RESEARCH ARTICLE Legume Research- An International Journal, Volume 48 Issue 6: 966-974 (June 2025)

Prediction of Faba bean (*Vicia faba* L.) Yield under Artificial Shade Constraints in Moroccan Agro-fields

Ghita Ajana[1,2], Salah-Eddine Laasli[3], Jamila El Figuigui[2], Meryem Benjelloun[2], Zain El Abidine Fatemi[1], Khalid Daoui[1]

10.18805/LRF-842

ABSTRACT

Background: Statistical regression models represent alternatives to process-based dynamic models for predicting the response of crop yields to climatic variations. This study chose stepwise and Lasso regression methods to predict Faba bean (*Vicia faba* L.) yield under different shading conditions in Moroccan agroecosystems.

Methods: Crop-related data were collected in six faba bean varieties that were widely grown in full sun (S1) and two shade treatments S2 (50%) and S3 (90%) two growing seasons. The same data set was used for both regression model development and validation.

Result: The results show that models based on the Lasso method were more accurate and precise (R-squared = 0.95; Root Mean Square Error (RMSE), = 102.6; Mean Absolute Percentage Error (MAPE). = 3.28%) than the stepwise approach (R^2overall = 0.86; RMSE overall = 214.8; MAPE overall = 6.88%) in terms of both shading levels and growing seasons. Positive correlations were spotted between yield and NP (number of pods per plant), NS (number of seeds per plant) and WSS (100-seed weight per plant) parameters in the best shading levels (S2 and S3). These parameters significantly contributed to the Faba bean yield prediction. These findings advance the field by highlighting the potential of integrating modern regression techniques into breeding programs, offering a replicable framework for other crops and conditions.

Key words: Faba bean, Lasso regression, Shading, Stepwise regression, Yield prediction.

INTRODUCTION

Faba bean (*Vicia faba* L.) is an important legume crop that constitutes one of the best sources of protein for human and animal nutrition. In consideration of its valuable nutritional qualities and climate adaptability, this important component of sustainable systems provides resilience in agricultural strategies and reformulation of diet (Nurmansyah *et al.,* 2024). This crop occupies a significant place in intercropping systems (ICS) in Morocco. The most common combination for this farming practice is legumes/cereals (Tamta *et al.,* 2019) or legumes/wheat (Chamkhi *et al.,* 2022). Intercropping leguminous crops is a promising way of enriching soils with biologically-fixed nitrogen, due to the legume-rhizobia symbiosis (Raji and Dörsch, 2020). However, in ICS, shading represents a potential constraint. Its effect, due to reduced light intensity, leads to changes in morphology, physiology, biomass and grain yield (Arenas Corraliza *et al.* 2020).

In a physiological context, a study conducted by Hu *et al.* (2023a) indicated that shading up to 60% caused a prominent decrease in SPAD entities in maize, leading to a considerable reduction in photosynthetic rate. Wang *et al.* (2021b) also reported that 50% shading notably increased chlorophyll contents (a, b and total) in peanut leaves. Nevertheless, chlorophyll levels under shaded conditions were still lower than those in unshaded conditions. Le soja [*Glycine max* (L.) Merr.] cultivé à l'ombre présentait des niveaux de proline plus élevés que ceux cultivés en plein soleil, ce qui indique que les plantes à l'ombre protégeaient mieux leurs cellules des dommages (Muhammad *et al.,* 2021).

[1]Agronomy and Plant Physiology Research Unit. Center Regional of Agricultural Research of Meknès. Km10, Road Haj Kaddour, BP S/40, Meknès 50001, Morocco.
[2]Laboratory of Functional Ecology and Environmental Engineering, Sidi Mohamed Ben Abdellah University, Faculty of Science and Technology. P.O. Box 2202, Road Imouzzer, Fez 30500, Morocco.
[3]Department of Plant and Environment Protection, Plant Protection Unit, National School of Agriculture, km. 10, Route Haj Kaddour, B.P. S/40, Meknès, 50001, Morocco.

Corresponding Authors: Ghita Ajana; Khalid Daoui, Agronomy and Plant Physiology Research Unit. Center Regional of Agricultural Research of Meknès. Km10, Road Haj Kaddour, BP S/40, Meknès 50001, Morocco; Laboratory of Functional Ecology and Environmental Engineering, Sidi Mohamed Ben Abdellah University, Faculty of Science and Technology. P.O. Box 2202, Road Imouzzer, Fez 30500, Morocco. Emails: ghita.ajana@usmba.ac.ma; khalid.daoui@inra.ma

How to cite this article: Ajana, G., Laasli, S.E., Figuigui, J.E., Benjelloun, M., Fatemi, Z.E.A. and Daoui, K. (2025). Prediction of Faba bean (*Vicia faba* L.) Yield under Artificial Shade Constraints in Moroccan Agro-fields. Legume Research. 48(6): 966-974. doi: 10.18805/LRF-842.

Submitted: 08-11-2024 **Accepted:** 08-03-2025 **Online:** 24-04-2025

According to Hu *et al.* (2023b), insufficient light reduced kernel number, filling rate and weight of summer maize kernels. In addition, several studies have shown that light interception by artificial shade structures at different stages during the breeding period reduces soybean yield, mainly by influencing pod number. These results suggest that yield is sensitive to shading during flowering and pod

formation (Rivest *et al.,* 2009). For this purpose, yield prediction under this stress is imperative, as it estimates the faba bean's economic productivity. Pre-harvest yield predictions are important for mitigating the effects of natural catastrophes and ensuring food security (Htun *et al.* 2023). Yield prediction using vegetation indices has been evaluated in some studies on wheat, maize and durum (Mohammadi *et al.,* 2024).

Many techniques have been developed for the selection of predictor variables such as Least absolute shrinkage and selection operator regression (Lasso) (Kumar *et al.* 2021). Lasso improves prediction quality by reducing the regression coefficient compared with prediction models fitted by non-penalized maximum likelihood methods (Gupta *et al.,* 2024). Tibshirani (1996) proposed Lasso, which can be used in the crop yield prediction technique. Lasso minimizes the sum of residual squares provided that the sum of the absolute value of the coefficient is less than a constant (Aravind *et al.,* 2022; Kumar *et al.,* 2021) reported that elastic Net and Lasso proved to be the best models for wheat yield prediction in different localities of northwest India. Statistical methods such as stepwise regression and correlation are also frequently used in crop yield prediction (Das *et al.,* 2018; Demirhan, 2024; Iniyan *et al.,* 2023). Correlation studies are useful for measuring the strength and direction of relationships between different traits and grain yield (Zarei *et al.,* 2013). Correlation is useful for revealing the magnitude and direction of the relationship between several yield component traits and grain yield (Islam *et al.,* 2020). Stepwise regression is a statistical method for analyzing the relationship between a unique response variable or so-called dependent variable (*i.e.* crop yield) and independent variables or so-called controlled variables (Pourshoaib *et al.,* 2022). According to (Saraka and Kouassi, 2019), this method was used to assess the relative contributions of each yield component to yield by eliminating the effect of non-effective characteristics in the yield regression model.

Thus, yield prediction models can rapidly identify the most resistant genotypes (Skobalski *et al.* 2024). Faba bean breeding programs aim to develop widely adaptable, disease-resistant, high-yielding genetic material (Mohammadi *et al.,* 2024). The Moroccan faba bean breeding program has been in operation for more than 37 years, playing a very important role in the improvement of seed yield and quality. Up to now, 25 major faba bean varieties and 10 minor faba bean varieties have been registered in the official Moroccan catalog, all of which are exclusively inbred lines (Chetto *et al.,* 2023). Variable selection in a multiple regression model plays an important role in prediction precision because (i) it makes the model easier to interpret, by removing redundant variables and adding no information (ii) it reduces the size of the problem to enable the algorithms to operate quicker, making it possible to handle high-dimensional data and (iii) it reduces over-fitting in the model.

Based on this previous work, the present study had a dual objective: (1) To determine if the yield of faba bean under shade stress can be predicted from these components. (2) To identify the yield components that correlate most strongly with yield. By identifying the most robust and easily measurable yield components. (3) Develop yield prediction models for *V. faba* using stepwise regression and Lasso regression. (4) Evaluate the accuracy of the field yield prediction models and select the best model for predicting bean yield under different shading treatments.

Moroccan plant breeding programs can employ indirect selection criteria to develop shade-tolerant varieties. These developed predictive models will contribute practically to the improvement of farming strategies in Morocco by showing which factors are important contributors to yield. These insights can guide farmers and agricultural planners in optimizing resource allocation, selecting suitable crop varieties and tailoring agronomic practices to local environmental conditions. This approach will equip breeders with valuable insights, enabling them to accurately identify and measure these varieties using precise analytical methods.

MATERIALS AND METHODS

The trials were carried out at the National Institute of Agricultural Research (INRA) experimental station in Douyet, Morocco, over two consecutive growing seasons: 2020-2021 (Y1) and 2021-2022 (Y2). The site, located at 34°04'N latitude and 5°07'W longitude, lies in the favorable agricultural zone of the Saïs plain, within the Province of Moulay Yaacoub, Fez-Meknes region, at an altitude of 416 meters. The soil is silty clay, consisting of 48.50% silt, 39.90% clay and 11.60% fine sand, predominantly dark Vertisols. Its chemical properties include a pH of 7.80, organic matter content of 3.63 mg/kg, 11.89 mg/kg of available P_2O_5 and 478.05 mg/kg of available K_2O. The region has a Mediterranean climate, with a dry season from May to October. Manual seeding was carried out on Y1 and Y2, respectively, where the area had been previously prepared by deep 3-disc plowing. Before this experiment, the land had been rotated with cereals. Then, fertilizers were applied and spread to improve soil quality, such as the base fertilizer NPK ″10:30:10″ (200 kg/ha). Plant-related biotic agents were intensively treated using commercialized pesticides (for pests and diseases) and cultural methods (*e.g.*, mechanical hoeing for weeds) to prevent yield loss. Irrigation was administered only when plants were at risk of perishing. Supplemental irrigation of 30 mm was applied once after plant emergence.

The experiment was arranged in a split-plot design with two replications. The main plots were assigned to three shade treatments (S1, S2 and S3), while six Faba bean varieties were assigned to the subplots. Each elementary plot measured 21.6 m × 4 m, comprising six rows per variety (36 rows per plot) with 0.60 m inter-row spacing. A 3-meter-wide alley separated the replicates and

treatments. Each plot covered 86.4 m^2, with a total experimental area of 811.8 m^2. The experiment was conducted on different plots within the station in each growing season. Subplots receiving the shading treatment were covered with Aluminet shading nets, made from high-density polyethylene with a metalized layer for added durability. The nets were stretched over wooden supports with metal wires and placed 1 meter above the ground. These structures shaded the plants from the top and sides, leaving the south side open for air circulation. The shading was designed to reduce the intensity of photosynthetically active radiation (PAR) by 50% and 90%, using a ZDS-10 Luxmeter (China), resulting in three treatments: no shading (S1), 50% shading (S2) and 90% shading (S3). These shading treatments were chosen based on another pot experiment conducted under olive trees' shading conditions. Sunlight reduction was measured in about 650 Lux and 115 Lux corresponding to the S1 and S3 shade net respectively. Shading was applied from flowering to maturity, starting 81 days after seeding in Y1 and 127 days after seeding in Y2. The delay in Y2 was due to harsher weather conditions compared to Y1 (*e.g.,* temperature increase of 1.7°C and precipitation decrease of 6.87 mm), which slowed Faba bean growth and delayed flowering. Yield per plant (YP) and its components included plant height, NTS (number of total stems per plant), NP (number of pods per plant), NS (number of seeds per plant) and WSS (100-seed weight per plant). To confirm the consistency of shade treatments, additional environmental variables (temperature, relative humidity and precipitations) were monitored throughout the experiment using specialized software (FieldClimate, METOS® by Pessl Instruments, Austria).

Stepwise regression is a combination of forward selection and backward elimination techniques. The model is built iteratively by adding significant variables and removing less significant ones based on predefined criteria such as p-value or Akaike Information Criterion (AIC). This regression was employed as the first modeling approach to determine the most significant variables contributing to yield prediction. The method used in this study was backward elimination, starting with a full model that included all predictors (dummy variables for shading and variety and the year variable). In each iteration, non-significant predictors (based on a significance level of 0.05) were removed, one by one, until a final model with only significant predictors was reached. The mathematical representation of the Stepwise Regression model is given as.

$$Yi = \beta_0 + \beta_1 X_{1i} + \beta_1 X_{2i} + ... + \beta_n X_{ni} + \varepsilon_i$$

Where,

Y_i = Predicted yield (kg/ha).

β_0 = Intercept.

β_n= Regression coefficients for each predictor (NP, NS and WSS).

X_{ni} = Represents the dummy variables for shading (S1, S2 and S3) and variety (V1 to V6), as well as the numeric variable for the year (Y1 and Y2).

ε = Error term.

Lasso regression is a regularization method that performs both variable selection and regularization. It minimizes the residual sum of squares subject to the sum of the absolute values of the coefficients being less than a constant. Unlike stepwise regression, Lasso automatically shrinks less important coefficients to zero, thus retaining only the most impactful variables. The lasso regression model is expressed as:

$$\min_{\beta_0, \beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (Yi - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

Where,

λ = Regularization parameter controlling the strength of the penalty.

The rest of the terms are the same as in the stepwise regression model.

To evaluate the performance of both the Stepwise and Lasso regression models, three key metrics were calculated: R-squared (R^2), root mean square error (RMSE) and mean absolute percentage error (MAPE). These metrics provided a comprehensive assessment of model accuracy and goodness of fit. The R-squared (R^2) statistic measures the proportion of variance in the observed yield data explained by the model was also calculated.

The root mean square error (RMSE) measures the square root of the average squared differences between the observed and predicted values and was calculated as.

$$RMSE = \sqrt{\frac{1}{n} \sum (Y_{obs} - Y_{pred})^2}$$

Where,

n = Number of observations.

RMSE provides an absolute measure of prediction error, with smaller values indicating better model performance.

The mean absolute percentage error (MAPE) expresses prediction accuracy as a percentage and was calculated as.

$$MAPE = \frac{1}{n} \sum \left| \frac{Y_{obs} - Y_{pred}}{Y_{obs}} \right| \times 100$$

All statistical analyses were performed using Python (version 3.11) programming language. The stats models package was used for implementing the Stepwise Regression model, utilizing the backward elimination approach to automatically select the most significant variables. The scikit-learn library was employed for the Lasso Regression model, with the LassoCV function used to perform cross-validation and determine the optimal regularization parameter λ. Data preprocessing, including the conversion of categorical variables such as shading

and variety into dummy variables, was handled using the Pandas library. For data visualization, Matplotlib and Seaborn libraries were employed to generate scatter plots that compare the predicted yields against observed yields for both regression techniques. These plots were further broken down by shading treatment to highlight differences in model performance across the different environmental conditions. Correlation analyses (Pearson type) were established using the Pandas library to reveal the relationships between yield parameters recorded in best shading conditions (S2 and S3) applied to different Faba bean varieties. In addition, a redundancy analysis (RDA) was conducted using *skbio.stats.ordination.rda* module to determine which predicting variables (yield parameters) were contributing the most to the precise yield prediction. All Statistical tests were established based on replicated treatments (n = 4) at a significance level of 5% (P<0.05).

RESULTS AND DISCUSSION

Faba bean yield forecast under shading conditions

Stepwise regression was accordingly implemented to assess the yield prediction potential of the Faba bean under the three shading conditions (Table 1). The predicted yield was essentially based on NP, NS and WSS parameters, which had significant trends in all shading levels (Table 1). The distribution of data points (crosses) across the plot follows a diagonal trend, indicating a close alignment between the predicted and actual yields. The model's performance was strong and accurate across the studied entities (R^2 = 0.84-0.88; $R^2_{overall}$ = 0.86; P < 0.001) (Table 1). Furthermore, an average de*via*tion of predicted yields from the actual yields was observed as most of the predicted entities were below the observed ones ($RMSE_{overall}$ = 214.8). The model had an average prediction error being less than 7% ($MAPE_{overall}$ = 6.88%) reflecting its precision. Notably, the differences in the slopes of the S1, S2 and S3 levels suggest that the model captures the varying effects of shading conditions on yield, with S3 generally yielding higher predictions.

The lasso regression was applied to optimize the yield prediction aspect of the studied shading levels (Table 2). Similar to the previous model, yield predictions were attributed to NP, NS and WSS parameters, with statistical significance observed for all treatments (P <0.0001) (Table 2). The data points exhibit a strong diagonal alignment, signifying an excellent agreement between predicted and actual yields. The model's performance demonstrated better accuracy and reliability across the shading levels, with R^2 values ranging from 0.93 to 0.96, culminating in an

Table 1: Stepwise regression model proprieties for faba bean yield prediction under different shading conditions.

Shade	Explanatory variables	Regression value (b)	Coefficient SE	t - value	P - value	Intercept	R^2
S1	NP	0.158779	0.089	8.334	0.0001	21.14	0.84
	NS	0.81277	0.046	17.824	0.0001		
	WSS	25.386244	1.595	15.918	0.0001		
Model		Yield = 21.140073 + 0.158779 NP + 0.812770 NS + 25.386244 WSS					
S2	NP	0.55677	0.159	3.509	0.0001	9.823	0.88
	NS	0.871396	0.064	13.591	0.0001		
	WSS	15.105639	0.102	13.704	0.0001		
Model		Yield = 9.823906 + 0.556770 NP + 0.871396 NS + 15.105639 WSS					
S3	NP	0.33944	0.11	6.112	0.0001	2.822	0.838
	NS	0.568543	0.029	19.879	0.0001		
	WSS	6.749858	0.82	8.236	0.0001		
Model		Yield = 2.822970 + 0.339440 NP + 0.568543 NS + 6.749858 WSS					

Table 2: Lasso regression model proprieties for faba bean yield prediction under different shading conditions.

Shade	Explanatory variables	Regression value (b)	Coefficient SE	t - value	P - value	Intercept	R^2
S1	NP	0.13374	0.254	6.354	0.0001	5.884	0.94
	NS	1.6559	0.566	15.248	0.0001		
	WSS	33.665	2.397	12.889	0.0001		
Model		Yield = 5.884123 + 0.13374 NP + 1.6559 NS + 33.665 WSS					
S2	NP	0.4998	0.477	3.176	0.0001	7.974	0.938
	NS	1.9661	0.321	10.274	0.0001		
	WSS	18.441	0.579	11.987	0.0001		
Model		Yield = 7.9741125 + 0.4998 NP + 1.9661 NS + 18.441 WSS					
S3	NP	0.657	0.381	4.223	0.0001	1.647	0.958
	NS	1.8874	0.166	17.421	0.0001		
	WSS	11.398	1.976	6.547	0.0001		
Model		Yield = 1.647471 + 0.6570 NP + 0.1.8874 NS + 11.3980 WSS					

$R^2_{overall}$ of 0.95 (P<0.001) (Table 2). The predicted yields show a highly minimal de*via*tion from actual values, as reflected in the much lower RMSE overall value (102.6) compared to the stepwise model (Table 3). The lasso model's prediction error was less than 3.5% ($MAPE_{overall}$ = 3.28%), highlighting the model's enhanced precision. The distinct slopes observed across shading levels (S1, S2, S3) suggest that the model accurately captures the nuanced impact of shading on yield, with S3 again yielding higher predictions overall.

The findings of this study offer significant insights into the impact of artificial shading on bean yield and highlight key factors influencing it. Stepwise and Lasso regression analyses demonstrated that NP, NS and WSS are crucial predictors of *V. faba* yield across all shading treatments. The exclusion of variables such as plant height, NTS and NP from the final models in specific treatments suggests that their influence may be limited under certain shading conditions. This could be attributed to the diminished relevance of these traits in altered light environments, as highlighted by recent research showing a significant reduction of plant height and stem characteristics of tropical tree species under high shading (Cifuentes and Moreno, 2022). In addition, our results align with those reported by (Chan Fu Wei and Molin, 2020), who found that models based on seed number effectively predict soybean yield with high accuracy (R^2 = 0.70). Furthermore, our findings were in line with other studies emphasizing the importance of seed number and seed weight in yield prediction (Wu *et al.,* 2017; Zhang *et al.,* 2023). Liu *et al.* (2021) similarly observed that in high shading conditions, both seed number and seed weight are pivotal for accurate yield predictions, reinforcing the reliability of these predictors across varying shading scenarios.

The high R^2 values achieved in stepwise analysis demonstrate the model's robustness in explaining the variability in faba bean yield. In this case, the decrease in R^2 value from S2 to S3 reflects the increasing complexity of predicting yield under higher shading intensities. It is often explained by the increased variability and stress of plants at a higher level of shading, as will be assumed. This assumption is also strengthened by some recent studies that prove the trend of lowered predictive accuracy over specific variables with shading intensification (Pratzer *et al.* 2023). Unaccounted yield variability ranged from 11.2% to 16.2% with stepwise regression and from 4.2% to 6.2% with Lasso. This would prove that multiple factors, other than variables influence the yield. Further research work has to be directed towards localizing, identifying and quantifying those factors responsible for this variability due to environmental, genetic diversity and management practices.

Faba bean yield prediction models were validated to evaluate the performance of both stepwise and Lasso regression aspects during Y1 and Y2. Results showed that the prediction error with the Lasso approach was 2.38 and 2.09, respectively which was less than the ones obtained with the stepwise approach (Table 4), which implied the high prediction accuracy of the Lasso model. In this study, the faba bean yield prediction model at the field scale developed by the Lasso method showed better performance and accuracy than that developed by the stepwise method. In a study conducted by Sharif *et al.* (2017) for modeling winter oilseed rape yield, Lasso was among the most accurate regression techniques among several others. This may be attributed to its ability to exclude variables that do not significantly contribute to the final yield. For such scenarios, where the true values of corresponding coefficients are likely near zero, regression techniques with variable selection capabilities often outperform traditional methods (Hastie *et al.,* 2009). However, stepwise regression was less performant in this study, possibly due to the inaccurate contribution of explanatory variables. Stepwise regression is primarily designed for low-dimensional problems. When the number of observations is not substantially larger than the number of explanatory variables, high variance and overfitting can compromise the predictive power of classical approaches. These results were in line with previous studies on regional yield prediction (Cai *et al.,* 2019; Cao *et al.,* 2021; Liu *et al.,* 2022). Kumar *et al.,* (2021) compared the performance of stepwise and Lasso regression models in wheat (*Triticum aestivum* L.) yield prediction using meteorological data. Their results demonstrated that Lasso regression was superior to stepwise regression in reducing data dimensionality.

The superior performance of the Lasso regression method highlights its potential as a valuable tool in precision agriculture, where high-dimensional datasets are common (Miller *et al.,* 2022; Pukrongta *et al.,* 2024). Accurate yield prediction can guide resource allocation, inform decision-making and support agricultural policies aimed at ensuring food security (Raihan, 2024). For instance,

Table 3: Comparison of stepwise and lasso regression parameters obtained for faba bean yield prediction.

Parameter	Stepwise regression	Lasso regression
R^2	0.86	0.95
RMSE	214.8	102.6
MAPE	6.88%	3.28%

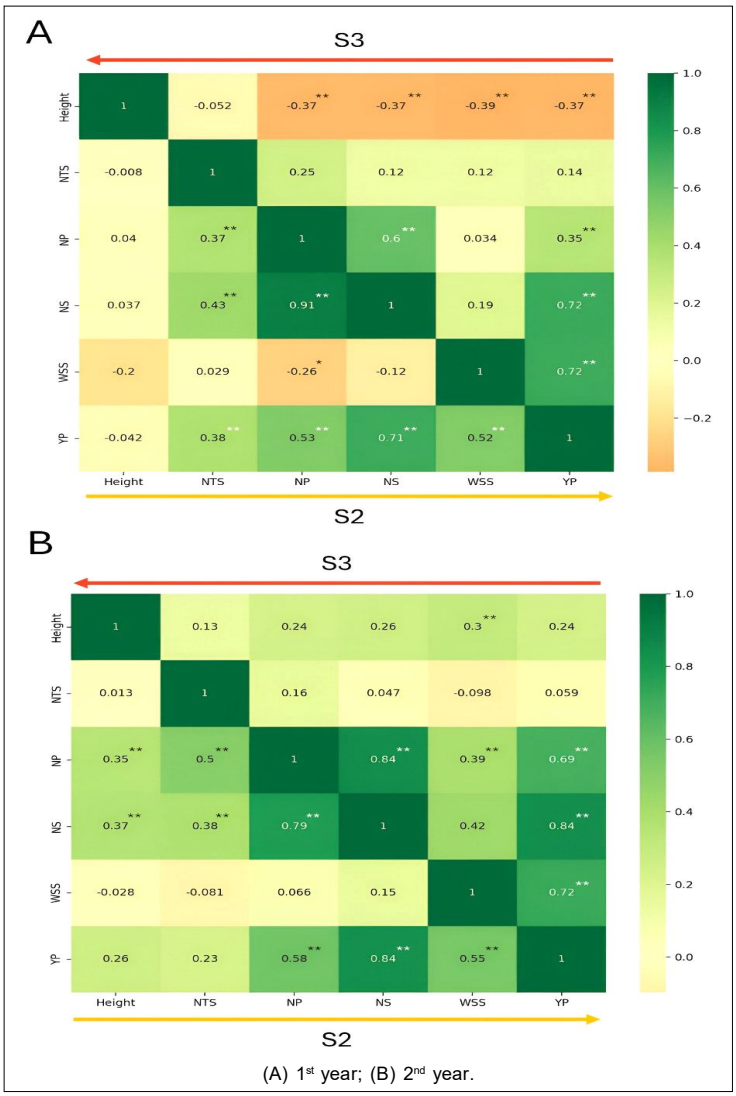
Table 4: Validation of stepwise and Lasso regression models for faba bean yield prediction during the 2020-2021 and 2021-2022 seasons.

Year	Observed yield (OY)	Stepwise regression		Lasso regression	
		Predicted yield (PY)	Error (%)	Predicted yield (PY)	Error (%)
Y1	1270	1155	6.77	1277	2.38
Y2	891	876	18.5	896	2.09

the ability to identify critical predictors from environmental, agronomic and management factors enhances the development of targeted interventions to optimize yield under variable conditions (Kumar *et al.,* 2015). Moreover, the implications extend to sustainable agriculture practices. Farmers and stakeholders can plan more efficiently, minimize input wastage and mitigate risks associated with adverse environmental conditions by accurately predicting their crop yield.

Relationships of faba bean yield prediction parameters under shading conditions

To delve deeper into the predictive yield dynamics, Pearson correlation analyses were accordingly conducted in the best possible shading conditions (S2 and S3) (Fig 1). Results revealed that under these conditions in Y1, Faba bean yield (YP) was positive and significantly correlated with NP, NS and WSS parameters (r = 0.35-0.72; $P < 0.001$), indicating that increasing these traits would increase Faba bean yield. In contrast, significant negative correlations were spotted between plant height and the YP (r = -0.39 to -0.37); $P < 0.001$) (Fig 1A). Furthermore, the transition from S2 to S3 levels further amplified these correlative effects. Regarding the second experimental year (Y2), the variables NP, NS and WSS still showed a significant positive correlation with YP in the S2 level, while a transitional shift was noticed into positive correlations between plant height and YP unlike Y1 (Fig 1B). The significant positive correlations observed between YP and NP, NS and WSS parameters in both Y1 and Y2 highlight the importance of these traits in determining yield. The increase in grain yield was attributed to an increase in the number of pods and seeds due to greater branching and more pods on the branches (Alharbi and Adhikari, 2020). The negative correlation between plant height and YP in Y1, particularly under S2 and S3 conditions suggests that taller plants



(A) 1st year; (B) 2nd year.

Fig 1: Correlation maps (Pearson type) displaying the relationships between Faba bean predictive yield components under best shading conditions (S2 and S3).

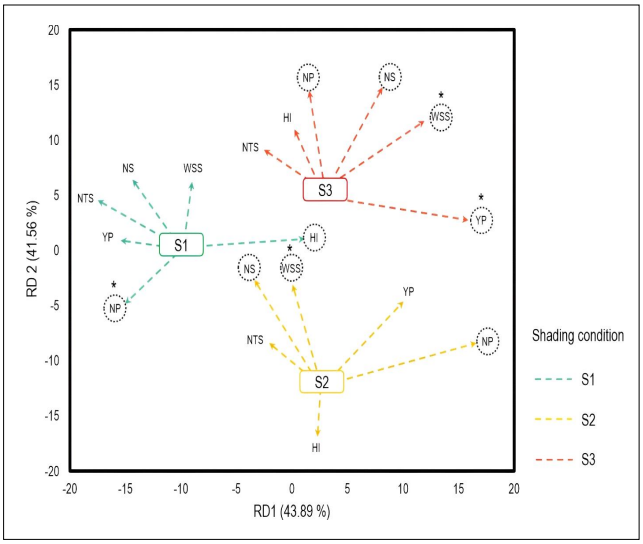


Fig 2: Redundancy Analysis (RDA) showcasing the overall dependencies of the shading conditions to the different studied variables for faba bean yield prediction.

may not necessarily be more productive in Faba bean. While taller plants often have a larger leaf area, excessive height can lead to increased lodging and reduced light interception at the lower canopy levels. Our results are similar to those of (Wadan and Abd el Shafi, 2014), who showed a negative correlation between plant height of wheat and grain yield. However, for barley (*Hordeum vulgare* L.), grain yield was positively correlated with plant height under all shading treatments (25% and 30%) (Arenas Corraliza *et al.* 2020). The shift in the correlation between plant height and YP from negative in Y1 to positive in Y2 under S2 conditions suggests that the relationship between plant height and yield can be influenced by environmental factors and specific growing conditions. In Y2, taller plants may have been able to utilize the available resources more efficiently, leading to increased yield. However, further research is needed to elucidate the underlying mechanisms driving this change. The number of branches is highly variable in legumes, but is an essential determinant of grain yield (Alharbi and Adhikari, 2020), however, no correlation was reported between YG and NTS under the two shading treatments. Furthermore, the impact of these traits on yield becomes more pronounced under higher levels of shading throughout the seasons as indicated by our study. This could be attributed to the increased competition for resources, such as light and nutrients, under more shaded conditions.

The contribution of the most important Faba bean parameters in each shading condition was further explored using an RDA. The results revealed that the variance in shading composition could be largely explained by both axes, with 43.89% of the variance explained by RD1 and 41.56% by RD2 (Fig 2). The third shading level (S3) had 4 major contributors including NP, NS, WSS and YP with the two latter components having more significance overall. Three contributors were detected for S2 (with WSS significance), while only the NP parameter was shown to be significantly relevant to the S1 level among the two detected contributors (Fig 2). The results of the RDA offer insightful contributions to our understanding of how the Faba bean responds to varying shading conditions. Key parameters (NP, NS, WSS and YP) emerged as major contributors, with WSS and YP having greater significance. These results are consistent with the evidence showing that yield-related traits, such as seed weight and total yield, are highly sensitive to reductions in light, with plants adjusting their resource allocation to maximize reproductive success under suboptimal conditions. A study by Wang *et al.* (2021) has demonstrated that high shading often induces changes in photosynthate partitioning, favoring reproductive organs over vegetative growth. This could explain the prominence of yield-related traits in the S3 shading condition. The differentiation of parameter importance across shading levels points to an adaptive response by Fab beans to varying light conditions, a concept supported by recent studies on crop plasticity (Mínguez and Rubiales, 2021). The findings reinforce the importance of NP, NS and WSS in determining yield, particularly under higher levels of shading.

CONCLUSION

In this study, the stepwise and lasso regression models were applied to predict the overall yield of faba bean varieties under 3 shading conditions over two growing seasons. Compared with the stepwise regression method, the Lasso models had higher accuracy and precision in yield prediction based on three main components (NP, NS and WSS). These entities proved to have the strongest correlations with yield, demonstrating their robustness and practical measurement in overall model prediction ability. These results can be directly applied to Moroccan plant breeding programs, providing breeders with precise

information on indirect selection criteria. Ultimately, this approach will improve the efficiency and accuracy of breeding efforts aimed at improving faba bean productivity under challenging light conditions, thus contributing to the development of more resistant crop varieties. However, more research has to be established in adopting, testing and validating new yield prediction models (including the ones integrating machine learning approaches) in the field conditions. These advancements will play a pivotal role in developing resilient crop varieties tailored to the environmental constraints of Moroccan and global agroecosystems.

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES

Alharbi, N.H., Adhikari, K.N. (2020). Factors of yield determination in faba bean (*Vicia faba* L.). Crop and Pasture Science. 71(4): 305 321.

Aravind, K., Vashisth, A., Krishnan, P., Das, B. (2022). Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models. Journal of Agrometeorology. 24(1): 18-25.

Arenas Corraliza, M.G., López-Díaz, M.L., Rolo, V., Moreno, G. (2020). Wheat and barley cultivars show plant traits acclimation and increase grain yield under simulated shade in Mediterranean conditions. Journal of Agronomy and Crop Science. 207(1): 100-119.

Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., Peng, B. (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agricultural and Forest Meteorology. 274: 144-159.

Cao, J., Zhang, Z., Luo, Y., Zhang, L., Zhang, J., Li, Z., Tao, F. (2021). Wheat yield predictions at a county and field scale with deep learning, machine learning and google earth engine. European Journal of Agronomy. 123: 126204.

Chamkhi, I., Cheto, S., Geistlinger, J., Zeroual, Y., Kouisni, L., Bargaz, A., Ghoulam, C. (2022). Legume-based intercropping systems promote beneficial rhizobacterial community and crop yield under stressing conditions. Industrial Crops and Products. 183: 114958.

Chan Fu Wei, M., Molin, J. (2020). Soybean yield estimation and its components: A linear regression approach. Agriculture. 10: 348.

Chetto, O., Fatemi, Z.E.A., Nabloussi, A. (2023). Estimation of outcrossing rate of faba bean under natural moroccan conditions. Legume Research- An International Journal. 46(10): 1295-1301. doi: 10.18805/LRF-734.

Cifuentes, L., Moreno, F. (2022). Trait coordination at leaf level explains the resistance to excess light stress in shade-tolerant tropical tree species. Tree Physiology. 42(7): 1325-1336.

Das, B., Nair, B., Reddy, V.K., Venkatesh, P. (2018). Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. International Journal of Biometeorology. 62(10): 1809-1822.

Demirhan, H. (2024). A deep learning framework for prediction of crop yield in Australia under the impact of climate change. Information Processing in Agriculture.

Gupta, S., Vashisth, A., Krishnan, P., Lama, A., Shivprasad, S.A., IMD, M. (2024). Weather based wheat yield prediction using machine learning. MAUSAM. 75: 639 648.

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Tibshirani, R., Friedman, J. (2009). Linear methods for regression. The Elements of Statistical Learning: Data Mining, Inference and Prediction. pp 43-99.

Htun, A.M., Shamsuzzoha, M., Ahamed, T. (2023). Rice yield prediction model using normalized vegetation and water indices from Sentinel-2A satellite imagery datasets. Asia-Pacific Journal of Regional Science. 7(2): 491-519.

Hu, J., Yu, W., Liu, P., Zhao, B., Zhang, J., Ren, B. (2023). Responses of canopy functionality, crop growth and grain yield of summer maize to shading, waterlogging and their combination stress at different crop stages. European Journal of Agronomy. 144: 126761.

Iniyan, S., Akhil Varma, V., Teja Naidu, C. (2023). Crop yield prediction using machine learning techniques. Advances in Engineering Software. 175: 103326.

Islam, S., Nualsri, C., Soonsuwon, W. (2020). Correlation and path analysis of phenological traits of Thai upland rice genotypes. 133: 142.

Kumar, A., Pathak, R.K., Gupta, S.M., Gaur, V.S. and Pandey, D. (2015). Systems biology for smart crops and agricultural innovation: filling the gaps between genotype and phenotype for complex traits linked with robust agricultural productivity and sustainability. Omics: A Journal of Integrative Biology. 19(10): 581-601.

Kumar, S., Attri, S.D., Singh, K.K. (2021). Comparison of Lasso and stepwise regression technique for wheat yield prediction. Journal of Agrometeorology. 21: 188-192.

Liu, Y., Heuvelink, G.B.M., Bai, Z., He, P., Xu, X., Ding, W. and Huang, S. (2021). Analysis of spatio-temporal variation of crop yield in China using stepwise multiple linear regression. Field Crops Research. 264: 108098.

Liu, Y., Wang, S., Wang, X., Chen, B., Chen, J., Wang, J., Huang, M., Wang, Z., Ma, L., Wang, P., Amir, M. and Zhu, K. (2022). Exploring the superiority of solar-induced chlorophyll fluorescence data in predicting wheat yield using machine learning and deep learning methods. Computers and Electronics in Agriculture. 192: 106612.

Miller, A., Panneerselvam, J., Liu, L. (2022). A review of regression and classification techniques for analysis of common and rare variants and gene-environmental factors. Neurocomputing. 489: 466-485.

Mínguez, M.I., Rubiales, D. (2021). Faba Bean. In: Crop Physiology Case Histories for Major Crops. Academic Press. pp 452-481.

Mohammadi, S., Uhlen, A.K., Lillemo, M., Ergon, Å., Shafiee, S. (2024). Enhancing phenotyping efficiency in faba bean breeding: Integrating UAV imaging and machine learning. Precision Agriculture. 25(3): 1502-1528.

Muhammad, A.A., Jiang, H., Shui, Z., Cao, X., Huang, X., Imran, S., Ahmad, B. *et al.* (2021). Interactive effect of shade and PEG-induced osmotic stress on physiological responses of soybean seedlings. Journal of Integrative Agriculture. 20(9): 2382-2394.

Nurmansyah, Y.H. Dewir, S.S. Alghamdi., Dewir, Y.H., Alghamdi, S.S. (2024). Influences of thidiazuron concentrations and exposure duration on axillary shoot proliferation of faba bean (*Vicia faba* L.) genotypes. Legume Research. 47(11). doi: 10.18805/LRF-817.

Pourshoaib, S.J., Rajabzadeh Ghatrami, E., Shamekhi, M.A. (2022). Comparing ultrasonic- and microwave-assisted methods for extraction of phenolic compounds from Kabkab date seed (*Phoenix dactylifera* L.) and stepwise regression analysis of extracts antioxidant activity. Sustainable Chemistry and Pharmacy. 30: 100871.

Pratzer, M., Fernández-Llamazares, Á., Meyfroidt, P., Krueger, T., Baumann, M., Garnett, S.T. and Kuemmerle, T. (2023). Agricultural intensification, Indigenous stewardship and land sparing in tropical dry forests. Nature Sustainability. 6(6): 671-682.

Pukrongta, N., Taparugssanagorn, A., Sangpradit, K. (2024). Enhancing crop yield predictions with PEnsemble 4: IoT and ML-Driven for Precision Agriculture. Applied Sciences. 14(8): 3313.

Raihan, A. (2024). A systematic review of Geographic information systems (GIS) in agriculture for evidence-based decision making and sustainability. Global Sustainability Research. 3(1): 1-24.

Raji, S.G., Dörsch, P. (2020). Effect of legume intercropping on N_2O emissions and CH 4 uptake during maize production in the Great Rift Valley, Ethiopia. Biogeosciences. 17(2): 345-359.

Rivest, D., Cogliastro, A., Vanasse, A., Olivier, A. (2009). Production of soybean associated with different hybrid poplar clones in a tree-based intercropping system in southwestern Québec, Canada. Agriculture, Ecosystems and Environment. 131(1): 51-60.

Saraka, Y., Kouassi, M. (2019). Assessment of morphological variability, heritability, genetic advances, correlation and stepwise regression for some yield component traits in four inbred lines of *Citrullus Mucosospermus* (Fursa). East African Scholars J. Biotechnol Genet. 1(6). doi: 10.36349/EASJBG.2019.v01i06.005.

Sharif, B., Makowski, D., Plauborg, F., Olesen, J.E. (2017). Comparison of regression techniques to predict response of oilseed rape yield to variation in climatic conditions in Denmark. European Journal of Agronomy. 82: 11-20.

Skobalski, J., Sagan, V., Alifu, H., Al Akkad, O., Lopes, F.A., Grignola, F. (2024). Bridging the gap between crop breeding and GeoAI/: Soybean yield prediction from multispectral UAV images with transfer learning. ISPRS Journal of Photogrammetry and Remote Sensing. 210: 260-281.

Tamta, A., Kumar, D., Ram, H., Meena, R., Meena, V., Yadav, M.R. (2019). Productivity and profitability of legume-cereal forages under different planting ratio and nitrogen fertilization. Legume Research. 42: 102-107. doi: 10.18805/LR-3992.

Tibshirani, R. (1996). Regression shrinkage and selection *via* the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology. 58(1): 267-288.

Wadan, A., Abd el Shafi, M. (2014). Regression and path analysis in Egyptian bread wheat. Journal of Agri-Food and Applied Sciences. 2: 139-148.

Wang, Y., Huang, R., Zhou, Y. (2021). Effects of shading stress during the reproductive stages on photosynthetic physiology and yield characteristics of peanut (*Arachis hypogaea* Linn.). Journal of Integrative Agriculture. 20(5): 1250 1265.

Wu, L., Zhang, W., Ding, Y., Zhang, J., Cambula, E., Weng, F., Liu, Z.H., Ding, C.-Q., Tang, S., Chen, L., Wang, S., Li, G. (2017). Shading contributes to the reduction of stem mechanical strength by decreasing cell wall synthesis in Japonica rice (*Oryza sativa* L.). Frontiers in Plant Science. 8: 881. doi: 10.3389/fpls.2017.00881.

Zarei, L., Cheghamirza, K., Farshadfar, E. (2013). Evaluation of grain yield and some agronomic characters in durum wheat (*Triticum turgidum* L.) under rainfed conditions. Australian Journal of Crop Science. 7(5): 609-617.

Zhang, Z., Yang, W., Li, Y., Zhao, Q., Dong, Y. (2023). Wheat-faba bean intercropping can control *Fusarium* wilt in faba bean under F. commune and ferulic acid stress as revealed by histopathological analysis. Physiological and Molecular Plant Pathology. 124: 101965.