Biosystems Engineering (2002) **83** (3), 319–326
doi:10.1016/S1537–5110(02)00174-5, available online at http://www.idealibrary.com on IDEAL®
PH—Postharvest Technology

Non-linear Multiple Regression Models to estimate the Drop Damage Index of Fruit

P. Menesatti[1]; G. Paglia[1]; S. Solaini[1]; A. Zanella[2]; R. Stainer[2]; C. Costa[3]; M. Cecchetti[3]

[1] Istituto Sperimentale per la Meccanizzazione Agricola, Via della Pascolare, 16, 00016 Monterotondo, Rome, Italy; e-mail of corresponding author: pmenesatti@tin.it
[2] Centro di Sperimentazione Agraria e Forestale Laimburg (BZ), Laimburg 39040 ORA (Bz) Italy; e-mail: angelo.zanella@provinz.bz.it
[3] Università Roma Torvergata, Laboratorio di Ecologia Sperimentale e Acquacoltura, Via di Passo Lombardo 430, Rome, Italy; e-mail: corrado_costa@libero.it

(Received 18 March 2001; accepted in revised form 16 July 2002)

The aim of the study was to determine the best reliable statistic model among linear and non-linear multiple regressions, to calculate an index of the fruit bruising susceptibility estimation that is the drop damage index (DDI). The DDI is determined by a specific numerical procedure where quantity damage estimation is transformed as a probability of damage occurrence, providing a single value that represents the drop height threshold for which the probability of damage is equal to a fixed value (5%). It is based on multiple regression analysis, considering the contribution of different fruit (physical–morphological) and impact (direct drop on rigid support) variables on the average damage. A modified (segmented curve or spline) linear regression and multiple non-linear models were compared in terms of a specific statistical performance score, using a specific fruit and cultivar (apple, *Golden* clone B). The highest statistically performing model presented the logarithmic algebraic operator in base of ten ($\log_{10}$) applied to both the drop height (H) and firmness (D) variable, without the constant: $\log_{10} H + \log_{10} D$. The DDI, calculated for the best model by a numerical procedure, estimated a drop height of 15 mm related to an apple damageability of 5%. The other logarithmic models, obtained by the application of both the operator $\log_{10}$ and $\ln_e$, presented lower values of the statistical performance score, followed by the square root models and lastly by the linear spline equation. The values of the DDI predicted by the logarithmic models are considered to be more fitting to the impact reality and to cultivar-specific damage sensitivity.

1. Introduction

The analysis of impact damage of fruits is a subject of high economic value and it is widely treated in literature (Bollen & Dela Rue, 1994; Brusewitz & Bartsch, 1989; Cappellini *et al.,* 1987; Hung, 1993; Schulte *et al.*, 1991). The most studied aspects are: finding impact standard methodologies (Kuang, 1998; Menesatti *et al*, 1999; Pang *et al.*, 1996; Studman & Banks, 1989a, 1989b; Siyami *et al.*, 1988), determining bruising susceptibility or resistance properties (Banks, 1991; Dela Rue, 1996; García *et al.*, 1995; Holt & Schoorl, 1977; Hyde, 1997; Klein, 1987; Mowatt & Banks, 1994; Ouyang & Studman, 1997; Schoorl & Holt, 1980; Studman, 1997) and setting limits that must not be exceeded during handling and transportation operations (Chen & Sun, 1981; García *et al.*, 1988; Holt *et al.*, 1981; Kunze *et al.*, 1975; Roudot *et al.*, 1991; Schulte *et al.*, 1990; Zhang & Brusewitz, 1991).

In a previous paper, Menesatti and Paglia (2001), have developed an instrumental and processing method to determine the drop damage index (DDI). This parameter is a synthetic index that allows the expression of fruit damage as a probability, in order to compare different agronomic techniques or cultivars. In the cited work, the authors have calculated the DDI on the basis of multiple linear regression model (MLR) through a specific procedure of elaboration. The DDI represents the maximum drop height in mm for which the probability of damage of the cultivar is equal to 5%.

Nevertheless, the functional model that links damage at a given impact level may not necessarily be considered linear. In fact, the authors have proposed a methodology of MLR adjustment obtaining models called *spline* or *segmented curve regression* to have null damage when the drop height is zero. The extension of this activity is to develop the expression for the proposed index utilising multiple non-linear regression models.

The aim of this paper is to compare, in terms of DDI, linear and non-linear bruising susceptibility estimation models for a specific fruit and cultivar (apple, *Golden*).

2. Materials and methods

The method to obtain the DDI through MLR models was widely discussed in a previous paper of the authors (Menesatti & Paglia, 2001).

Several distinct operating phases are involved in determining the index:

(1) drop impact procedure;
(2) instrumental measurement of the damage and associated variables (impactive, morphological);
(3) statistical analysis of acquired data (linear and non-linear regression models) and
(4) elaboration of DDI by specific numerical procedure.

2.1. *Drop impact and variables instrumental method*

A fruit samples of the apple cultivar 'Golden' clone 'B' were used. It was harvested in 1998 in Val Venosta (BZ), Italy. For each fruit, the morphological, impact, maturity and damage variables were measured, as indicated in *Fig. 1*.

Drop tests were conducted with the impact equipment which allows fruits to be released and fall freely onto a rigid surface. The distance between the steel plate and the lower edge of each single suspended piece of fruit identified the drop height. This value has been randomised in a range between 20 and 400 mm.In total, 113 fruits were impacted.

2.2. *Statistical analysis*

For each fruit, the dependent variable was represented by the average damage value that was calculated by averaging the three damage measurement: the minimum diameter, the maximum diameter and the depth.

All the measured variables were considered in the statistical development of the multiple linear regression models. The linear model of maximum correlation was determined, in order to satisfy the significance condition of each parameter and the absence of colinearity between the variables, together with the significance of coefficients for the variables (probability $P < 0{\cdot}05$).

Nevertheless, only two variables were found to be higher correlated in the linear regression calculations: the drop height H and the firmness D. A similar result was found also for almost all the non-linear regression models in a screening evaluation, where H and D contributed to almost all the variance explained by the models. So, to make a simplier comparison between the different regression models, the analysis, performed with software Statistica for Windows 2.0, was limited to these two variables. To evaluate the estimation capacity to compare linear and non-linear regression models, three statistical parameters were considered:

(1) the value of the adjusted coefficient of determination R^2;
(2) the value of the standard error of estimate (SEE) and
(3) the value of the Pearson coefficient p that represents the correlation coefficient between the observed damage values and the ones esteemed by the regression models, with the firmness value set as a constant and equal to 3 kg.

To choose the most significant model to calculate the DDI, an averaging parameter was applied: the statistical performance score of the regression models statistical performance score, (SPS). It represented the average for each regression model of the standardized value of each statistical parameter (R^2, SEE, p), considering a direct contribution of the R^2 and p (higher value, higher performance) and an inverse contribution of the SEE (lower value, higher performance).

2.2.1. *Linear regression models*

For DDI calculation, an adjustment of the multiple linear regression was used. This model, called spline or segmented curve regression, avoided any possibility of obtaining positive estimated damage, with zero drop height.

The normal form of the multiple linear regression model is represented by an equation of the following type:

$$Y_2' = hH + aD + C \quad (1)$$

where: Y_1' was the estimated damage value, H the drop height in mm, h the coefficient of drop height, D the firmness, a its coefficient and C the constant.

In order to avoid any possibility of obtaining by Eqn (1) positive estimated damage, with zero drop height, a correction parameter as a function of the drop height ($c_n H$) was applied both to D and C, with the exception of H itself. This can be explained by the fact that the independent variables express their 'effect' on damage only in presence of an impact, that is for $H > 0$.