ORIGINAL ARTICLE

Regression Analysis to Estimate Morphology Parameters of Pepper Plant in a Controlled Greenhouse System

Jayanta Kumar Basak[1] · Waqas Qasim[1] · Frank Gyan Okyere[1] · Fawad Khan[1] · Yong Jin Lee[1] · Jihoon Park[1] · Hyeon Tae Kim[1]

Abstract

Purpose Regression analysis to predict growth indices of plant is essential for understanding the relationship between the total leaf area, production of fresh weight and dry matter, and expansion of the plant growth.

Methods An experiment was conducted to develop regression models for estimating leaf area, fresh weight, and dry weight from measurements of plant height at the vegetative phase of hot pepper (*Capsicum annuum* Linnaeus) grown in biodegradable pots in a greenhouse. Five models were evaluated and compared: linear regression model, two-order polynomial regression model (P. order 2), three-order polynomial regression model (P. order 3), four-order polynomial regression model (P. order 4), and power regression model. The models were compared using the coefficient of determination (R^2), Pearson's correlation coefficient (r), root mean square error (RMSE), relative standard error (RSE), and mean absolute percentage error (MAPE).

Results Power regression involving plant height demonstrated the highest R-square among the other models with minimum error estimate for the expected leaf area ($R^2 > 0.96$, $r > 0.98$, RMSE < 1.2, RSE < 0.04, and MAPE < 11.8); however, P. order 2 had a more accurate calculation of the fresh weight ($R^2 > 0.98$, $r > 0.99$, RMSE < 0.26, RSE < 0.04, and MAPE < 16.07) and dry weight ($R^2 > 0.97$, $r > 0.98$, RMSE < 0.03, RSE < 0.02, and MAPE < 11.7) of the plant considering both the fit and degree of adjustment, and the interpretation of the model.

Conclusions This study creates scope for further experimentation on various species of crops by changing management practices under different environmental conditions to enhance knowledge and understanding of the growing patterns of plants.

Keywords Greenhouse · Morphology parameters · Pepper plants · Regression models · Vegetative phase

---

✉ Hyeon Tae Kim
  bioani@gnu.ac.kr

  Jayanta Kumar Basak
  basak.jkb@gmail.com

  Waqas Qasim
  waqasdocuments2012@gmail.com

  Frank Gyan Okyere
  okyerefrank200@gmail.com

  Fawad Khan
  fawad.pid@gmail.com

  Yong Jin Lee
 yowqqq@naver.com

  Jihoon Park
  qkrwlgndl@naver.com

[1] Department of Bio-systems Engineering, Gyeongsang National University (Institute of Agriculture & Life Science), Jinju 52828, South Korea

Introduction

This study aims to measure the performance of different regression models generated from plant height, leaf area, and the fresh and dry weights of pepper plants grown in biodegradable pots in a controlled greenhouse environment. Pepper is considered to be one of the most popular vegetables in South Korea. It is consumed both fresh and as an ingredient in many traditional Korean cuisines (Hahm et al. 2017; Sang et al. 2008; Park 1999). It not only plays an important role in the year-round vegetable supply but is also considered to be an important commercial-orientated crop. The most famous fermented food product in Korea is kimchi which is made using various ingredients of which the most important is red pepper.

It is essential to understand the growth state and morphological characteristics of the pepper plant to better understand their influence on the suitability of the plant for cultivation and overall yield. It is observed that plant varieties differ considerably in

Fig. 1 Images presenting the different growing periods of pepper plants in an experimental greenhouse. **a** Germination stage. **b** 2nd-week plant. **c** 6th-week plant. **d** 8th-week plant

(a) Germination stage

(b) 2nd week plant

(c) 6th week plant

(d) 8th week plant
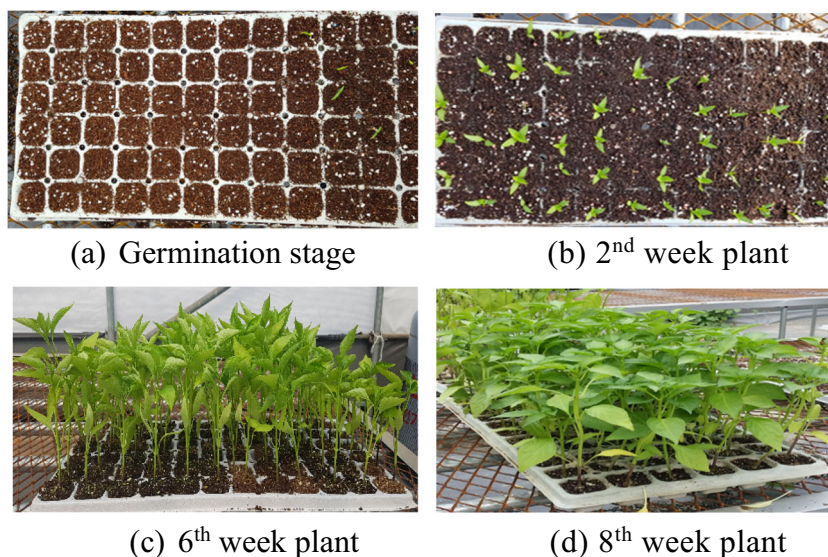
relative growth rate under identical or optimal conditions for both cultivated crops and wild species (Poorter and Bongers 2006; Van Kleunen et al. 2010; Gao et al. 2012). Moreover, Bonan (2008) reported microclimate factors also influence a wide range of important ecological processes, such as plant growth and soil nutrient cycling. Warm and high humid conditions are suitable for pepper plant growth; however, the plant requires dry weather at maturity stage. Pepper gives the best green yield and a better seed set at 21 to 27 °C during the day and 15 to 20 °C at night (Kahsay 2017). Lack of optimum temperature during the critical growth stage of the pepper plant causes substantial yield loss (Reddy and Kakani 2007).

The plant's characteristics, such as the time of each growth phase, plant height, fresh weight, dry weight, leaf area, the number of leaves, size of leaf, and number of branches, are used to describe the growth state (Salazar et al. 2018). Comparing the differences in morphological parameters, such as these among the same or various species, using mathematical models supports research on pepper cultivation techniques and improvement of yield and quality. In this study, regression models were developed to predict the leaf area, fresh weight, and dry matter of plants from measurements of plant height. The effectiveness of

these models at the vegetative phase is twofold. First, the models provide an enhanced knowledge and understanding of the growing pattern of pepper plants at the early stage. Second, they are useful for monitoring pepper growth using mathematical models to describe the relationships between the leaf area, the production of fresh and dry matter, and the expansion of total plant height.

Regression modeling is a generic term for a group of different statistical techniques using regression analysis. The fundamental purpose of these techniques is to observe the relationship between observed and predicted variables (Boldina and Beninger 2016). The most common type of regression analysis is linear regression, which is widely used in many different contexts in diverse fields. For example, in aquatic ecology, a number of papers used linear regression to reveal the species-area relationship (Willis et al. 1997; Peake and Quinn 1993). Schmid (2000) mentioned that there was a linear relationship between the population density and body size of benthic invertebrate species. Some studies were conducted based on regression analysis, such as the characterization of spatial patterning (Beninger and Boldina 2014; Seuront 2010) and the use of multiple fields, wherein allometric relations are prominent

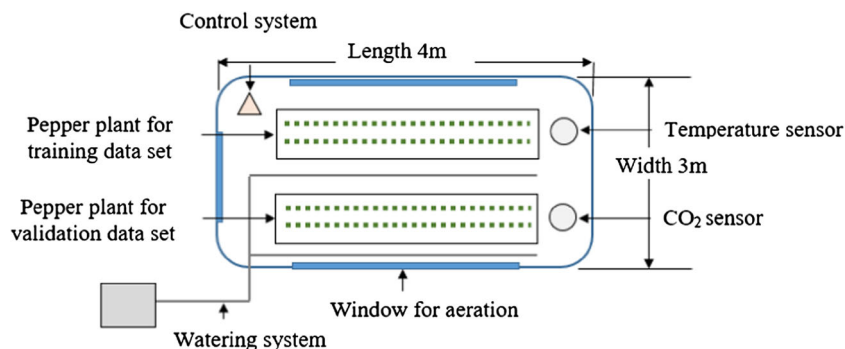**Fig. 2** Design and schematic diagram of the greenhouse

Control system

Length 4m

Pepper plant for training data set

Temperature sensor

Width 3m

Pepper plant for validation data set

CO$_2$ sensor

Window for aeration

Watering system

Table 1 Experimental values of PH, LA, FW, and DW (mean ± SD) for model preparation ($n = 20$)

Plant indicators	Weeks							
	1	2	3	4	5	6	7	8
PH (mm)	71.8 ± 3.77	93.65 ± 2.78	113 ± 4.15	157 ± 3.48	190 ± 7.05	221.3 ± 5.88	251 ± 3.38	272.4 ± 3.88
LA (cm^2)	2.55 ± 0.75	2.80 ± 0.37	4.55 ± 0.66	7.36 ± 0.64	9.11 ± 0.66	15.09 ± 0.65	16.64 ± 0.91	17.75 ± 0.65
FW (g)	0.14 ± 0.01	0.39 ± 0.04	0.68 ± 0.05	1.39 ± 0.08	1.99 ± 0.06	2.99 ± 0.09	3.50 ± 0.05	3.80 ± 0.07
DW (g)	0.02 ± 0.003	0.056 ± 0.01	0.085 ± 0.01	0.18 ± 0.02	0.25 ± 0.02	0.40 ± 0.05	0.49 ± 0.05	0.57 ± 0.03

PH, plant height; *LA* leaf area; *FW*, fresh weight; *DW*, dry weight

(Carey et al. 2013; Cranford et al. 2011; Hirst 2012). Regression analysis has been used for dynamic energy budget (DEB) modeling to show the relationship between phytoplankton cell size and abiotic factors (Duarte et al. 2012; Finkel et al. 2010). Similarly, linear regression models and polynomial regression models are also used in different research fields (Legendre and Legendre 2012). Moreover, it should be noted that the polynomial model is preferable because it can easily fit to the data; however, it is often difficult to ascribe any real meaning or rationale for higher-order polynomials, which is clearly undesirable (Austin 2007). In the case of a model to fit data, which is nonlinear in form, it is not possible to perform a polynomial regression. Therefore, power regression between the dependent and independent variables is also preferred in the current study (Glazier 2013; Kerkhoff and Enquist 2009; Packard 2013).

Such regression equations for estimating leaf area, fresh weight, and dry weight from plant height can reduce sampling effort and cost and may increase precision where samples are difficult to handle. Moreover, the non-destructive methods based on regression models have been used in a number of studies related to crop science because the measurement processes are quicker and easier to be performed (Karimi et al. 2009; Astegiano et al. 2001; Guo and Sun 2001; Antunes et al. 2008; Birch et al. 1998; Cho et al. 2007; Fascella et al. 2013; Bozhinova 2006; Ghoreishi et al. 2012).

The primary objective of the current study was to examine the relationship between the leaf area, production of fresh and dry matter, and expansion of total plant height of pepper at the vegetative stage. The study also evaluated the use of different regression models to provide more precise estimates of those parameters. Furthermore, it is also possible to understand the mechanism of various morphological parameters in pepper growth in a controlled greenhouse system. This study creates scope for further experimentation on various species of crops by changing management practices under different environmental conditions to enhance knowledge and understanding of the growing patterns of plants. Some limitation in accuracy is inevitable, but using a large number of samples in the experiment may minimize deviations. For more precise modeling in the future, environmental factors, crop management practices, and other growth factors should be included in the models.

Materials and Methods

Experimental Site

The field experiment was conducted from 15 January to 15 April 2018 at Gyeongsang National University in a controlled greenhouse equipped with an automatic control system to record ambient parameters. The dimensions of the greenhouse were 3 m (width) × 4 m (length) × 2.5 m (height). The mean values of temperature and the concentration of CO_2 were 16.4 °C and 365.2 ppm, respectively, with a range of 5.7–30.3 °C and 302–415 ppm, respectively, inside the greenhouse during the experimental period. Figure 1 shows the different growing periods of pepper plants in an experimental greenhouse.

Table 2 Fitted models to estimate predicted values of leaf area as a function of plant height

Equation type	Equation	R^2	r	RMSE	RSE	MAPE
Power	LA = 0.0023PH$^{1.6009}$	0.961	0.981	1.152	0.038	11.27
P. order 4	LA = − 3E−08PH^4 + 2E−05PH^3 − 0.0038PH^2 + 0.3538PH − 9.6996	0.972	0.939	169.19	828.3	835.29
P. order 3	LA = − 2E−06PH^3 + 0.0014PH^2 − 0.1609PH + 7.9937	0.968	0.982	5.116	0.757	35.34
P. order 2	LA = 0.0001PH^2 + 0.0333PH − 1.0935	0.962	0.981	2.112	0.129	16.19
Linear	LA = 0.0821PH − 4.5841	0.954	0.977	1.262	0.046	16.31

R^2, coefficient of determination; *r*, Pearson's correlation; *RMSE*, root mean square error; *RSE*, relative standard error; *MAPE*, mean absolute percentage error

Table 3 Fitted models to estimate predicted values of fresh weight as a function of plant height

Equation type	Equation	R^2	r	RMSE	RSE	MAPE
Power	$FW = 6E{-}06PH^{2.4057}$	0.974	0.984	0.279	0.042	15.23
P. order 4	$FW = -5E{-}09PH^4 + 3E{-}06PH^3 - 0.0006PH^2 + 0.0582PH - 2.049$	0.993	0.995	0.804	0.355	43.72
P. order 3	$FW = -4E{-}07PH^3 + 0.0002PH^2 - 0.0208PH + 0.6655$	0.991	0.998	1.162	0.743	52.04
P. order 2	$FW = 2E{-}05PH^2 + 0.0106PH - 0.8039$	0.988	0.994	0.249	0.034	15.04
Linear	$FW = 0.0191PH - 1.4142$	0.983	0.992	0.172	0.016	24.08

R^2, coefficient of determination; r, Pearson's correlation; *RMSE*, root mean square error; *RSE*, relative standard error; *MAPE*, mean absolute percentage error

Experimental Design

Pepper seeds were grown in two beds in an environmentally controlled greenhouse. Plants from one bed were chosen to comprise a training set which, in turn, was used to estimate the models (Fig. 2). The remaining plant bed was used as a validation set to measure the predictive ability of the fitted models. The windows of the greenhouse were opened for fresh air daily from 10 am to 6 pm based on the inside temperature of greenhouse. There were a total of 360 pepper plants in each of the beds of the greenhouse. Moreover, in this stud,y the temperature, humidity, and CO_2 concentrations were also measured daily using sensors (Lutron MCH-383SD, Electro Chemical Engineering, Melbourne, Australia) at three different locations and three distinct heights in each greenhouse.

Crop Management Practices

Seedlings of *Capsicum annuum* Linnaeus, a variety of pepper widely used in South Korea, were carefully sowed in same sized pots on 2 February 2018. In most or virtually all cases, all crop management practices were provided to each bed at the same time and at a fixed rate. Every seed was sowed into a single biodegradable pot then moved to the greenhouse. Fertilizer and irrigation applications were based on crop growth stages and soil nutrients. Different doses of fertilizer were applied at each growing stage. Fertilizer (N-P-K) application rate was 1:1:1 after 25 days of germination. The rate was 1:2:2 after 45 days one time

after germination during the experiment. The plant was watered thoroughly, with close attention paid to the base of the plant and the roots while applying the same quantity for each plant. Water was provided to the plants for 5 min per day (1.2 ℓ per day for 144 plants) by the sprinkler irrigation system during this experimental period.

Data Collection and Analysis

The plants were periodically examined after the germination stage to observe the changing pattern of morphological parameters. Twenty plants from each bed were collected randomly after 15 days of germination and their plant height, leaf area, fresh weight, and dry matter were measured over 8 weeks. The plant height of each sample was measured using a metric ruler, while the fresh and dry weights of the same plant were estimated using a digital balance (FX-300iWP, A&D Company Ltd, Tokyo, Japan) and drying oven (Shelves for 5E-DHG6310: 2 Layers, Changsha Kaiyuan Instruments Co., Ltd, Changsha, P. R., China). A number of papers were followed to select various temperature ranges and times for measuring the dry weight of the plants (Arshadullah and Zaid 2007; Cho et al. 2007; Karimi et al. 2009). In the experiment, sampled plants were studied for 8 weeks, and the dry weights of the plants were measured keeping the temperature at 80°C for 24 h. The leaf area was measured with the aid of a leaf area meter (CI-202, CID Bioscience, Camas, WA, USA) (Dalorima et al. 2018). The total number of pepper plants was classified into training and validation sets. Data were recorded from the 160 plants that were

Table 4 Fitted models to estimate predicted values of dry weight as a function of plant height

Equation type	Equation	R^2	r	RMSE	RSE	MAPE
Power	$DW = 9E{-}07PH^{2.3883}$	0.978	0.984	0.037	0.034	13.16
P. order 4	$DW = -6E{-}10PH^4 + 4E{-}07PH^3 - 8E{-}05PH^2 + 0.0085PH - 0.2894$	0.974	0.986	0.167	0.714	58.47
P. order 3	$DW = -2E{-}08PH^3 + 2E{-}05PH^2 - 0.0019PH + 0.0659$	0.972	0.986	0.043	0.046	14.53
P. order 2	$DW = 8E{-}06PH^2 + 0.0001PH - 0.0281$	0.972	0.986	0.034	0.029	11.38
Linear	$DW = 0.0028PH - 0.2156$	0.951	0.975	0.044	0.049	35.73

R^2, coefficient of determination; r, Pearson's correlation; *RMSE*, root mean square error; *RSE*, relative standard error; *MAPE*, mean absolute percentage error

Table 5 Suggested models to estimate predicted values of leaf area, fresh weight, and dry weight as a function of plant height

Relationship	Regression model	R^2	r	RMSE	RSE	MAPE
PH and LA	$LA = 0.0023PH^{1.6009}$	0.961	0.981	1.152	0.038	11.27
PH and FW	$FW = 2E{-}05PH^2 + 0.0106PH - 0.8039$	0.988	0.994	0.249	0.034	15.04
PH and DW	$DW = 8E{-}06PH^2 + 0.0001PH - 0.0281$	0.972	0.986	0.034	0.029	11.38

R^2, coefficient of determination; r, Pearson's correlation; *RMSE*, root mean square error; *RSE*, relative standard error; *MAPE*, mean absolute percentage error

used to comprise a training set which, in turn, was used to estimate the model. Data from another 160 plants were then used as a validation set to measure the predictive ability of a fitted model calculating, in this set, the coefficient of determination (R^2); Pearson's correlation coefficient (r); root mean square error (RMSE); relative standard error (RSE); and mean absolute percentage error (MAPE). Moreover, exploratory data analysis showed that the relationship between the dependent variables (leaf area, and fresh and dry weights of the plant) and predictor variables (plant height) did not always follow linear trends. Therefore, multiple linear regression models, including second-order, third-order, and fourth-order polynomial models and the power regression model were used to estimate the expected values. Prediction RMSE, R^2, r, RSE, and MAPE are summary statistics that allow for the comparison of those models and estimate their goodness of fit. To demonstrate the residual dispersion patterns on observed and expected values, standardized residual (error) figures were plotted. Standard statistical methods were used for data evaluation, including analysis of variance to practice completely randomized designs without the replication of samples with a significance level of $p < 0.05$. All statistical calculations were performed with Statistix10 (Analytical Software, Tallahassee, FL, USA), the Statistical Package for the Social Sciences (IBM SPSS Statistics 22.0.0.0, New York, NY, USA), and Origin Pro9.5.5 (OriginLab, Northampton, MA, USA).

Results and Discussion

Model Estimation

The evaluation of the plant height, leaf area, and fresh and dry weights as a function of time in the experimental period is presented in Table 1. Plant height, leaf area, and fresh and dry weights were significantly changed ($p < 0.05$) within the 8 weeks. The variations of those experimental values among the plants were very small; however, it was found that plant heights showed the maximum differences, followed by leaf area, then fresh weight and dry weight.

Summary statistics are provided for each of the variables evaluated for model development (Tables 3, 4, and 5). Experimental and predicted values were used in the regression models. Those models were validated by means of the leave-one-out cross-validation method (Perez et al. 2018), using five comparison criteria, namely, the highest cross-validated R^2 and r, and the least error of RMSE, RSE, and MAPE. Several studies only used two of the five comparison criteria, namely R^2 and RMSE, to perform a model of the adjusted versus reference line (Keramatloua et al. 2015; Fascella et al. 2013; Gao et al. 2012; Pompelli et al. 2012; Rouphael et al. 2010). Accurate models should reduce the RMSE by at least 2% (Clevers et al. 2008; Perez et al. 2018) and should show adjusted R^2 being above 0.7 (Yu et al. 2018).

Fig. 3 Standardized residuals (errors) plot for model development between plant height and leaf area
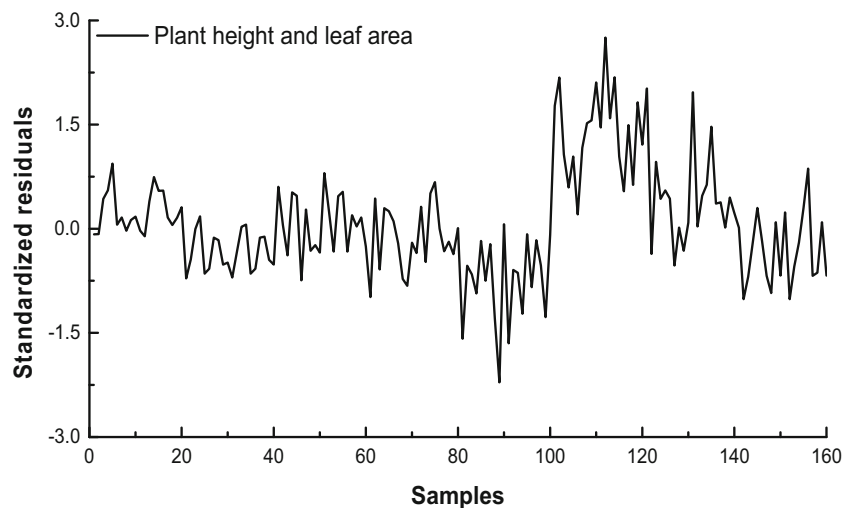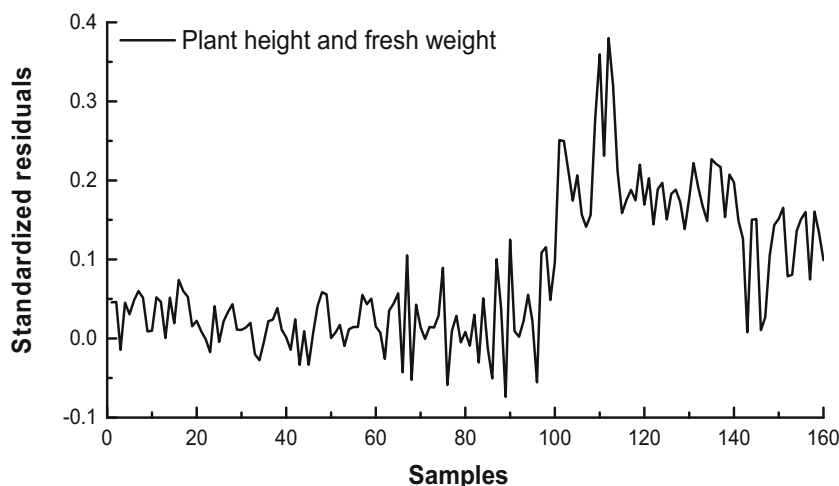
Fig. 4 Standardized residuals (errors) plot for model development between plant height and fresh weight

Five regression models showed the relationship between plant height and leaf area of the pepper plant (Table 2). High values of R^2 and r were observed ($R^2 > 0.95$ and $r > 0.93$), which adjusted the models for both parameters. However, a significant difference was observed among RMSE, RSE, and MAPE for all equations. Among the regression models, the lowest values of RMSE (1.15), RSE (0.038), and MAP (11.27) were found for the power formula, followed by linear, P. order 2, P. order 3, and P. order 4. From the results of observed and predicted values on the estimators, it was found that the power regression model was the best for estimating the leaf area using plant height. Some studies (Blanco and Folegatti 2005; Keramatloua et al. 2015; Fascella et al. 2013; Gao et al. 2012; Pompelli et al. 2012; Rouphael et al. 2010) frequently used the morphological parameters of the leaf, such as length and width, when developing regression estimators of leaf variables that are more difficult to measure. Salazar et al. (2018) found a positive correlation between leaf area and length and between leaf area and width, with correlation coefficients over 0.93. The study also showed that polynomial regressions involving both the length and width of

leaves provided very good models to estimate the expected area ($R^2 = 0.98$) and weight ($R^2 = 0.91$) of leaves.

Evaluation criteria for the five models described above were also estimated for both dependent variables of fresh and dry weights (Tables 3 and 4). In both cases, the P. order 2 model, which is a second-order polynomial on plant height, has a low MAPE compared with the other four models; however, RMSE, R^2, r, and RSE were almost similar. R^2 and r values obtained in estimating the fresh weight using the plant height data indicated that they were higher than 0.99 only for P. order 3 and P. order 4; however, the same two models based on RMSE, RSE, and MAPE values were not very accurate due to the increased estimation error. After analyzing the estimators of the fitted model, it was proposed that model P. order 2 was the best for fresh weight measurement. When dry weight was estimated from plant height considering R^2 and r, three models (P. order 3, P. order 4, and power models) had a higher R^2 and r; however, these models were not as precise as P. order 2 model because of the higher estimation error. Thus, the result of the study suggests that the P. order 2 model is the best fitted to estimate dry weight. The result was supported by

Fig. 5 Standardized residuals (errors) plot for model development between plant height and dry weight
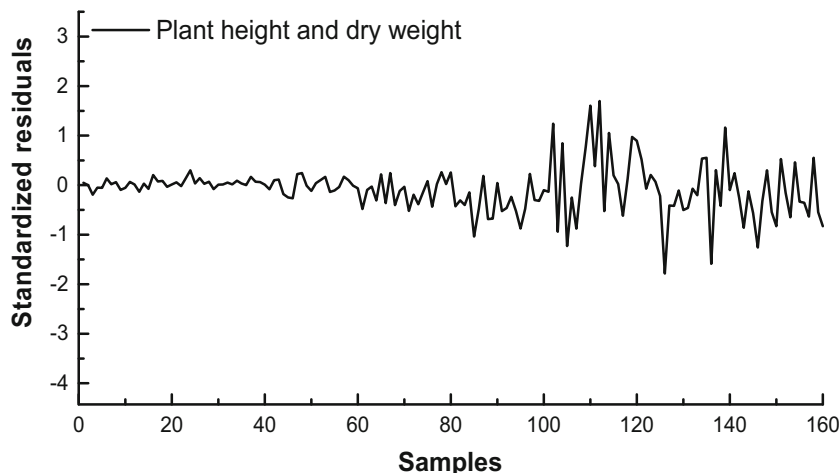
Table 6 Experimental values of PH, LA, FW, and DW (mean ± SD) for validation of model ($n = 20$)

Plant indicators	Weeks							
	1	2	3	4	5	6	7	8
PH (mm)	70.7 ± 3.15	91.1 ± 4.58	110 ± 4.80	152 ± 3.14	186 ± 3.15	217.7 ± 7.34	244.6 ± 7.22	264.1 ± 5.18
LA (cm^2)	2.49 ± 0.33	2.71 ± 0.36	4.49 ± 0.33	7.29 ± 0.80	8.93 ± 0.57	14.94 ± 0.73	16.42 ± 0.86	17.24 ± 0.66
FW (g)	0.13 ± 0.01	0.38 ± 0.03	0.64 ± 0.04	1.36 ± 0.06	1.95 ± 0.07	2.94 ± 0.07	3.47 ± 0.06	3.75 ± 0.08
DW (g)	0.02 ± 0.003	0.05 ± 0.01	0.08 ± 0.01	0.171 ± 0.01	0.25 ± 0.01	0.39 ± 0.04	0.47 ± 0.05	0.53 ± 0.05

the study of the pistachio plant (Karimi et al. 2009). Moreover, Williams and Martinson (2003) and Cho et al. (2007) found a good correlation between leaf weights by using only a single variable of either leaf length or leaf width.

The standardized residual plots (measurement of the strength of the difference between observed and expected values) are depicted in Figs. 3, 4, and 5 for the three best models (Table 5) for the estimation of the leaf area and fresh and dry weights of the plant. It is extremely useful to understand the residual dispersion pattern on a standardized scale because it easily allows detection of potential outliers (Qiana and Cuffney 2012). The standardized residual plot obtained from analysis of the plant data implied a significant relationship between observed and predicted values. In this situation, the reference and regression lines are too close to visually recognize differences among a set of competing models. When modeling the leaf area, and considering the RMSE, RSE, and MAPE (smaller is better), two models (power for leaf area estimation and P. order 2 for fresh and dry weight estimation) have lower RMSE, RSE, and MAPE. However, Figs. 3, 4, and 5, depicting that residual plots had several higher values for some samples, suggest that the error variance increased with plant height, whereas a distribution that revealed lower values indicated that error variance decreased with plant height.

Validation of the Model

The primary purpose of model validation of the study is to demonstrate that the five models are a reasonable representation of the actual system and that they focus on physiological behavior of the plant with enough fidelity to satisfy analysis objectives, and to confirm the accuracy of the model's representation of the experimental values. To validate the five regression equations, data (heights and their leaf areas, fresh and dry weights) of 160 plants that were not involved in the training set to develop the model (20 plants in each week) were analyzed as a validation set. For independent data, the measured plant height varied from 66 to 272 mm with a mean of 167 mm at 8 weeks. The leaf area of the validation set ranged from 1.98 to 17.74 cm^2 and the mean of the data set was 9.12 cm^2. Fresh weight and dry weight of this set varied from 0.125 to 3.88 g and from 0.013 to 0.61 g with a mean value of 1.83 g and 0.245 g, respectively (Table 6).

A regression model becomes over-fitted or more accurate when the prediction errors (RMSE, RSE, and MAPE) of the model are very low, instead of the relationship between independent and dependent variables. Therefore, the selection of the best model must be based not only on R^2 and r values but also on the use of accuracy measurement tools. Antunes et al. (2008) observed that models for *Coffea arabica* and *Coffea canephora*, even when showing high R^2 and high precision, produce underestimations of leaf area. A similar result was also found in the current study, showing that the P. order 4 model had a high R^2 value; however, the prediction error of this model was very high. Likewise, for the training data set, used in this study, it was also found that the power regression model was the best to predict leaf area because of its low RMSE, RSE and MAPE values on the validation data set, even when R^2 values were very high for polynomial regression models of different orders (Tables 2 and 7).

In the case of fresh weight and dry weight estimation, five regression models were generated from the training data set to

Table 7 Summary of the five models to predict leaf area. RMSE, r, RSE, and MAPE are calculated on the validation set

Equation type	Equation	r	RMSE	RSE	MAPE
Power	$LA = 0.0023PH^{1.6009}$	0.980	1.218	0.044	11.88
P. order 4	$LA = -3E-08PH^4 + 2E-05PH^3 - 0.0038PH^2 + 0.3538PH - 9.6996$	0.937	152.433	699.55	767.76
P. order 3	$LA = -2E-06PH^3 + 0.0014PH^2 - 0.1609PH + 7.9937$	0.980	4.645	0.649	31.90
P. order 2	$LA = 0.0001PH^2 + 0.0333PH - 1.0935$	0.980	2.252	0.152	16.95
Linear	$LA = 0.0821PH - 4.5841$	0.976	1.268	0.048	16.06

r, Pearson's correlation; *RMSE*, root mean square error; *RSE*, relative standard error; *MAPE*, mean absolute percentage error

Table 8 Summary of the five models to predict fresh weight. RMSE, r, RSE, and MAPE are calculated on the validation set

Equation type	Equation	r	RMSE	RSE	MAPE
Power	$FW = 6E{-}06PH^{2.4057}$	0.984	0.270	0.040	16.52
P. order 4	$FW = -5E{-}09PH^4 + 3E{-}06PH^3 - 0.0006PH^2 + 0.0582PH - 2.049$	0.996	0.804	0.362	46.19
P. order 3	$FW = -4E{-}07PH^3 + 0.0002PH^2 - 0.0208PH + 0.6655$	0.990	1.151	0.741	53.07
P. order 2	$FW = 2E{-}05PH^2 + 0.0106PH - 0.8039$	0.994	0.269	0.040	16.07
Linear	$FW = 0.0191PH - 1.4142$	0.991	0.187	0.019	27.53

r, Pearson's correlation; *RMSE*, root mean square error; *RSE*, relative standard error; *MAPE*, mean absolute percentage error

Table 9 Summary of the five models to predict dry weight. RMSE, r, RSE, and MAPE are calculated on the validation set

Equation type	Equation	r	RMSE	RSE	MAPE
Power	$DW = 9E{-}07PH^{2.3883}$	0.978	0.041	0.049	14.19
P. order 4	$DW = -6E{-}10PH^4 + 4E{-}07PH^3 - 8E{-}05PH^2 + 0.0085PH - 0.2894$	0.981	0.161	0.737	56.79
P. order 3	$DW = -2E{-}08PH^3 + 2E{-}05PH^2 - 0.0019PH + 0.0659$	0.981	0.045	0.056	14.25
P. order 2	$DW = 8E{-}06PH^2 + 0.0001PH - 0.0281$	0.981	0.037	0.039	11.78
Linear	$DW = 0.0028PH - 0.2156$	0.975	0.044	0.054	39.58

r, Pearson's correlation; *RMSE*, root mean square error; *RSE*, relative standard error; *MAPE*, mean absolute percentage error

determine the accuracy rate of those models that were best fitted for validation data set. The experimental data set for validation established that polynomial order 2 model performed better compared with the other four models to predict fresh and dry weights of pepper plants because of its low RMSE, RSE, and MAPE values (Tables 8 and 9). A similar result was also observed during model development for the same crop parameters. Table 8 shows that RMSE and RSE values were slightly higher for P. order 2 than for the linear regression, but the MAPE was less, while also including consideration of the model performance. In our study, whether the response variable was the fresh or dry weight of the plants, the values of all three evaluation criteria (RMSE, RSE, and MAPE) used to measure error were small for the P. order 2 regression model. The results obtained in this study demonstrate that fresh and dry weights of pepper plants can be estimated with minimum error using a polynomial of order 2 regression involving plant height. Standardized residuals were also analyzed by subtracting the observed responses from the predicted responses to determine the elements of variation unexplained by the three fitted models (Table 10 and Figs. 6, 7, and 8). The residuals in Figs. 6, 7, and 8 suggested that the models were drifting slowly to higher values as the investigation continued. It was also found that the size of the residuals changed as a function of a predictor's settings in regression model.

The results of the study show that plant height is a good variable to determine leaf area, fresh weight, and dry weight of pepper plants using the regression models. It was found that the area of pepper leaves is well-correlated to the plant height, manifested in a high coefficient of determination ($R^2 = 0.96$) and lower RMSE, RSE, and MAP between actual and predicted values using of power regression equation. However, the P. order 2 regression model had the highest coefficient of determination ($R^2 = 0.987$ for fresh weight and $R^2 = 0.962$ for dry weight) between predicted and actual values, and the lowest RMSE, RSE, and MAPE among the other equations (Table 10 and Figs. 9, 10, and 11).

Table 10 Validated models to estimate predicted values of leaf area, fresh weight, and dry weight as a function of plant height

Relationship	Regression model	r	RMSE	RSE	MAPE
PH and LA	$LA = 0.0023PH^{1.6009}$	0.980	1.218	0.044	11.88
PH and FW	$FW = 2E{-}05PH^2 + 0.0106PH - 0.8039$	0.994	0.269	0.040	16.07
PH and DW	$DW = 8E{-}06PH^2 + 0.0001PH - 0.0281$	0.981	0.037	0.039	11.78

r, Pearson's correlation; *RMSE*, root mean square error; *RSE*, relative standard error; *MAPE*, mean absolute percentage error

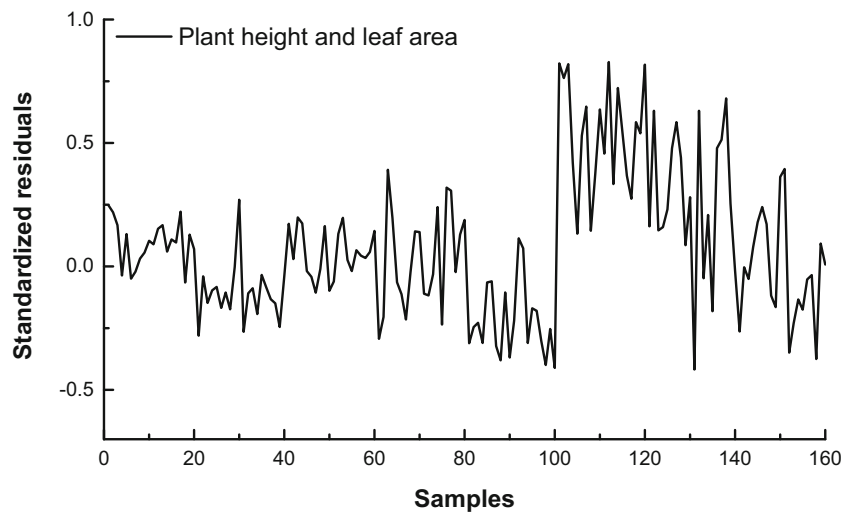**Fig. 6** Standardized residuals (errors) plot for model validation between plant height and leaf area

Fig. 7 Standardized residuals (errors) plot for model validation between plant height and fresh weight
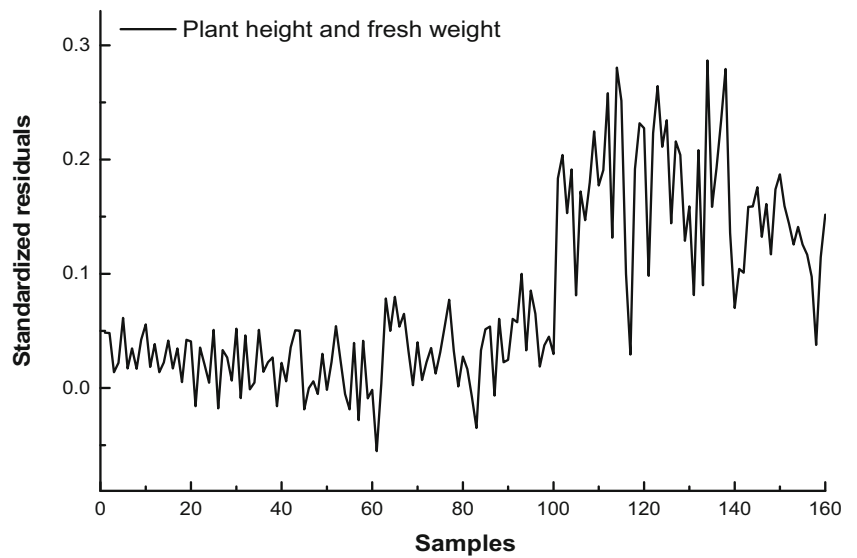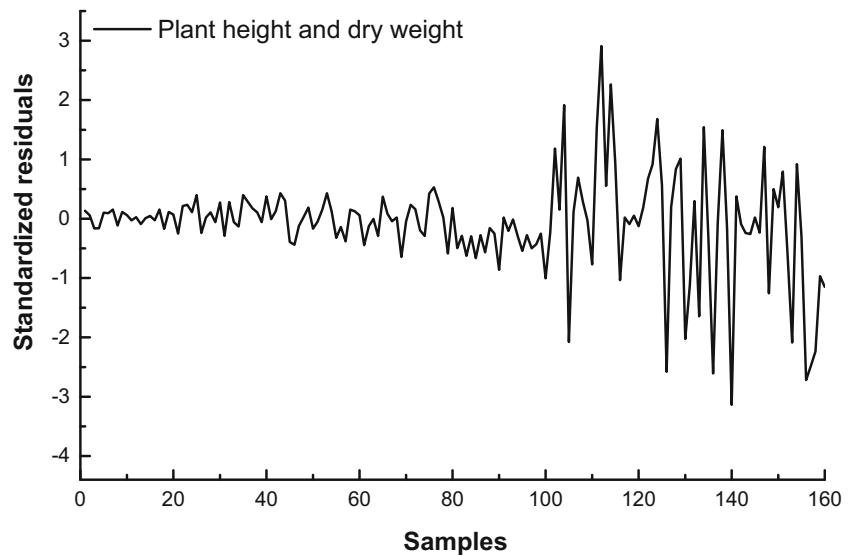

Fig. 8 Standardized residuals (errors) plot for model validation between plant height and dry weight
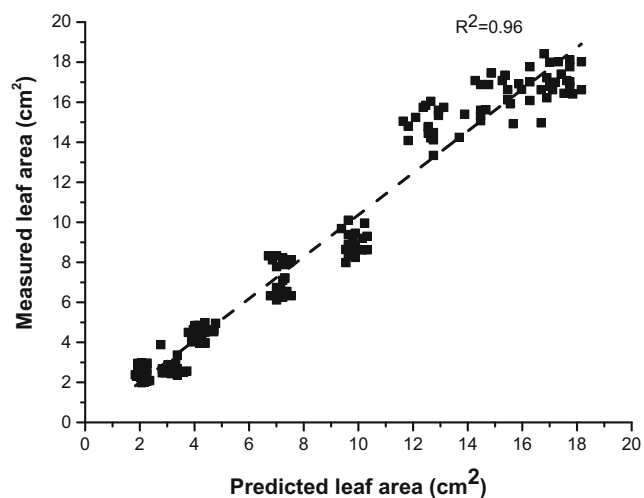
Fig. 9 Predicted values using the power regression model versus the actual values of leaf area (cm²)

Fig. 11 Predicted values using the P. order 2 regression model versus the actual values of dry weight (g/plant)

Conclusions

In this study, several regression models were presented and discussed along with newer and less commonly applied techniques to estimate leaf area, fresh weight, and dry weight using the height of pepper plants (*Capsicum annuum* Linnaeus) in a controlled greenhouse environment. While linear regression models are widely used in forecast-related studies in agriculture sectors, this technique may be inadequate to fully describe a complex and potentially nonlinear system found in plant morphology. In this study, five regression models were developed from a training data set; however, two models were suggested to predict leaf area and fresh and dry weights of the plant. Evaluation of the suggested models using a validation data set observed a strong positive relationship between predicted and measured data. The results of the study showed that pepper leaf area was estimated by a power regression model including the
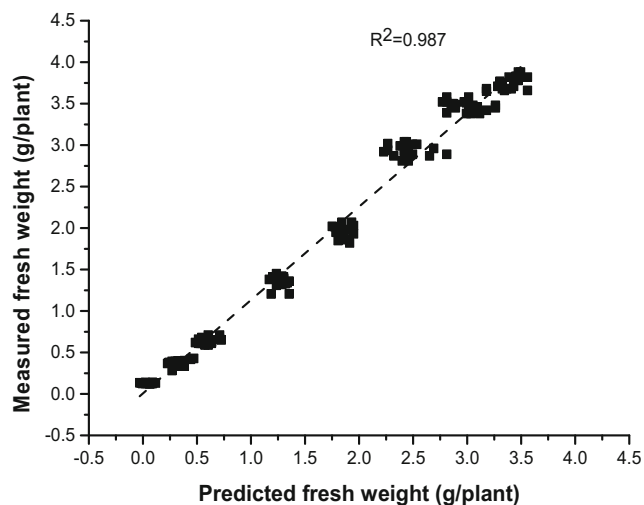

Fig. 10 Predicted values using the P. order 2 regression model versus the actual values of fresh weight (g/plant)
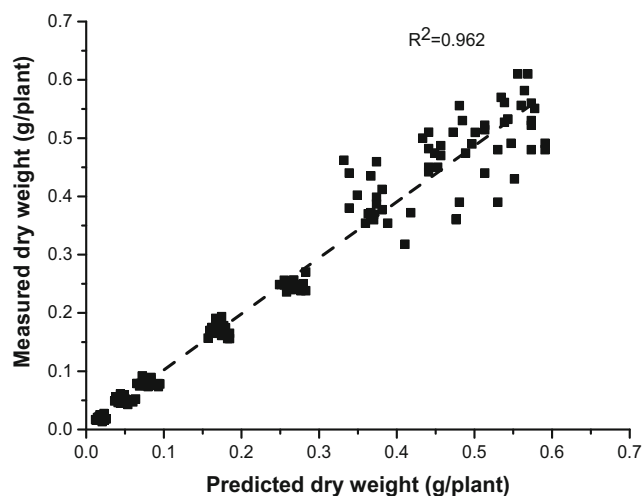
product of plant height. On the other hand, using the same predictor, fresh weight and dry weight can be calculated using the polynomial order 2 equation. Accuracy rates of these models were tested through different statistical tools (R^2 and r) and error calculation techniques (RMSE, RSE, and MAPE), demonstrating that the accuracy levels of the two suggested models were relatively high. Power regression involving plant height provided extremely good estimates for the expected leaf area ($R^2 > 0.96$, $r > 0.98$, RMSE < 1.2, RSE < 0.04, and MAPE < 11.8); however, P. order 2 provided greater accuracy in calculating the fresh weight ($R^2 > 0.98$, $r > 0.99$, RMSE < 0.26, RSE < 0.04, and MAPE < 16.07) and dry weight ($R^2 > 0.97$, $r > 0.98$, RMSE < 0.03, RSE < 0.02, and MAPE < 11.7) of plants considering the fit and degree of adjustment, and the interpretation of the model. The variable with the highest explanatory capability was used to develop a general equation to predict leaf area, fresh weight, and dry weight. These equations provide a very simple, more accurate, and time-saving tool to predict the growth of pepper plants. Some limitations in accuracy are inevitable, but using a larger number of samples in the experiment may minimize deviations. For more precise modeling, environmental factors, crop management practices, and other growth factors should be included in the models. The applicability of the suggested equations to other species of plants and changing management conditions in different environmental conditions should be tested. Furthermore, such observations are vital in understating the relationships between different treatments and the impacts on plant growth which, in turn, are key in determining the direction of optimum yield of crop.

Compliance with Ethical Standards

Conflict of Interest   The authors declare that there is no conflict of interest.

References

Antunes, W. C., Pompelli, M. F., Carretero, D. M., & DaMatta, F. M. (2008). Allometric models for non-destructive leaf area estimation in coffee (Coffeaarabica and Coffeacanephora). *Annals of Applied Biology, 153*(1), 33–40. https://doi.org/10.1111/j.1744-7348.2008.00235.x.

Arshadullah, M., & Zaid, S. A. R. (2007). Role of total plant dry weight in the assessment of variation for salinity tolerance in Gossypiumhirsutum (L.) Sarhad. *Journal of Agriculture, 23*(4), 857–866.

Astegiano, E. D., Favaro, J. C., & Bouzo, C. A. (2001). Estimación del area foliar en distintos cultivares de tomate (Lycopersiconesculentum Mill.) utilizandomedidasfoliareslineales. *Investigación Agraria Produccióny Protección Vegetales, 16*(2), 249–256.

Austin, M. (2007). Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling, 200*(1–2), 1–19. https://doi.org/10.1016/j.ecolmodel.2006.07.005.

Beninger, P. G., & Boldina, I. (2014). Fine-scale spatial distribution of the temperate in faunal bivalve Tapes (=Ruditapes) philippinarum (Adams and Reeve) on fished and unfished intertidal mudflats. *Journal of Experimental Marine Biology and Ecology, 457*, 128–134. https://doi.org/10.1016/j.jembe.2014.04.001.

Birch, C. J., Hammer, G. L., & Rickert, K. G. (1998). Improved methods for predicting individual leaf area and leaf senescence in maize (Zea mays). *Australian Journal of Agricultural Research, 49*(2), 249–262. https://doi.org/10.1071/a97010.

Blanco, F. F., & Folegatti, M. V. (2005). Estimation of leaf area for greenhouse cucumber by linear measurements under salinity and grafting. *Scientia Agricola, 62*(4), 305–309. https://doi.org/10.1590/s0103-90162005000400001.

Boldina, I., & Beninger, P. G. (2016). Strengthening statistical usage in marine ecology: linear regression. *Journal of Experimental Marine Biology and Ecology, 474*, 81–91. https://doi.org/10.1016/j.jembe.2015.09.010.

Bonan, G. (2008). *Ecological climatology* (2nd ed.). Cambridge: Cambridge University Press.

Bozhinova, R. P. (2006). Coefficients for determination of the leaf area in three Burley tobacco varieties. *Journal of Central European Agriculture, 7*(1), 7–12.

Carey, N., Sigwart, J. D., & Richards, J. G. (2013). Economies of scaling: more evidence that allometry of metabolism is linked to activity, metabolic rate and habitat. *Journal of Experimental Marine Biology and Ecology, 439*, 7–14. https://doi.org/10.1016/j.jembe.2012.10.013.

Cho, Y. Y., Oh, S., & Son, M. M. O. J. E. (2007). Estimation of individual leaf area, fresh weight, and dry weight of hydroponically grown cucumbers (Cucumissativus L.) using leaf length, width, and spad value. *ScientiaHorticulturae, 111*(4), 330–334. https://doi.org/10.1016/j.scienta.2006.12.028.

Clevers, J. G. P. W., Kooistra, L., & Schaepman, M. E. (2008). Using spectral information from the NIR water absorption features for the retrieval of canopy water content. *International Journal of Applied Earth Observation and Geoinformation, 10*(3), 388–397. https://doi.org/10.1016/j.jag.2008.03.003.

Cranford, P. J., Ward, J. E., & Shumway, S. E. (2011). Bivalve filter feeding: variability and limits of the aquaculture biofilter. In S. E.

Shumway (Ed.), *Shellfish aquaculture and the environment* (pp. 81–124). Hoboken: Wiley-Blackwell. https://doi.org/10.1002/9780470960967.ch4.

Dalorima, T., Khandaker, M. M., Zakaria, A. J., & Hasbullah, M. (2018). Impact of organic fertilizations in improving BRIS soil conditions and growth of watermelon (Citrullus Lanatus). *Bulgarian Journal of Agricultural Science, 24*(1), 112–118.

Duarte, P., Fernández-Reiriz, M. J., & Labarta, U. (2012). Modelling mussel growth in ecosystems with low suspended matter loads using a Dynamic Energy Budget approach. *Journal of Sea Research, 67*(1), 44–57. https://doi.org/10.1016/j.seares.2011.09.002.

Fascella, G., Darwich, S., & Rouphael, Y. (2013). Validation of a leaf area prediction model proposed for rose. *Chilean Journal of Agricultural Research, 73*(1), 73–76. https://doi.org/10.4067/s0718-58392013000100011.

Finkel, Z. V., Beardall, J., Flynn, K. J., Quigg, A., Rees, T. A. V., & Raven, J. A. (2010). Phytoplankton in a changing world: cell size and elemental stoichiometry. *Journal of Plankton Research, 32*(1), 119–137. https://doi.org/10.1093/plankt/fbp098.

Gao, M., Van der Heijden, G. W. A. M., Vos, J., Eveleens, B. A., & Marcelis, L. F. M. (2012). Estimation of leaf area for large scale phenotyping and modeling of rose genotypes. *Scientia Horticulturae, 138*, 227–234. https://doi.org/10.1016/j.scienta.2012.02.014.

Ghoreishi, M., Hossini, Y., & Maftoon, M. (2012). Simple models for predicting leaf area of mango (Mangiferaindica L.). *Advances in Biology and Earth Sciences, 2*(2), 845–853.

Glazier, D. S. (2013). Log-transformation is useful for examining proportional relationships in allometric scaling. *Journal of Theoretical Biology, 334*, 200–203. https://doi.org/10.1016/j.jtbi.2013.06.017.

Guo, D. P., & Sun, Y. Z. (2001). Estimation of leaf area of stem lettuce (Lactuca sativa varangustana) from linear measurements. *Indian Journal of Agricultural Science, 71*(7), 483–486.

Hahm, M. S., Son, J. S., Hwang, Y. J., Kwon, D. K., & Ghim, S. Y. (2017). Alleviation of salt stress in pepper (Capsicum annum L.) plants by plant growth-promoting rhizobacteria. *Journal of Microbiology and Biotechnology, 27*(10), 1790–1797. https://doi.org/10.4014/jmb.1609.09042.

Hirst, A. G. (2012). Intra specific scaling of mass to length in pelagic animals: ontogenetic shape change and its implications. *Limnology and Oceanography, 57*(5), 1579–1590. https://doi.org/10.4319/lo.2012.57.5.1579.

Kahsay, Y. (2017). Evaluation of hot pepper varieties (capsicum species) for growth, dry pod yield and quality at M/Lehke District, Tigray, Ethiopia. *International Journal of Engineering Development and Research, 5*(3), 15–27.

Karimi, S., Tavallali, V., Rahemi, M., Rostami, A. A., & Vaezpour, M. (2009). Estimation of leaf growth on the basis of measurements of leaf lengths and widths, choosing pistachio seedlings as model. *Australian Journal of Basic and Applied Sciences, 3*(2), 1070–1075.

Keramatloua, I., Sharifani, M., Sabouri, H., Alizadeh, M., & Kamkar, B. (2015). A simple linear model for leaf area estimation in Persian walnut (Juglansregia L.). *Scientia Horticulturae, 184*, 36–39. https://doi.org/10.1016/j.scienta.2014.12.017.

Kerkhoff, A. J., & Enquist, B. J. (2009). Multiplicative by nature: why logarithmic transformation is necessary in allometry. *Journal of Theoretical Biology, 257*(3), 519–521. https://doi.org/10.1016/j.jtbi.2008.12.026.

Legendre, P., & Legendre, L. (2012). *Numerical ecology* (3rd ed.). Amsterdam, Boston: Elsevier.

Packard, G. C. (2013). Fitting statistical models in bivariate allometry: scaling metabolic rate to body mass in mustelid carnivores. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 166*(1), 70–73. https://doi.org/10.1016/j.cbpa.2013.05.013.

Park, J.B. 1999. Red pepper and kimchi in Korea. Chile Pepper Institute: 8(1): 1–7.

Peake, A. J., & Quinn, G. P. (1993). Temporal variation in species-area curves for invertebrates in clumps of an intertidal mussel. *Ecography, 16*(3), 269–277. https://doi.org/10.1111/j.1600-0587.1993.tb00216.x.

Perez, J. R. D., Ordonez, C., Fernandez, A. B. G., Ablanedo, E. S., Valenciano, J. B., & Marcelo, V. (2018). Leaf water content estimation by functional linear regression of field spectroscopy data. *Biosystems Engineering, 165*, 36–46. https://doi.org/10.1016/j.biosystemseng.2017.08.017.

Pompelli, M. F., Antunes, W. C., Ferreira, D. T. R. G., Cavalcante, P. G. S., Wanderley-Filho, H. C. L., & Endres, L. (2012). Allometric models for non-destructive leaf area estimation of Jatrophacurcas. *Biomass and Bioenergy, 36*, 77–85. https://doi.org/10.1016/j.biombioe.2011.10.010.

Poorter, L., & Bongers, F. (2006). Leaf traits are good predictors of plant performance across 53 rain forest species. *Ecology, 87*(7), 1733–1743. https://doi.org/10.1890/0012-9658(2006)87[1733:LTAGPO]2.0.CO;2.

Qiana, S. S., & Cuffney, T. F. (2012). To threshold or not to threshold? That's the question. *Ecological Indicators, 15*(1), 1–9. https://doi.org/10.1016/j.ecolind.2011.08.019.

Reddy, K. R., & Kakani, V. G. (2007). Screening Capsicum species of different origins for high temperature tolerance by in vitro pollen germination and pollen tube length. *Scientia Horticulturae, 112*, 130–135. https://doi.org/10.1016/j.scienta.2006.12.014.

Rouphael, Y., Mouneimne, A. H., Ismail, A., Mendoza-De Gyves, E., Rivera, C. M., & Colla, G. (2010). Modeling individual leaf area of rose (Rosa hybrida L.) based on leaf length and width measurement. *Photosynthetica, 48*(1), 9–15. https://doi.org/10.1007/s11099-010-0003-x.

Salazar, J. C. B., Melgarejo, L. M., Bautista, E. H., Di Rienzoand, J. A., & Casanoves, F. (2018). Non-destructive estimation of the leaf weight and leaf area in cacao (Theobroma cacao L.). *ScientiaHorticulturae, 229*, 19–24. https://doi.org/10.1016/j.scienta.2017.10.034.

Sang, M. K., Chun, S. C., & Kim, K. D. (2008). Biological control of Phytophthora blight of pepper by antagonistic rhizobacteria selected from a sequential screening procedure. *Biology Control, 46*(3), 424–433. https://doi.org/10.1016/j.biocontrol.2008.03.017.

Schmid, P. E. (2000). Relation between population density and body size in stream communities. *Science, 289*(5484), 1557–1560. https://doi.org/10.1126/science.289.5484.1557.

Seuront, L. (2010). *Fractals and multifractals in ecology and aquatic science*. Boca Raton: CRC Press/Taylor & Francis, Boca Raton.

Van Kleunen, M., Weber, E., & Fischer, M. (2010). A meta-analysis of trait differences between invasive and non-invasive plant species. *Ecology Letters, 13*(2), 235–245. https://doi.org/10.1111/j.1461-0248.2009.01418.x.

Williams, L., & Martinson, T. E. (2003). Nondestructive leaf area estimation of 'Niagara' and 'De Chaunac' grapevines. *Scientia Horticulture, 98*(4), 493–498. https://doi.org/10.1016/s0304-4238(03)00020-7.

Willis, A. J., Begon, M., Harper, J. L., & Townsend, C. R. (1997). Ecology: individuals, populations, and communities. *The Journal of Ecology, 85*(3), 397–398. https://doi.org/10.2307/2960512.

Yu, P., Low, M. Y., & Zhoua, W. (2018). Design of experiments and regression modelling in food flavour and sensory analysis: a review. *Trends in Food Science & Technology, 71*, 202–215. https://doi.org/10.1016/j.tifs.2017.11.013.