# Journal of Agrometeorology

*(A publication of Association of Agrometeorologists)*

*ISSN : 0972-1665 (print), 2583-2980 (online)*
*Vol. No. 27 (4) : 407-414 (December - 2025)*
*https://doi.org/10.54386/jam.v27i4.3093*
*https://journal.agrimetassociation.org/index.php/jam*


***Research Paper***

# Remote sensing-based yield estimation of wheat crop at farm scale: A case study of Badsu village of Alwar district, Rajasthan


**SUDESH SINGH CHOUDHARY[1,2*] and MAHESH KUMAR JAT[1]**

[1]*Department of Civil Engineering, Malaviya National Institute of Technology Jaipur*
[2]*National Institute of Hydrology, Roorkee*
[*]*Corresponding Author: sudeshjat6994@gmail.com*



## ABSTRACT

Accurate wheat yield estimation at the farm scale is crucial for food security, market strategies, trade planning, and storage decisions. However, predicting crop production using remote sensing at farm scale presents significant challenges. This research aimed to develop a field-scale wheat yield prediction model using multi-temporal vegetation indices derived from Sentinel-2 MSI imagery for the rabi seasons of 2018–19 and 2019–20 from Badsu village in Alwar district, Rajasthan. Vegetation indices derived from cloud-free Sentinel-2 images spanning the crop growth cycle were processed to generate multiple vegetation indices, grouped into greenness, chlorophyll content, and dryness indicators. Spearman's rank correlation (ρ) assessed relationships between indices and wheat yield across various phenological stages and their combinations. Linear and multiple linear regression (MLR) models were developed using the most significant indices. Findings indicate that Wide Dynamic Range Vegetation Index (WDRVI), Normalized Green-Red Difference Index (NGRDI), and Normalized Difference Water Index-2 (NDWI2), representing greenness, chlorophyll, and water stress, respectively, exhibited strong correlations with yield, except during harvesting and crown root initiation. The best-performing model achieved an RMSE of 0.47 t $ha^{-1}$ and an $R^2$ of 0.74, demonstrating the effectiveness of remote sensing indices for precise wheat yield estimation at the field level in diverse agricultural Conditions.



***Keywords:*** Crop yield, MLR, Sentinel data, Spearman correlation, Wheat crop, Vegetation indices


Early prediction of crop biomass and yield is a cornerstone of modern agricultural management, playing a vital role in ensuring food security, optimizing resource use, and supporting economic stability. The ability to forecast crop performance at an early stage allows farmers and policymakers to swiftly address challenges such as pest outbreaks, nutrient shortages, or water stress before they escalate into significant yield losses. Timely interventions can make the difference between a successful harvest and crop failure, emphasizing the value of advanced yield prediction tools (Patel *et al.,* 2023). Furthermore, precise yield estimation is essential for crop insurance, enabling both insurers and farmers to make informed decisions about compensation in the event of adverse weather or other hazards (Van *et al.,* 2020). Early yield forecasts also underpin agricultural early warning systems, guiding farmers in the adoption of best agronomic practices and supporting national strategies for food security and rural development (Adedeji *et al.,* 2020).

Remote sensing (RS) has emerged as a transformative technology for crop monitoring and yield prediction, providing spatially explicit and timely information on plant health and development (Jhajharia, 2025). Over the decades, RS has proven effective in tracking crop growth, detecting stress, and estimating yield across large and small areas. Numerous studies have demonstrated the utility of RS-based frameworks for predicting crop yield and primary productivity, leveraging data from both multispectral and hyperspectral sensors (Murthy *et al.,* 2008; Khaki and Wang, 2019; Patel *et al.,* 2023). Empirical models that use RS-derived vegetation indices (VIs) have become popular, particularly for yield forecasting. These indices, such as the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and others, are calculated from canopy reflectance data and serve as proxies for key crop attributes like biomass, leaf area, and chlorophyll content (Misra *et al.,* 2020). The reflectance characteristics of a crop cano-


***Article info - DOI:*** *https://doi.org/10.54386/jam.v27i4.3093*
*Received:17 June 2025; Accepted: 11 August 2025; Published online : 1 December 2025*

py are influenced by its chemical, morphological, and physiological properties, enabling RS to detect stress and predict yield potential. Multispectral and hyperspectral remote sensing further enhance the ability to estimate canopy-level chlorophyll and nitrogen, which are critical indicators of crop health and yield (Goswami *et al.,* 2021).

Despite these advances, operationalizing RS-based yield prediction at the field scale remains challenging. Many satellite platforms, such as MODIS (with 250-meter resolution) or Landsat (30–60 meters), lack the spatial detail needed to monitor smallholder fields, which are common in countries like India where average farm size around 1.2 hectares (Elders *et al.,* 2022). High-resolution data are essential for accurate crop health assessment and yield estimation in such contexts. The Copernicus Sentinel-2 mission addresses this gap by providing freely available imagery at 10-meter resolution every five days, making it an ideal data for field-scale agricultural monitoring (Kumar *et al.,* 2022; Lonare *et al.,* 2022; Sisheber *et al.,* 2024).

In view of these, the study aims to develop a remote sensing-based empirical model for estimating wheat yield at the farm-scale using Sentinel-2 data, focusing on different vegetation indices along with multiple crop growth stages. This study will provide accurate and timely yield estimation information for assisting crop management and policymakers.

## MATERIALS AND METHODS

### *Study area and data collection*

This study was undertaken at Badsu village in Alwar district, located in the northeastern part of Rajasthan, India (Fig. 1) is characterized by a hot semi-arid climate (Köppen BSh). The study area lies between 27°13′30″ N, 76°55′00″ E and 27°13′00″ N, 76°53′70″ E. Two field surveys were conducted to collect the GPS locations of each crop field using GPS receivers. The spatial crop maps for wheat were generated using the collected wheat crop location data of 76 and 95 point for the rabi season of year 2018-19 and 2019-20, respectively are shown in Fig. 2. As per field survey information, the wheat crop was sown on 15$^{th}$ November and harvested on 31$^{st}$ March in all fields.

### *Satellite data - Sentinel -2 imagery*

The Sentinel-2 constellation consists of two optical satellites, Sentinel-2A and Sentinel-2B, which orbit in a complementary pattern to shorten the revisit interval from 10 days to 5 days, thereby significantly improving the frequency of image collection. Each satellite carries a Multi-Spectral Instrument (MSI) that records data across 13 spectral bands at different spatial resolutions: four key bands, B2 (blue), B3 (green), B4 (red), and B8 (near-infrared), are available at a high resolution of 10 meters, while six additional bands, including the red-edge (B5, B6, B7, B8A) and short-wave infrared (B11, B12), are provided at 20 meters, and the remaining three bands B1 (coastal aerosol), B9 (water vapor), and B10 (SWIR-cirrus) are captured at 60-meter resolution. In this study, surface reflectance products with less than 10% cloud cover were obtained from the ESA Copernicus Open Access Hub (https://browser.dataspace.copernicus.eu), wheat growing seasons (November 15$^{th}$ to March 31$^{st}$) for the years 2018–2019 and 2019–2020. To ensure consistency in spatial resolution, the 20-meter red-edge and short-wave infrared bands were resampled to 10 meters using the nearest neighbour interpolation technique, aligning all relevant bands for accurate and uniform analysis.

### *Vegetation indices*

Seventeen vegetation indices were computed from Sentinel-2 MSI (Table 1). The reflectance bands, i.e., visible, red-edge, near-infrared (NIR), and short-wave infrared (SWIR) bands, were considered for the computation of these vegetation indices. The selected indices represent three fundamental crop health parameters: greenness, chlorophyll content, and dryness. Greenness indices serve as robust indicators of crop biomass, with the presence of green vegetation strongly captured by indices that combine red and NIR spectral data. In the study, Normalized Difference Vegetation Index (NDVI), Transformed Normalized Difference Vegetation Index (TNDVI), Wide Dynamic Range Vegetation Index (WDRVI), and Soil Adjusted Vegetation Index (SAVI) were considered as presentations of greenness. The second group of indices focuses on chlorophyll or nitrogen content, which is closely tied to plant photosynthetic efficiency and fertilization status. Since chlorophyll pigments in leaves influence the reflectance of light in visible and red-edge spectral regions, therefore, Chlorophyll Index Red-edge (CIR), Normalized Green Red Difference Index (NGRDI), Normalized Difference Red Edge (NDRE), and Chlorophyll Vegetation Index (CVI) were selected to monitor these traits. The third parameter dryness, reflects water stress within crops. Dryness indices are based on the principle that water-stressed plants reduce stomatal opening, leading to altered canopy characteristics and elevated leaf temperatures; the Normalized Difference Water Index (NDWI) and Vegetation Condition Index (VCI) were applied to quantify these conditions. Collectively, these indices provide a comprehensive framework for evaluating wheat health and yield potential at the field scale.

### *Phenological stages*

In this study, seven phenological stages of wheat, crown root initiation (CRI), tillering, jointing, booting, milking, dough, and harvesting were considered to assess the performance of vegetation indices (Kumar *et al.,* 2018). For each stage, the average value of each vegetation index (VI) was computed per plot. Since crop yield is influenced not only by a single growth stage but also by conditions throughout the entire growing period, so for that combinations of multiple growth stages were generated (Table 2), and corresponding index values were calculated to capture their cumulative effects.

### *Sensitivity analysis of vegetation indices (VIs)*

The sensitivity analysis of various RS-derived VIs and observed yield with respective growth stages of the crop was analysed using Spearman's rank correlation coefficient (ρ). This non-parametric method was selected for its ability to capture monotonic relationships without assuming data normality. The strength of association was interpreted based on the absolute value of ρ, where values near ±1 indicate a strong correlation, whereas values near 0 indicate weak or no association (Xiao, 2019).

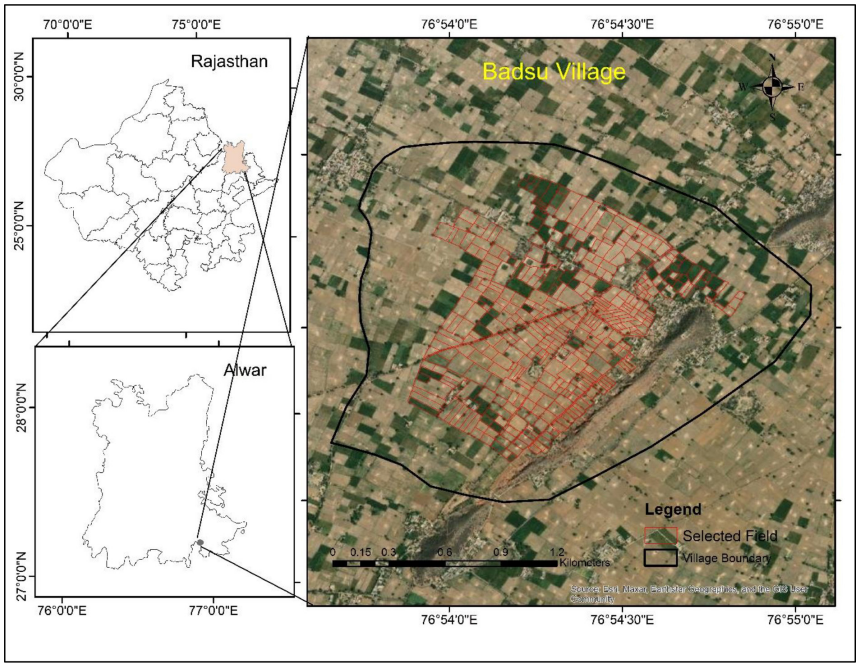


**Fig. 1:** Location map of Badsu village, Alwar, Rajasthan

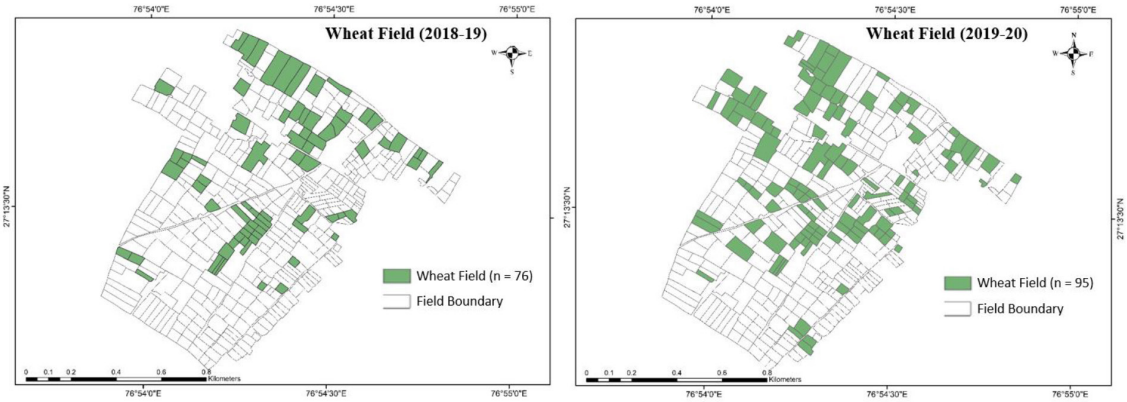


**Fig. 2:** Wheat crop observation location during (a) 2018-19 and (b) 2019-20

**Table 1:** Different vegetation index (VIs) used in this study

| S. No. | Index | Expression | Reference |
|---|---|---|---|
| 1 | NDVI | (NIR- RED)/(NIR+ RED) | Benedetti and Rossini (1993) |
| 2 | TNDVI | SQRT((NIR-RED)/(NIR+RED) +0.5) | Tucker (1979) |
| 3 | WDRVI | (a * NIR- RED)/(a * NIR+ RED) | Gitelson (2004) |
| 4 | SAVI | (1+L) * (NIR- RED)/(NIR+ RED +L) | Huete (1988) |
| 5 | CIR1 | (NIR/RE1)-1 | |
| 6 | CIR2 | (NIR/RE2)-1 | Gitelson *et al.,* (2003) |
| 7 | CIR3 | (NIR/RE3)-1 | |
| 8 | CIR4 | (NIR/RE4)-1 | |
| 9 | NDRE1 | (NIR-RE1)/(NIR+RE1) | |
| 10 | NDRE2 | (NIR-RE2)/(NIR+RE2) | Gitelson and Merzlyak (1994) |
| 11 | NDRE3 | (NIR-RE3)/(NIR+RE3) | |
| 12 | NDRE4 | (NIR-RE4)/(NIR+RE4) | |
| 13 | NGRDI | (GREEN – RED)/ (GREEN + RED) | |
| 14 | CVI | (NIR/GREEN) *(RED/GREEN) | |
| 15 | VCI | 100*(NDVI-NDVImin) / (NDVImax - NDVImin) | Hayes and Decker (1996) |
| 16 | NDWI1 | (NIR – SWIR1)/(NIR+SWIR1) | Bolton and Friedl (2013) |
| 17 | NDWI2 | (NIR – SWIR2)/(NIR+SWIR2) | |

**Table 2:** Combination of growth stages used in the study

| S. No. | Name of Assigned Combination | Description of Combination |
|---|---|---|
| 1 | Avg_all | Taking average of indices of all seven growth stages |
| 2 | Except_CRI | Average of indices of Tillering, Jointing, Booting, Milk, Dough and Harvesting growth stage |
| 3 | Except_harvesting | Average of indices of CRI, Tillering, Jointing, Booting, Milk and Dough growth stage |
| 4 | Except_har_cri | Average of indices of Tillering, Jointing, Booting, Milk and Dough growth stage |
| 5 | Except_cri_till_harve | Average of indices of Jointing, Booting, Milk and Dough growth stage |
| 6 | Except_cri_till_join_harve | Average of indices of Booting, Milk and Dough growth stage |
| 7 | Only_milk_dou | Average of indices of Milk and Dough growth stage |
| 8 | Only_booting_milk | Average of indices of Booting and Milk growth stage |
| 9 | Only_joint_boot | Average of indices of Jointing and Booting growth stage |
| 10 | Only_joint_boot_milk | Average of indices of Jointing, Booting and Milk growth stage |

$$\rho = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)}$$

Where $d_i^2$ = R(Xi) – R (Yi), is the difference between the two ranks of each sample; Xi is the remote sensing-derived VIs, Yi indicates observed crop yield at different growth stage periods, and n is the number of samples.

***Development of empirical models for crop yield prediction***

The empirical models have been developed based on the RS-derived VIs that describe plant growth characteristics, including greenness, chlorophyll, and dryness. The present study hypothesizes that crop yield is a function of remote sensing-derived VIs with time-scale. To test this, an initial set of 17 VIs was evaluated for correlation with observed wheat yield (t $ha^{-1}$) during the Rabi season for the year 2018-19. Spearman's rank correlation was used to identify sensitive VIs for crop yield prediction. VIs that exhibits strong and statistically significant correlations were selected for model development.

Using these selected VIs, both linear and multiple linear regression models were developed using curve-fitting techniques in Python environment. In the developed model, crop yield (t $ha^{-1}$) was considered as the dependent variable, while remote sensing-derived VIs were taken as independent variables. The developed regression models, i.e., linear, polynomial, exponential, and quadratic, are shown in Table 3. The model's accuracy was evaluated by comparing the predicted yield with the observed yield values of the wheat crop.

***Model performance evaluation***

The performance of the developed models was evaluated using standard statistical metrics, i.e., Root Mean Square Error (RMSE), Coefficient of Determination ($R^2$), and Mean Absolute Error (MAE). These metrics were calculated by comparing observed yields against model-simulated yields, enabling selection of the most robust prediction model.

## RESULTS AND DISCUSSION

***Relation between VIs, crop growth stage and crop yield***

The relationship between observed wheat yield and 17 VIs was examined across various growth stages using Spearman's rank correlation coefficient for the *Rabi* season (November to March) of the year 2018–19. The Spearman's rank correlation coefficient is presented using a heatmap (Fig. 3) illustrating how remote sensing-based VIs relate to yield with respect to various growth stages. Generally, the predictive strength of VIs varied significantly with growth stage, showing stronger correlations during mid to late phenological phases. Among four greenness indices, the WDRVI consistently exhibited a strong correlation with yield ($\rho = 0.79$), during the composite "only_joint_boot_milk" stage, revealing its effectiveness in capturing biomass and canopy cover. The NDWI2 showed the highest correlation ($\rho = 0.8$) as stress indices, particularly during the composite stage "Except_har_CRI", indicating its sensitivity to plant water content and drought stress. Within chlorophyll or nitrogen-related indices, the NGRDI emerged as the most reliable and sensitive ($\rho = 0.79$), at the same composite "only_joint_boot_milk" stage, highlighting its usefulness in detecting crop nutrient status. Overall, these findings emphasize the importance of selecting both the appropriate VIs and the optimal crop growth stage to enhance yield prediction accuracy. Indices such as WDRVI, NDWI2, and NGRDI, especially when derived from data spanning the jointing to milky stages, provide robust indicators of biomass accumulation, physiological health, and nutrient availability, making them reliable remote sensing-based VIs for wheat yield estimation and crop monitoring strategies.

***Performance of crop yield models***

Multiple crop yield models were developed using linear and multiple linear regression models based on selected VIs, i.e., WDRVI, NGRDI, and NDWI2, using polynomial curve fitting in a Python environment. Eight different models, including linear, quadratic, cubic, logarithmic, and exponential forms were tested. The performance metrics of models are presented in Table 4. The performance statistics revealed that for all three indices (WDRVI, NGRDI, and NDWI2), the M6 model (Third order polynomial) showed the highest $R^2$ and lower error (RMSE and MAE). For WDRVI-based derived models, the M6 model has the lowest RMSE and MAE,

**Table 3:** Different equations used in this study

| Model number | Model name | Expression |
|---|---|---|
| M1 | Boltzmann sigmoidal | $Y = P_{r1} + \frac{(P_{r2} - P_{r1})}{1 + e^{\frac{(P_{r3} - X)}{P_{r2}}}}$ |
| M2 | Gaussian | $Y = P_{r1} \times e^{-0.5(\frac{X - P_{r2}}{P_{r3}})^2}$ |
| M3 | Lorentzian | $Y = \frac{P_{r1}}{1 + (\frac{X - P_{r2}}{P_{r3}})^2}$ |
| M4 | First Order | $Y = P_{r1} + P_{r2} \times X$ |
| M5 | Second Order | $Y = P_{r1} + P_{r2} \times X + P_{r3} \times X^2$ |
| M6 | Third Order | $Y = P_{r1} + P_{r2} \times X + P_{r3} \times X^2 + P_{r4} \times X^3$ |
| M7 | Sigmoidal dose-response (Slope variable) | $Y = P_{r1} + \frac{P_{r2} - P_{r1}}{1 + 10^{\frac{P_{r3} - X}{P_{r4}}}}$ |
| M8 | Sigmoidal dose-response | $Y = P_{r1} + \frac{P_{r2} - P_{r1}}{1 + 10^{P_{r3} - X}}$ |

Where X – independent variable, Y – dependent variable, and $Pr_{1,2,3,4}$ are constants

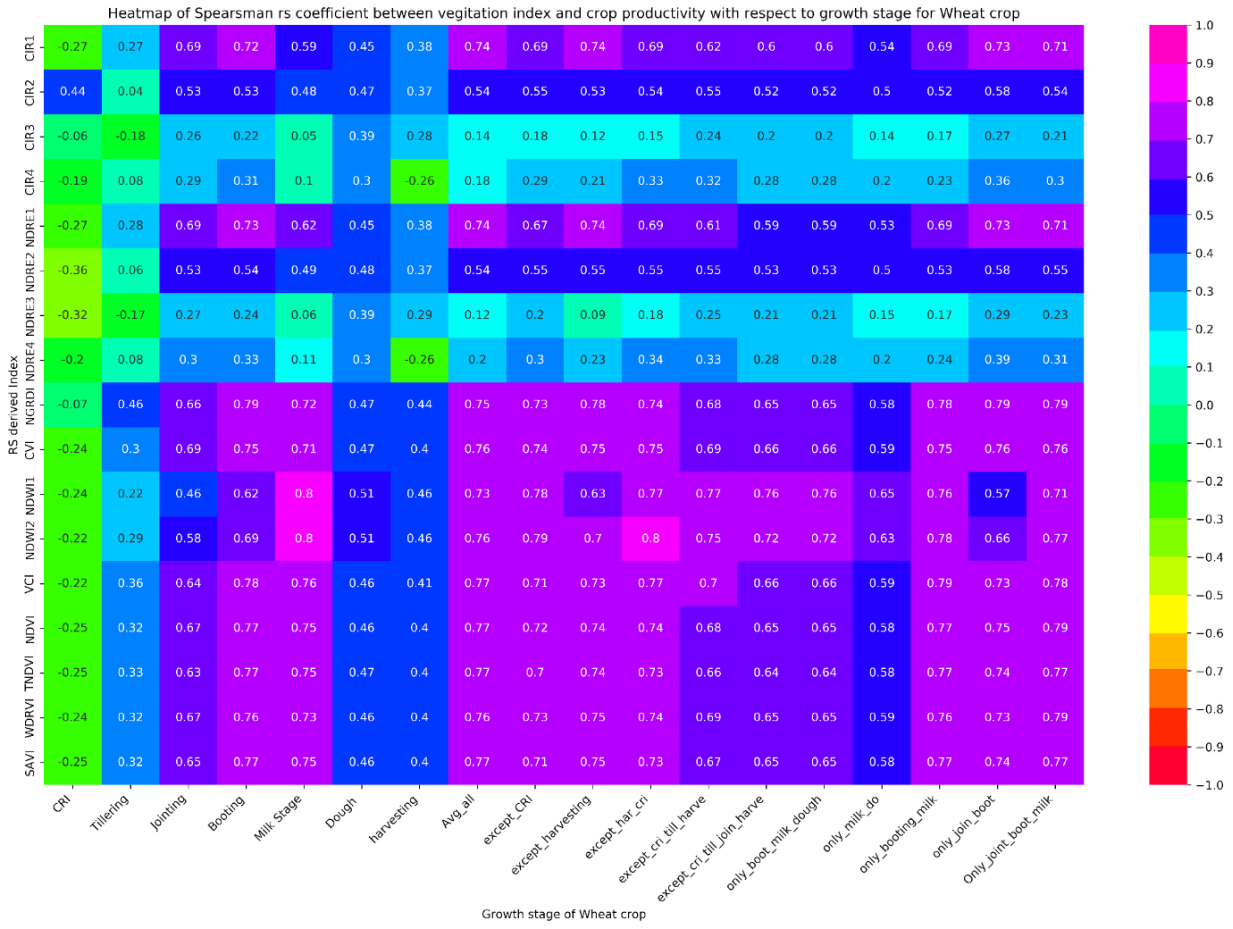


**Fig. 3:** Heatmap of Spearman's rank coefficient between VIs and crop yield to the growth stage of the wheat crop

with values of 0.43 and 0.34, respectively, and the highest $R^2$ of 0.64, which confirms its strength in capturing canopy biomass and greenness related to crop productivity. In the NGRDI-based models, the M6 model portrays the lowest RMSE and MAE of 0.40 and 0.32, respectively, and $R^2$ of 0.69, demonstrating its ability to capture chlorophyll content and nitrogen-related physiological traits.

**Table 4:** Statistical parameters of different models for WDRVI, NGRDI, and NDWI2

| Model | WDRVI | | | NGRDI | | | NDWI2 | | |
|---|---|---|---|---|---|---|---|---|---|
| | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE | RMSE | $R^2$ | MAE |
| M1 | 0.44 | 0.63 | 0.35 | 0.40 | 0.68 | 0.32 | 0.39 | 0.70 | 0.32 |
| M2 | 0.44 | 0.63 | 0.35 | 0.41 | 0.68 | 0.32 | 0.39 | 0.70 | 0.32 |
| M3 | 0.44 | 0.63 | 0.35 | 0.41 | 0.68 | 0.32 | 0.39 | 0.70 | 0.32 |
| M4 | 0.44 | 0.63 | 0.34 | 0.40 | 0.68 | 0.32 | 0.39 | 0.70 | 0.32 |
| M5 | 0.44 | 0.63 | 0.35 | 0.40 | 0.68 | 0.32 | 0.39 | 0.70 | 0.32 |
| M6 | **0.43** | **0.64** | **0.34** | **0.40** | **0.69** | **0.32** | **0.39** | **0.70** | **0.32** |
| M7 | 0.44 | 0.63 | 0.35 | 0.40 | 0.68 | 0.32 | 0.39 | 0.70 | 0.32 |
| M8 | 0.44 | 0.63 | 0.35 | 0.40 | 0.68 | 0.32 | 0.39 | 0.70 | 0.32 |

**Table 5:** Different multivariate models developed based on the M6 model combination

| Models | Equations | RMSE | $R^2$ | MAE |
|---|---|---|---|---|
| All indices | $Y=A+B\times p+C\times p^2+D\times p^3+E\times q+F\times q^2+G\times q^3+H\times r+I\times r^2+J\times r^3$ | **0.37** | **0.74** | **0.30** |
| Without cube | $Y=A+B\times p+C\times p^2+D\times q+E\times q^2+F\times r+G\times r^2$ | **0.38** | **0.72** | **0.31** |
| Without square | $Y=A+B\times p+C\times p^3+D\times q+E\times q^3+F\times r+G\times r^3$ | **0.38** | **0.72** | **0.31** |
| Without linear | $Y=A+B\times p^2+C\times p^3+D\times q^2+E\times q^3+F\times r^2+G\times r^3$ | 0.39 | 0.72 | 0.32 |
| Only linear | $Y=A+B\times p+C\times q+D\times r$ | 0.39 | 0.71 | 0.32 |
| Only square | $Y=A+B\times p^2+C\times q^2+D\times r^2$ | 0.40 | 0.70 | 0.32 |
| Only cube | $Y=A+B\times p^3+C\times q^3+D\times r^3$ | 0.40 | 0.69 | 0.32 |
| Multiply1 | $Y=A+B\times q+C\times q\times r+D\times p\times q\times r+E\times p\text{^}2$ | 0.40 | 0.70 | 0.32 |
| Multiply2 | $Y=A+B\times p+C\times q+D\times r+E\times q\times r+F\times q\times p+G\times p\times r$ | **0.38** | **0.72** | **0.31** |

Here Y = Crop yield (t $ha^{-1}$), p = WDRVI_only_joint_boot_milk, q = NGRDI_only_joint_boot_milk, and r = NDWI2_except_har_cri

The moderate $R^2$ and relatively low RMSE and MAE values indicate reliable predictive capability. Similarly, for the NDWI2-based derived model, the M6 model showed RMSE and MAE of 0.39 and 0.32, with $R^2$ of 0.70, which reflects the sensitivity of this index to canopy water content and crop water stress. Its strong correlation with yield is likely due to the influence of water availability on grain filling and biomass accumulation. A high $R^2$ and low RMSE suggest that NDWI2 effectively captures water-related stress factors influencing final yield, especially in semi-arid or variable rainfall regions.

Further, multivariate models were developed using the integration of three selected vegetation indices, WDRVI (greenness index), NGRDI (chlorophyll index), and NDWI2 (dryness index), that reflect key physiological and environmental factors affecting crop productivity. Based on the M6 (third-order polynomial) model, multiple combinations of multivariate models were generated as shown in Table 5. These combinations include, (a) All indices: This combination incorporates all parameters from the third-order polynomial equation for each index; (b) Without cube parameters: This combination excludes the cube parameters from the third-order polynomial equation; (c) Without square parameters: This combination excludes the square parameters, while retaining the linear and cubic parameters; (d) Without linear parameters: This combination excludes the linear parameters, while keeping only the square and cubic parameters; (e) Only linear, square, and cubic parameters: Each combination includes only the linear, square, or cubic parameters from all indices, respectively; (f) Multiplicative index combinations: In this combination, indices are multiplied together.

The results of the multivariate models showed that models incorporating all third-order parameters (all indices), without the cube parameter, without square parameters, and the multiplicative index combination performed better in estimating yield using the three indices. When all three VIs were incorporated in the multivariate model development, the RMSE decreased by up to 0.37 t $ha^{-1}$, and $R^2$ increased to 0.74. Based on RMSE, $R^2$, and MSE, the four models (all cubic, without Cube, without square, and the multiplicative index combination) were further used to estimate wheat yield for the Rabi season of year 2019-20. These combinations were tested to evaluate whether incorporating different mathematical relationships among the indices improves model performance or not, ultimately aiming to develop a robust multivariate yield prediction model.

***Validation of models***

The developed multivariate models were validated using observed wheat yield data for the Rabi season of year 2019–20. The comparison of different Models shows that the third-order polynomial-based (all_indices) combination model predicted yield with ±10% deviation from the 1:1 line, indicating a good agreement between observed and simulated crop yield (Fig. 4 and Table 6). Whereas, other models predicted yield showed over-prediction or under-prediction within the ±10% deviation. These patterns highlight the importance of including third-order polynomial components for reliable yield estimation at the field scale.

Among all tested combinations, the model incorporating all third-order polynomial forms (All index) of selected indices, i.e., WDRVI, NGRDI, and NDWI2, exhibited the highest accuracy, with

**Table 6:** Validation results of developed multivariate models

| Models | Equations | RMSE | $R^2$ | MAE |
|---|---|---|---|---|
| All index | $Y=A+B\times p+C\times p^2+D\times p^3+E\times q+F\times q^2+G\times q^3+H\times r+I\times r^2+J\times r^3$ | **0.47** | **0.74** | **0.37** |
| Without cube | $Y=A+B\times p+C\times p^2+D\times q+E\times q^2+F\times r+G\times r^2$ | 0.58 | 0.71 | 0.45 |
| Without square | $Y=A+B\times p+C\times p^3+D\times q+E\times q^3+F\times r+G\times r^3$ | 0.58 | 0.70 | 0.44 |
| Multiply2 | $Y=A+B\times p+C\times q+D\times r+E\times q\times r+F\times q\times p+G\times p\times r$ | 0.59 | 0.70 | 0.46 |

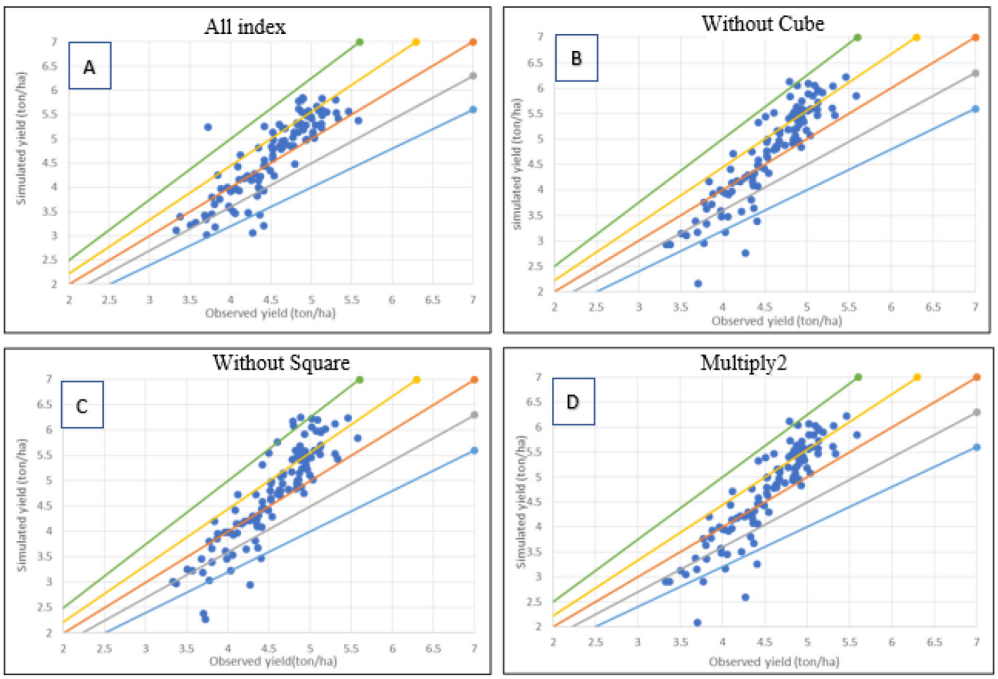


**Fig. 4 :** Observed versus simulated yield for model validation; (a) for all index based model; (b) for without cube-based model; (c) for without square based model; (d) for multiply2 based model

$R^2$ of 0.74 and RMSE of 0.47 t $ha^{-1}$ (Fig. 4). This model outperformed as compared to other combinations, i.e., without cube, without square, and the multiplicative index model. It demonstrates the effectiveness of capturing non-linear relations among corresponding VIs that represent canopy greenness, chlorophyll content, and crop dryness conditions.

Unlike other studies, such as Du *et al.,* (2025) demonstrated the performance of the yield prediction model with $R^2$ values between 0.57–0.6 and RMSE of 0.42–0.58, using dense time-series data, crop model, and machine learning algorithms. However, the developed multivariate model in this study offers a simpler and more interpretable polynomial framework that still achieves a robust accuracy. In addition, consideration of important physiological stages (tillering, jointing, booting, milking, dough), excluding CRI and harvesting, improves the model's yield prediction accuracy by avoiding noise from less informative periods.

## CONCLUSIONS

This study demonstrates that wheat yield can be accurately estimated at the farm scale in semi-arid regions by integrating Sentinel-2 satellite data-derived VIs such as WDRVI, NGRDI, and NDWI2 in a multivariate model. The results of the study depicted that consideration of VIs at critical phenological stages, particularly "only_joint_boot_milk" and "except_har_cri" are most informative for yield estimation modelling. The developed univariate polynomial models showed strong predictive performance with WDRVI ($R^2$ = 0.64), NGRDI ($R^2$ = 0.69), and NDWI2 ($R^2$ = 0.70), while a multivariate polynomial model further improved prediction accuracy ($R^2$ = 0.74). Furthermore, comparison of observed yield and simulated yield during the validation shows the robustness ($R^2$ = 0.74) of the developed multivariate model. The findings have significant practical implications, providing a cost-effective and scalable framework for field-scale yield forecasting in smallholder agricultural systems. This research offers a transparent and reproducible toolset that can be adopted by policymakers, agricultural extension agencies, and farmer collectives to improve data-driven decision-making and promote sustainable wheat production.

## ACKNOWLEDGEMENT

First author sincerely acknowledges the Ministry of Human Resource Development (MHRD), Government of India, for providing institutional Ph.D. support, which enabled me to undertake and complete this research work.

***Funding:*** Financial support received from MHRD is duly acknowledged.

***Data availability*:** Satellite data utilized in this study is accessible through open platforms. Field data and code can be made available upon reasonable request only.

***Conflict of Interests*:** The authors declare that there is no conflict of interest related to this article.

***Authors contribution*: S. S. Choudhary** - Data collection, Data Analysis, Conceptualization, Methodology, Visualization, Writing original draft, Writing-review; **M. K. Jat:** Supervision, Review, and editing.

***Disclaimer:*** The contents, opinions, and views expressed in the research article published in the Journal of Agrometeorology are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

***Publisher's Note***: The periodical remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

Adedeji, O., Olusola, A., James, G., Shaba, H. A., Orimoloye, I. R., Singh, S. K., and Adelabu, S. (2020). Early warning systems development for agricultural drought assessment in Nigeria. *Environ. Monit. Assess.,* 192: 1-21.

Benedetti, R. and Rossini, P., (1993). On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimate and forecast in Emilia Romagna. *Remote Sens. Environ.,* 45(3):.311-326.

Bolton, D.K. and Friedl, M.A., (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agric. Forest Meteorol.,* 173:74-84.

Du, X., Zhu, J., Xu, J., Li, Q., Tao, Z., Zhang, Y., Wang, H. and Hu, H. (2025). Remote sensing-based winter wheat yield estimation integrating machine learning and crop growth multi-scenario simulations. *Intern. J. Digital Earth,* 18(1):2443470.

Elders, A., Carroll, M. L., Neigh, C. S., D'Agostino, A. L., Ksoll, C., Wooten, M. R., and Brown, M. E. (2022). Estimating crop type and yield of small holder fields in Burkina Faso using multi-day Sentinel-2. *Remote Sensing Appl.: Soc. Environ.,* 27: 100820.

Gitelson, A., and M.N. Merzlyak. (1994). Quantitative estimation of chlorophyll-A using reflectance spectra-Experiments with autumn chestnut and maple leaves. *J. Photochem. Photobiol.,* B 22:247–252. doi:10.1016/1011-1344(93)06963-4

Gitelson, A.A. (2004). Wide Dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *J. Plant Physiol.,* 161: 165-173.

Gitelson, A.A., A. Viña, T.J. Arkebauer, D.C. Rundquist, G. Keydan, and B. Leavitt. (2003). Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys. Res. Lett.,* 30(5):1248-1253. doi:10.1029/2002GL016450

Goswami, S., Choudhary, S. S., Chatterjee, C., Mailapalli, D. R., Mishra, A., and Raghuwanshi, N. S. (2021). Estimation of nitrogen status and yield of rice crop using unmanned aerial vehicle equipped with multispectral camera. *J. Appl. Remote Sens.,* 15(4): 042407-042407.

Hayes, M.J. and Decker, W.L. (1996). Using NOAA AVHRR data to estimate maize production in the United States Corn Belt. *Remote Sens.,* 17(16):3189-3200.

Huete, A.R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.,* 25:295-309.

Jhajharia, K. (2025). Wheat yield prediction of Rajasthan using climatic and satellite data and machine learning techniques. *J. Agrometeorol.,* 27(1): 63-66. https://doi.org/10.54386/jam.v27i1.2807

Khaki, S., and Wang, L. (2019). Crop yield prediction using deep neural networks. *Front. Plant Sci.,* 10: 621.

Kumar, P., Prasad, R., Gupta, D.K., Mishra, V.N., Vishwakarma, A.K., Yadav, V.P., Bala, R., Choudhary, A. and Avtar, R. (2018). Estimation of winter wheat crop growth parameters using time series Sentinel-1A SAR data. *Geocarto Intern.,* 33(9):942-956.

Kumar, D. A., Neelima, T. L., Srikanth, P., Devi, M. U., Suresh, K., and Murthy, C. S. (2022). Maize yield prediction using NDVI derived from Sentinal 2 data in Siddipet district of Telangana state. *J. Agrometeorol.,* 24(2): 165-168. https://doi.org/10.54386/jam.v24i2.1635

Lonare, A., Maheshwari, B., and Chinnasamy, P. (2022). Village level identification of sugarcane in Sangali, Maharashtra using opensource data. *J. Agrometeorol.,* 24(3): 249-254. https://doi.org/10.54386/jam.v24i3.1688

Misra, G., Cawkwell, F., and Wingler, A. (2020). Status of phenological research using Sentinel-2 data: A review. *Remote Sens.,* 12(17): 2760.

Murthy, C. S., Sai, M. S., Kumari, V. B., Prakash, V. S., and Roy, P. S. (2008). Study of crop condition and assessment of agricultural drought in rabi season using IRS–AWiFS images. *J. Agrometeorol.,* 10(1): 19-26. https://doi.org/10.54386/jam.v10i1.1164

Patel, N. R., Pokhariyal, S., and Singh, R. P. (2023). Advancements in remote sensing-based crop yield modelling in India. *J. Agrometeorol.,* 25(3): 343-351. https://doi.org/10.54386/jam.v25i3.2316

Sisheber, B., Marshall, M., Mengistu, D., and Nelson, A. (2024). The influence of temporal resolution on crop yield estimation with Earth Observation data assimilation. *Remote Sens. Appl.: Soc. Environ.*, 36: 101272.

Tucker C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.,* 8(2): 127-150. https://doi.org/10.1016/0034-4257(79)90013-0

Van Klompenburg, T., Kassahun, A. and Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Comp. Electron. Agric.,* 177:105709.

Xiao, W. (2019). December. Novel online algorithms for nonparametric correlations with application to analyze sensor data. *In 2019 IEEE Intern. Conf. Big Data,* (404-412). DOI: 10.1109/BigData47090.2019.9006483.