Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Models for Predicting Fresh Fruit Yield from Tree Morphometric Parameters in Terminalia chebula Retz. Selections under Subtropical Himalayan Conditions

Sourav Singh
Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu
Manjot Singh Sidhu
manjotsidhu58@gmail.com

Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu
Kamal Kishor Sood
Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu
Sushil Kumar Gupta
Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu
Stanzin Khenrab
University of Ladakh
Tsering Lanzes
Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu

Article

Keywords: Chi-square validation, fresh fruit yield, Terminalia chebula, yield prediction

Posted Date: October 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10594560/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Reliable, non-destructive estimation of fresh fruit yield in long-rotation medicinal-fruit trees is essential for harvest planning, fertiliser scheduling and progeny evaluation in agroforestry systems. The present study was conducted during 2023 and 2024 on a seven-year-old plantation of *Terminalia chebula* Retz. comprising four geographically distinct selections, namely Jammu (S_1), Kallar (S_2), Pragpur (S_3) and Palodi (S_4), established at 4 m × 6 m spacing at the Experimental Farm of the Division of Silviculture and Agroforestry, SKUAST-Jammu, Chatha. Six fruit-bearing trees per selection were sampled for tree height, collar diameter, crown spread, number of branches and fresh fruit yield, and Pearson's correlation coefficients were computed among traits across selection means to characterise the across-selection co-variation among morphometric variables and yield. Selection-specific stepwise multiple linear regression models were subsequently developed to predict fresh fruit yield (kg tree^{-1}) from the four morphometric predictors. The chi-square goodness-of-fit test was used to validate predicted against observed yields. The correlation analysis revealed that no single morphometric trait was uniformly associated with fresh fruit yield across the four selections: yield was most strongly correlated with number of branches (r = + 0.77) and was moderately negatively correlated with collar diameter (r = − 0.48), while collar diameter and number of branches themselves were strongly inversely correlated (r = − 0.89). The four selections yielded markedly different predictor sets. In the Jammu selection, collar diameter and number of branches jointly explained 83.5 per cent of the variation in fresh fruit yield (R^2 = 0.835; χ^2 = 2.64, non-significant), with collar diameter (β = 0.717) being more influential than number of branches (β = 0.354). In Kallar, tree height alone accounted for 40.7 per cent of yield variation (R^2 = 0.407). In Pragpur, collar diameter alone explained 36.0 per cent of yield variation (R^2 = 0.360). The Palodi selection produced the most parameter-rich model, in which crown spread (β = 0.879), tree height (β = 0.840) and a partially compensatory negative coefficient on collar diameter (β = −0.658) together accounted for 85.0 per cent of yield variation (R^2 = 0.850). Non-significant chi-square values across all four selections confirmed close agreement between predicted and observed yields, thereby validating the models. Together, the across-selection correlation structure and the within-selection regression equations demonstrate that a single common predictor variable cannot reliably estimate fresh fruit yield of *T. chebula* across selections and that selection-specific regression equations are essential for accurate non-destructive yield estimation in genetically heterogeneous plantations of the species.

INTRODUCTION

Terminalia chebula Retz. (Family: Combretaceae), commonly known as chebulic myrobalan or haritaki, is one of the most economically and medicinally important deciduous trees of the Indian subcontinent. Its dried fruit forms one of the three ingredients of the classical Ayurvedic formulation *Triphala* and the species is widely employed in pharmaceutical, nutraceutical and traditional medicine industries [1, 2]. Beyond its medicinal value, *T. chebula* is increasingly being recognised as a promising multipurpose agroforestry tree on account of its hardiness, drought tolerance, ability to grow on marginal lands, leaf-fodder potential and the high commercial value of its fruit [3].

Despite this importance, *T. chebula* remains a comparatively under-domesticated species, with limited standardisation of agronomic protocols and yield-estimation procedures. As *T. chebula* is an out-crossed species, plantations established from seed-derived material exhibit considerable variability in tree morphology, phenology and fruit yield [4–6]. This genetic heterogeneity is further compounded by the use of contrasting geographical selections in modern plantations, where trees of provenances raised under common silvicultural management can differ substantially in their growth allometry and reproductive output.

In commercial horticulture and forestry, the destructive harvest of fruit from the entire stand is the only direct route to yield estimation, but it is labour-intensive, expensive and impractical for routine monitoring, progeny evaluation, fertilisation planning and pre-harvest market forecasting. Non-destructive estimation of fruit yield from easily measurable tree morphometric parameters – such as tree height, collar diameter, crown spread and branch number – is therefore widely recognised as a useful management tool in fruit and nut species [5, 7]. Such allometric and regression-based approaches have been extensively employed in apple, citrus, mango, pomegranate and aonla, but corresponding predictive frameworks are conspicuously lacking for *T. chebula*. Moreover, where regression models have been attempted in out-crossed tree species, the implicit assumption of a common predictor set across genotypes has rarely been tested. A complementary line of evidence, namely the across-selection correlation structure among morphometric traits and yield, can be used to establish whether such a pooled assumption is biologically justified.

The present investigation was therefore undertaken with the specific objectives of (i) characterising the across-selection co-variation among four morphometric predictors and fresh fruit yield using Pearson's correlation analysis, (ii) developing selection-specific stepwise multiple linear regression models to predict fresh fruit yield in four geographically distinct selections of *T. chebula* (Jammu, Kallar, Pragpur and Palodi) from the four morphometric predictors, and (iii) validating these models using the chi-square goodness-of-fit test between observed and predicted yields. The study tests the hypothesis that, owing to the genetic heterogeneity inherent in this out-crossed species, no single morphometric predictor will adequately describe yield across all selections, and that selection-specific equations will therefore provide more reliable non-destructive yield estimates than a pooled model.

MATERIALS AND METHODS

Experimental site

The study was conducted during 2023 and 2024 at the Experimental Farm of the Division of Silviculture and Agroforestry, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu (SKUAST-J), Chatha, located at 32° 39' 35" N latitude and 74° 48' 29" E longitude at an elevation of 295 m above mean sea level. The region experiences a sub-tropical climate with hot, dry summers (maximum temperature reaching up to 45°C during May-June), a humid monsoon (July-September) and cool winters (minimum temperature falling to approximately 1°C during December–January). Mean annual rainfall

ranges between 1000 and 1200 mm, of which 75–80 per cent is received during July to September. The soil of the site was sandy loam in texture, slightly alkaline in reaction (pH 7.56–7.58), medium in organic carbon (0.44–0.48 per cent), low in available nitrogen (235.1-238.5 kg ha^{-1}) and available phosphorus (15.4–15.7 kg ha^{-1}) and medium in available potassium (125.5-128.1 kg ha^{-1}).

Plant material and experimental layout

The investigation was carried out on a seven-year-old plantation of *T. chebula* comprising four selections drawn from four different geographical regions, viz., Jammu (S_1; Jammu, Jammu and Kashmir), Kallar (S_2; Bilaspur, Himachal Pradesh), Pragpur (S_3; Kangra, Himachal Pradesh) and Palodi (S_4; Sirmour, Himachal Pradesh). Each selection was represented by 12 plants established in a contiguous block at a uniform spacing of 4 m × 6 m. Within each selection, all trees bearing fruits were marked. Six fruit-bearing trees per selection (selected by simple random sampling in the Kallar selection, in which seven trees were bearing) were taken as the sampling units for the development of regression models. The same six trees per selection were re-measured in 2023 and 2024 to capture inter-annual variability.

Observations recorded

Four morphometric variables and one yield variable were recorded on each sample tree. Tree height (m) was measured from the base of the stem to the tip of the leader shoot using a Ravi altimeter. Collar diameter (cm) was measured at the root collar using a digital calliper. Crown spread (m) was estimated as the mean of two perpendicular measurements taken in the east–west (D_1) and north–south (D_2) directions of the tree using a metric tape and computed as $(D_1 + D_2) / 2$. Number of primary branches arising from the main stem was counted by direct visual enumeration. Fresh fruit yield per tree (kg $tree^{-1}$) was determined by harvesting all the fruits from each sample tree on 30 November of each study year and weighing them immediately on a digital electronic balance.

Statistical analysis and model validation

Prior to fitting the selection-specific regression equations, Pearson's product-moment correlation coefficients were computed among the four morphometric variables and fresh fruit yield using selection-mean values (n = 4 selections) to characterise the across-selection co-variation among traits and to verify whether any single morphometric variable was consistently associated with yield at the genotype level. The resulting correlation matrix was visualised as a colour-coded heatmap using a diverging blue–red palette centred at zero. Given the limited number of selections (n = 4, df = 2), correlations were interpreted in terms of magnitude and direction of association rather than tested for individual significance, the critical value of $|r|$ for $P < 0.05$ at two degrees of freedom being approximately 0.950 [8].

For each selection, a separate multiple linear regression model was developed using the stepwise selection procedure [9], with fresh fruit yield (kg tree^{-1}) as the dependent variable and the four morphometric variables – tree height (TH), collar diameter (CD), crown spread (CS) and number of branches (NB) – as candidate predictors. Variables were entered or removed using the conventional probability thresholds of $P \leq 0.05$ (entry) and $P \geq 0.10$ (removal). The performance of each model was characterised by the coefficient of determination (R^2), the adjusted coefficient of determination (adjusted R^2), the unstandardised and standardised regression coefficients (β) and the associated probability levels. Standardised coefficients were used to compare the relative importance of predictors within the same model. The goodness of fit between observed and predicted values of fresh fruit yield per tree was tested using the chi-square (χ^2) statistic, with a non-significant χ^2 value ($P > 0.05$) indicating that the predicted distribution closely conformed to the observed distribution and thereby validating the model [10].

RESULTS

Variation in morphometric and yield variables across selections

Mean values of the four morphometric variables and fresh fruit yield in the four selections, averaged over the two years of study, are presented in Table 1. Tree height varied significantly across selections, ranging from 3.80 m in S_1 (Jammu) to 5.93 m in S_2 (Kallar). Collar diameter ranged from 11.85 cm in S_1 (Jammu) to 16.08 cm in S_3 (Pragpur). Crown spread was maximum in S_3 (Pragpur; 4.96 m) and minimum in S_4 (Palodi; 3.65 m). Number of branches per tree did not vary significantly across selections, with an overall mean of 6.46. Fresh fruit yield was significantly higher in S_2 (Kallar; 21.82 kg tree^{-1}) and S_1 (Jammu; 19.19 kg tree^{-1}) than in S_3 (Pragpur; 15.44 kg tree^{-1}) and S_4 (Palodi; 13.67 kg tree^{-1}). Importantly, the rank order of selections for fresh fruit yield ($S_2 > S_1 > S_3 > S_4$) did not coincide with the rank order for any single morphometric variable, indicating that no individual predictor would, on its own, faithfully represent fruit-yield variation across the selections.

Table 1
Mean values (over 2023 and 2024) of tree morphometric parameters and fresh fruit yield in four selections of *Terminalia chebula* under subtropical Jammu conditions.

Selection	Tree height (m)	Collar diameter (cm)	Crown spread (m)	No. of branches	Fresh fruit yield (kg/tree)
S_1 (Jammu)	3.80	11.85	4.68	7.00	19.19
S_2 (Kallar)	5.93	13.30	4.14	6.67	21.82
S_3 (Pragpur)	5.09	16.08	4.96	6.00	15.44
S_4 (Palodi)	4.89	13.59	3.65	6.17	13.67
$CD_{0.05}$	*0.81*	*2.51*	*0.78*	*NS*	*5.49*

Pearson's correlation structure across selections

The Pearson's correlation matrix computed across the four selections means revealed that no single morphometric variable was strongly and consistently associated with fresh fruit yield (Fig. 2; Table 2). At the across-selection level, fresh fruit yield was most strongly and positively correlated with number of branches (r = + 0.77), followed by weakly positive correlations with tree height (r = + 0.23) and crown spread (r = + 0.17), whereas collar diameter showed a moderate negative correlation with yield (r = − 0.48). Among the predictors themselves, collar diameter and number of branches were strongly and negatively correlated (r = − 0.89), indicating a clear inverse relationship between radial stem growth and branch production at the selection level. Tree height was moderately positively correlated with collar diameter (r = + 0.45) and weakly negatively correlated with both crown spread (r = − 0.30) and number of branches (r = − 0.38), while crown spread was only weakly correlated with the other predictors ($|r| \leq 0.33$).

With only four selections (df = 2), individual correlations did not attain conventional levels of statistical significance (critical $|r| \approx 0.950$ for $P < 0.05$), but the pattern of correlations was biologically informative. In particular, the moderate negative correlation between collar diameter and yield at the across-selection level confirmed that genotypes with the largest radial stem growth (notably the Pragpur selection) were not the highest-yielding, whereas the strong negative correlation between collar diameter and number of branches suggested a trade-off between vegetative thickening and branching that varied systematically among selections. Crucially, the variable showing the strongest positive correlation with yield across selections – number of branches – was not the variable retained by the stepwise procedure in three of the four within-selection regression models (see subsequent sections), illustrating that across-selection trait–yield relationships cannot be assumed to mirror within-selection relationships. Together, these patterns provided a clear rationale for fitting separate stepwise regression equations within each selection rather than relying on a single pooled model.

Table 2
Pearson's correlation matrix among tree morphometric parameters and fresh fruit yield computed across the four selections means (n = 4) of *Terminalia chebula*.

Variable	TH	CD	CS	NB	FFY
Tree height (TH)	**1.000**	+ 0.449	−0.299	−0.383	+ 0.231
Collar diameter (CD)	+ 0.449	**1.000**	+ 0.327	−0.888	−0.483
Crown spread (CS)	−0.299	+ 0.327	**1.000**	+ 0.071	+ 0.168
No. of branches (NB)	−0.383	−0.888	+ 0.071	**1.000**	+ 0.768
Fresh fruit yield (FFY)	+ 0.231	−0.483	+ 0.168	+ 0.768	**1.000**

TH = tree height; CD = collar diameter; CS = crown spread; NB = number of branches; FFY = fresh fruit yield. Correlations computed on selection-mean values (n = 4). Critical |r| ≈ 0.950 for P < 0.05 at df = 2.

Stepwise regression model for the Jammu selection (S_1)

In the Jammu selection, the stepwise procedure retained collar diameter (CD) and number of branches (NB) as significant predictors of fresh fruit yield, while tree height and crown spread were excluded at the entry threshold (Table 3). The fitted equation explained 83.5 per cent of the variation in fresh fruit yield per tree (R^2 = 0.835; adjusted R^2 = 0.798) and the predicted distribution did not differ significantly from the observed distribution (χ^2 = 2.64, P > 0.05). The corresponding equation is:

Y = − 7.930 + 1.589 CD + 1.183 NB

Based on standardised coefficients, collar diameter (β = 0.717) was the more influential predictor compared with number of branches (β = 0.354), implying that radial stem growth contributed more strongly to fresh fruit yield in the Jammu selection than the morphological architecture of the canopy.

Table 3
Stepwise regression model predicting fresh fruit yield (kg tree^{-1}) in the Jammu selection (S_1) of *Terminalia chebula*.

Variable	Unstandardised coefficient	Standardised coefficient (β)	P-value
Constant	−7.930	–	0.086
Collar diameter (cm)	1.589	0.717	0.001
Number of branches	1.183	0.354	0.040

R^2 = 0.835; Adjusted R^2 = 0.798; χ^2 = 2.64, non-significant (P > 0.05).

Stepwise regression model for the Kallar selection (S_2)

In contrast to the Jammu selection, tree height was the only variable retained by the stepwise procedure in Kallar, with collar diameter, crown spread and number of branches all failing to meet the entry threshold (Table 4). The fitted equation accounted for 40.7 per cent of the variation in fresh fruit yield (R^2 = 0.407; adjusted R^2 = 0.348) and the predicted yields conformed well to the observed yields (χ^2 = 11.39, P > 0.05). The corresponding equation is:

$$Y = -24.377 + 7.487\ TH$$

The relatively modest R^2 of the Kallar model, despite the high mean fruit yield of the selection, suggests that other unmeasured factors – such as the number of fruit-bearing shoots, canopy density or year-to-year flowering intensity – may contribute substantially to yield in Kallar trees beyond what is captured by gross stem dimensions.

Table 4

Stepwise regression model predicting fresh fruit yield (kg $tree^{-1}$) in the Kallar selection (S_2) of *Terminalia chebula*.

Variable	Unstandardised coefficient	Standardised coefficient (β)	P-value
Constant	−24.377	–	0.198
Tree height (m)	7.487	0.638	0.026

R^2 = 0.407; Adjusted R^2 = 0.348; χ^2 = 11.39, non-significant (P > 0.05).

Stepwise regression model for the Pragpur selection (S_3)

In the Pragpur selection, collar diameter alone emerged as the single significant predictor of fresh fruit yield, explaining 36.0 per cent of the variation (R^2 = 0.360; adjusted R^2 = 0.295). The goodness-of-fit test indicated close agreement between predicted and observed yields (χ^2 = 4.17, P > 0.05). The corresponding equation is:

$$Y = -2.878 + 1.140\ CD$$

The dominance of collar diameter in this selection is consistent with its biology: although Pragpur trees recorded the largest collar diameter (16.08 cm) and crown spread (4.96 m) among the four selections, their fruit yield was lower than that of S_1 and S_2, indicating that radial growth is a meaningful but not a sufficient determinant of reproductive output in Pragpur.

Table 5

Stepwise regression model predicting fresh fruit yield (kg tree^{-1}) in the Pragpur selection (S_3) of *Terminalia chebula*.

Variable	Unstandardised coefficient	Standardised coefficient (β)	P-value
Constant	−2.878	–	0.719
Collar diameter (cm)	1.140	0.600	0.039

R^2 = 0.360; Adjusted R^2 = 0.295; χ^2 = 4.17, non-significant (P > 0.05).

Stepwise regression model for the Palodi selection (S_4)

The Palodi selection produced the most parameter-rich model among the four selections, with crown spread (CS), tree height (TH) and collar diameter (CD) jointly retained as significant predictors of fresh fruit yield (Table 6). These three variables together accounted for 85.0 per cent of the variation in yield (R^2 = 0.850; adjusted R^2 = 0.794) and the goodness-of-fit test confirmed close agreement between observed and predicted yields (χ^2 = − 0.09, P > 0.05). The fitted equation is:

$$Y = -8.727 + 1.037\,CS + 0.777\,TH - 0.735\,CD$$

Based on standardised coefficients, crown spread (β = 0.879) emerged as the most influential positive driver of fresh fruit yield, closely followed by tree height (β = 0.840). Notably, collar diameter contributed a negative coefficient (β = −0.658, P = 0.029) once the effects of crown spread and tree height had been accounted for. This partial regression coefficient does not imply that thicker stems intrinsically depress fruit yield in Palodi; rather, it suggests that, conditional on the canopy and height already represented in the model, additional radial stem growth is associated with disproportionately lower yield, possibly reflecting carbon partitioning towards stem-wood at the expense of fruiting in the largest-bole Palodi individuals.

Table 6

Stepwise regression model predicting fresh fruit yield (kg tree^{-1}) in the Palodi selection (S_4) of *Terminalia chebula*.

Variable	Unstandardised coefficient	Standardised coefficient (β)	P-value
Constant	−8.727	–	0.040
Crown spread (m)	1.037	0.879	0.002
Tree height (m)	0.777	0.840	0.002
Collar diameter (cm)	−0.735	−0.658	0.029

R^2 = 0.850; Adjusted R^2 = 0.794; χ^2 = − 0.09, non-significant (P > 0.05).

DISCUSSION

The four selection-specific regression equations developed in the present study, taken together with the Pearson's correlation matrix at the across-selection level, collectively demonstrate that no single morphometric variable can be regarded as a universal predictor of fresh fruit yield in *Terminalia chebula*. Three different predictor sets were retained across the four selections: collar diameter together with number of branches in Jammu; tree height alone in Kallar; collar diameter alone in Pragpur; and crown spread, tree height and collar diameter jointly in Palodi. This heterogeneity in predictor set is consistent with the genetic and morphological variability previously reported in *T. chebula* across regions [4–6] and reflects the fact that the species is naturally out-crossed, giving rise to numerous genotypes with diverse allocation strategies for stem, canopy and reproductive growth [11].

The Pearson's correlation analysis at the across-selection level provided important context for interpreting the four selection-specific regression equations. The moderate negative correlation between collar diameter and fresh fruit yield across selection means ($r = -0.48$) indicates that, at the genotype level, the relationship between radial stem growth and reproductive output is not monotonically positive across the species – a finding that is fully consistent with the contrasting roles played by collar diameter within each selection (a strong positive predictor in Jammu and Pragpur, but a suppressor-like negative partial predictor in Palodi). Likewise, the strong negative correlation between collar diameter and number of branches ($r = -0.89$) at the across-selection level suggests that the same physiological trade-off between bole thickening and branch proliferation that underlies the Palodi result also operates among selections. Importantly, the variable that emerged as the strongest yield correlate across selection means – number of branches ($r = +0.77$) – was retained as a significant predictor only in the Jammu within-selection equation, illustrating clearly that across-selection relationships among trait means cannot be assumed to mirror within-selection trait–yield relationships in genetically heterogeneous *T. chebula* plantations.

The Jammu selection produced a relatively strong two-predictor model, in which radial stem growth (collar diameter) was the dominant driver of fresh fruit yield, with number of branches providing additional explanatory power. This pattern is biologically plausible: in this comparatively short-statured selection (mean tree height of 3.80 m) trees with thicker stems are likely to possess a larger photosynthate reserve and greater hydraulic capacity, both of which scale with reproductive output, while a higher number of primary branches simply provides more fruiting sites. Saleem *et al.* [3] also documented superior fruit characteristics for Jammu provenance material under conditions broadly comparable to the present study.

In Kallar, the dominance of tree height as the sole predictor likely reflects the unusually upright architecture of this selection, which recorded the maximum mean height (5.93 m) but a comparatively modest crown spread (4.14 m). The relatively low R^2 value of 0.407 suggests that gross stem dimensions capture only a part of the yield variability in this selection; reproductive output in Kallar appears to be governed additionally by factors not included in the present predictor set, such as the proportion of reproductive shoots, flowering intensity or year-to-year masting behaviour. In Pragpur, the very large collar diameter (16.08 cm) and crown spread (4.96 m) coupled with a comparatively modest

fruit yield (15.44 kg tree^{-1}) indicate that this selection invests disproportionately in vegetative biomass relative to reproductive output and that radial stem growth alone explains roughly one-third of among-tree yield variability.

The Palodi model is the most informative of the four, both in terms of explanatory power (R^2 = 0.850) and biological interpretation. The dominant positive contribution of crown spread (β = 0.879), supported by tree height (β = 0.840), is consistent with the well-established notion that the photosynthetically active canopy area is the proximate source of carbohydrates for fruit development. The apparently anomalous negative coefficient on collar diameter (β = −0.658) should not be interpreted in isolation: in a multiple regression framework it represents the partial effect of collar diameter at constant crown spread and tree height. Such suppressor-like or compensatory partial coefficients are well known in tree allometry and arise when two or more correlated predictors are retained simultaneously; biologically, the result implies that, among Palodi trees of comparable canopy size and height, those with disproportionately thick stems tend to allocate carbon preferentially to stem-wood, with a corresponding reduction in fruit yield. This within-selection finding mirrors the across-selection negative correlation between collar diameter and yield (r = − 0.48) observed in the Pearson analysis [12].

The non-significant chi-square goodness-of-fit values obtained for all four selections (χ^2 = 2.64 for Jammu, 11.39 for Kallar, 4.17 for Pragpur and − 0.09 for Palodi; all P > 0.05) provide a robust validation of the equations, demonstrating that the distribution of predicted yields conforms statistically to the distribution of observed yields. This means that, despite the differing R^2 values of the individual models, each is internally consistent and can be used for non-destructive yield estimation within its own selection.

From an applied standpoint, the present results have three immediate implications. First, *T. chebula* plantations established with material of mixed or unknown provenance should not be evaluated using a pooled yield-prediction model; rather, separate equations should be developed for each identifiable selection, and the across-selection correlation structure should be reported alongside the within-selection equations to make the lack of a common predictor explicit. Second, the rank order of selections for fresh fruit yield ($S_2 > S_1 > S_3 > S_4$) demonstrates that yield can vary by approximately 60 per cent between the best and the poorest selection under identical management and any region-wide expansion of *T. chebula* cultivation should therefore prioritise the Kallar and Jammu selections. Third, the contrasting predictor sets suggest that future selection programmes can use easily measurable morphometric variables to indirectly screen for high-yielding individuals, but the screening variable should be matched to the selection – collar diameter and branch number in Jammu-type material, tree height in Kallar-type material, collar diameter in Pragpur-type material and crown spread together with tree height in Palodi-type material.

CONCLUSION

Selection-specific stepwise multiple linear regression models developed in the present study, supported by an across-selection Pearson's correlation analysis, provide validated, non-destructive equations for predicting fresh fruit yield in four selections of *Terminalia chebula* under subtropical Jammu conditions. The correlation analysis revealed that no single morphometric trait was uniformly associated with yield across selections: number of branches showed the strongest positive across-selection correlation with yield (r = + 0.77) and collar diameter a moderate negative one (r = − 0.48), while collar diameter and number of branches themselves were strongly inversely correlated (r = − 0.89). The within-selection regression analyses confirmed and refined this picture: collar diameter and number of branches were the dominant predictors in Jammu; tree height in Kallar; collar diameter in Pragpur; and crown spread, tree height and collar diameter in Palodi. Non-significant chi-square goodness-of-fit values across all four equations confirmed close agreement between observed and predicted yields. The absence of a common predictor set across selections, evident both in the across-selection correlation structure and in the four within-selection regression equations, indicates that, in a genetically heterogeneous out-crossed species such as *T. chebula*, selection-specific yield-prediction equations are essential and a single pooled model is inappropriate. The proposed equations can be used by orchard managers, agroforestry practitioners and tree-breeding programmes for non-destructive yield estimation, harvest planning and indirect selection of superior individuals within each selection.

Declarations

Funding

This research received no external funding.

Author Contribution

S.S. conceived the study, conducted the investigation, collected and analysed the data, prepared the figures and tables, and wrote the original manuscript. M.S.S. contributed to the study design, interpreted the results, and critically revised the manuscript. K.K.S. supervised the research, provided scientific guidance, interpreted the findings, and critically revised the manuscript. S.K.G. contributed to the methodology, provided scientific guidance, and critically revised the manuscript. S.K. and T.L. contributed to the review and editing of the manuscript. All authors reviewed and approved the final manuscript.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

1. Bag, A., Bhattacharyya, S. K. & Chattopadhyay, R. R. The development of Terminalia chebula Retz. (Combretaceae) in clinical research. *Asian Pac. J. Trop. Biomed.* **3** (3), 244–252 (2013).
2. Cock, I. E. The medicinal properties and phytochemistry of plants of the genus Terminalia (Combretaceae). *Inflammopharmacology* **23** (5), 203–229. 10.1007/s10787-015-0246-z (2015).
3. Saleem, M., Sood, K. K., Gupta, S. K., Raina, N. S. & Gupta, L. M. Suitable agro-techniques for germination and vegetative propagation of superior germplasm of Terminalia chebula Retz. – a multipurpose agroforestry tree. *Range Manage. Agrofor.* **35** (1), 121–127 (2014).
4. Navhale, V. C., Sonone, N. G., Jangam, P. S., Jadhav, S. T. & Bhave, S. G. Genetic variability and selection of candidate plus trees in chebulic myrobalan (Terminalia chebula). *Electron. J. Plant. Breed.* **2** (1), 157–163 (2011).
5. Nakandalage, N. et al. Morphological characterization of Terminalia chebula (Aralu) and their propagation techniques. *Int. J. Minor. Fruits Med. Aromatic Plants*. **7** (1), 84–88. 10.53552/ijmfmap.2021.v07i01.011 (2021).
6. Mahadeva, S. M. et al. Unveiling the genetic potential of Terminalia chebula Retz. accessions from North-East and South India: a morphometric analysis. *Trees Forests People*. **18**, 100669. 10.1016/j.tfp.2024.100669 (2024).
7. Sharma, D., Thakur, S. & Jha, S. K. Characterization of population variation for fruit and pulp in Terminalia chebula. *Agroforest Syst.* **90** (7), 361–369 (2015).
8. Dormann, C. F. et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46 (2013).
9. Gomez, K. A. & Gomez, A. A. *Statistical Procedures for Agricultural Research* 2nd edn (Wiley, 1984).
10. O'Brien, R. M. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* **41**, 673–690. 10.1007/s11135-006-9018-6 (2007).
11. Paul, K. I. et al. Testing the generality of above-ground biomass allometry across plant functional types at the continent scale. *Glob. Change Biol.* **22**, 2106–2124 (2016).
12. Tu, Y. K., Gunnell, D. & Gilthorpe, M. S. Simpson's Paradox, Lord's Paradox, and Suppression Effects are the same phenomenon – the reversal paradox. *Emerg. Themes Epidemiol.* **5**, 2. 10.1186/1742-7622-5-2 (2008).

Figures

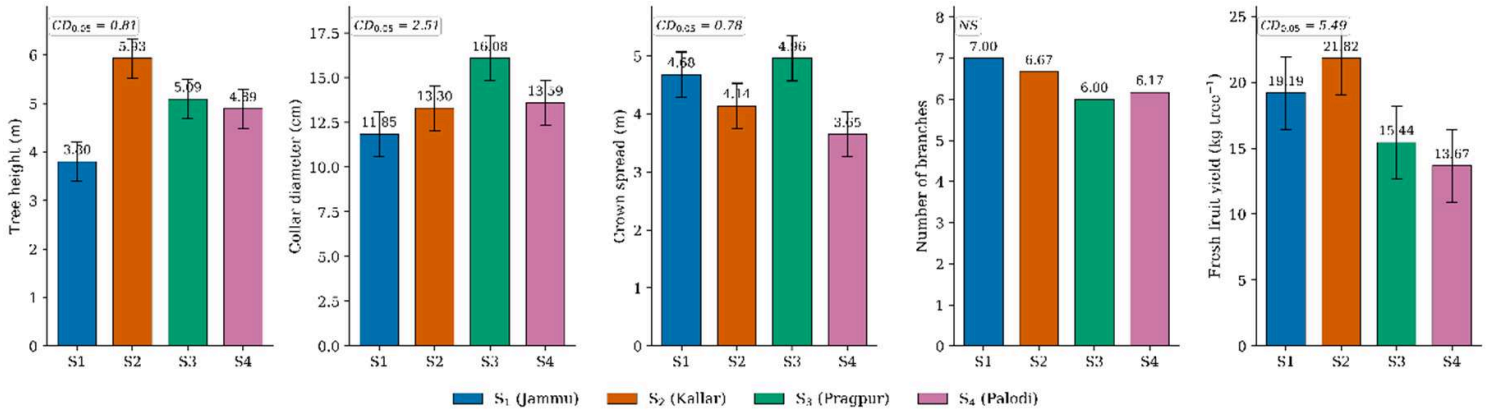


Figure 1

Tree morphometric parameters and fresh fruit yield in four selections of *Terminalia chebula* under subtropical Jammu conditions.

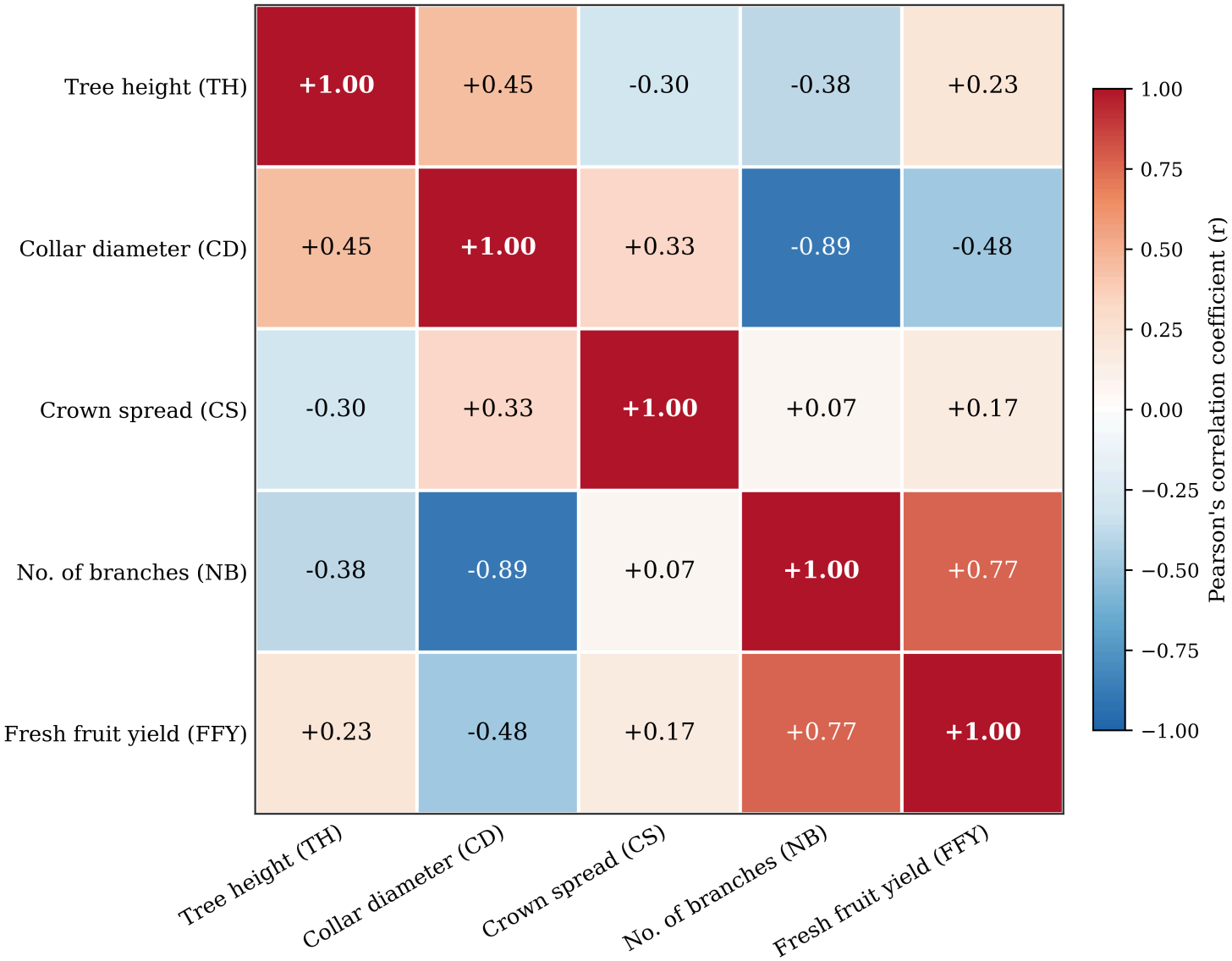


Figure 2

Pearson's correlation heatmap among tree height (TH), collar diameter (CD), crown spread (CS), number of branches (NB) and fresh fruit yield (FFY) computed across selection means (n = 4) of *Terminalia chebula*. Cell shading corresponds to the magnitude and sign of the Pearson correlation coefficient (r); blue tones indicate negative associations, red tones indicate positive associations and white indicates near-zero correlation.

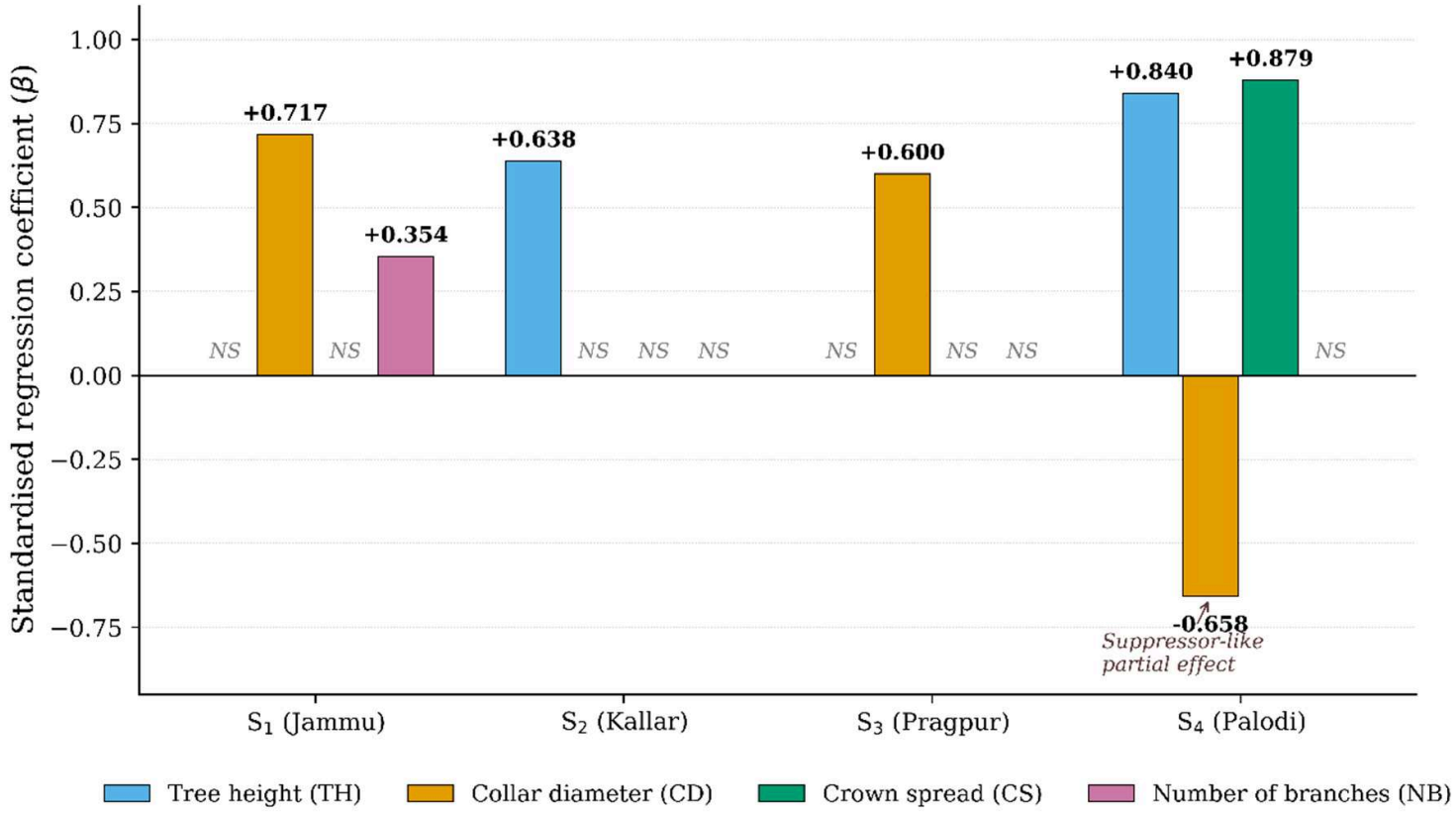


Figure 3

Standardised regression coefficients (β) of tree morphometric predictors of fresh fruit yield across four selections of *Terminalia chebula*.

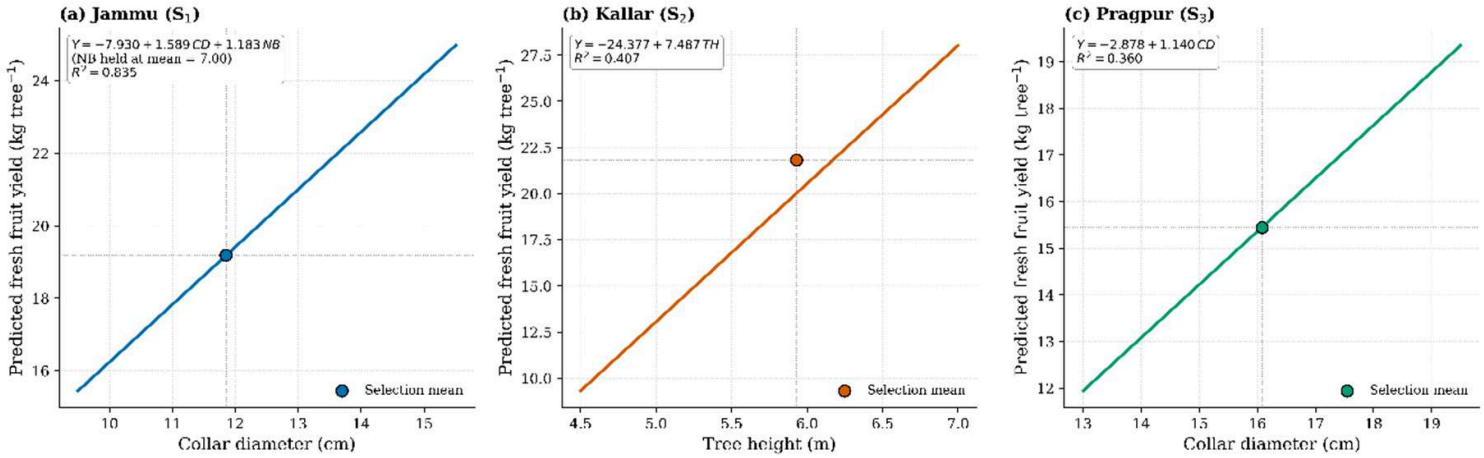


Figure 4

Predicted fresh fruit yield as a function of the dominant morphometric predictor in three selections of *Terminalia chebula*. In (a) Jammu, (b) Kallar and (c) Pragpur.