

Estimating Fresh Weight of Individual Pea Shoots Using Measurable Morphological Characteristics

Yun Kong, Xiangyue Kong, and Youbin Zheng

School of Environmental Sciences, University of Guelph, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada

Additional index words. biomass, model, multi-linear regression, nondestructive method, *Pisum sativum*

Abstract. Nondestructive estimation of individual shoot fresh weight (FW) from its measurable morphological traits is useful for a wide variety of purposes in pea shoot production. To predict individual shoot FW, nine regression models in total were developed, including two power models using stem diameter (SMD) or stem length (SML) as a variable, and seven linear models using part or all the following variables: SMD, SML, leaflet length (LL), leaflet width (LW), stipule length (SEL), and stipule width (SEW). Among the nine models, the 6-variable linear equation had the highest coefficient of determination, $R^2 = 0.92$, indicating it is most effective at explaining the variation in FW. The linear equations including only one variable, SMD or SML, were equally the least effective as nonlinear equations (i.e., power models). This finding suggests that there was a linear rather than nonlinear relationship between FW and the morphological variables. During stepwise regression, SEW and LW together were first removed from the 6-variable linear models without reducing the R^2 , and then SEL, SMD, SML were further removed one-by-one, which reduced the R^2 from 0.92 to 0.90, 0.85, and 0.71, respectively. The result suggests that SMD, SML, SEL, and LL were the most important four predictor variables for multivariable linear regression models to estimate FW, an idea that was also supported by path analysis. For the four linear models with 1–4 predictor variables from stepwise regression, the prediction accuracy of FW was evaluated based on the agreement between the predicted and measured values using another independent dataset. The 4- and 3-variable linear models (i.e., $FW = -1.437 + 0.276 \text{ SMD} + 0.010 \text{ SML} + 0.022 \text{ LL} + 0.013 \text{ SEL}$ and $FW = -1.383 + 0.308 \text{ SMD} + 0.011 \text{ SML} + 0.030 \text{ LL}$, respectively) were selected for their more accurate prediction than 1- and 2-variable linear models and relatively simpler forms than a 6-variable linear model. Although the prediction accuracy can be potentially affected by air temperature, light conditions, and harvesting time, the multilinear regression model is an effective approach for estimating fresh weight of individual pea shoots using its measurable morphological traits.

Pea shoots are young, tender vine tips of garden peas (*Pisum sativum* L.), and they have been recognized as a popular specialty vegetable in some parts of Asia and Africa (Lim, 2012). They are eaten fresh, stir fried, lightly steamed, or sautéed (Miles et al., 2018). Nowadays, consumption of pea shoots as a part of a healthy diet has become increasingly popular worldwide (Miles et al., 2018; Santos et al., 2014), because they are believed to be rich in health-promoting phytochemicals, especially anti-

oxidants such as vitamin C, carotenoids, and various phenolic compounds (Liu et al., 2014; Santos et al., 2014).

Pea shoots can be harvested many times throughout the growing season (Miles et al., 2018). However, pea shoot production has been also limited by a few factors. One of the limiting factors is that the harvesting is labor intensive. Currently, pea shoots are mainly harvested manually, which contributes greatly to production costs in labor-expensive regions (Miles et al., 2018). Another factor is the fluctuation of yield among different harvests, because the growth of pea shoots is not uniform. This growth factor increases the difficulty in predicting crop yield before each harvest.

Nondestructive estimation of individual shoot fresh weight (FW), using its measurable morphological traits, is useful for a wide variety of purposes in pea shoot production. For example, it could potentially be used for robotic harvesting in identifying and selecting the proper target shoots with the aid of a vision system (Ruiz-Altisent et al., 2004), because quick estimation of plant sizes from

their two-dimensional images is becoming possible (Golzarian et al., 2011). In addition, it is important to predict the yield of pea shoots before each harvest, so the growers can determine if they are likely to fulfill market orders in advance of harvest. Developing a nondestructive approach to estimate individual shoot weight is the first step to predict crop yield, because a destructive approach is not appropriate for long-lived plants with complex structures (Spencer et al., 2006).

Both nonlinear and linear models have been used for estimation of plant biomass from measurable morphological traits. For nonlinear models, power functions, with the general form of $y = ax^b$, are very useful in relating plant weight to its height or diameter (Fehrmann and Kleinn, 2006; Pilli et al., 2006; Spencer et al., 2006). In the above functions, x is normally breast-height diameter for estimating aboveground biomass of trees (Pilli et al., 2006), and plant height for estimating aboveground biomass of aquatic plants (Spencer et al., 2006; Thursby et al., 2002). However, the power functions in the above studies are used for estimating dry weight of aboveground biomass. This raised a question about the adaptation of the nonlinear functions to estimate FW of shoot tips (e.g., pea shoots). Nevertheless, in the estimation of fresh biomass of tea's shoot tip, a linear relationship has been obtained between shoot weight and some morphological variables such as shoot length, leaf number, and bud length (Sun et al., 2007). However, tea is a woody plant, and its shoot tip has different morphological traits from pea shoots, so even when using a linear regression model, the included variables may be different between the two species.

Although quite a few studies have been performed to estimate plant biomass (especially dry biomass) from measurable morphological traits by using both nonlinear and linear models, the related information has so far been unavailable for pea shoots. The objective of this study was to develop effective regression models for estimating FW of individual pea shoots based on its measurable morphological traits, by screening and selecting suitable model forms and variables to include.

Material and Methods

Snow peas (*Pisum sativum* L. 'China six') were grown in three greenhouse compartments located in the Edmund C. Bovey building at the University of Guelph, Guelph, ON, Canada (lat. 43°33'N, long. 80°15'W). One compartment had higher air temperature (22 to 24 °C) than the other two compartments (12 to 14 °C). For the high-temperature compartment, plants were grown under no supplemental lighting (SL). For each low-temperature compartment, plants were grown under five light conditions: no SL and four SL light levels (i.e., 50, 80, 110, and 140 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$). The SL was provided from light-emitting diode lights (Pro Series 325;

Received for publication 14 Apr. 2020. Accepted for publication 8 May 2020.

Published online 17 June 2020.

This research was supported by the Natural Sciences and Engineering Research Council of Canada.

We thank David Llewellyn for providing technical support in this work, and Devdutt Kamath for helping with editing the manuscript.

Y.Z. is the corresponding author. E-mail: yzheng@uoguelph.ca.

This is an open access article distributed under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

LumiGrow, Inc., Emeryville, CA) that had a photon flux density ratio of 4:1 for red to blue light, and peak wavelengths of 665 and 440 nm, respectively. The lights, controlled by an Argus system (Argus Control Systems Ltd., Surrey, BC), were turned on/off at 16/0.5 h before dusk, to achieve a 16-h photoperiod.

Samples were collected from the pea shoots harvested on 19 Jan., 30 Jan., and 11 Feb. 2015 (96 samples from each harvest). For each harvest, 16 pea shoots were sampled separately from the high-temperature compartment (i.e., no SL), and from each of the five zones with different light conditions (i.e., no SL, and SL levels of 50, 80, 110, and 140 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) in the two low-temperature compartments. Because there was no visible morphological difference among the four levels of SL (i.e., 50, 80, 110, and 140 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$), the pea shoots sampled from these four zones were mixed together and considered as those from SL. A total of 288 pea shoots were collected from the above three greenhouse compartments.

Typically, each of the harvested pea shoots had one fully unfolded leaf and a shoot tip (Fig. 1), except for those harvested on 11 Feb., when there were flower buds or flowers on some pea shoots. For each harvested shoot, its FW was determined by a scale, and then its shoot length (SML), leaflet length (LL) and width (LW), and stipule length (SEL) and width (SEW) were measured by a ruler, and the shoot diameter (SMD) by a caliper (Fig. 1). The SML was measured from the cutting position to the shoot tip. The SMD was measured at the cutting position. Considering the irregular shape of a pea stem, two diameter measurements at cross directions were performed for each shoot, and these measurements were presented as an average value of two measurements. The leaf size traits (i.e., LL, LW,

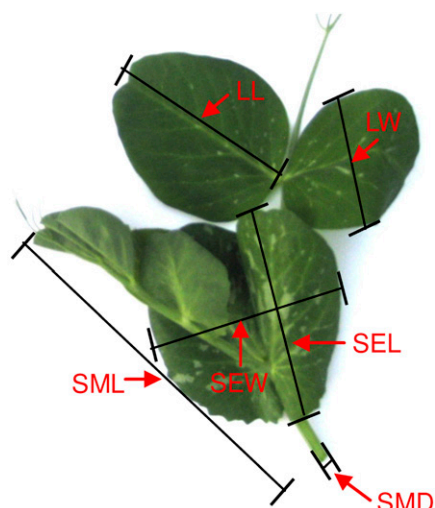


Fig. 1. Harvested individual pea shoot and measured morphological traits. SML = stem length; SMD = stem diameter; LL = leaflet length; LW = leaflet width; SEL = stipule length; SEW = stipule width.

SEL, and SEW) were measured at maximum length or width position.

For each measured variable, the minimum, first quartile, median, third quartile,

and maximum values, and the coefficient of variation (cv) were calculated for the samples from each of the two temperature conditions (i.e., high vs. low), the two light conditions

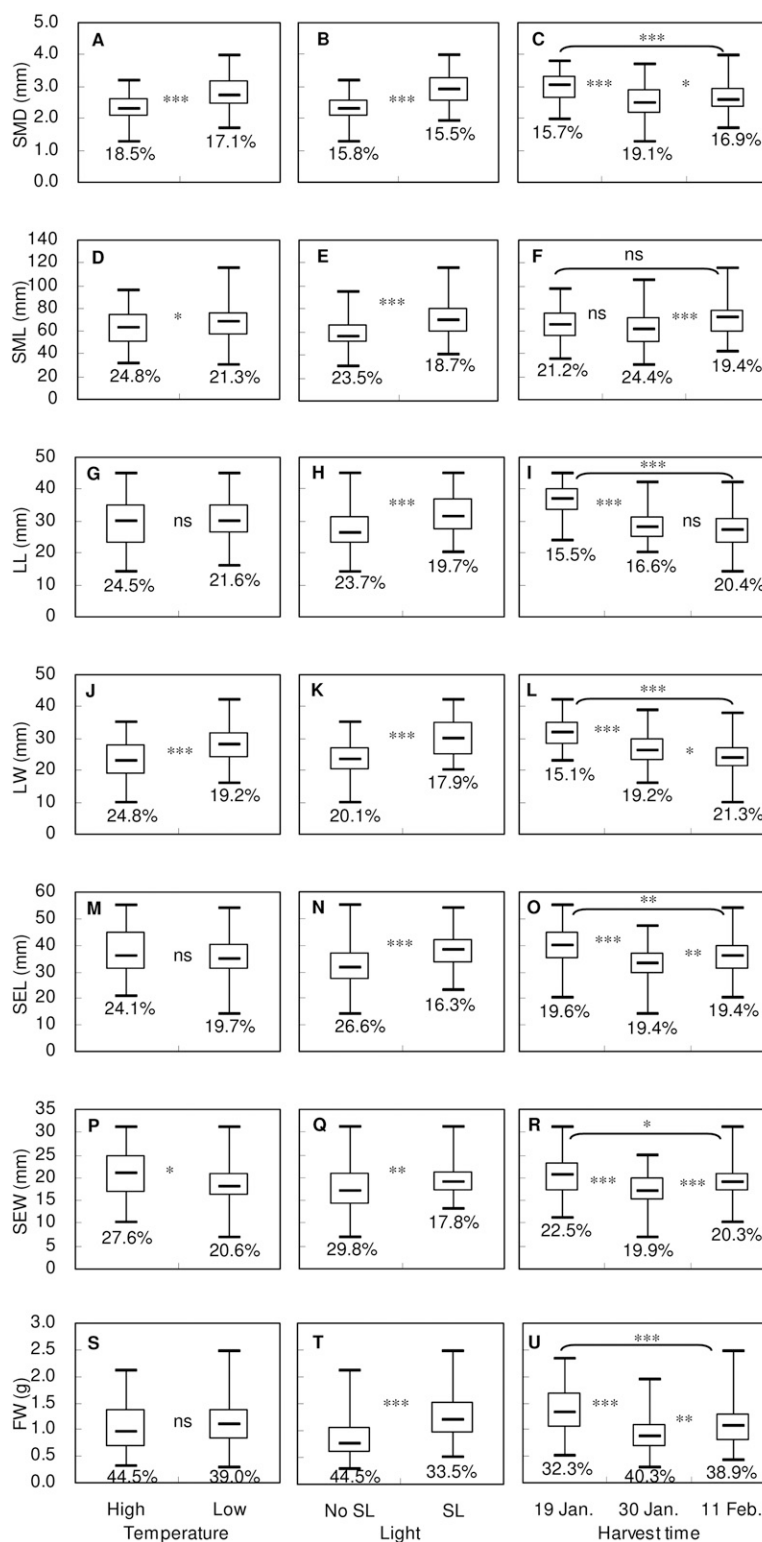


Fig. 2. Descriptive statistics of the measured variables in pea shoots sampled from different environment conditions and different harvest time. The bottom and top of the whiskers are the minimum and maximum values, respectively. The bottom and top of the boxes are the first quartile and third quartile value, respectively. The short lines inside the boxes are the median values; $n = 48$ and 240 for high and low temperature; 96 and 192 for no SL and SL, and 96 for each harvest date, respectively. The percentage number at the bottom of each box is cv value. Inside each panel, between every two environment conditions, or harvest dates, the mark of ns, *, **, or *** indicates that the difference is not significant, or significant at $P \leq 0.05$, 0.01 , or 0.001 , respectively.

(i.e., no SL vs. SL), or the three harvest dates (i.e., 19 Jan., 30 Jan., and 11 Feb.), and then presented as box-plot graphs. The differences in each variable between different greenhouse environment conditions or dif-

ferent harvest dates were compared using a *t* test.

Using a randomization procedure, half of the samples (i.e., 144 samples) collected from each greenhouse environment condition (i.e.,

temperature or light) or harvest date were used for model construction, and another half of the samples were used for model validation. For the data for model construction or validation, the minimal, maximum, and mean values, the standard deviation (SD), and the CV were calculated for each variable.

For the model construction data, the correlation between all the variables was analyzed first. Then, path analysis was carried out to determine the direct and indirect effects of the six variables (i.e., SMD, SML, LL, LW, SEL, and SEW) on the FW of pea shoots. After that, five linear models with 6, 4, 3, 2, and 1 predictor variables were developed using stepwise regression. In addition, using SML or SMD as a single predictor variable, two nonlinear models (i.e., power function) and two linear models were constructed. The coefficient of determination (R^2), the root mean square error (RMSE), *F*-value, and *P* value were calculated for each equation using Data Processing System Software (DPS version 7.05; Refine Information Tech. Co., Hangzhou, China).

Four linear models with 1–4 predictor variables from stepwise regressions were selected for evaluation of prediction accuracy using the validation data. The pea shoot FW values obtained using the above four linear models were plotted against the measured FW values using linear regression ($Y = kX$; *k* is slope), and the accuracy of the models was evaluated based on the R^2 and the *k* value for the regressions. Also, the agreement between predicted values and measured ones was evaluated by calculating the relative bias, estimated by the mean of differences between the measured and estimated values (MD) and the standard deviation of the differences (SDD) (Mokhtarpour et al., 2010).

Based on the above results, two of the linear models with better prediction accuracy were selected to evaluate the model flexibility using pea shoots sampled from different environmental conditions or different harvest dates. For the selected two linear models, three goodness of fit indicators, the mean absolute error (MAE), RMSE, and the index of agreement (*d*) were calculated using the formulas provided by Mokhtarpour et al. (2010), and compared among different environmental conditions or different harvest dates. Also, for the pea shoots harvested from different environmental conditions or different harvest dates, the predicted FW values

Table 1. Descriptive statistics of the measured variables used to derive and validate the models for estimating fresh weight of pea shoots.

Variable	Shoot no.	Minimum	Maximum	Mean	SD	CV (%)
Data used for the generation of models (n = 144 shoots)						
SMD (mm)	144	1.7	4.0	2.7	0.5	18.2
SML (mm)	144	31.0	115.0	68.1	14.7	21.6
LL (mm)	144	18.0	45.0	31.0	6.7	21.6
LW (mm)	144	14.0	41.0	27.7	5.5	20.0
SEL (mm)	144	14.0	55.0	36.2	7.8	21.6
SEW (mm)	144	7.0	31.0	19.0	4.5	23.5
FW (g)	144	0.3	2.5	1.2	0.5	40.5
Data used for the validation of models (n = 144 shoots)						
SMD (mm)	144	1.3	3.8	2.7	0.5	18.5
SML (mm)	144	30.0	105.0	64.8	14.5	22.3
LL (mm)	144	14.0	45.0	30.2	6.8	22.6
LW (mm)	144	10.0	42.0	27.6	6.1	22.1
SEL (mm)	144	20.0	53.0	35.9	7.0	19.6
SEW (mm)	144	10.0	31.0	18.9	4.0	21.3
FW (g)	144	0.3	2.1	1.1	0.4	39.6

SD = standard deviation; CV = coefficient of variation; SMD = stem diameter; SML = stem length; LL = leaflet length; LW = leaflet width; SEL = stipule length; SEW = stipule width; FW = fresh weight.

Table 2. Correlation coefficients among fresh weight and measured morphological characteristics of individual pea shoots (n = 144).

	SML (mm)	LL (mm)	LW (mm)	SEL (mm)	SEW (mm)	FW (g)
SMD (mm)	0.51**	0.73**	0.75**	0.64**	0.63**	0.83**
SML (mm)		0.48**	0.50**	0.48**	0.49**	0.73**
LL (mm)			0.88**	0.74**	0.69**	0.84**
LW (mm)				0.60**	0.58**	0.79**
SEL (mm)					0.93**	0.79**
SEW (mm)						0.76**

SMD = stem diameter; SML = stem length; LL = leaflet length; LW = leaflet width; SEL = stipule length; SEW = stipule width; FW = fresh weight.

**Significant at $P \leq 0.01$.

Table 3. Path coefficients of measured morphological characteristics of pea shoots indicating direct and indirect effects on individual shoot fresh weight (n = 144).

Plant traits	Direct effect	Indirect effect via					
		SMD	SML	LL	LW	SEL	SEW
SMD	0.29	—	0.17	0.23	0.00	0.12	0.02
SML	0.33	0.15	—	0.15	0.00	0.09	0.01
LL	0.31	0.21	0.16	—	0.00	0.14	0.02
LW	0.00	0.22	0.17	0.27	—	0.11	0.02
SEL	0.19	0.19	0.16	0.23	0.00	—	0.03
SEW	0.03	0.19	0.16	0.21	0.00	0.17	—

SMD = stem diameter; SML = stem length; LL = leaflet length; LW = leaflet width; SEL = stipule length; SEW = stipule width; FW = fresh weight.

Table 4. Regression models selected for estimation of fresh weight of pea shoots (n = 144).

Regression models		R^2	RMSE	<i>F</i> -value	Pr > F
1.	FW = -1.433 + 0.273 SMD + 0.010 SML + 0.022 LL + 0.000 LW + 0.011 SEL + 0.003 SEW	0.92	0.14	250.96	<0.0001
2.	FW = -1.437 + 0.276 SMD + 0.010 SML + 0.022 LL + 0.013 SEL	0.92	0.14	381.30	<0.0001
3.	FW = -1.383 + 0.308 SMD + 0.011 SML + 0.030 LL	0.90	0.15	408.55	<0.0001
4.	FW = -1.138 + 0.013 SML + 0.044 LL	0.85	0.18	398.87	<0.0001
5.	FW = -0.657 + 0.059 LL	0.71	0.25	347.34	<0.0001
6.	FW = 0.769 SMD - 0.949	0.68	0.26	306.97	<0.0001
7.	FW = 0.023 SML - 0.417	0.53	0.32	162.14	<0.0001
8.	FW = 0.177 SMD ^{1.839}	0.69	0.31	311.34	<0.0001
9.	FW = 0.003 SML ^{1.374}	0.53	0.35	163.06	<0.0001

R^2 = coefficient of determination; RMSE = the root mean square error; Pr = probability; SMD = stem diameter; SML = stem length; LL = leaflet length; LW = leaflet width; SEL = stipule length; SEW = stipule width; FW = fresh weight.

from the two linear models were plotted against the measured FW values, and 1:1 line (i.e., $Y=X$), as well as the t test comparison of the mean predicted vs. measured values, was used to examine a possible overestimation or underestimation by the models.

Results

Model construction. The pea shoots sampled from different environment conditions or different harvest dates showed differences in most of the growth traits, despite having similar cv for each trait (Fig. 2). Under high vs. low temperature, pea shoots had smaller SMD, SML, and LW, and larger SEW, but no differences in the other traits. For the pea shoots under no SL vs. SL, all the tested growth traits were smaller. All the traits except SML were smaller for the pea shoots harvested on 30 Jan. and 11 Feb. compared with 19 Jan. Most traits except LL and LW were larger for the pea shoots harvested on 11 Feb. compared with 30 Jan.

The shoots used in this study displayed considerable morphological variation, and they were representative of harvested pea shoots (Table 1). Data dispersion can be considered as strong (i.e., $cv > 30\%$) for FW, or moderate (i.e., $15\% < cv \leq 30\%$) for the other variables. All variables of the pea shoots sampled for model validation presented values close to those sampled for model construction. Consequently, the values used for the validation should not extrapolate those used for the modeling.

Correlation coefficients for FW and measurable morphological characteristics of pea shoots sampled for modeling are given in Table 2. FW was positively correlated to all the measured morphological traits, showing the largest correlation coefficient with LL. The highest correlation among the measured morphological characteristics of pea shoots was obtained between SEL and SEW.

Path coefficient analysis partitioned correlation coefficients among FW and measurable morphological characteristics of pea shoots sampled for modeling into direct and indirect effects (Table 3). Based on the direct positive effects on FW, the order of first four factors was $SML > LL > SMD > SEL$. However, each of the direct effects from the above four parameters was less than half of the correlation coefficients between FW and the respective measurable morphological traits. The direct effects of SMD and SEL on FW were of similar magnitude to their indirect effects on FW via LL and SML. The direct effects of SEW and LW on FW were negligible, but it was evident that both SEW and LW had great indirect effects on FW via the other four parameters (i.e., LL, SML, SMD, and SEL). The relatively small residual effect (0.29) suggested that nearly all the measurable morphological characteristics contributed to FW were included in this study except for sampling error.

Through step-by-step regression, five linear models (Eq. [1]–[5]), which included the number of variables (i.e., measurable mor-

phological traits) gradually reduced from 6 to 4, 3, 2, and 1 in sequence, were constructed for the estimation of FW (Table 4). Also, the single measurable morphological parameter, i.e., SMD or SML, was used to construct

linear and nonlinear regression models as comparison. All the five linear regression models from step-by-step regression (Eq. [1]–[5]) have higher R^2 values than the two nonlinear regression models (Eq. [8]–[9]).

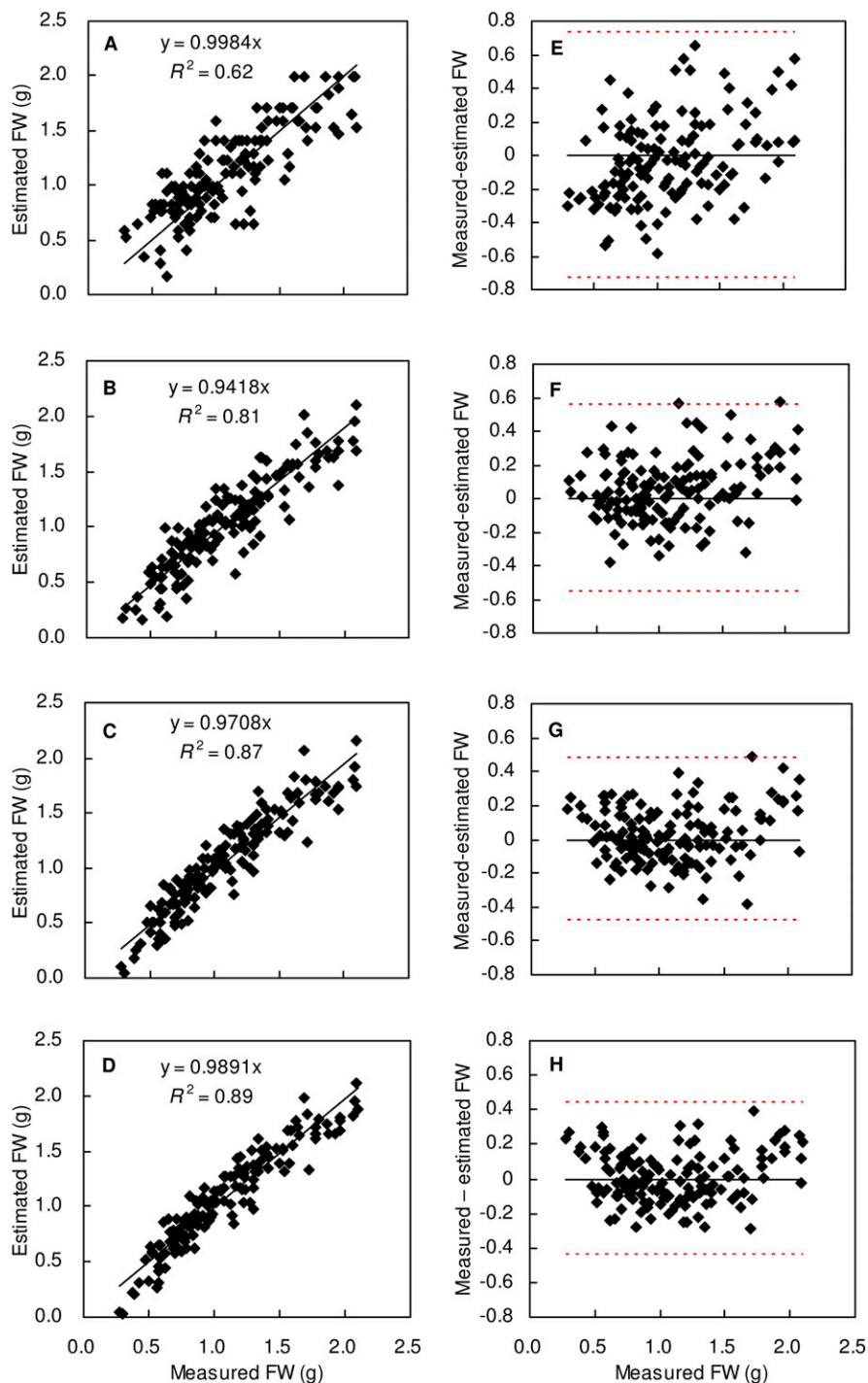


Fig. 3. A comparison of measured and estimated fresh weight (FW), and the difference between measured and estimated fresh weight (FW) of pea shoots using linear regression models with different variable numbers; $n = 144$. A–D corresponds to the linear regression models (i.e., Eq. [5], [4], [3], and [2], respectively), and the same is E–H. Eq. [5], [4], [3], and [2] have one variable (i.e., $FW = -0.657 + 0.059 LL$), two variables (i.e., $FW = -1.138 + 0.013 SML + 0.044 LL$), three variables (i.e., $FW = -1.383 + 0.308 SMD + 0.011 SML + 0.030 LL$), and four variables (i.e., $FW = -1.437 + 0.276 SMD + 0.010 SML + 0.022 LL + 0.013 SEL$), respectively. In E–H, the middle solid line represents the mean of differences between the measured and estimated value (MD) and the standard deviation of the differences (SDD), and the top dashed line represents the $MD + 3SDD$, and the bottom dashed line the $MD - 3SDD$.

Even if the nonlinear regression models (Eq. [8]–[9]) were transformed into linear models (Eq. [6]–[7]) with the same single variable included, the value of R^2 decreased little, especially for the models with SML included. It appeared that a linear relationship existed between FW and the measurable morphological traits.

For the linear models, due to the same values of R^2 and RMSE, the regression Eq. [2] with four variables (i.e., SML, LL, SMD, and SEL) was as equally effective as the Eq. [1] with all the six variables at explaining the variation in FW (Table 4). In fact, the value of R^2 in Eq. [3] was still ≥ 0.90 after removing another variable, SEL, from Eq. [2], which only decreased R^2 by 0.02 and increased RMSE by 0.01. However, when one or two more variables were removed from Eq. [3], the value of R^2 decreased below 0.90.

Model validation. To validate the developed linear regression models (Eq. [2]–[5]) for estimating the FW of individual pea shoots, measured and estimated data were compared in Fig. 3A–D. There was a linear relationship ($Y = kX$) between the measured FW and estimated value from all the four models. When the number of included variables varied from one to four, the values of R^2 increased from 0.62 to 0.89—and reached above 0.85 in models (2) and (3), with four and three included variables, respectively. In addition to greater values of R^2 , models (2) and (3) had greater k values (i.e., closer to 1) than model (4). Although the k value was the closest to 1 for model (5) with one included variable, its R^2 value was the lowest among the above four models. This suggested that there were stronger agreements between the measured and estimated FW values for models (2) and (3) than models (4) and (5).

Comparing to the way based on k and R^2 values, a similar result was achieved for the agreement evaluation of the above four linear models using the calculated relative bias, which was estimated by the MD and SDD (Fig. 3E–H). For the tested four models, all MD values were close to zero, and above 97% of the differences between observed and predicted values fell within $MD \pm 3$ SDD, indicating that the differences were distributed normally around zero. However, when the number of included variables increased from one to four, the range of $MD \pm 3$ SDD became narrower. In other words, following the order of models (5), (4), (3), and (2), the SDD values decreased gradually, and the differences distribution became closer to zero. Above 99% of the differences between the measured and predicted values were distributed within $MD \pm 3$ SDD for models (2) and (3) rather than models (4) and (5). This suggested that models (2) and (3) showed better agreements between the measured and estimated values than models (4) and (5).

Based on the agreement evaluation of four models, models (2) and (3) were selected for further evaluation using pea shoots sampled from different environmental conditions or different harvest dates. For the three goodness of fit indicators (i.e., d , MAE, and

RMSE), in most cases, d was higher, and MAE and RMSE were lower for Eq. [2] than Eq. [3] (Table 5). This result suggested that the model from Eq. [2] had better prediction accuracy than the model from Eq. [3]. However, Eq. [2] and [3] showed a similar trend in

the variation of d , MAE, and RMSE with different environmental conditions or harvest dates. Generally, for both Eq. [2] and [3], a lower d , and higher MAE and RMSE were obtained for the pea shoots harvested from high vs. low temperature, no SL vs. SL, and

Table 5. Accuracy evaluation of the 4- and 3-variable models (Eq. [2] and [3], respectively) developed to predict the fresh weight of pea shoots from different environment conditions and different harvest dates.

	MAE		RMSE		d		Shoot samples
	E2 ^a	E3	E2	E3	E2	E3	
			<i>Temperature</i>				
High	0.16	0.20	0.18	0.22	0.79	0.74	24
Low	0.11	0.12	0.14	0.15	0.84	0.84	120
			<i>Light</i>				
No SL	0.14	0.15	0.17	0.18	0.78	0.76	48
SL	0.11	0.12	0.13	0.15	0.83	0.82	96
			<i>Harvest time</i>				
19 Jan.	0.11	0.12	0.13	0.15	0.85	0.82	48
30 Jan.	0.11	0.11	0.13	0.14	0.81	0.81	48
11 Feb.	0.14	0.16	0.17	0.20	0.76	0.74	48

^aE2 = Eq. [2] (i.e., $FW = -1.437 + 0.276$ SMD + 0.010 SML + 0.022 LL + 0.013 SEL); E3 = Eq. [3] (i.e., $FW = -1.383 + 0.308$ SMD + 0.011 SML + 0.030 LL).

MAE = the mean absolute error; RMSE = the root mean square error; d = the index of agreement; SL = supplemental lighting.

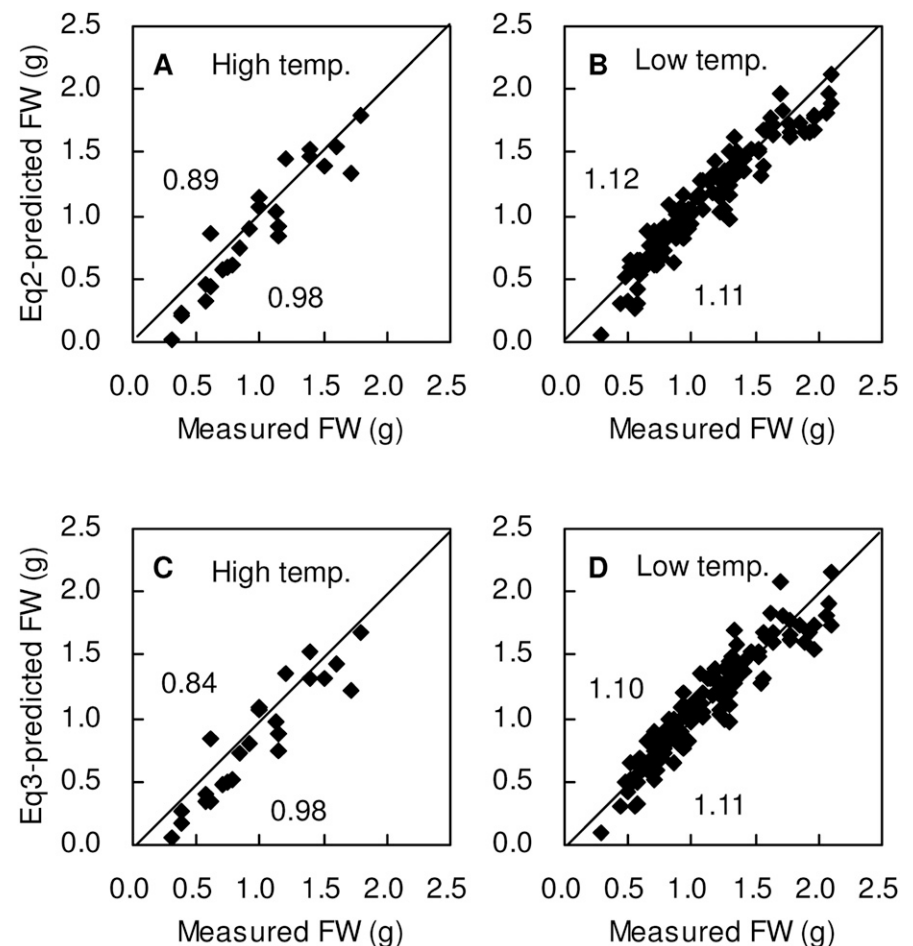


Fig. 4. Evaluation of the models, Eq. [2] and [3], developed to predict fresh weight (FW) of pea shoots from different temperature conditions; $n = 24$ and 120 for pea shoots harvested from high temperature and low temperature, respectively. The diagonal line through origin is 1:1 line (i.e., $Y = X$). Eq. [2] and [3] have four variables (i.e., $FW = -1.437 + 0.276$ SMD + 0.010 SML + 0.022 LL + 0.013 SEL) and three variables (i.e., $FW = -1.383 + 0.308$ SMD + 0.011 SML + 0.030 LL), respectively. The data nearby Y and X axis are means of predicted and measured FW, respectively.

11 Feb. vs. 19 and 30 Jan., suggesting a lower prediction accuracy of these two models under these conditions.

For the models from both Eq. [2] and [3], there was a potential underestimation of FW for the pea shoots sampled under high temperature or no SL condition, or those harvested on 11 Feb. (Figs. 4–6). For the pea shoots sampled under high temperature or no SL, or those harvested on 11 Feb., most of the data were under 1:1 line (Figs. 4–6). Nevertheless, there were no statistical differences between the means of predicted vs. measured values for the pea shoots sampled from each environment condition or harvest date (data not shown). This finding indicated that the underestimation of the two models was still in the acceptable range.

Discussion

Nonlinear or linear regression models. It has long been noted that power function, $y = ax^b$, can be used to estimate total biomass from predictive biometric variables such as breast-height diameter, or height in trees, or aquatic plants (Pilli et al., 2006; Spencer

et al., 2006). However, in the present study on pea shoots, the R^2 values were similar for linear or nonlinear relationships between FW and SMD or SML, indicating that their relationships should be linear. Similar results have been reported in the estimation of shoot mass of eight browse plant species, where the transformation of twig length by squaring did not significantly improve estimates (Schuster, 1965). It is well known that any nonlinear relationship may behave as an approximate linear relationship over a narrower range. Possibly, the present linear relationships were due to the narrower ranges of SMD or SML of pea shoots relative to trees and aquatic plants (e.g., *Arundo donax*).

In the present study, multivariable linear models were better than one-variable linear or nonlinear models for estimating pea shoot FW from the measured morphological traits. Similarly, a combination of predictor factors is usually superior to a single factor for estimating shoot biomass of browse plants and tea plants (Rutherford, 1979; Schuster, 1965; Sun et al., 2007). Deferring from pea shoots, tree biomass normally can be estimated by a single variable, i.e., stem size

(Pilli et al., 2006). The difference between trees and pea shoots could be explained by the role of the stem in its contribution to biomass. For trees, more than 80% of total aboveground biomass is stem biomass (Pilli et al., 2006). However, for pea shoots, both stems and leaves contribute greatly to shoot biomass (Wu et al., 2006), and thus the stems exert a smaller effect on the biomass estimation compared with trees.

Screening predictor variables. In the present study, six predictor variables, including both stem and leaf traits (i.e., SMD, SML, LL, LW, SEL, and SEW), were considered for developing multivariable linear models to estimate FW of pea shoots at the beginning. To avoid including some unimportant morphological traits, it is best to screen the crucial predictor variables for multivariable linear regression models, which could normally be realized by path analysis and stepwise regression (Güler et al., 2001; Payandeh and Wood, 1988). In the present study, path analysis showed that the two variables, SEW and LW, had small direct effects on FW. This result was also confirmed by the stepwise regression, where after removing SEW and LW, the remaining four variables (i.e., SML, SMD, SEL, and LL) can fully explain the variation of pea shoot FW. For tea shoots, the most important variables are shoot length, number of leaves, and bud length in a linear regression model for shoot weight estimation (Sun et al., 2007). The difference between pea and tea shoots might result from their different organographic composition of shoot tip parts (Niklas, 2005).

For model development, although multiple predictor variables gave significantly better predictions in most cases, a single variable may be of some value in practical sampling (Schuster, 1965). In the present study, it is also necessary to further screen one crucial predictor variable from the above four variables (i.e., SML, SMD, SEL, and LL). In the process of stepwise regression, after SEL, SMD, and SML were removed successively from a 4-, 3-, and 2-variable model, the multiple-variable linear model finally had only a single variable: LL. The removing order was generally negatively related to the direct effect on fresh biomass imposed by these variables, except during the last step where LL rather than SML was kept, although the latter had a higher direct effect than the former (Table 3). The reason may be that among all the tested morphological traits, LL as a single variable has the highest correlation with shoot FW (Table 2).

Accuracy and flexibility of the model. The best model has a close degree of fit to the data set (i.e., estimation accuracy), while it has the least number of predictor variables (Mokhtarpour et al., 2010). For estimation of plant biomass, including fewer predictor variables in the prediction models can reduce sampling time demand, but it may also potentially reduce estimation accuracy. Consequently, modeling plant biomass is at least a compromise of estimation accuracy and sampling time demand (Verwijst and Telenius,

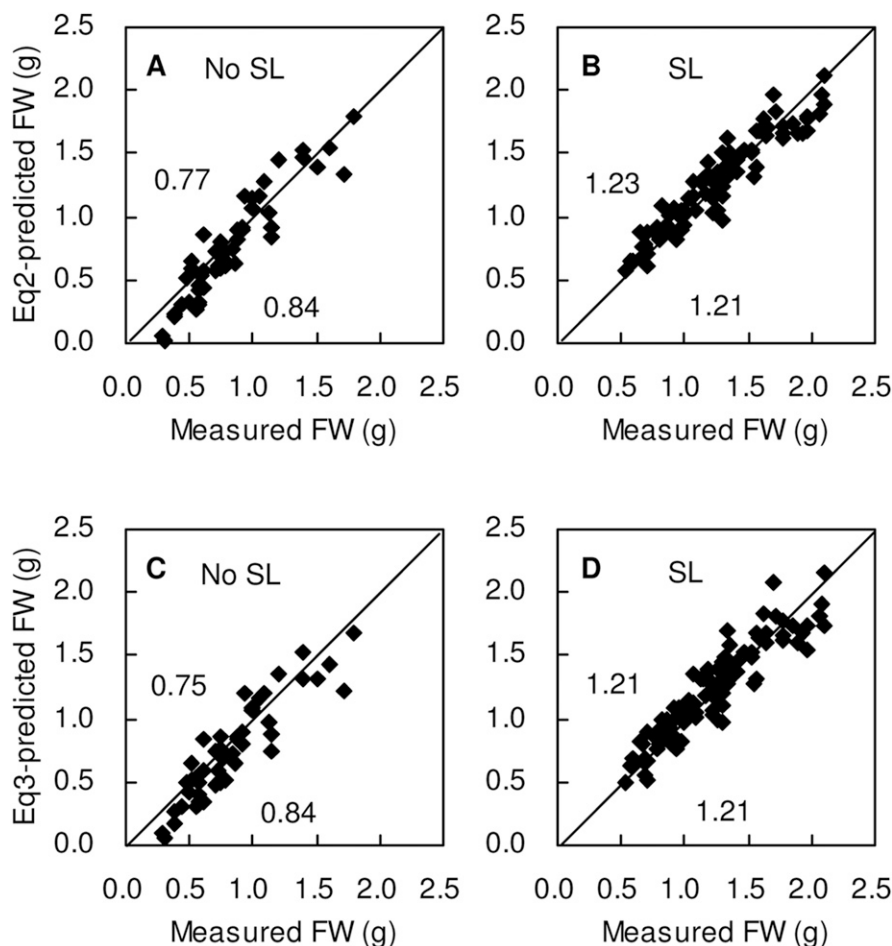


Fig. 5. Evaluation of the models, Eq. [2] and [3], developed to predict fresh weight (FW) of pea shoots from different light conditions. SL = supplemental lighting; n = 48 and 96 for pea shoots harvested from no SL and SL, respectively. The diagonal line through origin is 1:1 line (i.e., $Y = X$). Eq. [2] and [3] have four variables (i.e., $FW = -1.437 + 0.276 \text{ SMD} + 0.010 \text{ SML} + 0.022 \text{ LL} + 0.013 \text{ SEL}$) and three variables (i.e., $FW = -1.383 + 0.308 \text{ SMD} + 0.011 \text{ SML} + 0.030 \text{ LL}$), respectively. The data nearby Y and X axis are means of predicted and measured FW, respectively.

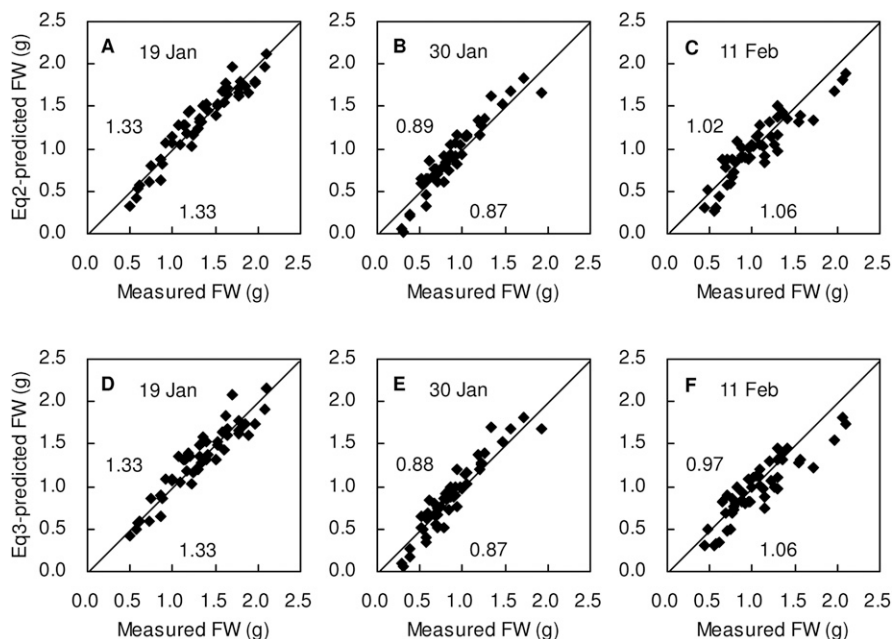


Fig. 6. Evaluation of the models, Eq. [2] and [3], developed to predict fresh weight (FW) of pea shoots from different harvest dates; $n = 48$ for pea shoots sampled from each harvest date. The diagonal line through origin is 1:1 line (i.e., $Y = X$). Eq. [2] and [3] have four variables (i.e., $FW = -1.437 + 0.276 \text{ SMD} + 0.010 \text{ SML} + 0.022 \text{ LL} + 0.013 \text{ SEL}$) and three variables (i.e., $FW = -1.383 + 0.308 \text{ SMD} + 0.011 \text{ SML} + 0.030 \text{ LL}$), respectively. The data nearby Y and X axis are means of predicted and measured FW, respectively.

1999). This is the reason why the regression models with 4 or 3 variables, rather than those with 6, 2, or 1 variables, were selected for the estimation of pea shoot FW in the present study. The main difference between 4- and 3-variable regression models was whether stem diameter was included as a predictor variable. Although a 4-variable model is more accurate for estimation of pea shoot FW than a 3-variable model, stem diameter measurement increases the time demand during the developing of the models. For pea shoots, the cross section of the stem is not round, so the diameter measurement needs more than one measurement to increase accuracy. Similarly, in irregularly shaped tree trunks, or browse plant twigs, several measurements of diameter at different directions are required to achieve an average value (Barnes et al., 1976; Rutherford, 1979; Telfer, 1969). However, this difficulty in diameter measurement could be reduced with the aid of machine vision in robots using image processing technology. Therefore, the selection between 4- or 3-variable regression models to estimate pea shoot FW would depend on available technologies and required accuracy.

A good model should also have some degrees of flexibility; if a model is sensitive to fluctuations in the data set, it has little predictive value (Mokhtarpour et al., 2010). In the present study, for the 4- or 3-variable models, the average values of predicted vs. measured pea shoot FW was not statistically different, regardless of different environment conditions or harvesting dates. This result suggests that there are some degrees of flex-

ibility in these models. However, a general trend in underestimation of pea shoot FW from the regression models was found for the last harvest (i.e., 11 Feb.), high temperature, or no SL (Figs. 4–6). For the last harvest, flower buds or flowers began to appear in some harvested pea shoots, but this organ was not included as one of the predictor variables in this study. Under high vs. low temperature, fewer (24 vs. 120) samples were collected to develop the prediction models, so the coefficients in the prediction models might match the morphological variables of the pea shoots under low temperature. Under no SL vs. SL, the measured FW of pea shoots was 30% smaller, and this was distributed in a range of smaller values (Fig. 2), which contributed to the underestimation to some degree (Fig. 3). However, all the above problems could be addressed by adjusting the model coefficients under different environmental conditions or at different harvest dates. A previous study indicated that in the linear function for tea shoot tip, the coefficients of models are affected by soil water level (Sun et al., 2007). For pea shoots, further research may be necessary for studying the response of model coefficients to the above factors.

In summary, to estimate FW of individual pea shoot, five linear models using all or part of the six variables (including SML, SMD, SEL, SEW, LL, and LW) were developed through stepwise regression. Also, two power and two linear regression models using SMD or SML as a single variable were constructed. The former five linear regression models (with 6, 4, 3, 2, and 1 variables) were more effective at explaining the variation in FW

than the latter two nonlinear and two linear 1-variable regression models, suggesting a linear relationship between FW and the morphological traits. From the former five linear regression models, the 3- and 4-variable linear models were selected for their more accurate prediction than the 1- and 2-variable linear models, and relatively simpler forms than 6-variable linear model. Despite a possibility of underestimating FW for the pea shoots sampled under high temperature or no SL condition, or those harvested at late time, the prediction accuracy of the two models was statistically acceptable. This implies the possibility and effectiveness of using the measurable morphological traits of pea shoot for estimation of its fresh biomass through the multilinear modeling.

Literature Cited

- Barnes, D.L., B.V. Lloyd, and L. McNeill. 1976. The use of shoot dimensions to estimate the leaf mass or leaf area of certain indigenous trees in Rhodesia. *Proc. Ann. Congr. Grassl. Soc. South. Afr.* 11:47–50.
- Fehrmann, L. and C. Kleinn. 2006. General considerations about the use of allometric equations for biomass estimation on the example of Norway spruce in central Europe. *For. Ecol. Mgt.* 236:412–421.
- Golzarian, M.R., R.A. Frick, K. Rajendran, B. Berger, S. Roy, M. Tester, and D.S. Lun. 2011. Accurate inference of shoot biomass from high-throughput images of cereal plants. *Plant Methods* 7, doi: 10.1186/1746-4811-1187-1182.
- Güler, M., M.S. Adak, and H. Ulukan. 2001. Determining relationships among yield and some yield components using path coefficient analysis in chickpea (*Cicer arietinum* L.). *Eur. J. Agron.* 14:161–166.
- Lim, T.K. 2012. *Pisum sativum*, p. 849–866. In: *Edible medicinal and non-medicinal plants*. Springer, Dordrecht.
- Liu, W., Q. Yang, Z. Qiu, and J. Zhao. 2014. Effects of light intensity and nutrient addition on growth, photosynthetic pigments and nutritional quality of pea seedlings. *Acta Hort.* 1037:391–396.
- Miles, C.A., J.O. Dea, C. Daniels, and J. King. 2018. Pea shoots. Washington State University Cooperative Extension. <<http://pubs.cahnr.wsu.edu/publications/pubs/pnw567/>>.
- Mokhtarpour, H., C.B.S. Teh, G. Saleh, A.B. Selamat, M.E. Asadi, and B. Kamkar. 2010. Non-destructive estimation of maize leaf area, fresh weight, and dry weight using leaf length and leaf width. *Commun. Biom. Crop Sci.* 5:19–26.
- Niklas, K.J. 2005. Modelling below- and above-ground biomass for non-woody and woody plants. *Ann. Bot.* 95:315–321.
- Payandeh, B. and J.E. Wood. 1988. Identifying factors affecting plantation performance in boreal forests of Ontario. *New For.* 2:73–87.
- Pilli, R., T. Anfodillo, and M. Carrer. 2006. Towards a functional and simplified allometry for estimating forest biomass. *For. Ecol. Mgt.* 237:583–593.
- Ruiz-Altisent, M., J. Ortiz-Canavate, and C. Valero. 2004. Fruit and vegetables harvesting systems, p. 261–285. In: R. Dris and S.M. Jain (eds.), *Production practices and quality assessment of food crops*. Springer, Dordrecht.

- Rutherford, M.C. 1979. Plant-based techniques for determining available browse and browse utilization: A review. *Bot. Rev.* 45:203–228.
- Santos, J., M. Herrero, J.A. Mendiola, M.T. Oliveira, E. Ibáñez, C. Delerue-Matos, and M. Oliveira. 2014. Assessment of nutritional and metabolic profiles of pea shoots: The new ready-to-eat baby-leaf vegetable. *Food Res. Intl.* 58:105–111.
- Schuster, J.L. 1965. Estimating browse from twig and stem measurements. *J. Range Manage.* 18:220–222.
- Spencer, D.F., P.-S. Liow, W.K. Chan, G.G. Ksander, and K.D. Getsinger. 2006. Estimating *Arundo donax* shoot biomass. *Aquat. Bot.* 84:272–276.
- Sun, Y., X. Liu, and Y. Jiang. 2007. Research on the measure of the growth quantity of tea plant. *Chin. Agric. Sci. Bull.* 23:124–126.
- Telfer, E.S. 1969. Twig weight-diameter relationships for browse species. *J. Wildl. Manage.* 33:917–921.
- Thursby, G.B., M.M. Chintala, D. Stetson, C. Wigand, and D.M. Champlin. 2002. A rapid, non-destructive method for estimating above-ground biomass of salt marsh grasses. *Wetlands* 22:626–630.
- Verwijst, T. and B. Telenius. 1999. Biomass estimation procedures in short rotation forestry. *For. Ecol. Mgt.* 121:137–146.
- Wu, X., Y. Lu, M. Ma, and T. Wu. 2006. Analysis of yield and quality of pea shoots in different leaf-type species. *J. Shanghai Jiaotong Univ. (Agr. Sci.)*. 24:260–263.