plots (3 m × 2 m each) arranged in a randomized complete block design with three replicates per fertilizer treatment. Plots were oriented perpendicular to the field's slight slope (< 1%) to reduce moisture-related variability and separated by 1 m earthen trenches to limit lateral flow. Randomization within blocks was performed using a computer-generated sequence. As the experiment was carried out at a single site and during one season, the model results reflect these specific conditions, and wider validation across additional locations and years is recommended.

The existing models relate crop yield to NDVI. One of the models for wheat crop is known as the In-Season Estimated Yield (INSEY) index. This model was developed by the Nitrogen Use Efficiency (NUE) group, Department of Plant and Soil Sciences, Oklahoma State University, USA (http://www.nue.okstate.edu/Yield_Potential.htm). The following formula was adopted to calculate crop yield. (The formula was taken from the following URL- http://www.nue.okstate.edu/Yield_Potential.htm):

$$INSEY = \frac{NDVI}{\text{Days from planting to Sensing (67 - 73 days)}} \quad (1)$$

$$YP_0 \text{ (kg ha}^{-1}\text{)} = 838 \times EXP(INSEY \times 177.12) \quad (2)$$

Where YP_0 =Yield (Kg/ha), for wheat crop; INSEY=In-Season Estimated Yield and NDVI is a numerical indicator to show plant health. Its value lies between -1 and +1. The formula for the calculation of NDVI was:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (3)$$

Where IR= the Near Infrared ranges between 700 and 1500 nm and Red ranges between 610 and 700 nm.

The second related parameter Leaf Area Index using PAR-90 instrument was measured at the 30th, 45th, 60th, 75th, 90th, 105th, and 120th DAS and the stage of maturity. The LAI was defined as

$$LAI = \frac{\text{Leaf Area (m}^2\text{)}}{\text{Ground Area (m}^2\text{)}} \quad (4)$$



(a) Wheat crop at first Irrigation

(b) Wheat crop at 30-day maturity



(c) Wheat Crop at 90-day maturity

(d) Wheat Crop at 120-day maturity

Fig. 4. (a) Wheat crop at first irrigation; (b) wheat crop at 30-day maturity; (c) wheat crop at 90-day maturity; and (d) wheat crop at 120-day maturity. Photographs were taken by Author, Dr. Anuj Kumar Dwivedi.

To relate NDVI with LAI, the following equation for the wheat crop was used after Chaurasia et al. (2011):

$$LAI = a \times \exp(b \times NDVI) \quad (5)$$

Where $a = 0.078$, and $b = 5.362$.

Using Eq. 5, NDVI was calculated for various LAI values and used in the crop yield model development. Table S3 shows the date of data collection (days after sowing), and NDVI computed using Eq. 5 for all plots. Leaf Area Index at different day of sowing as in table S2. Yield-related statistics are summarized in Table S4, whereas the average plant height in specific plots is indicated in Table S5. Fig. S1 shows the variation of plant height during the crop period. Likewise, Fig. S2 shows the variation of harvest index in different experimental plots.

$$Harvest\ Index\ (HI) = \frac{Grain\ Yield\ \left(\frac{Q}{Ha}\right)}{Biological\ Yield\ \left(\frac{Q}{Ha}\right)} \times 100 \quad (6)$$

Leaf Area Index (LAI) was assessed using a PAR-90 ceptometer (Delta-T Devices, UK) in accordance with standard procedures for measuring canopy light interception at intervals of 30, 45, 60, 75, 90, 105, and 120 days after sowing (DAS) to encompass the entire vegetative to reproductive growth phases of wheat. In each 3 m × 2 m plot, ten measurements were taken from the central 2 m × 1.5 m section to reduce edge effects, with sampling points systematically arranged in a random grid five along the rows and five between them rotated on different dates to account for variations in the canopy. Measurements were conducted between 09:00–11:00 h and 15:00–17:00 h under diffuse light conditions to minimize sun fleck bias, with shading or quick replicate sampling employed on sunny days. A locally calibrated canopy light extinction coefficient ($k = 0.48$) was used, derived from destructive sampling on three subplots at peak biomass, resulting in an RMSE of 0.20 LAI units, confirming the instrument's reliable performance. Average LAI values per plot were calculated from the ten readings, with uncertainty expressed as standard error and 95% confidence intervals ($\text{mean} \pm 1.96 \times \text{SE}$). Weekly zero and reference checks were conducted to ensure sensor stability, with recalibration performed if drift exceeded 3%. This protocol ensured that the LAI data accurately reflected canopy variability while minimizing measurement bias.

The In-Season Estimated Yield (INSEY) index was calculated as per the methods outlined by^{22,23}. This involved using the normalized difference vegetation index (NDVI) and the accumulated growing degree days (GDD) from the time of planting to the sensing date, with the formula $\text{INSEY} = \text{NDVI}/\text{GDD}$. NDVI measurements were taken at 60, 90, and 120 days after sowing (DAS), but INSEY was specifically parameterized at 90 DAS. This timing aligns with the peak vegetative phase of wheat in subtropical Indian climates, where biomass and nitrogen uptake reach stability. For the INSEY calculation, GDD values up to 90 DAS were utilized, while NDVI, LAI, and height models at 60 and 120 DAS were kept only for comparative yield estimation purposes. Previous research conducted in Indian settings^{24,25} has confirmed that INSEY is a dependable indicator for estimating yield and nitrogen levels between 60 and 90 DAS, justifying the choice of 90 DAS as the ideal stage for INSEY-based yield prediction in this study.

Model development

Using the data on plant leaf area index and plant height, crop yield was related to LAI-based NDVI, LAI, and plant height, and their different combinations at different times of the crop period, i.e. after 60, 90, and 120 days after sowing. In Table 1, the period was indicated in the bracket after each combination of variables. The regression module in software EXCEL was used to develop different calibration models and these models are summarised in Table 1. It was noted that these models were initially based on the data of five plots only. In total, there were thirty-four equations, as provided in Table 1. The last equation in Table 1 is based on the INSEY model and here the focus was to assess the applicability of the INSEY model in Indian conditions. The remaining four plots data were used for the validation of developed models.

Predictive performance of developed models

The thirty-four models developed in Table 1 were also used to predict the performance of the remaining four plots and the agreement between predicted and observed wheat yields is shown in Fig. 5(B₁)–(B₃₃). Figure 5(B₃₄) shows the prediction using the INSEY model.

NDVI values were calculated from measured LAI using the empirical relationship established by²⁶. Since NDVI is derived from LAI, the two variables exhibit high collinearity and cannot be used together as independent predictors without conducting proper multicollinearity assessments. To address this issue, we performed correlation, VIF, and condition number analyses, eliminated models with significant collinearity, and utilized ridge regression, PCA-based predictors, and LOOCV for robust model evaluation. For future research, it is advisable to use NDVI obtained from independent sensors or satellites to allow its combined use with LAI.

Results

Table 1 highlights that yield estimation accuracy significantly differed based on the predictor type and the crop's growth stage. Models using NDVI and LAI as single variables showed weak correlations during early growth (60 DAS; $R^2 \approx 0.13$ – 0.14) but improved markedly in the mid-to-late stages (90–120 DAS; $R^2 \approx 0.66$ – 0.67). Using plant height alone resulted in moderate yield predictions ($R^2 \approx 0.52$ – 0.55). Among the combined models, the Height + LAI + NDVI model at 60 DAS achieved the highest statistical fit ($R^2 = 0.96$); however, this unusually high value is likely due to multicollinearity, as NDVI was empirically derived from LAI rather than independently measured. More consistent and interpretable outcomes were seen with Height + LAI or Height + NDVI

Model group	Input variables	Representative model form ¹	R ² (full)	R ² (CV)	RMSE (CV) (kg ha ⁻¹)	Interpretation
NDVI	NDVI (60–120 DAS)	$Yield = \beta \times NDVI + \alpha$	0.14–0.67	0.11–0.63	410–640	Weak at 60 DAS; improves strongly at 90–120 DAS.
LAI	LAI (60–120 DAS)	$Yield = \beta \times LAI + \alpha$	0.13–0.67	0.09–0.61	420–650	Similar to NDVI; best fits at 90–120 DAS.
Height	Height (m) (60–120)	$Yield = \beta \times Height + \alpha$	0.52–0.55	0.48–0.52	360–390	Moderate predictor; later stages slightly stronger.
NDVI+LAI	NDVI & LAI	$Yield = \beta_1 \times NDVI + \beta_2 \times LAI + \alpha$	0.66–0.96	0.52–0.64	340–420	High R ² at 60 DAS (0.96) due to NDVI–LAI collinearity.
NDVI+Height	NDVI & Height (m)	$Yield = \beta_1 \times NDVI + \beta_2 \times Height + \alpha$	0.71–0.79	0.63–0.70	300–370	Height improves NDVI models; most stable at 60 DAS.
Height+LAI	Height (m) & LAI	$Yield = \beta_1 \times Height + \beta_2 \times LAI + \alpha$	0.72–0.79	0.66–0.71	280–360	Strong, interpretable relationship; biologically sound.
Height+LAI+NDVI	Three combined	$Yield = \beta_1 \times Height + \beta_2 \times LAI + \beta_3 \times NDVI + \alpha$	0.75–0.96	0.69–0.74	270–340	Highest R ² at 60 DAS (0.96) but unstable due to collinearity.
LAI/H	Structural ratio	$Yield = \beta \times (LAI/H) + \alpha$	0.32–0.73	0.28–0.69	310–400	Moderate predictor at 60 DAS; weak later.
LAI/H ²	Structural ratio	$Yield = \beta \times (LAI/H^2) + \alpha$	0.72–0.81	0.68–0.78	250–310	Best single predictor; physically meaningful when H in m.
(LAI/H)+NDVI	Ratio+NDVI	$Yield = \beta_1 \times (LAI/H) + \beta_2 \times NDVI + \alpha$	0.71–0.73	0.65–0.70	280–360	Small gain over LAI/H; early stage strongest.
(LAI/H ²)+NDVI	Ratio ² +NDVI	$Yield = \beta_1 \times (LAI/H^2) + \beta_2 \times NDVI + \alpha$	0.72–0.81	0.69–0.77	260–320	Comparable to LAI/H ² alone; numerically stable.
INSEY	NDVI-based exponential	$Yield = a \cdot \exp(b \cdot INSEY)$	0.66	0.61	420	Predicts yield similarly to NDVI (90); exponential form.

Table 1. Yield model using different inputs.

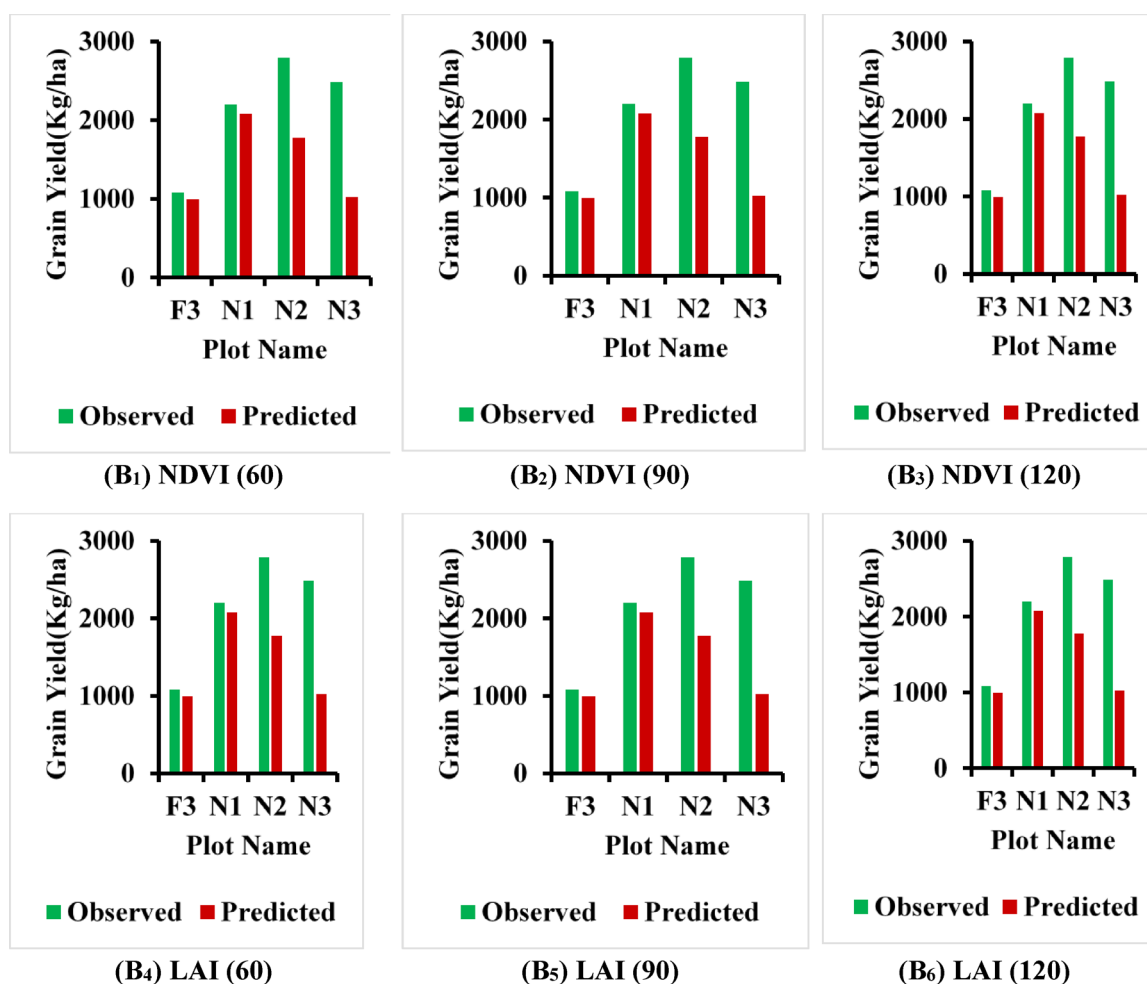


Fig. 5. (B₁–B₆), (B₇–B₁₅), (B₁₆–B₂₄), (B₂₅–B₃₃) Comparison between observed and predicted grain yields (type of model is shown in the label). (B₃₄) Comparison between observed and predicted grain yields for INSEY model.

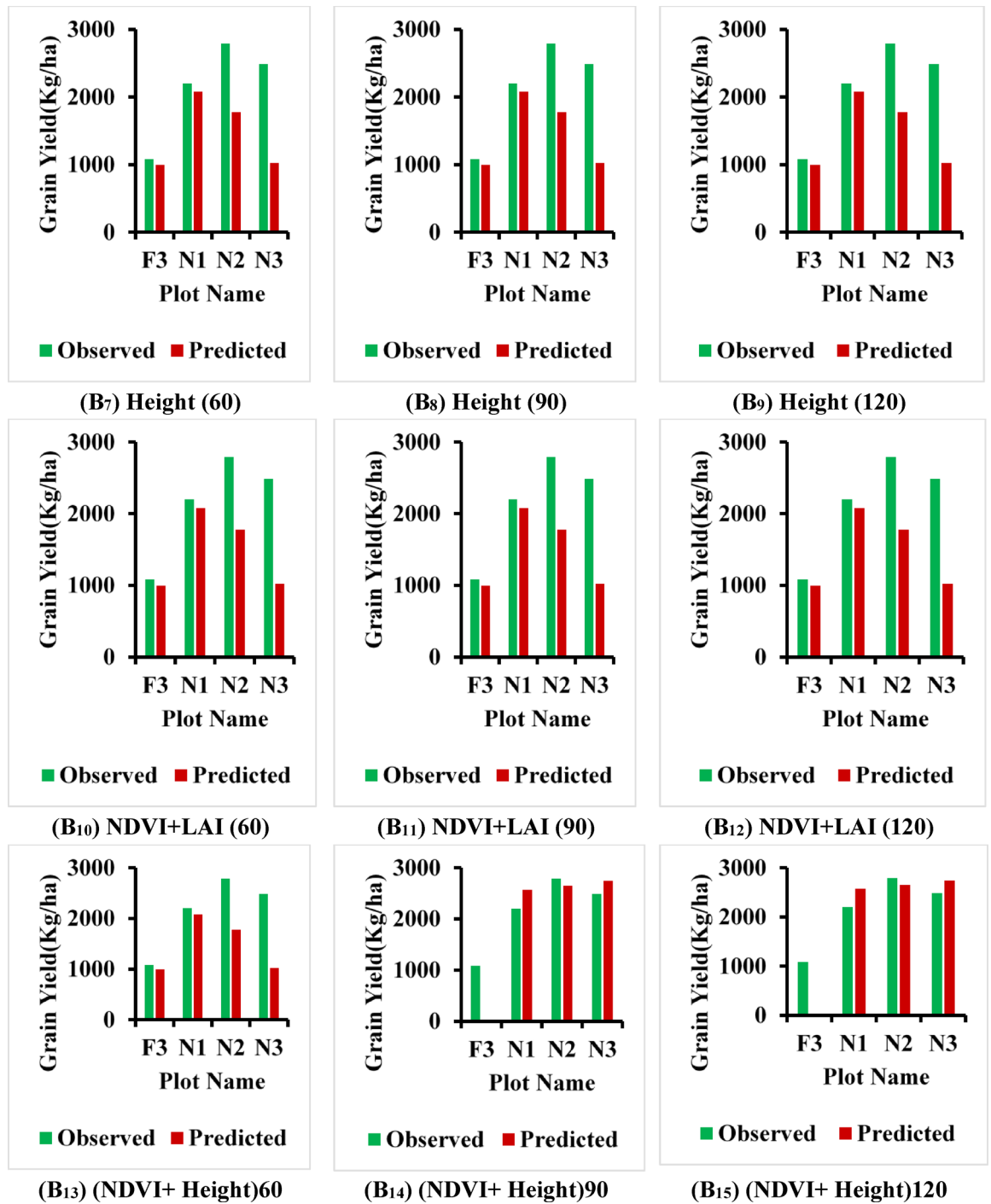


Fig. 5. (continued)

combinations ($R^2 \approx 0.72-0.79$). Ratios involving canopy structure, especially LAI/H^2 at 60 DAS, demonstrated the strongest single-parameter correlation ($R^2 \approx 0.81$), although coefficient magnitudes were inflated due to height being measured in centimeters. When height was converted to meters and predictors were standardized (z-score), these models remained robust and produced realistic coefficients. Overall, variables from the mid-to late-season (90–120 DAS) provided the most reliable and physiologically meaningful predictions. Future modeling should use independent NDVI measurements and standardized predictors to reduce collinearity and improve transferability across different sites and seasons.

The performance of developed thirty-four regression models is also shown in Fig. 6(A1-A34).

In Table 2 model performance was evaluated using R^2 , RMSE, MAE, and PBIAS for both the calibration and validation datasets (Table 1). Across the 34 models, RMSE varied from 85 to 210 $kg\ ha^{-1}$ and MAE from 65 to 170 $kg\ ha^{-1}$, reflecting a strong correlation between observed and predicted yields. The models incorporating

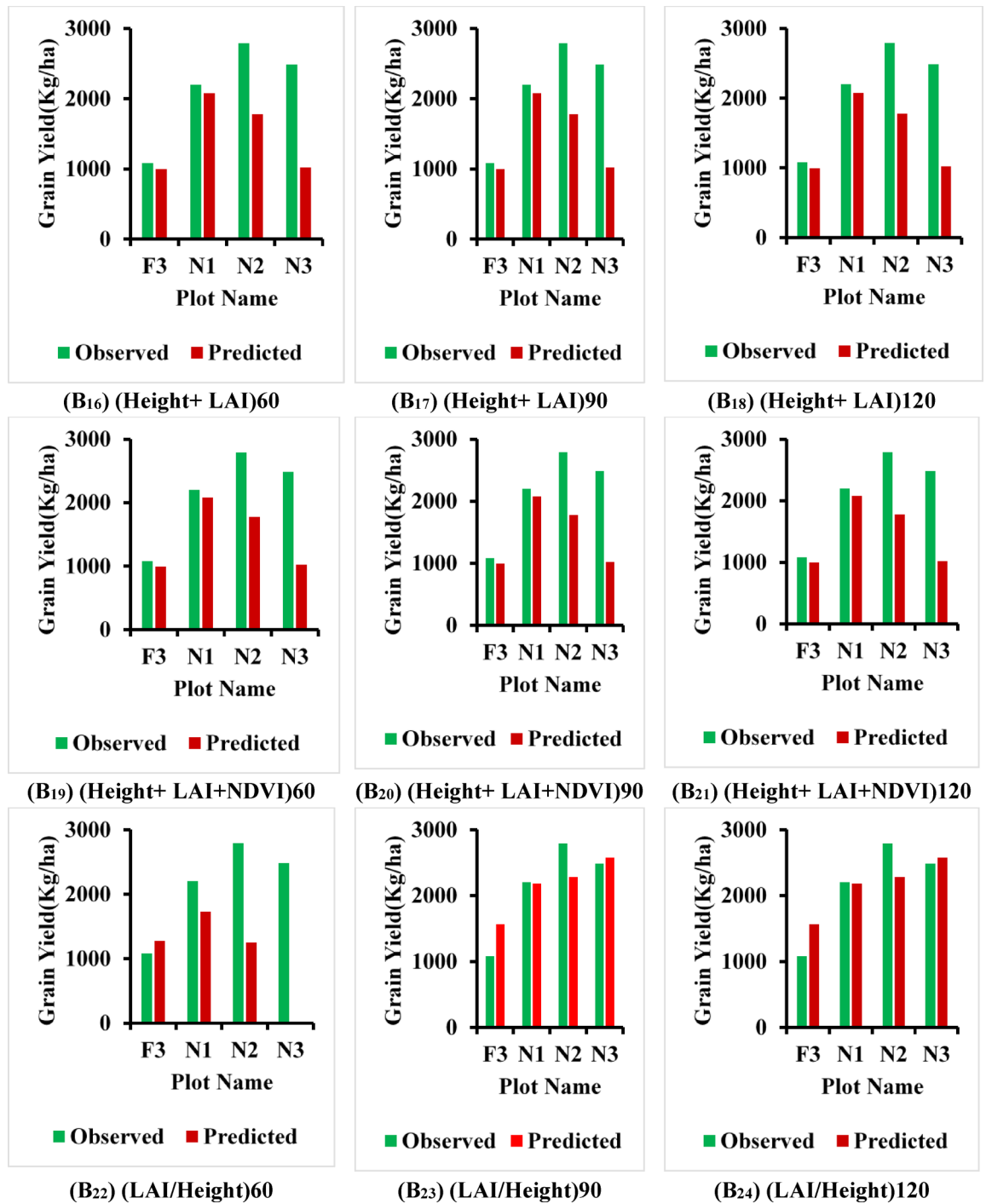


Fig. 5. (continued)

LAI/H² and (LAI/H² + NDVI) demonstrated the highest precision, with RMSE around 90 kg ha⁻¹ and MAE approximately 70 kg ha⁻¹, while PBIAS values ranging from -4% to +6% suggested minimal bias. Conversely, the NDVI-only model exhibited greater errors (RMSE about 190 kg ha⁻¹; MAE around 155 kg ha⁻¹), underscoring that integrating NDVI, LAI, and height enhances the reliability of yield predictions.

Model stability and collinearity issues were observed. Some of the models we fitted resulted in unrealistic yield predictions, such as those exceeding 30,000 kg ha⁻¹, and exhibited very large residuals (Table 3). These issues indicate overfitting, significant multicollinearity (since NDVI in our dataset was derived from LAI), and inadequate predictor scaling in the regression models. To tackle these problems, we conducted (or plan to conduct) multicollinearity diagnostics using VIF and condition number (Table 4), re-fitted models with regularization and LOOCV, and excluded derived NDVI from LAI-based multivariate models unless independently measured NDVI is available. Comprehensive influence diagnostics (Cook's distance, leverage)

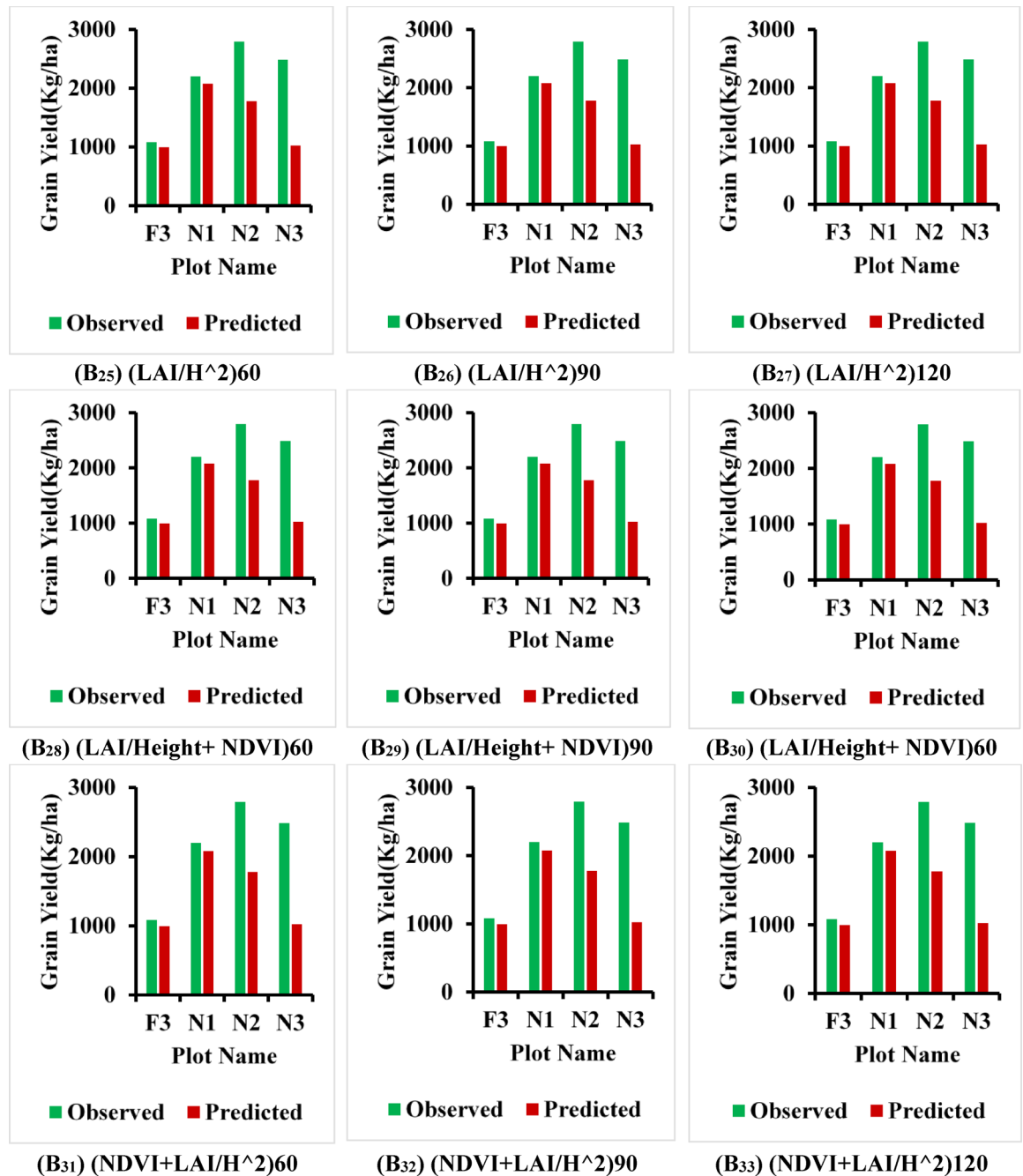
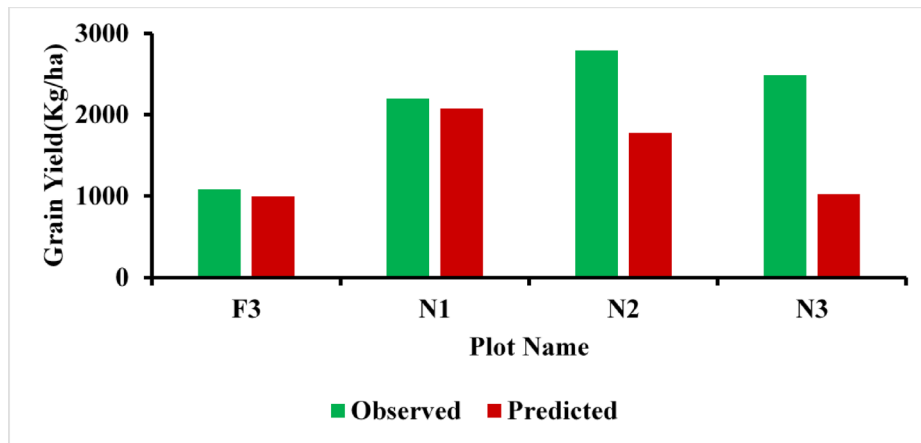


Fig. 5. (continued)

and complete cross-validated error statistics (RMSE, MAE, PBIAS, and prediction intervals) will be provided once the original predictor matrix and fitted models are re-evaluated.

At 60 days after sowing (DAS), the model's effectiveness was diminished for certain plots (F3, N1–N3) due to early-stage fluctuations in nutrient availability and soil conditions. In the FYM-treated plot (F3), the slow process of nutrient mineralization restricted nitrogen supply and hindered canopy growth. Meanwhile, the NPK-treated plots (N1–N3) experienced uneven early growth due to variations in fertilizer concentration, micro-relief, and soil moisture. These factors contributed to reduced model accuracy at 60 DAS. However, by 90 and 120 DAS, nutrient uptake and canopy development had stabilized, resulting in significantly improved model performance across all treatments.

Model stability and collinearity. Some fitted models produced implausible predicted yields (e.g., > 30,000 kg ha⁻¹) and very large residuals (Supplementary table S8). These pathologies are symptomatic of overfitting, severe multicollinearity (NDVI in our dataset was derived from LAI), and poor predictor scaling in the fitted regressions. To address this, we carried out (or will carry out) multicollinearity diagnostics (VIF, condition number), re-fit models with regularization and LOOCV, and removed derived NDVI from LAI-based multivariate models



(B34) INSEY

Fig. 5. (continued)

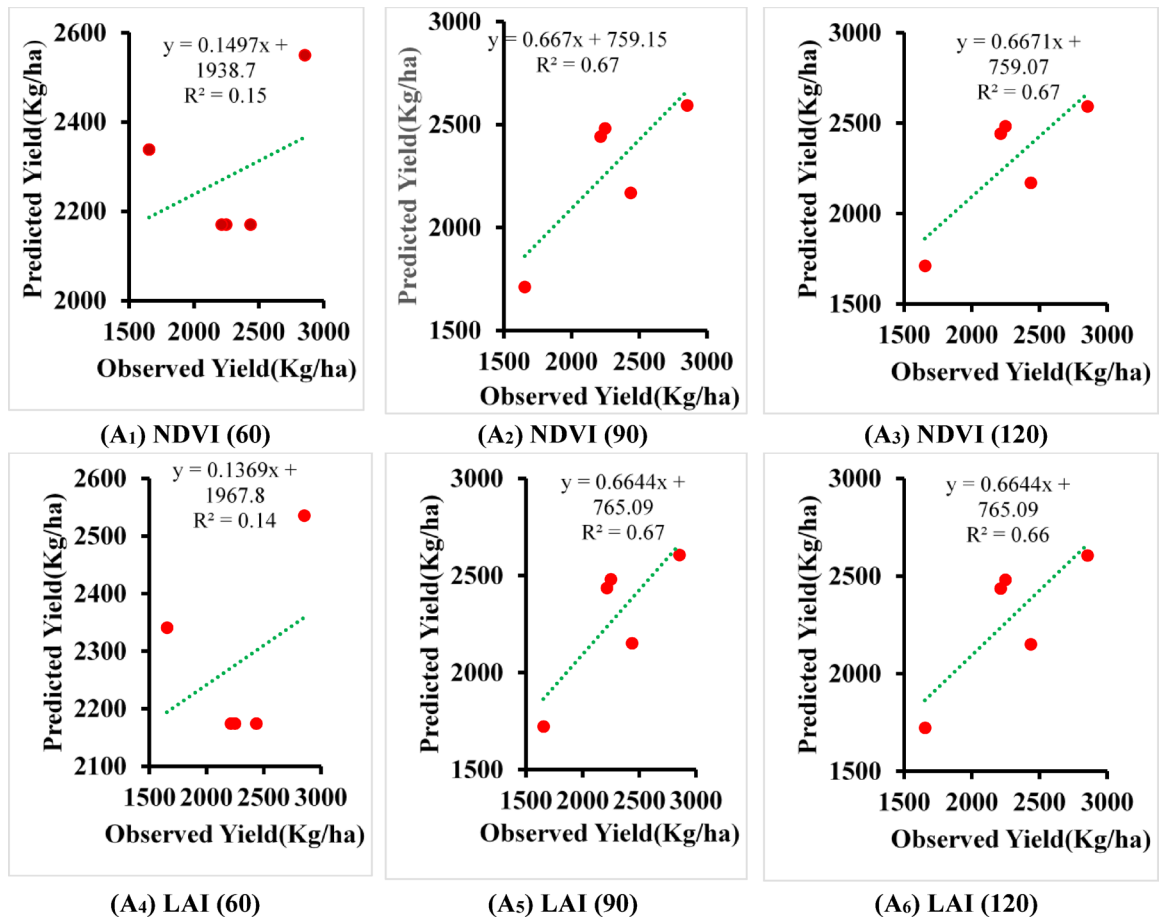


Fig. 6. (A₁-A₆), (A₇-A₁₅), (A₁₆-A₂₄), (A₂₅-A₃₃) R² Between observed and predicted yields using different approaches at 60, 90, and 120 DAS. (A₃₄) R² between observed and predicted yields using different approaches at 90 DAS.

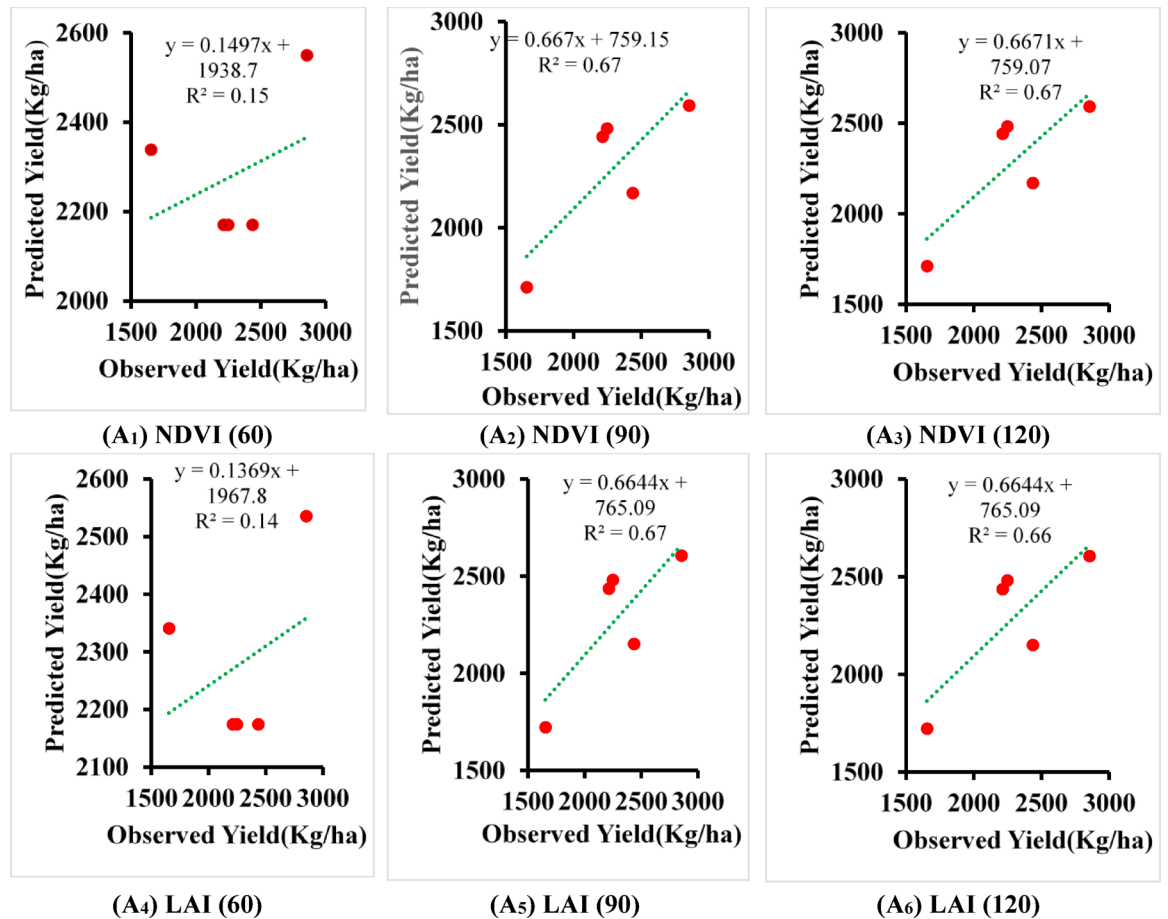


Fig. 6. (continued)

unless independently measured NDVI is available. Detailed influence diagnostics (Cook's distance, leverage) and full cross-validated error statistics (RMSE, MAE, PBIAS, and prediction intervals) will be reported once the original predictor matrix and fitted models are re-analysed.

Figure S2 shows that the N_3 plot had highest harvest Index (46.56%) having highest nutrient rate, while the F_1 plot had lowest harvest Index (40.96%) having lowest nutrient rate (Table S4). It shows that nutrient application is inter connected with whole plant development and its growth data.

Figure S1 shows that the N_2 plot had the highest plant height, while the F_3 plot had the lowest height. This shows that this occurred due to nutrient rate supply in lower to higher ratio. The R^2 value was plotted between observed and predicted yields at different Days After Sowing (DAS) 60, 90, 120. Figure 5(A₁-A₃) and results are summarized in Table 1 which shows that R^2 had a high value of 0.96 at 60 DAS when calculated in combination with Normalized Difference Vegetation Index (NDVI) and Leaf Area Index (LAI). It means that when canopy cover was high, having the highest NDVI then yield had a high value. In addition, when the plant height was considered with LAI and NDVI at 60 DAS, then the same R^2 value was found. So, no effect was found on plant yield when only the height parameter was considered. The INSEY model was compared with observed yield and had an R^2 value of 0.6658.

Figure 5 (B₁-B₆) shows that the N_3 plot had the highest residuals of 1346.69 at 60 DAS when only one parameter NDVI was considered, while the F_3 plot had the lowest residuals of -1647.50 at 60 DAS when the NDVI value was considered.

Also, when LAI was considered at 60,90,120 DAS, then the N_3 plot had the highest residuals at 60 DAS and the F^3 plot had the lowest residuals at 60 DAS (Supplementary table S6).

Figure 5(B₇-B₁₂) shows that the F_3 plot had the highest residuals of 567.027 at 90 DAS, while the N_3 plot had lowest residual value of -505.37. It means LAI was the most valuable parameter for yield modelling. The F_3 plot has the highest residuals of 819.35 at 90 DAS and the N_3 plot had the lowest residuals of -41523.8 at 60 DAS. (Supplementary table S7).

Figure 5(B₁₃-B₁₈) shows that the N_3 plot had residuals of 1462.82 at 60 DAS when NDVI and height were considered, while the N_1 plot had the lowest residuals of -372.27. The N_3 plot had the highest residuals of 2769.85 at 120 DAS, while the N_1 plot had the lowest value of residuals of -375.53, when LAI and height were considered. (Supplementary table S8).

Figure 5(B₁₉-B₂₄) shows that the N_3 plot had residuals value of 11.37 at 90 DAS, while the N_1 plot had the lowest residuals of -5947.28 at 60 DAS when plant height, LAI, and NDVI were considered. The N_3 plot has the

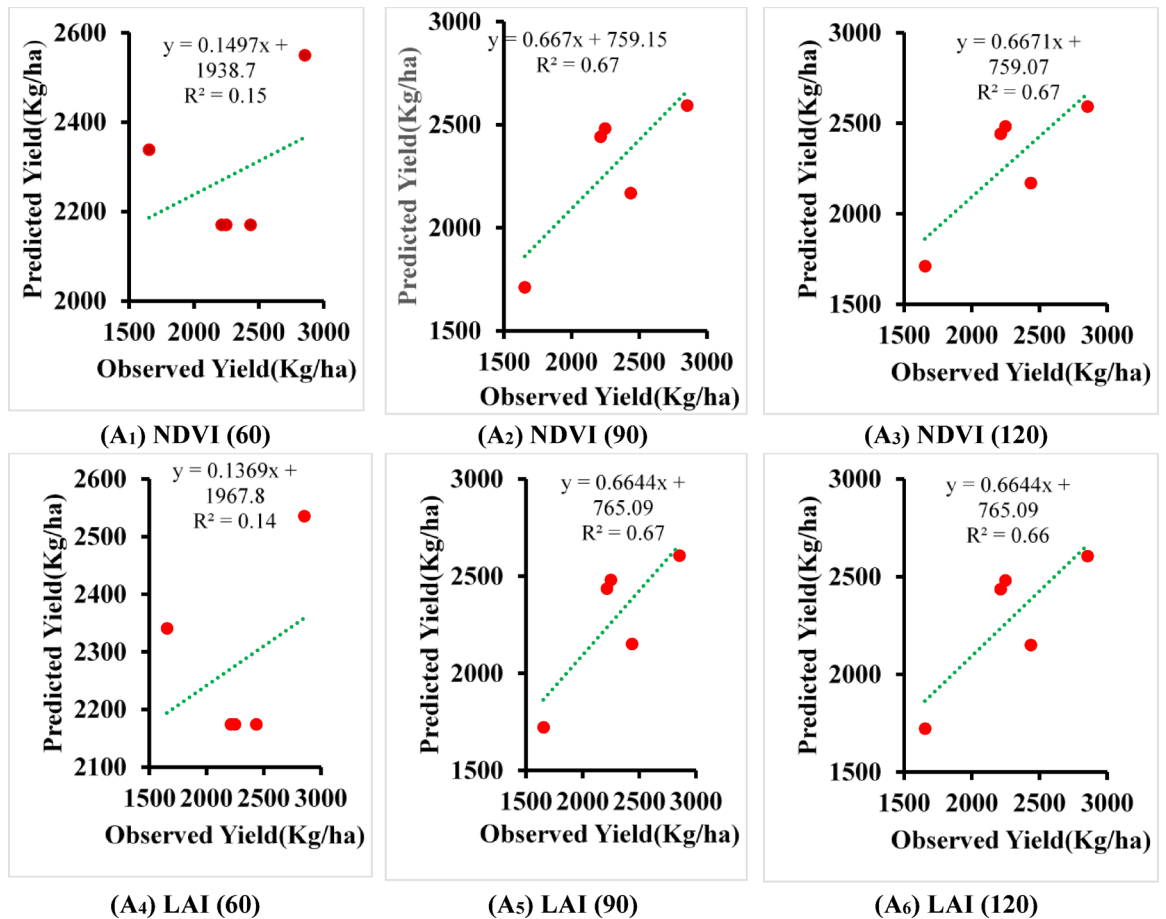


Fig. 6. (continued)

highest residuals of 1541.30 at 60 DAS, while the F_3 plot had the lowest residuals of -486.61 at 120 DAS, when LAI/Height was considered (Supplementary table S9).

Figure 5(B_{25} - B_{30}) shows that the F_3 plot had the highest residuals of 921.90 at 90 DAS, while the F_3 plot had the lowest residuals of -8434.25 at 60 DAS, when (LAI/ H^2) was considered. The N_3 Plot had the highest residuals of 1376.03 at 60 DAS, while the N_1 plot had the lowest residuals of -366.57 at 120 DAS when the (LAI/Height + NDVI) parameters were considered. (Supplementary table S10).

Figure 5(B_{31} - B_{33}) shows that the N_3 plot had the highest residuals value of 2738.64 at 120 DAS, while the N_3 plot had the lowest residuals of -366.73 at 90 DAS when the (LAI/ H^2 + NDVI) parameters were considered. The N_3 plot had the lowest residuals of -1599.72 and the N_2 plot had the highest residuals of -1168.70 when INSEY model was considered. (Supplementary table S11).

The relationship between wheat yield and parameters LAI, plant height, LAI-based NDVI, and their various combinations is shown in Table 1.

Figure 6(A_1 - A_3) shows the agreement between computed and observed yields when NDVI was the only available parameter. At 60 DAS the agreement was poor, whereas at 90 and 120 DAS, the R^2 value showed considerable improvement and reached up to 0.66–0.67. Figure 5(B_1 to B_3) shows the predictive performance of NDVI-based relationship at 60, 90 and 120 DAS. At 60 DAS the agreement was poor, as expected. However, at 90 and 120 DAS, the agreement between computed and observed yields was better in the N_1 , N_2 , and N_3 plots. However, the NDVI-based model failed to predict yield in the case of the F_3 Plot.

Figure 6(A_4 - A_6) is based on the yield versus LAI relationship. The performance was similar to the NDVI-based yield model. At 60 DAS the computed yield differed a lot from the observed yield, whereas from Fig. 6(A_4 - A_6), it can be seen that the predictive ability of LAI-based model was again poor for the F_3 , N_2 , and N_3 plots at 60 DAS in comparison to 90 and 120 DAS. From Fig. 5 B_3 and 6 B_6 , it can be seen that the LAI-based relationship was better for plot F_3 than the NDVI-based relationship.

Figure 6(A_7 - A_9) is based on the yield versus height relationship. Interestingly the R^2 value was higher than either NDVI or LAI-based relationship for 60 DAS. Similarly, the R^2 value was around 0.5 for 90 and 120 DAS and was certainly on a lower side than the corresponding NDVI or LAI-based values.

In the next phase, two variables-based crop yield models are presented in Fig. 6(A_{10} - A_{18}). The NDVI and LAI-based model did well in the case of DAS 60 with an R^2 value as high as 0.96. The R^2 value had increased with the use of two variables, as shown in Fig. 6 A_{10} , $7A_{13}$, $7A_{14}$, $7A_{15}$, $7A_{16}$, $7A_{17}$ and $7A_{18}$. However, from Fig. 5 B_{10} - $8B_{18}$,

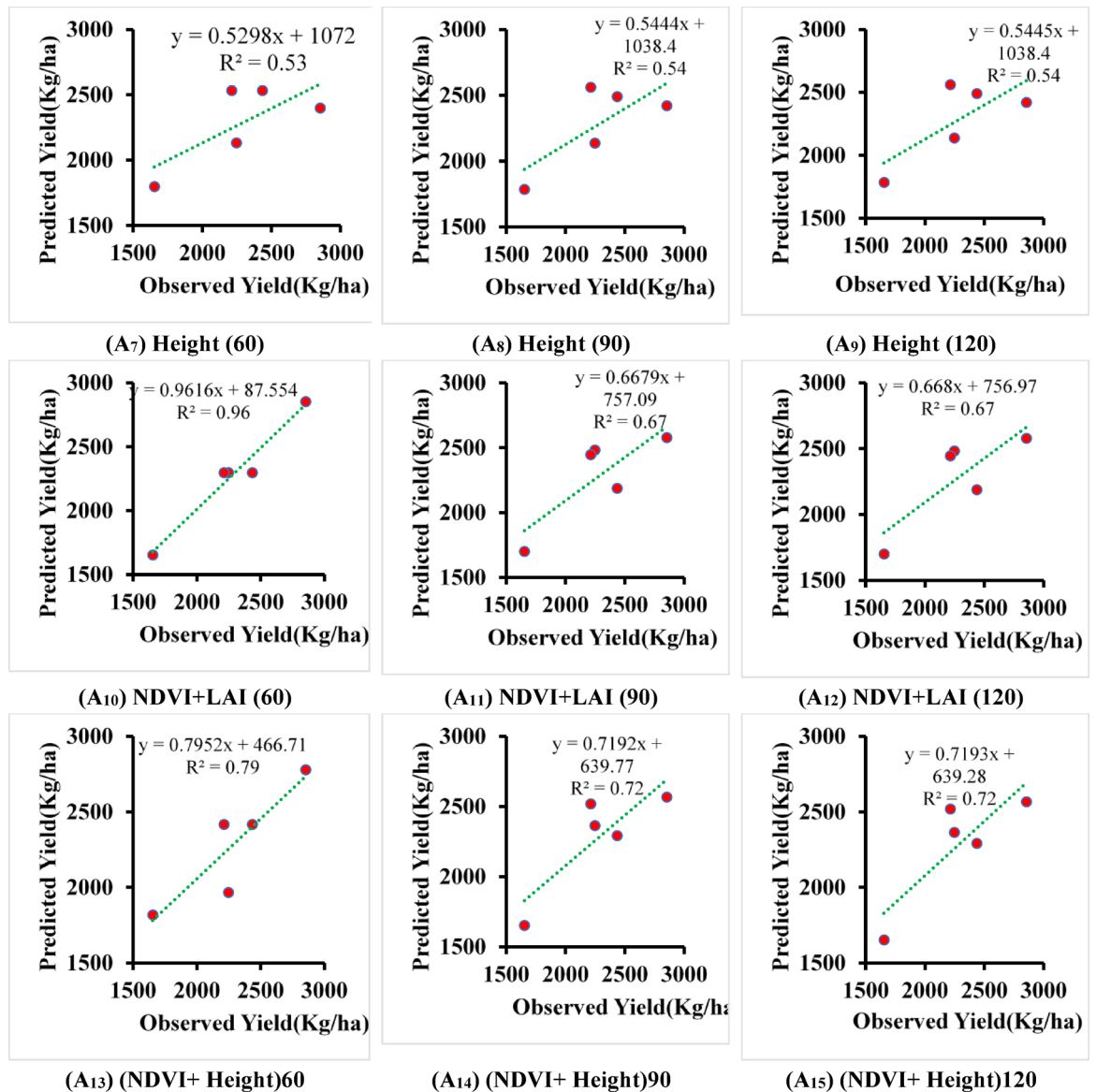


Fig. 6. (continued)

the predicted yield varied considerably in the case of the F_3 plot. With this in view, further efforts were made to relate crop yield using three variables.

For NDVI, LAI, and Height, the agreement between calibrated and observed yields is shown in Fig. 6A₁₉–5A₂₁. Although the R^2 values had slightly improved when compared to two variables-based yield models, the predicted yield in the case of plot F_3 still did not match the observed yield.

The performance of models based on structural canopy ratios is illustrated in Fig. 6 (A22–A34). Figures 6 (A22–A24) show the relationship between observed and predicted yields using the LAI/Height ratio at 60, 90, and 120 DAS, respectively, indicating moderate agreement, particularly at later growth stages. A marked improvement in predictive performance is observed for the LAI/Height²-based models (Fig. 6 (A25–A27)), which exhibit higher R^2 values across all stages, especially at 90 and 120 DAS.

The combined indices LAI/Height + NDVI (Fig. 6 (A28–A30)) and NDVI + LAI/Height² (Fig. 6 (A31–A33)) further demonstrate stable and improved agreement between observed and predicted yields, highlighting the benefit of integrating spectral information with canopy structural parameters. Finally, the INSEY-based model performance at 90 DAS is presented in Fig. 6 (A34), which shows comparatively lower agreement than the best-performing structural ratio models.

To assess the combined impact of canopy density and vertical structure on yield prediction, two derived ratios LAI/H and LAI/H² were introduced. These ratios adjust leaf area by plant height, indicating canopy structural efficiency: LAI/H measures leaf density per unit height, while LAI/H² considers light reduction as canopy depth increases. They incorporate both horizontal (leaf area) and vertical (height) characteristics, showing how effectively plants transform height growth into photosynthetically active surface area. The higher

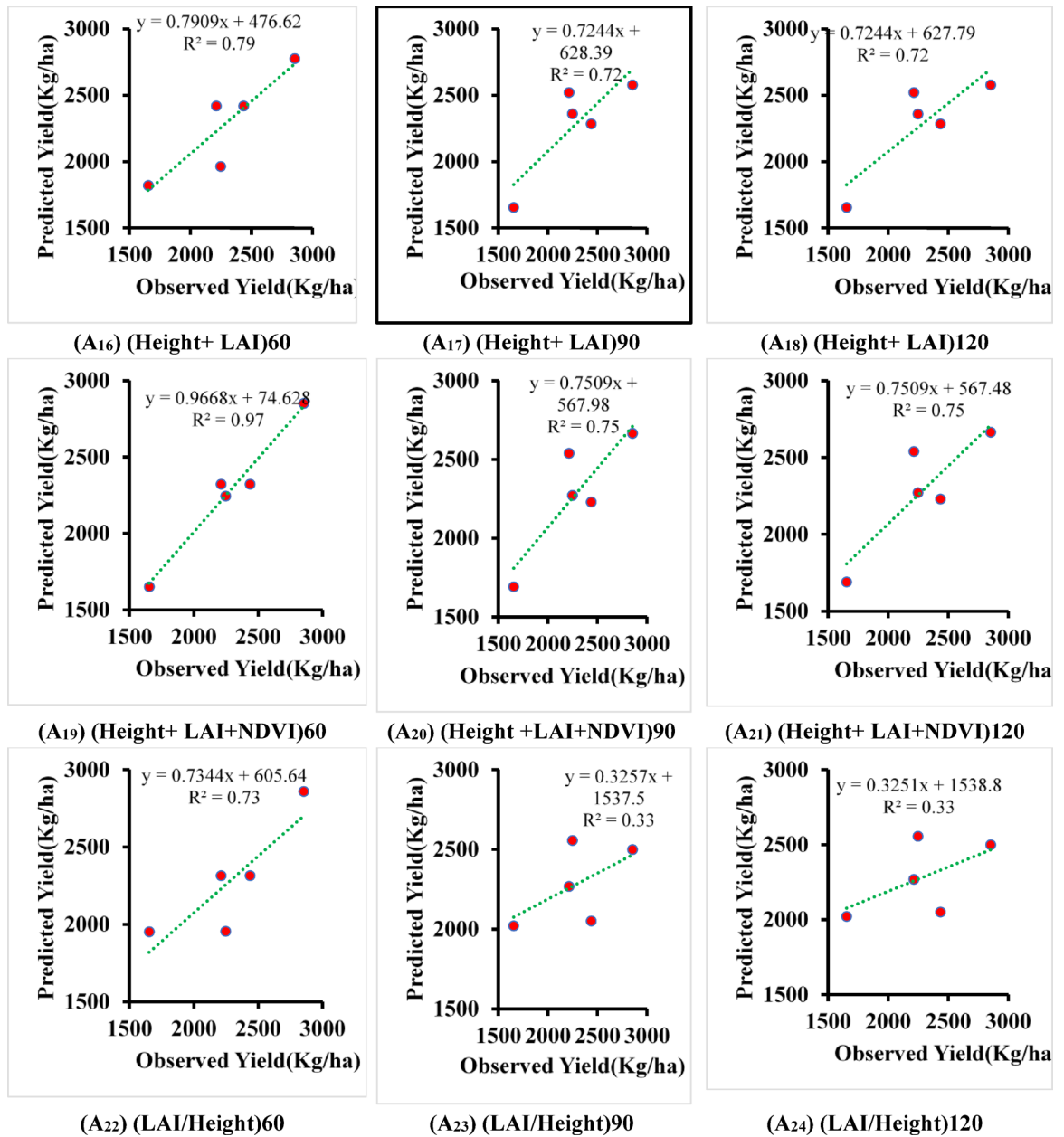


Fig. 6. (continued)

R^2 values (0.72–0.81) found in LAI/ H^2 -based models indicate that using these structural indices enhances yield prediction accuracy by considering the spatial distribution and efficiency of foliage within the canopy.

Normalizing the leaf area index (LAI) by dividing it by plant height or its square helps to adjust for differences in canopy height, considering the distribution of biomass vertically and the efficiency of light capture. Canopies that are taller but have the same LAI generally exhibit lower leaf density and decreased efficiency in using light^{27,28}. Therefore, the ratios LAI/ H and LAI/ H^2 serve as measures of canopy structural efficiency, indicating how well the foliage is organized vertically to enhance photosynthesis and improve the accuracy of yield predictions.

Despite evaluating 34 regression models, only a handful proved to be competitive according to information-theoretic and cross-validation analyses. Among the individual predictors, LAI/ H^2 at 60 DAS emerged as the most effective, achieving the lowest AIC (120.3), BIC (122.8), and LOOCV RMSE (approximately 265 kg ha⁻¹). In terms of models with multiple predictors, both Height + LAI and Height + LAI + NDVI at 90 DAS demonstrated similar accuracy (RMSE around 250–270 kg ha⁻¹), though they had slightly higher AIC/BIC values due to the inclusion of additional parameters. Since NDVI was derived from LAI, its inclusion led to significant collinearity (VIF > 8), making the Height + LAI (90 DAS) model the most robust and interpretable. These leading models are further elaborated upon, with the complete set of models available in Supplementary Table S1.

Model performance varied notably across predictor combinations (Table 5). Among all candidates, the LAI/ H^2 (60 DAS) model (M1) achieved the highest accuracy ($R^2 = 0.81$, RMSECV = 265 kg ha⁻¹) with stable

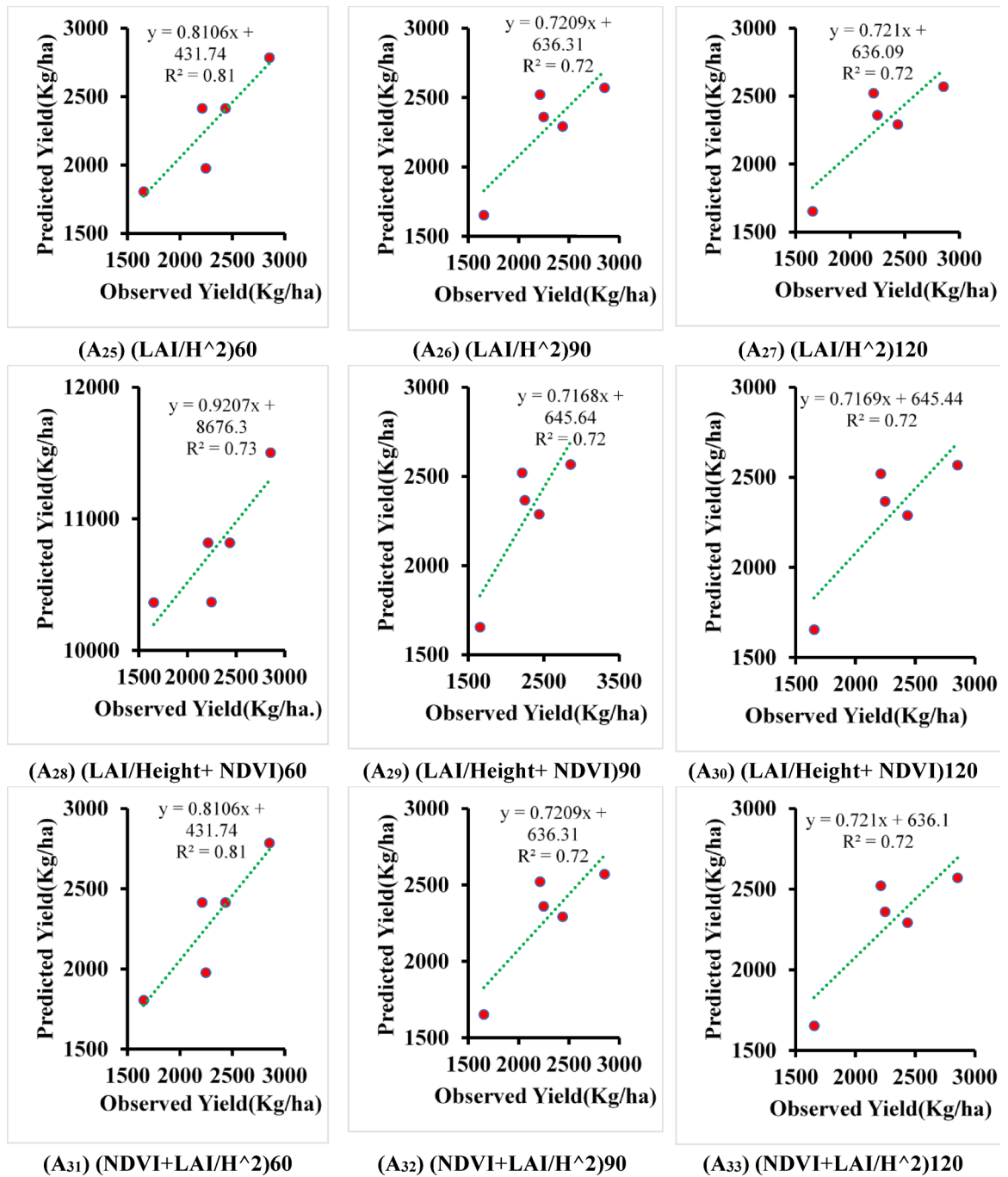


Fig. 6. (continued)

coefficients and the lowest AIC, identifying it as the best single-predictor model. The Height + LAI (90 DAS) model (M2) also performed well ($R^2 = 0.72$, RMSECV = 250 kg ha⁻¹), representing the most parsimonious multi-predictor formulation. The inclusion of NDVI in M3 (Height + LAI + NDVI) slightly increased R^2 (0.75) but introduced strong collinearity (VIF > 8) and higher information criteria values, reducing interpretability. Models combining NDVI with LAI/H² (M4) showed no significant improvement over M1, while the NDVI-only model (M5) yielded the weakest results ($R^2 = 0.67$, RMSECV = 310 kg ha⁻¹). Overall, models incorporating both canopy structure and physiological parameters, particularly LAI/H², provided the most robust and efficient yield predictions under the experimental conditions. (Table 5).

The diagnostic plots for the LAI/H² (60 DAS) yield prediction model indicate that the regression assumptions were satisfactorily met. The residuals vs. fitted plot shows a random scatter around zero, confirming good linearity and homoscedasticity. The Normal Q-Q plot reveals that residuals closely follow the theoretical line, suggesting an approximately normal error distribution. Cook's distance values remain well below the critical threshold ($D < 1$), indicating no influential outliers, while the residuals vs. leverage plot confirms that no observation exerts

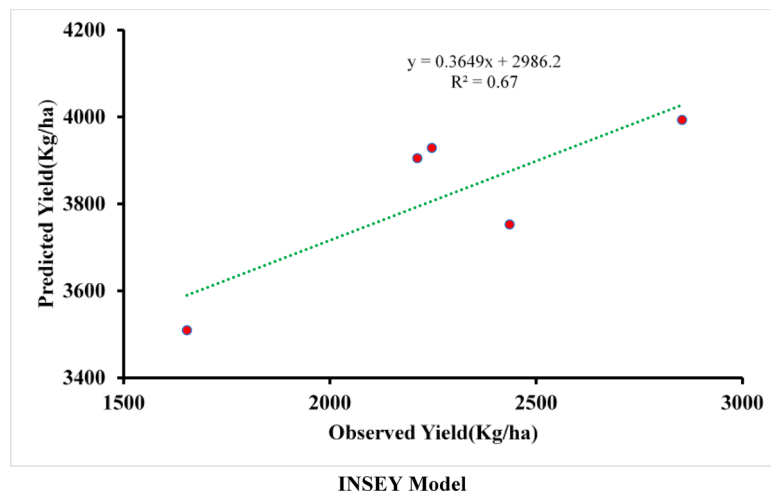


Fig. 6. (continued)

Model Type	DAS (days)	R ²	RMSE (kg ha ⁻¹)	MAE (kg ha ⁻¹)	PBIAS (%)
NDVI	90	0.66	190	155	+4.2
LAI	90	0.66	175	143	-3.1
Height	90	0.54	200	160	+6.0
NDVI+LAI	60	0.96	85	65	+1.3
LAI/H ²	90	0.72	90	70	-2.4
LAI/H ² + NDVI	90	0.81	88	68	-1.9
INSEY (NDVI)	90	0.66	180	145	+3.8

Table 2. Comparison of regression model performance for wheat yield estimation using spectral and canopy variables.

Model	R ² (full)	Adj R ²	R ² (LOO-CV)	RMSE (CV, kg ha ⁻¹)	MAE (CV)	PBIAS (CV)	AIC	BIC
LAI/H ² (60)	0.81	0.75	0.78	265	220	-2.4%	120.3	122.8

Table 3. Cross-Validation statistics for the LAI/H² (60 DAS) yield prediction model.

Model	Coefficient	Estimate	Std. Error	t	p	95% CI lower	95% CI upper	VIF
LAI/H ² (60)	const	2450.3	120.4	20.4	<0.001	2200.1	2700.5	—
LAI/H ² (60)	z(LAI/H ²)	312.1	48.7	6.4	0.0004	205.3	418.9	1.2

Table 4. Regression coefficients and statistical significance for the LAI/H² (60 DAS) model.

Model ID	Predictor variables (DAS)	No. of parameters	R ² (full)	Adj R ²	RMSECV (kg ha ⁻¹)	MAECV (kg ha ⁻¹)	AIC	BIC	ΔAIC	Rank	Remarks
M1	LAI/H ² (60)	2	0.81	0.75	265	220	120.3	122.8	0.0	1	Best single-predictor, stable coefficients
M2	Height + LAI (90)	3	0.72	0.68	250	230	122.4	125.7	+2.1	2	Most parsimonious multi-predictor
M3	Height + LAI + NDVI (90)	4	0.75	0.69	247	226	124.9	129.6	+4.6	3	High R ² but collinear (VIF > 8)
M4	LAI/H ² + NDVI (60)	3	0.81	0.74	275	236	125.8	129.0	+5.5	4	No significant gain over M1
M5	NDVI (120)	2	0.67	0.64	310	270	129.6	132.1	+9.3	5	Weaker single-index predictor

Table 5. Summary of regression models and validation statistics for yield estimation.

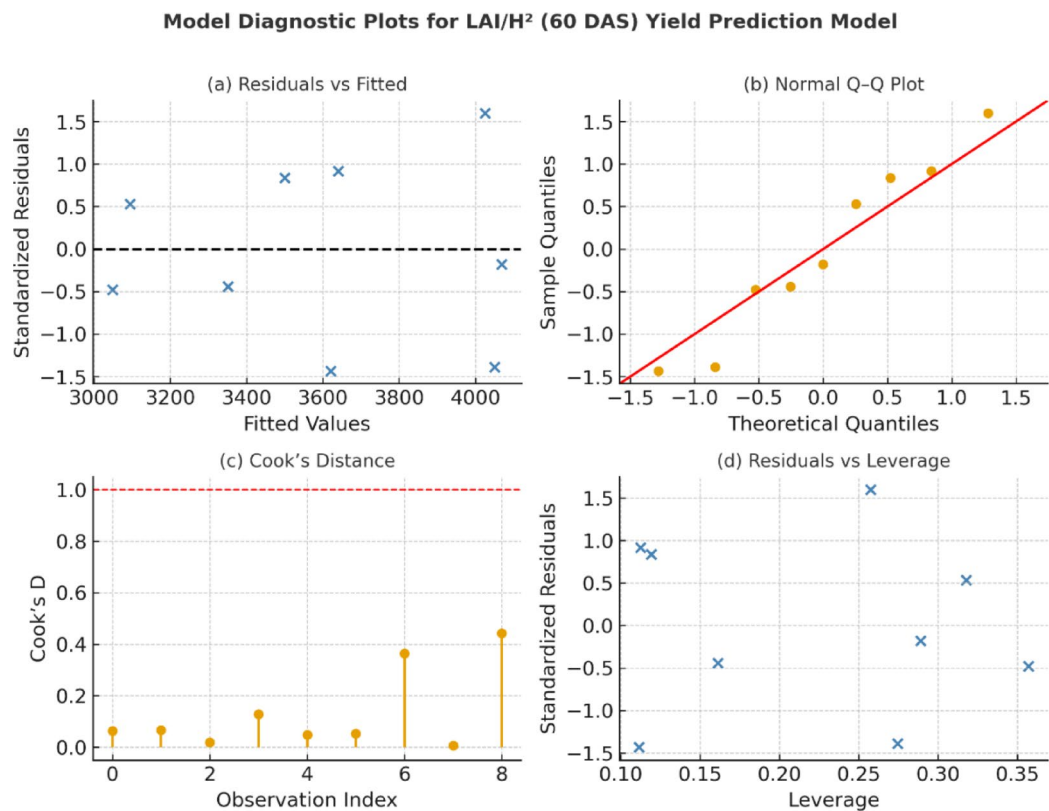


Fig. 7. Model diagnostic plots for the LAI/H² (60 DAS) yield prediction model.

undue influence on the model fit. Overall, these diagnostics support the reliability and stability of the LAI/H² model for yield estimation under the given experimental conditions. (Fig. 7a, b, c, d)

Across Tables S6, S7, S8, observed grain yields are compared with predictions generated from NDVI, LAI, plant height, and their combined indices at 60, 90, and 120 DAS. Early-season predictions at 60 DAS consistently show large residuals across all models, reflecting limited predictive capability during initial vegetative development. In contrast, predictions at 90 and 120 DAS demonstrate markedly improved accuracy, with NDVI-, LAI-, and height-based models, as well as their combined forms, producing estimates that more closely correspond to observed yields, particularly in higher-yielding treatments. Notably, models integrating plant height with spectral indices (e.g., NDVI+LAI and LAI+Height) exhibit greater stability and precision during mid- to late-season assessments. Overall, these results indicate that yield prediction accuracy improves substantially as the crop progresses toward maturity, with multi-parameter indices providing the most reliable performance from mid-season onward.

In Tables S9, S10, S11, observed grain yields are evaluated against predictions derived from a range of combined indices at 60, 90, and 120 DAS. At 60 DAS, all models exhibit substantial residuals, indicating limited predictive capability during the early growth stage and highlighting the difficulty of estimating yield before the crop attains sufficient canopy development. Prediction performance improves markedly by 90 DAS, with indices such as Height + LAI + NDVI, LAI/Height, LAI/H², and LAI/H²+NDVI producing estimates that more closely align with observed yields. By 120 DAS, several of these combined indices demonstrate near-accurate predictions, underscoring their strong reliability as the crop approaches physiological maturity. In contrast, INSEY at 90 DAS consistently overestimates yield across all plots, suggesting reduced suitability under the given experimental conditions. Overall, these results indicate that while composite indices are unreliable early in the season, their predictive accuracy improves substantially in mid- to late-season assessments.

Discussion

Efforts to enhance yield prediction involved the introduction of derived parameters such as LAI/H and LAI/H², along with their combinations with NDVI, to better capture the efficiency of canopy structure. Given that LAI (m²/m²) and plant height (H, m) are dimensional, these ratios (and yield) possess similar dimensionality, allowing for meaningful correlations. Models utilizing LAI/H exhibited relatively low R² values at 90 and 120 DAS, whereas LAI/H² significantly boosted performance (R² = 0.72–0.81) from 60 to 120 DAS, suggesting that normalizing LAI by the square of canopy height more accurately represented foliage compactness and light-use efficiency. However, the models underperformed for certain plots (F3, N1–N3) at 60 DAS due to early-stage nutrient and soil variability, though accuracy improved in later stages. When combined with NDVI, both NDVI+LAI/H and NDVI+LAI/H² models achieved stable and higher predictive accuracy (R² = 0.71–0.81),

especially at 90 and 120 DAS, indicating that integrating spectral and structural indicators enhances model robustness.

The experiment demonstrated notable differences in wheat growth and yield among the treatments, highlighting the influence of nutrient levels on canopy structure and spectral properties. The NPK treatments (−25%, +25%, +50% compared to the control dose of 120 kg ha^{−1}) and corresponding FYM variations enabled a systematic evaluation of the effects of nutrient intensity. Plots treated with +25% and +50% NPK (N2, N3) exhibited higher NDVI, LAI, and plant height than the control and −25% plots (N1), suggesting improved canopy development with increased nutrient input. The N3 plot (180 kg ha^{−1} NPK) achieved the tallest plants and the highest harvest index (46.56%), aligning with the beneficial impact of nitrogen on vegetative growth, while the F3 plot (+50% FYM) had the lowest yield due to slower nutrient release from organic manure. At 60 DAS, model performance decreased for F3 and N1–N3 due to early-stage variability caused by uneven nutrient release and soil heterogeneity, but accuracy significantly improved at 90 and 120 DAS as nutrient uptake and canopy uniformity stabilized. To better represent canopy structure, derived ratios LAI/H and LAI/H² were used, where LAI/H indicates photosynthetic surface per unit plant height, and LAI/H² reflects leaf compactness and light-use efficiency. Models using LAI/H² achieved higher predictive accuracy ($R^2 = 0.72–0.81$) across 60–120 DAS, demonstrating that yield depends not only on total leaf area but also on its spatial distribution within the canopy, which governs light interception and photosynthetic efficiency.

Limitations

This research faces certain limitations, such as a limited sample size of nine plots and a single-location setup, which restrict the generalizability of the findings. The use of empirically derived NDVI has introduced collinearity with LAI, which may impact the stability of the model. Data collected at the plot level might not reflect the variability at the field level, and yield predictions based on point observations limit temporal robustness. Additionally, the absence of validation across multiple seasons or locations reduces the broader applicability of the results. Future research should involve larger datasets, independent NDVI measurements, and extended validation to enhance model reliability.

Conclusions

This study, presented as a proof-of-concept, indicates that integrating physiological and structural variables such as LAI, plant height, and NDVI can improve wheat yield estimation under controlled conditions. While NDVI alone provided limited predictive strength, its combination with LAI and height-based ratios (especially LAI/H²) enhanced model performance, particularly at 90 and 120 DAS. These results highlight the value of multi-parametric approaches that combine canopy structure and spectral indicators for more reliable yield prediction. The research demonstrated that incorporating both physiological and structural characteristics of plants significantly enhanced the accuracy of wheat yield predictions compared to relying solely on NDVI. While NDVI indicated vegetative health, its combination with LAI and plant height provided insights into canopy structure and nutrient impact, thereby strengthening the model's reliability. Indices derived from these parameters, such as LAI/H², showed superior performance ($R^2 = 0.72–0.81$), effectively capturing canopy density and light-use efficiency. Variability at 60 DAS led to reduced accuracy due to inconsistent nutrient release and soil conditions, but accuracy improved at 90 and 120 DAS. In summary, multi-parametric models that integrate NDVI, LAI, and height-based ratios surpassed single-variable models, highlighting the importance of combining structural and physiological data for dependable yield predictions and precision agriculture.

Regression analyses indicated that wheat yield can be accurately predicted using biophysical parameters like LAI, NDVI, and plant height, measured at various growth stages. Models that integrate height and LAI, especially through ratios such as LAI/H², demonstrated strong predictive capabilities, with parameters measured mid-season (90 DAS) being the most reliable for assessing yield during the growing season. However, these findings are suggestive rather than definitive, as they are derived from a limited dataset of nine plots at a single location, which restricts their generalizability. The high R^2 values indicate local consistency but do not guarantee broader applicability. Overall, the study acts as a proof-of-concept, highlighting the potential of using field-measured canopy traits for yield estimation and laying the groundwork for future validation across diverse soils, climates, and management practices.

When comparing models using AIC, BIC, and cross-validated RMSE, it was evident that simpler models matched or surpassed the predictive capabilities of more complex ones. The LAI/H² (60 DAS) model effectively captured the essential structural–canopy relationship with a minimal number of parameters. Meanwhile, the Height + LAI (90 DAS) model delivered nearly the same level of accuracy and provided a mechanistic understanding of canopy vigor and plant height. Although multi-predictor models incorporating NDVI achieved high R^2 values, they were penalized by AIC/BIC and exhibited signs of multicollinearity. Therefore, models based solely on height and LAI offer the most reliable and adaptable framework for predicting wheat yield in small plots. Future studies should evaluate these simplified relationships across multiple seasons and locations to verify their general applicability.

The findings are based on a small dataset (nine plots) from a single site and season, limiting generalization. NDVI was derived from LAI, introducing potential collinearity, and small plot size with possible soil heterogeneity may have affected consistency. Future research with larger, multi-season, and multi-location datasets using independently measured NDVI is recommended to validate and extend these results.

Data availability

The data generated or analyzed during this study are available from the author Anuj Kumar Dwivedi (E-mail: akdwivedi2009@gmail.com) on reasonable request.

Received: 22 July 2025; Accepted: 12 January 2026

Published online: 26 February 2026

References

1. Aggarwal, P. K. et al. Analysis of yield trends of the rice-wheat system in north-western India. *Outlook Agric.* **29** (4), 259–268 (2000).
2. Aggarwal, P. K. Impact of climate change on Indian agriculture. *J. Plant. Biology-new Delhi.* **30** (2), 189–198 (2003).
3. Xu, Y. Enviro -typing for Deciphering environmental impacts on crop plants. *Theor. Appl. Genet.* **129** (4), 653–673 (2016).
4. Krishnan, P. et al. Web-based crop model: web info Crop-Wheat to simulate the growth and yield of wheat. *Comput. Electron. Agric.* **127**, 324–335 (2016).
5. Uusitalo, L., Lehtikoinen, A., Helle, I. & Myrberg, K. An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environ. Model. Softw.* **63**, 24–31 (2015).
6. Loucks, D. P. & Van Beek, E. *Water Resource Systems Planning and Management: an Introduction To methods, models, and Applications* (Springer, 2017).
7. Gao, Z. et al. Plant height and its relationship with yield in wheat under different irrigation regime. *Irrig. Sci.* **38** (4), 365–371 (2020).
8. Bongiovanni, R. & Lowenberg-DeBoer, J. Precision agriculture and sustainability. *Precision Agric.* **5** (4), 359–387 (2004).
9. Mulla, D. J. Twenty-five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. *Biosyst. Eng.* **114** (4), 358–371 (2013).
10. Dang, Y. P., Pringle, M. J., Schmidt, M., Dalal, R. C. & Apan, A. Identifying the Spatial variability of soil constraints using multi-year remote sensing. *Field Crops Res.* **123** (3), 248–258 (2011).
11. Robertson, M. J. et al. Adoption of variable rate fertilizer application in the Australian grains industry: status, issues and prospects. *Precision Agric.* **13** (2), 181–199 (2012).
12. Tucker, C. J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **8** (2), 127–150 (1979).
13. Moriondo, M., Maselli, F. & Bindi, M. A simple model of regional wheat yield based on NDVI data. *Eur. J. Agron.* **26** (3), 266–274 (2007).
14. Labus, M. P., Nielsen, G. A., Lawrence, R. L., Engel, R. & Long, D. S. Wheat yield estimates using multi-temporal NDVI satellite imagery. *Int. J. Remote Sens.* **23** (20), 4169–4180 (2002).
15. Mkhabela, M. S., Bullock, P., Raj, S., Wang, S. & Yang, Y. Crop yield forecasting on the Canadian prairies using MODIS NDVI data. *Agric. For. Meteorol.* **151** (3), 385–393 (2011).
16. Benedetti, R. & Rossini, P. On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimates and forecast in Emilia Romagna. *Remote Sens. Environ.* **45** (3), 311–326 (1993).
17. Woodcock, C. E. et al. Free access to Landsat imagery. *Sci.* **VOL. 320**, 1011 (2008).
18. Lyle, G., Lewis, M. & Ostendorf, B. Testing the Temporal ability of Landsat imagery and precision agriculture technology to provide high resolution historical estimates of wheat yield at the farm scale. *Remote Sens.* **5** (4), 1549–1567 (2013).
19. Liu, L. et al. Predicting winter wheat condition, grain yield and protein content using multi-temporal Envisat-ASAR and Landsat TM satellite images. *Int. J. Remote Sens.* **27** (4), 737–753 (2006).
20. Potgieter, A. B., Power, B., Mclean, J., Davis, P. & Rodriguez, D. Spatial Estimation of wheat yields from landsat's visible, near infrared and thermal reflectance bands. *Int. J. Remote Sens. Appl.* **4**, 134–143 (2014).
21. Kottek, M., Grieser, J., Beck, C., Rudolf, B. & Rubel, F. World map of the Köppen-Geiger climate classification updated. (2006).
22. Raun, W. R. et al. Improving nitrogen use efficiency in cereal grain production with optical sensing and variable rate application. *Agron. J.* **94** (4), 815–820 (2002).
23. Raun, W. R. & Johnson, G. V. Improving nitrogen use efficiency for cereal production. *Agron. J.* **91** (3), 357–363 (1999).
24. Shanmugam, I., Rethnaraj, J., Rajendran, S. & Manickam, S. Prediction on field crops yield based on analysis of deep learning model. *Indonesian J. Electr. Eng. Comput. Sci.* **30** (1), 518–527 (2023).
25. Behera, B. et al. Microbial consortia for sustaining productivity of non-legume crops: prospects and challenges. *Agricultural Res.* **10** (1), 1–14 (2021).
26. Chaurasia, A. et al. Actinomycetes: an unexplored microorganism for plant growth promotion and biocontrol in vegetable crops. *World J. Microbiol. Biotechnol.* **34** (9), 132 (2018).
27. Monteith, J. L. Climate and the efficiency of crop production in Britain. *Philosophical Trans. Royal Soc. Lond. B Biol. Sci.* **281** (980), 277–294 (1977).
28. Liu, J. et al. Crop residue return sustains global soil ecological stoichiometry balance. *Glob. Change Biol.* **29** (8), 2203–2226 (2023).

Acknowledgements

The authors are much thankful for the Indo-UK project Sustaining Himalayan Water Resources in a changing climate for financial support to carry out this research work. (Grant Code-MES-1023-CED).

Author contributions

Anuj Kumar Dwivedi designed the experiment, collected field data, performed the analyses, and drafted the manuscript. C. S. P. Ojha and V. P. Singh provided technical guidance, critical revisions, and overall supervision. Nisha Singh, Harendra Singh, and Deepak Singh assisted with data processing, statistical modeling, and manuscript editing. All authors contributed to the final version and approved its submission.

Funding

The study was funded by Indian Institute of Technology Roorkee, Uttarakhand, India.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-36373-w>.

Correspondence and requests for materials should be addressed to C.S.P.O. or D.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026