Mass modelling by dimension attributes for Mango (Mangifera indica cv. Zebdia) relevant to post-harvest and food plants engineering

Amr Mossad 1,2, Waleed Kamel ElHelew 1*, Hemat E. Elsheshetawy 3, Vittorio Farina 4

(1 Agricultural Engineering Department, Faculty of Agriculture, Ain Shams University, Cairo, Egypt; 2 Agricultural Engineering Department, College of Food and Agricultural Sciences, King Saud University, Riyadh, Saudi Arabia; 3 Department of Food Science, Faculty of Agriculture, Ain Shams University, Cairo, Egypt; 4 Department of Agricultural and Forestry Sciences, Università degli Studi di Palermo, Palermo, Italy)

Abstract: Mass identification of mango fruits from their dimension attributes remains challenging. This is because of the unregulated shapes of these fruits. Therefore, this research aims at creating mathematical models that can demonstrate the relationship between the fruit’s mass and dimension attributes. Hence, these models can be used in post-harvest engineering systems. The researchers used 100 mango fruits (Mangifera indica cv. Zebdia) to determine the mathematical relationship between the fruits’ weight and dimension attributes. The researcher measured and photographed the dimensions of these fruits and processed the image captured for each fruit using a computer program to find the fruit’s dimensions. The results obtained led to the development of six mathematical models to predict a fruit’s mass from the dimensions. Given these results, the mathematical model based on the fruit’s length shows the best performance in the mass prediction (Pearson’s r=0.87). One can infer that a fruit’s mass could be obtained from its dimensions. This conclusion is not generalizable to other mango cultivars. Thus, the researcher recommends conducting further studies of other cultivars to develop a unified mathematical model. This will be helpful in developing modern post-harvest engineering systems.

Keywords: bioprocess technology, fruit sorting, image processing, physical attributes


1 Introduction

Mango (Mangifera indica L.) is a species of the Anacardiaceae family. Egypt is the one of world’s top ten mango producers (FAOSTAT, 2014). Cultivated areas of mango have been increased during the last few years with a focus on the local varieties that can generate a value to the territory. Although currently only small amounts of Egyptian mangoes are traded globally, it is representing a noticeable increase over the quantities traded on the international market recently. This increasing number of traded mangoes has placed more attention on studying the aspects of fruit quality. Additionally, Mango has several unique characteristics that set them different from any other fruits; one of these characteristics is that it ripens very quickly (Islas-Osuna et al., 2010). Therefore, the identification of correlation among these properties makes quality control and sorting procedures of the fruits more easy (Jannatizadeh et al., 2008). This is in order to satisfy consumer preferences and to reduce fruit losses during transportation and handling (Valero and Ruiz-Altisent, 2000).

The proper design of machines for fruit transportation, handling, cleaning and sorting requires a good understanding of fruit physical attributes and their relationships (Mirzaee et al., 2008). Among these attributes, length, width, thickness, volume, and mass are
the most important factors in fruit classification (Bahnasawy et al., 2004 and Mohsenin, 1986). Commonly, fruits grading is done based on their mass, size and volume. The most convenient fruit grading comparing with other grading systems is mass-sizing systems for the reason that using these systems are easier and more economical (Peleg, 1985). Therefore, the relationship between mass and geometric attributes is needed for designing and fabricating equipment for handling, transporting, processing, storage and assessing fruit quality (Khoshnam et al., 2007).

Fruit sizing by mass is recommended especially for the products with irregular shapes (Tabatabaeefar and Rajabipour, 2005). The mechanical sizing mechanism is tedious, while the electrical sizing mechanism is expensive. Hence, promoting new sizing estimation techniques with a comprehensive classification system to determine dimensional size is needed (Moreda et al., 2009). There is, therefore, a general tendency to use the image-based grading for classification of fruits with irregular shape such as mango. This technique is considered as a non-destructive method of inspection and grading of fruits that can distinguish fruit mass from its shape characteristics (Brosnan and Sun, 2002). It also provides precisely and rapidly applied sorting systems that either improve the classification procedure success or speed up the process (Cubero et al., 2011; Kleynen et al., 2003; Polder et al., 2003). Thus, mass modeling of fruit based on its dimensions is useful in grading systems that use machine vision and image processing techniques. Recently, many studies were carried out on mango fruit grading based on image processing such as (Ganiron Jr, 2014; Cavalcanti Mata et al., 2010; Chuang and Rahman, 2006; Yimyam et al., 2005 and Razak et al. 1987)

By knowing the fruit diameter or volume, its mass may be calculated using empirical equations (Sitkei, 1987). Therefore, the mass modeling of the mango fruit determines models for predicting other properties for grading especially with its irregular shape. Several studies attempt to create a correlation between mass and physical attributes for mango. Guzmán-Estrada et al., 1996, presented a new model for mass estimation of “Manila” mangoes with high coefficient of determination (R² = 0.93). Teoh and Syaifudin, 2004, developed an algorithm for mango size grading with high correlation between the measured area by image analysis and the actual weight of mango (R² = 0.934). Also, Chalidabhongse et al., 2006, proposed a vision system that can extract 2D and 3D visual properties of mango such as size (length, width, and thickness), projected area, volume, and surface area from images and use them in sorting.

No detailed studies concerning mass modeling of Zebdia mango has been performed up to now. Therefore, the objective of this research is to develop robust equations that can estimate mango mass based on its dimension attributes. These equations can then be applied to design and develop post-harvest handling and grading systems in food factories.

2 Materials and methods

2.1 Plant materials and measurements

Plant material of 100 fresh-harvested mango fruits were obtained from two growing areas depending on the land type in Egypt. The first area of fruit samples was from a farm adjacent to Ismailia city as an example of the new desert lands whereas the second was in Sharqia province as an example of old alluvial lands. Fruit samples were transferred to the Central Laboratory of Agricultural Engineering Department, Faculty of Agriculture, Ain Shams University, to be subjected to measurements analysis. Samples were selected randomly considering that they are free of injuries and damaged fruits.

For each mango fruit, five linear dimensions were measured by using a digital caliper with an accuracy of 2 digits, including length (L), maximum width (W_max), minimum width (W_min), maximum thickness (T_max) and minimum thickness (T_min). Figure 1 depicts the considerable dimensions of the fruit. Water
displacement method was used for determining the measured volume (V) of fruits. The mass of the fruit was taken using the digital balance with the accuracy of 3 digits. After that, fruit was photographed under laboratory conditions using a Canon camera (model Ixus 9515). Measurements were performed at room temperature ranged between 25ºC -30ºC.

![Mango fruit with considerable dimensions](image)

**Figure 1 Mango fruit with considerable dimensions**

### 2.2 Digital image analysis

Dimensions of each fruit obtained by digital analysis of fruits images using the model incorporated in the program Image J (http://rsb.info.nih.gov/ij/). The Image J is considered as one of the best open-source programs for digital image analysis, which is written in Java language. This program can calculate the area, measure distances and angles of the pixels through defining a specific scale of the image by users. To reach the dimensions of the fruits, the following methodology according to the software manual was considered:

1. Converting the colored images to grayscale (8-byte) one.
2. Determining the image scales by taking a reference scale of the image.
3. Measuring the required dimensions by drawing a line between the desired points.

### 2.3 Data validation and analysis

Actual dimensions were compared with dimensions obtained from the digital image analysis. This is to check the data accuracy and excludes the anomalous values in order to have reliable mathematical models that can estimate the mass of the mango fruits. The mean value, standard deviation (SD) and coefficient of variation (CV) were calculated for the data set. All data were processed initially by using Microsoft Excel 2007. Meanwhile, analyzing the data and estimating the mango mass models based on its dimension attributes were performed using the IBM SPSS 20 (SPSS20, 2014). An analysis of variance between group (ANOVA) for both mass and dimension attributes was performed. All these attributes are tested for statistically significant differences at 1% confidence level. The capability of these models to estimate the fruit mass was examined using the correlation coefficient (Pearson’s r) between the actual and estimated fruit mass. Furthermore, the accuracy of the models was evaluated by using a series of data testing methods. These include the coefficient of the determination ($R^2$), Root Mean Square Error (RMSE) and Mean Relative Deviation (MRD) according to the following equations:

\[
R^2 = \frac{\sum(M_{\text{est}}-M)^2}{\sum(M-M)^2} \tag{1}
\]

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(M-M_{\text{est}})^2}{n}} \tag{2}
\]

\[
\text{MRD} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{M-M_{\text{est}}}{M} \right| \tag{3}
\]

\[\text{CV} = \frac{\text{Standard Deviation}}{\text{Mean}}\]

Where, n is the number of samples; M and $M_{\text{est}}$ are the measured and estimated mass of Zebdia mango fruit, respectively; and $\bar{M}$ the mean mass of measured Zebdia mango fruit.
3 Results and discussion

3.1 Results

3.1.1 Dimension attributes of mango and analysis

A summary of the descriptive statistics for the dimension attributes of Zebdia mango, including the maximum and minimum values as well as mean standard deviation and CV of each attribute is shown in Table 1. According to the data in Table, the effect of all dimension attributes on the mass of mango fruit is statistically significant at 1% probability level. In general, the different dimensional attributes of the Zebdia mango fruit show low variations, resulting in low CV that is ranging from 0.08 to 0.13 for L and T respectively. While, M and V are demonstrating a higher variation with a CV value of 0.15 for both attributes. The mean values of dimensional attributes of the studied mango fruit were 127.19 (±10.02), 89.83 (±8.93), 82.74 (±8.90), 78.72 (±9.90) and 70.60 (±8.86) for L, $W_{\text{max}}$, $W_{\text{min}}$, $T_{\text{max}}$ and $T_{\text{min}}$, respectively. Meanwhile, the mean values of volume and mass were 364.79 (±54.63) and 379.23 (±56.55), respectively. Table 2 shows the analysis of variance (ANOVA) test. The significance level of all dimension attributes is less than the critical value of 0.05. Hence, the effects of dimension attributes on the fruit mass would be statistically significant.

Table 1 Dimension attributes of Zebdia mango fruits at harvest time

<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Unit</th>
<th>No. of sample</th>
<th>Max.</th>
<th>Min.</th>
<th>Mean (±SD)</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L</td>
<td>mm</td>
<td>100</td>
<td>154.46</td>
<td>107.23</td>
<td>127.19 (±10.02)</td>
<td>0.08</td>
</tr>
<tr>
<td>2</td>
<td>$W_{\text{max}}$</td>
<td>mm</td>
<td>100</td>
<td>116.97</td>
<td>72.19</td>
<td>89.83 (±8.93)</td>
<td>0.10</td>
</tr>
<tr>
<td>3</td>
<td>$W_{\text{min}}$</td>
<td>mm</td>
<td>100</td>
<td>108.59</td>
<td>61.34</td>
<td>82.74 (±8.90)</td>
<td>0.11</td>
</tr>
<tr>
<td>4</td>
<td>$T_{\text{max}}$</td>
<td>mm</td>
<td>100</td>
<td>114</td>
<td>62.04</td>
<td>78.72 (±9.90)</td>
<td>0.13</td>
</tr>
<tr>
<td>5</td>
<td>$T_{\text{min}}$</td>
<td>mm</td>
<td>100</td>
<td>99.81</td>
<td>54.15</td>
<td>70.60 (±8.86)</td>
<td>0.13</td>
</tr>
<tr>
<td>6</td>
<td>V</td>
<td>cm³</td>
<td>100</td>
<td>488</td>
<td>261</td>
<td>364.79 (±54.63)</td>
<td>0.15</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>g</td>
<td>100</td>
<td>523</td>
<td>286</td>
<td>379.23 (±56.55)</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 2 Analysis of variance (ANOVA)

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>L Between Groups</td>
<td>273944.886</td>
<td>80</td>
<td>3424.311</td>
<td>24.837</td>
</tr>
<tr>
<td>L Within Groups</td>
<td>2619.563</td>
<td>19</td>
<td>137.872</td>
<td></td>
</tr>
<tr>
<td>L Total</td>
<td>276564.449</td>
<td>99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W_{\text{max}} Between Groups</td>
<td>269827.632</td>
<td>80</td>
<td>3372.845</td>
<td>19.803</td>
</tr>
<tr>
<td>W_{\text{max}} Within Groups</td>
<td>3236.072</td>
<td>19</td>
<td>170.320</td>
<td></td>
</tr>
<tr>
<td>W_{\text{max}} Total</td>
<td>273063.704</td>
<td>99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W_{\text{min}} Between Groups</td>
<td>266432.067</td>
<td>80</td>
<td>3330.401</td>
<td>23.225</td>
</tr>
<tr>
<td>W_{\text{min}} Within Groups</td>
<td>2724.604</td>
<td>19</td>
<td>143.400</td>
<td></td>
</tr>
<tr>
<td>W_{\text{min}} Total</td>
<td>269156.671</td>
<td>99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_{\text{max}} Between Groups</td>
<td>272262.391</td>
<td>80</td>
<td>3403.280</td>
<td>25.686</td>
</tr>
<tr>
<td>T_{\text{max}} Within Groups</td>
<td>2517.439</td>
<td>19</td>
<td>132.497</td>
<td></td>
</tr>
<tr>
<td>T_{\text{max}} Total</td>
<td>274779.831</td>
<td>99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_{\text{min}} Between Groups</td>
<td>269165.482</td>
<td>80</td>
<td>3364.569</td>
<td>22.563</td>
</tr>
<tr>
<td>T_{\text{min}} Within Groups</td>
<td>2833.256</td>
<td>19</td>
<td>149.119</td>
<td></td>
</tr>
<tr>
<td>T_{\text{min}} Total</td>
<td>271998.738</td>
<td>99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V Between Groups</td>
<td>311113.688</td>
<td>80</td>
<td>3888.921</td>
<td>208.580</td>
</tr>
<tr>
<td>V Within Groups</td>
<td>354.250</td>
<td>19</td>
<td>18.645</td>
<td></td>
</tr>
<tr>
<td>V Total</td>
<td>311467.938</td>
<td>99</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2 demonstrates the frequency of measured mass and dimension attributes occurrence for Zebdia mango fruits. These histograms are intended to show the variation in distribution of each measured attribute. As shown in this figure, all histograms are somewhat bell-shaped; therefore, they can be assumed as a normal distribution. Their general shapes tend to be a right-skewed distribution with median values of $M$, $L$, $W_{\text{max}}$ and $T_{\text{max}}$ are 367 g, 126.6, 88.7 and 76.9 mm, respectively.

### 3.1.2 Fruit mass mathematical models

In order to estimate mass models, two general mathematical classifications based on single or multiple variable regression models of mango dimensional characteristics, Length ($L$), Width ($W_{\text{max}}$ and $W_{\text{min}}$) and Thickness ($T_{\text{max}}$ and $T_{\text{min}}$), were considered. The relations are as follows:

a. Linear model depends on the dimensions of the fruit:

$$M = \alpha D - \beta \quad (4)$$

b. Linear model depends on the volume of the fruit:

$$M = \alpha V + \beta \quad (5)$$

c. Complex model based on all major dimensions of the fruit:
Where, \( M \) is the fruit mass in grams; \( D \) is one of the fruit dimensions; and \( \alpha \) and \( \beta \) are empirical constants.

Table 3 presents the best-obtained models for estimating the mass of Zebdia mango fruits based on the measured dimension attributes and their empirical constants. The performance of the predictions resulting from these models by using statistically computed measures for goodness of fit (\( R^2 \), RMSE and MRD) are also shown in the same table. Ideally, \( R^2 \) values of all models are ranging between 0.84 and 0.97, which indicate that the correlation between the measured and estimated data is satisfactory. Meanwhile, it is also evident that the MRD values for all models based on dimension attributes are close to 0.045, with the exception of the mass model based on fruit volume (0.020). The RMSE indicates the absolute fit of the model. It has different values for the various models of mass estimation. Among these models, the mass model based on volume had the lowest RMSE (9.12), while the models for estimating the mass based on all dimensions and \( T_{\text{min}} \) had the highest RMSE of 22.90 and 22.87 respectively.

Regarding the model validation, Figure 3 depicts the relationship between the actual fruit mass and estimated fruit mass derived from different models. This figure encompasses a five-scatter plots representing the five models for estimating mango fruit mass. The scatter points are grouped into a clear linear shape with a positive correlation type.

Table 3 The best models estimating the mass of Zebdia mango fruit based on dimension attributes

<table>
<thead>
<tr>
<th>No.</th>
<th>Model</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( R^2 )</th>
<th>RMSE</th>
<th>MRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( M_{\text{est}} = \alpha L - \beta )</td>
<td>5.2482</td>
<td>288.29</td>
<td>0.87</td>
<td>20.79</td>
<td>0.044</td>
</tr>
<tr>
<td>2</td>
<td>( M_{\text{est}} = \alpha W_{\text{max}} - \beta )</td>
<td>5.8525</td>
<td>146.49</td>
<td>0.85</td>
<td>21.62</td>
<td>0.047</td>
</tr>
<tr>
<td>3</td>
<td>( M_{\text{est}} = \alpha W_{\text{min}} - \beta )</td>
<td>5.8281</td>
<td>103.01</td>
<td>0.84</td>
<td>22.51</td>
<td>0.049</td>
</tr>
<tr>
<td>4</td>
<td>( M_{\text{est}} = \alpha T_{\text{max}} - \beta )</td>
<td>5.2937</td>
<td>37.485</td>
<td>0.86</td>
<td>22.22</td>
<td>0.046</td>
</tr>
<tr>
<td>5</td>
<td>( M_{\text{est}} = \alpha T_{\text{min}} - \beta )</td>
<td>5.887</td>
<td>36.383</td>
<td>0.85</td>
<td>22.87</td>
<td>0.047</td>
</tr>
<tr>
<td>6</td>
<td>( M_{\text{est}} = \alpha V + \beta )</td>
<td>1.0216</td>
<td>6.5484</td>
<td>0.97</td>
<td>9.12</td>
<td>0.020</td>
</tr>
<tr>
<td>7</td>
<td>( M_{\text{est}} = \alpha \ln(L \times W_{\text{max}} \times T_{\text{max}}) - \beta )</td>
<td>17.787</td>
<td>827.85</td>
<td>0.86</td>
<td>22.90</td>
<td>0.046</td>
</tr>
</tbody>
</table>
Figure 3  Relationship between actual fruit mass (g) and estimated fruit mass (g) for zebdia mango fruit; based on different attributes.
Figure 4  Correlation between digital measured image dimensions and manually measured dimensions (mm) for the main dimensions of 100 Zebdia mango fruits

(A) L, mm, (B) \( W_{\text{max}} \), mm, (C) \( T_{\text{max}} \), mm
3.2 Discussion

All histograms give some convenient level of precision with right-skewed shape. According to these histograms, there is almost perfectly normal distribution for all the dimension attributes. It is noted that, the count of fruits with a mass greater than 400g has a low frequency (Figure 2A). As well as, the count of fruit based on length more than 140 mm, maximum width more than 100 mm and maximum thickness more than 90 mm have a low frequency as depicting in Figures 2B, C and D. This means that the majority of Zebdia mango fruit is homogeneous. This homogeneity gives the fruit a peculiarly during post harvesting as well as grading in food products processes.

The proposed models to estimate fruit mass from their dimension attributes are linear models. As seen in Table 3 the model no. 6 which is based on fruit volume is the best model for estimating the mass. This model follows the linear equations that has the highest $R^2 = 0.97$ with lowest MRD = 0.02 and RMSE = 9.12. Therefore, the volume-based model could be a general model that has a good ability for mass estimation. Similar results concerning the fruit mass estimation based on volume were reported by many researchers for different fruit species. Jahromi et al., (2007), suggested a new model for Bergamot (Citrus medica) mass estimation. The proposed model is given by the relationship between fruit mass and volume with high accuracy. The equation $M=0.52V+44.72$ describes this relationship with $R^2=0.99$. Other attempts conducted by Khoshnam et al., (2007) to estimate the mass of pomegranate using the measured volume, the equation was reported as $M=0.96V+4.20$ with $R^2$ of 0.99. In addition, Khanali et al., (2007), proposed a new mass model for tangerines as $M=0.99V-5.52$ with $R^2$ of 0.96. Even though this possibility of estimating the fruit mass for many species using their volumes with high accuracy, the mass modelling based on fruit volume is not reasonable. This is because the process remains tedious and time consuming. Thus, it is convenient adopting new models for fruit mass estimation using the dimensional measures.

The mass models based on the dimensional characteristics of Zebdia mango reveal that the highest and lowest $R^2$ are 0.86 and 0.84, respectively. The lowest mass estimation accuracy appears in the model No. 3 that obtained from a fruit minimum width. This model has the lowest value of $R^2$ (0.84) with a relatively high value of the RMSE and MRD. Therefore, the minimum width-based model is less sensitive in estimating the mass of Zebdia mango fruits. Whereas, the models No. 2 and 5 are equal in terms of $R^2$ and MRD values 0.85 and 0.047, respectively. However, the model No. 2 remains better than model No. 5 due to the relative low value of RMSE.

The performance of the model typically summarizes the discrepancy between the measured and estimated values under the model conditions (Table 3). The main goal of testing the model performance is to maintain the low values of errors and coupling them with high values of $R^2$. The results of correlation between the measured and estimated values of mass and fruit dimension attributes show a very strong positive relationship between the actual data and model output according to Pearson’s $r$ at a $P$ value $\leq 0.01$. These results of Pearson’s are significant with a strong positive correlation ($r>0.9$) for all fruit mass models. The best regression models are used to determine the fruit mass based on the dimension attributes. The measured and estimated mass values are shown in Figure 3. Clearly, the estimated mass values were very close to the actual fruit mass in all fruit mass estimating models. Nonetheless, all these models are performing better in the mass range between 280 to 400 g. Accordingly, with the data in Table 3, the model of high performance was produced by the volume with lowest RMSE and MRD.

The ultimate purpose of this research is developing new models that could estimate the mass of Zebdia mango fruit by photographing it. The application possibility of these models in predicting the mass by
recognizing the main fruit dimensional attributes of the digital image is investigated. Figure 4 shows the fruit recognition accuracy. Notably, there are slight differences between the digital measured image and manual measured for fruit dimensional attributes. Thus, the effect of measured L, W_max, and T_max demonstrate the high correlation of image-based values. This raises the ability of two-dimensional images in predicting the mass of mangoes. As well as, the two-dimensional images can reproduce a raster model of the fruit (Yimyam et al., 2005) and create three-dimensional models (Chalidabhongse et al., 2006).

4 Conclusions

Despite the difficulty of modeling the mass of mango fruit, which has a shape that standard dimension attributes cannot approximate, there is a dependable correlation between the size attributes and mass. This facilitates the description of cultivar properties; however, the fruit mass and dimension attributes provide useful information with regard to post-harvest and grading processes in food factories. In this study, the researcher presented mass modeling of Zebdia mango fruits based on their dimension attributes. The results showed that four attributes, namely L, W_max, T_max and V, are most acceptable to characterize the mass of mango fruits. From the results, one can also conclude the following:

(1) The ideal model overall for estimating the mass of Zebdia cultivar are based on the fruit’s volume (R^2=0.97).

(2) The best dimension-based model to estimate fruit mass is the one based on length, whereas the model based on the minimum width was the least-accurate model in fruit mass prediction.

(3) The two-dimensional images show a good ability to estimate the mass of Zebdia mangoes since they demonstrate a high correlation when compared to the measured dimension values.

Finally, as different mango cultivars vary considerably in their dimension attributes, the equations found in this study are not generalizable the other mango cultivars. Thus, the equations for estimating fruit mass must undergo testing using other cultivars to determine their specific coefficients.

References


