Annals of Applied Biology ISSN 0003-4746

RESEARCH ARTICLE

Allometric models for non-destructive leaf area estimation in coffee (*Coffea arabica* and *Coffea canephora*)

W.C. Antunes, M.F. Pompelli, D.M. Carretero & F.M. DaMatta

Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

Keywords
Coffea arabica; *Coffea canephora*; estimating model.

Correspondence
F.M. DaMatta, Departamento de Biologia Vegetal, Universidade Federal de Viçosa, 36570-000 Viçosa, Minas Gerais, Brazil.
Email: fdamatta@ufv.br

Received: 15 October 2007; revised version accepted: 28 January 2008.

doi:10.1111/j.1744-7348.2008.00235.x

Abstract

We aimed to evaluate the currently used allometric models, as well as to propose a reliable and accurate model using non-destructive measurements of leaf length (L) and/or width (W), for estimating the area of leaves of eight field-grown coffee cultivars. For model construction, a total of 1563 leaves were randomly selected from different levels of the tree canopies and encompassed the full spectrum of measurable leaf sizes (0.3–263 cm^2) for each genotype. Power models better fit coffee leaf area (LA) than linear models. To validate the model, an independent data set of 388 leaves was used. We demonstrated that the currently used allometric models are biased, underestimating the area of a coffee leaf. We developed a single power model $\left(Y = \beta_0 x^{\beta_1}\right)$ based on two leaf dimensions [LA = 0.6626 $(LW)^{1.0116}$; standard errors: $\beta_0 = 0.0064$, $\beta_1 = 0.0019$; $R^2 = 0.996$] with high precision and accuracy, random dispersion pattern of residuals and also unbiased, irrespective of cultivar and leaf size and shape. Even when the L (but not width) alone was used as the single leaf dimension, the power model developed still predicted with good accuracy the LA but at the expense of some loss of precision, as particularly found for 8% of the leaves sampled with length-to-width ratios below 2.0 or above 3.0.

Introduction

Among 103 species of the *Coffea* genus (Davies *et al.*, 2006), *Coffea arabica* L. (arabica coffee) and *Coffea canephora* Pierre ex Froehner (robusta coffee) dominate the world coffee trade, being responsible for about 65% and 35%, respectively, of world coffee production. Coffee is one of the most important commodities in international agricultural trade generating over US$90 billion each year and involving about 500 million people in its management, from cultivation to final product for consumption. It is currently grown in some 80 countries in four continents. Brazil is the world's largest coffee producer followed alternately by Colombia and Vietnam. Many African countries including Uganda, Burundi, Rwanda and Ethiopia have coffee as their main source of foreign exchange (DaMatta & Ramalho, 2006).

Leaf area (LA) is a key variable for most agronomic and physiological studies involving plant growth, light interception, photosynthetic efficiency, evapotranspiration and responses to fertilisers and irrigation (Blanco & Folegatti, 2005). Therefore, LA strongly influences crop growth and productivity, and estimation of LA is a fundamental component of crop growth models (Lizaso *et al.*, 2003). However, the measurement of the surface area of a large number of leaves is often costly, time consuming and destructive. A modelling approach involving linear relationships between LA and one or more dimensions of the leaf is an inexpensive, rapid, reliable and a non-destructive alternative for accurately measuring LA (Williams & Martinson, 2003; Lu *et al.*, 2004). However, in many studies, the adequacy of the model assumptions for estimating LA has not been carefully examined. In this regard, small or minor violations of the underlying assumptions can invalidate the inferences drawn from the analysis in a major way. A simple and effective method for detecting model deficiencies in regression analysis is the examination of residual plots. Residual analysis may lead to suggestions of structure or point to information in data that might be missed or overlooked if the analysis is based only on summary statistics (Chatterjee & Hadi, 2006).

The most commonly used method employing leaf dimensions for estimating the area of a coffee leaf was proposed by Barros *et al.* (1973). They obtained a linear model without intercept in which the LA (dependent variable) can be calculated as the product of the leaf length (L) and width (W) (independent variables) multiplied by a *K* coefficient (0.667). Another two similar equations [LA = 0.63LW, with the *K* coefficient derived as the ratio of LA to the product of L and W (Awantramani & Gopalakrishna, 1965); and LA = 0.220 + 0.649LW, with coefficients estimated using least-square linear regression analysis (Rey & Alvarez, 1991)] were also proposed for estimating the area of a coffee leaf. However, these equations have been chosen based only on values obtained for the coefficient of determination (R^2) and the standard error of estimates, without assessing their accuracy. In fact Tavares-Júnior *et al.* (2002), comparing the method of Barros *et al.* (1973) against a standard method for measuring LA, found that it was biased, significantly underestimating the true LA. In addition, these models were just developed for a single arabica coffee genotype, and hence no information is available on whether or not such models can be successfully extrapolated to other coffee genotypes. From the above, a simple and rigorously tested accurate model for LA estimation of coffee is necessary. In this study, we thus aimed to evaluate the current models (Barros *et al.*, 1973; Rey & Alvarez, 1991), as well as to propose a reliable and accurate model using non-destructive measurements of L and/or W, for estimating the surface area of leaves of different sizes from different cultivars of both arabica and robusta coffee.

Materials and methods

For model construction, a total of 1563 healthy leaves were collected in the Coffee Germplasm Bank of the Federal University of Viçosa (20°47′S, 20°52′W, 700 m a.s.l.), southeastern Brazil, in February 2007 (end of the growing season). Leaves were sampled from eight coffee genotypes including those from arabica coffee (cv. Yellow Bourbon, $n = 189$; cv. Catuaí 2147, $n = 144$; cv. Mundo Novo 2190, $n = 230$ and cv. Typica, $n = 176$), robusta coffee (cv. Conilon 513, $n = 187$ and cv. CC 3580, $n = 174$) as well as from the Hybrid of Timor (a natural interspecific cross between *C. arabica* and *C. canephora*, $n = 226$) and the segregant population F_2 421-4 also derived from interspecific crossing between those species ($n = 226$). To validate the model, an independent data set of 388 leaves was taken for both species and their hybrids in November 2007 (growing season). For constructing and validating the model, the leaves sampled were randomly selected from different levels of the tree canopy, removed from the branches and brought to the laboratory. The sampled leaves encompassed the full spectrum of measurable leaf sizes (0.3–263 cm^2) for each genotype. As a rule, coffee leaves tend to be elliptical, accuminate and shortly acute at the base; generally, leaves of robusta coffee are bigger than those of arabica coffee.

Leaf L was measured from the leaf tip to the point at which the lamina is attached to the petiole. Leaf W corresponded to the maximum W perpendicular to the blade mid-rib. Measurements were made to the nearest millimetre. The actual area of an individual leaf (taken as a reference) was measured with an area meter (area measurement system; Delta-T Devices, Cambridge, UK).

Several linear and non-linear regression models between L and W dimensions and LA were run for each genotype. Equality of a set of linear regression models among genotypes was examined using the test for model identity (slopes and intercepts) described by Graybill (2000). Here, we present only linear and power relationships because they better fit coffee LA than other tested models. When power models $\left(Y = \beta_0 x^{\beta_1}\right)$ were used, both dependent and independent variables were subjected to logarithmic transformation before analysis. To compare the estimated LA with the observed LA, graphical procedures described by Graybill (2000) were adopted. Statistical criteria for model selection were based on the *F*-test, coefficient of determination, standard error of estimates and dispersion pattern of residuals. These criteria allowed us to evaluate the occurrence of bias and model precision and accuracy. In this regard, bias denotes the statistical difference between a population mean or test results and an accepted reference or true value, accuracy represents the overall distance between the estimated (or observed) value and the true value and

Table 1 Analysis of variance for model $\left(Y = \beta_0 x^{\beta_1}\right)$ identity test in which x is the product of length and width. Both dependent and independent variables were log-transformed before analysis. Data derived from the calibration data set ($n = 1563$ leaves). Source of variation (SV), degrees of freedom (d.f.), sum of squares (SS), mean squares (MS) and calculated *F* (F_{calc})[a]

SV	d.f.	SS	MS	F_{calc}
Parameters	(16)	(3377.2199)	—	
Reduction (β_s')	2	3377.1977	—	
Reduction (H_o)	14	0.0222	0.00159	1.26
Residual	1546	1.9518	0.00126	
Total	1562	3379.1717		

[a] $F_{5\%}(14, \infty) = 1.69$.

Table 2 Statistical models, regression coefficients, standard errors of estimates (SE), coefficients of determination adjusted for the degrees of freedom $\left(R_a^2\right)$, degrees of freedom of residuals (R-d.f.), residual sum of squares (R-SS) and calculated F (F_{calc}) and equations of leaf area as a function of linear dimensions of leaves (length, L, and width, W) of diverse coffee genotypes

Model	Coefficients	SE	R_a^2	R-d.f.	R-SS	F_{calc}	Estimator of LA $(\hat{Y})$
$Y = \beta_1 x + e_i$[a]	$\beta_1 = 0.667$	—	0.9889	430	—	—	$(\hat{Y}) = 0.667LW$
$Y = \beta_0 + \beta_1 x + e_i$[b]	$\beta_0 = 0.2197$ $\beta_1 = 0.649$	—	0.9425	878	—	—	$(\hat{Y}) = 0.2197 + 0.649LW$
$Y = \beta_1 x + e_i$	$\beta_1 = 0.70125$	0.000797	0.9957	1562	9855.56	360 018.6	$(\hat{Y}) = 0.70125LW$
$Y = \beta_0 + \beta_1 x + e_i$	$\beta_0 = -0.52665$ $\beta_1 = 0.70604$	0.001165 0.094246	0.9958	1561	9659.45	367 120.6	$(\hat{Y}) = -0.52665 + 0.70604LW$
$Y = \beta_0 x^{\beta_1} e_i$	$\beta_0 = 0.66256$ $\beta_1 = 1.01156$	0.006137 0.001871	0.9959	1561	9615.62	368 800.9	$(\hat{Y}) = 0.66256(LW)^{1.01156}$
$Y = \beta_0 x^{\beta_1} e_i$	$\beta_0 = 0.21318$ $\beta_1 = 2.11756$	0.005606 0.009113	0.9781	1561	49 827.86	69 044.8	$(\hat{Y}) = 0.21318L^{2.11756}$
$Y = \beta_0 x^{\beta_1} e_i$	$\beta_0 = 2.11246$ $\beta_1 = 1.86425$	0.030150 0.006936	0.9833	1561	38 076.51	91 018.5	$(\hat{Y}) = 2.11246W^{1.86425}$

[a] Barros *et al.* (1973).
[b] Rey & Alvarez (1991).

precision denotes the variation among individual measurements (variance) (Walther & Moore, 2005).

All the statistical analyses were performed using the software Statistica version 7.0 (StatSoft, Tulsa, OK, USA), DataFit version 8.0.32 (Oakdale Engineering, Oakdale, PA, USA) and Minitab 14 (Minitab Inc., State College, PA, USA).

Results

Separate regression models that estimate LA from L, W and the L–W product did not differ significantly ($\alpha = 0.05$) among genotypes for the L–W power model we developed (Table 1) and for the other power models (Table 2). Therefore, data for these genotypes were pooled and single regression models were fitted to the combined data. The coefficient of determination adjusted for the degrees of freedom $\left(R_a^2\right)$ for all tested models was highly significant, exceeding 97.5% (Table 2). Noteworthy, the linear models without (LA = 0.667LW) and with (LA = 0.220 + 0.649LW) intercept, respectively, developed by Barros *et al.* (1973) and by Rey & Alvarez (1991) showed good precision (low SE; Table 2) but were biased (−0.051 and −0.074, respectively, $P < 0.0001$), leading to a significant underestimation of LA (Fig. 1). The linear model (LA = 0.701LW) we developed had also good precision but its bias (0.002, $P < 0.0001$) led to an overestimation of LA (Fig. 1). It is

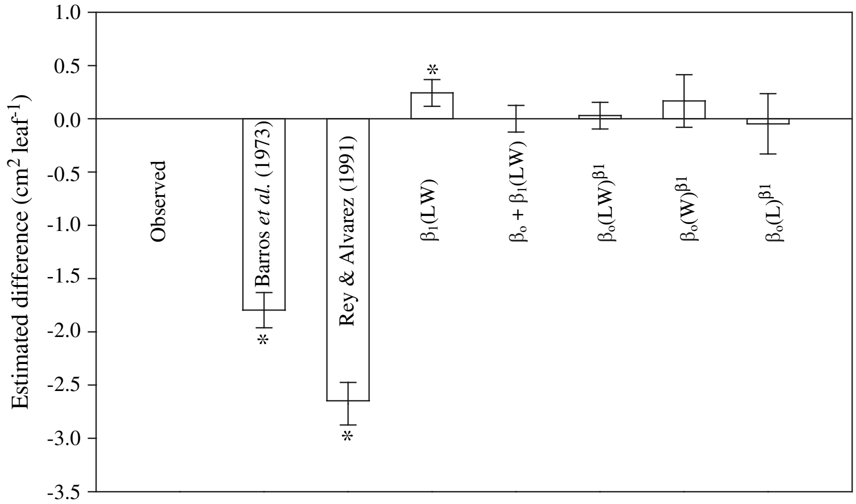

Figure 1 Statistical analysis of the deviation of the estimated area from the observed area for an individual leaf. Leaf area for coffee was estimated using several models in which β_0 and β_1 are coefficients. Vertical bars denote means and spreads denote 95% confidence intervals of the difference. L, length; W, width (see further details in the text).

also worth noting that the model assuming an elliptical shape of leaves failed to adequately predict LA, that is, it also led to an overestimation of LA (data not shown) probably because coffee leaves are not perfectly elliptical.

The biased model LA = 0.701LW was modified by introducing a constant β_0 (-0.527). Compared with this model, the modified model (LA = $-0.527 + 0.706$LW) showed similar and high R_a^2 value but was more precise, accurate and unbiased (Table 2; Fig. 2A). Nonetheless, the dispersion pattern of residuals did not follow a normal distribution (i.e. a heteroscedastic behaviour) when the L–W product approaches zero (Fig. 2A, inset). This feature translates into underestimation of LA, especially for small leaves with L < 5 cm or W < 2 cm (relative errors of 60%).

The best estimation of the area of a coffee leaf could be obtained through the power model LA = 0.6626 $(\mathrm{LW})^{1.0116}$. This model, in addition to showing high R_a^2, accuracy and precision, and lack of bias (Table 2; Figs 1

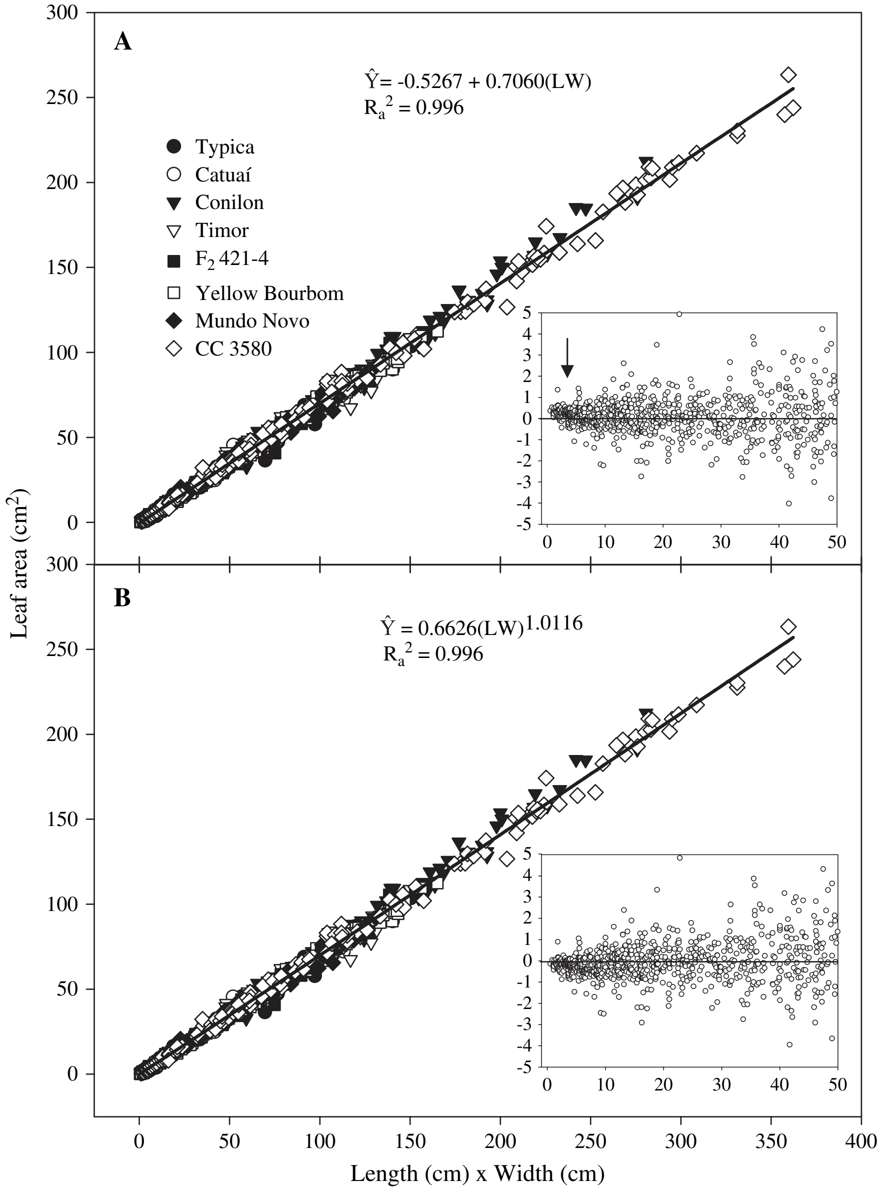

Figure 2 The relationship between estimated leaf area and leaf dimensions (L, length and W, width) for eight coffee genotypes using a linear (A) and a power (B) models. The initial portion of analysis of dispersion pattern of residuals for the respective models is shown in the insets. Arrow indicates an underestimated area for small leaves when using the linear model (A, inset).

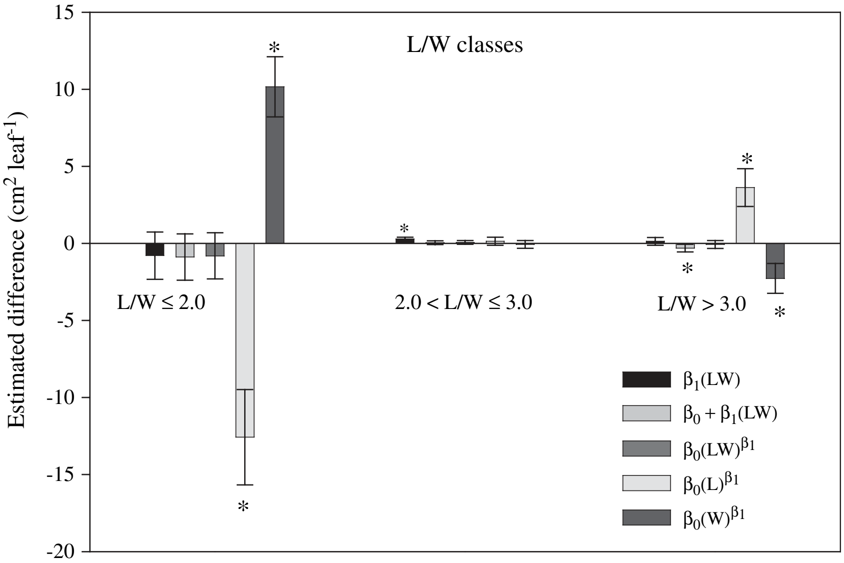

Figure 3 Difference between the real leaf area and the calculated leaf area for the different models and classes of length-to-width (L/W) ratio. Vertical bars denote means and spreads denote 95% confidence intervals. Statistical differences ($P < 0.05$, *t*-test) are denoted by asterisks.

and 2B), could also rigorously estimate LA regardless of leaf dimensions (Fig. 2B, inset, Fig. 3). The reduction of estimation errors was manifested through the increasing F_{calc} and the decreasing sum of residual squares compared with the linear regression models (Table 2).

A simplification of the L–W power model was tested by taking a single leaf dimension (Fig. 4A), either L (LA = $0.2132L^{2.1176}$) or W (LA = $2.1125W^{1.8643}$) (Table 2). When developing such a model, we noted that the residual scatter plot had a heteroscedastic behaviour (shown

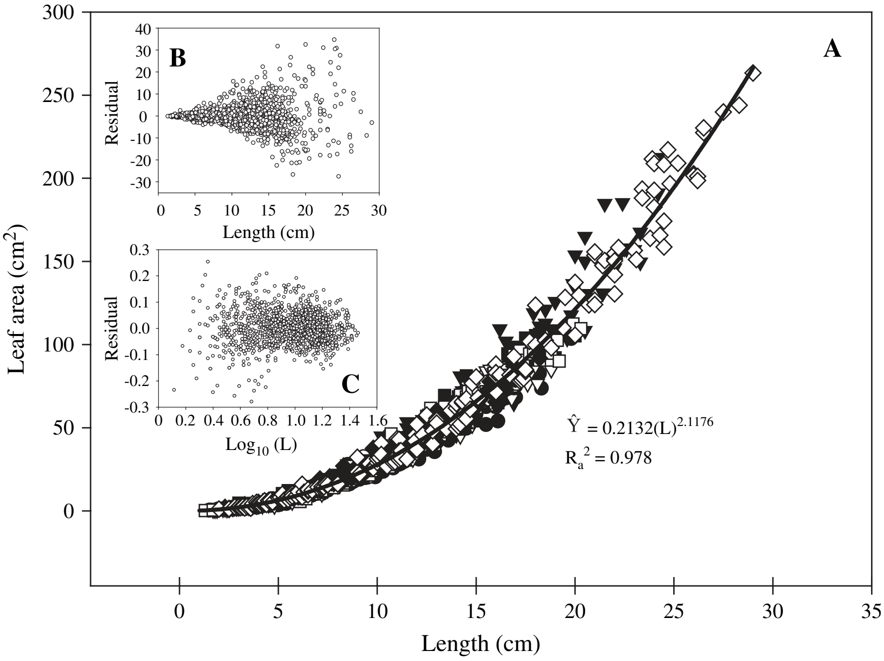

Figure 4 The relationship between estimated leaf area and leaf length (L) for eight coffee genotypes using a power model (A). Symbols as in Fig. 2. In the insets, dispersion pattern of residuals relative to the model $\left(Y = \beta_0(L)^{\beta_1}\right)$ (B) and dispersion pattern of log-transformed residuals of that model presenting a homoscedastic distribution (C) are shown.

for L in Fig. 4B but also valid for W, data not shown). In this context, the ordinary least square method for estimating the model coefficients cannot be reliably used unless corrective action for removing heteroscedasticity needs to be adopted (Chatterjee & Hadi, 2006). This was performed by logarithmic transformation of data resulting in residual scatter plot with random mean and variance distribution (Fig. 4C). The developed power models based on single dimension factors when compared with the L–W power model showed a slight decrease in R_a^2 and lower precision but with similar accuracy (Table 2; Fig. 1). In fact, the LA estimated by the power models strongly agreed with the observed values, with $R^2 = 0.977$, 0.982 and 0.996, respectively for L, W and L–W models (Fig. 5).

In the model validation analysis, the β_0 and β_1 coefficients were compared using the calibration and validation data sets. As shown in Fig. 6, the predictive power of the L and L–W models performed well, whereas the W model failed to adequately predict the area of an individual leaf.

Despite being appropriate for estimating LA, the power model using L as the single leaf dimension factor may be limited (which was not the case of the two leaf dimension model), particularly for leaves with L/W ratio <2.0 or >3.0; in this range, an underestimation and an overestimation of LA were found for that model (being the opposite when using W as the single factor; Fig. 3). It should be mentioned, however, that only 3.3% of the sampled leaves randomly showed L/W ratio <2.0 and 4.7% with L/W ratio >3.0 (data not shown).

Because we calibrated a model derived from a large sample size, the minimum number of leaves necessary to develop our model was examined. This was performed using the Minitab software that randomly selected 10 samples for each predefined sample size, and then the β_0 and β_1 coefficients obtained for each sample size were compared with the coefficients found for the entire data set. It is feasible to reduce the amount of leaf samples to about 200 to develop a reliable regression model to predict the area of coffee leaves (Fig. 7).

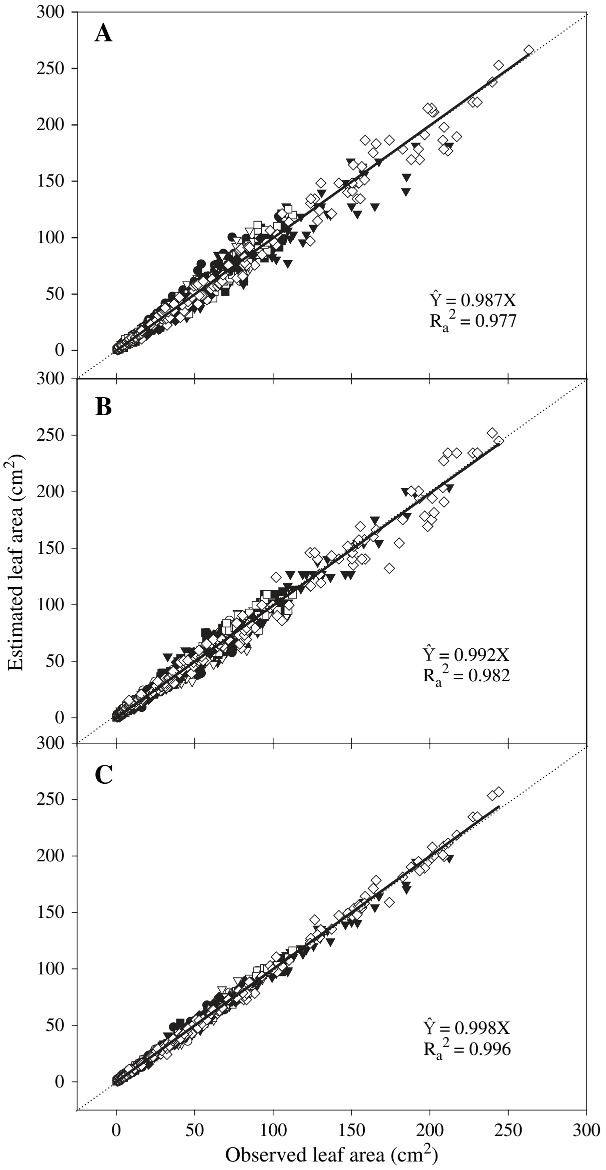

Figure 5 The relationship between estimated and observed leaf area for eight coffee genotypes. Symbols as in Fig. 2. Leaf area was estimated using power models (see Table 2) based on both single leaf dimensions, length (A) or width (B) and the product of two leaf dimensions, length and width (C). Dotted line represents the 1:1 relationship.

Discussion

In this paper, we described how LA can be accurately estimated from simple non-destructive measurements in coffee, for which a power model incorporating either the leaf L alone or both leaf L and W together was developed. We also demonstrated that the currently used linear allometric models for estimating the area of a coffee leaf are inappropriate. Although showing relatively high R_a^2 and high precision, they are not accurate, causing an underestimation of LA.

Regardless of its high apparent precision and accuracy, the linear model with intercept we developed has also proven to be inadequate, particularly for small leaves with L/W ratio >3.0 (Figs 2B and 4). It should be mentioned that when the dispersion pattern of residuals follows a heteroscedastic behaviour and no corrective action is taken, application of ordinary least squares to the raw data will result in estimated coefficients that lack precision in a theoretical sense. Furthermore, the

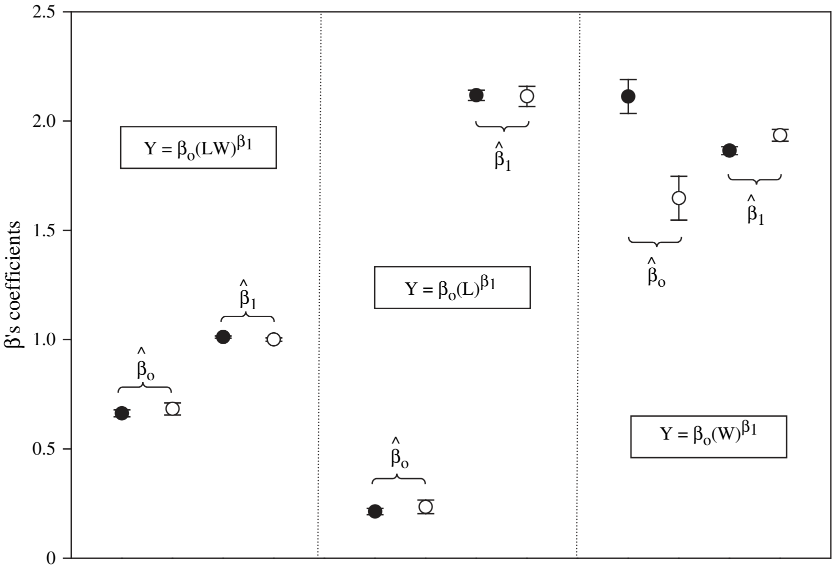

Figure 6 Statistical comparisons of the β_0 and β_1 coefficients of the power models $\left(Y = \beta_0(x)^{\beta_1}\right)$ using the calibration (black circles; n = 1563) and the validation (white circles; n = 388) data sets. Vertical bars represent the 99% confidence intervals for β coefficient means.

estimated standard errors of the regression coefficients are often underestimated, giving a false perception of accuracy (Chatterjee & Hadi, 2006). In addition, the introduction of a constant – intercept – in linear models always causes problems for estimating LA, particularly for the smaller, expanding leaves, a fact that can be safely neglected with increasing leaf size. By contrast, the power model based on two leaf dimensions could be used to estimate LA with high precision and accuracy, independently of leaf size. This model was considered the most adequate for estimating LA of other perennial crops like black pepper (Kandiannan *et al.*, 2002), grapevine (Williams & Martinson, 2003) and dwarf coconut tree (Sousa *et al.*, 2005).

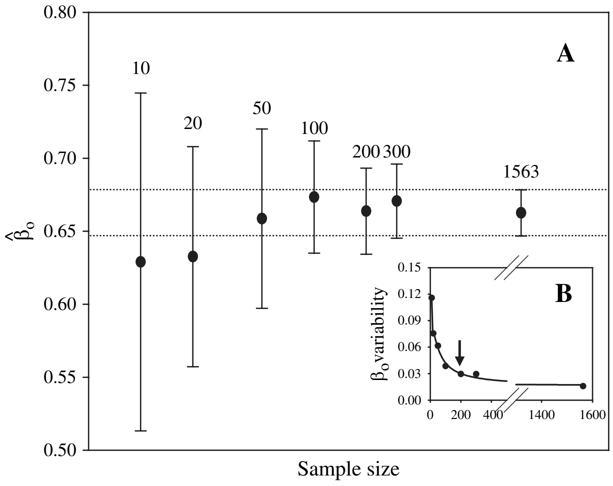

Figure 7 β_0 estimation in the model = $\beta_0(LW)^{\beta_1}$ as a function of sample size. Vertical bars represent the 99% confidence intervals for the β_0 coefficient means. The numbers above the bars denote the sample size (n = 10) (A). The entire data set was composed by a sample of 1563 leaves. In the inset, inverted polynomial regression for β_0 deviation as a function of leaf size showing the approximate minimum sample size (indicated by the arrow) to develop a regression model to reliably predict the area of coffee leaves (B).

Differences in leaf shape may occur even within individuals, as found in *Salix viminalis* (Verwijst & Wen, 1996). In the current study, however, we did not find consistent differences in leaf shape (as judged from the L/W ratio; Verwijst & Wen, 1996) within and between coffee species (data not shown). In contrast, random differences in leaf shape among the sampled leaves were detected. Whereas caution must be exercised when using the L model for estimating the area of leaves with L/W ratios below 2.0 or above 3.0, the L–W power model allowed reliable LA estimations regardless of differences in leaf shape (Fig. 3). Nonetheless, bearing that limitation in mind, the single leaf dimension power model incorporating L may be an interesting option because it requires measurement of only one leaf dimension, thus simplifying measurement procedures (Williams & Martinson, 2003), an important aspect specially in field when a large number of leaves has to be monitored. Furthermore, when using the L model there is obviously no need to measure W, which is not as easily gauged as L because the need of considering an imaginary perpendicular line to the leaf L and also because fully expanded coffee leaves are not always perfectly flat. These facts must be judiciously considered when measuring W to obtain reliable LA estimations using the L–W model.

In conclusion, we developed simple predictive models to accurately estimate the area of leaves for all the genotypes we investigated, including the most widely grown cultivars of both arabica and robusta coffee in Brazil. Power models better fit coffee LA than linear models, and thus the earlier proposed models should be avoided for LA estimation. Compared with the L–W power model, the L-based model also allowed reliable LA estimations but at the expense of some loss of precision, as particularly found for 8% of the sampled leaves (L/W < 2.0; L/W > 3.0). In contrast, irrespective of cultivar and leaf developmental stages and shapes, we demonstrated that the L–W power model is an excellent and non-destructive tool for studying leaf growth and development in coffee.

Acknowledgements

Scholarships granted by the Brazilian Government (W.C. Antunes and M.F. Pompelli – CNPq, and D.M. Carretero – CAPES) are gratefully acknowledged. Special thanks are due to Geraldo de Jesus for his assistance in the selection of leaves in Germplasm Bank of the Federal University of Viçosa. This study was partially funded by the Minas Gerais State Research Foundation, Brazil (Grant CAG APQ-3995-3.08/7).

References

Awantramani D.A., Gopalakrishna H.K. (1965) Measurement of leaf area (*Coffea arabica* L.). *Indian Coffee*, **29**, 61–64.

Barros R.S., Maestri M., Vieira M., Braga-Filho L.J. (1973) Determinação da área de folhas do café (*Coffea arabica* L. cv. 'Bourbon Amarelo'). *Revista Ceres*, **20**, 44–52.

Blanco F.F., Folegatti M.V. (2005) Estimation of leaf area for greenhouse cucumber by linear measurements under salinity and grafting. *Scientia Agricola*, **62**, 305–309.

Chatterjee S., Hadi A.S. (2006) *Regression Analysis by Example*. Hoboken, NJ, USA: Wiley.

DaMatta F.M., Ramalho J.D.C. (2006) Impacts of drought and temperature stress on coffee physiology and production: a review. *Brazilian Journal of Plant Physiology*, **18**, 55–81.

Davies A.P., Govaerts R., Bridson D.M., Stoffelen P. (2006) An annotated taxonomic conspectus of genus *Coffea* (Rubiaceae). *Botanical Journal of the Linnean Society*, **152**, 465–512.

Graybill F.A. (2000) *Theory and Application of the Linear Model*. Pacific Grove, CA, USA: Duxbury Press.

Kandiannan K., Kailassam C., Chandaragiri K., Sankaran N. (2002) Allometric model for leaf area estimation in black pepper (*Piper nigrum* L.). *Journal of Agronomy and Crop Science*, **188**, 138–140.

Lizaso J.I., Batchelor W.D., Westgate M.E. (2003) A leaf area model to simulate cultivar-specific expansion and senescence of maize leaves. *Field Crops Research*, **80**, 1–17.

Lu H.-Y., Lu C.-T., Wei M.-L., Chan L.-F. (2004) Comparison of different methods for nondestructive leaf area estimation in taro. *Agronomy Journal*, **96**, 448–453.

Rey R., Alvarez P. (1991) Evaluación de diferentes ecuaciones de regresión en la estimación del área foliar del cafeto en vivero a partir de sus medidas lineales. *Agrotecnia de Cuba*, **23**, 69–74.

Sousa E.F., Araújo M.C., Posse R.P., Detmann E., Bernardo S., Berbert P.A., Santos P.A. (2005) Estimating the total leaf area of the green dwarf coconut tree (*Cocos nucifera* L.). *Scientia Agricola*, **62**, 597–600.

Tavares-Júnior J.E., Favarin J.L., Dourado-Neto D., Maia A.H.N., Fazuoli L.C., Bernardes M.S. (2002) Análise comparativa de métodos de estimativa de área foliar em cafeeiro. *Bragantia*, **61**, 199–203.

Verwijst T., Wen D.Z. (1996) Leaf allometry of *Salix viminalis* during the first growing season. *Tree Physiology*, **16**, 655–660.

Walther B.A., Moore J.L. (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.

Williams L. III, Martinson T.E. (2003) Nondestructive leaf area estimation of 'Niagara' and 'DeChaunac' grapevines. *Scientia Horticulturae*, **98**, 493–498.