

## NUMBER OF LEAVES NEEDED TO MODEL LEAF AREA IN JACK BEAN PLANTS USING LEAF DIMENSIONS

## NÚMERO DE FOLHAS PARA MODELAR A ÁREA FOLIAR DE FEIJÃO DE PORCO POR DIMENSÕES FOLIARES

Alberto CARGNELUTTI FILHO<sup>1</sup>; Marcos TOEBE<sup>2</sup>; Cláudia BURIN<sup>3</sup>;  
Bruna Mendonça ALVES<sup>4</sup>; Ismael Mario Márcio NEU<sup>4</sup>

1. Engenheiro Agrônomo, Professor do Departamento de Fitotecnia, Doutor em Agronomia, Universidade Federal de Santa Maria - UFSM, Santa Maria, RS, Brasil. alberto.cargnelutti.filho@gmail.com; 2. Engenheiro Agrônomo, Doutorando em Agronomia - UFSM, Santa Maria, RS, Brasil; 3. Engenheira Florestal, Mestranda em Agronomia - UFSM, Santa Maria, RS, Brasil; 4. Engenheira Agrônoma, Doutorando em Agronomia - UFSM, Santa Maria, RS, Brasil; 5. Acadêmico de Agronomia - UFSM, Santa Maria, RS, Brasil.

**ABSTRACT:** Leaf area estimation models based on linear leaf dimensions are an important method because their application is not destructive to the leaves. For these models to be reliable, it is important that the estimation of model parameters is accurate, and for that to occur, the models must be generated using an adequate sample size (number of leaves). The objective of this study was to determine the number of leaves necessary to accurately model the leaf area of jack beans (Y), determined by digital photos, according to the width of the central leaflet (x), by a power model ( $Y = ax^b$ ) generated through an iterative process. Accordingly, an experiment was performed in a 256 m<sup>2</sup> area. A total of 745 leaves were randomly collected at six different crop development stages (29, 43, 57, 73, 87 and 101 days after emergence). Each leaf was comprised of a left, central and right leaflet. The width of the central leaflet (x) was measured on the 745 leaves. Leaf area (sum of the area of the left, central and right leaflets; Y) was then determined using a digital photo method. The number of leaves necessary for the estimation of the parameters a and b and the coefficient of determination ( $R^2$ ) of the power model were determined through resampling with replacement. The power model ( $\hat{Y} = 4.2049x^{1.8215}$ ,  $R^2 = 0.9701$ ), based on the width of the central leaflet was determined to be adequate for estimating jack bean leaf area. Data collected from a sample of 200 leaves were determined to be sufficient for constructing an accurate power model for the leaf area of jack beans (Y) as a function of the width of the central leaflet (x), based on determinations of leaf area using digital photos.

**KEYWORDS:** *Canavalia ensiformis*. Sample size. Power model.

### INTRODUCTION

Jack bean (*Canavalia ensiformis*) crops have a high nitrogen fixation capacity and weed control efficiency (FERNANDES et al., 1999), which reduces soil and nitrogen loss (PANSACK et al., 2008), and also have heavy-metal phytoremediation potential for elements such as lead (PEREIRA et al., 2010) and copper (ZANCHETA et al., 2011). The seeds of the jack bean are nutritional and can be used in human and animal food (BENÍTEZ et al., 2013; MARIMUTHU; GURUMOORTHY, 2013; SASIPRIYA; SIDDHURAJU, 2013) as long as they are correctly processed. Extracts from raw and processed jack bean seeds also exhibit antioxidant properties (SOWNDHARARAJAN et al., 2011), and the urease produced by the plant may be an important alternative remedy for the control of fungal diseases (POSTAL et al., 2012) and insects (DEFFERRARI et al., 2011).

Leaf area is important for determining plant growth and has an direct relationship with light interception, photosynthetic efficiency, evapotranspiration rates and plant response to fertilizers and irrigation (BLANCO; FOLEGATTI, 2005). Therefore, leaf area should be measured accurately. Leaf area can be measured directly or indirectly. In general, the direct measurement of the leaf area is destructive and requires special equipment,

such as leaf area integrators, scanners or digital cameras (BLANCO; FOLEGATTI, 2005), which are expensive and overly complicated for use in simpler and more basic studies (DEMIRSOY et al., 2005). Alternatively, leaf area can be estimated indirectly using leaf area estimation models based on linear leaf dimensions such as length, width or length × width.

Models generated based on the linear leaf dimensions are generally accurate, as has been shown for grapevines (WILLIAMS; MARTINSON, 2003), strawberry (DEMIRSOY et al., 2005), coffee (ANTUNES et al., 2008), Barbados nut (POMPELLI et al., 2012), forage turnip (CARGNELUTTI FILHO et al., 2012a), velvet bean (CARGNELUTTI FILHO et al., 2012b) and jack bean (TOEBE et al., 2012). These models allow for the quantification of leaf area in a non-destructive way with a lower financial investment. Additionally, leaf area can be estimated for different growth and crop development periods.

Of the different types of leaf area estimation models that have been studied for use with agricultural crops, power models ( $Y = ax^b$ ), in which leaf area (Y) is a function of a linear leaf dimension (x) (such as length, width or length × width), have been determined to be accurate and adequate for several crops. Antunes et al. (2008), Cargnelutti Filho et al. (2012a) and Pompelli et al.

(2012) tested several linear and non-linear models and found that power models based on the product of leaf length  $\times$  width were the most precise models for estimating leaf area in coffee, forage turnip and the Barbados nut, respectively. Williams and Martinson (2003) tested nine different leaf area estimation models for two grapevine cultivars and found that a power model based on leaf width was the best-fitting model. Power models based on the width of the central leaflet were found to be adequate for the estimation of leaf area in velvet bean (CARGNELUTTI FILHO et al., 2012b) and jack bean (TOEBE et al., 2012) crops.

Antunes et al. (2008) and Pompelli et al. (2012) found that the best models were based on the product of leaf length  $\times$  width. These authors suggested that, although models based on a single dimension (length or width) exhibit high accuracy, they may produce biased estimates, especially in the case of small or large leaves, and the residuals may not follow a normal distribution. However, Williams and Martinson (2003), Cargnelutti Filho et al. (2012b) and Toebe et al. (2012) all recommend using models based on a single linear leaf dimension (width) due to the ease of measurement and the high level of accuracy. According to Williams and Martinson (2003) and Zhang and Pan (2011), using a single linear leaf dimension also prevents potential problems of colinearity between the independent variables in the model.

For leaf area estimation models to have high accuracy and reliability, and to avoid bias in the models, it is necessary to use a representative sample (a certain number of leaves) from the plant population used to generate the model. Zhang and Pan (2011) used samples of between 202 and 476 leaves to generate leaf area estimation models for a tree species. Pompelli et al. (2012) used 1,200 leaves from Barbados nut plants to generate leaf area models and used two independent samples of 300 leaves for the validation of the generated models. According to these authors, previous models for the Barbados nut, which were generated using samples of 89 and 250 leaves, may have not estimated leaf area correctly due an insufficient number of leaves. The authors aimed to determine the sample size (number of leaves) necessary for generating an accurate model, and they found 415 leaves to be an adequate sample size for the Barbados nut. Antunes et al. (2008) used 1,563 leaves to generate a leaf area estimation model for coffee and 388 leaves for validating the model. These authors also concluded that approximately 200 leaves was an adequate sample size for generating power models for this species. For the velvet bean (CARGNELUTTI FILHO et al., 2012b) and jack bean (TOEBE et al., 2012), samples of 650 and 605 leaves, respectively, were used to generate leaf area models, and samples of 140 leaves were used for validation of the models.

Leaf area estimation models for the jack bean have been previously generated and validated. A power model ( $\hat{Y} = 3.7046x^{1.8747}$   $R^2 = 0.9757$ ) based on the width of the central leaflet ( $x$ ), was found to adequately estimate leaf area ( $Y$ ), determined through the use of digital photos (TOEBE et al., 2012). However, these authors did not determine the adequate sample size (number of leaves) needed for the generation of the model. In addition, the parameters  $a$  and  $b$  and the coefficient of determination ( $R^2$ ) of the  $Y = ax^b$  function were determined following a logarithmic transformation for linearization of the  $Y = ax^b$  function (STEEL et al., 1997). A similar procedure has been used by other authors (WILLIAMS; MARTINSON, 2003; ANTUNES et al., 2008). The use of an iterative process for nonlinear models may have resulted in a better fit of the models generated in those studies, minimizing the error sum of squares, especially when the number of leaves was small. Thus, the objective of our study was to determine the sample size (number of leaves) needed for the construction of a power model ( $Y = ax^b$ ) generated by an iterative process that would estimate leaf area ( $Y$ ) in jack beans as a function of the width of the central leaflet ( $x$ ), using digital photos to measure total leaf area.

## MATERIAL AND METHODS

### Field experiment

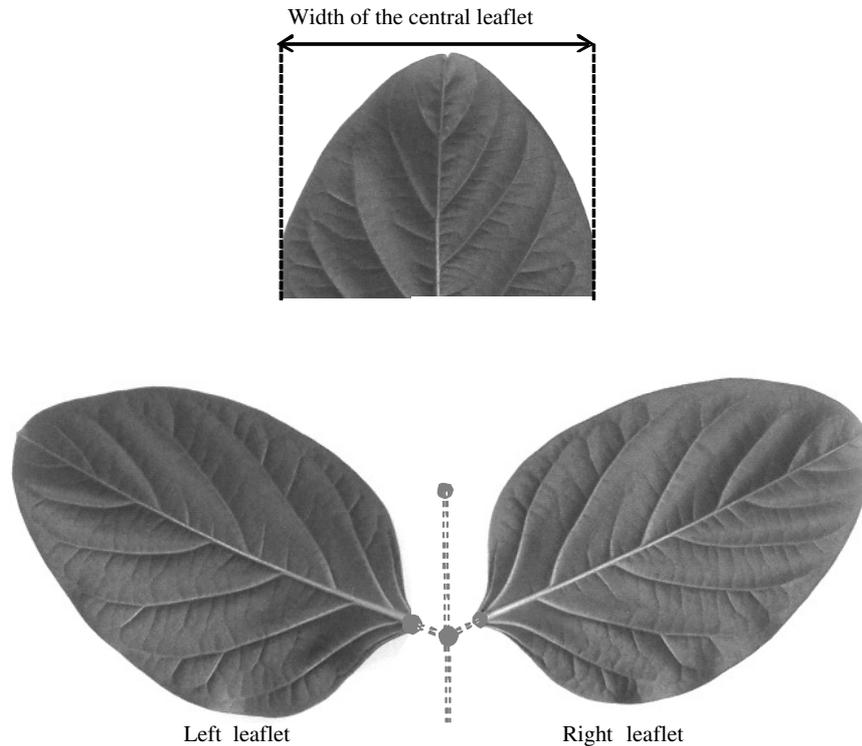
In this study, a field experiment was performed in a 256 m<sup>2</sup> area planted with jack bean (*Canavalia ensiformis*). The spacing between rows was 0.5 m, with 0.125 m between plants in the row, for a total of 16 plants m<sup>-2</sup>. Sowing was performed on November 12, 2010 with a base fertilizer application of 40 kg ha<sup>-1</sup> N, 150 kg ha<sup>-1</sup> P<sub>2</sub>O<sub>5</sub> and 100 kg ha<sup>-1</sup> K<sub>2</sub>O. The day of emergence was November 22, 2010, at which time 50% of the seeds had emerged. Ninety and 95 leaves were collected 29 and 43 days after emergence (DAE), respectively. At each of the remaining sampling times (57, 73, 87 and 101 DAE), 140 leaves were randomly collected in the experimental area. A total of 745 leaves, each comprised of three leaflets (left, central and right), were collected.

### Variable measurement and model generation

The width of the central leaflet ( $x$ ) of each of the 745 leaves was measured with a millimeter ruler (Figure 1). All leaves (each comprised of three leaflets) were photographed with a Sony DSC-W110 digital camera. The resulting images were processed using Sigma Scan Pro v. 5.0 (Jandel Scientific, 1991) software to determine leaf area (sum of the leaf area of the left, central and right leaflets;  $Y$ ) according to the digital photo method. Measures of central tendency and variability and kurtosis and skewness coefficients were calculated for each measured variable ( $x$  and  $Y$ ) for each sampling period (29, 43, 57, 73, 87 and 101 DAE) and using the data from the

745 leaves. Additionally, the normality of the data was tested using the Kolmogorov-Smirnov test, and frequency

histograms and scatterplots were generated.



**Figure 1.** Representation of a leaf of the jack bean (*Canavalia ensiformis*), comprised of a left, central and right leaflet. The location of the measurement of the width of the central leaflet ( $x$ ) is shown.

Leaf area for the 745 leaves ( $Y$ ), measured using the digital photos, was used to construct a power model ( $Y = ax^b$ ) for leaf area as a function of the width of the central leaflet ( $x$ ). The model was generated using an iterative process until convergence was achieved. The data set generated from this sample of 745 leaves was previously used in a preliminary study by Toebe et al. (2012). In that study, the authors used data from 605 leaves to generate linear, quadratic and power models (by linearization of the model variables) of leaf area of jack beans as a function of the width, length and length  $\times$  width of the central leaflet (TOEBE et al., 2012). Data from 140 leaves were used in model validation. Using digital photos to determine leaf area, the authors concluded that a power model ( $Y = ax^b$ ) as a function of the central leaflet ( $x$ ) was the most adequate model for estimating leaf area ( $Y$ ). In this complementary study, our objective was to determine the number of leaves needed to model  $Y$  as a function of  $x$  using a power model ( $Y = ax^b$ ). We used data from all the leaves in the sample ( $n=745$ ) to generate our models so that the models were based on the most representative sample possible. Furthermore, the model had been validated in the previous study and needed no further validation.

#### Determination of sample size (number of leaves)

In this study, the sample size (number of leaves) required to model the leaf area of jack bean ( $Y$ ) (determined using digital photos) as a function of the width of the central leaflet ( $x$ ) using a power model ( $Y = ax^b$ ) was determined through resampling with replacement. Estimation of the model parameters was performed using an iterative process. For resampling, we used 791 simulated sample sizes, starting with an initial sample size of 10 leaves (which was considered the minimum sample size for construction of the model). The remaining sample sizes were tested in increments of one, up to a sample size of 800 leaves. Therefore, sample sizes of 10, 11, 12, ..., 800 leaves were tested.

For each simulated sample size, 3,000 resamplings with replacement were performed. Estimates of the parameters  $a$  and  $b$  and the coefficient of determination ( $R^2$ ) of the power model ( $Y = ax^b$ ) were obtained for each resampling. To minimize the error sum of squares, the estimates of  $a$ ,  $b$  and  $R^2$  were obtained by iteration until convergence was achieved. The values established for the convergence criterion were: iterations=200; step size=1; tolerance=0.0000000001. Thus, for each sample size, 3,000 estimates of  $a$ ,  $b$  and  $R^2$  were obtained and were determined the 2.5% percentile, mean and 97.5% percentile. The 95% confidence interval

(ACI) was calculated as the difference between the 97.5% percentile and 2.5% percentile.

For  $a$ ,  $b$  and  $R^2$ , the ACI for the smallest sample size (10 leaves,  $ACI_{10}$ ) was considered to be the reference condition of 100%, i.e., the maximum ACI value, representing minimal accuracy of the estimates of  $a$ ,  $b$  and  $R^2$ . The accuracy gain ( $AG_i$ , %) was calculated by adding  $i$ th leaves ( $i = 1, 2, \dots, 790$  leaves, respectively, for the sample sizes 11, 12, ..., 800 leaves), according to  $AG_i = 100 - (ACI_i/ACI_{10}) \times 100$ , where  $ACI_i$  is the size of the 95% confidence interval for the sample sizes 11, 12, ..., 800 leaves. Was considered in this study, the accuracy gain ( $AG_i$ ) minimum of 79.41% for the estimates of  $a$ ,  $b$  and  $R^2$  as criterion for defining the sample size.

The 2.5% percentile, the mean, and the 97.5% percentile of  $a$ ,  $b$  and  $R^2$  were plotted by sample size in 10-leaf intervals for better visual representation. Because the results for each sample size were too extensive to present in a table, the accuracy gain for each 10-leaf interval was shown. Statistical analyses were performed in R (R Development Core Team, 2012) and Microsoft Office Excel<sup>®</sup> software.

## RESULTS

### Characterization of the data and the leaf area model

For the 745 jack bean leaves, the mean width of the central leaflet was 10.01 cm, and the minimum and maximum widths were 2.70 and 17.10 cm, respectively (Table 1). The average total leaf area (sum of the areas of the left, central and right leaflets) was 298.16 cm<sup>2</sup>, and the minimum and maximum total leaf area were 22.20 and 790.58 cm<sup>2</sup>, respectively. For each of the six sampling periods (29, 43, 57, 73, 87 and 101 DAE) and across all sampling periods, there was less variation in the width of the central leaflet ( $x$ ) than in total leaf area ( $Y$ ). The coefficients of variation for the 745-leaf sample were 29.84 and 49.84% for width of the central leaflet and total leaf area, respectively. Given the large variation in leaf size, the data for the width of the central leaflet were considered to be adequate for generating a leaf area estimation model, considering that our sample included leaves of several different sizes and from different developmental stages.

**Table 1.** Number of leaves ( $n$ ), minimum, maximum, mean, coefficient of variation (CV), variance, kurtosis, skewness and Kolmogorov-Smirnov test p-value for width of the central leaflet ( $x$ ) and leaf area (sum of the areas of the left, central and right leaflets) ( $Y$ ), determined using digital photos of 745 leaves of jack bean (*Canavalia ensiformis*) plants.

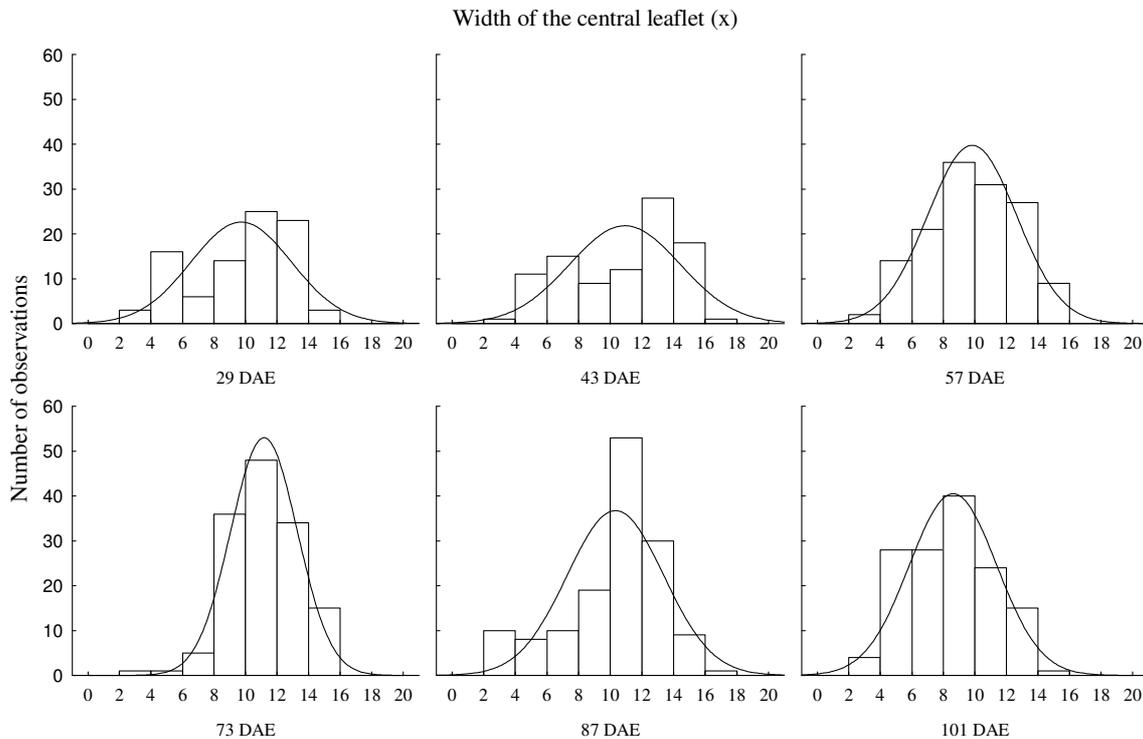
Variable	n	Minimum	Maximum	Mean	CV(%)	Variance	Kurtosis <sup>(1)</sup>	Skewness <sup>(2)</sup>	p-value
----- 29 days after emergence -----									
x	90	3.20	15.90	9.67	32.83	10.08	2.07 <sup>ns</sup>	-0.48 <sup>ns</sup>	0.15
Y	90	22.20	622.36	276.91	55.50	23620.06	1.91*	-0.12 <sup>ns</sup>	0.17
----- 43 days after emergence -----									
x	95	3.90	17.10	10.87	31.97	12.07	1.87*	-0.38 <sup>ns</sup>	0.04
Y	95	36.05	790.58	342.83	55.95	36794.16	1.83*	0.00 <sup>ns</sup>	0.07
----- 57 days after emergence -----									
x	140	3.80	15.20	9.77	28.76	7.89	2.17*	-0.09 <sup>ns</sup>	0.58
Y	140	41.45	568.23	284.01	49.15	19485.21	2.00*	0.22 <sup>ns</sup>	0.23
----- 73 days after emergence -----									
x	140	3.50	15.80	11.11	18.96	4.44	3.52 <sup>ns</sup>	-0.26 <sup>ns</sup>	0.90
Y	140	38.22	624.04	336.63	34.87	13782.94	2.68 <sup>ns</sup>	0.38 <sup>ns</sup>	0.15
----- 87 days after emergence -----									
x	140	2.70	16.20	10.27	29.60	9.24	3.17 <sup>ns</sup>	-0.86*	0.00
Y	140	24.10	670.91	314.12	46.06	20929.11	2.66 <sup>ns</sup>	-0.25 <sup>ns</sup>	0.37
----- 101 days after emergence -----									
x	140	3.80	15.10	8.56	32.21	7.60	2.13*	0.12 <sup>ns</sup>	0.49
Y	140	42.18	640.53	241.20	54.27	17137.30	2.74 <sup>ns</sup>	0.55*	0.43
----- All samplings -----									
x	745	2.70	17.10	10.01	29.84	8.93	2.38*	-0.36*	0.00
Y	745	22.20	790.58	298.16	49.84	22085.57	2.36*	0.14 <sup>ns</sup>	0.27

<sup>(1)</sup>\* Kurtosis different from three, according to a t-test ( $p \leq 0.05$ ). <sup>ns</sup> Non-significant; <sup>(2)</sup>\* Skewness different from zero, according to a t-test ( $p \leq 0.05$ ). <sup>ns</sup> Non-significant.

The data for width of the central leaflet showed mesokurtic behavior (kurtosis not different from 3) in 50% of the samplings performed (three samplings) and platykurtic behavior (kurtosis lower than 3) on the remaining samplings periods. Furthermore, in five of the samplings periods (83.33% of the data), the skewness did not differ from zero, and in four of the samplings periods (66.67% of the data) the data had a normal distribution ( $p > 0.05$ ) (Table 1). Therefore, in general, the data for the width of the central leaflet were well fitted to a normal distribution. For the total data set from 745 leaves, a kurtosis different from three, a skewness different from zero and a non-normal distribution were observed. These conditions of the rejection of the hypothesis of kurtosis equal to three, a skewness equal to zero and normality of the data resulted from the high number of observations in the data set ( $n = 745$  leaves).

For leaf area (Y), the kurtosis was not different from three and the skewness was not different from zero in

three samplings (50% of the data) and five samplings periods (83.33% of the data), respectively (Table 1). The data from each of the six samplings periods had a normal distribution ( $p > 0.05$ ). Across the entire sample of 745 leaves, the skewness was not different from zero, and the data were well fitted to a normal distribution curve. Thus, even with a high number of observations ( $n = 745$  leaves), the data for leaf area had a normal distribution. Consistent with our previous inferences, width of the central leaflet (x) and leaf area (Y) for the leaves collected during the six sampling periods (29, 43, 57, 73, 87 and 101 DAE) were, in general, well fitted to a normal distribution (Table 1). This result can also be observed in the frequency histograms for these two variables (Figures 2 and 3). Therefore, our data set appeared to be sufficient for generating models of leaf area as a function of the width of the central leaflet and for the study of sample size.



**Figure 2.** Frequency histogram and normal distribution curve of the width of the central leaflet (x), in cm, of the jack bean (*Canavalia ensiformis*) leaves collected at 29, 43, 57, 73, 87 and 101 days after emergence (DAE).

The scatterplot of width of the central leaflet (x) against leaf area (Y) for the entire data set from the 745 jack bean leaves showed a non-linear association between the two variables (Figure 4). This result indicates that the power model was adequate, in accordance with the previous report by Toebe et al. (2012). In this study, the power model generated using an iterative process was  $\hat{Y} = 4.2049x^{1.8215}$ , and the high coefficient of determination ( $R^2$

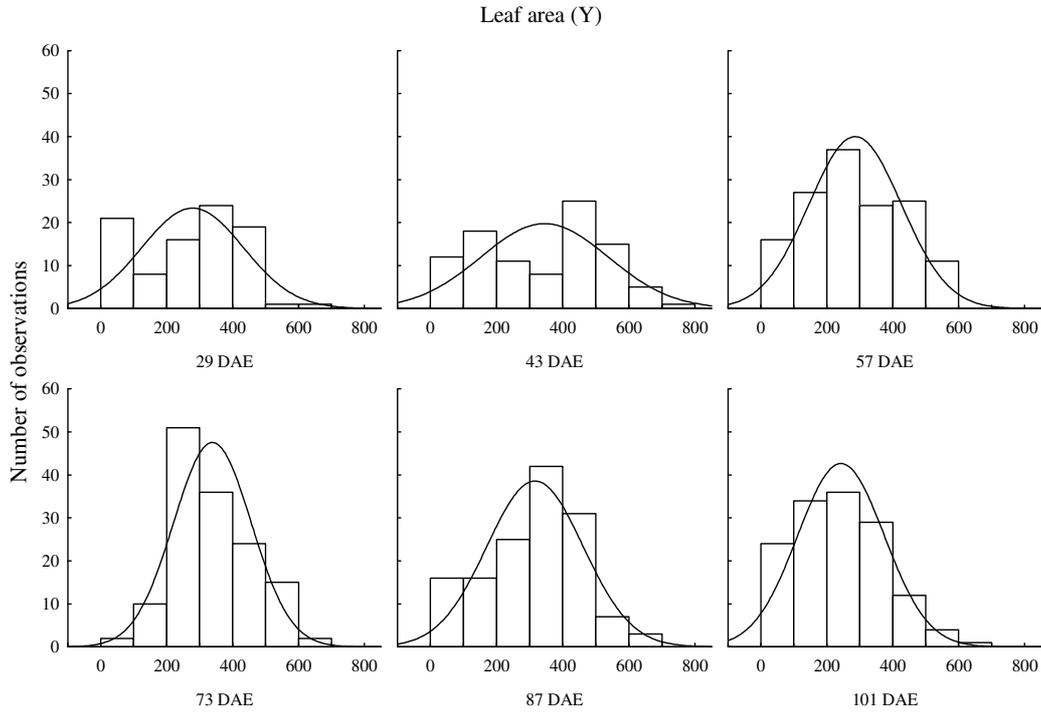
= 0.9701) indicated that the accuracy of the model was high.

**Sample size (number of leaves)**

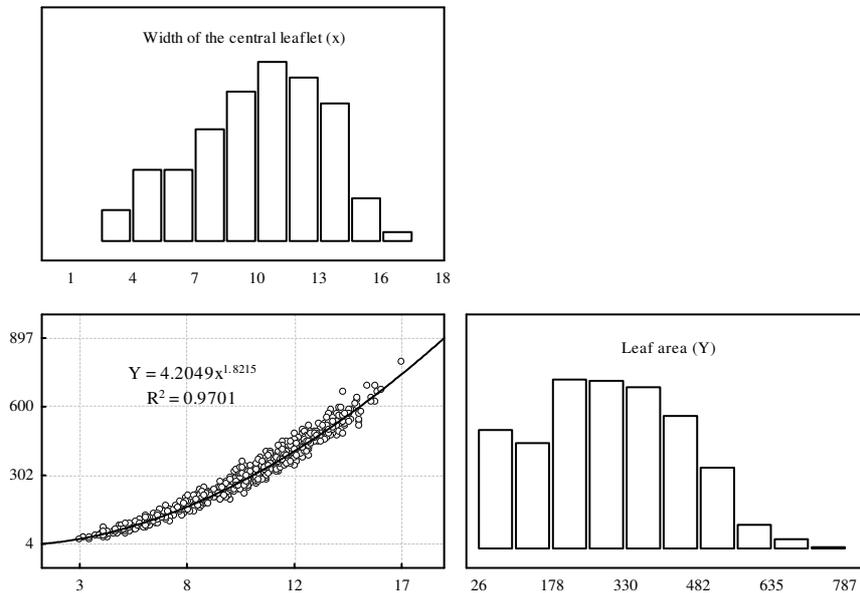
Using the average of the 3,000 estimates for parameters a and b and the coefficient of determination ( $R^2$ ) of the power model ( $Y = ax^b$ ) for the leaf area of jack bean (Y), generated by the resampling of 800 leaves with replacement, the following model was obtained:  $\hat{Y} =$

$4.2026x^{1.8220}$  with  $R^2 = 0.9702$ . The estimates obtained through resampling were similar to the estimates produced

by the model generated from the data set derived from the sample of 745 leaves ( $\hat{Y} = 4.2049x^{1.8215}$ ,  $R^2 = 0.9701$ ).



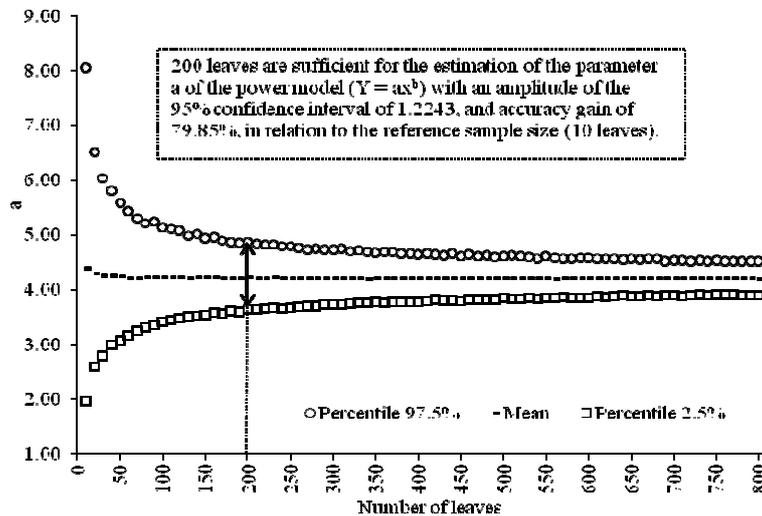
**Figure 3.** Frequency histogram and normal distribution curve of the leaf area (axis X), in  $\text{cm}^2$ , of the jack bean (*Canavalia ensiformis*) leaves collected at 29, 43, 57, 73, 87 and 101 days after emergence (DAE).



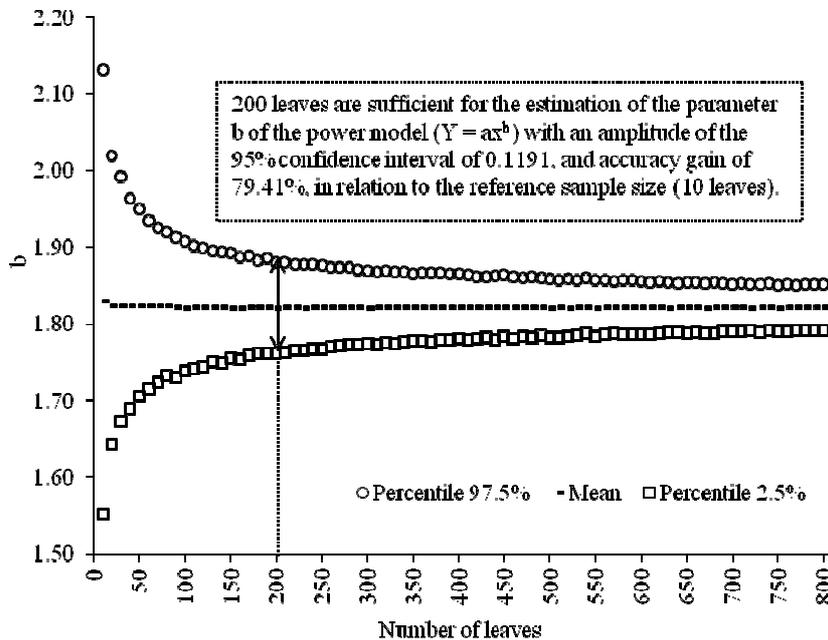
**Figure 4.** Frequency histograms for width of the central leaflet (x) (cm) and leaf area (Y) ( $\text{cm}^2$ ) (at diagonal) and corresponding scatterplots for width of the central leaflet (x) and leaf area (Y). A power model for the estimation of Y as a function of x, generated using data from 745 leaves of the jack bean (*Canavalia ensiformis*) plant collected on six different growth periods (29, 43, 57, 73, 87 and 101 days after emergence), and the coefficient of determination of the model ( $R^2$ ) are shown.

For the 3,000 resamples of 10 leaves (the smallest sample size used in this study), the 95% confidence interval (ACI) for a, b and R<sup>2</sup> was 6.0757, 0.5786 and 0.0819, respectively, and the mean for a, b and R<sup>2</sup> was 4.3915, 1.8296 and 0.9697, respectively (Figures 5, 6 and 7 and Table 2). At the opposite end of the range of sample

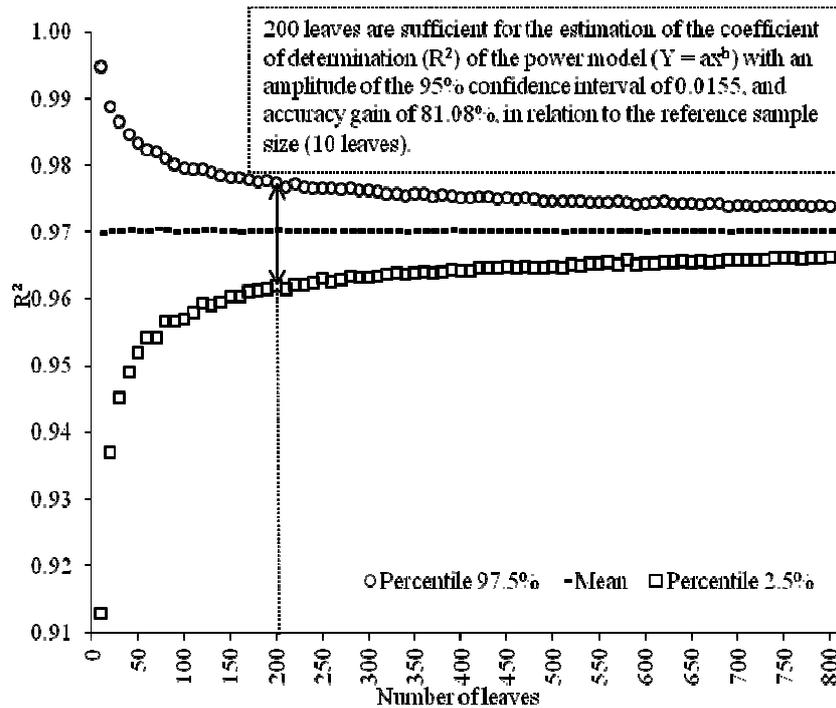
sizes, for the 3,000 resamplings of 800 leaves (the largest sample size used), the 95% confidence interval (ACI) for a, b and R<sup>2</sup> was 0.6219, 0.0604 and 0.0076, respectively. The mean for a, b and R<sup>2</sup> was 4.2026, 1.8220 and 0.9702, respectively (Figures 5, 6 and 7 and Table 2).



**Figure 5.** Percentile 2.5%, mean and percentile 97.5% for 3,000 estimates of parameter a in the power model ( $Y = ax^b$ ) for the leaf area (Y) of the jack bean (*Canavalia ensiformis*) as a function of the width of the central leaflet (x), generated using a digital photo method for leaf area. Percentiles and mean are shown for each sample size (number of leaves – ten in ten leaves).



**Figure 6.** Percentile 2.5%, mean and percentile 97.5% for 3,000 estimates of parameter b in the power model ( $Y = ax^b$ ) for the leaf area (Y) of the jack bean (*Canavalia ensiformis*) as a function of the width of the central leaflet (x), generated using a digital photo method for leaf area. Percentiles and mean are shown for each sample size (number of leaves – ten in ten leaves).



**Figure 7.** Percentile 2.5%, mean and percentile 97.5% for 3,000 estimates of the coefficient of determination ( $R^2$ ) in the power model ( $Y = ax^b$ ) for the leaf area ( $Y$ ) of the jack bean (*Canavalia ensiformis*) as a function of the width of the central leaflet ( $x$ ), generated using a digital photo method for leaf area. Percentiles and mean are shown for each sample size (number of leaves – ten in ten leaves).

**Table 2.** Values for the 95% confidence interval (ACI) and accuracy gain ( $AG_i$ , %) of the estimates of parameters  $a$  and  $b$  in the power model ( $Y = ax^b$ ) of leaf area for the jack bean (*Canavalia ensiformis*). Leaf area was determined using digital photos, and the model represents leaf area as a function of the width of the central leaflet ( $x$ ). The coefficient of determination ( $R^2$ ) for the sample sizes 10, 11, 12, ..., 800 leaves is also shown.

N° leaves	ACI	ACI	ACI	AG <sub>i</sub>	AG <sub>i</sub>	AG <sub>i</sub>	N° leaves	ACI	ACI	ACI	AG <sub>i</sub>	AG <sub>i</sub>	AG <sub>i</sub>
	a	b	R <sup>2</sup>	a	b	R <sup>2</sup>		a	b	R <sup>2</sup>	a	b	R <sup>2</sup>
10	6.0757	0.5786	0.0819	-	-	-	410	0.8586	0.0848	0.0109	85.87	85.34	86.69
20	3.9194	0.3762	0.0518	35.49	34.98	36.70	420	0.8272	0.0810	0.0107	86.38	85.99	86.99
30	3.2462	0.3196	0.0413	46.57	44.76	49.50	430	0.8201	0.0796	0.0106	86.50	86.25	87.10
40	2.8189	0.2742	0.0356	53.60	52.60	56.53	440	0.8520	0.0828	0.0103	85.98	85.69	87.39
50	2.5126	0.2442	0.0314	58.65	57.78	61.61	450	0.8175	0.0808	0.0102	86.55	86.04	87.49
60	2.2606	0.2199	0.0281	62.79	61.99	65.62	460	0.8269	0.0799	0.0104	86.39	86.19	87.35
70	2.0485	0.2017	0.0277	66.28	65.14	66.12	470	0.8063	0.0772	0.0103	86.73	86.65	87.47
80	1.8969	0.1860	0.0244	68.78	67.86	70.18	480	0.8068	0.0784	0.0103	86.72	86.46	87.36
90	1.8767	0.1817	0.0235	69.11	68.59	71.35	490	0.7806	0.0761	0.0101	87.15	86.85	87.71
100	1.7253	0.1687	0.0227	71.60	70.84	72.28	500	0.7730	0.0754	0.0098	87.28	86.97	88.00
110	1.6534	0.1602	0.0216	72.79	72.31	73.65	510	0.7853	0.0751	0.0099	87.07	87.02	87.93
120	1.6009	0.1544	0.0200	73.65	73.31	75.54	520	0.7859	0.0759	0.0095	87.06	86.88	88.37
130	1.4689	0.1442	0.0199	75.82	75.08	75.71	530	0.7451	0.0723	0.0097	87.74	87.50	88.20
140	1.4972	0.1449	0.0190	75.36	74.95	76.77	540	0.7370	0.0720	0.0091	87.87	87.55	88.87
150	1.3990	0.1375	0.0178	76.97	76.23	78.23	550	0.7666	0.0743	0.0093	87.38	87.15	88.70
160	1.3771	0.1335	0.0179	77.33	76.92	78.10	560	0.7270	0.0706	0.0091	88.03	87.79	88.94
170	1.3398	0.1296	0.0168	77.95	77.60	79.46	570	0.6979	0.0677	0.0094	88.51	88.30	88.55
180	1.2531	0.1219	0.0164	79.38	78.93	80.02	580	0.7238	0.0701	0.0087	88.09	87.89	89.42
190	1.2540	0.1228	0.0162	79.36	78.77	80.21	590	0.7265	0.0708	0.0090	88.04	87.77	89.04
200	1.2243	0.1191	0.0155	79.85	79.41	81.08	600	0.7204	0.0707	0.0090	88.14	87.78	89.07
210	1.2060	0.1177	0.0153	80.15	79.65	81.35	610	0.6995	0.0681	0.0091	88.49	88.23	88.84
220	1.1582	0.1124	0.0151	80.94	80.58	81.50	620	0.6972	0.0680	0.0092	88.53	88.25	88.81
230	1.1482	0.1117	0.0146	81.10	80.69	82.12	630	0.6853	0.0667	0.0089	88.72	88.48	89.15
240	1.1282	0.1109	0.0142	81.43	80.84	82.70	640	0.6651	0.0646	0.0087	89.05	88.83	89.39
250	1.1186	0.1089	0.0136	81.59	81.18	83.44	650	0.6787	0.0660	0.0088	88.83	88.59	89.27

260	1.0633	0.1034	0.0140	82.50	82.12	82.92	660	0.6769	0.0655	0.0086	88.86	88.67	89.50
270	1.0390	0.1021	0.0136	82.90	82.35	83.40	670	0.6771	0.0661	0.0088	88.86	88.57	89.21
280	1.0345	0.1008	0.0132	82.97	82.57	83.88	680	0.6753	0.0669	0.0086	88.89	88.44	89.51
290	1.0054	0.0975	0.0131	83.45	83.15	84.01	690	0.6370	0.0624	0.0081	89.52	89.21	90.05
300	0.9925	0.0963	0.0131	83.66	83.35	84.06	700	0.6406	0.0620	0.0083	89.46	89.28	89.92
310	1.0086	0.0961	0.0127	83.40	83.39	84.43	710	0.6452	0.0630	0.0082	89.38	89.12	89.94
320	0.9644	0.0940	0.0120	84.13	83.75	85.35	720	0.6234	0.0600	0.0081	89.74	89.63	90.13
330	0.9623	0.0942	0.0119	84.16	83.72	85.50	730	0.6382	0.0619	0.0083	89.49	89.30	89.91
340	0.9338	0.0911	0.0116	84.63	84.26	85.78	740	0.6174	0.0607	0.0079	89.84	89.50	90.30
350	0.9028	0.0873	0.0119	85.14	84.91	85.44	750	0.6214	0.0603	0.0080	89.77	89.57	90.20
360	0.9183	0.0894	0.0117	84.89	84.54	85.71	760	0.6214	0.0611	0.0079	89.77	89.43	90.40
370	0.9122	0.0898	0.0115	84.99	84.47	86.01	770	0.6105	0.0593	0.0081	89.95	89.75	90.13
380	0.8867	0.0867	0.0115	85.41	85.02	85.90	780	0.6241	0.0604	0.0078	89.73	89.57	90.53
390	0.8900	0.0852	0.0111	85.35	85.28	86.47	790	0.6140	0.0593	0.0078	89.89	89.75	90.53
400	0.8659	0.0843	0.0109	85.75	85.43	86.62	800	0.6219	0.0604	0.0076	89.76	89.56	90.77

$AG_i = 100 - (ACI_i/ACI_{10}) \times 100$ , where  $ACI_i$  is the size of the 95% confidence interval for the sample sizes 11, 12, ..., 800 leaves, and  $ACI_{10}$  is the size of the 95% confidence interval for the reference sample size, i.e., 10 leaves.

Although the mean estimates for  $a$ ,  $b$  and  $R^2$  were similar for the 10 leaf and 800 leaf samples, the larger confidence intervals obtained for the 10 leaf sample compared to the 800 leaf sample indicate that the estimation of the model parameters with the 10 leaf sample had a lower accuracy. This result indicates that an insufficient sample size may result in biased estimates of leaf area and that models generated using data from a small number of leaves should not be used to determine leaf area. Generating an accurate model requires first defining an adequate sample size.

The size of the 95% confidence intervals (ACI) for the estimates of  $a$ ,  $b$  and  $R^2$  gradually decreased with an increasing number of leaves in the sample (Figures 5, 6 and 7 and Table 2). This result was expected, and it indicates that estimation accuracy, and consequently, the chance of obtaining more reliable models, improves with an increasing number of leaves in the sample. A visual evaluation of Figures 5, 6 and 7 shows that there was a pronounced decrease in the ACI up to a sample size of approximately 200 leaves. Thereafter, ACI continued to decrease, but the decreases were smaller, indicating that the additional workload required for measuring more than 200 leaves results in negligible improvement to the accuracy of the estimation of the model parameters. A visual evaluation suggests that a sample of 200 leaves is sufficient to generate a strong power model for estimating  $a$ ,  $b$  and  $R^2$ .

An increase in sample size from 10 to 20 leaves resulted in accuracy gains of 35.49% [ $100 - (3.9194/6.0757) \times 100$ ], 34.98% [ $100 - (0.3762/0.5786) \times 100$ ] and 36.70% [ $100 - (0.0518/0.0819) \times 100$ ], for the estimates of  $a$ ,  $b$  and  $R^2$ , respectively. An increase from 10 to 30 leaves resulted in accuracy gains of  $a=46.57%$ ,  $b=44.76%$  and  $R^2=49.50%$  (Table 2). The accuracy gains with an increasing number of leaves in the sample were similar for  $a$ ,  $b$  and  $R^2$ . In addition, the gains were more pronounced with an increase from 10 to 20 leaves than with an increase from 20 to 30

leaves, and so on. Accuracy gains close to 80% ( $a=79.85%$ ,  $b=79.41%$  and  $R^2=81.08%$ ) were obtained from increasing the sample size from 10 to 200 leaves. With an increase from 10 to 800 leaves, the accuracy gains were approximately 90% ( $a=89.76%$ ,  $b=89.56%$  and  $R^2=90.77%$ ). Therefore, an increase in accuracy of only approximately 10% was gained when the sample size increased from 200 to 800 leaves (an increase of 600 leaves). Although researchers should aim to use the largest number of leaves possible to obtain maximize the reliability of the models, 200 leaves appears to be a reasonable sample size for estimating the parameters of a power model for jack bean leaves.

## DISCUSSION

### Characterization of the data set and the leaf area model

Although our experiment was performed in a single environment, we suggest that our data set was representative of leaf size in jack bean crops because we used a large number of leaves ( $n = 745$ ), we took measurements at six different crop development stages and our leaf sample showed large variation in leaf width and area. Therefore, we suggest that our data set was reliable and adequate for generating models and determining an adequate sample size (number of leaves). To construct models and analyze sample size, Pompelli et al. (2012) used 1,200 Barbados nut leaves from different sites and different seasons, and Antunes et al. (2008) used 1,563 coffee leaves from eight different genotypes. Demirsoy et al. (2005) used two strawberry cultivars to generate leaf area models and observed that the models could be used to accurately estimate leaf area for seven different tested cultivars. Blanco and Folegatti (2005) generated leaf area estimation models for cucumber based on linear dimensions of the leaves and observed that models generated from data corresponding to different levels of salinity and/or grafting conditions could be used for any

conditions (of grafting or salinity). This indicated that their data were representative of leaf size for cucumber plants.

In this study, the data used to generate the model and perform the analysis of sample size generally showed a good fit to a normal distribution curve, with the exception of a few cases in which there were slight deviations from normality. These deviations from the normal distribution curve may have been associated with the high number of observations in our data set ( $n = 745$  leaves), as previously described by Toebe et al. (2012). Therefore, we suggest that the data used in this study were reliable for generating leaf area estimation models and analyzing sample size.

Although we used the same data set as Toebe et al. (2012), the small differences in the  $a$ ,  $b$  and  $R^2$  estimates between our model ( $\hat{Y} = 4.2049x^{1.8215}$ ,  $R^2 = 0.9701$  with  $n = 745$  leaves) and the model obtained by Toebe et al. (2012) ( $\hat{Y} = 3.7046x^{1.8747}$ ,  $R^2 = 0.9757$  with  $n = 605$  leaves) can be attributed to the following: 1) In our study, the estimates were obtained using the data from all leaves ( $n=745$  leaves), whereas Toebe et al. (2012) used a random sample of 605 leaves for the generation of the model and used the remaining 140 leaves for the validation of the model, and 2) In our study, the estimates for  $a$ ,  $b$  and  $R^2$ , were obtained using an iterative process until convergence was achieved, with the goal of minimizing the error sum of squares, whereas Toebe et al. (2012) used a different process. The value of Pearson's linear correlation coefficient for the leaf area estimates of the two models was  $r = 0.9999$  ( $p=0.000$ ), indicating that both models can be used. However, the use of our model ( $\hat{Y} = 4.2049x^{1.8215}$ ,  $R^2 = 0.9701$ ), which was generated using an iterative process using the entire data set (which is more representative of jack bean leaf size), should be favored over the Toebe et al. (2012) model. Power models for the estimation of the leaf area based leaf width have also been shown to be highly accurate for other crops, as observed by Williams and Martinson (2003) for two grapevine cultivars ( $R^2 \geq 0.9632$ ) and by Cargnelutti Filho et al. (2012b) for the velvet bean ( $R^2=0.9886$ ).

### Sample size (number of leaves)

In this study, a sample of 10 leaves was insufficient to build accurate power models for the estimation of leaf area for the jack bean. The increase in the number of leaves used to generate the model reduced the confidence intervals for the parameters  $a$  and  $b$  and the coefficient of determination of the model, resulting in accuracy gains. There was a pronounced increase in the accuracy gain with increases in the sample size up to a sample of 200 leaves, beyond which the gain was negligible (Figures 5, 6 and 7 and Table 2). Based on this information, the use of a 200 leaf sample is recommended for generating power models for leaf area estimation based on the width of the central leaflet of jack bean leaves.

Models generated with data from a smaller number of leaves can be inaccurate and their parameters may be biased. Models generated with a larger number of leaves are more accurate, but the accuracy gain resulting from increasing the sample size beyond 200 leaves tends to be progressively smaller, and does not compensate for the extra resources that must be used to obtain measurements from the larger sample size. A sample size of 200 leaves was also considered optimal for the estimation of coffee leaf area (ANTUNES et al., 2008). Pompelli et al. (2012), however, recommended the use of approximately 415 leaves for accurate estimation of leaf area for the Barbados nut. According to these authors, previously published models using 89 and 250 leaf samples may not have been accurate and may have resulted in biased parameter estimates.

Although the sample size determined in this study to be adequate for the jack bean plant was similar to the sample size determined to be adequate for coffee ( $n = 200$  leaves; Antunes et al., 2008), it was smaller than the sample size suggested by Pompelli et al. (2012) for the Barbados nut. In addition to differences in modeling methods, different sample sizes may affect the independent variables used and the results for tested crops. Considering the high importance of leaf area estimation models in agronomy studies, the high number of models that are generated and the specific characteristics of each crop, we suggest that further studies of sample size must be completed to attain strong models for other agricultural crops. Such studies would lend support and reliability to published models.

### CONCLUSIONS

For the jack bean plant, the power model  $\hat{Y} = 4.2049x^{1.8215}$ ,  $R^2 = 0.9701$  based on the width of the central leaflet is an adequate model for estimating leaf area as determined using digital photos.

Measurements from a sample of 200 leaves was sufficient for generating accurate power models for estimating the leaf area ( $Y$ ) of jack beans as a function of the width of the central leaflet ( $x$ ), with leaf area determined using digital photos.

### ACKNOWLEDGEMENTS

The authors would like to thank the National Council for the Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)) and the Coordination for the Improvement of Higher Education Personnel (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)) for granting scholarships to our project. We thank the Rio Grande do Sul Research Foundation (Fundação de Amparo à Pesquisa do Estado

do Rio Grande do Sul) for financial support and a scientific initiation scholarship.

**RESUMO:** Modelos de estimação de área foliar de plantas em função das dimensões lineares das folhas são importantes, principalmente, por não haver necessidade de destruição das folhas. Para modelos fidedignos, é importante que as estimativas de seus parâmetros sejam precisas, e, para isso, devem ser gerados a partir de um número adequado de folhas. O objetivo deste trabalho foi determinar o número de folhas necessário para modelar a área foliar de feijão de porco determinada por fotos digitais (Y) em função da largura do limbo do folíolo central da folha (x), por meio do modelo potência ( $Y = ax^b$ ) gerado por processo iterativo. Para isso foi conduzido um experimento numa área de 256 m<sup>2</sup>, no qual, em seis períodos de desenvolvimento da cultura (29, 43, 57, 73, 87 e 101 dias após a emergência) foram coletadas, aleatoriamente, o total de 745 folhas. Cada folha é composta pelos folíolos esquerdo, central e direito. Nas 745 folhas foi mensurada a largura do limbo do folíolo central (x). A seguir, determinou-se a área foliar (soma da área dos folíolos esquerdo, central e direito) por meio do método de fotos digitais (Y). O número de folhas, necessário para a estimação dos parâmetros a e b do modelo potência e do coeficiente de determinação do modelo (R<sup>2</sup>), foi determinado por reamostragens, com reposição. Em feijão de porco, o modelo potência ( $\hat{Y} = 4,2049x^{1,8215}$ , R<sup>2</sup> = 0,9701) da largura do limbo do folíolo central é adequado para estimar a área foliar obtida por fotos digitais. Mensurar 200 folhas é suficiente para construir modelos precisos do tipo potência, da área foliar de feijão de porco determinada por fotos digitais (Y) em função da largura do limbo do folíolo central da folha (x).

**PALAVRAS-CHAVE:** *Canavalia ensiformis*. Dimensionamento amostral. Modelo potência.

## REFERENCES

- ANTUNES, W. C.; POMPELLI, M. F.; CARRETERO, D. M.; DAMATTA, F. M. Allometric models for non-destructive leaf area estimation in coffee (*Coffea arabica* and *Coffea canephora*). **Annals of Applied Biology**, Warwick, v. 153, n. 1, p. 33-40, 2008. <http://dx.doi.org/10.1111/j.1744-7348.2008.00235.x>
- BENÍTEZ, V.; CANTERA, S.; AGUILERA, Y.; MOLLÁ, E.; ESTEBAN, R. M.; DÍAZ, M. F.; MARTÍN-CABREJAS, M. A. Impact of germination on starch, dietary fiber and physicochemical properties in non-conventional legumes. **Food Research International**, Amsterdam, v. 50, n. 1, p. 64-69, 2013. <http://dx.doi.org/10.1016/j.foodres.2012.09.044>
- BLANCO, F. F.; FOLEGATTI, M. V. Estimation of leaf area for greenhouse cucumber by linear measurements under salinity and grafting. **Scientia Agricola**, Piracicaba, v. 62, n. 4, p. 305-309, 2005. <http://dx.doi.org/10.1590/S0103-90162005000400001>
- CARGNELUTTI FILHO, A.; TOEBE, M.; BURIN, C.; FICK, A. L.; CASAROTTO, G. Estimativa da área foliar de nabo forrageiro em função de dimensões foliares. **Bragantia**, Campinas, v. 71, n. 1, p. 47-51, 2012a. <http://dx.doi.org/10.1590/S0006-87052012000100008>
- CARGNELUTTI FILHO, A.; TOEBE, M.; BURIN, C.; FICK, A.L.; NEU, I.M.M.; FACCO, G. Estimação da área foliar de mucuna cinza por meio de método não destrutivo. **Ciência Rural**, Santa Maria, v. 42, n. 2, p. 238-242, 2012b.
- DEFFERRARI, M. S.; DEMARTINI, D. R.; MARCELINO, T. B.; PINTO, P. M.; CARLINI, C. R. Insecticidal effect of *Canavalia ensiformis* major urease on nymphs of the milkweed bug *Oncopeltus fasciatus* and characterization of digestive peptidases. **Insect biochemistry and molecular biology**, Amsterdam, v. 41, n. 6, p. 388-399, 2011.
- DEMIRSOY, H.; DEMIRSOY, L.; ÖZTÜRK, A. Improved model for the non-destructive estimation of strawberry leaf area. **Fruits**, Paris, v. 60, n. 1, p. 69-73, 2005. <http://dx.doi.org/10.1051/fruits:2005014>
- R Development Core Team. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, 2012.
- FERNANDES, M. F.; BARRETO, A. C.; EMÍDIO FILHO, J. Fitomassa de adubos verdes e controle de plantas daninhas em diferentes densidades populacionais de leguminosas. **Pesquisa Agropecuária Brasileira**, Brasília, v. 34, n. 9, p. 1593-1600, 1999. <http://dx.doi.org/10.1590/S0100-204X1999000900009>

Jandel Scientific, 1991. User's Manual.

MARIMUTHU, M.; GURUMOORTHY, P. Physicochemical and functional properties of starches from Indian Jack bean (*Canavalia ensiformis*), an underutilized wild food legume. **Journal of Chemical and Pharmaceutical Research**, Rajasthan, v. 5, n. 1, p. 221-225, 2013.

PANSAK, W.; HILGER, T. H.; DERCON, G.; KONGKAEW, T.; CADISCH, G. Changes in the relationship between soil erosion and N loss pathways after establishing soil conservation systems in uplands of Northeast Thailand. **Agriculture, Ecosystems & Environment**, Amsterdam, v. 128, n. 3, p. 167-176, 2008. <http://dx.doi.org/10.1016/j.agee.2008.06.002>

PEREIRA, B. F. F.; ABREU, C. A.; HERPIN, U.; ABREU, M. F. ; BERTON, R. S. Phytoremediation of lead by jack beans on a Rhodic Hapludox amended with EDTA. **Scientia Agricola**, Piracicaba, v. 67, n. 3, p. 308-318, 2010. <http://dx.doi.org/10.1590/s0103-90162010000300009>

POMPELLI, M. F.; ANTUNES, W.C.; FERREIRA, D.T.R.G.; CAVALCANTE, P.G.S.; WANDERLEY-FILHO, H.C.L.; ENDRES, L. Allometric models for non-destructive leaf area estimation of *Jatropha curcas*. **Biomass and Bioenergy**, Amsterdam, v. 36, n. 1, p. 77-85, 2012. <http://dx.doi.org/10.1016/j.biombioe.2011.10.010>

POSTAL, M.; MARTINELLI, A. H. S.; BECKER-RITT, A. B.; LIGABUE-BRAUN, R.; DEMARTINI, D. R.; RIBEIRO, S. F. F.; PASQUALI, G.; GOMES, V. M.; CARLINI, C. R. Antifungal properties of *Canavalia ensiformis* urease and derived peptides. **Peptides**, Amsterdam, v. 38, n. 1, p. 22-32, 2012. <http://dx.doi.org/10.1016/j.peptides.2012.08.010>

SASIPRIYA, G.; SIDDHURAJU, P. Evaluation of growth performance, serum biochemistry and haematological parameters on broiler birds fed with raw and processed samples of *Entada scandens*, *Canavalia gladiata* and *Canavalia ensiformis* seed meal as an alternative protein source. **Tropical animal health and production**, Dordrecht, v. 45, n. 3, p. 811-820, 2013.

SOWNDHARARAJAN, K.; SIDDHURAJU, P.; MANIAN, S. Antioxidant activity of the differentially processed seeds of Jack bean (*Canavalia ensiformis* L. DC). **Food Science and Biotechnology**, New York, v. 20, n. 3, p. 585-591, 2011. <http://dx.doi.org/10.1007/s10068-011-0083-9>

STEEL, R. G. D.; TORRIE, J. H.; DICKEY, D. A. **Principles and procedures of statistics a biometrical approach**, 3rd ed. McGraw-Hill, Nova York, 1997.

TOEBE, M.; CARGNELUTTI FILHO, A.; BURIN, C.; FICK, A.L.; NEU, I.M.M.; CASAROTTO, G.; ALVES, B. M. Modelos para a estimação da área foliar de feijão de porco por dimensões foliares. **Bragantia**, Campinas, v. 71, n. 1, p. 37-41, 2012. <http://dx.doi.org/10.1590/S0006-87052012005000010>

WILLIAMS, L.; MARTINSON, T. E. Nondestructive leaf area estimation of "Niagara" and "DeChaunac" grapevines. **Scientia Horticulturae**, Amsterdam, v. 98, n. 4, p. 493-498, 2003. [http://dx.doi.org/10.1016/S0304-4238\(03\)00020-7](http://dx.doi.org/10.1016/S0304-4238(03)00020-7)

ZANCHETA, A. C. F.; ABREU, C. A. DE; ZAMBROSI, F. C. B.; ERISMANN, N. DE M.; LAGÔA, A. M. M. A. Fitoextração de cobre por espécies de plantas cultivadas em solução nutritiva. **Bragantia**, Campinas, v. 70, n. 4, p. 737-744, 2011.

ZHANG, L.; PAN, L. Allometric models for leaf area estimation across different leaf-age groups of evergreen broadleaved trees in a subtropical forest. **Photosynthetica**, Praha, v. 49, n. 2, p. 219-226, 2011. <http://dx.doi.org/10.1007/s11099-011-0027-x>