DEVELOPMENT OF NEURAL NETWORK MODEL FOR PREDICTING KIWIFRUIT INTERNAL QUALITY INDEXES

A. Addo, A. Bart-Plange and K. Dzisi
Department of Agricultural Engineering
Kwame Nkrumah University of Science & Technology
Kumasi, Ghana

ABSTRACT

Neural network models were developed to predict the internal quality indexes of kiwifruit (Actinidia deliciosa var. 'Hayward' and 'Koryoku'). Networks were developed to predict soluble solids content, titratable acidity, glucose, and fructose in kiwifruit stored at 2 and 7°C. Input data consists of storage time, package carbon dioxide and oxygen concentrations, weight loss, skin surface lightness and hue of three pieces of kiwifruit sealed with a sachet of ethylene absorbent in a low-density polyethylene film (30 mm thick). The networks were trained and tested with data collected over a period of 24 weeks at 14 day interval.. The networks were generally able to predict the variations in soluble solids, titratable acidity, fructose and glucose over time. For comparison purposes, multiple linear regression models were also estimated. They also perform well, particularly in following the trends in the data over time. Accuracy, as assessed by root mean square error in predicted value by neural network models was better than statistical regression. There was no significant difference ($p<0.05$) between the predicted neural network and statistical regression models data and experimental data for soluble solids, titratable acidity and fructose except for glucose. Measurement from the outside by means of the optical properties when used can make this modelling approach a useful tool for online sorting of fruits into various nutritional quality categories.

Keywords: Neural network, Kiwifruit, Internal quality indexes, Statistical regression

1. INTRODUCTION

Consumers purchase a product based on their perception of the quality of the product and their willingness to pay the marked price for the level of quality (Thai and Shewfelt, 1991).

Quality has been defined as "the composite of those characteristics that differentiate individual units of a product, and have significance in determining the degree of acceptability by the buyer" (Kramer and Twigg, 1970).

This statement demonstrates the importance of the perception of the buyer or consumer, especially when one considers that consumption of fruits and vegetables has increased in the past two decades (Shewfelt, 1990).

Further, the demand for high quality food products is occurring worldwide. Thus, to remain successful in a more competitive market, quality maintenance and control are essential to the fruit and vegetable sectors of the economy. To deliver optimal quality products to the consumer, we must understand, predict, and possibly control quality changes in a reliable manner. In other words, we need accurate tools for predicting quality changes (Thai, 1993).

Statistical methods such as Bayesian statistical classifier or multivariate linear regression have been used in modelling. However, these statistical methods are limited to the implementation of simple models, i.e., one in which the number of classifying parameters cannot be numerous. The easiest method deals with the problem of linearity, where the decision variable could be computed through a linear combination of the descriptive parameters (Park *et al.*, 1994).

Neural network (NN) can be viewed as a computer system that is made up of several simple and highly interconnected processing elements similar to the neuron structure found in the human brain (McClelland *et al.*, 1986). These elements process information by their dynamic state response to inputs. Problems which are normally not solvable by traditional algorithmic approaches can frequently be solved using an NN approach (Davidson and Lee, 1991).

NN can be used to solve problems in which the input and output values are known, but the relationship between the inputs and the outputs is not well understood. These conditions are commonly found in many agricultural applications (Elizando *et al.*, 1994).

Thai and Shewfelt (1991) used neural networks and statistical regression to find the mathematical relationships linking human sensory judgments to colour of tomato and peach and used a backpropagation neural network in their study. In their results, they found that the relationship to be mostly linear and thus the two methods gave approximately the same results. Park *et al.* (1994) used a backpropagation neural network with one hidden layer with different variable hidden nodes to predict the quality of beef. They used ultrasonic spectral features as input data and found that the neural network model performed better than the traditional statistical multivariate regression model.

The purpose of this study was to develop neural network and statistical regression models to predict the internal quality indexes of kiwifruit, and to compare their predictive performances.

2. MATERIALS AND METHODS

2.1 Plant materials

Kiwifruit (*Actinidia chinensis var.* 'Hayward' and 'Koryoku') were purchased from the Japan Agricultural Cooperatives in Kagawa and Ishikawa prefectures. The fruits were stored at 2°C for 24 h prior to packaging. Three individual fruits and a 10 g sachet of ethylene absorbent ($KMnO_4$) were sealed in 30 mm thick low-density polyethylene film. There was replication of each variety. The surface areas of the bags were 0.020 and 0.025 m^2 for 'Hayward' and 'Koryoku', respectively. Quality attributes of the fruit were assessed at 14 days interval during cool storage for 24 weeks. Headspace carbon dioxide and oxygen concentrations were measured. Fruits were removed from storage and assessed after being held for 1 h at 20°C, and the quality attributes of fruits measured were weight loss, skin colour, soluble solids, titratable acidity and glucose and fructose contents.

2.2 Analysis of headspace carbon dioxide and oxygen concentrations

Headspace gas composition of each package was analyzed using gas chromatography (Shimadzu Co., Model 8A, Japan) equipped with thermal conductivity detector and molecular sieve and silica gel columns were used for oxygen and carbon dioxide analyses, respectively. On each sampling day, 1 ml of gas was extracted by syringe through a silicon septum attached to the film. Helium was used as carrier gas at flow rates of 60 and 40 mL/min for carbon dioxide and oxygen, respectively. The column, injection and detector temperatures were maintained at 50°C as specified by the manufacturer.

2.3 Measurement of weight loss

Each fruit was marked and weighed with electronic balance before sealing the bags. After removal from storage, fruits were re-weighed, and the weight loss expressed as the percentage difference of the initial weight.

2.4 Measurement of external skin colour

Skin colour was measured using a colorimeter (Minolta Co., Model CR-200, Japan) with a 1-cm diameter specimen port. The instrument was standardized to a white tile of known Hunter L, a, and b values supplied by the manufacturer. Two readings were taken on opposite sides of the middle portion of the fruit and the values converted to the hue angle, q, where $q = \tan^{-1}(b/a)$ (Little, 1975).

2.5 Measurement of soluble solids and titratable acidity

Soluble solids (SS) content of juice as centrifuged for sugar analysis was determined by a temperature compensated hand-held refractometer (Atago-1E). The readings were expressed in %Brix. Titratable acidity (TA) was determined by titrating 2 mL of juice in 80 mL water with 0.1 N NaOH to an endpoint of pH 8.2.

Equation 1 was used to calculate the titratable acidity.

$$TA = \frac{V_1 \times N \times Meq \times 100}{V_2} \qquad (1)$$

where,
V_1 = volume of NaOH used (mL)
N = normality of NaOH
Meq = acid milli-equivalent factor for citric acid = 0.064
V_2 = volume of fruit juice used (mL)

2.5 Measurement of glucose and fructose contents

Pulp of each fruit was macerated in a mortar and filtered through cheesecloth. The juice was centrifuged at 3000 revolutions per minute, and an aliquot of supernatant filtered through a 0.45 mm Millipore filters. Samples (10 mL) were injected into the sampling port of a HPLC (Japan Spectrophotometer, Japan) equipped with a refractive index detector (830-RI) and a carbohydrate column (Shodex Ion Pac KC-801) maintained at 50°C. The mobile phase of the column was distilled water at a flow rate of 1 mL/min and analysis time was 12 min. Chromatographic data were collected with Borwin (JMBS Developments, Le Fontanil, France) chromatographic data system software (version 1.21). The analytes were quantitatively determined by the external standard method using peak areas.

2.6 Neural network models

Fruit quality data for the two varieties of kiwifruit collected in the laboratory were used for the model development. NeuralWorks Professional II/Plus (NeuralWare, Inc., Pittsburgh, Pa.) software was used for neural network model development. The numerical values of the selected variables were normalized between 0.1 and 0.9 to constrain the values within the linear portion of the sigmoid curve (Smith, 1996).

Equation 2 was used to normalize and denormalize the network input and target data.

$$Tar = Tmin + \frac{(Val - Val_{min})}{(Val_{max} - Val_{min})}(T\max - T\min) \qquad \mathbf{(2)}$$

where, Tar = function of the raw variable (Val)
T_{max} = 0.9
T_{min} = 0.1
Val_{max} = maximum value of dependent variable
Val_{min} = minimum value of dependent variable

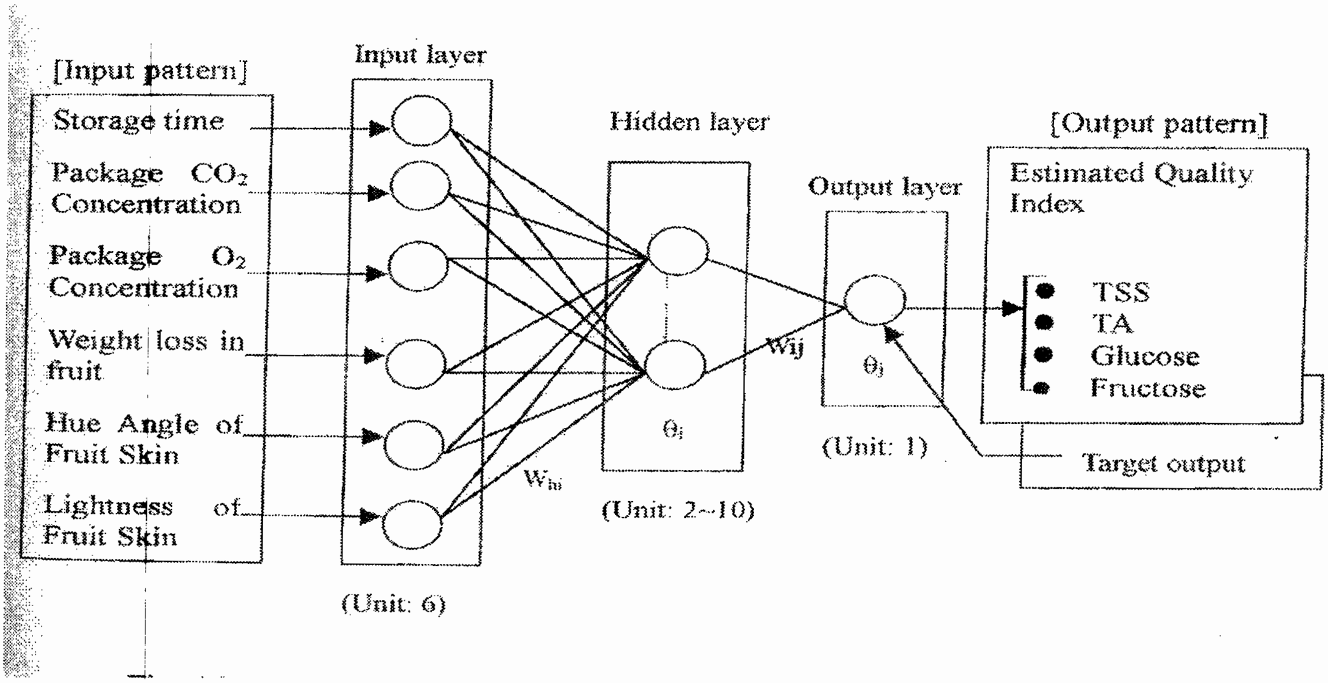
The backpropagation neural network algorithm was selected in the neural network model development because it has been used successfully for problems of supervised learning in which a continuous output value is reported (Elizando *et al.*, 1994).

A three-layer (one hidden layer) network with six inputs (storage time, headspace carbon dioxide and oxygen, concentrations, weight loss, hue and lightness) was developed. The output layer of the network comprised a single node representing total soluble solids (SS), titratable acidity (TA), glucose or fructose (Figure 1).

training was repeated until the network's performance was satisfactory, i.e., until the room mean square (RMS) error satisfied the preset requirement of 0.1 or the maximum number of iteration of 20,000 was reached.

To reduce the 'overtraining' phenomena, which cause overprediction of the models, a software ('savebest') was used. Savebest (NeuralWare, Inc., Pittsburg, Pa) is a user control program which periodically performs test commands during training and saves the network only if the test results have improved since the last test was run. The final network saved by "savebest' is one which produces the lowest error on the test data set (Park *et al.*, 1994).

Figure 1. Schematic diagram of a three layer neural network with 6 inputs.



Each input node was connected to each hidden node and each hidden node was connected to the output node.

The neural network model development consisted of training various networks in which the set of inputs were varied. Storage time, carbon dioxide and oxygen concentrations, weight loss were present in all the input combinations because of their established effects on the physiology of fruits in the modified atmosphere packaging whereby ethylene was scrubbed. Lightness and hue angle values were included in other models, because studies show that changes in internal qualities of fruits (sugar content and acidity content) are related to external colour change during maturation (Vigl, 1995). Different network parameters were used and the best selected for each simulation. The processing elements in the hidden layer were varied between 2 and 10. Seventy-two data sets were available from the experiment and 50% was used for network training and the other 50% for testing. The

2.7 Statistical regression models

In determining the regression models for calculating each quality parameter, the variables selected for the neural network model development were also used as inputs for the development of the regression model. The equations obtained from the development of the regression models were then applied to the data set used as the test set in the neural network model development (Parmer *et al.*, 1997). The statistical technique used was multivariate regression analysis from the statistical package SPSS Release 7 (SPSS Inc.). RMS errors for training and regression analysis were calculated using Microsoft Excel 97 package.

3. RESULTS AND DISCUSSION

Figure 2 shows a graph of neural network and statistical regression results for changes in SS of 'Koryoku' at 2°C when all the inputs were fed to the network during test

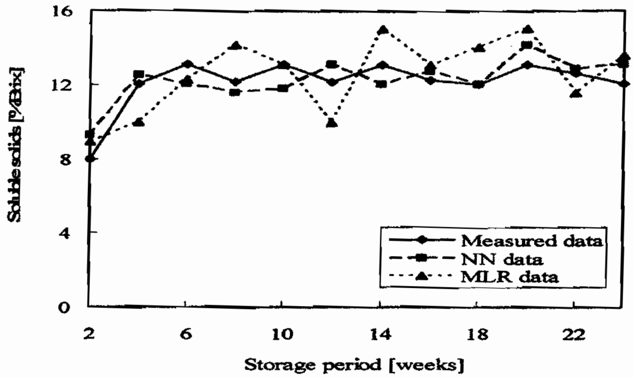


Fig. 2. Comparison of measured and estimated data by neural network (NN) and multivariate linear regression (MLR) values of total soluble solids content with 6 inputs to network and multivariate regression equation

ing. There was no significant difference at (p<5%) when neural network or statistical regression output data were tested against measured data. However, the predictive capabilities of NN and statistical regression models as determined by RMS errors show that neural network had a better predictive capability than statistical regression (Table 1). The NN and statistical regression model RMS errors were 0.175 and 0.203 respectively. Similarly, for 'Hayward' the neural network and statistical regression model RMS errors were 0.138 and 0.173 respectively. Furthermore, when 4 inputs (storage time, carbon dioxide and oxygen concentrations, and weight loss) were fed to the network NN models performed better than statistical regression models.

In the prediction model for TA values of 'Hayward' at 2° C (Figure 3), the RMS error for the neural network (0.190) was the same as that for statistical regression (0.191) when all the 6 variables were used as inputs for storage temperature of 2°C, whereas those for 'Koryoku' were 0.138 and 0.255 respectively.

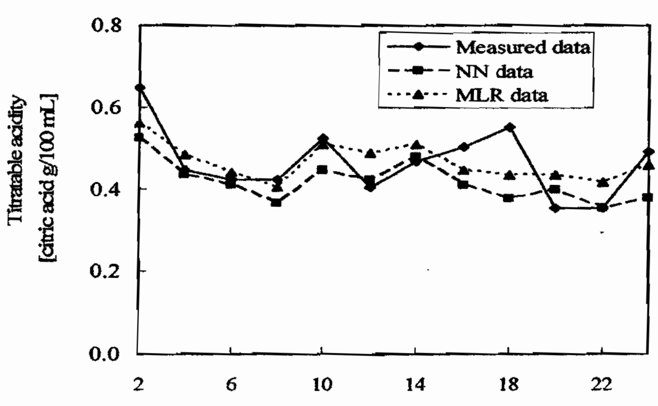


Fig. 3. Comparison of measured and estimated data by neural network and multivariate linear regression (MLR) values of titratable acidity with 6 inputs to network and multivariate regression equation.

In the case of 'Hayward' at 7°C, the RMS error for neural networks (0.164) was better than for statistical regression (0.238). When storage time, carbon dioxide and oxygen concentrations, and weight loss were fed to both models, at 2°C the neural network (0.172) prediction accuracy was better than statistical regression (0.188). Similar performance was observed in both models at 7° C. Neural network models for TA with different combinations of input variables performed better than statistical regression models. This implies that neural network performed better than statistical regression in predicting SS and TA. Statistical analysis results showed no significant difference between estimated data and measured data at $p<0.05$.

Changes in fructose and glucose values are shown in Figures 4 and 5. The predictive capabilities of NN and statistical regression models are listed in Table 2. For glucose in 'Hayward' at 2°C, the RMS error for neural network (0.157) was better than statistical regression (0.182) when the 6 inputs were fed to the network. Also, when 4 inputs were used, the neural networks (0.150) performed better than statistical regression (0.219).

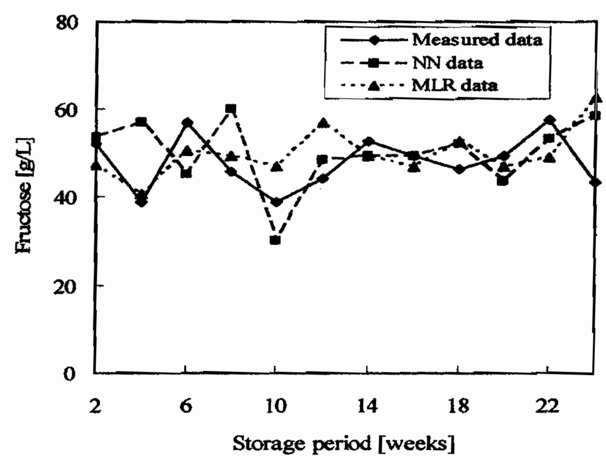


Fig. 4. Comparison of measured and estimated data by neural network (NN) and multivariate linear regression (MLR) values of fructose content with 6 inputs to network and multivariate regression equation.

For fruits stored at 7°C, the RMS errors for neural networks were 0.198 and 0.197, respectively, when 6 and 4 inputs were used. Those values were better than regression model value of 0.207 with both inputs. The results in Table 2 show that for fructose at 2°C, when the 6 inputs were used the RMS error for statistical regression was 0.171 whereas

that for neural network was 0.167. Similarly, when 4 inputs were used the statistical regression model had RMS error of 0.224 whereas neural networks value was 0.133.

Furthermore, neural networks performed better than statistical regression when models were developed with 4 inputs (0.176 for neural networks and 0.198 for statistical

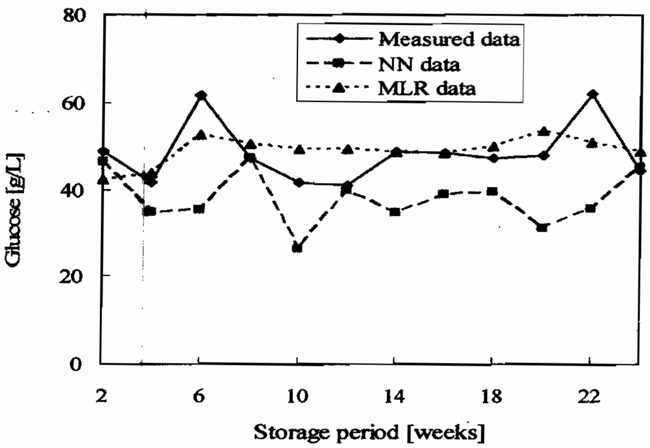


Fig. 5. Comparison of measured and estimated data by neural network (NN) and multivariate linear regression (MLR) values of glucose content with 6 inputs to network and multivariate regresson equation.

regression) and 6 inputs (0.173 for networks and 0.199 for regression) for storage at 7°C.

4. CONCLUSIONS

The results presented in this article show that:

Neural network and statistical regression models can predict nutritional quality changes in fruits, but the performance of neural network as judged by RMS error is better than statistical regression.

Data for future modelling must be expanded to include similar cultivars over different seasons and storage temperatures.

Although destructive method was used to analyse the fruits for internal quality indexes, measurement from the outside by means of the optical properties which have been proven useful in many research laboratories (Bellon *et al.* 1993; Gunasekaran *et al.* 1985; Kawano 1994), can make the modelling a useful tool for online sorting of fruits into various nutritional quality categories.

REFERENCES

Bellon, V., Vigneau, J. L. and Leclercq, M. (1993). Feasibility and Performances of a New Multiplexed, Fast and Low-cost Fiber-optic NIR Spectrometer for the Online Measurement of Sugar in Fruits. *Applied Spectroscopy* 47, 1079-1083

Davidson, C. S. and Lee, R. H. (1991). Artificial neural networks for automated agriculture. *Proceedings of the 1991 Symp. on Automated Agriculture for the 21st Century*, St. Joseph, Mich.: ASAE, pp.106-115.

Elizando, D.A., McClendon, R. W. and Hoogenboom, G. (1994). Neural network models for predicting flowering and physiological maturity of soybean. *Transactions of the ASAE* 37, 981-988.

Gunasekaran, S., Paulsen, M. R., and Shove, G. C. (1985). Optical Methods for Nondestructive Quality Evaluation of Agricultural and Biological Materials. *Journal of Agricultural Engineering Research* 32, 209-241.

Kawano S. (1994). Non-destructive NIR Quality Evaluation of Fruits and Vegetables in Japan. NIR News 5, 10-12.

Kramer, A. and Twigg, B. A. (1970). Fundamentals of Quality Control for the Food Industry. AVI Division of Van Nostrand, Inc.

Little, A. A. (1975). A research note on a tangent. *J. Food Sci.* 40, 410-411.

McClelland, J. L., Rumelhart, D. E. and the PDP Research Group. (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. I Foundations. Cambridge, Mass.: MIT Press.

Park, B., Chen, Y. R., Whittaker, A. D., Miller, R. K. and Hale D. S. (1994). Neural network modeling for beef sensory evaluation. *Transactions of the ASAE* 37, 1547-1553.

Parmer, R.S., McClendon, R. W., Hoogenboom, G., Blankenship, P. D., Cole, R. J. and Donner, J. W. (1997). Estimation of aflatoxin contamination in preharvest peanuts using neural networks. *Transactions of the ASAE* 40, 809-813.

Shewfelt, R. L. (1990). Quality of fruits and vegetables. *Food Technol.* 44, 99-106.

Smith, M. (1996). Neural networks for statistical modeling. International Thompson Computer Press, Boston, MA.

Thai, C. N. and Shewfelt, R. L. (1991). Modeling sensory color quality of tomato and peach: Neural networks and statistical regression. *Transactions of the ASAE* 34, 950-955.

Thai, C. N. (1993). "Modeling quality characteristics". In: Postharvest Handling: A Systems Approach. R.L. Shewfelt and S.E. Prussia (Eds.). Academic Press, Inc. pp. 167-185.

Vigl, J. (1995). Bargt das aussere aussehen des apfels auch für die innere qualität? *Obstbau Wienbau.* 32(11): 295-297. [English abstract in Postharvest News and Information. 7, 196-198].

Table 1 Comparison of RMS errors between neural network model and statistical regression model for total soluble solids and titratable acidity of fruit

Model Input Parameter	Temp. (°C)	Data Set	Total soluble Solids		Titratable Acidity	
			Hayward	Koryoku	Hayward	Koryoku
Storage time, CO_2, O_2, weight loss, lightness, hue	2	Training Test	0.105 (0.150) 0.138 (0.173)	0.138 (0.167) 0.175 (0.203)	0.156 (0.194) 0.190 (0.191)	0.206 (0.168) 0.138 (0.255)
	7	Training Test	0.107 (0.123) 0.138 (0.151)	0.139 (1.326) 0.133 (0.299)	0.154 (0.157) 0.164 (0.238)	0.163 (0.137) 0.161 (0.233)
Storage time, CO_2, O_2, weight loss	2	Training Test	0.109 (0.135) 0.154 (0.179)	0.139 (0.185) 0.197 (0.212)	0.159 (0.196) 0.172 (0.188)	0.188 (0.169) 0.155 (0.214)
	7	Training Test	0.119 (0.151) 0.120 (0.147)	0.125 (0.137) 0.141 (0.292)	0.154 (0.158) 0.181 (0.234)	0.183 (0.141) 0.152 (0.256)

Table 2 Comparison of RMS errors between neural network model and statistical regression model for glucose and fructose of fruit stored in low-density polyethylene bags with ethylene absorbent.

Model Input Parameter	Temp. (°C)	Data Set	Glucose		Fructose	
			Hayward	Koryoku	Hayward	Koryoku
Storage time, CO_2, O_2, weight loss, lightness, hue	2	Training Test	0.115 (0.176)* 0.157 (0.182)	0.158 (0.164) 0.199 (0.266)	0.1437 (0.112) 0.1667 (0.171)	0.198 (0.153) 0.198 (0.218)
	7	Training Test	0.146 (0.181) 0.198 (0.207)	0.170 (0.197) 0.219 (0.329)	0.131 (0.154) 0.173 (0.199)	0.179 (0.207) 0.225 (0.310)
Storage time, CO_2, O_2, weight loss	2	Training Test	0.141 (0.128) 0.150 (0.219)	0.232 (0.164) 0.193 (0.252)	0.154 (0.128) 0.133 (0.224)	0.142 (0.175) 0.177 (0.206)
	7	Training Test	0.149 (0.187) 0.197 (0.207)	0.173 (0.204) 0.218 (0.323)	0.129 (0.159) 0.176 (0.198)	0.183 (0.222) 0.225 (0.293)

*Figures in parenthesis refer to regression model values