Scientia Horticulturae 229 (2018) 19–24

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

# Non-destructive estimation of the leaf weight and leaf area in cacao (*Theobroma cacao* L.)

Juan Carlos Suárez Salazar[a,b,*], Luz Marina Melgarejo[b], Ervin Humprey Durán Bautista[a], Julio A. Di Rienzo[c], Fernando Casanoves[d]

[a] Universidad de la Amazonia, Facultad de Ingeniería, 180001 Florencia, Caquetá, Colombia
[b] Universidad Nacional de Colombia, Facultad de Ciencias, Bogotá D.E., Colombia
[c] Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba, 5016, Córdoba, Argentina
[d] Unidad de Biometría, Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), 30501, Turrialba, Costa Rica

ARTICLE INFO

*Keywords:*
Multiple regression analysis
Leaf morphology
Model validation

ABSTRACT

Regression models to predict leaf area and leaf weight in cacao (*Theobrama cacao* L.) were fitted using leaves from cultivated plants under nursery conditions and from plants of commercial plantations, both located in the Amazon Investigations Center CIMAZ at Macagual, Caquetá, Colombia. A total of 2895 leaves were collected in such a way to cover a wide range of leaf sizes. Width, length, weight, and leaf area were measured for each leaf. The total number of leaves was randomly divided into training and validation sets. The training set was used for model fitting and selection, the other for measuring model prediction ability. Leaf area and leaf weight were modeled using different linear regression models based on length and width of leaf. Polynomial regressions involving both length and width of leaves provided very good models to estimate the expected area ($R^2$ = 0.98) and weight ($R^2$ = 0.91) of leaves.

## 1. Introduction

Monitoring crop growth uses mathematical models to describe the relationship between the growth of a plant, the production of dry matter, and the expansion of total leaf area. The study of the time course of these variables can be followed either directly or through functions of them, such as Relative Growth Rate (RGR), Crop Growth Rate (CGR), Net Assimilation Rate (NAR), Leaf Area Duration (LAD), Leaf Area Relationship (LAR), and Leaf Area Index (LAI). The study thought time of these variables allows for the evaluation of the effects of different environmental conditions and management on crop growth.

Most studies that determine growth under controlled conditions require destructive samples, but this is impossible in experiments that aim to monitor the growth of the same leaf over time. By using regression models, leaf area and leaf weight can be predicted using variables that do not require destructive sampling techniques. Leaf area and leaf weight have been demonstrated to be good proxies of plant growth from which growth parameters can be estimated.

Different methods of measurement have been developed to calculate leaf area; however, these procedures require the extraction of leaves, preventing these from healing over time. Non-destructive sampling allows for the repeated measurement of leaves over time while avoiding the biological alteration characteristic of destructive methods (De Swart et al., 2004). Rouphael et al. (2007) working on sunflowers and Tsialtas et al. (2008) working on the cabernet sauvignon grapevine, modeled leaf area and weight, applying linear regression models used as predictors of leaf length and/or leaf width. These methods are more economical, quick, and reliable, as well as non-destructive. Through leaf area and weight estimations, the agronomic and physiological behavior of plants with respect to the availability of radiation and water, as well as sowing schemes, can be explained (Blanco and Folegatti, 2005).

Regression models have been widely used to estimate the area and weight of leaves in a variety of crops such as pepper (Rojas-Lara et al., 2008), avocado (Calderón et al., 2009), coffee (Antunes et al., 2008), peach (Espinoza-Espinoza et al., 1998), maize (Birch et al., 1998), orange (Avanza et al., 2004; Hernández-López et al., 2004), papaya (Cardona et al., 2009), rose (Fascella et al., 2013), tobacco (Bozhinova, 2006), cassava (Burgos et al., 2010), and mango (Ghoreishi et al., 2012, Calderón et al., 2009). The aim of this study was to evaluate different models to provide precise estimates of leaf area and leaf weight in *Theobroma cacao* L. using non-destructive techniques.

* Corresponding author at: Universidad de la Amazonia, Facultad de Ingeniería, 180001 Florencia, Caquetá, Colombia.
*E-mail address:* juansuarez1@gmail.com (J.C. Suárez Salazar).

http://dx.doi.org/10.1016/j.scienta.2017.10.034
Received 1 August 2017; Received in revised form 16 October 2017; Accepted 17 October 2017
0304-4238/ 

## 2. Materials and methods

Information concerning the morphological variables of leaves was collected from plants of the *Theobroma cacao* L. from a mixture of four clones (CCN51, EET8, IMC67 and TCH565), cultivated in nursery as well as in commercial plantations, both located at the *Centro de Investigación Amazónica* CIMAZ at Macagual (1° 37′ N, 75° 36′ W), Caquetá, Colombia. A total of 2890 leaves without apparent damage were collected from 440 plants randomly selected from the plantations (plants aging from 0.5 to 4 years old), and from 360 plants randomly selected at the nursery (plants aged six months old). We systematically chose two leaves per plant in nursery (bottom and top of the plant) and five leaves per plant in plantation covering five height strata of the plant.

The fresh weight of each leaf was taken using an Ohaus Scout electronic balance (100 ± 0.001 g). Each leaf was also scanned using the HP ScanJet Pro 2500 scanner. Program ImageJ was used to calculate morphometric measurements such as length, width, perimeter, and leaf area from its scanned image (Ferreira and Rasband, 2012). Summary statistics are provided for each of the morphometric variables evaluated. Pairwise Pearson's and Spearman's correlation coefficients were also calculated among the set of morphometric variables.

Exploratory data analysis showed that relationship among dependent (leaf weight and leaf area) and predictor (leaf length and leaf width) variables always described curvilinear trends. Therefore, multiple linear regression models, including second order polynomials on length, width or both variables, were used to estimate the expected value of the weight and area of leaves. The fitted models are listed below (Table 1). Variable Y represents leaf area or leaf weight whenever applicable. In all cases, the heteroscedastic version of the model performed better (lower AIC) than the homoscedastic one. The heteroscedasticity was modeled with a power function of the predicted value. Models were fitted using function gls of library nlme (Pinheiro et al., 2016) of R language version 3.4.0 (R Core Team, 2017) using the interface of InfoStat (Di Rienzo et al., 2017).

From the total of the 2890 leaves recorded, 1465 were randomly chosen to comprise a training set which in turn was used to estimate the models. The remaining 1425 leaves were used as a validation set to measure the predictive ability of a fitted model calculating, in this set, the $R^2$ and prediction mean square error (PMSE).

Prediction mean square error (PMSE) and $R^2$ are summary statistics that allow for the comparison of models and suggest how good the fitting is. However, they do not tell anything about trends in departure of observed to expected values, across the regions of prediction space. A simple and effective way to visualize any trouble regarding this issue is to draw a scatter plot of observed vs. expected values, and overlap a reference line (y = x) and the regression line of observed vs. expected values. A departure of the regression line of observed vs. expected from the reference line is evidence of bias.

Sometimes, reference and regression lines are too close to visually recognize differences among a set of competing models. We calculated the area between those lines (ABL) in the range of observed dependent variables as a measurement of bias (an R script implementation of this function is provided in the Appendix A).

**Table 2**
Summary statistics for morphometric variables of leaves of *Theobroma cacao* L. The sample was taken to cover a wide range of leave sizes.

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Weight (g) | 0.69 | 0.38 | 0.04 | 1.99 |
| Area (cm$^2$) | 60.34 | 30.47 | 3.59 | 165.47 |
| Length (cm) | 15.47 | 4.50 | 2.34 | 29.08 |
| Width (cm) | 5.59 | 1.49 | 1.11 | 10.47 |

## 3. Results

Summary statistic are provided for each of the morphometric variables evaluated (Table 2). Pairwise Pearson's and Spearman's correlation coefficients were also calculated among the set of morphometric variables (Fig. 1). Pearson's and Spearman's correlation coefficients were very similar (concordance between these correlations coefficients was 0.95).

Regression coefficients for the six models described above were estimated for both dependent variables − weight and leaf area (Tables 3 and 4). In both cases, Model 2, which is a second order polynomial on leaf length and leaf width, without their cross product have a PMSE as low as for Model 1, but require one less parameter, not diminishing $R^2$. Exclusion of leaf length or leaf width increased PMSE, and reduced $R^2$.

A departure of the regression line of observed vs. expected from the reference line is evidence of bias (Figs. 2 and 3 for weight and area respectively). In this case reference and regression line are too close to visually recognize differences among a set of competing models. When modeling weight, and considering the PMSE (smaller is better) and $R^2$, three models (Model 1, Model 2 and Model 5) have similar performance. However, Model 5 has a smaller number of parameters and the regression line of observed vs. predicted is closer to the reference line (lesser ABL, Fig. 2). Thus, we propose Model 5 as the best for weight estimation.

When modeling area, and considering the PMSE (smaller is better) and $R^2$, two models (Model 1, Model 2) have the lower PMSE and higher $R^2$. However, Model 2 has a smaller number of parameters and the regression line of observed vs. predicted is closer to the reference line (lesser ABL, Fig. 3). Thus, we propose Model 2 as the best for area estimation.

## 4. Discussion

Calculating leaf area and weight via regression methods is an inexpensive and useful tool for the investigation of the physiology and agronomic behavior of crops under different management conditions, including fertilization, water availability, and instances of contrasting conditions such as crop rehabilitation via agroforestry. For this reason, the morphological parameters of the leaf, such as length and width, have been frequently used when developing regression estimators of leaf variables that are more difficult to measure (Keramatloua et al.,

**Table 1**
Fitted models to estimate expected values of weight and area of leaves of cacao (Theobroma cacao L.) as a function of the length and width of the leaves.

| | |
|---|---|
| (MODEL 1) | $Y_i = \beta_0 + \beta_1 length_i + \beta_2 width_i + \beta_3 length_i^2 + \beta_4 width_i^2 + \beta_5 length_i \times width_i + e_i$ |
| (MODEL 2) | $Y_i = \beta_0 + \beta_1 length_i + \beta_2 width_i + \beta_3 length_i^2 + \beta_4 width_i^2 + e_i$ |
| (MODEL 3) | $Y_i = \beta_0 + \beta_1 length_i + \beta_2 length_i^2 + e_i$ |
| (MODEL 4) | $Y_i = \beta_0 + \beta_1 width_i + \beta_2 width_i^2 + e_i$ |
| (MODEL 5) | $Y_i = \beta_0 + \beta_1 length_i + \beta_2 width_i + \beta_3 width_i^2 + e_i$ |
| (MODEL 6) | $Y_i = \beta_0 + \beta_1 length_i + \beta_2 length_i^2 + \beta_3 width_i + e_i$ |

$e_i \sim N(0, \sigma_i^2 = (\hat{Y}_i)^{\rho});\quad cor(e_i, e_{i'}) = 0 \ \ \forall\ (i \neq i')$