https://doi.org/10.1590/1678-4499.20230269

BASIC AREAS
Article

A non-destructive method for leaflet area prediction of *Spondias tuberosa* Arruda: an approach to regression models

Patrycia Elen Costa Amorim[1], Dalbert de Freitas Pereira[1], Raíres Irlenizia da Silva Freire[1], Agda Malany Forte de Oliveira[1], Vander Mendonça[1], João Everthon da Silva Ribeiro[1,*]

1. Universidade Federal Rural do Semi-Árido – Departamento de Ciências Agronômicas e Florestais – Mossoró (RN), Brazil.

Received: Nov. 23, 2023 | **Accepted:** Feb. 26, 2024

Section Editor: Alberto Cargnelutti Filho

***Corresponding author:** j.everthon@hotmail.com

How to cite: Amorim, P. E. C., Pereira, D. F., Freire, R. I. S., Oliveira, A. M. F., Mendonça, V. and Ribeiro, J. E. S. (2024). A non-destructive method for leaflet area prediction of *Spondias tuberosa* Arruda: an approach to regression models. Bragantia, 83, e20230269. https://doi.org/10.1590/1678-4499.20230269

ABSTRACT: Umbu (*Spondias tuberosa* Arruda, Anacardiaceae) is a fruit species native to the semi-arid region of Brazil and economically important for several regions. The objective of this study was to determine equations to estimate the leaflet area of *S. tuberosa* genotypes. A total of 1,000 leaflets was collected from four genotypes (250 leaflets of each genotype) of *S. tuberosa*. In each leaflet collected, the length, width, and leaflet area were measured, and the product between length and width was calculated. Linear, linear with intercept, power, and exponential regression models were used to fit the equations. The criteria for choosing the best equation were based on Pearson's correlation coefficients, Akaike's information criteria, Willmott's concordance indices, and root mean square error. The product-adjusted equations between length and width can be used to estimate the leaflet area of all *S. tuberosa* genotypes. The leaflet area of the species (pooled data) can be estimated accurately and quickly from equations obtained with the linear models without intercept ($\hat{y} = 0.6911*LW$) and power ($\hat{y} = 0.7127*LW^{0.9888}$).

Key words: Anacardiaceae, biometrics, allometric equations, leaflet length, leaflet width.

INTRODUCTION

Leaves are the structure with the most significant interaction with biotic and abiotic factors, participating in critical physiological processes of plants, such as light interception, carbon fixation, gas exchange, and plant defense (Legris 2023). During the evolution of species, due to different environmental gradients, this structure developed various sizes and shapes to allow survival and perpetuation (Lusk et al. 2019, He et al. 2020).

Leaf area is among the factors that condition plant performance in natural and agricultural ecosystems, which is related to photosynthetic capacity (relative leaf growth rate, net assimilation rate, and photosynthetic efficiency), biomass production, competition, nutrition, soil-plant relationship, and, as a result, leaf area is used as a parameter in plant physiology and production (Macário et al. 2020, Adji et al. 2021, Boyaci and Küçükönder 2022).

Direct, destructive, or non-destructive methods are used to determine leaf area, as well as indirect (non-destructive) methods (Ribeiro et al. 2022). The determination by direct and destructive methods, such as the use of graph paper, gravimetric method, leaf discs, and bench gauges, requires the extraction of leaves, which, consequently, compromises research with reduced samples, the evaluation of leaves until the end of the cycle, the development of the plant, and makes successive analyses unfeasible (Pohlmann et al. 2021).

Direct and non-destructive methods are carried out with precise and easy-to-use equipment, such as portable foliar scanners. However, they require expensive acquisition and complex maintenance, which sometimes becomes unfeasible. On the other hand, indirect (non-destructive) methods assess leaf area while preserving leaves, allowing successive sampling over time in a fast, practical, and low-cost way (Schmildt et al. 2023). Commonly, indirect methods involve using linear

leaf dimensions, such as the length and width of leaves or the product of these dimensions, applied to regression equations consolidated by statistical modeling studies, determining the values referring to leaf area with high precision. From knowledge of regression equations, it is possible to determine the leaf area using easily accessible graded tools, sparing leaf tissue, not compromising development, and allowing successive evaluations over time (Ribeiro et al. 2022).

Several studies have used linear dimensions to determine leaf area with precision. Salazar et al. (2018) proposed the use of the product of length and width using polynomial regressions to estimate the leaf area of cacao (*Theobroma cacao* L.) with a coefficient of determination of 98%, and this use is reported for several species, such as *Ceiba glaziovii* (Ribeiro et al. 2022), *Eustoma grandiflorum* (Dias et al. 2022), watermelon (Rouphael et al. 2010), *Forsythia viridissima, Ligustrum lucidum, Ligustrum sinense, Osmanthus fragrans, Syringa oblata* var. *alba* (Shi et al. 2019), and *Malus domestica* cultivars (Boyaci and Küçükönder 2022).

Spondias tuberosa is a fruit tree native to Brazil, belonging to the Anacardiaceae family, endemic to the Brazilian semi-arid region and that can reach 4 to 6 m in height with a crown of 10 to 15 m in diameter (Lima et al. 2018). It is adapted to regions with annual rainfall of 400 to 800 mm and temperatures of 12 to 38°C. It has tuberous roots, called xylopods, which allow it to tolerate recurrent water stress in the Brazilian semi-arid region (Menezes et al. 2017). Although this species has broad ecological and economic importance, non-destructive methods for determining the leaflet area of *S. tuberosa* have not yet been reported. Therefore, the objective of this study was to determine equations to estimate the leaflet area of *S. tuberosa* genotypes.

MATERIAL AND METHODS

The study used genotypes of *S. tuberosa* from the Rafael Fernandes Experimental Farm, belonging to the Universidade Federal Rural do Semi-Árido, located in Mossoró, Rio Grande do Norte, Brazil. The climate of the region, according to Köppen, is BSh type (Alvares et al. 2013), very hot, semi-arid, steppe-type climate, with an average temperature of 27.8°C, with a rainy season in April, May, and June, with an average annual rainfall of the region of 555 mm (Climate-Data 2023).

The genotypes of *S. tuberosa* used in the study were Esperança, Macaúbas, Livramento Cavaco, and Ribeira de Pombal. For each genotype, 250 expanded leaflets were collected, free of pests, diseases, and other biotic or abiotic factors (Fig. 1), totaling 1,000 leaflets sampled. The collected material was stored in plastic bags and kept in the shade to prevent excessive water loss through transpiration. Then, the leaflets were separated and scanned using a flatbed scanner (model Samsung Xpress SL-C480FW) at 600 DPI resolution.

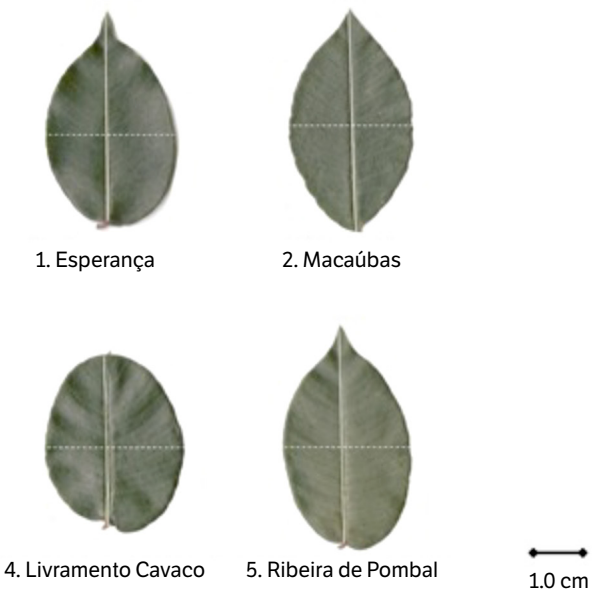

Figure 1. Leaflets of *Spondias tuberosa* genotypes.

The digitized leaflets were processed in the ImageJ software with image contrast, according to Ribeiro et al. (2022). After processing, the length (L) was determined, consisting of the distance from the leaflet apex to the insertion of the peduncle, the width (W), the maximum measurement perpendicular to the midrib, and the leaflet area (LA). From the L and W data, the product between these parameters (LW) was calculated.

Linear ($\hat{y} = \beta_0 + \beta_1 \cdot x + \varepsilon_i$), linear with intercept ($\hat{y} = \beta_1 \cdot x + \varepsilon_i$), power ($\hat{y} = \beta_0 \cdot x^{\beta_1} + \varepsilon_i$), and exponential ($\hat{y} = \beta_0 \cdot \beta_1^{\,x} + \varepsilon_i$) regression models were used, where $\hat{y}$ corresponds to the estimation of leaflet area (LA) as a function x (leaflets dimensions–L, W, and LW).

The selection of the best models and equations was based on the following criteria: coefficient of determination (R^2) (Eq. 1), Pearson's correlation coefficient (r) (Eq. 2), Akaike's information criterion (AIC) (Eq. 3), Willmott's agreement index (*d*) (Eq. 4), and root mean square error (RMSE) (Eq. 5).

$$R^2 = 1 - \frac{\sum_{i=_1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=_1}^{n}(y'_i)^2} \tag{1}$$

$$r = \frac{\sum_{i=_1}^{n}(y_i - \bar{y}_i)(x_i - \bar{x}_i)}{\sqrt{\sum_{i=_1}^{n}(x_i - \bar{x}_i)^2 \sum_{i=_1}^{n}(y_i - \bar{y}_i)^2}} \tag{2}$$

$$AIC = -2 \ln L \left(x \backslash \hat{\theta}\right) + 2\,(p) \tag{3}$$

$$d = 1 - \frac{\sum_{i=_1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=_1}^{n}(|\hat{y}'_i| + |y_i|)^2} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=_1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{5}$$

where: R^2: coefficient of determination; r: Pearson's correlation coefficient; AIC: Akaike's information criterion; *d*: Willmott's concordance index; RMSE: root mean square error; ŷi: estimated leaflet area values; y'i: observed leaflet area values; mean yi of the observed values; y'I: ŷi - y; y'i: yi - y; L(x\θ): maximum-likelihood function; p: number of model parameters; n: number of observations; xi and yi: observations of variables y and x; y and x: mean of variables y and x.

Descriptive analyses were performed to determine minimum, mean, and maximum values, standard deviation, amplitude, standard error, and coefficient of variation. Normality was verified using the Shapiro-Wilk's test (Shapiro and Wilk 1965). R software was used in statistical analyses (R Core Team 2023). The observed and estimated leaflet area values were compared using Student's t-test for paired samples ($p \leq 0.01$).

RESULTS

The analyzed genotypes showed high variation of maximum and minimum values and high total amplitude (difference between maximum and minimum values) for L, W, LW, and observed LA (Fig. 2). Genotype 2 (Macaúbas) had the lowest maximum values for L (5.27 cm), W (3.11 cm), LW (16.39 cm^2), and real LA (10.66 cm^2). In addition, it presented the lowest minimum values for W of 1.11 cm, LW of 1.27 cm^2, and real LA of 1.27 cm^2. Consequently, the averages were also the lowest among all genotypes, with a mean L of 3.20 cm, W of 2.20 cm, LW of 7.10 cm^2, and real LA of 4.95 cm^2 (Fig. 2).

In contrast, genotype 4 (Ribeira de Pombal) had the highest maximum values for L of 8.05 cm, W of 4.46 cm, LW of 35.92 cm^2, and real LA of 22 cm^2, also presenting the highest minimum values for L and LW of 1.90 and 1.68 cm^2, respectively. The highest median values were 5.20-cm length (L), 3.25-cm width (W), 16.54 cm^2 for LW, and 11.70 cm^2 for real LA. The other genotypes (1–Esperança, and 3–Livramento Cavaco) showed slight variation with similar maximum, minimum and median values (Fig. 2).

The smallest amplitude of leaflet L was verified by genotype 2 (Macaúbas) at 3.906 cm, and the largest one by genotype 4 (Ribeira de Pombal), at 6.161 cm (Fig. 2a). The W of the leaves ranged from 2 (Macaúbas) to 3.58 cm (Ribeira de Pombal) (Fig. 2b). The product obtained from the ratio (LW) ranged from 14.677 (Macaúbas) to 34.2427 cm^2 (Ribeira de Pombal) (Fig. 2c). The variation for real LA was from 8.7927 (Macaúbas) to 20.8327 cm^2 (Ribeira de Pombal) (Fig. 2d). The highest coefficients of variation were obtained for LW (41.68%) and real LA (41.58%).

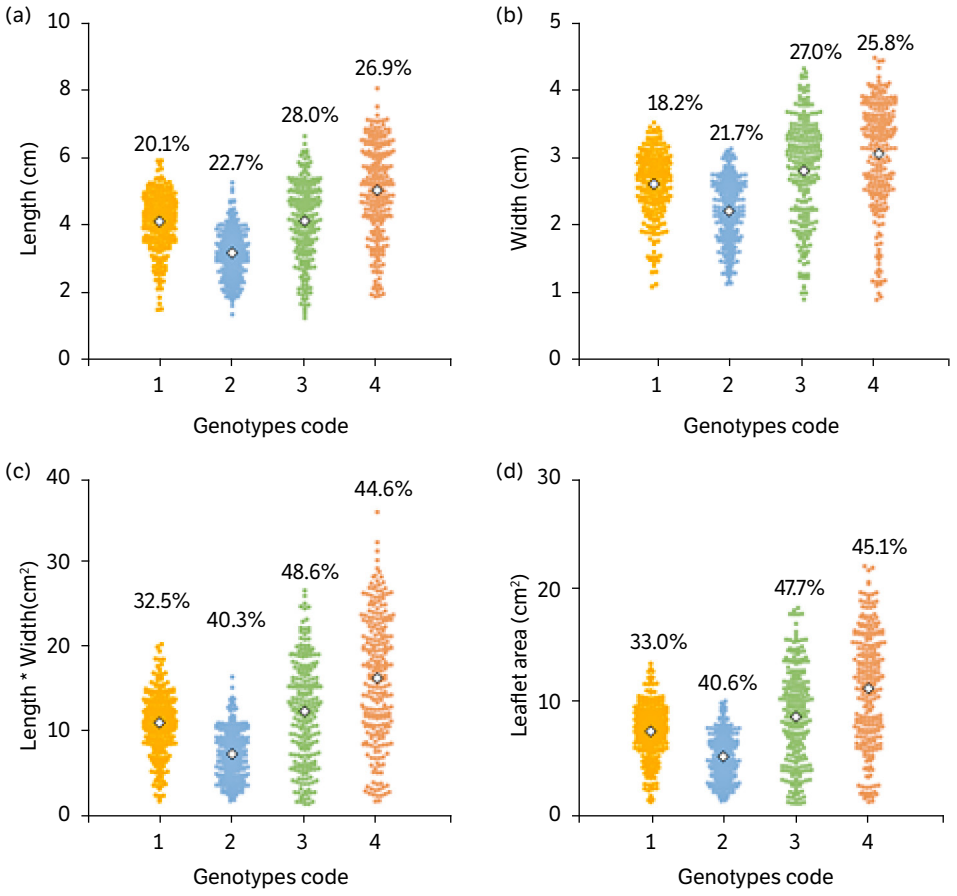

Figure 2. Descriptive analysis of the morphological parameters (a) length, (b) width, (c) length*width, and (d) leaflet area in four *Spondias tuberosa* genotypes (genotypes code: 1–Esperança; 2–Macáubas; 3–Livramento Cavaco; 4–Ribeira de Pombal). The numbers above the dots refer to the coefficients of variation.

The histogram of dispersion between the independent variables L, W, LW, and real LA indicate different relationships between them, suggesting adjustments of linear and nonlinear models (Fig. 3).

The principal component (PC) analysis of the four genotypes of *S. tuberosa* (Esperança, Macaúbas, Livramento Cavaco, and Ribeira de Pombal), based on the linear dimensions of the leaflets (L, W, and LW) and observed LA, is presented in Fig. 4. PC1 explained 70.05% of the PC analysis, while PC2 explained 27.24% of it, totaling 97.29% of the total variability (Fig. 4a). The genotypes did not show significant variability in the linear dimensions of the leaflets to be considered distinct, indicating a single group. Figure 4b shows the proximity and high correlation between the product of leaflet LW and observed LA.

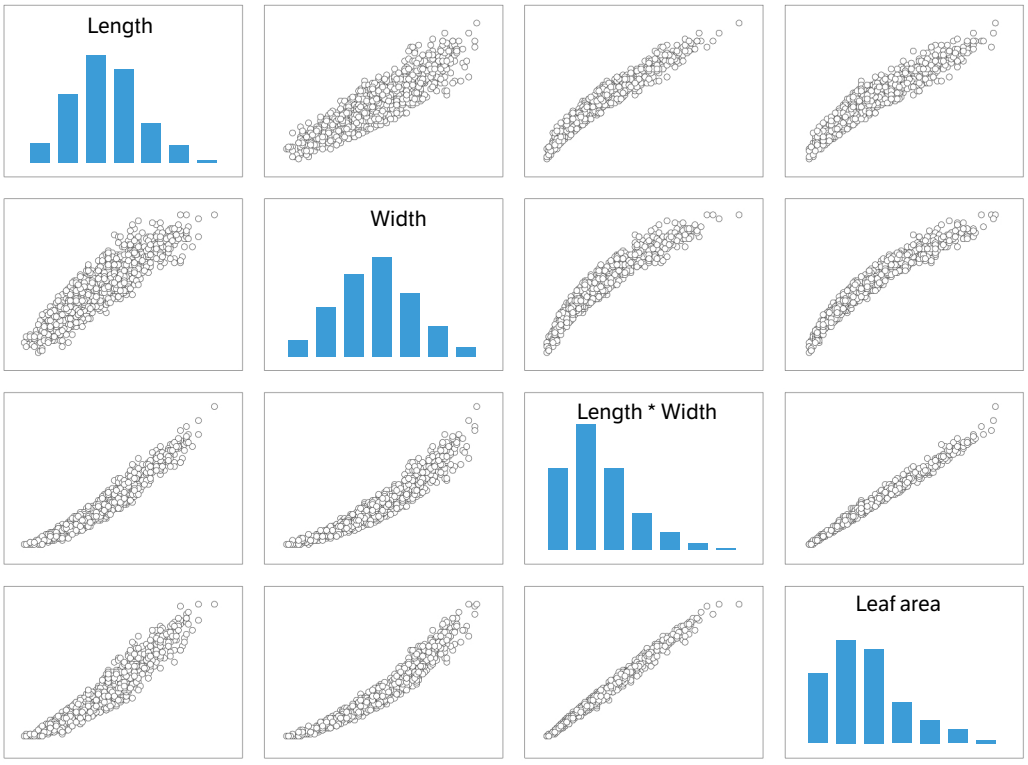

Figure 3. Matrix with frequency histogram and scatterplot of length, width, product of length and width, and leaflet area (grouped data) of four *Spondias tuberosa* genotypes (Esperança, Macáubas, Livramento Cavaco e Ribeira de Pombal).

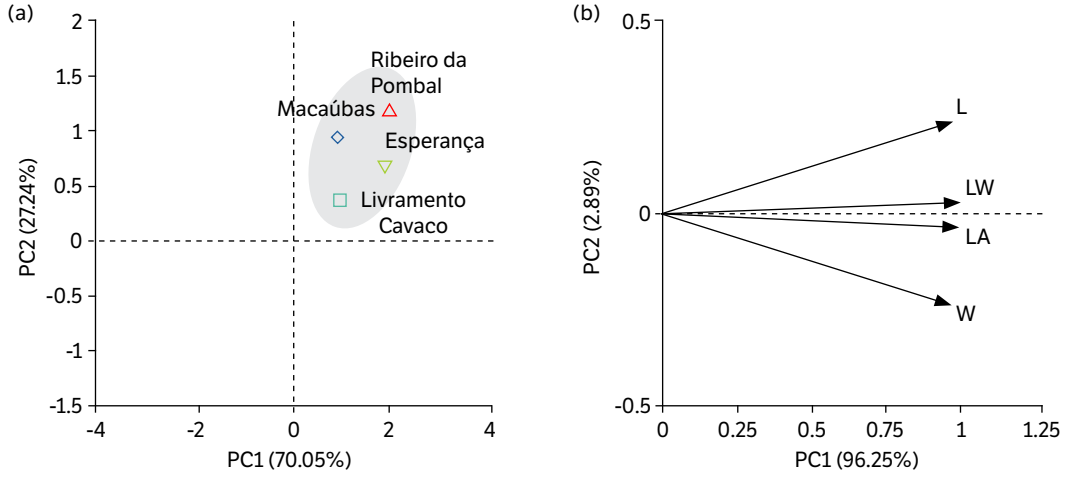

Figure 4. (a) Principal component analysis of the morphological parameters in four *Spondias tuberosa* genotypes (Esperança, Macáubas, Livramento Cavaco e Ribeira de Pombal). (b) Loading plot graph.

The regression models and equations obtained to estimate the individual and grouped leaflet area of the four genotypes of *S. tuberosa* are presented in Table 1. According to the selection criteria, model 4 (linear regression without intercept) and model 7 (power), using LW, were the most appropriate and accurate to estimate the leaflet area of genotypes 1 (Esperança), 2 (Macaúbas), 3 (Livramento Cavaco), and 4 (Ribeira de Pombal), presenting the highest values of R^2, r, d, and lower values of AIC and RMSE (Table 1).

Table 1. Regression model, equations, coefficient of determination (R^2), Pearson correlation coefficient (r), Akaike information criterion (AIC), Willmott's concordance index (*d*) and mean squared error (RMSE) obtained as a function of measurements of leaflet dimensions of four genotypes of *Spondias tuberosa* Arruda (Esperança, Macáubas, Livramento Cavaco, and Ribeira de Pombal).

Equation code	Model	Coefficients		R^2	r	AIC	*d*	RMSE	Estimator of LA ($\hat{y}$)
		β_0	β_1						
Genotype 1 (Esperança)									
1	$\hat{y} = \beta_0 + \beta_1 \cdot L + \varepsilon_i$	-3.557	2.666	0.8334	0.9132	711.2	0.9533	0.9804	$\hat{y} = -3.557 + 2.666 \cdot L$
2	$\hat{y} = \beta_0 + \beta_1 \cdot W + \varepsilon_i$	-5.202	4.872	0.9124	0.9553	549.2	0.9767	0.7109	$\hat{y} = -5.202 + 4.872 \cdot W$
3	$\hat{y} = \beta_0 + \beta_1 \cdot LW + \varepsilon_i$	0.1879	0.658	0.9780	0.9889	200.9	0.9944	0.3562	$\hat{y} = 0.1879 + 0.6589 \cdot LW$
4	$\hat{y} = \beta_1 \cdot LW + \varepsilon_i$	----	0.674	0.9978	0.9889	205.7	0.9944	0.3611	$\hat{y} = 0.6744 \cdot LW$
5	$\hat{y} = \beta_0 \cdot L^{\beta 1} + \varepsilon_i$	0.8128	1.550	0.8344	0.9135	710.6	0.9530	0.9794	$\hat{y} = 0.8128 \cdot L^{1.5509}$
6	$\hat{y} = \beta_0 \cdot W^{\beta 1} + \varepsilon_i$	1.256	1.840	0.9180	0.9581	533.6	0.9780	0.6893	$\hat{y} = 1.256 \cdot W^{1.840}$
7	$\hat{y} = \beta_0 \cdot LW^{\beta 1} + \varepsilon_i$	0.7352	0.965	0.9783	0.9891	198.2	0.9944	0.3543	$\hat{y} = 0.7352 \cdot LW^{0.9659}$
8	$\hat{y} = \beta_0 \cdot \beta_1^{L} + \varepsilon_i$	1.544	1.450	0.8166	0.9036	737.5	0.9459	1.0330	$\hat{y} = 1.544 \cdot 1.450^{L}$
9	$\hat{y} = \beta_0 \cdot \beta_1^{W} + \varepsilon_i$	1.143	2.021	0.9030	0.9502	577.9	0.9729	0.7526	$\hat{y} = 1.143 \cdot 2.021^{W}$
10	$\hat{y} = \beta_0 \cdot \beta_1^{LW} + \varepsilon_i$	2.863	1.087	0.9030	0.9502	512.4	0.9729	0.7526	$\hat{y} = 2.863 \cdot 1.087^{LW}$
Genotype 2 (Macaúbas)									
1	$\hat{y} = \beta_0 + \beta_1 \cdot L + \varepsilon_i$	-3.813	2.519	0.9092	0.9537	527.7	0.9758	0.7048	$\hat{y} = -3.813 + 2.519 \cdot L$
2	$\hat{y} = \beta_0 + \beta_1 \cdot W + \varepsilon_i$	-4.420	4.777	0.9495	0.9745	384.5	0.9869	0.5255	$\hat{y} = -4.420 + 4.777 \cdot W$
3	$\hat{y} = \beta_0 + \beta_1 \cdot LW + \varepsilon_i$	0.2495	0.650	0.9788	0.9894	172.6	0.9946	0.3404	$\hat{y} = 0.2495 + 0.6507 \cdot LW$
4	$\hat{y} = \beta_1 \cdot LW + \varepsilon_i$	----	0.676	0.9965	0.9894	191.0	0.9944	0.3549	$\hat{y} = 0.6765 \cdot LW$
5	$\hat{y} = \beta_0 \cdot L^{\beta 1} + \varepsilon_i$	0.5723	1.706	0.9098	0.9538	527.8	0.9754	0.7050	$\hat{y} = 0.5723 \cdot L^{1.7062}$
6	$\hat{y} = \beta_0 \cdot W^{\beta 1} + \varepsilon_i$	1.323	1.891	0.9630	0.9813	309.3	0.9904	0.4504	$\hat{y} = 1.323 \cdot W^{1.891}$
7	$\hat{y} = \beta_0 \cdot LW^{\beta 1} + \varepsilon_i$	0.7729	0.943	0.9800	0.9899	160.4	0.9948	0.3320	$\hat{y} = 0.7729 \cdot LW^{0.9430}$
8	$\hat{y} = \beta_0 \cdot \beta_1^{L} + \varepsilon_i$	1.132	1.509	0.8798	0.9380	601.0	0.9651	0.8190	$\hat{y} = 1.132 \cdot 1.509^{L}$
9	$\hat{y} = \beta_0 \cdot \beta_1^{W} + \varepsilon_i$	0.9257	2.279	0.9458	0.9725	408.0	0.9850	0.5515	$\hat{y} = 0.9257 \cdot 2.2795^{W}$
10	$\hat{y} = \beta_0 \cdot \beta_1^{LW} + \varepsilon_i$	2.452	1.099	0.9458	0.9725	548.6	0.9850	0.5515	$\hat{y} = 2.452 \cdot 1.099^{LW}$
Genotype 3 (Livramento Cavaco)									
1	$\hat{y} = \beta_0 + \beta_1 \cdot L + \varepsilon_i$	-3.452	2.709	0.8791	0.9378	557.0	0.9670	0.7252	$\hat{y} = -3.452 + 2.709 \cdot L$
2	$\hat{y} = \beta_0 + \beta_1 \cdot W + \varepsilon_i$	-4.069	4.233	0.9259	0.9623	434.1	0.9805	0.5677	$\hat{y} = -4.069 + 4.233 \cdot W$
3	$\hat{y} = \beta_0 + \beta_1 \cdot LW + \varepsilon_i$	0.0753	0.703	0.9840	0.9920	49.0	0.9959	0.2636	$\hat{y} = 0.0753 + 0.7037 \cdot LW$
4	$\hat{y} = \beta_1 \cdot LW + \varepsilon_i$	----	0.712	0.9977	0.9920	49.9	0.9959	0.2651	$\hat{y} = 0.7126 \cdot LW$
5	$\hat{y} = \beta_0 \cdot L^{\beta 1} + \varepsilon_i$	0.7365	1.661	0.8740	0.9349	569.1	0.9645	0.7429	$\hat{y} = 0.7365 \cdot L^{1.6616}$
6	$\hat{y} = \beta_0 \cdot W^{\beta 1} + \varepsilon_i$	1.124	1.904	0.9364	0.9677	396.5	0.9833	0.5267	$\hat{y} = 1.124 \cdot W^{1.904}$
7	$\hat{y} = \beta_0 \cdot LW^{\beta 1} + \varepsilon_i$	0.7554	0.973	0.9845	0.9922	43.0	0.9960	0.2604	$\hat{y} = 0.7554 \cdot LW^{0.9734}$
8	$\hat{y} = \beta_0 \cdot \beta_1^{L} + \varepsilon_i$	1.118	1.593	0.8344	0.9135	641.0	0.9496	0.8573	$\hat{y} = 1.118 \cdot 1.593^{L}$
9	$\hat{y} = \beta_0 \cdot \beta_1^{W} + \varepsilon_i$	0.7795	2.302	0.9262	0.9624	435.6	0.9799	0.5694	$\hat{y} = 0.7795 \cdot 2.3020^{W}$
10	$\hat{y} = \beta_0 \cdot \beta_1^{LW} + \varepsilon_i$	2.141	1.122	0.9262	0.9624	478.3	0.9799	0.5694	$\hat{y} = 2.141 \cdot 1.122^{LW}$
Genotype 4 (Ribeira de Pombal)									
1	$\hat{y} = \beta_0 + \beta_1 \cdot L + \varepsilon_i$	-6.486	3.510	0.9054	0.9517	925.2	0.9747	1.5327	$\hat{y} = -6.486 + 3.510 \cdot L$
2	$\hat{y} = \beta_0 + \beta_1 \cdot W + \varepsilon_i$	-7.341	6.107	0.9175	0.9580	891.4	0.9781	1.4320	$\hat{y} = -7.341 + 6.107 \cdot W$
3	$\hat{y} = \beta_0 + \beta_1 \cdot LW + \varepsilon_i$	0.1950	0.677	0.9880	0.9939	412.2	0.9969	0.5471	$\hat{y} = 0.1950 + 0.6778 \cdot LW$
4	$\hat{y} = \beta_1 \cdot LW + \varepsilon_i$	----	0.687	0.9980	0.9939	415.5	0.9969	0.5529	$\hat{y} = 0.6878 \cdot LW$
5	$\hat{y} = \beta_0 \cdot L^{\beta 1} + \varepsilon_i$	0.7171	1.675	0.9075	0.9526	921.2	0.9748	1.5203	$\hat{y} = 0.7171 \cdot L^{1.6759}$
6	$\hat{y} = \beta_0 \cdot W^{\beta 1} + \varepsilon_i$	1.195	1.959	0.9463	0.9728	785.3	0.9860	1.1571	$\hat{y} = 1.195 \cdot W^{1.959}$
7	$\hat{y} = \beta_0 \cdot LW^{\beta 1} + \varepsilon_i$	0.7513	0.970	0.9882	0.9941	406.6	0.9970	0.5409	$\hat{y} = 0.7513 \cdot LW^{0.9708}$
8	$\hat{y} = \beta_0 \cdot \beta_1^{L} + \varepsilon_i$	2.125	1.369	0.8813	0.9387	986.8	0.9656	1.7342	$\hat{y} = 2.125 \cdot 1.369^{L}$
9	$\hat{y} = \beta_0 \cdot \beta_1^{W} + \varepsilon_i$	1.546	1.859	0.9353	0.9671	835.6	0.9822	1.2802	$\hat{y} = 1.546 \cdot 1.859^{W}$
10	$\hat{y} = \beta_0 \cdot \beta_1^{LW} + \varepsilon_i$	4.308	1.056	0.9353	0.9671	879.2	0.9822	1.2802	$\hat{y} = 4.308 \cdot 1.056^{LW}$
10	$\hat{y} = \beta_0^{*} \beta_1^{LW} + \varepsilon_i$	3.482	1.067	0.9393	0.9691	4061.2	0.9833	1.0109	$\hat{y} = 3.482 \cdot 1.067^{LW}$

continue...

Table 1. Continuation...

Equation code	Model	Coefficients		R^2	r	AIC	*d*	RMSE	Estimator of LA (ŷ)
		β_0	β_1						
				Pooled group					
1	$\hat{y} = \beta_0 + \beta_1 \cdot L + \varepsilon_i$	-5.489	3.243	0.8925	0.9447	4219.0	0.9709	1.3321	$\hat{y} = -5.489 + 3.243 \cdot L$
2	$\hat{y} = \beta_0 + \beta_1 \cdot W + \varepsilon_i$	-6.418	5.529	0.9175	0.9600	3829.3	0.9792	1.1376	$\hat{y} = -6.418 + 5.529 \cdot W$
3	$\hat{y} = \beta_0 + \beta_1 \cdot LW + \varepsilon_i$	0.0783	0.685	0.9890	0.9945	1399.4	0.9972	0.4253	$\hat{y} = 0.0783 + 0.6855 \cdot LW$
4	$\hat{y} = \beta_1 \cdot LW + \varepsilon_i$	----	0.691	0.9975	0.9945	1406.8	0.9972	0.4269	$\hat{y} = 0.6911 \cdot LW$
5	$\hat{y} = \beta_0 \cdot L^{\beta 1} + \varepsilon_i$	0.6198	1.757	0.9075	0.9521	4047.9	0.9748	1.2429	$\hat{y} = 0.6198 \cdot L^{1.7576}$
6	$\hat{y} = \beta_0 \cdot W^{\beta 1} + \varepsilon_i$	1.124	1.976	0.9521	0.9757	3223.0	0.9876	0.8900	$\hat{y} = 1.124 \cdot W^{1.976}$
7	$\hat{y} = \beta_0 \cdot LW^{\beta 1} + \varepsilon_i$	0.7127	0.988	0.9890	0.9945	1397.1	0.9972	0.4249	$\hat{y} = 0.7127 \cdot LW^{0.9888}$
8	$\hat{y} = \beta_0 \cdot \beta_1^{L} + \varepsilon_i$	1.622	1.436	0.8786	0.9373	4394.1	0.9645	1.4299	$\hat{y} = 1.622 \cdot 1.436^{L}$
9	$\hat{y} = \beta_0 \cdot \beta_1^{W} + \varepsilon_i$	1.245	1.958	0.9393	0.9691	3537.5	0.9833	1.0109	$\hat{y} = 1.245 \cdot 1.958^{W}$
10	$\hat{y} = \beta_0 \cdot \beta_1^{LW} + \varepsilon_i$	3.482	1.067	0.9393	0.9691	4061.2	0.9833	1.0109	$\hat{y} = 3.482 \cdot 1.067^{LW}$

The genotype grouping showed results similar to the ones from the individual analysis, in which the best models were 4 (linear regression without intercept), and 7 (power) (Table 2). In these models, the R^2 was higher than 0.98, suggesting that at least 98% of the variations in the leaflet area of *S. tuberosa* were explained by the adjusted equations. Thus, the best estimate of the leaflets area of the species can be obtained by the equations $\hat{y} = 0.6911 \cdot LW$ and $\hat{y} = 0.7127 \cdot LW^{0.9888}$ (Table 1).

Based only on the L and W linear dimensions of the leaflets, models 1 and 2 (simple linear regression), 5 and 6 (power), and 8 and 9 (exponential) showed the lowest R^2 values when the genotypes were evaluated individually or grouped, which confirms that the use of LW better fits these models compared to the isolated use of L or W (Table 1).

The visual analysis of the dispersion of residues indicates that the linear model without intercept and the power model have a positive relationship between LA and LW, with a low dispersion of the data and a residual homogeneity, confirming the applicability of the proposed models (R^2 = 0.9975 and 0.9891) (Fig. 5).

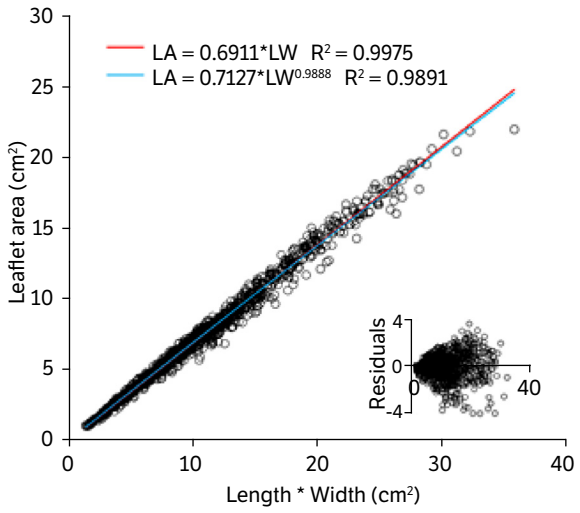

Figure 5. Relationship between the observed leaflet area (LA) and the product between length and width (LW) of *Spondias tuberosa* leaves (pooled data) from the linear models without intercept and power. The dispersion analysis of the residues is presented in the insertion.

The chosen models (power and linear without intercept) to estimate the LA of *S. tuberosa* genotypes showed a high correlation with observed LA (R^2 = 0.9891) (Figs. 6a and 6c). It was verified that there was no significant difference between the LA estimated by the chosen models and the observed LA (Figs. 6b and 6d). Therefore, the power and linear models without intercept can be used to estimate the LA of *S. tuberosa* using the LW product.

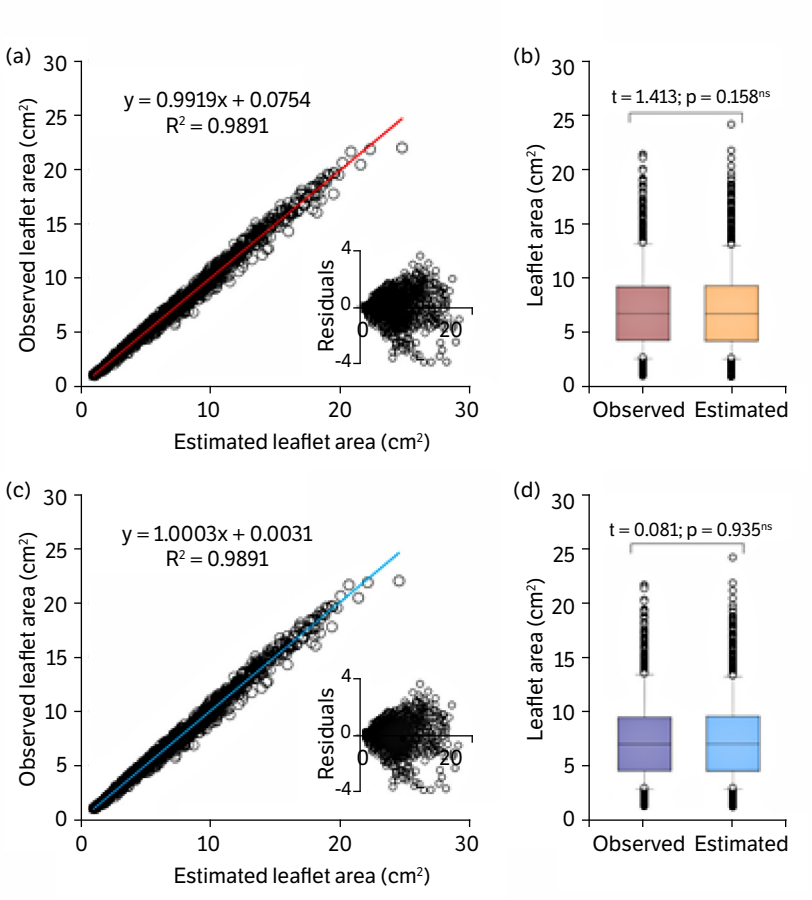

Figure 6. Relationship and comparison of observed leaflet area and estimated leaflet area using (a and b) the linear model without intercept, and (c and d) power as a function of the product of the width and length of leaves of umbuzeiro (*Spondias tuberosa*).

DISCUSSION

The studied genotypes of *S. tuberosa* showed greater L than W. Only Ribeira de Pombal had similar L and W values. This LW ratio is critical, as it is responsible for the constitution of the observed LA, influencing leaf size and active photosynthetic area, directly interfering in the photosynthetic capacity of the species (Li et al. 2020). There was a linear association for LW. As a result, linear and nonlinear models of the linear type without intercept and power were generated and tested to estimate the LA in each linear dimension. A similar result was obtained by Ribeiro et al. (2020) evaluating non-destructive methods for determining leaf area in *Erythroxylum pauferrense* Plowman, in which they observed linear and nonlinear adjustments for the variables L, W, LW, and observed LA.

Variation between leaflets dimensions is common, especially among genotypes of the same species. This represents a survival strategy in species native to the semi-arid northeast, which helps in conditions of biotic and abiotic stress. The significant variability found among the studied genotypes is considered a positive factor for this study, because it has different leaf sizes, can indicate good data distribution, and thus obtains models with greater representativeness and precision for each genotype (Shi et al. 2019).

PC analysis revealed that the four genotypes had common characteristics since PC1 and PC2 showed maximum overlap and no distinct separation (Fig. 4a), reinforcing the low variation in dimensions between *S. tuberosa* leaflets, considered a beneficial characteristic. In Fig. 4b, the angle of the vectors reflected the relationship between LA and the dimensions L, W, and LW, exhibiting a positive correlation since the angle between them did not exceed 90º, indicating a negative correlation

(Al-Naggar et al. 2020). However, the dimension that most influence the grouping of genotypes and determination of LA for the species is LW, precisely because of its proximity to LA.

The results also showed that the individual and grouped LA of the *S. tuberosa* genotypes can be estimated from the linear regression models without intercept and power derived from the LW of the leaflets (Table 1). These models were more accurate and reliable based on the higher quality of the data and more adjusted parameters, including R^2, which was higher than 0.97. According to Williams and Martinson (2003), the accepted models should have R^2 greater than 0.95, indicating greater predictive capacity and lower dispersion. The results of the present study corroborate those observed for *Vitis vinifera* (Buttaro et al. 2015), *Eustoma grandiflorum* (Dias et al. 2022), and *Malus domestica* (Boyaci and Küçükönder 2022), in which the most appropriate models were the linear model without intercept and power to estimate LA.

The study proved that, when using LW, the relationship becomes linear and with better criteria, suggesting more efficiency in estimating the LA of *S. tuberosa*, compared only to single dimensions, except for the 9 (exponential) model, that used W. A similar result was observed by Goergen et al. (2021) when they studied the allometric relationship and modeling of LA estimation by the non-destructive method in *Salvia hispanica* culture.

It is noteworthy that, in these equations with only L or W, there were greater practicality and speed in data collection in the field, suggesting savings in the number and time of measurements (Santos et al. 2016), but they also had low precision and fit the models (Bosco et al. 2012, Souza et al. 2015), since it did not obtain a LA that represents the entire LA of the species. This is consistent with results found by several authors for crops such as *Moringa oleifera* (Macário et al. 2020), *Juglans regia* (Keramatlou et al. 2015), *S. hispanica* (Goergen et al. 2021), and *M. domestica* (Boyaci and Küçükönder 2022). Pohlmann et al. (2021), in a study predicting the LA of common bean (*Phaseolus vulgaris*), found that the equations that best fit were those with the product (LW), with R^2 above 0.94.

It was possible to adjust a single equation (general model) for each chosen model for the species due to the morphological similarities of the leaflets of the *S. tuberosa* genotypes analyzed, already confirmed by PC analysis. These proposed equations ($\hat{y} = 0.6911 \cdot LW$ and $\hat{y} = 0.7127 \cdot LW^{0.9888}$) can achieve cost-effective measurements, allowing farmers or researchers to cheaply, quickly, and reliably perform non-destructive or repeated measurements for crop LA determination.

According to Guimarães et al. (2019), when working with many accessions or genotypes that have not yet been studied, mathematical equations involving groups of genotypes are highly relevant. In addition, the calibration of the model based on a large number of genotypes is immensely important since the shape of the leaf can vary between different genetic materials (Rouphael et al. 2010); however, in the research, there was no such significant variation.

Thus, the equations described will be of great value to facilitate future studies in the area of phytopathology, agronomy, and physiological growth of *S. tuberosa*, considering that LA is one of the most critical measures to evaluate vegetative growth and estimate the yield potential of the crop, because it is linked to the interception of light by the photosynthetic apparatus, conversion into chemical energy and, consequently, an increase in plant dry matter (Keramatlou et al. 2015, Taiz et al. 2017, Goergen et al. 2021).

The results revealed that the LA observed in the chosen models (power and linear without intercept) showed a high correlation with the estimated LA, with R^2 close to 1. This information reaffirms that the models can be used to estimate the LA of *S. tuberosa* using the LW product. Similar results were found for *Sesamum indicum* (Ribeiro et al. 2023), *Fagopyrum esculentum* (Cargnelutti Filho et al. 2021), and *Cajanus cajan* (Cargnelutti Filho et al. 2015).

CONCLUSION

PC analysis revealed that the genotypes of *S. tuberosa* were similar in terms of biometric parameters. LW was the biometric parameter that provided the best adjustments of the linear models without intercept and power.

The LA of genotypes 1 (Esperança), 2 (Macaúbas), 3 (Livramento Cavaco), and 4 (Ribeira de Pombal) can be estimated from the equations $\hat{y} = 0.6911 \cdot LW$ and $\hat{y} = 0.7127 \cdot LW^{0.9888}$ with precision.

CONFLICT OF INTEREST

Nothing to declare.

AUTHORS' CONTRIBUTION

Conceptualization: Amorim, P. E. C. and Ribeiro, J. E. S.; **Formal Analysis:** Pereira, D. F. and Freire, R. I. S.; **Funding Acquisition:** Mendonça, V. and Ribeiro J. E. S.; **Investigation: A**morim, P. E. C., Pereira, D. F., Freire, R. I. S. and Oliveira, A. M. F.; **Methodology:** Ribeiro, J. E. S.; **Project Administration:** Mendonça, V. and Ribeiro, J. E. S.; Resources: Ribeiro, J. E. S.; **Supervision:** Ribeiro, J. E. S.; **Validation:** Ribeiro, J. E. S.; **Visualization:** Mendonça, V. and Ribeiro, J. E. S; **Writing – Original Draft:** Amorim, P. E. C., Pereira, D. F., Freire, R. I. S., Oliveira, A. M. F., Mendonça, V. and Ribeiro, J. E. S.; **Writing – Review and Editing:** Amorim, P. E. C., Pereira, D. F., Freire, R. I. S., Oliveira, A. M. F., Mendonça, V. and Ribeiro, J. E. S.

DATA AVAILABILITY STATEMENT

All dataset were generated and analyzed in the current study.

FUNDING

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
Finance Code 001

ACKNOWLEDGMENTS

We thank the Universidade Federal Rural do Semi-Árido for technical support during the research.

REFERENCES

Adji, B. I., Akaffou, D. S., Kouassi, K. H., Houphouet, Y. P., De Reffye, P., Duminil, J., Jaeger, M. and Sabatier, S. (2021). Allometric models for non-destructive estimation of dry biomass and leaf area in *Khaya senegalensis* (Desr.) A. Juss., 1830 (Meliaceae), Pterocarpus erinaceus Poir., 1804 (Fabaceae) and *Parkia biglobosa*, Jack, R. Br., 1830 (Fabaceae). Trees, 35, 1905-1920. https://doi.org/10.1007/s00468-021-02159-y

Al-Naggar, A. M. M., Shafik, M. M. and Musa, R. Y. M. (2020). Genetic diversity based on morphological traits of 19 maize genotypes using principal component analysis and GT biplot. Annual Research & Review in Biology, 35, 68-85. https://doi.org/10.9734/arrb/2020/v35i230191

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. D. M. and Sparovek, G. (2013). Köppen's climate classification map for Brazil. Meteorologische Zeitschrift, 22, 711-728. https://doi.org/10.1127/0941-2948/2013/0507

Bosco, L., Bergamaschi, H., Cardoso, L. S., Paula, V. A. and Casamali, B. (2012). Seleção de modelos de regressão para estimar a área foliar de macieiras 'Roayal Gala' e 'Fuji Suprema' sob tela antigranizo e em céu aberto. Revista Brasileira de Fruticultura, 34, 504-514. https://doi.org/10.1590/S0100-29452012000200024

Boyaci, S. and Küçükönder, H. A. (2022). A research on non-destructive leaf area estimation modeling for some apple cultivars. Erwerbs-Obstbau, 64, 1-7. https://doi.org/10.1007/s10341-021-00619-w

Buttaro, D., Rouphael, Y., Rivera, C. M., Colla, G. and Gonnella, M. L. (2015). Simple and accurate allometric model for leaf area estimation in *Vitis vinifera* L. genotypes. Photosynthetica, 53, 342-348. https://doi.org/10.1007/s11099-015-0117-2

Cargnelutti Filho, A., Pezzini, R. V., Neu, I. M. M. and Dumke, G. E. (2021). Estimation of buckwheat leaf area by leaf dimensions. Semina Ciências Agrárias, 42, 1529-1548. https://doi.org/10.5433/1679-0359.2021v42n3Supl1p1529

Cargnelutti Filho, A., Toebe, M., Alves, B. M. and Burin, C. (2015). Leaf area estimation of pigeonpea by leaf dimensions. Ciência Rural, 45, 1-8. https://doi.org/10.1590/0103-8478cr20140551

[Climate-Data] (2023). Dados climáticos para cidades mundiais. Clima Mossoró-RN, 2023. Available at: https://pt.climate-data.org/america-do-sul/brasil/rio-grande-do-norte/mossoro-4448/. Accessed on: Nov. 16, 2022.

Dias, M. G., Silva, T. I. D., Ribeiro, J. E. D. S., Grossi, J. A. S. and Barbosa, J. G. (2022). Allometric models for estimating the leaf area of lisianthus (*Eustoma grandiflorum*) using a non-destructive method. Revista Ceres, 69, 7-12. https://doi.org/10.1590/0034-737X202269010002

Goergen, P. C., Lago, I., Schwab, N. T., Alves, A. F., Freitas, C. P. D. O., and Selli, V. S. (2021). Allometric relationship and leaf area modeling estimation on chia by non-destructive method. Revista Brasileira de Engenharia Agrícola e Ambiental, 25, 305-311. https://doi.org/10.1590/1807-1929/agriambi.v25n5p305-311

Guimarães, M. J. M., Coelho Filho, M. A., Gomes Junior, F. A., Silva, M. A. M., Alves, C. V. O. and Lopes, I. (2019). Modelos matemáticos para a estimativa da área foliar de mandioca. Revista de Ciências Agrárias, 62, 1-5. https://doi.org/10.22491/rca.2019.3015

He, J., Reddy, G. V., Liu, M. and Shi, P. (2020). A general formula for calculating surface area of the similarly shaped leaves: Evidence from six Magnoliaceae species. Global Ecology and Conservation, 23, e01129. https://doi.org/10.1016/j.gecco.2020.e01129

Keramatlou, I., Sharifani, M., Sabouri, H., Alizadeh, M. and Kamkar, B. (2015). A simple linear model for leaf area estimation in Persian walnut (*Juglans regia* L.). Scientia Horticulturae, 184, 36-39. https://doi.org/10.1016/j.scienta.2014.12.017

Legris, M. (2023). Light and temperature regulation of leaf morphogenesis in Arabidopsis. New Phytologist, 240, 2191-2196. https://doi.org/10.1111/nph.19258

Li, Y., Reich, P. B., Schmid, B., Shrestha, N., Feng, X., Lyu, T., Maitner, B. S., Xu, X., Li, Y. C., Zou, D., Tan, Z. H., Su, X., Tang, Z., Guo, Q., Feng, X., Enquist, B. J. and Wang, Z. (2020). Leaf size of woody dicots predicts ecosystem primary productivity. Ecology Letters, 23, 1003-1013. https://doi.org/10.1111/ele.13503

Lima, M. A., Silva, S. M., Oliveira, V. R., Oliveira, S. E. and Brito, E. S. (2018). Umbu: *Spondias tuberosa*. In S. Rodrigues, E. O. Silva and E. S. Brito (Eds.). Exotic Fruits (p. 427-433). London: Academic Press. https://doi.org/10.1016/B978-0-12-803138-4.00057-5

Lusk, C. H., Grierson, E. R. and Laughlin, D. C. (2019). Large leaves in warm, moist environments confer an advantage in seedling light interception efficiency. New Phytologist, 223, 1319-1327. https://doi.org/10.1111/nph.15849

Macário, A. P. S., Ferraz, R. L. D. S., Costa, P. D. S., Brito Neto, J. F. D., Melo, A. S. D. and Dantas Neto, J. (2020). Allometric models for estimating *Moringa oleifera* leaflets area. Ciência e Agrotecnologia, 44, 005220. https://doi.org/10.1590/1413-7054202044005220

Menezes, P. H. S., Souza, A. A., Silva, E. S., Medeiros, R. D., Barbosa, N. C. and Soria, D. G. (2017). Influência do estádio de maturação na qualidade físico-química de frutos de umbu (*Spondias tuberosa*). Scientia Agropecuaria, 8, 73-78. https://doi.org/10.17268/sci.agropecu.2017.01.07

Pohlmann, V., Lago, I., Lopes, S. J., Martins, J. T. D. S., Rosa, C. A. D., Caye, M. and Portalanza, D. (2021). Estimation of common bean (*Phaseolus vulgaris*) leaf area by a non-destructive method. Semina Ciências Agrárias, 42, 2163-2180. https://doi.org/10.5433/1679-0359.2021v42n4p2163

R Core Team (2023). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Ribeiro, J. E. D. S., Coêlho, E. D. S., Figueiredo, F. R. A. and Melo, M. F. (2020). Non-destructive method to estimate leaf area of *Erythroxylum pauferrense* (Erythroxylaceae) from linear dimensions of leaf blades. Acta Botánica Mexicana, 127, e1717. https://doi.org/10.21829/abm127.2020.1717

Ribeiro, J. E. D. S., Coêlho, E. S., Oliveira, A. K. S., Silva, A. G. C., Lopes, W. D. A. R., Oliveira, P. H. A. and Silveira, L. M. (2023). Artificial neural network approach for predicting the sesame (*Sesamum indicum* L.) leaf area: A non-destructive and accurate method. Heliyon, 7, e17834. https://doi.org/10.1016/j.heliyon.2023.e17834

Ribeiro, J. E. D. S., Coêlho, E. D. S., Pessoa, Â. M. D. S., Oliveira, A. K., Oliveira, A. M., Barros Júnior, A. P. and Nunes, G. H. D. S. (2022). Nondestructive method for estimating the leaf area of sapodilla from linear leaf dimensions. Revista Brasileira de Engenharia Agrícola e Ambiental, 27, 209-215. https://doi.org/10.1590/1807-1929/agriambi.v27n3p209-215

Rouphael, Y., Mouneimne, A. H., Rivera, C. M., Cardarelli, M., Marucci, A. and Colla, G. (2010). Allometric models for non-destructive leaf area estimation in grafted and ungrafted watermelon (*Citrullus lanatus* Thunb.). Journal of Food, Agriculture & Environment, 8, 161-165.

Salazar, J. C. S., Melgarejo, L. M., Bautista, E. H. D., Di Rienzo, J. A. and Casanoves, F. (2018). Non-destructive estimation of the leaf weight and leaf area in cacao (*Theobroma cacao* L.). Scientia Horticulturae, 229, 19-24. https://doi.org/10.1016/j.scienta.2017.10.034

Santos, J. C. C., Costa, R. N., Silva, D. M. R., Souza, A. A., Moura, F. B. P., Silva Junior, J.M. and Silva, J. V. (2016). Use of allometric models to estimate leaf area in *Hymenaea courbaril* L. Theoretical and Experimental Plant Physiology, 28, 357-369. https://doi.org/10.1007/s40626-016-0072-8

Schmildt, E. R., Oliveira, V. S. and Arantes, S. D. (2023). Modelagem da área foliar individual. São José dos Pinhais: Brazilian Journals. Available at: https://biblioteca.incaper.es.gov.br/digital/bitstream/item/4457/1/modelagemdaareafoliarindividual.pdf. Accessed on: Nov. 18, 2022.

Shapiro, S. S. and Wilk, M. B. (1965). Analysis of variance test for normality. Biometrika, 52, 591-611. https://doi.org/10.2307/2333709

Shi, P., Liu, M., Yu, X., Gieli, J. and Ratkowsky, D. A. (2019). Proportional relationship between leaf area and the product of leaf length and width of four types of special leaf shapes. Forests, 10, 178. https://doi.org/10.3390/f10020178

Souza, M. C., Amaral, C. L., Habermann, G., Costa, A., Alves, P. L. and Costa, F. B. (2015). Non-destructive model to estimate the leaf area of multiple Vochysiaceae species. Brazilian Journal of Botany, 38, 903-909. https://doi.org/10.1007/s40415-015-0176-4

Taiz, L., Zeiger, E., Møller, I. M. and Murphy, A. (2017). Fisiologia e desenvolvimento vegetal. Porto Alegre: Artmed.

Williams, L. and Martinson, T. E. (2003). Nondestructive leaf area estimation of 'Niagara' and 'DeChaunac' grapevines. Scientia Horticulturae, 98, 493-498. https://doi.org/10.1016/S0304-4238(03)00020-7