Mechanistic and Statistical Models to Forecast the Australian Macadamia Crop

D.G. Mayer[a], R.A. Stephenson[b], K.H. Jones[c], J.S. Yee-Yet[d], A.W. Dunstan[e], D.J.D. Bell[f], K.E. Delaney[a], K.J. Wilson[g] and J. Wilkie[h]

[a] *Queensland Beef Industry Institute, L.M.B. 4, Moorooka Qld 4105. (mayerd1@dpi.qld.gov.au)*

[b] *Queensland Horticulture Institute, S.C.M.C. Box 5083, Nambour Qld 4560.*

[c] *Australia Macadamia Society Ltd, 113 Dawson Street, Lismore NSW 2480.*

[d] *Climate Impacts and Natural Resource Systems, Queensland Department of Natural Resources and Mines, Meiers Road, Indooroopilly Qld 4068.*

[e] *Member, Australia Macadamia Society Ltd, R16/356 Blunder Road, Durack Qld 4077.*

[f] *Hidden Valley Plantations, Alf's Pinch Road, Beerwah Qld 4519.*

[g] *Gray Plantations, P.O. Box 306, Clunes NSW 2480.*

[h] *AGRIMAC Processors, 1 Northcott Crescent, Alstonville NSW 2477.*

Abstract: This project aims at producing two levels of predictions for the Australian macadamia industry. The first is overall longer-term forecasts, based on tree census data of growers in the Australian Macadamia Society (AMS). This data set currently covers about 70% of total production, and is supplemented by our best estimates of non-AMS new plantings. Given these tree numbers, average yields per tree are needed to complete the forecasts. Yields from regional variety trials were initially used, but were found to be consistently higher than average growers' yield data. Hence, a statistical model was developed using historical growers' yields, taken from the AMS database. This model allowed for the effects of tree age, variety, year, region, and tree spacing, and explained 67% of the total variation in the data. The second component of this forecasting project is an annual fine-tuning of these overall estimates, which accounts for the effects of the previous year's climate on production. This fine-tuning is based on historical yields, measured as the percentage difference between expected and actual production. The dominant climatic variables are observed temperature, rainfall and solar radiation. Water stress and waterlogging events were estimated by running a soil water-balance model, but these terms were shown to have only a minor effect. All models showed good agreement within the historical data - the jackknife cross-validation R^2 values ranged up to 97%. However, projections of the 2001 crop varied widely between models. The reasons for this are currently unclear, and exploratory multivariate analyses shed few insights.

Keywords: Crop forecast; Macadamia; Statistical model; Climate

1. INTRODUCTION

Production of macadamia nuts in Australia has been steadily increasing as new areas are planted and existing trees age, from around 4,400 tonnes in 1987 to 34,500 tonnes in 1999. However, the crop for last year (2000) was only 29,100 tonnes, indicating the high degree of year-to-year variability. With most orchards having reasonably good management and degree of pest control, this variability is generally attributed to climatic factors in the year prior to harvest. In order to facilitate future marketing and export demands, the macadamia industry needs the ability to anticipate and manage both future production increases and this inherent seasonal variability.

Two approaches have been adopted in this project. Firstly, the longer-term 'expected' yields are estimated from existing tree numbers and estimated yields. With this tree-crop having a considerable delay from planting to significant levels of production, reasonably accurate predictions out to about five years are possible [Scott, 1992]. The second research approach is to take these estimates and 'fine-tune' them, by considering the effect of the previous year's key climatic factors.

2. OVERALL MODEL

Very fortunately, the Australian Macadamia Society regularly conducts a census of its members. This has resulted in a database containing tree numbers by age, variety, planting density and location. The current production from these recorded trees is around 70% of the total crop, so 'scaling up' future projections should not introduce major errors. The only exception here is if new plantings (by newer investors in this industry, and not yet AMS members) are disproportionately represented. This appears to be the case in some regions, particularly around Bundaberg. Hence, we are incorporating best estimates of these numbers of young trees, from industry and government personnel.

Given these tree numbers, future production is then dependent on their patterns of yield increase as trees age. We initially tried using regional variety trials to estimate these relationships, but these data were considered inappropriate, as comparisons showed them to be markedly higher than yields observed on growers' properties. Fortunately, the tree census of the AMS also included historical yields (1996 to 2000), for each block of trees. To use in the regression analysis, these data were scrutinised for blocks with predominantly the same age, variety type, and planting density, resulting in 812 data points.

Plotted against age, these data displayed quite a deal of scatter [Mayer and Stephenson, 2000], but much of this was attributable to known effects. A 13-parameter multiplicative bent-stick model explained 67% of the variation, and produced interpretable fitted constants. Taking the best production region (northern NSW) as the standard, 'good' regions (other NSW sites, Glasshouse, Gympie, Bundaberg) averaged 90% of the standard yield, for any given age, variety and density. The other production regions (Atherton Tablelands, tropical Qld, other southeast Qld, WA) averaging 75%. Similarly, against the top commercial varieties, those nominated as 'medium varieties' averaged 94%, and the 'poor varieties' only 70%.

Orthogonal to these contrasts was the interaction between tree age and planting density, as illustrated in Figure 1. This shows a logical pattern – for these well-managed orchards, production begins at about the fifth year, and increases linearly until tree crowding occurs (when adjacent trees start touching and growing into each other). Naturally, this happens earlier with the higher-density plantings. There is a range of pruning and canopy management options available after this, but the individual trees have 'filled' the available area and only increase their yields by small amounts.

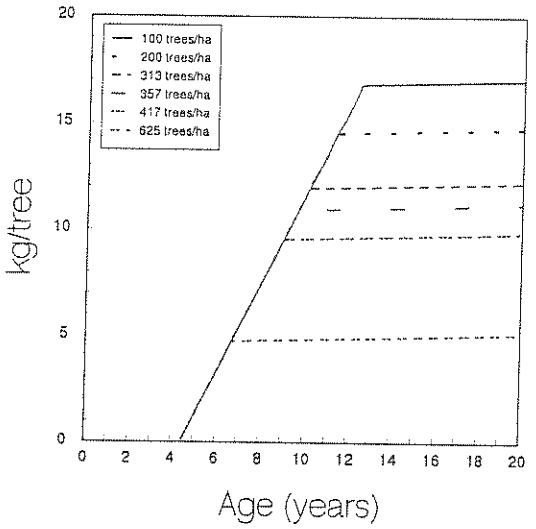

Figure 1. Average yield patterns (kg/tree), for commercial varieties in northern NSW.

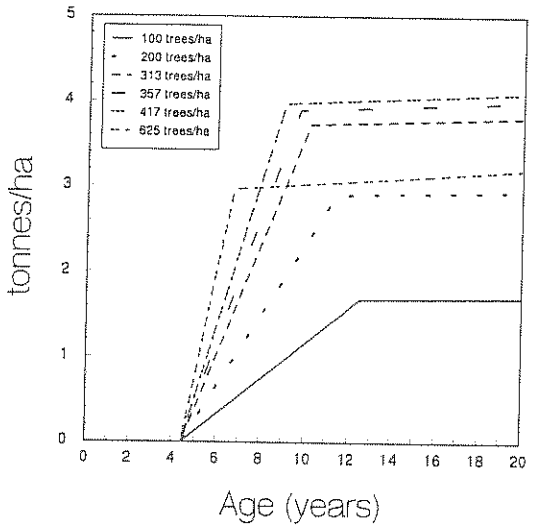

Figure 2. Average yield patterns (t/ha), for commercial varieties in northern NSW.

Of obvious interest is the integration of the yield patterns in Figure 1 with the planting densities to estimate yield per hectare, as illustrated in Figure 2. Here, the areas under each curve represent the cumulative yields over time, but are particular to these varieties in this region. Of course, a full economic analysis (incorporating the costs of purchasing, planting, and maintaining these trees over time, along with loan payments and other sundries) would be required to estimate optimal density, which would vary with differing assumptions.

In summary, the long-term predictions are simply an integration of existing tree numbers and expected future yield patterns, scaled upwards to account for non-census plantings. The predicted macadamia crop for 2001 is 36,000 tonnes.

3. ANNUAL FINE-TUNING

The hypothesis adopted here is that historical deviations about expected annual production, standardised to a percentage deviance [Mayer and Stephenson, 2000], are primarily a result of climatic effects. We can thus screen for correlations with measured climatic variables, using general linear models. This approach has previously been adopted for data from Hawaii [Liang et al., 1983] and Australia [Stephenson et al., 1986]. In these studies, temperature, rainfall and stress-days (measured via an evapo-transpiration index) proved important. Water-logging and water stress events have also been implicated in yield losses. As we have no actual data on the distribution of these events over past years, they were estimated for macadamia areas from a verified soil-water model, using best-tuned soil and plant parameters, actual climate records, and a 100-year 'burn-in' period to negate any effect of the initial soil water profiles.

For these analyses of the annual percentage deviance, the independent (X) variables screened included –

Bienniality term. Initial time-series analyses indicated the presence of a reasonable ($r = -0.52$) lag-1 autocorrelation in the model, with no evidence ($r = -0.01$) of an additional two-year effect. To allow for this possible biennial bearing pattern, the percentage difference of the previous year's crop was included. To ensure all models could be directly compared, the first observation (the 1987 crop) was excluded from all models.

Climatic and soil moisture terms. These were considered on a 'physiological year' basis, which for each year's crop is compared against data from 1st April in the previous year to the end of March in the year of that crop. Meteorological and soil water variables were taken at four locations representative of the major production areas, and then overall weighted averages used (weighted according to historical AMS production figures for these areas, being northern NSW 66%, Bundaberg 17%, Glasshouse 9% and Gympie 8%). The monthly values screened, along with seasonal and annual averages or sums, were -

- temperature (minimum, maximum and average)
- rainfall, ln(x+1) transformed
- evaporation
- solar radiation
- soil water index
- number of water-stress days per month [defined as days with less than 15% plant-available-water-capacity (PAWC), which equals field capacity minus wilting point]
- number of waterlogged days per month (days when PAWC > 95%)

Climatic indices. Field-crop modellers use a range of climatic indicators, as this has the benefit of increasing lead-time. We took these on an 'actual-year' basis, ie, the indices for a calendar year were correlated against the following year's crop. The indices used included the monthly average Southern Oscillation Index (SOI) values, and the SOI phase (five discrete levels, as used by the Agricultural Production Systems Research Unit, Queensland Centre for Climate Applications). To investigate possible longer-term effects, a lag of one year (ie, data from the previous year) was also included.

Two types of models were developed – one using actual climate data, and the other with the climatic indices. As with many climatic data sets, the (assumedly independent) X-variables were moderately to highly correlated, introducing ambiguity into the interpretation (ie, there is no single best model). Forward stepwise regression using critical graphical evaluation was conducted. Here, the residuals at each stage were plotted against the best contender X-variates, with the next being chosen by consideration of the overall pattern, rather than the contribution of only one or two influential or high-leverage points. This was done to guard against over-fitting, and to aid cross-validation. In circumstances where adjacent months or seasons contributed similarly (as measured by the fitted coefficients), these were pooled into further composite X-variables and added to the models as a single degree of freedom term. In some models, the significant factors were screened for contrasts (usually binary), based on graphical patterns. With only 13 observations and hence 12

(adjusted) total degrees of freedom, a desired minimum of 6 residual degrees of freedom was targeted to guard against over-parameterisation. In the final modelling stages, the best X-variables were tested in all-subsets regressions.

Our initial models (both with four regression, and hence eight residual, degrees of freedom) were –

(Climate data) -
Diff% = 388 – 1.05 Lag1Diff% + 7.84 Tav4 – 29.0 Tav8to11 – 1.91 Wstress2; R^2 = 83%,

where Diff% is annual difference from expected,
Lag1Diff% is from the preceding year,
Tav is average temperature,
Wstress is the number of water-stress days,
and the appended numbers represent month.

(Climatic indices) -
Diff% = -10.6 + 22.1 (if SOIph7 is 2) + 17.7 (if SOIph6prev is 3) + 5.7 (if SOIph11prev is 1) – 6.3 (if SOIph10prev is 5); R^2 = 98%,

where SOIph is SOI phase, the number represents the month, and 'prev' indicates the previous year.

These and other trialed models fitted surprisingly well. To check these relationships, we conducted jackknife cross-validations. Here, in turn each observation is left out and the model fitted to the remaining 12 years' data, and that model is then used to predict the observation that was held back. This jackknife cross-validation R^2 is thus expected to reflect the degree of fit for future predictions of unknown observations. Figure 3 shows the degree of fit across years from these cross-validations, for the best climate index and actual data models. The cross-validation R^2 values were 97 and 96% respectively. Note in particular here the agreement of the extreme 1990 percentage difference – the models' predictions (which both turned out to be very good) were effectively extrapolations, well beyond their underlying data. Figure 4 then shows how these translate to the prediction of actual production.

Despite this good agreement with the historical data, these two models gave very different predictions for this year (2001), namely –10.6% (32,200 tonnes) and +7.0% (38,500 tonnes) respectively. This discrepancy was also noted in other trialed models, and led to a wider screening of possible models, as summarised in Table 1. The wide range of predicted values here is obviously of concern. It has been suggested that this may be due to the 'climate package' of the year April 2000 to March 2001 being unlike anything observed in our historical data (which was limited to 1986 onwards).

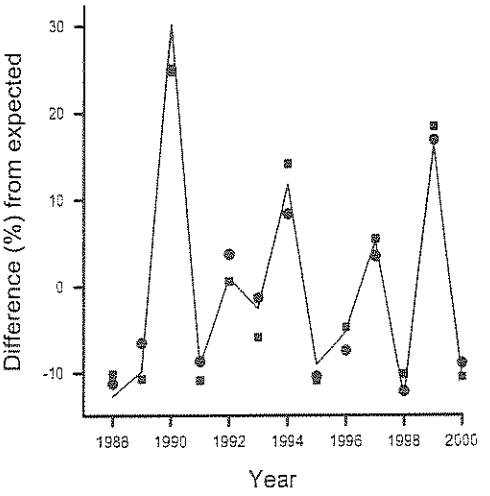

Figure 3. Cross-validation predictions of percent differences over time (line), for climatic indices model (squares), and climate data model (circles).

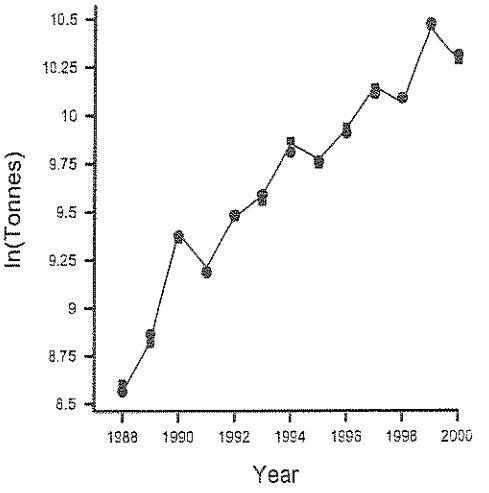

Figure 4. Cross-validation predictions of annual crop (ln-scale) over time, for climatic indices model (squares), and climate data model (circles).

Faced with these discrepancies, and as a check on some of our assumptions, we conducted a poll of 32 macadamia growers and consultants. This sample was spread geographically across growing areas in Australia. In mid-January and mid-March (just pre-harvest), they were asked to estimate the average percentage change from last year. Individual responses varied considerably, from –50% to +40%, with regional averages between –14% and +30%. Applied (by regions) to the tree census data, these percentages can be used to form a total crop prediction, which came out at 33,400 tonnes.

4. MULTIVARIATE ANALYSES OF CLIMATE PATTERNS

As an investigation of the clustering of the types of years using all climatic data, both multi-dimensional scaling (with Euclidean distances and in two dimensions) and principal components analysis were conducted. Neither was overly successful at reducing the 153 climate variables down into only two dimensions. The multi-dimensional scaling gave a residual stress of 0.39 (considerably above the target of 0.2 to 0.3), and the principal components analysis explained 38% of the variation. Results are shown in Figures 5 and 6 respectively.

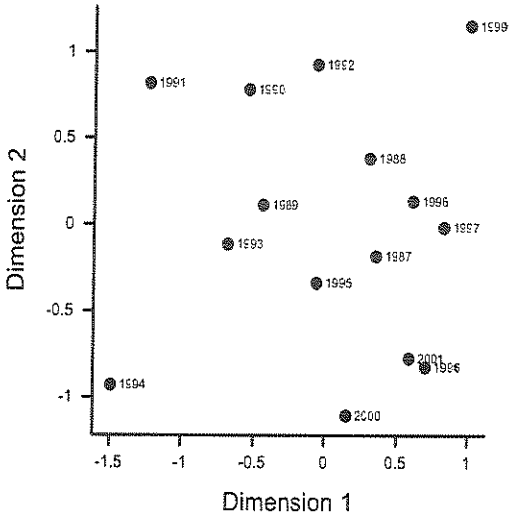

Figure 5. Multi-dimensional scaling results, using monthly, seasonal and annual climate data.

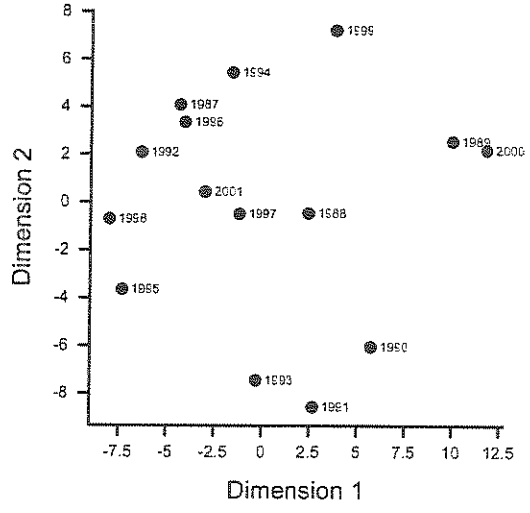

Figure 6. Principal components analysis, using monthly, seasonal and annual climate data.

We had expected that these two multivariate techniques would produce somewhat similar patterns. In Figure 5 the current year (2001) is in one corner, very close to 1996 (which had a crop deviance of -5%). However, in Figure 6 it is 'mid-cluster', with 1997 (+5%) closest, but surrounded by 1992 (+1%), 1996 (-5%), 1995 (-9%), and 1988 and 1998 (both –13%).

Faced with this uncertainty, we then conducted a discriminant analysis, where the percent deviance of each historical year (as graphed in Figure 3) was classified as positive, negative, or average (near zero). Results are graphed in Figure 7. Here, 2001 plots very close to 1992 (+1%), and it is clearly closer to the group centroid of the 'average' years, than either 'positive' or 'negative'.

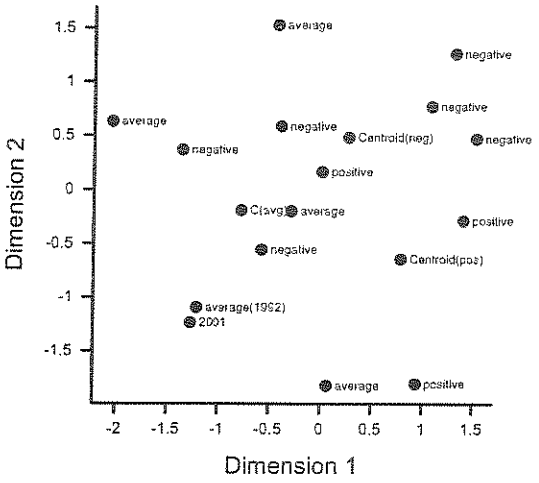

Figure 7. Discriminant analysis of historical years (classified as positive, negative or average).

5. DISCUSSION

In predicting the annual 2001 macadamia crop, we have an underlying trend which gives an overall expectation of about 36,000 tonnes. About this, different fine-tuning climate models produced widely differing deviations, in both positive and negative directions. Multivariate analyses of climatic patterns produced little consistency, and overall indicated that a near-zero deviation is probably more likely.

The median predicted deviation (from the 33 fitted statistical models of Table 1) was +5.1%, with the empirical 90% interval ranging between –11.4 and +16.1%. These convert to a median predicted crop of 37,800 tonnes (with the quite wide 90% range being 31,900 to 41,800 tonnes). This compares with a prediction of 33,400 tonnes by the growers.

Table 1. Statistical models and predictions for the 2001 crop deviation.

Model basis	Model	Terms	R^2 (%)	% dev. (pred.)
SOI	1	SOI phase (Jul), SOI phase (prev. June)	88.5	-14.53
	2a	SOIph(Jul), SOIph(prev. Jun), SOIph(prev. Nov)	95.2	-11.37
	2b	SOIph(Jul), SOIph(prev. Jun), SOIph(prev. Nov), SOIph(prev. Oct)	98.3	-10.57
Biennial-ity	1a	Lag1Diff%, Tav4, Tav8-11	63.2	5.13
	1b	Lag1Diff%, Tav4, Tav8-11, Wstress2	83.0	6.99
	1c	Lag1Diff%, Tav4, Tav8-11, Wstress2, lnRain10	94.4	5.42
	1d	Lag1Diff%, Tav4, Tav8-11, Wstress2, lnRain10, Tmin1	96.6	6.23
	2a	Lag1Diff%, TavSpr, Wlog8	59.4	5.54
	2b	Lag1Diff%, TavSpr, Wlog8, lnRainAnn	82.4	-0.68
	2c	Lag1Diff%, TavSpr, Wlog8, lnRainAnn, Evap12	92.8	-1.80
	2d	Lag1Diff%, TavSpr, Wlog8, lnRainAnn, Evap12, lnRain3	97.7	-2.84
Climate	1a	lnRain11, lnRain2&3, Tav1	83.3	13.43
	1b	lnRain11, lnRain2&3, Tav1, TminSpr	92.1	13.18
	1c	lnRain11, lnRain2&3, Tav1, Tmin8-11, Radn11	95.9	17.14
	1d	lnRain11, lnRain2&3, Tav1, Tmin8-11, Radn11, Radn6	98.6	16.13
	2a	Tav4, Tmin12, Evap8	80.6	3.60
	2b	Tav4, Tmin12, Evap8, Radn9	90.2	-5.06
	2c	Tav4, Tmin12, Evap8, Radn9, lnRain12	97.2	1.45
	2d	Tav4, Tmin12, Evap8, Radn9, lnRain12, Evap10	98.9	0.82
	3a	Tav4, Radn11, Tmin8&9	77.7	17.43
	3b	Tav4, Radn11, Tmin8&9, Wstress12	84.9	9.65
	3c	Tav4, Radn11, Tmin8&9, Wstress12, lnRain2	91.3	9.49
	3d	Tav4, Radn11, Tmin8&9, Wstress12, lnRain2, Wstress1	94.9	8.43
	3e	Tav4, Radn11, Tmin8&9, Wstress12, lnRain2, Wstress1, Evap4	97.1	4.64
All	Av.	average of (Bienniality 1b, Climate 1b, Climate 1c)	-	12.44
Biennial-ity	3b#	Lag1Diff%,Tav4,Tav8-11,Wstress2	56.7	3.15
	4a#	Lag1Diff%,Radn6&7,EvapSum,lnRain6,Swix5&6	98.4	-9.19
	4b#	Lag1Diff%,Radn6&7,EvapSum,lnRain6,Swix5&6,Radn1,SwixSum	99.8	-12.71
Climate	4b#	lnRain11,lnRain2&3,Tav1,TminSpr	86.6	5.57
	4c#	lnRain11,lnRain2&3,Tav1,Tmin8-11,Radn11	89.7	11.36
	5a#	Tav4&5,SwixSpr,TminAnn,Tmin7	95.5	1.35
	5b#	Tav4&5,SwixSpr,TminAnn,Tmin7,lnRain4,Tmin9	99.2	3.06
All	Av.#	average of (Bienniality 3b#, Climate 4b#, Climate 4c#)	-	6.69

Models using Dunoon climate data only.

On reflection, we feel that the growers are being conservative in their forecasts, or perhaps they have underestimated the magnitude of last year's decline. The overall (tree-census) expectation is around 36,000 tonnes, and 24 of the 33 statistical models predict a value larger than this.

For future years, we hope that the different climate models will perform similarly – certainly more confidence could be placed on these predictions if the range was smaller. In particular, this is likely to occur if the observed year aligns (climatically) with one of the historical years. This obviously did not happen in this past year.

6. ACKNOWLEDGMENTS

We are grateful to the Australia Macadamia Society Ltd and the Horticultural Research and Development Corporation for funding.

7. REFERENCES

Liang, T., W.P.H. Wong, and G. Uehara, Simulating and mapping agricultural land productivity: An application to macadamia nut, *Agricultural Systems,* 11, 225-253, 1983.

Mayer, D.G., and R.A. Stephenson, Macadamia crop forecasting, Proceedings Annual Conference, Australian Macadamia Society Ltd, 26-28 October 2000, Gold Coast, 27-30, 2000.

Scott, F.S., Jr., Methodology for projecting orchard crop production: A case study of macadamias, Proceedings First International Macadamia Research Conference, Kaiua-Kona, Hawaii, 28-30 July 1992, Ed. H. C. Bittenbender, 30-37, 1992.

Stephenson, R.A., B.W. Cull, and D.G. Mayer, Effects of site, climate, cultivar, flushing, and soil and leaf nutrient status on yields of macadamia in south-east Queensland, *Scientia Horticulturae,* 30, 227-235, 1986.