Md. Matiur Rahaman[1] / Md. Asif Ahsan[1] / Zeeshan Gillani[1,2] / Ming Chen[1]

Digital Biomass Accumulation Using High-Throughput Plant Phenotype Data Analysis

[1] Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou 310058, China, E-mail: mchen@zju.edu.cn

[2] COMSATS Institute of Information Technology – MA Jinnah Campus, Computer Science 1km Defense Road, Lahore 54000, Pakistan

Abstract:

Biomass is an important phenotypic trait in functional ecology and growth analysis. The typical methods for measuring biomass are destructive, and they require numerous individuals to be cultivated for repeated measurements. With the advent of image-based high-throughput plant phenotyping facilities, non-destructive biomass measuring methods have attempted to overcome this problem. Thus, the estimation of plant biomass of individual plants from their digital images is becoming more important. In this paper, we propose an approach to biomass estimation based on image derived phenotypic traits. Several image-based biomass studies state that the estimation of plant biomass is only a linear function of the projected plant area in images. However, we modeled the plant volume as a function of plant area, plant compactness, and plant age to generalize the linear biomass model. The obtained results confirm the proposed model and can explain most of the observed variance during image-derived biomass estimation. Moreover, a small difference was observed between actual and estimated digital biomass, which indicates that our proposed approach can be used to estimate digital biomass accurately.

1 Introduction

The advent of next-generation sequencing technology has had a major impact on genomics, and the genomic analysis has become routine for most agricultural crop species [1], [2], [3], [4]. Due to the increased availability of high-throughput genotyping platforms, there are many economically important crop varieties that have since been sequenced and annotated. It has significantly contributed to the increase in agricultural productivity. However, satisfing the demand of a growing world population still presents a tremendous challenge for crop improvement [5]. Although genomics techniques have been advancing rapidly, conventional plant phenotyping lags far behind compared to current genotyping systems. To relieve this bottleneck, various research institutes have been established or many plant research laboratories are planning to establish their own phenotyping system. They employ robotics, automation and use various imaging techniques [6], [7], [8], [9], [10]. This advanced phenotyping aims at a quantification of photosynthesis, development, architecture, growth or biomass productivity of single plants by accelerating plant breeding programs [10], [11]. Automated, high-throughput phenotyping facilities have enabled to grow hundreds to thousands of plants maintained in a controlled environment and automatically photographed each day from the standard position. This image data is analysed via image analysis algorithm and software to extract phenotypic traits.

To characterize plant architecture and performance, image analysis methods have become more popular. This method has the capacity to measure many dynamically morphological and physiological traits of a given individual [12]. Plant phenotyping is the quantitative or qualitative study of these traits at any organizational level, in a particular genomic expression state and environment [13]. Generally, plant phenotypic traits of interest can be classified as physiological, structural, or performance-related. Performance-related traits are defined by the complex traits (such as shoot fresh/dry weight, yield) which eventually determine plant performance

in terms of biomass and yield. Plant biomass is a vital trait in the study of functional plant biology and growth analysis. Second, repeated measurements of plant biomass are the source for the calculation of growth rates and net primary production [14], [15]. Thus, plant biomass analysis is a basis for unraveling a number of complex questions of plant growth, development and response to the environment.

There are several techniques to measure plant biomass depending on the available budget, required accuracy, structure and composition of the vegetation, and also numerous disciplines of plant biology [16], [17]. However, it is difficult to make a generally applicable statement regarding the best technique for biomass estimation. The standard method for biomass determination of individual plant is defined to measure shoot fresh (SFW) biomass or the oven-dried shoot dry (SDW) biomass [14], [18]. For the dry shoot biomass, the plant is harvested and oven-dried at the end of the experiment. It is considered one of the widely acceptable measures for studying the biomass of an individual plant [17]. These customary plant biomass measuring methods are destructive. Consequently, they require many individuals to be cultivated for repeated measurements that are labor-intensive and time-consuming. Imaging-based phenotyping has enabled the non-destructive assessment of plant responses to the environment over time, and allows determination of plant biomass without having to harvest the whole plant [19], [20], [21], [22].

A non-destructive method based on digital image analysis addresses not only above-ground fresh and oven-dried biomass. It has also been applied for estimating above-ground forest and canopy biomass for remote sensing, satellite, and airborne images [23], [24], [25], [26], [27], [28], [29], [30], [31], [32]. For that reason, imaging techniques are now the most commonly used method for estimating biomass in ecology and agriculture.

A number of linear and non-linear functions are used to model biomass accumulations [15], [17], [32], [33], [34], [35]. However, non-linear models are complicated due to the fact that higher order model coefficients are insignificant [17]. The models predominantly cited in literature to estimate the biomass of a plant were generated by destructively measured parameters as response variables, and parameters derived from image analysis as predictor variables [15], [17], [34].

Several image-based biomass studies considered linear methods for estimating the biomass as a linear function of plant area [17], [36] which perform better than non-linear models, such as quadratic, cubic and power methods [15], [17], [27]. However, the estimation error from this model is large, prohibiting accurate estimation of the biomass of plants. Color pixel-based traits and a mixed variable (area × days) were also used in the image-based biomass model as predictor variables, where the response variable SDW was destructively measured [15], [17]. For the accurate inference of biomass and to bridge the genotype to phenotype gap, it is crucial to recognize more significant traits, particularly for the stressed plants. Plant compactness is an important phenotypic trait that reflects plant density and architecture [9]. Thus, there is a need to develop such a method for estimating biomass that takes into consideration plant compactness as well. Although, Yang et al. [9] used biomass models by considering plant compactness; however, the response variable of those models was destructively measured.

The objective of the present study is to develop a generalized linear model to estimate accurate plant biomass without using destructively measured parameters. We attempt to develop a linear biomass estimation model to estimate plant biomass based on image-derived phenotypic traits. We have demonstrated our model that uses mixed variables of plant area and their age and plant compactness, which significantly reduces estimation errors. This model can be used to acquire more insights by accurately estimating the plant biomass using the non-destructive approach from high-throughput phenotype images.

2 Materials and Methods

2.1 Experiment and Data Description

We analyzed a barley image dataset downloaded from http://iapg2p.sourceforge.net/modeling/#dataset. The Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany generated this high-throughput phenotype data set [35]. The summarized description of that experiment is given below:

LemnaTec HTS-Scanalyzer 3D platform was used to screen 16 German two-rowed spring barley cultivars (cv.) and two parents of a DH-mapping population (cv. Morex and cv. Barke) for vegetative drought tolerance. Plants grew under controlled greenhouse conditions and were phenotyped on a daily basis over the entire experimental phase using the fully automated phenotyping system consisting of conveyer belts, a weighing and watering station, and three imaging sensors.

The experiments were performed consecutively from May to July 2011. The experiment consisted of two treatments: well-watered (control treatment) and water limited (drought stress treatment) over a period of 58 days. Drought stress was applied by withholding water from 27th day after sowing until 44th day. Stressed

plants were re-watered on the 45th day. Control plants remained well watered at a field capacity of 90 %. After the stress period (27th–44th days), all plants were re-watered to 90 % field capacity (FC) and kept well-watered again up to 58 days. The greenhouse growth conditions were set to 18 °C and 16 °C during the day and night, respectively. The daylight period lasted ~13 h, starting at 7 AM.

During each treatment, six plants per DH parent and nine plants per core set cultivar were tested. A total of 312 plants were used for this study. One top view image and three side view images were obtained per plant at different angles. SFW and SDW measured manually when plants were harvesting at the 58th day [35].

2.2   Image Analysis Description

Chen et al. [35] performed image analysis through IAP software to extract quantitative information from the barley plant images [37]. These phenotype images were exported and analyzed using the barley analysis pipeline with optimized parameters. Image processing operations included steps: pre-processing, to prepare the images for segmentation; segmentation, to divide the image into foreground and background section accordingly, and feature extraction. The analyzed features were exported in .csv file format. A detailed description of this data set is available at [35].

The extracted features i.e. plant pixel area from all side and top view images were summed to give the projected shoot area [2], [17], [19], [21], [38]. Extracted plant pixel area from top and side view images were also used to calculate a volume (unit: voxel), termed as "digital biomass" that corresponds to a pixel volume and defined as [22],

$$\text{Digital biomass} = \sqrt{\text{average pixel side area}^2 \times \text{top area}}$$

Digital biomass was used as a proxy of the quantitative estimator for plant fresh biomass (SFW).

3   Model Developments

A schematic workflow for development of the biomass model based on high-throughput plant imaging is shown in Figure 1. It shows the image-based biomass model construction steps ranging from the experiment to the model to be developed.
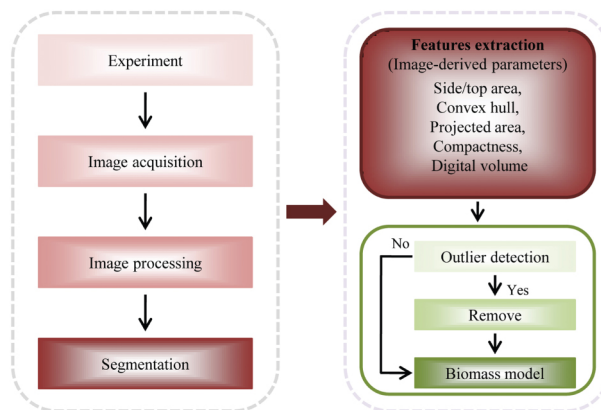

Figure 1: Workflow for the image-derived biomass model construction. High-throughput imaging data from automated phenotyping system require image storage and image processing (Left). Then expected features/phenotypic traits need to be extracted from the segmented images. Eliminating outliers in the phenotypic data is another key pre-processing step before model construction, and then need to perform the biomass model for estimating image-derived biomass (Right).

We have considered a linear model

$$D_b = a_0 + a_1 \times A + e_0 \tag{1}$$

where we defined D_b as digital volume (response variable), A as projected shoot area (predictor variable), a_0 as model intercept, a_1 as model coefficient and e_0 as model error. The biomass estimation error of this model is

large, unable to explain observed variances, and moreover there is a big difference between actual and estimated biomass. To improve the model [1], Golzarian et al. [17] proposed a linear predictive model based on the concept of plant specific weight (PSW), defined as the plant weight per total projected shoot area.

We have generalized this linear biomass model based on plant compactness. This phenotypic trait provides meaningful information on plant architecture in addition to the commonly recognized agronomic traits such as plant height, tiller number and green leaf area [9]. This is the reason why we chose plant compactness as a predictor to improve the biomass model. Plant compactness was calculated as the square of plant border length divided by the projected side or top area [39].

We extended equation (1) by including trait (predictor) plant compactness, compactness × days, and area × days. Then the associated equations with our proposed predictive biomass models can be written as,

$$\text{Model 1: } D_b = a_0 + a_1 \times A + a_2 \times PC + e_0$$

$$\text{Model 2: } D_b = a_0 + a_1 \times A + a_2 \times PC \times HD + e_0$$

$$\text{Model 3: } D_b = a_0 + a_1 \times A + a_2 \times A \times HD + a_3 \times PC + e_0$$

Where PC and HD are plant compactness and plant age in days after planting, respectively.

The performance of these models was assessed through five-fold cross-validation technique [9], [17]. For comparison of model fitting and model superiority, we used the following model assessment criteria:

1. The Pearson correlation coefficient (PCC; r) between the predicted biomass and the observed biomass [26], [36],

$$r = \frac{\sum_{i=1}^{n} \left(D_{bi} - \overline{D_b} \right) \left(\widehat{D_{bi}} - \widetilde{D_b} \right)}{\sqrt{\sum_{i=1}^{n} \left(D_{bi} - \overline{D_b} \right)^2} \sqrt{\sum_{i=1}^{n} \left(\widehat{D_{bi}} - \widetilde{D_b} \right)^2}}$$

2. The coefficient of determination, R^2 [26], [31], [34], [35], [40], [41],

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(D_{bi} - \widehat{D_{bi}} \right)^2}{\sum_{i=1}^{n} \left(D_{bi} - \overline{D_b} \right)^2}$$

3. The root mean squared relative errors, RMSRE [34],

$$\text{RMSRE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{(D_{bi} - \widehat{D_{bi}})}{D_{bi}} \right)^2}$$

where, D_{bi}, observed biomass; $\widehat{D_{bi}}$, predicted biomass; $\overline{D_b}$, mean value of the observed biomass; $\widetilde{D_b}$, mean value of the predicted biomass; n, the number of data points.

3.1 Cross-Validation Technique

Cross-validation is a standard technique for assessing the prediction error of a model [17], [31], [42]. In cross-validation, observations are randomly assigned indices, integer 1 to M, and the dataset is partitioned into M approximately equal-sized parts. Let $m : \{1, \ldots, N\} \rightarrow \{1, \ldots, M\}$ be an indexing function that indicate the partition to which observation is allocated by the randomization. The fitted function denoted by $\hat{f}^{-m}(x)$ computed by removing m-th part from the data, then the cross-validation error is given by:

$$CV(\hat{f}) = \frac{1}{N} \sum_{i=1}^{N} L(y_i, \hat{f}^{-m(i)}(x_i))$$

Thus, cross-validation error is the average of the loss function (L), evaluated using model trained on different subsets of the data [43]. The superscript $-m(i)$ means model f is trained without the training patterns in the same partition of the dataset as pattern i. By applying this technique, obtained estimation errors were used to validate the performance of the predictive models.

Cross-validation is a robust method and preferred over the R^2 statistic. R^2 inevitably increases with additional predictors within one dataset. However, cross-validation error decreases only as long as the additional predictor improves the prediction accuracy of the model in an independent dataset [17]. The cross-validation analysis was performed using the R package "DAAG" [44]. All statistical analysis was performed using the R software.

4   Results

The models in this study were developed using a barley plant phenotype data set collected in the experiment explained earlier. We constructed three models, where digital volume is a function of inputs of area and compactness; area and compactness × HD; and area, area × HD and compactness. The coefficients of each model were estimated using regression analysis. All of these coefficients contribute significantly to the predicted value of digital biomass (Table 1). The average PCC, R^2, and RMSRE with standard error (SE) of two treatment categories are given in Table 2.

Table 1: Significance of regression coefficients of the proposed model.

Treatment	Model	Coefficients	Coefficients value	SE	t-value	Sig.
Control	Model 1	a_0	−2.306351	0.021996	−104.85	0.00
		a_1	1.661959	0.003353	495.62	0.00
		a_2	−0.195737	0.004939	−39.63	0.00
	Model 2	a_0	−2.52E+00	7.74E−02	−32.518	0.00
		a_1	1.56E+00	8.04E−03	193.485	0.00
		a_2	−5.99E−05	7.50E−05	−0.798	0.01
	Model 3	a_0	−1.33E+00	6.52E−02	−20.34	0.00
		a_1	1.55E+00	7.49E−03	207.66	0.00
		a_2	6.37E−04	4.01E−05	15.9	0.00
		a_4	−1.90E−01	4.73E−03	−40.2	0.00
Stress	Model 1	a_0	−2.377423	0.053426	−44.5	0.00
		a_1	1.665137	0.006043	275.55	0.00
		a_2	−0.187755	0.006353	−29.56	0.00
	Model 2	a_0	−1.39E+00	7.74E−02	−17.93	0.00
		a_1	1.44E+00	7.22E−03	199.56	0.00
		a_2	7.95E−04	4.06E−05	19.61	0.00
	Model 3	a_0	−5.53E−01	5.86E−02	−9.436	0.00
		a_1	1.50E+00	6.03E−03	248.293	0.00
		a_2	9.11E−04	2.07E−05	44.04	0.00
		a_4	−2.20E−01	4.99E−03	−44.103	0.00

Table 2: The comparison of three proposed models using PCC, R^2, and RMSRE.

Treatment	Model	PCC	SE	R^2	SE	RMSRE	SE
Control	Model 1	0.97	0.0001	0.94	0.0001	0.0061	0.0003
	Model 2	0.97	0.0001	0.94	0.0001	0.0067	0.0003
	Model 3	0.99	0.0001	0.98	0.0001	0.0061	0.0003
Stress	Model 1	0.98	0.0002	0.96	0.0004	0.0066	0.0006
	Model 2	0.97	0.0002	0.94	0.0004	0.0072	0.0006
	Model 3	0.99	0.0002	0.98	0.0004	0.0064	0.0006

According to Table 2, Model 1 and Model 2 provide almost the same resulting PCC (0.97 ± 0.0001) and R^2 (0.94 ± 0.0001) value, and the PCC and R^2 value of Model 3 is 0.99 ± 0.0001 and 0.98 ± 0.0001 for control data set on average. For the stress dataset resulting values, Model 1 provides the PCC (0.98 ± 0.0002) and R^2

(0.96 ± 0.0004), Model 2 provides the PCC (0.97 ± 0.0002) and R^2 (0.94 ± 0.0004) and Model 3 provides the PCC (0.99 ± 0.0002) and R^2 (0.98 ± 0.0004). The estimation error (RMSRE) of Model 2 is 0.0067 ± 0.0003, where Model 1 and Model 2 RMSRE are 0.0061 ± 0.0003 for the control dataset. In stress datasets, Model 3 produces smaller RMSRE (0.0064 ± 0.0006) than Model 1 and Model 2 (Table 2).

Among these three proposed models, Model 3 performed better, when area, area × HD and compactness are considered as predictors in the biomass model. However, we found that Model 1 and Model 2 provide almost similar results, but still less superior than Model 3 according to the estimation error. Therefore, we propose to use Model 3, which looks like an integrated biomass model of Yang et al. [9] and Golzarian et al. [17] for digital biomass estimation.

4.1   Performance Evaluations of the Proposed Model

We compared our proposed model (Model 3) with the following linear models, which are available for biomass estimation from two-dimensional images [17], [24], [36],

$$\text{Model A} : D_b = a_0 + a_1 \times A + e_0$$

$$\text{Model B} : D_b = a_0 + a_1 \times A + a_2 \times A \times \text{HD} + e_0$$

The resulting root mean squared relative errors after applying cross-validation technique from Model A, B and the proposed one are compared in Figure 2. Figure 2 shows the RMSRE of three cases of the experiment: before stress period (Figure 2A and B), during stress period (Figure 2C and D) and during the recovery period (Figure 2E and F). The resulting PCC (SE) and R^2 (SE) are given in Table 3.
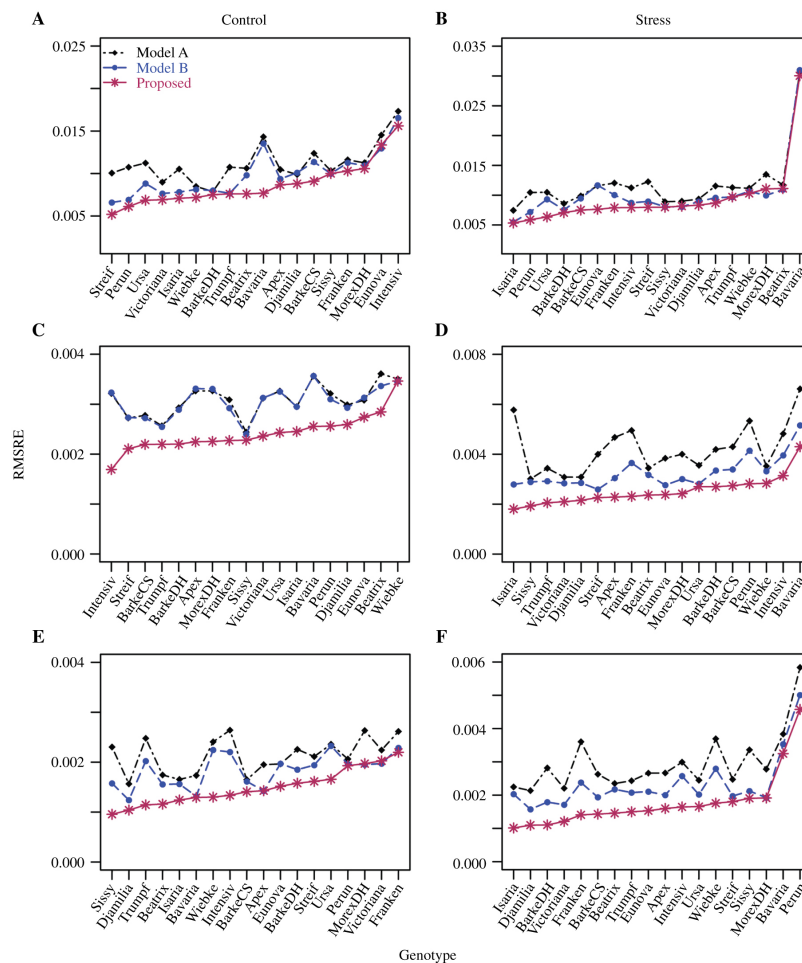

Figure 2: Performance evaluations of the proposed image-derived biomass model. Performance evaluations of the proposed model in different barley cultivars using RMSRE, under different water condition of two treatments.

Table 3: The comparison of Model A, Model B and our proposed model using PCC and R^2 values.

Period	Model	Control				Stress			
		PCC	SE	R^2	SE	PCC	SE	R^2	SE
Before stress	Model A	0.94	0.0007	0.88	0.0012	0.93	0.0015	0.86	0.0027
	Model B	0.95	0.0005	0.90	0.0011	0.94	0.0015	0.88	0.0028
	Proposed	**0.98**	**0.0004**	**0.96**	**0.0010**	**0.98**	**0.0015**	**0.96**	**0.0026**
Stress	Model A	0.93	0.0004	0.86	0.0002	0.95	0.0012	0.90	0.0023
	Model B	0.94	0.0002	0.88	0.0002	0.96	0.0005	0.92	0.0009
	Proposed	**0.99**	**0.0001**	**0.98**	**0.0002**	**0.99**	**0.0003**	**0.98**	**0.0002**
Recovery	Model A	0.93	0.0017	0.86	0.0033	0.92	0.0039	0.84	0.0071
	Model B	0.95	0.0011	0.90	0.002	0.94	0.0028	0.88	0.0049
	Proposed	**0.98**	**0.0010**	**0.96**	**0.0018**	**0.99**	**0.0024**	**0.98**	**0.0043**

Before the stress period, the RMSRE of the proposed model is smaller than Model A (Figure 2A) and Model B (Figure 2B) for each barley cultivar. The RMSRE of the stress period is illustrated in Figure 2C and D. Here, the proposed model produces notably smaller RMSRE than Model A and Model B. The RMSRE of Model A and Model B are also much higher during the recovery period than the proposed model (Figure 2E and F). We observed that in all cases the proposed model produces a significantly smaller (p-value < 0.00001) RMSRE.

Similar results were obtained for the PCC and R^2 (Table 3). Table 3 describes the average PCC and R^2 values with SE of different water condition. In all cases, the proposed models' PCC and R^2 average values are higher than the Model A and Model B with smaller SE. These findings show that the proposed model provides better results when compared to Model A and Model B.

By modeling biomass as a function of plant area, plant age, and compactness, a small difference was observed between actual and predicted biomass for drought-stress and control plants as shown in Figure 3A ($R^2 \geq 0.99$). It indicates that the proposed model explained 99 % of the dataset observed in all barley cultivars. The estimation bias of different barley cultivars under different water conditions is shown in Figure 3C and D. It is the average overall images of the image derived digital biomass $_{(predicted)}$ −digital biomass $_{(observed)}$. We observed that the estimated biases of the proposed model are close to zero for each barley cultivar. Maximum bias has obtained 0.001 for the cultivars 'Bavaria' for both the controlled and stressed plants, respectively (Figure 3C and D).
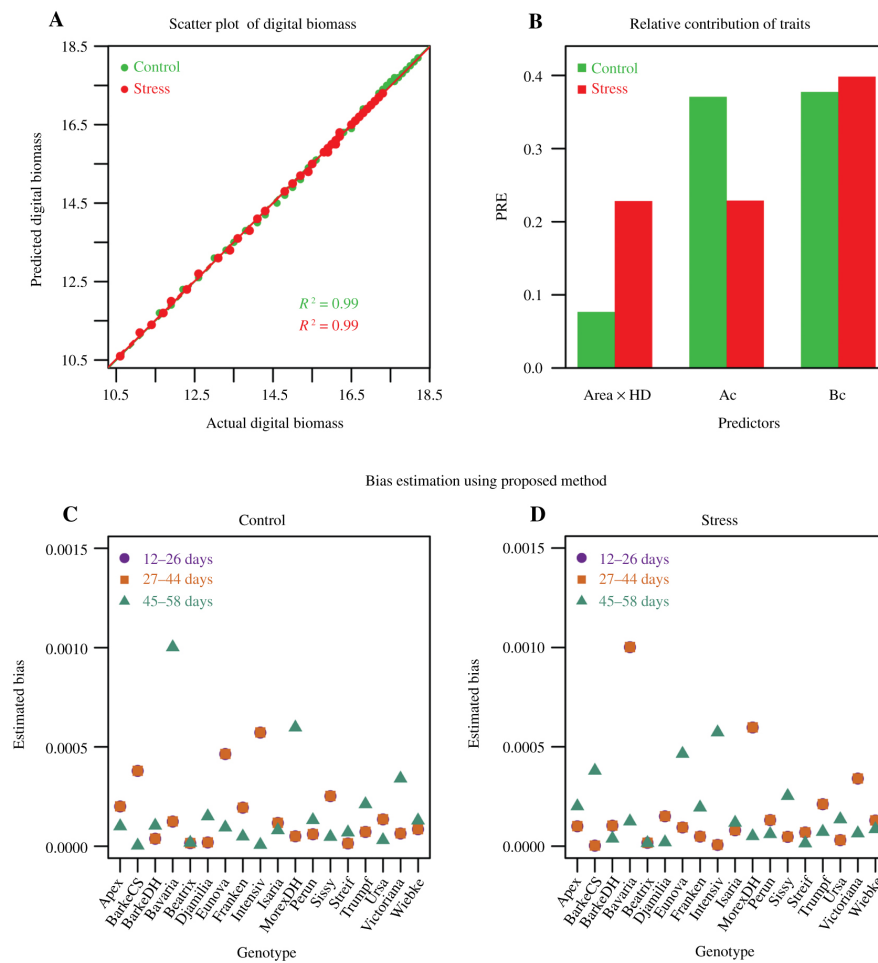
Figure 3: Models accuracy and the relative contribution of traits. (A) Scatter plot of image-derived actual biomass compared with the estimated values using the proposed model for the controlled and stressed plants, (B) the relative contribution of model predictors (traits) area × days and compactness for the image derived biomass model used in this study, (C,D) bias estimation using the proposed model in different barley cultivars under different water condition of two treatments.

Scatter plots of manual versus image-derived predicted biomass highlighted that the digital biomass predicted from images using our proposed approach is significantly correlated with the manually measured biomass (Figure 4). The correlation values between digital biomass and manually measured biomass ranged from 0.73 to 0.91 (Figure 4A and B for SFW and Figure 4C and D for SDW).
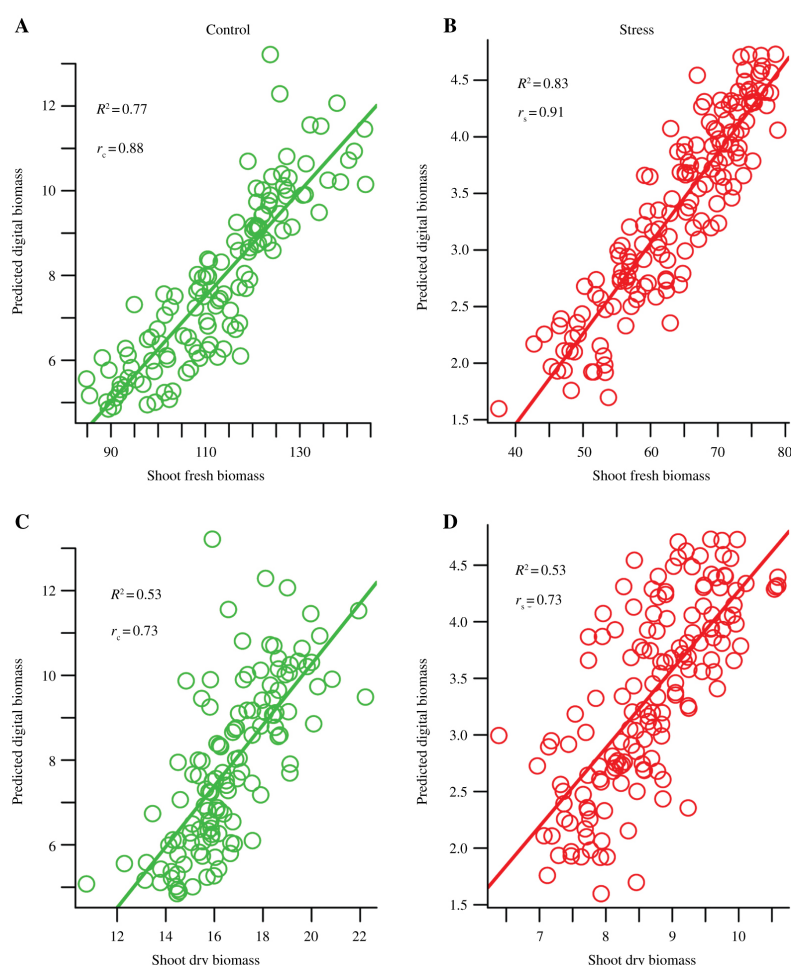
Figure 4: Correlation between image-derived predicted biomass and manually measured biomass. Scatter plots of image-derived predicted biomass with manual measurements. The manual measurements include (A,B) shoot fresh biomass (SFW) and (C,D) shoot dry biomass (SDW) when plants were harvested at day 58.

Prediction performance according to the plant age can be seen graphically in the bar plots of Figure 5 where average biomass estimation error is higher when plant age is 12–26 days. The standard error (SE) is also much higher in this time period. The estimation error gradually decreases as the plant age increases with smaller SE (Figure 5). In our study, when the plant age is between 27 and 58 days, the proposed model provides more accurate inference about digital biomass. Digital biomass estimation error is significantly smaller at the age 45–58 days on average. These results proved that the better prediction of digital biomass from images is also depending on plant age. Another observation is that the prediction error of control plants is smaller than stress plants in all cases (Figure 5).
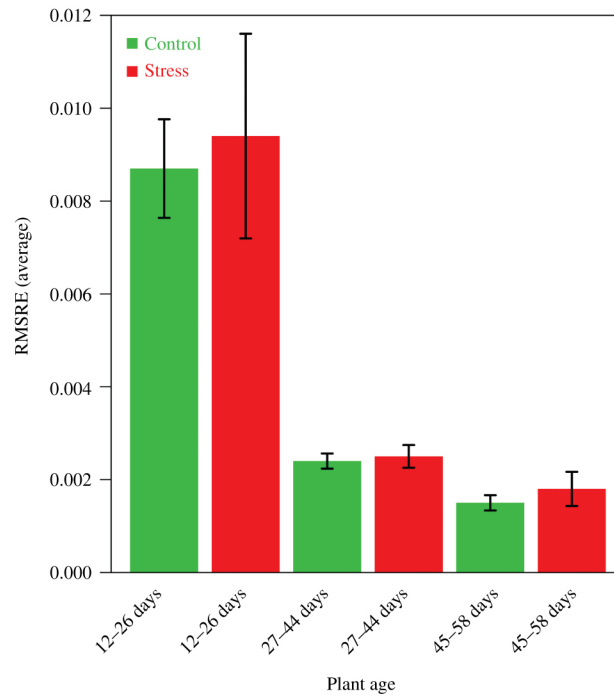
Figure 5: Prediction accuracy of the proposed model in terms of time after planting (plant age). Average estimation error (RMSRE) of the proposed model in terms of plant age.

4.2   The Relative Contribution of Predictors in Biomass Model

To assess the relative contribution of the predictor (phenotypic traits) in our proposed model, we compared models according to the approach of Judd et al. [45]. We formulate proportional reduction of error (PRE) with the following equation that represents the effect size of model predictor,

$$ \text{PRE} = \frac{(\text{RSS}_{(i)} - \text{RSS}_{(j)})}{\text{RSS}_{(i)}}; \ i, j = 1, 2, 3; i \neq j. $$

where, RSS is the residuals sum of the square of i-th and j-th model [45].

Figure 3B describes predictor area $\times HD$ has 7.67–22.81 % contribution in explaining the variance of digital biomass. Although Golzarian et al. [17] described that this predictor has a significant contribution to estimate destructively measured biomass. In case we considered phenotypic trait compactness as an additional predictor in Model A and Model B denoted by trait A_c and B_c, respectively, the trait compactness has 22.89–39.81 % contributions in explaining the variance of digital biomass (Figure 3B). PRE-values for the proposed model are 0.38 and 0.40 for two treatment plants, respectively. This measurement allowed a deeper understanding of plant compactness importance in biomass prediction [9].

5   Discussion

Plant growth and development studies demonstrated that automated digital imaging is a powerful tool to relieve the plant phenotyping bottleneck [2], [9], [22], [35], [46], [47], [48], [49], [50]. Plant reveal complex phenotypic traits, and thus the main challenge is to analyze and model phenotypic traits that bridge the genotype-phenotype gap [22], [51].

In the emerging period of plant phenomics there is a need to improve existing methods or develop new ones to resolve this analytical bottleneck [10], [12], [35], [52]. It has been realized that estimation of plant biomass is important from phenotype images of cereal plants. Therefore, a practical analysis framework or approach for the estimation of biomass using image-derived traits is needed. Image-based biomass estimation methods developed so far include total harvesting of plants or harvesting sample during measurement SFW and SDW as well [15], [17].

We proposed a framework to estimate plant biomass in terms of digital image analysis. Since, most protocols of the traditional methods for biomass determination of individual plants are destructive [9], [14], [15], [17],

[18]. There are difficulties with that method in measuring dynamic responses of plant growth under environment, and to collect seed from the individuals being measured [21]. We designed image-based non-destructive approach which allows determination of biomass without harvesting the whole plant.

We have used the digital volume which is highly correlated with destructively measured plant complex traits [9], [21], [22], [35]. Based on a proof of concept from the recent non-destructive biomass study, we constructed a linear model for biomass determination [21], [22], [34], [35]. For digital biomass prediction, the noticeable improvement was achieved by adding plant compactness into the model. It has significantly reduced bias in biomass estimation of cereal plants, which is the major reason of the estimation error. By contrast, other biomass estimation models resulted in large estimation error. From the analysis of results, we found that plant compactness has a good impact on digital biomass prediction.

The digital biomass predicted using our proposed model is highly correlated with the real biomass (SFW/SDW). Comparing our proposed model with the existing models, it is evident that the analyzed results confirmed the proposed model which performed well in all cases, and the model improvement is consistent and significant.

Overall, analysis results confirmed the idea that the plant compactness, which was used as an additional input for the proposed model, plays a significant role in reducing the error for estimating digital biomass. Our proposed approach provides a practical method for estimating digital biomass as a substitute method for the conventional methods. Finally, the proposed model performed better than a conventional model with smaller prediction errors. The overall performance of the new model was superior to that of the traditional model, and there were significant differences between the traditional and new model during digital biomass analysis.

6   Conclusion

In this study, we focus on a method for the accurate inference of digital biomass from high-throughput phenotype images. Our proposed model employs information obtained from image-derived traits of plants and their age. It enables the high-throughput non-destructive estimation of biomass for cereal plants regardless of whether or not plants are stressed. The method has been tested using imaged barley data sets under drought stress treatment. Comparing the obtained results from our model with the results of the existing models, we conclude that the proposed model in most cases performs better than the existing ones. The obtained results based on the presented approach demonstrated that the proposed biomass model is robust and accurate. This would be useful to advance our views for the accurate estimation of digital biomass in high-throughput image analysis.

Author Contributions: MMR contributed to the conception and the development of the method, prepared the manuscript and analyzed the results. MAA and ZG prepared and revised the manuscript. MC contributed to the design and conception of the project, critically read and approved the final manuscript. All authors read and approved the final manuscript.

Conflict of interest statement: Authors state no conflict of interest. All authors have read the journal's Publication ethics and publication malpractice statement available at the journal's website and hereby confirm that they comply with all its parts applicable to the present scientific work.

References

[1] Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery . Nat Rev Genet. 2011;12:745–55.

[2] Neilson EH, Edwards AM, Blomstedt CK, Berger B, Moller BL, Gleadow RM. Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C-4 cereal crop plant to nitrogen and water deficiency over time. J Exp Bot. 2015;66:1817–32.

[3] Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. Cell. 2013;155:27–38.

[4] Edwards D, Batley J, Snowdon RJ. Accessing complex crop genomes with next-generation sequencing. Theor Appl Genet. 2013;126:1–11.

[5] Takeda S, Matsuoka M. Genetic approaches to crop improvement: responding to environmental and population changes. Nat Rev Genet. 2008;9:444–57.

[6] Granier C, Aguirrezabal L, Chenu K, Cookson SJ, Dauzat M, Hamard P, et al. PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. New Phytol. 2006;169:623–35.

[7] Skirycz A, Vandenbroucke K, Clauw P, Maleux K, De Meyer B, Dhondt S, et al. Survival and growth of Arabidopsis plants given limited water are not equal. Nat Biotechnol. 2011;29:212–4.

[8] Tisne S, Serrand Y, Bach L, Gilbault E, Ben Ameur R, Balasse H, et al. Phenoscope: an automated large-scale phenotyping platform offering high spatial homogeneity.. Plant J. 2013;74:534–44.

[9] Yang WN, Guo ZL, Huang CL, Duan LF, Chen GX, Jiang N, et al. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat Commun. 2014;5:5087.

[10] Rahaman MM, Chen D, Gillani Z, Klukas C, Chen M. Advanced phenotyping and phenotype data analysis for the study of plant growth and development. Front Plant Sci. 2015;6:619.

[11] Walter A, Studer B, Kolliker R. Advanced phenotyping offers opportunities for improved breeding of forage and turf species. Ann Bot-London. 2012;110:1271–9.

[12] Granier C, Vile D. Phenotyping and beyond: modelling the relationships between traits. Curr Opin Plant Biol. 2014;18:96–102.

[13] Dhondt S, Wuyts N, Inze D. Cell to whole-plant phenotyping: the best is yet to come. Trends Plant Sci. 2013;18:433–44.

[14] Cornelissen JH, Lavorel S, Garnier E, Diaz S, Buchmann N, Gurvich DE, et al. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. Aust J Bot. 2003;51:335–80.

[15] Tackenberg O. A new method for non-destructive measurement of biomass, growth rates, vertical biomass distribution and dry matter content based on digital image analysis. Ann Bot. 2007;99:777–83.

[16] Catchpole WR, Wheeler CJ. Estimating Plant Biomass – a Review of Techniques. Aust J Ecol. 1992;17:121–31.

[17] Golzarian MR, Frick RA, Rajendran K, Berger B, Roy S, Tester M, et al. Accurate inference of shoot biomass from high-throughput images of cereal plants. Plant Methods. 2011;7:2.

[18] Schwinning S, Weiner J. Mechanisms determining the degree of size asymmetry in competition among plants. Oecologia. 1998;113:447–55.

[19] Rajendran K, Tester M, Roy SJ. Quantifying the three main components of salinity tolerance in cereals. Plant Cell Environ. 2009;32:237–49.

[20] Furbank RT, Tester M. Phenomics – technologiess to relieve the phenotyping bottleneck. Trends Plant Sci. 2011;16:635–44.

[21] Hairmansis A, Berger B, Tester M, Roy SJ. Image-based phenotyping for non-destructive screening of different salinity tolerance traits in rice. Rice. 2014;7:16.

[22] Neumann K, Klukas C, Friedel S, Rischbeck P, Chen DJ, Entzian A, et al. Dissecting spatiotemporal biomass accumulation in barley under different water regimes using high-throughput image analysis. Plant Cell Environ. 2015;38:1980–96.

[23] Sher-Kaul S, Oertli B, Castella E, Lachavanne J-B. Relationship between biomass and surface area of six submerged aquatic plant species. Aquat Bot. 1995;51:147–54.

[24] Dietz H, Steinlein T. Determination of plant species cover by means of image analysis. J Veg Sci. 1996;7:131–6.

[25] Rutchey K, Vilchek L. Air photointerpretation and satellite imagery analysis techniques for mapping cattail coverage in a northern Everglades impoundment. Photogramm Eng Rem S. 1999;65:185–91.

[26] Montes N, Gauquelin T, Badri W, Bertaudiere V, Zaoui EH. A non-destructive method for estimating above-ground forest biomass in threatened woodlands. Forest Ecol Manag. 2000;130:37–46.

[27] Paruelo JM, Lauenroth WK, Roset PA. Technical note: estimating aboveground plant biomass using a photographic technique. J Range Manage. 2000;53:190–3.

[28] Lu D, Mausel P, Brondizio E, Moran E. Above-ground biomass estimation of successional and mature forests using TM images in the Amazon Basin. In: Richardson DE, van Oosterom P , eds. Advances in spatial data handling, 10th International Symposium on Spatial Data Handling. Springer, 2002:183–96.

[29] Lim KS, Treitz PM. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. Scand J Forest Res. 2004;19:558–70.

[30] Proisy C, Couteron P, Fromard F. Predicting and mapping mangrove biomass from canopy grain analysis using Fourier-based textural ordination of IKONOS images. Remote Sens Environ. 2007;109:379–92.

[31] Vega C, Vepakomma U, Morel J, Bader JL, Rajashekar G, Jha CS, et al. Aboveground-biomass estimation of a complex tropical forest in india using Lidar. Remote Sens-Basel. 2015;7:10607–25.

[32] Flombaum P, Sala OE. A non-destructive and rapid method to estimate biomass and aboveground net primary production in arid environments. J Arid Environ. 2007;69:352–8.

[33] Kuyah S, Dietz J, Muthuri C, Jamnadass R, Mwangi P, Coe R, et al. Allometric equations for estimating biomass in agricultural landscapes: I. Aboveground biomass. Agric Ecosyst Environ. 2012;158:216–24.

[34] Busemeyer L, Ruckelshausen A, Muller K, Melchinger AE, Alheit KV, Maurer HP, et al. Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. Sci Rep-UK. 2013;3. DOI. 10.1038/srep02442.

[35] Chen DJ, Neumann K, Friedel S, Kilian B, Chen M, Altmann T, et al. Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. Plant Cell. 2014;26:4636–55.

[36] Leister D, Varotto C, Pesaresi P, Niwergall A, Salamini F. Large-scale evaluation of plant growth in Arabidopsis thaliana by non-invasive image analysis. Plant Physiol Biochem. 1999;37:671–8.

[37] Klukas C, Chen DJ, Pape JM. Integrated analysis platform: an open-source information system for high-throughput plant phenotyping. Plant Physiol. 2014;165:506–18.

[38] Harris BN, Sadras VO, Tester M. A water-centred framework to assess the effects of salinity on the growth and yield of wheat and barley. Plant Soil. 2010;336:377–89.

[39] Jansen M, Gilmer F, Biskup B, Nagel KA, Rascher U, Fischbach A, et al. Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in Arabidopsis thaliana and other rosette plants. Funct Plant Biol. 2009;36:902–14.

[40] Montes JM, Technow F, Dhillon BS, Mauch F, Melchinger AE. High-throughput non-destructive biomass determination during early plant development in maize under field conditions. Field Crop Res. 2011;121:268–73.

[41] Sanquetta CR, Wojciechowski J, Dalla Corte AP, Behling A, Pellico Netto S, Rodrigues AL, et al. Comparison of data mining and allometric model in estimation of tree biomass. BMC Bioinformatics. 2015;16:247.

[42] Li N, Xie GD, Zhang CS, Xiao Y, Zhang BA, Chen WH, et al. Biomass resources distribution in the terrestrial ecosystem of China. Sustainability-Basel. 2015;7:8548–64.

[43] Lu ZQ. The elements of statistical learning: data mining, inference, and prediction. J R Stat Soc A Stat. 2010;173:693–4.

[44] Bay TF, Schoney RA. Data-analysis with computer-graphics – production-functions. Am J Agric Econ. 1982;64:289–97.

[45] Judd CM, McClelland GH, Ryan CS. Data analysis: a model comparison approach. Routledge, 2011.

[46] Junker A, Muraya MM, Weigelt-Fischer K, Arana-Ceballos F, Klukas C, Melchinger AE, et al. Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. Front Plant Sci. 2014;5:770.

[47] Zhang X, Hause RJ, Borevitz JO. Natural genetic variation for growth and development revealed by high-throughput phenotyping in Arabidopsis thaliana. G3 (Bethesda). 2012;2:29–34.

[48] Slovak R, Goschl C, Su XX, Shimotani K, Shiina T, Busch W. A scalable open-source pipeline for large-scale root phenotyping of Arabidopsis. Plant Cell. 2014;26:2390–403.

[49] Campbell MT, Knecht AC, Berger B, Brien CJ, Wang D, Walia H. Integrating image-based phenomics and association analysis to dissect the genetic architecture of temporal salinity responses in rice. Plant Physiol. 2015;168:1476–U697.

[50] Nagel KA, Bonnett D, Furbank R, Walter A, Schurr U, Watt M. Simultaneous effects of leaf irradiance and soil moisture on growth and root system architecture of novel wheat genotypes: implications for phenotyping. J Exp Bot. 2015;66:5441–52.

[51] Chen D, Chen M, Altmann T, Klukas C. Bridging genomics and phenomics. In: Chen M Hofestädt R , eds. Approaches in integrative bioinformatics. Springer, 2014:299–333.

[52] Moore CR, Johnson LS, Kwak IY, Livny M, Broman KW, Spalding EP. High-throughput computer vision introduces the time axis to a quantitative trait map of a plant growth response. Genetics. 2013;195:1077.