Fruit Fresh Mass–Diameter Relationship for ‘Royal Gala’ Apple across Seasons and among Fruit Production Regions of New Zealand

H.N. De Silva¹, D.S. Tustin², W.M. Cashmore², C.J. Stanley³, G. Lupton³, and S.J. McCartney⁴

The Horticulture and Food Research Institute of New Zealand, New Zealand

Additional index words. *Malus × domestica*, fruit, mass, prediction

Abstract. A number of mass–diameter equations were compared for their potential use in indirect measurement of fruit masses of ‘Royal Gala’ apple (*Malus × domestica*). The fruit fresh-mass–diameter relationship changed with time during the season, hence no single function fitted the data well. Smooth piecewise functions that assume different relationships for intervening segments of a curve bounded by knots on the x-axis are particularly useful for modeling such data. The curve is said to be smooth because the first derivative of the function is continuous on the interval, including the knots. Two such equations, a three-parameter piecewise power function and a five-parameter spline exponential function, provided good fits to data. For both equations, the estimated mean bias on individual fruit predictions was within 5% of predicted mass over the two validating data sets. As for the precision conditional on no bias, a sample size of 20 fruit gave standard errors within 2.5% of mean predicted mass. These precisions are adequate to meet the industry requirements for monitoring fruit mass through the growing season. There was evidence of a seasonal difference in the estimated bias, but we were unable to confirm that this variation resulted from seasonal differences in fruit shape. Application of these two equations to data from other regions suggested that divergence from the estimated functional form may in fact be greater under increasingly different climatic conditions. Hence, further investigations to identify possible sources of these differences are necessary before the proposed equations can be applied across climatically different regions.

Methods for accurate prediction and forecasting of fruit crop volumes, size distribution, and quality attributes increasingly are required as tools for achieving competitive advantages for fresh-fruit marketing services. Prediction and forecasting methodology ideally should provide estimates with known precision that can be calculated using the smallest sets of easily collected, simple measurements. Consequently, an initial step in establishing forecasting systems requires an evaluation of methods for sampling fruit attributes, especially where nondestructive methods can be used for estimating attributes normally measured destructively, e.g., fruit mass. Early-season fruit growth rate may be a key parameter in models that forecast harvest distributions. In measurement of fruit growth rates, nondestructive methods provide markedly reduced variation for a given sample size. To explain further, let $\bar{x}_1$ and $\bar{x}_2$ be mean fruit masses of independent samples taken from a

tree at two successive time points. Then the variance of mean growth, $(\bar{x}_2 - \bar{x}_1)$ is: $\text{var}(\bar{x}_2 - \bar{x}_1) = \text{var}(\bar{x}_1) + \text{var}(\bar{x}_2) - 2 \text{cov}(\bar{x}_1, \bar{x}_2)$. For independent samples the covariance term is equal to zero, hence: $\text{var}(\bar{x}_2 - \bar{x}_1) = [\text{var}(X_1) + \text{var}(X_2)]/n$, where n is the sample size (equal at the two sampling times). For dependent measurements, i.e., n fruit measured (repeatedly) at times 1 and 2, the individual fruit's growth, $dX = X_2 - X_1$, has variance, $\text{var}(dX) = \text{var}(X_1) + \text{var}(X_2) - 2 \text{cov}(X_1, X_2)$. Then, $\bar{dX}$, the mean growth, has variance, $\text{var}(\bar{dX}) = \text{var}(dX)/n$. Growth being a predominantly deterministic process, the $\text{cov}(X_1, X_2)$, in the case of repeated measurements on the same fruit, will be strongly positive. Hence, the variance of mean growth from a single sample measured repeatedly $[\text{var}(\bar{dX})]$ will be much smaller than that from measurements of two independent (destructive) samples $[\text{var}(\bar{x}_2 - \bar{x}_1)]$.

In monitoring systems, fruit size can be measured nondestructively by some indirect measurements of dimensions. With apple, the maximum fruit diameter frequently is used to monitor changes in size with time, and a mass–diameter equation is used to convert diameters into corresponding fresh masses. It may be argued that growth rate can be expressed on measured scale itself (Δ diameter per unit time). However, because the mass–diameter relationship is not likely to be mathematically linear, same growth rates obtained on a diameter scale may reflect widely different rates of

increase in mass. As such, rate of increase in diameter can lead to confusion, e.g., when comparing growth rates on a given time interval between seasons. An equation describing the mass–diameter relationship would be most useful if a single function can be applied for the range of diameter measurements from fruit set to harvest. Furthermore, for general use, the equation needs to be consistent across seasons as well as among regions. The objectives of this study were to compare several mass–diameter equations for the apple cultivar ‘Royal Gala’, and, using data sets collected from three regions over several years, to select the most appropriate equation for general use and test its validity.

Materials and Methods

Source data. Three-year-old central-leader trees of ‘Royal Gala’ apple on MM.106 rootstock were located on orchards in three fruit-growing regions, Hawkes Bay (lat. 39°S), Nelson (lat. 41°S), and Canterbury (lat. 44°S), New Zealand. Twenty trees at each site were selected for monitoring fruit growth throughout the season over three consecutive seasons: 1991–92, 1992–93, and 1993–94. Each year, on a single date at midbloom, all 2-year-old spurs with the king bloom just opening were thinned to retain just the king flower, which was hand-pollinated. All other flower clusters on the trees were removed so that all fruit that set were of a similar age. A total of only 50 to 100 fruit per tree was retained to ensure that fruit growth was determined primarily by environmental conditions representative of the growing region rather than by competition among fruit. This has implications for applicability to typical growing conditions.

Thirty of the flower-bearing spurs were tagged at pollination and were used subsequently to provide fruit samples for destructive measurement throughout the duration of fruit development. One or two fruit per tree were collected every 3 to 4 d during the first 7 weeks after pollination and then about every 2 weeks until harvest. Fruit diameter at the widest equatorial point and fruit fresh and dry mass were measured from every individual fruit on each sampling date. The number of days after pollination was recorded for each sampling date.

Mass–diameter equations. The scatter plots of fruit mass vs. fruit diameter from a representative data set (Hawkes Bay, 1992–93) indicated an exponential type relationship (Fig. 1). Also, the variability in mass for a fixed diameter value appeared to increase with increasing diameter. Hence, some form of variance-stabilizing transformation was needed for fitting linear models. Ideally, the curve must pass through the origin but, if very small fruit are of no interest, this requirement may not be necessary. Several equations were considered. The power function is given by:

$$w = B_0 \times d^{B_1} \\ \log(w) = \log(B_0) + B_1 \times \log(d) \quad [1]$$

where w and d are the mass (g) and maximum diameter (cm) of fruit, respectively. It can be

Received for publication 8 Nov. 1996. Accepted for publication 31 Mar. 1997. The cost of publishing this paper was defrayed in part by the payment of page charges. Under postal regulations, this paper therefore must be hereby marked *advertisement* solely to indicate this fact.

¹Private Bag 11030, Palmerston North; to whom reprint requests should be addressed.

²Private Bag 1401, Havelock North.

³P.O. Box 220, Motueka.

⁴Private Bag 4704, Christchurch.

fitted easily as a linear model by taking the logarithm on both sides [1]. The function assumes that the mass–diameter curve passes through the origin (0,0). When a particular form of a function is not consistent across the entire domain, piecewise smoothing can be used for a better fit. The technique allows variance in function parameter values across segments bounded by the knots to obtain the best fit. The curve is said to be smooth because the first derivative of function is continuous on the interval, including the knots. A piecewise power function with a single knot at $d = d_0$ may be written as:

$$w = B_0 \times d^{B_1} \times \left(\frac{d}{d_0}\right)^{B_2} +$$

$$\log(w) = \log(B_0) + B_1 \times \log(d) + B_2 \times [\log(d) - \log(d_0)]_+ \quad [2]$$

where the “+” subscript in the equation implies the corresponding term is included only when $d > d_0$. The linear form of Eq. [2] can be fitted as a first-order spline function. The piecewise power function [2] is a three-parameter model, whereas the ordinary power function [1] has only two parameters. Eqs. [3] and [4] are exponential polynomials of third and fourth order (cubic and quartic exponential):

$$\log(w) = B_0 + B_1 \times d + B_2 \times d^2 + B_3 \times d^3 \quad [3]$$

$$\log(w) = B_0 + B_1 \times d + B_2 \times d^2 + B_3 \times d^3 + B_4 \times d^4 \quad [4]$$

that have four and five parameters, respectively. Finally, we consider a five-parameter, smooth piecewise exponential polynomial, Eq. [5], with a single knot at $d = d_0$, where d_0 is not necessarily the same value as in the piecewise power function [2]:

$$\log(w) = B_0 + B_1 \times d + B_2 \times d^2 + B_3 \times d^3 + B_4 \times (d - d_0)^3_+ \quad [5]$$

A smooth piecewise polynomial is generally referred to as a spline. Since the above equation is an exponential polynomial, we will refer to it as the spline exponential hereafter. Note that the cubic and quartic exponentials and the spline exponential permit y intercepts; i.e., they are not forced to pass through the origin (0,0). As such, they are likely to fit better for large fruit, which are of primary interest.

The Statistical Analysis System (SAS) software package (SAS Institute, Cary, N.C.) was used for equation fitting and evaluation.

Results and Discussion

Evaluation of equations. Sample sizes and timing of the sequential fruit measurements varied among the nine data sets. For purposes of evaluating the candidate equations, a data set with a wide range of diameters and an even spread of sampling times, i.e., the Hawkes Bay 1992–93 data set, was selected. The power function [1] and the two exponential poly-

nomials [3] and [4] were fitted using the standard linear regression least-squares procedure in SAS. The spline procedure (*Proc Transreg*) of SAS was used for fitting the piecewise power function [2] and the spline exponential [5]. The spline procedure was run iteratively with varying values of d_0 and the optimum value estimated using the residual sums-of-squares criterion.

The R^2 values and residual plots were used to evaluate the goodness-of-fit of the various equations (Table 1). Since in all five equations the linear predictor was $\log(w)$, the ratio of the difference in residual sums of square per degree of freedom to the error mean square of the maximal model can be used to test the goodness-of-fit of parsimonious models with a lesser number of parameters. The optimal d_0 values were found to be ≈ 1.5 and 3.0 cm for the piecewise power function and the spline exponential, respectively. The residuals (r) on the scale of linear predictor is $r = \log(w) - \log(p) = \log(w/p)$, where p is the predicted value on the original scale. The back-transformed residuals, after subtracting one $[(w/p) - 1]$, are defined here as proportional errors. The plots of fitted curves of equations (excluding the cubic exponential), with observed data from Hawkes Bay 1992–93, are given in Fig. 1. The R^2 values were very high for all five fitted equations (Table 1). Apart from the obvious lack of fit of the power function (Fig. 1), other equations provided good fit to the data. The piecewise power function is a three-parameter

Fig. 1. Different functional relationships fitted to apple fruit mass–diameter data of 1992–93 season from Hawkes Bay region.

Table 1. Estimates of parameters and their standard errors (in parentheses) for the five equations fitted to apple mass–diameter data.

Equation	Parameter					R^2
	B_0	B_1	B_2	B_3	B_4	
1. Power	1.038 (0.021)	2.418 (0.017)				0.981
2. Piecewise ² power	0.856 (0.016)	1.951 (0.029)	0.943 (0.053)			0.989
3. Cubic exponential	–3.287 (0.033)	3.423 (0.049)	–0.587 (0.017)	0.0385 (0.0016)		0.993
4. Quartic exponential	–3.571 (0.041)	4.195 (0.089)	–1.076 (0.051)	0.145 (0.011)	–0.0074 (0.0007)	0.994
5. Spline ² exponential	–3.591 (0.042)	4.219 (0.088)	–1.035 (0.046)	0.1057 (0.0066)	–0.1008 (0.0097)	0.995

²The knot is at diameter = 1.5 and 3.0 cm for piecewise power and spline exponential functions, respectively.

model in contrast to the quartic exponential and the spline exponential (Fig. 1), which are five-parameter-models. We tested the goodness-of-fit of the piecewise-power-function model against the spline exponential. The difference in residual sums of squares was significant ($180 \sim F_{2,381}$, $P < 0.01$). The quartic exponential gave a slightly lower R^2 value than the spline exponential (Table 1). In comparison with the piecewise-power function, the spline exponential appeared to fit better at the very small end of the fruit-size range (<2 cm in diameter) in the first 4 weeks after pollination. The definite advantage of the piecewise power function is that it is likely to be more robust as a predictive equation because it has fewer parameters. It is also a more biologically plausible model, since the curve passes through the origin (0,0). Based on results presented in this section we selected the piecewise power and the spline exponential models for validation with remaining data sets from the Hawkes Bay site.

Model validation. For purposes of validation, the two equations selected (with the parameter estimates obtained using the Hawkes Bay 1992–93 data) were each used to predict individual fruit masses for the Hawkes Bay data sets from 1991–92 and 1993–94. The curve of each equation is overlaid on a scatter plot of actual mass vs. diameter values (Fig. 2). Generally, there was good agreement between predicted and observed values. We define the relative bias on the i^{th} prediction as $(w_i - p_i)/p_i$, where w_i = actual fruit mass and p_i = predicted mass. Let the prediction error $e_i = \log(w_i/p_i)$, and the mean and standard deviation of e_i values are given by m_E and SD_E respectively. Then, estimates of average relative bias and precision for predictions can be expressed as:

$$\begin{aligned} \text{Bias} &= \exp(m_E) - 1 \\ PE_i &= \exp(m_E) \times SD_E \\ PE_M &= \exp(m_E) \times (SD_E/\sqrt{n}) \end{aligned}$$

where PE_i = precision for an individual prediction given in terms of proportional error, and PE_M = precision on a sample of size n . For validation of selected equations, bias is the more important issue. A significant bias would indicate that parameter values in the equations, or the functional forms describing the relationship, vary for different data sets. On the contrary, the precision can be controlled by changing the sample size. A lower precision (i.e., higher variance) would make the equations less useful in predicting an individual fruit mass based on diameter; however, by increasing the sample size, the required precision can be achieved for the predicted mean.

The observed proportional errors up to ≈ 1.5 cm in diameter were highly variable. The fruit of 1.5-cm diameter weighed only ≈ 2 g, and measurement errors on such small fruit are likely to be high relative to size measured. Hence, in order to provide reasonable estimates of bias and precision for mass predictions, we restricted the data to fruit diameters

Fig. 2. Predicted fruit-mass curves of piecewise power function and spline exponential function together with the observed values of Hawkes Bay 1991–92 and 1993–94 data.

>1.5 cm. The mean bias given as proportional error were within 5% of the predicted mass for both 1991–92 and 1993–94 data sets (Table 2). There was some evidence of a small seasonal difference with observed fruit masses being slightly higher than masses predicted by both

equations for the 1993–94 data (Fig. 2). Such seasonal differences may well arise from systematic bias on fruit diameter and mass measurements, particularly if data are collected by different individuals. Differences in fruit shape of 'Royal Gala' apple fruit between seasons

have not been reported in the literature, although fruit shape of 'Delicious' apple was found to be more elongate-conic if grown in a warm day–cool night climate compared with a hot day–warm night climate (Westwood and Burkhart, 1968). Considering that the mass predictions were based on a single measurement of fruit diameter, both equations proved to be very useful. Table 2 provides estimates of precision under varying sample sizes. Accordingly, conditional on no bias, a sample size of 20 fruit should provide a precision within 2.5% in predicting mean fruit mass.

Regional differences. The piecewise power function and the spline exponential (still with parameter estimates obtained using the Hawkes Bay 1992–93 data) were used to predict fruit masses from the Nelson and Canterbury re-

Table 2. Estimates of mean relative bias and precision obtained by using the fitted (using Hawkes Bay 1992–93 data) piecewise power and the spline exponential functions on validating data sets of the 1991–92 and 1993–94 seasons from the Hawkes Bay region.

Equation ^a	Relative bias	Precision error for varying sample sizes			
		n = 1	n = 5	n = 10	n = 20
		1991–92			
Piecewise power	0.004	0.081	0.036	0.026	0.018
Spline exponential	−0.035	0.082	0.037	0.026	0.018
		1993–94			
Piecewise power	0.048	0.110	0.049	0.035	0.025
Spline exponential	0.022	0.089	0.040	0.028	0.020

^aThe piecewise power and spline exponential functions refer to Eqs. [2] and [5] in text, respectively.

Fig. 3. Predicted fruit-mass curves of piecewise power function together with the observed values of Nelson (left) and Canterbury (right) fruit-grower regions, in 1991–92 (top), 1992–93 (middle) and 1993–94 (bottom) seasons.

gions. Since the plotted graphs of the two equations were very similar, only results of the piecewise power function are presented (Fig. 3). For Nelson, the 1991–92 fruit masses were clearly lower than masses predicted by the function (Fig. 3 and Table 3). In contrast, 1992–93 and 1993–94 mass–diameter scatter plots were in good agreement with the equation. Note that the value for mean bias for Nelson 1993–94 (Table 3) is likely to be less accurate due to fewer observations. Similar plots for Canterbury data indicated that predictions clearly were biased, but not in a consistent way among the three seasons.

The range of precision error or bias for any single year was similar for Nelson and Hawkes Bay, with precision error <2.5% and bias usually much <5% (Table 3). The climate in these two regions is quite similar. Incidence of increasing bias seemed to coincide with wider differences between regional climates, illustrated by the bias in predictions for Canterbury data sets (Table 3). For 1991–92 data, observed fruit masses were slightly lower than predicted. The reverse trend is seen with 1992–93 and 1993–94 data sets (Fig. 3). There is no consistency in the prediction bias across the three seasons. These results provide some indication that the functional form of the fruit mass–diameter relationship could differ among regions, and be affected by differences in climate. Further, the bias within a region can be in opposite directions depending on the season. These aspects need to be investigated further before the two equations described here can be applied across climatically different regions.

The fruit shape or density of 'Royal Gala' apple may differ slightly among seasons and regions leading to prediction bias in some data sets (Westwood, 1962). In this study, fruit numbers per tree were reduced to limit the effects of factors other than climatic conditions. Large differences in mean fruit mass at harvest were evident between seasons as well as among regions. Similarly, the length : diameter ratio of fruit as well as fruit : dry-matter

Table 3. Estimates of mean proportional bias of individual fruit-mass predictions from data sets derived from Nelson and Canterbury fruit-growing regions together with the standard errors of precision (in parentheses) for a sample size of 20.

Season	Nelson ^a		Canterbury ^a	
	Piecewise power	Spline exponential	Piecewise power	Spline exponential
1991–92	–0.022 (0.014)	–0.038 (0.011)	0.008 (0.019)	–0.020 (0.013)
1992–93	0.006 (0.017)	–0.010 (0.015)	0.087 (0.018)	0.068 (0.018)
1993–94	0.026 (0.015)	–0.011 (0.013)	0.073 (0.015)	0.063 (0.015)

^aThe piecewise power and spline exponential functions refer to Eqs. [2] and [5] in text, respectively.

Table 4. Changes in fresh mass, length : diameter (L : D) ratio, and percent dry matter of commercially mature 'Royal Gala' apples, when grown under nonlimiting, crop-load conditions in Hawkes Bay, Nelson, and Canterbury regions.

Region	Fresh mass (g)			L : D ratio			Dry matter (%)		
	1992	1993	1994	1992	1993	1994	1992	1993	1994
Hawkes Bay	159	181	214	0.90	0.94	0.95	13.7	15.4	14.8
Nelson	168	185	200	0.89	0.97	0.92	14.0	15.3	13.2
Canterbury	140	156	158	0.91	0.93	0.95	12.9	15.1	12.2

percentage varied among seasons and regions although in no consistent patterns (Table 4). Differences in length : diameter ratios did not fit the patterns described for 'Delicious' apple, where greater elongation was correlated with cooler growing conditions (Westwood and Burkhart, 1968). The variation in length : diameter ratio and dry-matter content of apples, however, could account for the bias in prediction of mean fruit mass from diameter measurements observed with some of the present data sets.

From this initial evaluation of suitable models for predicting fresh mass using fruit diameter measurements of 'Royal Gala' apple, two equations provide some promise. Prediction precision may be sufficient to use the piecewise power function or the spline exponential as tools with sufficient accuracy to meet industry requirements for monitoring fruit mass through the growing season in the Nelson and Hawkes Bay regions. The bias observed with data sets from Canterbury suggests that the models should not be applied across widely differing climatic regions until the causes of the deviations from the models are elucidated and the appropriate range of

regions and climates for applying the models can be established.

The fruit-mass predictions and validations were made using populations of fruit selected for uniformity with respect to age of fruit (flowers bloomed and pollinated on a single day) on trees thinned at bloom to eliminate crop-load influence on fruit growth and final fruit size. This approach was taken to establish a maximum potential precision for predicting fruit mass from diameter measurements collected throughout the growing season. For practical application in the fruit industry, such precise sampling methods probably cannot be used. Our next objective is to examine the effects on prediction precision and bias of the models when using less selective sampling methods for fruit mass and diameter collected from trees bearing economic yields.

Literature Cited

- Westwood, M.N. 1962. Seasonal changes in specific gravity and shape of apple, pear and peach fruits. *Proc. Amer. Soc. Hort. Sci.* 80:90–96.
Westwood, M.N. and D.J. Burkhart. 1968. Climate influences shape in 'Delicious'. *Amer. Fruit Grower* 88(6):26.