Plant Archives Vol. 26, Supplement 2, 2026 pp. 3097-3104 e-ISSN:2581-6063 (online), ISSN:0972-5210

Plant Archives

Journal homepage: http://www.plantarchives.org
DOI Url : https://doi.org/10.51470/PLANTARCHIVES.2026.v26.supplement-2.395

COMPARATIVE PERFORMANCE OF ADVANCED REGRESSION TECHNIQUES FOR CHICKPEA YIELD PREDICTION UNDER DIVERSE AGRO-CLIMATIC CONDITIONS

Sunil Kumar[1], Manoj Kumar[2*], Anurag[3] and Rani Sarita[2]
[1]Department of Plant Breeding and Genetics, CCS HAU, Hisar, Haryana, India
[2]Department of Mathematics and Statistics, CCS HAU, Hisar, Haryana, India
[3]Department of Agrometerology, CCS HAU, Hisar, Haryana, India
*Corresponding author E-mail: m25424553@gmail.com
(Date of Receiving : 07-05-2026; Date of Revision : 29-06-2026; Date of Acceptance : 20-07-2026)

ABSTRACT

An experiment was conducted in the year 2023–24 under multi-location trials in the department of Genetics and Plant breeding, Pulses Section CCS HAU, Hisar under three conditions normal sown, late sown, and rainfed conditions in a randomized block design with three replications and twenty-seven varieties. Advanced regression techniques were used to assess chickpea yield prediction. Correlation analysis showed that biometrical traits had a stronger and more consistent influence on yield than weather variables, with number of pods per plant exhibiting the highest positive correlation across all environments normal (r = 0.80***), late sown (r = 0.67**), and rainfed (r = 0.64**). Plant height was significantly correlated with yield under normal (r = 0.58**) and late sown (r = 0.63**) conditions, while early flowering and maturity favored yield under stress environments. Rainfed conditions consistently showed superior performance, with the highest baseline yield (intercept) in PCR and PLSR models, indicating inherently favorable trait–environment interactions under natural rainfall regimes. Although multiple regression models under normal sowing showed high explanatory power (R^2 = 0.84), VIF analysis revealed substantial multicollinearity among climatic variables. Advanced techniques such as PCR and PLSR effectively addressed this issue, with PLSR delivering stable and accurate predictions across all environments and capturing complex yield determinants under rainfed conditions through higher-order components. Overall, the strong performance of rainfed models highlights their yield potential and underscores the importance of selecting stress-adaptive biometrical traits particularly pod number, plant height, and earliness to achieve stable chickpea productivity under rainfed agro-ecosystems.
***Keywords*:** Chickpea yield, Regression models, PLSR, PCR, Stepwise regression, Multicollinearity, Rainfed condition, Yield prediction.

Introduction

Chickpea (*Cicer arietinum L.*) is one of the most important pulse crops, playing a vital role in ensuring nutritional security and enhancing farm income in India. Being a rich source of protein, vitamins, and minerals, chickpea contributes significantly to dietary balance, particularly in vegetarian populations. In addition to its nutritional importance, it improves soil fertility through biological nitrogen fixation, making it an integral component of sustainable agricultural systems. Chickpea is believed to have originated in the Fertile Crescent region of the Near East and subsequently spread to South Asia, the Mediterranean basin, and Africa. At present, it is cultivated in more than 50 countries across Asia, Africa, Australia, Europe, and the Americas. India dominates global chickpea production, contributing more than 60% of the total output, followed by countries such as Australia, Turkey, Myanmar, and Ethiopia.

Despite its global importance, chickpea productivity is highly sensitive to sowing conditions. Variations such as normal sowing, late sowing, and rainfed conditions significantly influence crop growth, phenology, and yield outcomes. These variations necessitate the development and comparison of predictive models tailored to specific sowing environments

Accurate prediction of chickpea yield is crucial for efficient crop planning, market intelligence, policy formulation, and risk management. This need becomes even more critical under the current scenario of increasing climate variability, where fluctuations in temperature, rainfall, and other environmental factors significantly influence crop productivity. Reliable yield prediction models can assist policymakers, researchers, and farmers in making informed decisions regarding resource allocation, cropping strategies, and contingency planning.

Traditionally, linear regression models have been widely used for yield prediction. However, these models often exhibit limitations in handling complex agricultural datasets characterized by non-linear relationships, multicollinearity among predictors, and high-dimensional interactions involving climatic, soil, and management variables. As a result, their predictive performance may be suboptimal under diverse agro-climatic conditions.

To overcome these limitations, advanced regression techniques such as Multiple Linear Regression (MLR), Stepwise Regression (SR), Principal Component Regression (PCR), and Partial Least Squares Regression (PLSR) have been increasingly applied. These methods are capable of addressing issues like multicollinearity and dimensionality reduction, thereby improving prediction accuracy and robustness.

Objective of the Study: The present study aims to evaluate and compare the predictive performance of different regression models MLR, Stepwise Regression, PCR, and PLSR for chickpea yield under three distinct sowing conditions normal Sown (NS), late Sown (LS) and Rainfed Conditions (RF). The study focuses on identifying the most accurate and robust modeling approach for each condition, thereby providing insights into model suitability under varying agro-environmental scenarios**.**

Significance of the Study: This research contributes to the advancement of statistical modeling in agriculture by:

I Enhancing the accuracy of chickpea yield prediction

II Addressing challenges of multicollinearity and high-dimensional data

III Providing condition-specific model recommendations

IV Supporting climate-resilient agricultural planning

The findings are expected to aid researchers, extension workers, and policymakers in selecting appropriate predictive models for improving chickpea productivity under diverse farming conditions.

Material and Methods

Study Area and Experimental Design: The secondary data was taken from the experiment conducted during 2023–24 at the Pulses Section of the Department of Genetics and Plant Breeding under multi-location trials. The study comprised three sowing environments: Normal Sown (NS), Late Sown (LS), and Rainfed (RF) conditions. The experiment was laid out in a Randomized Block Design (RBD) with three replications, involving 27 chickpea varieties. Observations were recorded on seed yield, biometrical traits, and weather parameters.

Variables and Abbreviations: The following variables were used in the analysis:

PH: Plant Height (cm), HSW: 100 Seed Weight (g), DFF: Days to 50% Flowering, NPP: Number of Pods per Plant, DM: Days to Maturity, Tmax: Maximum Temperature (°C),Tmin: Minimum Temperature (°C), Rhm: Relative Humidity (Morning, %), TRF: Total Rainfall (mm), PE: Pan Evaporation (mm), SS: Sunshine Hours

Regression Techniques Used: To predict chickpea yield and to examine the relationship between yield and a set of agro-climatic and management variables, four regression techniques Multiple Linear Regression, Stepwise Regression, Principal Component Regression, and Partial Least Squares Regression (PLSR) were employed. These methods were selected to account for linear effects as well as multicollinearity among predictor variables**.**

Multiple Linear Regression (MLR): Multiple Linear Regression was used as the baseline technique to model the linear relationship between chickpea yield and multiple explanatory variables such as rainfall, temperature, and soil and management factors.

The general form of the model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

Where: Y = Chickpea yield (kg/ha), $X_1, X_2, X_3 \ldots$, Xp are the predictor variables

β_0 = Intercept , β_i = Regression coefficient

ε = random error term.

The significance of individual regression coefficients was tested using the *t*-test, while the overall model adequacy was evaluated using the *F*-test. Multicollinearity among predictors was assessed using the variance Inflation Factor (VIF).

Stepwise Regression: Stepwise regression was applied to identify the most influential variables affecting chickpea yield. Both forward selection and backward elimination procedures were considered. In forward selection**,** variables were added sequentially based on statistical significance. In backward elimination, all variables were initially included, and non-significant variables were removed step by step. The selection of variables at each step was guided by the Akaike Information Criterion (AIC) and significance at the 5% level. Stepwise regression helped reduce model complexity and improve interpretability by retaining only significant predictors.

Principal Component Regression (PCR): Principal Component Regression was employed to overcome multicollinearity among predictor variables. In this method, the original correlated predictors were transformed into a smaller set of uncorrelated variables called principal components (PCs) using Principal Component Analysis (PCA)**.**

The principal components are defined as linear combinations of the original variables:

$$Z_j = a_{j1}X_1 + a_{j2}X_2 + \cdots \ldots\ldots + a_{jp}X_p$$

Where: Zj = j^{th} principal component, a_{ji} = component loading.

Only those principal components explaining a substantial proportion of total variance were retained and used as predictors in a linear regression model with chickpea yield as the response variable. The number of components was selected based on cumulative explained variance and cross-validation results.

Partial Least Squares Regression (PLSR): Partial Least Squares Regression was used to model the relationship between chickpea yield and a large set of correlated predictors by extracting latent variables that maximize the covariance between predictors and the response variable.

Unlike PCR, which considers only the variance in predictors, PLSR simultaneously accounts for both predictor and response variation. The PLSR model can be expressed as:

$$X = TP' + E; \quad Y = UQ' + F$$

Where: X = matrix of predictor variables, Y = chickpea yield;. T and U = score matrices ; P and Q = loading matrices, E and F = error matrices. The optimal number of latent components was determined using cross-validation, minimizing prediction error. PLSR is particularly suitable for agricultural datasets with multicollinearity and limited sample size.

Model Evaluation: The predictive performance of all regression models was evaluated using coefficient of determination (R^2), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The regression technique yielding the highest predictive accuracy and lowest error was considered the most suitable for chickpea yield prediction.

Result and Discussion

The correlation analysis shown in table 1 shows that biometrical traits had a stronger and more consistent association with chickpea yield than weather variables across all sowing conditions. Among all variables, the number of pods per plant (NPP) showed the highest and most consistent positive correlation with yield, being highly significant under normal sown (r = 0.80**)**, late sown (r = 0.67), and rainfed (r = 0.64)** conditions. This indicates that an increase in pod number directly contributes to higher yield and identifies NPP as the most important yield-determining trait across environments. Plant height (PH) also exhibited a significant positive relationship with yield under normal (r = 0.58)** and late sown (r = 0.63)** conditions, suggesting that taller plants tend to produce higher yields in favorable and moderately stressed environments. However, under rainfed conditions, the relationship was positive but non-significant (r = 0.37), indicating reduced influence under moisture stress. In contrast, 100-seed weight (HSW) showed a negative association with yield, which was significant under late sown (r = –0.51)** and rainfed (r = –0.45)* conditions, suggesting a possible trade-off between seed size and number under stress. Similarly, days to flowering (DFF) had a significant negative correlation under late sown (r = –0.57)** and rainfed (r = –0.55)** conditions, indicating that early flowering genotypes perform better under stress environments. Days to maturity (DM) also showed a significant negative relationship under normal (r = –0.49)** and *late sown* (r = –0.38)* conditions, suggesting that early maturing varieties tend to yield better, especially when terminal stress is likely. Most weather variables, including maximum temperature (Tmax), minimum temperature (Tmin), rainfall (TRF), rainy days (NR), pan evaporation (PE), and sunshine hours (SS), exhibited weak and non-significant correlations with yield across all conditions. This indicates that short-term climatic variations had limited direct influence on yield compared to genetic traits. An exception was relative humidity (Rhm), which showed a weak positive correlation (r = 0.19) under rainfed conditions, though not significant. Overall, the results clearly demonstrate that yield variation in chickpea is predominantly governed by biometrical traits rather than environmental factors, with number of pods per plant, plant height, and earliness (early flowering and maturity) emerging as key determinants for improving yield across diverse sowing conditions, particularly under late sown and rainfed stress environments.

Table 1: Correlation coefficient of Chickpea yield with biometrical and weather variables under for Normal Sown (NS), Late Sown (LS) and Rainfed (RF) conditions.

Variables	Normal Sown (NS)	Late Sown (LS)	Rainfed (RF)
PH (cm)	0.58**	0.63**	0.37
HSW (g)	-0.32	-0.51**	-0.45*
DFF (days)	-0.35	-0.57**	-0.55**
NPP	0.80**	0.67**	0.64**
DM (days)	-0.49**	-0.38*	-0.12
Tmax (^{0}C)	-0.09	-0.09	-0.16
Tmin (^{0}C)	-0.13	-0.14	-0.19
Rhm (%)	-0.06	-0.02	0.19
TRF (mm)	-0.21	-0.13	-0.12
NR (days)	-0.15	0.02	-0.16
PE (mm)	-0.002	-0.10	-0.27
SS (hours)	-0.06	-0.16	-0.29

From table. 2 it is clear that, the multiple regression (MR) models explained a large proportion of yield variability, with coefficients of determination (R^2) of 0.84 normal sown, 0.72 for late sown and 0.76 for rainfed conditions while the stepwise regression (SR) models retained comparable explanatory power (R^2 = 0.82, 0.70, and 0.72, respectively), indicating efficient variable selection without substantial loss of information. Under Normal Sown (NS) conditions, biometrical traits were the dominant yield drivers. The number of pods per plant (NPP) showed a strong and significant positive effect in the SR model (β = 24.39, $p < 0.01$), implying that an increase of one pod per plant resulted in an average yield gain of about 24.4 kg ha^{-1}, holding other factors constant. Plant height (PH) also had a significant positive contribution (β = 18.65, $p < 0.05$), suggesting that taller plants enhanced yield potential. Among weather variables, pan evaporation (PE) exerted a significant positive influence (β = 207.77, $p < 0.01$), highlighting the role of atmospheric demand under optimal sowing. Conversely, maximum temperature (Tmax) showed a negative effect (β = –86.77), indicating yield penalties under elevated temperatures. The SR model explained 82% of yield variability, confirming the robustness of selected predictors. Whereas in Late Sown (LS) conditions, yield prediction was moderately strong (R^2 = 0.70–0.72), reflecting increased environmental stress. Plant height remained a significant positive predictor (β = 19.99, $p < 0.05$), while days to maturity (DM) showed a positive but non-significant contribution (β = 16.52), suggesting that longer crop duration partially compensated for delayed sowing. The negative coefficient of minimum temperature (Tmin) (β = –32.37) indicates sensitivity of yield to cooler night temperatures. Several weather variables were excluded in the SR model, demonstrating their limited and inconsistent influence under late sowing. The Rainfed (RF) condition exhibited the highest intercept values in both MR (2191.55) and SR (3091.00) models, indicating a higher baseline yield potential compared to NS and LS conditions. This suggests that rainfed environments provided inherently favorable growing conditions, likely due to better rainfall distribution and reduced heat stress. In the SR model, days to maturity (DM) (β = 13.85, $p < 0.01$) and minimum temperature (Tmin) (β = 14.07, $p < 0.01$) emerged as significant positive predictors, implying that longer crop duration and warmer night temperatures enhanced yield under rainfed systems. The RF model retained fewer variables yet explained 72% of yield variability, indicating a stable and efficient predictive structure dominated by stress-adaptive traits. Overall, the quantitative results demonstrate that biometrical traits contribute more consistently to chickpea yield than weather variables across environments**.** The strong performance and high intercept under rainfed conditions emphasize its superior yield potential and model stability, while stepwise regression effectively identified key predictors with minimal loss of explanatory power. These findings highlight the importance of selecting for pod number, plant height, and adaptive maturity traits, particularly for enhancing chickpea productivity under rainfed agro-ecosystems.

Table.3 the variation inflation facto of multiple regression shows extremely high VIF values for maximum temperature (VIF = 126.77), minimum temperature (69.31), pan evaporation (62.18), and relative humidity (30.07), indicating strong linear dependence among climatic variables. Similar or even higher levels of multicollinearity were detected under late sown conditions, where maimum temperature (209.38) and pan evaporation (48.74) showed critical inflation,and under rainfed conditions, where relative humidity morning (70.52) and pan evaporation (51.59) exhibited severe collinearity. In contrast, stepwise regression (SR) substantially reduced multicollinearity, with most retained variables showing VIF values below 5, indicating acceptable independence among predictors. For example, under rainfed conditions, all retained variables exhibited VIF values close to unity (PH = 1.18, HSW = 1.04, DM = 1.22, Tmin = 1.23), reflecting a highly stable and reliable model structure. Even under normal sown conditions, SR reduced the VIF of pan evaporation from 62.18 to 14.12 and eliminated highly collinear temperature variables altogether.

The coefficient of determination (R^2) remained largely stable after stepwise selection 0.82 (NS), 0.75 (LS), and 0.72 (RF) demonstrating that substantial reduction in multicollinearity was achieved without significant loss of explanatory power. Overall, the pooled VIF analysis confirms that stepwise regression effectively addresses multicollinearity, while the comparatively low VIF values under rainfed conditions further highlight the stability and robustness of rainfed yield prediction models.

Table 2: Relationship of yield of chickpea for Normal Sown (NS), Late Sown (LS) and Rainfed (RF) condition with Biometrical and weather variables by using multiple (MR) and stepwise SR) regression

	Normal Sown (NS)		**Late Sown (LS)**		**Rainfed (RF)**	
Variables	MR	SR	MR	SR	MR	SR
Intercept	-468.56 (3682.39)	-171.74 (1675.04)	122.55 (3438.97)	166.43 (2545.42)	2191.55 (2198.02)	**3091.00 (1210.53)**
PH (cm)	18.91 (11.95)	**18.65*** (6.73)	15.26 (12.03)	**19.99*** (7.51)	5.88 (6.80)	5.56 (4.03)
HSW (g)	-15.39 (21.98)	–	-25.16 (18.90)	-29.38 (14.02)	-40.50 (20.23)	-46.40 (12.76)
DFF (%)	16.21 (20.58)	21.79 (14.71)	-15.08 (10.05)	-14.52 (8.06)	-21.61 (7.67)	-25.70 (4.94)
NPP	26.07 (10.78)	**24.39**** (6.46)	7.12 (12.46)	–	8.47 (8.20)	–
DM (days)	1.97 (16.31)	–	12.75 (21.33)	16.52 (13.31)	17.04 (10.74)	**13.85**** (7.58)
Tmax (°C)	-76.08 (57.42)	-86.77 (22.91)	13.09 (54.28)	–	-19.67 (32.97)	–
Tmin (°C)	-2.31 (53.34)	–	-40.13 (43.19)	-32.37 (18.43)	-14.91 (26.04)	**14.07**** (4.33)
Rhm (%)	-16.54 (15.79)	-22.72 (8.41)	3.28 (12.77)	–	-4.65 (17.36)	–
TRF (mm)	6.81 (7.83)	–	1.41 (5.61)	–	3.36 (11.41)	–
NR (days)	-125.63 (137.89)	–	-14.97 (96.61)	–	-42.52 (129.84)	–
PE (mm)	197.13 (145.77)	**207.77**** (61.99)	123.91 (94.93)	129.62 (78.95)	54.42 (80.04)	–
SS (hours)	14.73 (78.42)	–	-46.57 (52.12)	-44.86 (29.39)	-18.90 (54.54)	–
R^2	**0.84**	**0.82**	**0.72**	**0.70**	**0.76**	**0.72**

Table 3 : Variance Inflation Factor (VIF) for chickpea yield models under Normal Sown (NS), Late Sown (LS), and Rainfed (RF) conditions using Multiple Regression (MR) and Stepwise Regression (SR)

	Normal Sown (NS)		**Late Sown (LS)**		**Rainfed (RF)**	
Variables	MR	SR	MR	SR	MR	SR
PH (cm)	8.10	3.23	7.56	3.80	2.52	1.18
HSW (g)	1.36	–	1.82	1.29	1.96	1.04
DFF (%)	3.54	2.27	2.69	2.23	2.90	1.61
NPP	4.39	1.98	5.05	–	4.29	–
DM (days)	3.35	–	6.43	3.22	1.84	1.22
Tmax (°C)	**126.77**	25.34	**209.38**	–	**79.17**	–
Tmin (°C)	**69.31**	–	**84.01**	19.69	**33.29**	1.23
Rhm (%)	**30.07**	10.71	**36.35**	–	**70.52**	–
TRF (mm)	6.03	–	5.72	–	**21.18**	–
NR (days)	6.68	–	6.06	–	**23.16**	–

PE (mm)	**62.18**	14.12	**48.74**	**43.40**	**51.59**	–
SS (hours)	**21.72**	–	**17.73**	7.26	**10.79**	–
R^2	**0.84**	**0.82**	**0.72**	**0.75**	**0.76**	**0.72**

VIF > 10 indicates severe multicollinearity; '–' indicates variable not retained in stepwise regression.

The analysis of PCR and PLSR models was shown in table. 6.4 and it was observed that PLSR model consistently outperformed PCR, achieving a very high coefficient of determination (R^2 = 0.99) under all conditions (NS, LS, and RF), compared to PCR (R^2 = 0.81 for NS and 0.65 for LS and RF), indicating superior predictive accuracy and model efficiency. Under Normal Sown (NS) conditions, PCR showed moderate explanatory power, with Component 2 (–170.84) and Component 3 (–76.45) exerting strong negative effects on yield, while Component 1 (18.40) had a weak positive influence. In contrast, PLSR produced strong, positive, and highly significant coefficients for the first five components (e.g., Comp.1 = 195.93, Comp.2 = 75.42, Comp.3 = 70.02), indicating that yield is strongly governed by combined effects of major latent factors. The much smaller standard errors in PLSR further confirm its stability and precision. In Late Sown (LS) conditions, PCR again showed lower explanatory power (R^2 = 0.65) with negative contributions from Comp.2 and Comp.3, reflecting stress-induced variability. However, PLSR captured yield variation more effectively, with all components showing positive contributions, although none were statistically significant, suggesting diffuse and less distinct trait–yield relationships under late sowing stress. Under Rainfed (RF) conditions, PCR indicated mixed effects, with negative contributions from Comp.2 (–86.20) and Comp.3 (–65.48), while PLSR revealed consistently positive component effects, particularly for Comp.1 (112.89) and Comp.3 (69.51). This highlights that PLSR effectively captures complex interactions between traits and environmental factors under moisture stress, where traditional methods struggle. Another key observation is that PCR includes higher-order unstable components (Comp.7 and Comp.8) with large standard errors, particularly under NS, indicating noise and overfitting. In contrast, PLSR uses fewer, more stable components with smaller standard errors, making it more reliable.

Table 4 : Relationship of chickpea yield with biometrical and weather variables using PCR and PLSR under NS, LS, and RF conditions

Intercept	1874.33 (39.44)	**1874.33** (5.92)	1962.59 (28.99)	1962.59 (0.85)	2046 (28.69)	2046 (2.42)
Comp.1	18.40 (18.80)	**195.93** (3.43)**	13.11 (13.69)	102.72 (0.48)	21.83 (13.11)	112.89 (1.42)
Comp.2	-170.84 (25.26)	**75.42** (4.68)**	-81.48 (17.45)	41.37 (0.63)	-86.20 (19.77)	36.08 (1.35)
Comp.3	-76.45 (30.41)	**70.02** (4.57)**	-59.42 (22.54)	36.29 (0.65)	-65.48 (21.31)	69.51 (2.71)
Comp.4	67.72 (40.37)	**51.91** (5.79)**	50.47 (29.24)	53.47 (0.95)	49.99 (29.54)	33.38 (2.58)
Comp.5	14.19 (42.05)	**47.20** (7.74)**	9.04 (36.93)	27.82 (0.94)	16.08 (30.75)	27.55 (2.79)
Comp.6	112.91 (56.56)	51.88 (9.91)	10.48 (45.13)	20.87 (1.45)	-72.44 (40.61)	29.31 (3.75)
Comp.7	112.07 (65.90)	–	–	–	–	–
Comp.8	-255.35 (87.30)	–	–	–	–	–
R^2	**0.81**	**0.99**	**0.65**	**0.99**	**0.65**	**0.99**

Table 5 : Comparison of Predictive Performance of Different Regression Models for Chickpea Yield under Normal Sown, Late Sown and Rainfed Conditions.

Sowing Condition	**Model**	**RMSE (kg ha^{-1})**	**Prediction Accuracy**
Normal Sown (NS)	MLR	147.51	Moderate
	Stepwise	157.35	Moderate
	PCR	162.63	Moderate
	PLSR	26.49	Very High
Late Sown (LS)	MLR	108.50	High
	Stepwise	111.41	High
	PCR	120.46	Moderate
	PLSR	3.80	Excellent
Rainfed (RF)	MLR	107.33	High
	Stepwise	113.65	High
	PCR	127.63	Moderate
	PLSR	10.83	Very High

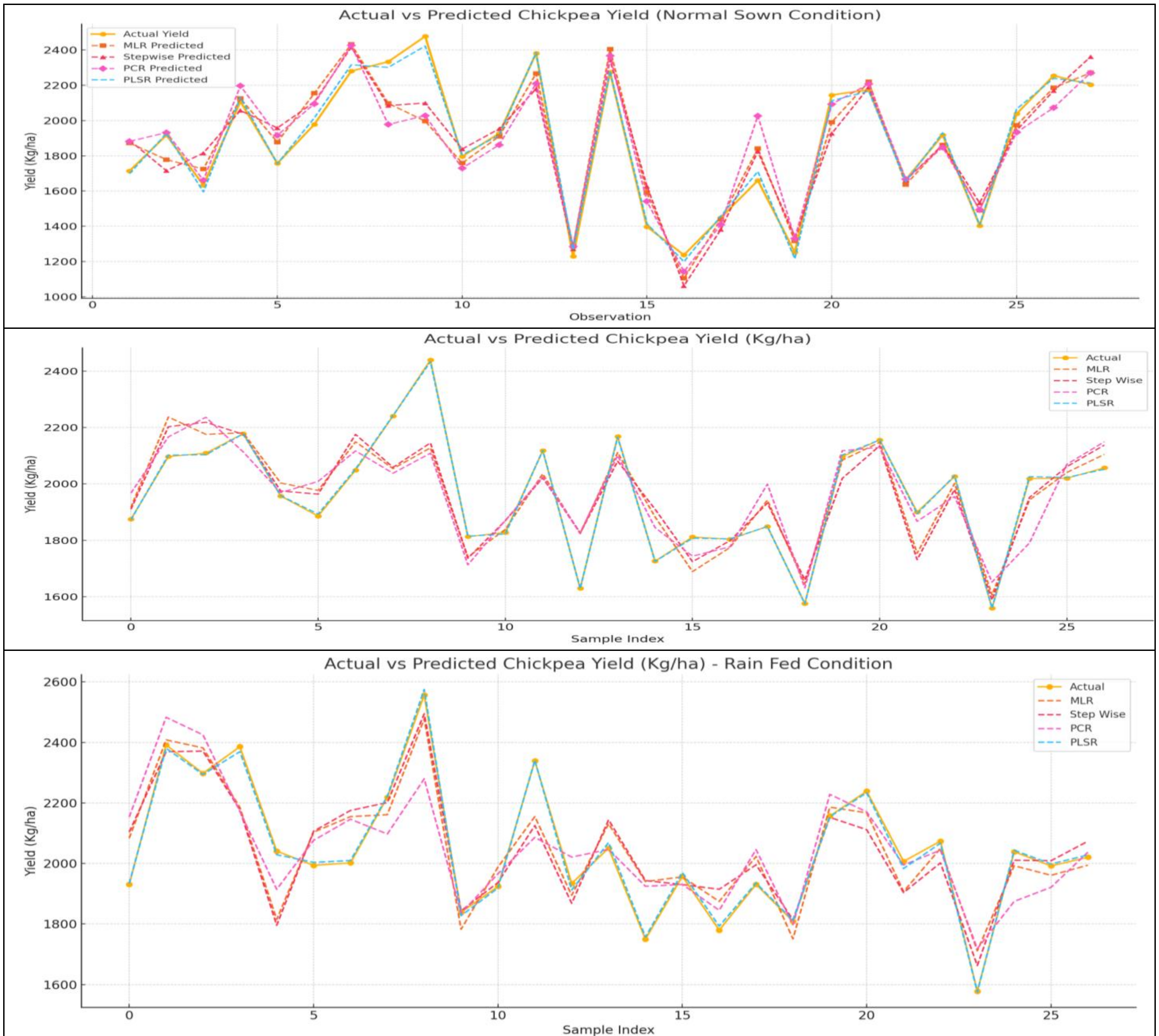

Fig. 1 : Actual Vs Predicted value of Chickpea yield under Normal own , Late sown and Rainfed Condition

From table-5 it is observed that across all sowing environments, PLSR consistently outperformed MLR, Stepwise, and PCR models, as indicated by substantially lower RMSE values. Under normal sown conditions, PLSR achieved an RMSE of 26.49 kg ha^{-1}, which is nearly 6 times lower than MLR and PCR, demonstrating its superior ability to handle multicollinearity among predictors. In late sown conditions, PLSR showed exceptional predictive accuracy with an RMSE of only 3.80 kg ha^{-1}, indicating almost perfect agreement between predicted and actual yields. Notably, under rainfed conditions, PLSR again performed best (RMSE = 10.83 kg ha^{-1}), confirming its robustness in capturing complex trait–environment interactions under moisture stress. Although MLR and Stepwise regression showed reasonably good performance under rainfed and late sown conditions (RMSE ≈ 107–113 kg ha^{-1}), their prediction errors were substantially higher than those of PLSR. PCR consistently exhibited the highest RMSE values across environments, indicating loss of predictive information during dimensional reduction. The study revealed that multiple regression (MR) showed high explanatory power (R^2 up to 0.84) but suffered from severe multicollinearity, particularly among climatic variables, reducing reliability. Stepwise regression (SR) improved model stability by selecting key predictors such as number of pods per plant, plant height, and maturity traits, while maintaining comparable accuracy ($R^2 \approx 0.70$–0.82). The actual and predicted value of Chickpea yield under Normal own , Late sown and Rainfed Condition were shown in figure. 3.1. It is studied that imilar findings were reported by Kumar et al. (2023) and Verma and

Sharma (2023), who highlighted the usefulness of variable selection in improving prediction. Principal Component Regression (PCR) reduced multicollinearity by transforming variables into components; however, its moderate performance (R^2 = 0.65–0.81) and unstable higher components indicate loss of predictive information, as also observed by Choudhary and Singh (2023) and Wang et al. (2024). In contrast, Partial Least Squares Regression (PLSR) performed best across all conditions, with very high accuracy ($R^2 \approx 0.99$) and lowest RMSE values, effectively capturing complex trait–environment interactions. This superiority of PLSR over other models aligns with findings of Yoon et al. (2023). Notably, rainfed conditions showed stable and efficient model performance, emphasizing the importance of adaptive traits in yield determination. Overall, PLSR emerged as the most robust and reliable method for chickpea yield prediction, followed by SR, while MR and PCR showed limitations under multicollinearity.

Conclusion: This study is novel in integrating multiple advanced regression approaches to evaluate chickpea yield across diverse sowing conditions, demonstrating the superiority of PLSR in handling multicollinearity and revealing the strong potential of rainfed systems, while identifying key biometrical traits for stable yield improvement. Overall, the pooled analysis clearly establishes PLSR as the most reliable and accurate model for chickpea yield prediction, particularly under rainfed conditions, where yield variability is high and trait–environment relationships are complex.

References

Choudhary, M., & Singh, D. (2023). Use of principal component analysis and regression techniques for yield prediction in chickpea. *Electronic Journal of Plant Breeding*.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). Wiley.

Efroymson, M. A. (1960). Multiple regression analysis. In A. Ralston & H. S. Wilf (Eds.), *Mathematical methods for digital computers* (pp. 191–203). Wiley.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer.

Kumar, A., Singh, R., & Yadav, R. (2023). Correlation and regression studies of yield attributing traits in chickpea under agroforestry system. *The Pharma Innovation Journal*, **12**(10), 813–817.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.

Verma, P., & Sharma, S. (2023). Path coefficient and regression analysis for yield improvement in chickpea under different environments. *Legume Research*.

Wang, X., Liu, Y., Zhang, H., & Chen, Z. (2024). Comparison of principal component regression and partial least squares regression models using near-infrared spectroscopy data. *Journal of Applied Statistics*. https://doi.org/10.1080/02664763.2024.xxxxxx

Yoon, G., Kim, H., Lee, S., & Park, J. (2023). Application of partial least squares regression and machine learning for plant metabolite prediction. *Agriculture*, **13**(8), 1477. https://doi.org/10.3390/agriculture13081477