horticultura brasileira

Research

HERNANDÉZ-FERNANDÉZ, IA; JARMA-OROZCO, A; POMPELLI, MF. 2021. Allometric models for non-destructive leaf area measurement of stevia: an in depth and complete analysis. *Horticultura Brasileira* 39: 205-215. Doi: http://dx.doi.org/10.1590/s0102-0536-20210212

Allometric models for non-destructive leaf area measurement of stevia: an in depth and complete analysis

Israel A Hernández-Fernandéz [1]; Alfredo Jarma-Orozco [1]; Marcelo F Pompelli [2]*

[1]Regional Group of Participatory Research of Small Farmers of the Colombian Atlantic Coast INVEPAR, Faculty of Agricultural Sciences, University of Cordoba, Montería, Córdoba, Colombia; ajarma@correo.unicordoba.edu.co; [2]Laboratório de Ecofisiologia Vegetal, Departamento de Botânica, Centro de Biociências, Universidade Federal de Pernambuco (UFPE), Recife-PE, Brasil; mfpompelli@gmail.com (*corresponding author)

ABSTRACT

Leaf area measurement is pivotal for plant physiologists. Hence, accurate measurement of their leaf area is incredibly relevant in agronomic terms. The plant *Stevia rebaudiana* is a sucrose-free plant species that is now vital to the global production of sucrose-free foods. Here, we estimated *S. rebaudiana* leaf area using a nondestructive methodology comprising allometric equations. Through leaf length (L), leaf width (W), and/or their product (LW) the leaf area was determined. One thousand leaves were sampled from four distinct *S. rebaudiana* genotypes for model construction. Linear or power models were generated, and the best equation was selected using a statistical criterion. The statistical criteria indicated that the linear models best suited all genotypes tested, included a function of LW, exhibited increased stability, and precisely estimated coefficients. ANOVA revealed that both generalized and combined equations were feasible. Nevertheless, grouping all genotypes into a single model was not possible as the genotype leaf architectures were very dissimilar.

Keywords: *Stevia rebaudiana*, general allometric model, leaf area estimation, linear and power models, morphological leaf traits, statistical analysis.

RESUMO

Modelos alométricos não destrutivos para estimação da área foliar de stevia: uma análise completa e profunda

A medição da área foliar é de extrema importância para os estudos de fisiologia vegetal. Assim, é fundamental a medição precisa de sua área foliar nos estudos agronômicos. *Stevia rebaudiana* é uma espécie produtora de esteviosídeo, uma substância utilizada como adoçante natural, mundialmente conhecida em produtos *sacarose-free*. A área foliar de *S. rebaudiana* foi estimada por metodologia não destrutiva compreendendo equações alométricas que foram determinadas através do comprimento da folha (L), largura da folha (W) e / ou seu produto (LW). Foram amostradas mil folhas de quatro genótipos distintos de *S. rebaudiana* para construção do modelo. Modelos lineares e não lineares foram gerados e a melhor equação foi selecionada usando-se critérios estatísticos robustos, os quais indicaram que os modelos lineares se adequavam melhor a todos os genótipos testados, exibiam maior estabilidade e precisos coeficientes estimados. A ANOVA revelou que a equação generalizada é viável; no entanto, o agrupamento de todos os genótipos em um único modelo é contraindicado, pois as arquiteturas genotípicas das folhas são muito distintas.

Palavras-chave: *Stevia rebaudiana*, modelo alométrico geral, alometria foliar, modelos lineares e potenciais, características morfológicas foliares, análises estatísticas, plantas livres de açúcares.

Received on September 18, 2020; accepted on April 21, 2021

Stevia rebaudiana is a small perennial plant, reaching a height of 65-80 cm. Various stevia species contain sweetening compounds, such as diterpene glycosides but *S. rebaudiana* has the highest concentrations of these substances, whose are of considerable interest for sucrose-free food production. Stevia plants contain steviol glycosides, that are ~300 times sweeter than sucrose at 0.4% (w/v) (Kinghorn & Soejarto, 1991). The incidences of cardiovascular diseases, diabetes, hypercholesterolemia, and obesity have increased globally (Guerrero *et al.*, 2018). Consequently, the demand for non-caloric sweeteners has increased and sugar consumption has decreased (Putnik *et al.*, 2020). In addition to being natural sweeteners, stevia leaf lowers blood pressure, improves gastrointestinal function, and protects against dental caries (Kinghorn & Soejarto, 1991). The growing worldwide demand for natural sweeteners has induced large-scale *S. rebaudiana* cultivation.

Humankind has learned to cultivate various plant species. The leaves of horticultural plants, such as lettuce, watercress, celery, chard, and spinach may be directly consumed. Other horticultural plants such as basil, marjoram, parsley, coriander, onion, and bay leaf are cultivated for condiments and food preservation. Still other horticultural plants such as lemongrass, chamomile, lemon, orange, eucalyptus, and stevia leaves can be consumed as

such, see *e.g.*, their use in Paraguay and other countries where they are used as tea or for sweetening purposes but are sources of commercially important compounds. Hence, in the leaves of these species are the main organs of commercial interest. For this reason, a major agroeconomic objective in the cultivation of these plants is to maximize their leaf area (LA). The leaf is the main photosynthetic organ for most plants. Thus, accurate LA determination is vital to the production of these crops (Antunes *et al.*, 2008; Pompelli *et al.*, 2012). It is not possible to confirm that greater amounts of sweetening compounds are directly proportional to the leaf development. However, it is quite plausible to argue that plants with a larger leaf area are more likely to produce a greater amount of sweeteners, since the synthesis of sweeteners starts in the stroma of the chloroplasts (Totté *et al.*, 2000; Ceunen & Geuns, 2013c). Larger leaf area denotes more chloroplasts and consequently more sweeteners can be produced. Notwithstanding, LA measurement is critical in ecophysiological, agricultural, and ecological research (Pandey, 2011). LA has been widely used to describe growth, productivity, photosynthetic efficiency, soil salinity and acidity, heat transfer, carbon, nutrients, and water exchange into the atmosphere. In turn, these properties influence crop yield (Antunes *et al.*, 2008; Pompelli *et al.*, 2012, 2019).

The direct measurement of individual LA is both, laborious, expensive, time-consuming, and constrained by logistical factors. A modeling approach is crucial for the evaluation of continuous changes in LA and growth (Pompelli *et al.*, 2012). Accurate LA estimation is required to understand and model ecosystem function (Antunes *et al.*, 2008). New tools and machines, such as hand scanners and laser optic apparatuses, and more recently (Adhikari *et al.*, 2020), low-cost smartphone software have been developed for LA measurements. However, some of them are too expensive and complex for basic and simple studies (Demirsoy, 2009). Allometry establishes quantitative relationships among characteristic dimensions such as LA, and volume that are not readily determined directly. Allometric models estimate the leaf area in a non-destructive way and are useful in horticultural experiments. In addition, they enable LA measurements of the same leaf throughout the growth period and may, therefore, reduce data variability (Antunes *et al.*, 2008; Pompelli *et al.*, 2012, 2019). The use of simple linear measurements to predict LA for horticultural plants eliminates the need for costly LA meters (Antunes *et al.*, 2008; Pandey, 2011; Pompelli *et al.*, 2012;Adhikari *et al.*, 2020). Modeling the linear relationships among LA and other leaf dimensions rapidly, reliably, inexpensively, accurately and nondestructively measure LA (Antunes *et al.*, 2008; Pompelli *et al.*, 2012; Adhikari *et al.*, 2020). The development of statistical regression models from linear leaf measurements to predict total and individual LA has been useful for growth and development studies (Achten *et al.*, 2010; Adhikari *et al.*, 2020). Thus, simple linear measurements, such as leaf length (L) and leaf width (W) are used in allometric equations to model the observed LA (Achten *et al.*, 2010). Nondestructive allometry for LA determination has been a subject of intensive research, particularly for plants of high economic value, such as grapevine (Teobaldelli *et al.*, 2020), purging nut (Achten *et al.*, 2010; Pompelli *et al.*, 2012), among others.

In numerous studies, however, the adequacy of the model assumptions for LA estimation has not been critically assessed. Even minor violations of the underlying assumptions could invalidate the inferences drawn from the analysis (Chatterjee & Hadi, 2006). Two prior studies described allometric models to estimate *S. rebaudiana* LA. Ramesh *et al.* (2007) used only 80-300 leaves to construct their allometric models, and Lima Filho & Malavolta (1986) applied only 70 leaves for that purpose. The major flaw in both models was the lack of regression coefficients (b_n), stability testing and, by extension, model validation. Furthermore, the studies did not conduct morphological analyses of the *S. rebaudiana* leaves. This error was serious as *S. rebaudiana* presents distinct and widely varied leaf morphotypes (Hastoy *et al.*, 2019). Here, then, we propose the following hypotheses: (*i*) the current allometric equations for estimating *S. rebaudiana* LA are unbiased; (*ii*) the linear models are reliable for estimating *S. rebaudiana* LA; (*iii*) a generalized equation for estimating *S. rebaudiana* LA is feasible; and (*iv*) leaf length is the factor that contributes the most in the variation to the LA in various genetic materials.

MATERIAL AND METHODS

Plant material and environmental conditions

Four *Stevia rebaudiana* genotypes were studied: a commercial (Morita II) and three experimental (clones 4, 16, and 18), which were obtained by natural pollination in a controlled greenhouse after selection among 25 distinct started genotypes (Aramendiz-Tatis *et al.*, 2021). The *S. rebaudiana* leaves were collected from the experimental campus of the Facultad de Ciencias Agrícolas of Universidad de Córdoba (8º47'37"N, 75º51'51"W, 15 m altitude). The region has a mean annual rainfall of 1,346 mm, a relative humidity of 84%, and a mean annual temperature of 27.4ºC. It is characterized by a tropical wet climate according to the Köppen's climate classification system.

Plant samples and processing

The *S. rebaudiana* individuals selected for leaf collection belonged to various age classes assuming to have wide genetic variability. One thousand healthy developing and mature leaves were harvested per genotype. The leaves were randomly sampled from various parts of the plants and measured to develop the best fitting model for predicting *S. rebaudiana* LA. The leaves were scanned at 1,200 dpi × 1,200 dpi resolution using the HP PSC1410 (HP Corp., Palo Alto, CA, USA), and the images were analyzed with ImagePro® Plus (v. 4.5.0.29; Media Cybernetics, Silver Spring, MD, USA) as described in detail by Pompelli *et al.* (2012). The leaf samples covered the broadest possible dimensional ranges (Table 1).

Several linear and nonlinear regression models using width (W) and/or length (L) were developed for each plant genotype. The following statistical criteria were used in model selection: (*i*) F-test, (*ii*) sample-adjusted coefficients of determination, (*iii*) stability and standard errors of estimates, and (*iv*) residual dispersion patterns. The data exploration protocol of Zuur *et al.* (2009) was used to verify statistical assumptions such as normality and independence of errors. The Durbin-Watson criteria (Durbin & Watson, 1950) were applied to select accurate, parsimonious equations. In this manner, model bias, precision, simplicity, and accuracy could be evaluated (Walther & Moore, 2005).

Model validation

Five hundred leaves of different genotypes were sampled from various *S. rebaudiana* plants grown under the same conditions as those used for model construction. Previously selected models were re-estimated using new validation samples, and the data were compared by calibrating models. The estimated leaf areas (ELA) obtained with the models were plotted against the measured leaf areas (MLA) and linear regression curves were plotted.

Model identity test

To assess whether a given model could accurately estimate LA for all genotypes with distinct leaf morphology, the following statistical null hypothesis was set up: H_0: $\beta_1 = \beta_2 = \beta_n$, where β_1, β_2, and β_n are regression coefficients. The null hypothesis postulated that there were no differences among coefficients based on the differences between the sums of squares of the complete model. Rejection of H_0 implied evidence for the acceptance of the alternative hypothesis, namely, significant differences among models for all genotypes.

Principal component analysis

The morphological parameters were used to run a principal component analysis (PCA) in the PCA function of Minitab v. 18.1.0.0 (Minitab LLC, Pennsylvania State University, University Park, PA, USA). The PCA summary function calculated the proportion of variance in each morphological parameter explained by each principal component axis. For hierarchical clustering, Pearson's correlations were used to compare similarities among genotypes via the "cor" function in R (R Core Team, Murray Hill, NJ, USA). The complete linkage method and the Euclidean distance were used for hierarchical clustering with the R index in Minitab.

Statistical data analysis

Statistical analyses were performed in Statistica v. 8.0 (StatSoft, Tulsa, OK, USA), DataFit v. 8.0.32 (Oakdale Engineering, Oakdale, PA, USA), SigmaPlot for Windows v. 11.0 (Systat Software, Inc., San Jose, CA, USA) and R v. 3.3.3 (CoreTeam, 2020). All other calculations, statistical analyses, and graph generation were performed in GerminaR (Lozano-Isla *et al.*, 2020). To identify significant differences among factors the ANOVA was performed. With a Student-Newman-Keuls test the means were compared. $P<0.001$ was considered statistically significant.

RESULTS AND DISCUSSION

The large sample size used in this study (n = 1,000) revealed a diversity of leaf size ranging from expanding to fully expanded leaves. The leaf length ranges were 1.74-5.49 cm (genotype 4), 2.56-6.40 cm (genotype 16), 2.27-6.07 cm (genotype 18), and 2.37-7.03 cm (Morita II). The leaf width ranges were 0.95-2.59 cm, 0.98-3.27 cm, 1.10-2.88 cm, and 1.01-2.81 cm for the same respective genotypes. Thus, the average LA were 3.77±1.10 cm^2, 4.41±1.24 cm^2, 5.71±1.54 cm^2, and 6.29±1.62 cm^2 for genotypes 4, 18, Morita, and 16, respectively (Table 1). Several studies (Antunes *et al.*, 2008; Pompelli *et al.*, 2012) reported that using small sample sizes to construct leaf area allometric models can generate biased equations. A few studies used only expanded leaves (Lima Filho & Malavolta, 1986; Ramesh *et al.*, 2007; Achten *et al.*, 2010). However, this practice is not recommended for the construction of allometric models because agricultural treatments must be applied to whole plants and not merely the leaves alone. Hence, certain size selection criteria must be met, and the management and interpretation of horticultural characteristics become difficult. Pompelli *et al.* (2012) compared the methods of Achten *et al.* (2010) against a standard LA measurement method and found that the former were biased and substantially underestimated LA. For this reason, we were unsuccessful at validating the allometric equations proposed by Ramesh *et al.* (2007) and Lima Filho & Malavolta (1986). Nevertheless, these equations were selected only based on their coefficients of determination (R^2) and their standard errors of the estimate. The accuracy of those equations was not tested.

At least 125 equations were generated per genotype. Of these, at least 36 exhibited satisfactory biological behavior. To reduce complexity, we

Table 1. Means ± standard deviations (SD), minimum (Min) and maximum (Max) values for the leaf length and width and leaf area of 1,000 independent *Stevia rebaudiana* leaves. Cordoba, University of Cordoba, 2020.

Genotypes	Leaf length (cm)			Leaf width (cm)			Leaf area (cm²)		
	$\bar{x}$	Min	Max	$\bar{x}$	Min	Max	$\bar{x}$	Min	Max
4	3.18 ± 0.52	1.74	5.49	1.64 ± 0.28	0.95	2.59	3.77 ± 1.10	1.34	8.06
16	3.94 ± 0.52	2.56	6.40	2.29 ± 0.36	0.98	3.27	6.29 ±1.62	1.96	12.15
18	3.72 ± 0.60	2.27	6.07	1.84 ± 0.27	1.10	2.88	4.41 ± 1.24	1.83	10.13
Morita II	4.54 ± 0.80	2.37	7.03	1.88 ± 0.28	1.01	2.81	5.71 ± 1.54	1.92	11.23

disqualified second- to sixth-order exponential, logarithmic, and polynomial equations. Therefore, we presented only nine main equations including those that were linear with a zero intercept, linear with non-zero intercept, and power models. All of them were generated from linear measurements of L, W, or both (Table 2). The coefficients of

Table 2. Statistical models, regression coefficients (β_0 and β_1), standard errors of estimates (SE), coefficients of determination adjusted for the degrees of freedom (R_a^2), mean square error (MS_{es}), calculated F (F_{calc}), P value, and equations of leaf area as a function of linear dimensions of leaves (length, L, and width, W) of *Stevia rebaudiana* leaves. Cordoba, University of Cordoba, 2020.

Equation Number	Model	Coefficients		SE	R_a^2	MS_{res}	F_{calc}	P	Estimator of LA ($\hat{Y}$)*
		β_0	β_1						
Genotype 04									
#1	$Y = \beta_1*W + \varepsilon_i$	---	2.3337	0.5652	0.9783	0.3194	47,369.91	< 0.0001	$\hat{Y} = 2.3337*(W)$
#2	$Y = \beta_0 + \beta_1*W + \varepsilon_i$	-2.2004	3.6349	0.4278	0.8498	0.1830	5,654.56	< 0.0001	$\hat{Y} = -2.2004 + 3.6349*(W)$
#3	$Y = \beta_0*W^{\beta 1} + \varepsilon_i$	1.6931	1.5861	0.4216	0.8541	0.1777	5,850.84	< 0.0001	$\hat{Y} = 1.6931*(W)^{1.5861}$
#4	$Y = \beta_1*L + \varepsilon_i$	---	1.200	0.6770	0.9694	0.4583	32,714.97	< 0.0001	$\hat{Y} = 1.5861*(L)$
#5	$Y = \beta_0 + \beta_1*L + \varepsilon_i$	-1.9275	1.7903	0.6028	0.7019	0.3633	2,353.08	< 0.0001	$\hat{Y} = -1.9275 + 1.7903*(L)$
#6	$Y = \beta_0*L^{\beta 1} + \varepsilon_i$	0.6640	1.4921	0.6024	0.7022	0.3629	2,357.02	< 0.0001	$\hat{Y} = 0.6640*(L)^{1.4921}$
#7	$Y = \beta_1*LW + \varepsilon_i$	---	0.7041	0.2742	0.9941	0.0752	204,489.09	< 0.0001	$\hat{Y} = 0.7041*(LW)$
#8	$Y = \beta_0 + \beta_1*LW + \varepsilon_i$	0.2629	0.6590	0.2634	0.9431	0.0694	16,555.51	< 0.0001	$\hat{Y} = 0.2629 + 0.6590*(LW)$
#9	$Y = \beta_0*LW^{\beta 1} + \varepsilon_i$	0.7996	0.9295	0.2619	0.9437	0.0686	16,753.86	< 0.0001	$\hat{Y} = 0.7996*(LW)^{0.9295}$
Genotype 16									
#1	$Y = \beta_1*W + \varepsilon_i$	---	2.7849	0.7880	0.9843	0.6209	67,055.07	< 0.0001	$\hat{Y} = 2.7849*(W)$
#2	$Y = \beta_0 + \beta_1*W + \varepsilon_i$	-3.2219	4.1584	0.6061	0.8606	0.3673	6,177.70	< 0.0001	$\hat{Y} = -3.2219 + 4.1584*(W)$
#3	$Y = \beta_0*W^{\beta 1} + \varepsilon_i$	1.6398	1.6086	0.5714	0.8762	0.3265	7,073.60	< 0.0001	$\hat{Y} = 1.6398*(W)^{1.6086}$
#4	$Y = \beta_1*L + \varepsilon_i$	---	1.6151	1.0396	0.9734	1.0808	38,099.16	< 0.0001	$\hat{Y} = 1.6151*(L)$
#5	$Y = \beta_0 + \beta_1*L + \varepsilon_i$	-4.0469	2.6251	0.8975	0.6947	0.8055	2,274.01	< 0.0001	$\hat{Y} = -4.0469 + 2.6251*(L)$
#6	$Y = \beta_0*L^{\beta 1} + \varepsilon_i$	0.6748	1.6228	0.9025	0.6913	0.8145	2,237.70	< 0.0001	$\hat{Y} = 0.6748*(L)^{1.6228}$
#7	$Y = \beta_1*LW + \varepsilon_i$	---	0.6879	0.3267	0.9965	0.1067	394,912.94	< 0.0001	$\hat{Y} = 0.6879*(LW)$
#8	$Y = \beta_0 + \beta_1*LW + \varepsilon_i$	0.1608	0.6715	0.3244	0.9601	0.1052	24,054.20	< 0.0001	$\hat{Y} = 0.1608 + 0.6715*(LW)$
#9	$Y = \beta_0*LW^{\beta 1} + \varepsilon_i$	0.7344	0.9717	0.3239	0.9602	0.1049	24,131.71	< 0.0001	$\hat{Y} = 0.7344*(LW)^{0.9717}$
Genotype 18									
#1	$Y = \beta_1*W + \varepsilon_i$	---	2.4423	0.6737	0.9774	0.4539	45,186.89	< 0.0001	$\hat{Y} = 2.4423*(W)$
#2	$Y = \beta_0 + \beta_1*W + \varepsilon_i$	-3.6120	4.3701	0.4306	0.8794	0.1854	7,286.99	< 0.0001	$\hat{Y} = -3.6120 + 4.3701*(W)$
#3	$Y = \beta_0*W^{\beta 1} + \varepsilon_i$	1.4446	1.8115	0.4103	0.8905	0.1683	8,127.39	< 0.0001	$\hat{Y} = 1.4446*(W)^{1.8115}$
#4	$Y = \beta_1*L + \varepsilon_i$	---	1.2014	0.6871	0.9765	0.4721	43,406.05	< 0.0001	$\hat{Y} = 1.2014*(L)$
#5	$Y = \beta_0 + \beta_1*L + \varepsilon_i$	-2.4716	1.8491	0.5648	0.7925	0.3190	3,816.62	< 0.0001	$\hat{Y} = -2.4716 + 1.8491*(L)$
#6	$Y = \beta_0*L^{\beta 1} + \varepsilon_i$	0.5789	1.5371	0.5598	0.7962	0.3134	3,903.05	< 0.0001	$\hat{Y} = 0.5789*(L)^{1.5371}$
#7	$Y = \beta_1*LW + \varepsilon_i$	---	0.6310	0.2744	0.9954	0.0753	277,489.22	< 0.0001	$\hat{Y} = 0.6310*(LW)$
#8	$Y = \beta_0 + \beta_1*LW + \varepsilon_i$	0.2957	0.5919	0.2614	0.9555	0.0683	21,474.14	< 0.0001	$\hat{Y} = 0.2957 + 0.5919*(LW)$
#9	$Y = \beta_0*LW^{\beta 1} + \varepsilon_i$	0.7177	0.9375	0.2612	0.9556	0.0682	21,512.56	< 0.0001	$\hat{Y} = 0.7344*(LW)^{0.9717}$
Genotype Morita II									
#1	$Y = \beta_1*W + \varepsilon_i$	---	3.0836	0.8454	0.9786	0.7147	47,893.19	< 0.0001	$\hat{Y} = 3.0836*(W)$
#2	$Y = \beta_0 + \beta_1*W + \varepsilon_i$	-3.5863	4.9530	0.6582	0.8171	0.4332	4,463.98	< 0.0001	$\hat{Y} = -3.5863 + 4.9530*(W)$
#3	$Y = \beta_0*W^{\beta 1} + \varepsilon_i$	1.9863	1.6575	0.6479	0.8228	0.4197	4,639.20	< 0.0001	$\hat{Y} = 1.9863*(W)^{1.6575}$
#4	$Y = \beta_1*L + \varepsilon_i$	---	1.2661	0.9249	0.9745	0.8554	39,853.34	< 0.0001	$\hat{Y} = 1.2661*(L)$
#5	$Y = \beta_0 + \beta_1*L + \varepsilon_i$	-1.3778	1.5604	0.8936	0.6628	0.7986	1,965.07	< 0.0001	$\hat{Y} = -1.3778 + 1.5604*(L)$
#6	$Y = \beta_0*L^{\beta 1} + \varepsilon_i$	0.8607	1.2472	0.8935	0.6629	0.7983	1,965.92	< 0.0001	$\hat{Y} = 0.8607*(L)^{1.2472}$
#7	$Y = \beta_1*LW + \varepsilon_i$	---	0.6538	0.4861	0.9922	0.2363	146,893.27	< 0.0001	$\hat{Y} = 0.6538*(LW)$
#8	$Y = \beta_0 + \beta_1*LW + \varepsilon_i$	0.6246	0.5872	0.4543	0.9129	0.2064	10,468.67	< 0.0001	$\hat{Y} = 0.6246 + 0.5872*(LW)$
#9	$Y = \beta_0*LW^{\beta 1} + \varepsilon_i$	0.8458	0.8866	0.4517	0.9138	0.2041	10,597.18	< 0.0001	$\hat{Y} = 0.7344*(LW)^{0.9717}$

determination adjusted for degrees of freedom (R_a^2) were in the ranges of 0.7019-0.9941 (genotype 4), 0.6913-0.9965 (genotype 16), 0.7925-0.9954 (genotype 18), and 0.6628-0.9922 (Morita II) (Table 2). Models #1 and #7 did not efficiently estimate LA for any of the tested genotypes, as they generated extremely high calculated F (F_{calc}) values (≤394,912.94) for the LA estimation of genotype 16 (equation #7). Figure 1 shows that equation #2 was ineffective at estimating any genotype here. Model #1 slightly overestimated LA, whereas model #2 strongly overestimated it. Equation #7 significantly underestimated the LA for all genotypes, except 16. Equation #4 significantly overestimated the LA for genotype 18. The first selection disqualified equations #1, #2, and #7 from validation but did not permit the exclusion of equation #4, as only the LA of genotype 18 was overestimated by ~38%.

Hence, only equations #3, #4, #5, #6, #8, and #9 remained in the validation process (Figure 2). Allometric LA estimation was perfectly suited to all genotypes, and the variation was

Table 3. Variance analysis for linear models ($Y = \beta_0 + \beta_1 X$), where X is LW product, using the set of leaves adjustment of the four distinct *Stevia rebaudiana* leaves: Clone 4, Clone 16, Clone 18, and Clone Morita II (n = 4,000). The dependent and independent variables were log-transformed for the analysis, following recommendations for statistical standardization of the data for variance reduction in accord of Zuur *et al.* (2009). Cordoba, University of Cordoba, 2020.

Variation font	Degrees of Freedon	Sum of squares	Mean square	F_{calc}
Parameters	10	2,424.622	-	
Reduction (βs)	2	1,976.070	-	
Reduction (H_0)	8	448.552	56.069	12.736*
Residual	3,989	17,560.854	4.402	
Total	3,999	19,985.477		

* $F_{0.01}$ (8; 3,939) = 2.516.

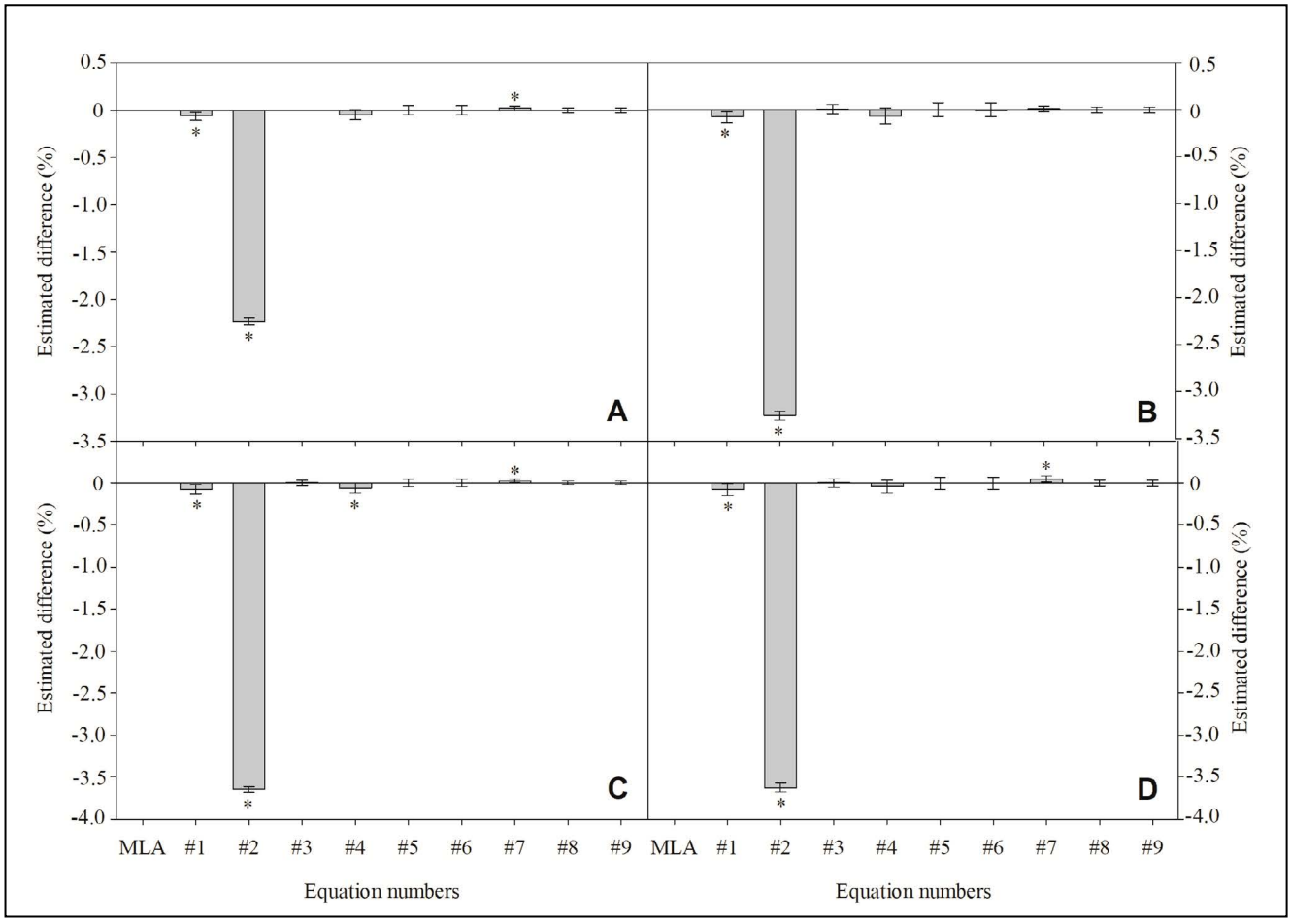


Figure 1. Deviation of the estimated area from the observed leaf area (LAo) for individual leaves. Leaf areas for the genotype 4 (A), 16 (B), 18 (C), and Morita II (D) of *Stevia rebaudiana* were estimated using several models in which β_0 and β_1 are coefficients. Vertical bars denote means and spreads denote 99% confidence intervals of the difference (distribution of t-test). For details of the equations and its equation number, see Table 2. The asterisks (*) denotes biased equations. Cordoba, University of Cordoba, 2020.

not genotype-dependent. However, analysis of the residual dispersion patterns showed that model #3 slightly overestimated the LA of the Morita II by ~1.1% (Figure 2A). Nevertheless, this error was insignificant relative to a sample size of 1,000 leaves. Therefore, equation #3 remained in the *S. rebaudiana* LA estimation analysis. In contrast, analysis of the residual dispersion pattern for equation #4 disclosed a slight overestimation of LA

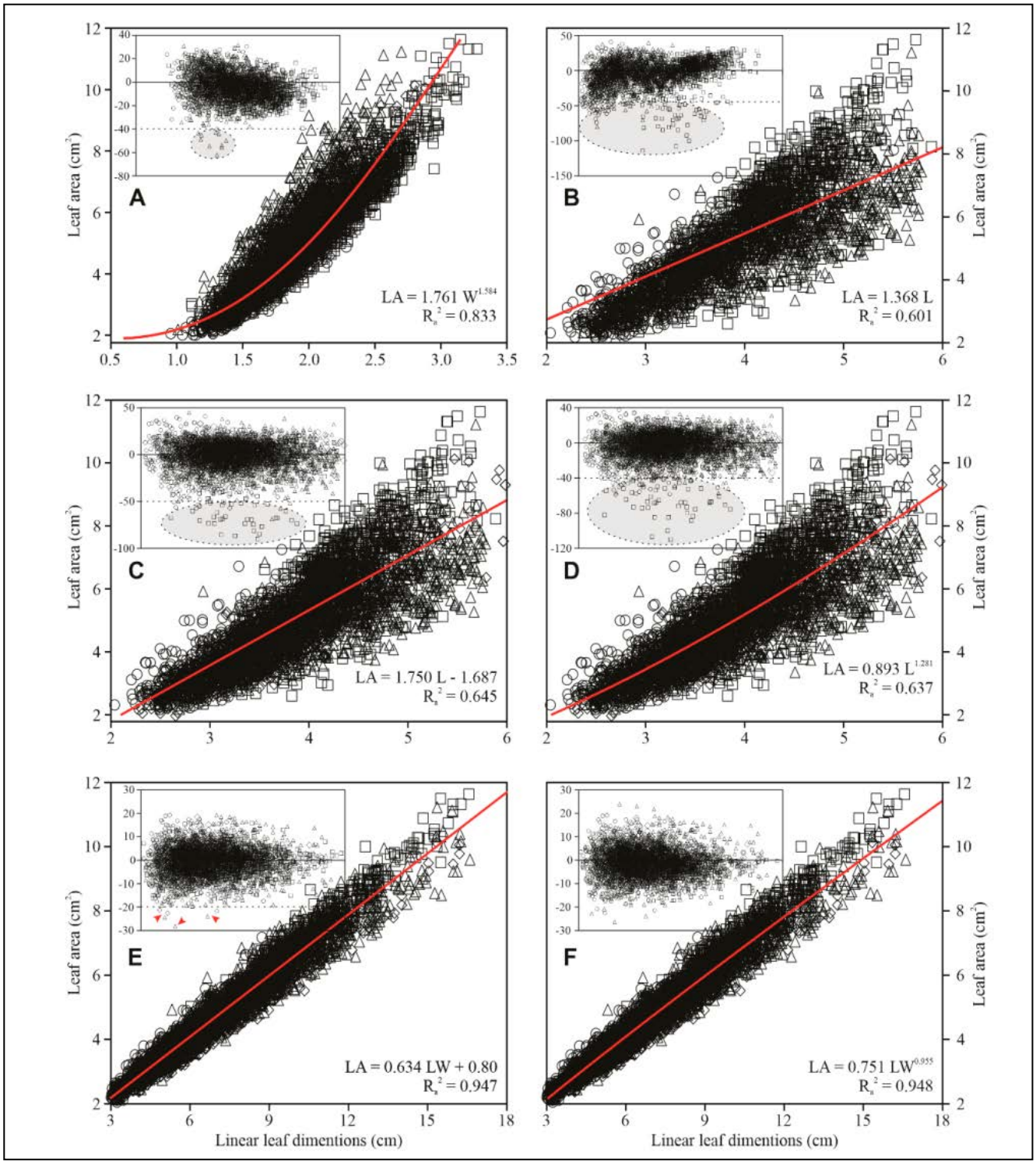


Figure 2. Regression curves for leaf area and linear leaf dimensions: width (A; equation #3), length (B-D; equation #4, #5, and #6), and product of leaf width and leaf length (E-F; equations #8, and #9) for *Stevia rebaudiana* leaves: Genotypes 4 (circles), 16 (squares), 18 (diamond), and Morita II (triangle), using linear (B, C, and E) and power (A, D, and F) models. The dispersion pattern of residuals for the respective models is shown in the insets. The shaded area denotes the region of highly biased data in the equations causing an overestimation of *Stevia rebaudiana* leaves, while data pointed with arrows denote a slight overestimation of *Stevia rebaudiana* leaves. All graphs were generated with 1,000 independent leaves of each *Stevia rebaudiana* genotypes. Cordoba, University of Cordoba, 2020.

(4.4%) for genotype 16 (Figure 2B), especially for the more expanded leaves, as well as LA overestimations of ~1.3% and ~2.1% for Morita II and genotype 4, respectively. With 75% LA significant overestimation, we considered model 4 too biased for estimation and disqualified it from the subsequent analyses. Model #5 promoted an overestimation of the LA of genotype 16 by 3.5% (Figure 2C). We reanalyzed β_0 for this model (-1.93, -4.05, -2.47, and -1.38 for genotypes 4, 16, 18, and Morita II, respectively) and concluded that the negative β_0 would negative ELA even if the leaf length was zero. This biological condition is invalid, and this model was disqualified from the subsequent analyses. Model #6 overestimated LA for genotypes 4 and 16 by 1.9% and 4.5%, respectively, and nonsignificantly overestimated LA for genotype 18 and Morita II (Figure 2D). Thus, model #6 was disqualified for the analysis of genotypes 4 and 16, but it was used to analyze genotype 18 and Morita II.

The equations that were retained for the subsequent analyses were #3, #8, and #9 for genotypes 4 and 16 (Figure 3) and #3, #6, #8, and #9 for genotypes 18 and Morita II (Figure 4). They satisfied the requirements of lower F_{calc} values, higher sample-adjusted coefficients of determination, greater stability and standard errors of the estimate, and non-biased residual dispersion

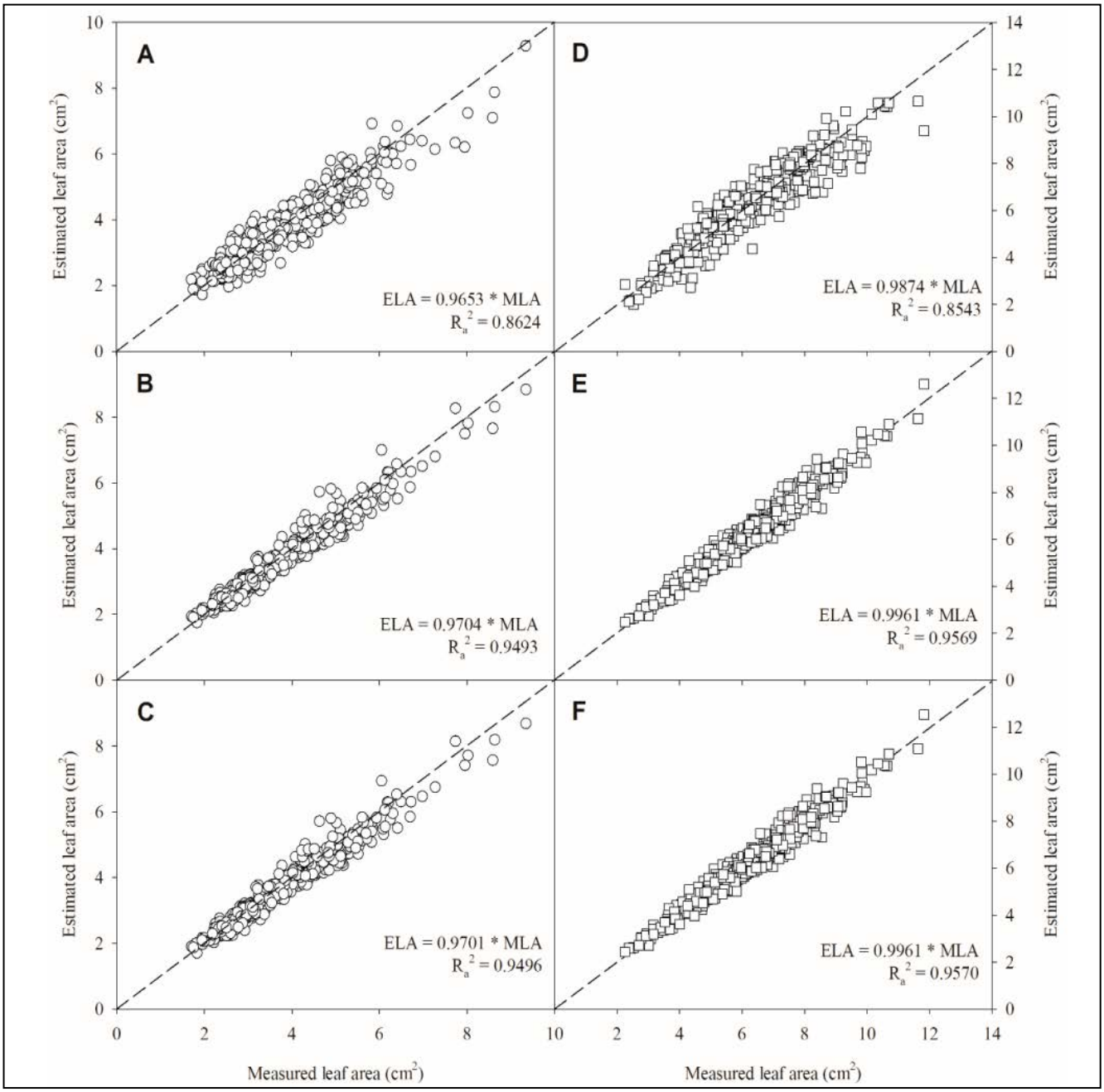


Figure 3. The relationship between estimated and measured area for the *Stevia rebaudiana* leaves. Leaf area was estimated in accord of equations #3 (A, D), #8 (B, E), and #9 (C, F) for genotype 4 (A-C), and genotype 16 (D-F) (for more details of these equations, see Table 2). Dotted line represents the 1:1 relationship. n = 500. Cordoba, University of Cordoba, 2020.

patterns. Reliability and non-biased of the allometric models proposed for LA estimation were exhibited via linear regressions between the MLA and the ELA for each of them (Figures 3 and 4). Accuracy of the β_0 and β_1 coefficients increased as R_a^2 approached unity. Hence, these equations could make unbiased estimates of LA for the second dataset. A careful analysis of Figures 3 and 4 showed that equations #3, #8, and #9 had linear regression coefficients near 1. Some were >0.9960 and R_a^2 ranged from 0.8543 (Figure 3D) to 0.9976 (Figure 4C). Thus, the equations proposed here for *S. rebaudiana* LA estimation were reliable and non-biased. In contrast, equation 6 was inappropriate for *S. rebaudiana* LA estimation, as its coefficients were unstable. While the linear regression between linear measurements and MLA proposed by equation #6 generated high linear determination coefficients (0.9904 and 0.9921 for genotypes 18 and Morita II, respectively; Table 2), the R_a^2 values of MLA *versus* ELA was only 0.6817 (Figure 4B), and 0.5123 (Figure 4F), respectively. Thus, LA estimates for a new population may not be the highest-

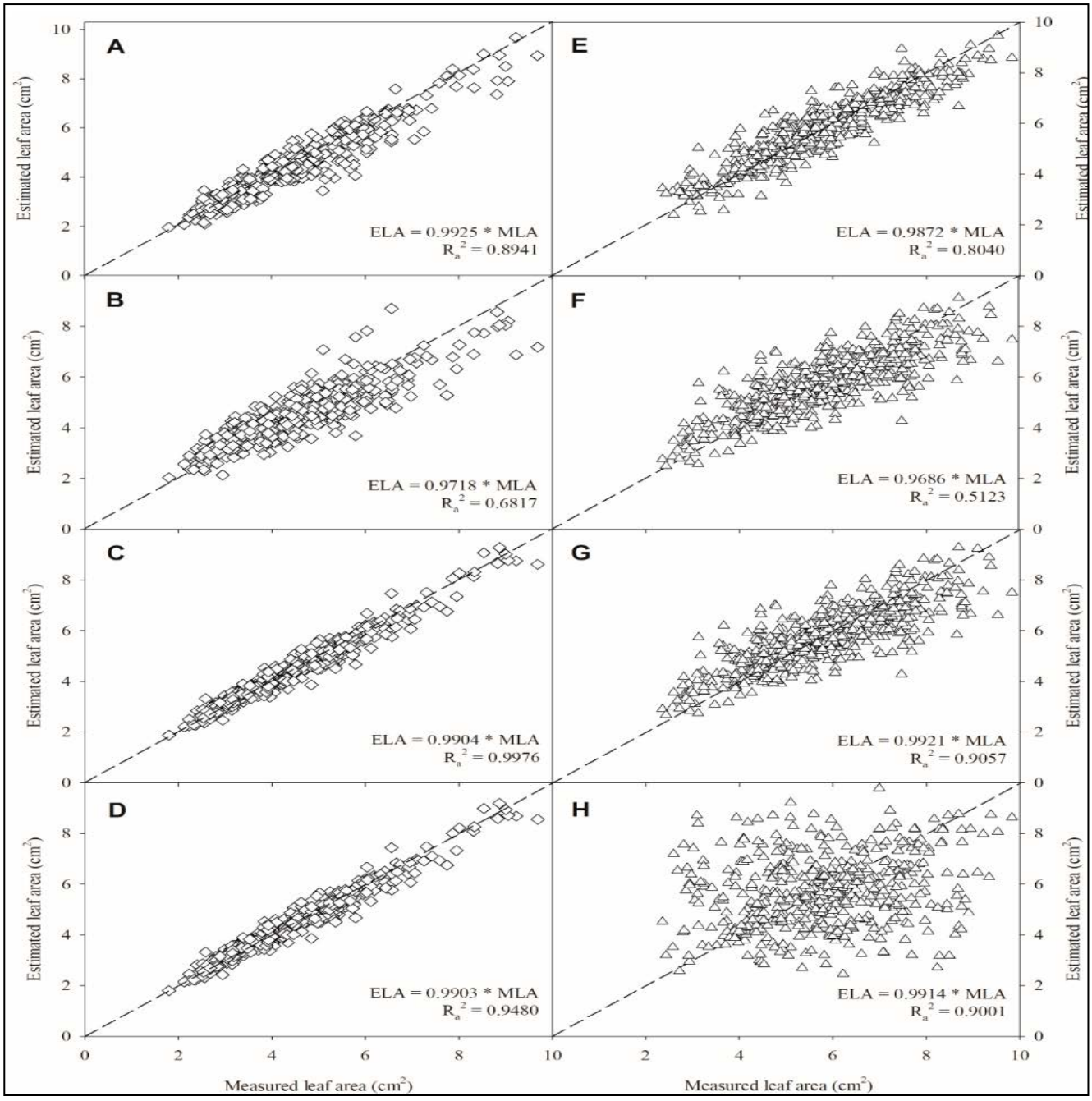
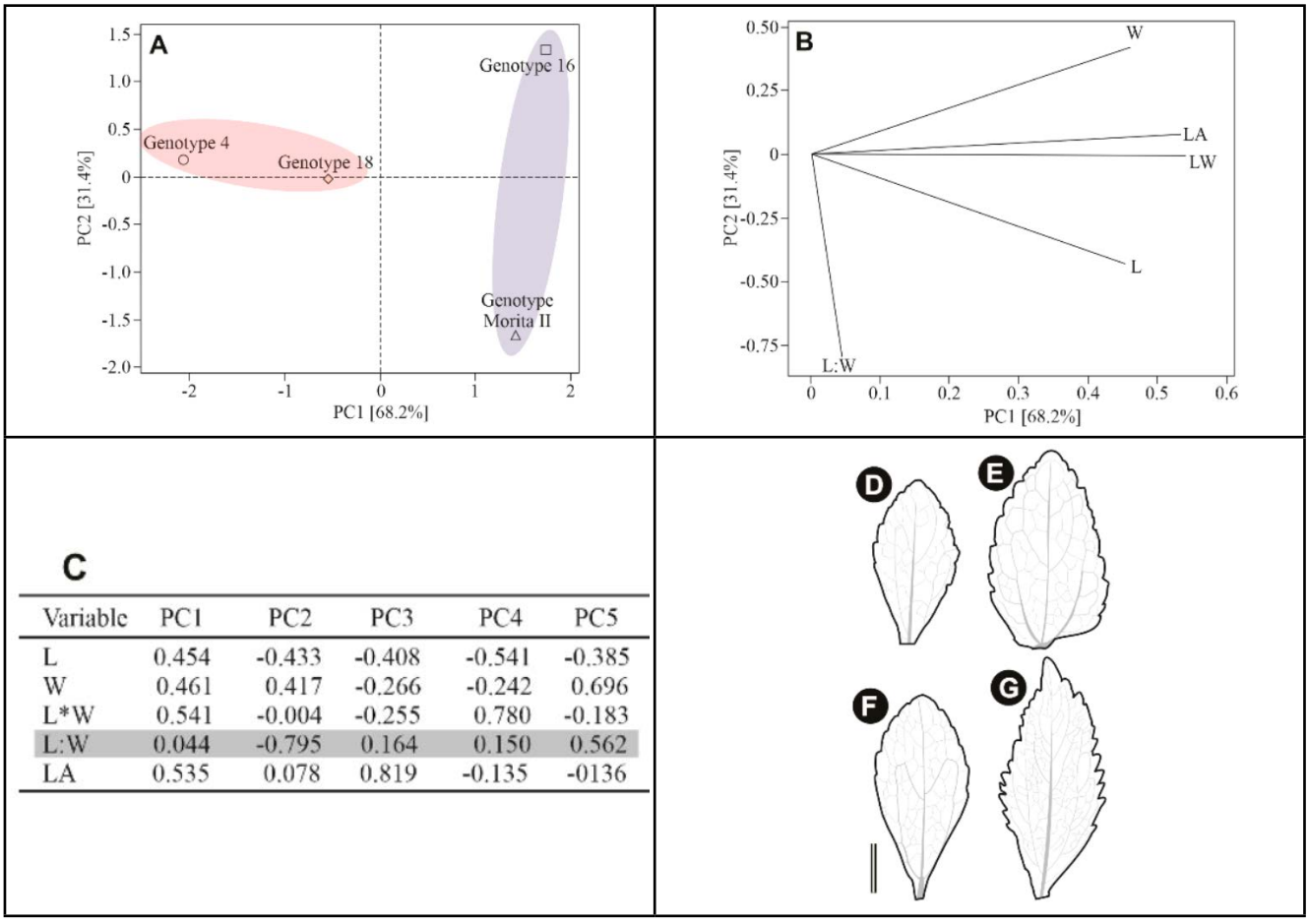


Figure 4. The relationship between estimated and measured area for the *Stevia rebaudiana* leaves. Leaf area was estimated in accord of equations #3 (A, E), #6 (B, F), #8 (C, G), and #9 (D, H) for genotype 18 (A-D), and Morita II (E-H) (for more details of these equations, see Table 2). Dotted line represents the 1:1 relationship. n = 500. Cordoba, University of Cordoba, 2020.

quality ELA in the process of building allometric models. For this reason, equation #6 was disqualified from any further *S. rebaudiana* LA estimation.

The β's coefficient stability in model validation was confirmed by analyzing the coefficients. β_0 and β_1 were comparatively more homogeneous in equations #8, and #9 but significantly more dispersed in equation #3 (data not shown). However, equation #8 presented lower confidence intervals for both β's than that of equation #9. As the R_a^2 for equations #8 and #9 were terribly similar across all genotypes, both of the estimated *S. rebaudiana* LA were with good accuracy and without bias. Nevertheless, one objective of the present study was to develop a relatively simple, non-biased equation. For this reason, we argued that the linear equation $Y = \beta_0 + \beta_1 * LW + \varepsilon_i$ type is the best fit for *S. rebaudiana* LA estimation. Then, the statistical tests and validation of the mathematical models for *S. rebaudiana* LA estimation indicate that equation #8 is the most suitable for this purpose.

In this study we described how simple nondestructive *Stevia rebaudiana* leaf measurements can accurately estimate LA. The currently used linear allometric models for estimating *S. rebaudiana* leaf area (Lima Filho & Malavolta,1986; Ramesh *et al.*, 2007) are inappropriate. In this study, we used 1,000 leaves per genotype to build the models and another 500 leaves per genotype to validate them while Ramesh *et al.* (2007) used only 80-300 leaves to construct their models. The selection based only in R^2 leads to false perception of accuracy (Chatterjee & Hadi, 2006), mainly in case of previous allometric equation (Lima Filho & Malavolta, 1986; Ramesh *et al.*, 2007) where the R^2 ranged between 0.75 to 0.83. Our high R_a^2 differed from those reported by Toebe *et al.* (2019) and Oliveira *et al.* (2019) for *Cucurbita moschata* and *Pyrus communis*, respectively. The LA estimate could be calculated accurately only for certain stages of leaf development, resulting in low R^2, possibly a small data set. In contrast, the high R^2 obtained for the estimation of the leaf area of *C. moschata* (Toebe *et al.*, 2019) was derived from a very complex linear measurement of three sections per leaf. A similar analysis was



Variable	PC1	PC2	PC3	PC4	PC5
L	0.454	-0.433	-0.408	-0.541	-0.385
W	0.461	0.417	-0.266	-0.242	0.696
L*W	0.541	-0.004	-0.255	0.780	-0.183
L:W	0.044	-0.795	0.164	0.150	0.562
LA	0.535	0.078	0.819	-0.135	-0136

Figure 5. A. Principal component analysis (PCA) of the morphological parameters in four *Stevia rebaudiana* genotypes: 4, 16, 18, and Morita II. The large circles represent the two clusters formed by the Euclidean distance method considering ~65% of similarity. B. Loading plot graph. In Loading plot, the direction and length of the lines are directly proportional to variables importance in separating groups. Drawings of the genotype 4 (D), 16 (E), 18 (F), and Morita II (G) of *Stevia rebaudiana* showing their morphotypes and leaf venation patterns. Scale for figure D-G are 1 cm. PC1, principal component 1; PC2, principal component 2. C. The morphological parameters contributions in each principal components. L: leaf length. W: leaf width. L*W: product by leaf length and leaf width. L:W: leaf length : leaf width ratio. LA: leaf area. Cordoba, University of Cordoba, 2020.

conducted on *P. communis* (Oliveira *et al.*, 2019). However, none of the aforementioned studies verified residual distribution patterns or assessed β's stability.

In another way, an allometric equation was developed for each genotype, but this approach is neither scientifically nor agronomically practical. Thus, it was essential to generate a generalized model that could reliably estimate LA for all genotypes. An identity test and the estimated β_0 and β_1 for the equations generated for each genotype to create a generalized model encompassing all genotypes in a single sample consisted of 4,000 leaves. Only Morita II made biased estimates of a greater positive amplitude of β_0 and greater amplitude of β_1 as compared to the β's values produced by the generalized equation or those returned by the equations for each genotype. Thus, they could be estimated by a single equation, namely, $LA = 0.2798 + 0.6341LW + \varepsilon_i$ [$R_a^2 = 0.9471$; RMSE (root mean square error) = 0.3947]. However, the ANOVA (Table 3) showed that a F_{calc} was 5.8-fold higher than $F_{standard\ (8;\ 3,939)}$. For this reason, H0 was rejected, implying that the genotypes presented with distinct leaf morphologies could not be grouped into a single allometric model for *S. rebaudiana* LA estimation. The present study confirmed that the average leaf morphologies were distinct for the genotypes because L:W for the broader leaves of genotype 16 was smaller (1.75±0.27) than that for the narrower leaves of Morita II (2.44±0.37). Genotypes 4 and 18 shared similar morphological characteristics and had L:W of 1.96±0.27 and 2.03±0.22, respectively (Figure 5A).

Power models based on leaf L or W were the most suitable for estimating LA in perennials such as coffee (Antunes *et al.*, 2008), purging nut (Pompelli *et al.*, 2012) and Suriname cherry (Pompelli *et al.*, 2018). For pear (Oliveira *et al.*, 2019) and squash (Toebe *et al.*, 2019), however, linear models were best suited for estimating LA. We demonstrated that separate L and W each had a relatively high R_a^2 value and a narrow residual dispersion pattern. The models did not estimate *S. rebaudiana* leaf area with high precision and without bias. These limitations may invalidate these models (Chatterjee & Hadi, 2006). Recently, certain scholars (Oliveira *et al.*, 2019; Toebe *et al.*, 2019) described a best-fit equation to estimate LA for various horticultural plant species when both L and W were factored into the model. Absence of bias, homoscedastic residual scatter, and high stability of the estimated coefficients and the lack of other deficiencies were realized for the best-fit LA estimation equations applicable to any horticultural plant species whose leaves are the main organ of economic interest. Then, we proposed a model wherein LA= 0.2629 + 0.6590*LW; LA= 0.1608 + 0.6715*LW; LA= 0.2957 + 0.5919*LW; and LA= 0.6246 + 0.5872*LW for the estimation of the LA of 4, 16, 18, and Morita II *S. rebaudiana* genotype, respectively (R_a^2= 0.9431, 0.9601, 0.9555, and 0.9129). The generalized model encompassing all genotypes was LA= 0.2798 + 0.6341*LW (R_a^2= 0.9471; RMSE= 0.3947), which accurately estimated ~93% of all *S. rebaudiana* LA without bias and irrespective of genotype.

PCA attempted to cluster the genotypes and demonstrated that they could not be grouped. Figure 5A shows that Morita II and genotype 16 shared similar phylogenetic characteristics as did genotypes 4 and 18. PC1 and PC2 accounted for 99.6% and revealed that the evaluated characteristics were widely distributed. A Euclidean distance similarity of ≥66% indicated that genotypes 4 and 18 could not be grouped with genotypes 16 and Morita II. In the quest for the parameters rendering genotype grouping infeasible, we verified that L:W describing leaf stretching in Morita II strongly influenced (-0.795) the PC2 axes as compared to the other genotypes (Figure 5C). This phenomenon was confirmed by drawing of the leaf morphotypes (Figures 5D-G), that confirms that L:W stretching made it impossible to group all genotypes or construct a generalized model for *S. rebaudiana* LA estimation. The PCA indicated wide morphological variation among *S. rebaudiana* leaves (Figure 5A). This phenomenon was previously addressed (Hastoy *et al.*, 2019). Genotypes 16 and Morita II shared similar phylogenetic features as did genotypes 4 and 18. Most of the stevia species are short day plants, meaning that after a number of long nights, the plants will flower and reduce sweetener synthesis. It has been shown that growth and sweetener production continue under long days, obtained by interruption of the long nights by RED light (Ceunen & Geuns, 2013a). Based on this, Aramendiz-Tatis *et al.* (2021) describe that new *S. rebaudiana* genotypes are required to delay flowering, increase annual harvest, and augment stevioside production (Ashok *et al.*, 2019). According to these authors, Morita II has a short vegetative phase and its first flower buds appear within a few days. In contrast, genotypes 4, 16, and 18 were selected as they have late-flowering phases and, therefore, relatively larger annual harvests. Ceunen & Geuns (2013b) and Ceunen & Geuns (2013a) reported that preventing flowering by interruption of the long nights by red light (e.g., by drones) is the easiest way.

Here, we developed simple predictive models to estimate the leaf areas of several *S. rebaudiana* genotypes including the globally distributed Morita II. Linear models fit *S. rebaudiana* LA better than power models. Therefore, the previously proposed LA estimation models should be avoided. Leaf morphology is similar among *S. rebaudiana* congenerics. Nevertheless, the published equations could not accurately estimate leaf area. The equations were validated by various statistical criteria and showed high coefficients of determination, coefficient stability, low sums of squares, and high simplicity. The allometric equations described herein can achieve cost-effective *S. rebaudiana* LA measurements and facilitate future research into the physiology and ecology of this agroeconomically important plant. In the present study, there were close relationships between the actual and model-predicted leaf area. A rapid and simple generalized equation was developed to predict *S. rebaudiana* leaf area, namely, LA=

0.2798 + 0.6341* LW (R_a^2= 0.9471; RMSE= 0.3947). This simple model can generate results as accurate as those produced by costly apparatus and more complex estimation models. A principal component analysis disclosed that certain genotypes had features in common, whereas others did not. The methodology formulated here and the data generated may help advance *S. rebaudiana* breeding with the objectives of late flowering and increased numbers of leaves per square meter. In this way, annual yield and stevioside production may be augmented.

ACKNOWLEDGEMENTS

The authors thank the Universidad de Córdoba, Colombia for funding this research. This study was financed by the Regional Group of Participatory Research of Small Farmers of the Colombian Atlantic Coast INVEPAR, Project Manejo Agronómico de Clones de *Stevia rebaudiana* en el Caribe Colombiano [Grants no. FCA-02-17]. Special thanks to Drs. Luis Rodríguez-Paéz and Carlos Espitia-Romero for their assistance and logistical support.

REFERENCES

ACHTEN, WMJ; MAES, WH; REUBENS, B; MATHIJS, E; SINGH, VP; VERCHOT, L; MUYS, B. 2010. Biomass production and allocation in *Jatropha curcas* L. seedlings under different levels of drought stress. *Biomass Bioenergy* 34: 667-676.

ADHIKARI, R; LI, C; KALBAUGH, K; NEMALI, K. 2020. A low-cost smartphone controlled sensor based on image analysis for estimating whole-plant tissue nitrogen (N) content in floriculture crops. *Computadores e Eletronicos na Agricultura* 169: 205173.

ANTUNES, WC; POMPELLI, MF; CARRETERO, DM; DAMATTA, FM. 2008. Allometric models for non-destructive leaf area estimation in coffee (*Coffea arabica* and *Coffea canephora*). *Annals of Applied Biology* 153: 33-40.

ARAMENDIZ-TATIS, HA; POLO, DJP; GONZÁLEZ, AIP. 2021. Caracterización morfoagronómica de 25 clones de estevia (*Stevia rebaudiana* Bertoni.), en condiciones del valle del Sinú medio. *Revista de Investigação Agricola e Ambiental* in press.

ASHOK, K; SANATSUJAT, S; BIKRAM, S; SANJAY, K. 2019. CSIR-IHBT-ST-02 (IC0624506: INGR17053), a Stevia (*Stevia rebaudiana*) germplasm with delayed flowering by 120 days. Prolonged vegetate phase with more numbers of harvest per year. *Indian Journal of Plant Genetic Resources* 32: 260-261.

CEUNEN, S; GEUNS, JMC. 2013a. Influence of photoperiodism on the spatio-temporal accumulation of steviol glycosides in *Stevia rebaudiana* (Bertoni). *Plant Science* 198: 72-82.

CEUNEN, S; GEUNS, JMC. 2013b. Spatio-temporal variation of the diterpene steviol in *Stevia rebaudiana* grown under different photoperiods. *Phytochemistry* 89: 32-38.

CEUNEN, S; GEUNS, JMC. 2013c. Steviol glycosides: chemical diversity, metabolism, and function. *Journal of Natural Product* 76: 1201-1228.

CHATTERJEE, S; HADI, AS. 2006. *Regression Analysis by Example*. Hoboken: John Wiley & Sons.

CORETEAM. 2020. The R Project for Statistical Computing. Bell Laboratories. https://www.r-project.org/about.html. Accessed on April 15, 2020.

DEMIRSOY, H. 2009. Leaf area estimation in some species of fruit tree by using models as a non-destructive method. *Fruits* 64: 45-51.

DURBIN, J; WATSON, GS. 1950. Testing for serial correlation in least squares regression I. *Biometrika* 37: 409-428.

GUERRERO, AB; EMETERIO, LS; DOMEÑO, I; IRIGOYEN, I; MURO, J. 2018. Steviol glycoside content dynamics during the growth cycle of *Stevia rebaudiana* Bert. *American Journl of Plant Science* 9: 892-901.

HASTOY, C; COSSON, P; CAVAIGNAC, S; BOUTIÉ, P; WAFFO-TEGUO, P; ROLIN, D; SCHURDI-LEVRAUD, V. 2019. Deciphering performances of fifteen genotypes of *Stevia rebaudiana* in southwestern France through dry biomass and steviol glycoside evaluation. *Indian Crop Production* 128: 607-619.

KINGHORN, AD; SOEJARTO, DD. 1991. Stevioside. In: NABORS, LO; GELARDI, RC (eds). *Alternative Sweeteners, 2nd editions, revised and expanded*. New York: Academic Press, p.157-171.

LIMA FILHO, OP; MALAVOLTA, E. 1986. Estimativa de área foliar em estevia (*Stevia rebaudiana* (Bert.) Bertoni). Unipi. http://www.lni.unipi.it/stevia/stevia/estimativadearea.htm. Accessed May 17, 2020.

LOZANO-ISLA, F; ALFARO, OB; SANTANA, DG; RANAL, MA; POMPELLI, MF. 2020. Germination indexes for seed germination variables for ecophysiological studies [R package GerminaR version 1.4.2]. R Development Core Team 2020, Federal University of Pernambuco.

OLIVEIRA, VS; SANTOS, KTH; MORAIS, AL; SANTOS, GP; SANTOS, JSH; SCHMILDT, O; CZEPAK, MP; GONTIJO, I; ALEXANDRE, RS; SCHMILDT, ER. 2019. Non-destructive method for estimating the leaf area of pear cv. 'Triunfo'. *Journal of Agricultural Science* 11: 14-21.

PANDEY, SK. 2011. A simple, cost-effective method for leaf area estimation. *Journal of Botany* 2011: 658240.

POMPELLI, MF; ANTUNES, WC; FERREIRA, DTRG; CAVALCANTE, PPGS; WANDERLEY-FILHO, HCL; ENDRES, L. 2012. Allometric models for non-destructive leaf area estimation of the *Jatropha curcas*. *Biomass Bioenergy* 36: 77-85.

POMPELLI, MF; MENDES, KR; RAMOS, MV; SANTOS, JNB; YOUSSEF, DTA; PEREIRA, JD; ENDRES, L; JARMA-OROZCO, A; GOMES, RS; JARMA-ARROYO, B; SILVA, ALJ; SANTOS, MA; ANTUNES, WC. 2019. Mesophyll thickness and sclerophylly among *Calotropis procera* morphotypes reveal water-saved adaptation to environments. *Journal of Arid Land* 11: 795-810.

PUTNIK, P; BEZUK, I; BARBA, FJ; LORENZO, JM; POLUNIĆ, I; BURSAĆ, DK. 2020. Sugar reduction: *Stevia rebaudiana* Bertoni as a natural sweetener. In: BARBA, FJ; PUTNIK, P; KOVAČEVIĆ, DB (eds). *Agri-Food Industry Strategies for Healthy Diets and Sustainability*. San Diego: Academic Press, p.123-152.

RAMESH, K; RAMAWAT, N; SINGH, V. 2007. Leaf area distribution pattern and non-destructive estimation methods of leaf area for *Stevia rebaudiana* (Bert.) Bertoni. *Asian Journal of Plant Science* 6: 1037-1043.

TEOBALDELLI, M; ROUPHAEL, Y; GONNELLA, M; BUTTARO, D; RIVERA, CM; MUGANU, M; COLLA, G; BASILE, B. 2020. Developing a fast and accurate model to estimate allometrically the total shoot leaf area in grapevines. *Scientia Horticulturae* 259: 108794.

TOEBE, M; SOUZA, RR; MELLO, AC; MELO, PJ; SEGATTO, A; CASTANHA, AC. 2019. Leaf area estimation of squash 'Brasileirinha' by leaf dimensions. *Ciência Rural* 49: e20180932.

TOTTÉ, N; CHARON, L; ROHMER, M; COMPERNOLLE, F; BABOEUF, I; GEUNS, JMC. 2000. Biosynthesis of the diterpenoid steviol, an ent-kaurene derivative from *Stevia rebaudiana* Bertoni, via the methylerythritol phosphate pathway. *Tetrahedron Letters* 41: 6407-6410.

WALTHER, BA; MOORE, JL. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* 28: 815-829.

ZUUR, AF; ELENA, NI; ELPHICK, CS. 2009. A protocol for data exploration to avoid common statistical problems. *Method Ecol Evolution*1: 3-14.