Available online at www.sciencedirect.com
ScienceDirect
Agricultural Systems 91 (2006) 159–170
AGRICULTURAL SYSTEMS
www.elsevier.com/locate/agsy

Annual forecasting of the Australian macadamia crop – integrating tree census data with statistical climate-adjustment models

D.G. Mayer [a,*], R.A. Stephenson [b], K.H. Jones [c], K.J. Wilson [d], D.J.D. Bell [e], J. Wilkie [f], J.L. Lovatt [g], K.E. Delaney [a]

[a] *Animal Research Institute, Department of Primary Industries and Fisheries, L.M.B. 4, Moorooka Qld 4105, Australia*
[b] *Maroochy Research Station, Department of Primary Industries and Fisheries, S.C.M.C. Box 5083, Nambour Qld 4560, Australia*
[c] *Australian Macadamia Society Ltd., 113 Dawson Street, Lismore NSW 2480, Australia*
[d] *Gray Plantations, P.O. Box 306, Clunes NSW 2480, Australia*
[e] *Hidden Valley Plantations, Alf's Pinch Road, Beerwah Qld 4519, Australia*
[f] *AGRIMAC Macadamias, 1 Northcott Crescent, Alstonville NSW 2477, Australia*
[g] *Bundaberg Research Station, Department of Primary Industries and Fisheries, 49 Ashfield Road, Kalkie Qld 4670, Australia*

Received 17 February 2005; received in revised form 19 January 2006; accepted 8 February 2006

Abstract

To facilitate marketing and export, the Australian macadamia industry requires accurate crop forecasts. Each year, two levels of crop predictions are produced for this industry. The first is an overall longer-term forecast based on tree census data of growers in the Australian Macadamia Society (AMS). This data set currently accounts for around 70% of total production, and is supplemented by our best estimates of non-AMS orchards. Given these total tree numbers, average yields per tree are needed to complete the long-term forecasts. Yields from regional variety trials were initially used, but were found to be consistently higher than the average yields that growers were obtaining. Hence, a statistical model was developed using

* Corresponding author. Tel.: +61 7 33629574; fax: +61 7 33629429.
E-mail address: david.mayer@dpi.qld.gov.au (D.G. Mayer).

0308-521X/$ - see front matter
doi:10.1016/j.agsy.2006.02.004

growers' historical yields, also taken from the AMS database. This model accounted for the effects of tree age, variety, year, region and tree spacing, and explained 65% of the total variation in the yield per tree data. The second level of crop prediction is an annual climate adjustment of these overall long-term estimates, taking into account the expected effects on production of the previous year's climate. This adjustment is based on relative historical yields, measured as the percentage deviance between expected and actual production. The dominant climatic variables are observed temperature, evaporation, solar radiation and modelled water stress. Initially, a number of alternate statistical models showed good agreement within the historical data, with jack-knife cross-validation R^2 values of 96% or better. However, forecasts varied quite widely between these alternate models. Exploratory multivariate analyses and nearest-neighbour methods were used to investigate these differences. For 2001–2003, the overall forecasts were in the right direction (when compared with the long-term expected values), but were over-estimates. In 2004 the forecast was well under the observed production, and in 2005 the revised models produced a forecast within 5.1% of the actual production. Over the first five years of forecasting, the absolute deviance for the climate-adjustment models averaged 10.1%, just outside the targeted objective of 10%.

Keywords: Crop forecast; Nearest-neighbour; Tree yield

1. Introduction

The production of macadamia nuts in Australia has been steadily increasing as new areas are planted and existing trees age, from around 4000 tonnes in 1987 to 44,000 tonnes in 2004 (Fig. 1). Notably, the crops in 2002 and 2003 were under 30,000 tonnes, reflecting the high degree of year-to-year variability. With most orchards having reasonably good management and pest control, this variability is generally attributed to climatic factors in the year prior to harvest. To facilitate future marketing and export contracts, the macadamia industry needs to anticipate and manage both future production increases and this inherent annual variability. The objective of this project was to produce forecasts for the industry's total production each year, with a target of $\pm10\%$ deviance.

Many agricultural systems are affected by multiple sources of influence. Here, statistical model-selection methods have been used to determine the relative importance of these independent influences, and to estimate their effects (Garcia-Paredes et al., 2000; Deng et al., 2005). The fitted coefficients of these statistical models can then also be used for forecasting purposes (Chatfield, 2005).

We adopted two main stages for this macadamia forecasting project. Firstly, the longer-term 'expected' yields are estimated from existing tree numbers, assumed new plantings and estimated yields. Because of the considerable delay in achieving significant levels of production after planting, reasonable predictions out to about five years are possible (Scott, 1992). Beyond this time frame, the effect of (unknown) future plantings starts to impact on the accuracy. Our second stage was to take these estimates and 'fine-tune' them each year, by considering the effect of the previous

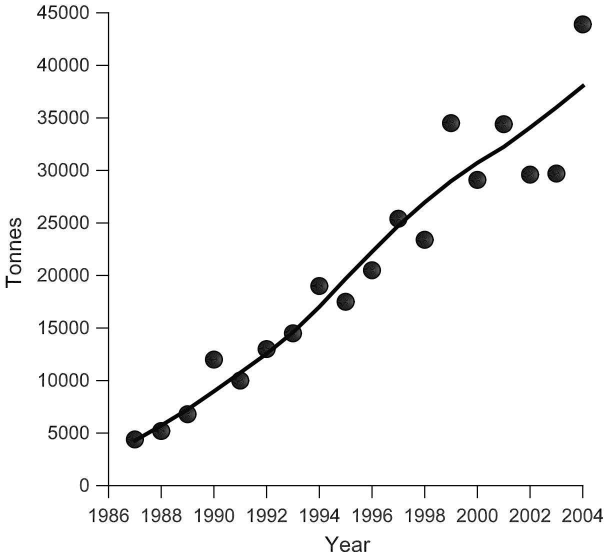

Fig. 1. Australian macadamia production – historical annual totals (dots), and long-term forecasting model (line).

year's key climatic factors using statistical models. A parallel task to this climate adjustment is a survey of the growers and industry pest scouts (professional entomologists, who regularly inspect orchards to advise on spraying requirements), to make progressive forecasts of the coming crop.

In this paper, the historical development and refinement of each of these stages is outlined in turn, along with discussions on alternate and supporting methods. This is followed by an evaluation of the relative success of the crop forecasts, for the years of 2001–2005.

2. Materials and methods

2.1. Long-term forecasting model

The Australian Macadamia Society (AMS) 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 better than 70% of the total crop. Long-term forecasts for the total crop are based on these trees, and 'scaled-up' for the unsurveyed portion of the industry. This should not introduce any major errors; the only problem here is if new plantings (by newer investors in this industry, and perhaps not yet AMS members) are disproportionately represented. This only appears to be the case in one particular region, namely Bundaberg. Overall, we are incorporating best estimates of these numbers of young trees, from both industry and government personnel.

Given tree numbers, future production is dependent on patterns of yield increases as these trees age. Initially, yield data from the regional variety trials were investigated to estimate these relationships, but comparisons (Mayer and Stephenson, 2000) showed these yields to be markedly higher than those observed on growers' properties. Fortunately, the AMS tree census also included historical yields (1996–2004), for each block of trees. These data were scrutinised for blocks with predominantly the same age, variety type and planting density, resulting in 2698 data points. These were used as the basis for estimating expected yield patterns. In summary, the long-term forecasting model is an integration of existing tree numbers and assumed future plantings with expected future yield patterns, scaled upwards to account for non-AMS orchards.

2.2. Annual adjustment for climate

The observed variation in annual yields (Fig. 1) can be standardised to an annual percentage deviance, by comparing observed production with expected (Mayer and Stephenson, 2000). Expected production for each historical year was estimated by hind-casting the scaled long-term model. The key assumption adopted for the climate adjustment is that these historical deviations are primarily a result of climatic effects. Given the short-term nature of these adjustment models, we have ignored any potential yield improvements from newer varieties or improved technologies.

For each historical year we have access to daily meteorological data including temperature, rainfall, solar radiation and evaporation. These were available as interpolated surfaces (Jeffrey et al., 2001) covering the macadamia production regions. The approximate centre-point of each region was selected, with these data then being spatially integrated by using weights proportional to each region's overall production. From these basic data, a number of physiologically-important derived variables were also calculated, namely degree-days above the lower (cold) threshold of 15 °C, degree-days above the upper (hot) threshold of 30 °C, and degree-days either side of the optimal temperature for photosynthesis, of 26 °C (Allan and De Jagar, 1979). In addition, climatic indices were considered, namely the monthly Southern Oscillation Index (SOI) and SOI phases. These have the advantage of being reasonably correlated with future climate, both in Australia (Stone and Auliciems, 1992) and in other countries (Stone et al., 1996).

Water-logging and water-stress events have also been implicated in yield losses in macadamias. As we have no actual data on the distribution of these events over past years, they were modelled for the key macadamia areas from a verified soil–water model (McKeon et al., 1990), 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. An index of water-stress is estimated by accumulating the number of days per month when the modelled plant-available-water-capacity (PAWC) is below 15%. Similarly, waterlogging is assumed to occur when PAWC is greater than 95%.

These data were all considered on a 'physiological year' basis, which for each year's crop commences on 1st April in the previous year and ends in March in the year of that crop. This cut-off date was proposed by producers in the industry, and implicitly assumes that the crop amount is 'fixed' by the end of March and unaffected by subsequent climate. Whilst there is some effect of climate on the ease and effectiveness of harvesting, the broad timing of harvesting (about April to August) indicates this effect should only be minor.

Bienniality of bearing is also known to occur in many tree-crops, including macadamias. Initial time-series analyses indicated the presence of a reasonable ($r = -0.52$; $P < 0.05$) 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 deviance of the previous year's crop was included as an independent term in the models. Since 2002, we have adopted the more conservative approach of including this as a 'positive only' effect, whereby a large crop in one year is realistically expected to decrease the crop in the following year, but the converse does not necessarily apply.

Each year, general linear models (GenStat, 2000) were initially used to screen for correlations between the relative yields and the measured climatic variables. 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 proved important. We adopted step-forward selection to build multiple models. This modelling process faced the problem of having many potential predictors, some of which were highly correlated. Hence there were usually a number of alternate models with similar degrees of fit. In selecting the next variable at each step, we took into account the degree of statistical improvement, the distribution and patterns of the partial graphs, and any macadamia physiological interpretations. Only those models with better than 60% of the variation explained (R^2) were used in the forecasting process. To guard against overfitting and to ensure adequate degrees of freedom for the residual, a maximum of five model terms was set.

In the initial years of the project, monthly climate data were used in model selection. This did cause some problems regarding the number of potential predictors, correlations amongst these, and some selection of adjacent months in different models which were probably accounting for the same climatic effect. To somewhat alleviate these problems, we investigated and then adopted a move to integrate the data into 'macadamia-physiology' periods, namely 'floral initiation' (April and May), 'winter' (June to August; note that this industry is in the Southern Hemisphere), 'flowering and nut set' (September and October), 'premature nut fall and nut growth' (November and December), and 'oil accumulation' (January and February). The month of March is now not included under this approach, as whenever monthly data were considered it rarely appeared in any of the selected models.

In 2003, it was further proposed that the Bundaberg region may behave differently to the rest of the Australian production regions, due to its drier and hotter climate and its heavy reliance on irrigation. Principal components analysis of the climate data confirmed this, with Bundaberg appearing as a relative outlier compared to a cluster containing the other regions. Fortunately, we could obtain accurate regional production data over time for the Bundaberg region, and these records confirmed this different behaviour. Hence, from 2004 the Bundaberg region was divided off and suites of climate-adjustment models were developed separately for it and for the rest of the industry. Given this region's reliance on irrigation, two extra variables were derived for the Bundaberg models, namely the percentage surface-water and ground-water irrigation allocations.

As a complementary exercise to the statistical climate-adjustment models, we also instigated a survey of key industry growers and pest scouts. They are asked to progressively supply estimates of how each year's crop compares with that of the previous year. Their replies are averaged regionally and applied to our estimated regional production breakdown, with these being summed to provide an overall industry total. The annual forecast from these sources is taken as at March, which is the same time as our climate-adjustment model forecasts are made.

2.3. Multivariate analyses of annual climate patterns

To assist with the interpretation of the forecasts from the climate adjustment model suites, the meteorological data (by macadamia-physiology periods) were subjected to principal components analysis. The dominant vectors were used to determine which historical years were 'closest' (climatically) to each year to be forecast. For the 2004 and 2005 forecasts, these alignments of annual climate patterns were also used to investigate nearest-neighbour methods, using two to eight-dimensional representation of the Euclidean distances, and a range of different weighting schemes and numbers of neighbours. These exploratory nearest-neighbour forecasts were used more as confirmation of the climate-adjustment models, rather than being taken as alternate forecasts.

3. Results

3.1. Long-term forecasting model

Plotted against age, the yield per tree data displayed quite a deal of scatter, as shown in Mayer and Stephenson (2000). Initial exploratory general linear model analyses found that much of this scatter was attributable to known effects. We have subsequently fitted a 13-parameter multiplicative bent-stick model to these data. This incorporated an interaction between age and planting density, which is multiplied by regional and varietal main effects. This model explained 65% of the total variation, and produced interpretable fitted constants. The production regions of northern New South Wales (NSW) and Bundaberg gave the highest yields per tree. The regions of Glasshouse, Gympie and other NSW areas averaged 90% of these yields, for any given age, variety and density. The other production regions (predominantly the Atherton Tablelands and the more tropical areas of Queensland) averaged 81%. Similarly, against the top-rated commercial varieties, those nominated as 'medium varieties' averaged 95%, and the 'poor varieties' 90%.

Additional to these multiplicative main effects was the key interaction between tree age and planting density, as illustrated in Fig. 2. This shows a logical pattern – for these well-managed orchards, production begins on average in the fourth year. This production increases linearly until tree crowding occurs, when adjacent trees start growing into each other, competing for light. 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 effectively 'filled' the available area and only increase their yields by smaller amounts. Some contention remains regarding the fitted asymptotic phase, with some industry personnel suggesting that yields decline slightly in aging orchards. Limited research on these patterns (McFadyen et al., 2004, 2005) show mixed results, so a flat asymptote appears appropriate.

The next step is the integration of the yield patterns in Fig. 2 with the planting densities, to estimate overall yields per hectare, as illustrated in Fig. 3. The areas under each curve represent cumulative yields over time. These patterns are used to form the long-term forecasts, which are then fine-tuned each year to account for the climatic effects.

3.2. Annual adjustment for climate

In the first year of forecasting (for the 2001 crop), the 'best' model was selected from the climate data by considering both the degree of fit and the physiological importance of the fitted independent terms. Similarly, a best-fit model was selected for the climatic indices. These two models were subjected to jack-knife cross-validation. Here each historical year is deleted in turn, and these models re-fitted to the remaining data and then used to predict the missing year. This exercise indicated that

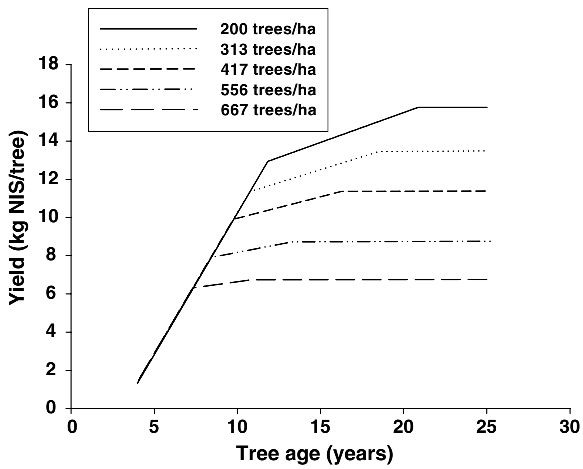

Fig. 2. Average yield patterns (kg nut-in-shell per tree) by age and planting density, for commercial varieties in regions with the highest yields.

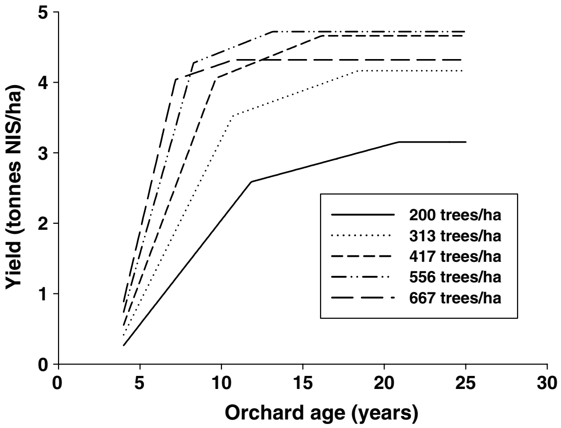

Fig. 3. Average yield patterns (tonnes nut-in-shell per ha) by age and planting density, for commercial varieties in regions with the highest yields.

a good degree of predictability could be expected (Fig. 4). The cross-validation R^2 values were both greater than 96%. Note in particular the very good agreement of the extreme 1990 percentage deviance – the models' predictions were effectively extrapolations, with both being beyond the values observed in their underlying data.

However, the respective forecasts of the 2001 macadamia crop from these two models diverged markedly, at +7.0% for the climate-adjustment model, and −10.6% for the climatic indices model. This is not uncommon in the field of

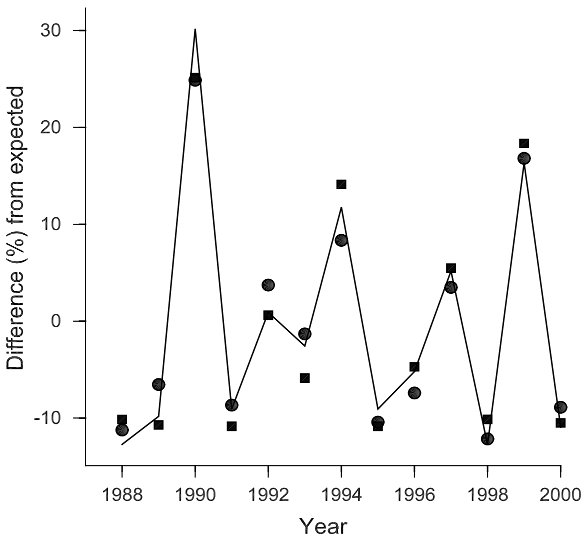

Fig. 4. Cross-validation hind-casts of annual percent deviance for the initial climatic indices model (squares), and climate data model (dots), against observed values (line).

time-series forecasting (Chatfield, 2005). To investigate possible reasons, we developed suites of step-forward models for each data type, with different initial terms and numbers of variates. Those using the climate data tended to be more consistent, mostly giving positive deviance percentages (indicating a 'good' production year), whereas the models using climatic indices varied quite widely, in both the positive and negative directions. The climatic indices models were thus discarded from further consideration. In each subsequent year, these suites of prediction models (usually 50 or more) were used to investigate the likely stability of the forecasts (in terms of common direction), the overall average, and the empirical 90% confidence intervals (namely, the 5th to 95th percentiles).

3.3. Multivariate analyses of annual climate patterns

The first two dimensions of the principal components analysis, which totalled a reasonable 42% of the total variation, are shown in Fig. 5. It is interesting to note that the first year of forecasting, namely 2001, appears amid a cluster of positive years. The model suites of that time predicted an average deviance of +5%, and actual production in this year came in at +3%. The next two years of 2002 and 2003 were somewhat remote from any other years. Hence, this representation does not appear to offer much assistance for these years, as their annual climates had not been experienced in our historical time-series. The model suite deviances for these two years both averaged −7%, against the observed values of −13% and −18%, respectively. Despite largely being extrapolations, these forecasts were still

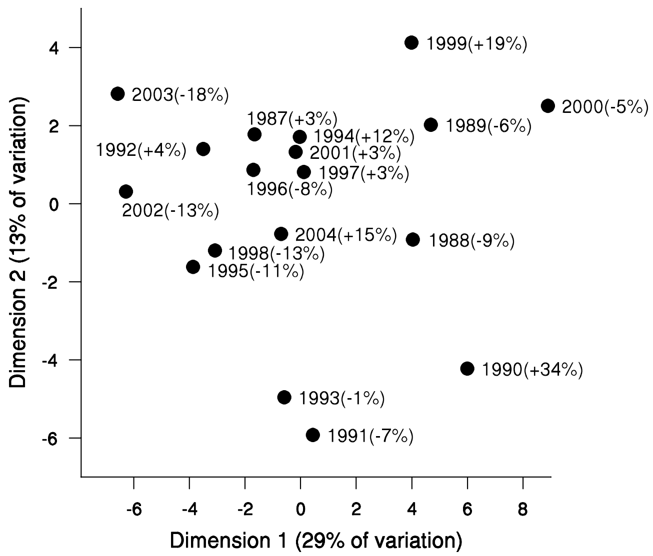

Fig. 5. Principal components representation of meteorological data for 1987–2004, with percentage deviance for each year's macadamia crop in brackets.

in the right direction. All indications for 2004 were for a negative percentage deviance – whilst being in a 'sparse' area of Fig. 5, this year is surrounded by negative years. The climate-adjustment models averaged −7.8%, and 32 out of the 36 exploratory nearest-neighbour models gave a negative forecast of the percentage deviance. However, actual production proved to be the highest on record, which in deviance terms was +15%. In 2005, the climate-adjustment models were almost all negative, in agreement with the nearest-neighbour results. The suite of these climate adjustment models averaged −15.3%, and the actual production for that year came in at −19.4%.

3.4. Overall

Fig. 6 shows the relative performance of all the forecasts made thus far. In 2001 the growers gave a conservative forecast, almost right on the long-term expectation. Actual production was higher, but not as high as the forecast from the climate-adjustment models. For the next two years both forecasts were very similar – in the right direction (i.e., less than the long-term expectation), but overly-optimistic. Given these two years of disappointing yields, the growers and pest scouts forecasts for 2004 (these are now estimated separately) were down again, and the climate forecast was somewhat above these – but none of these were realistically close to the actual record production. In 2005 the growers and the pest scouts forecasts were closer to the long-term expectation, but over-estimated actual production by 15.4% and

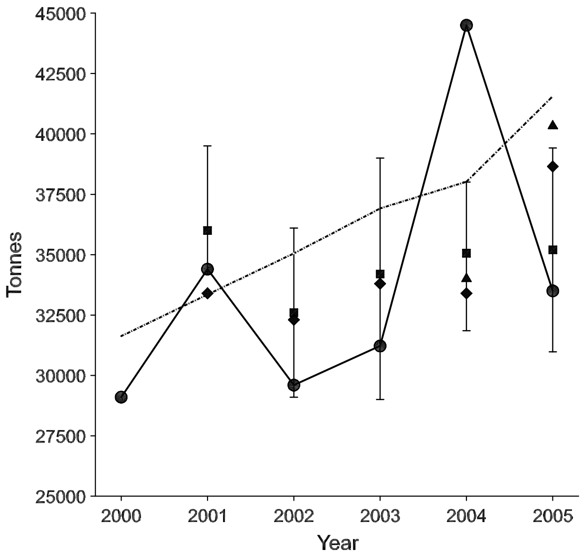

Fig. 6. Recent Australian macadamia production (dots joined by solid line), long-term model expectations (dash–dot line), climate-adjustment model forecasts (squares, with empirical 90% confidence intervals), growers' forecasts (diamonds), and pest scouts' forecasts (triangles; 2004 and 2005 only).

20.4%, respectively. The forecast from the climate-adjustment model suite was 5.1% above the actual value.

Considering the five years of forecasting (2001–2005), the absolute deviance of actual production averaged 15.7% from the long-term model forecasts. Hence, whilst this underlying increasing trend may be useful for future planning, it is not sufficiently accurate for market allocation on an annual basis. The growers' forecasts averaged 12.1% from actual, so this is not a great improvement. In 2004 and 2005, we estimated the pest scouts' forecasts separately, as it was expected that their regular close inspections of the orchards might produce more accurate forecasts. However, thus far these have actually proved to be worse than the growers' estimates. Overall, the climate-adjustment models have been the most accurate, with a five-year average absolute deviance of 10.1% – just outside the project's objective of 10%.

4. Discussion

Time-series forecasting from statistical models 'is fraught with problems and is not for the faint-hearted' (Chatfield, 2005, p. 133). These exercises often produce disappointing results (Chatfield, 2005). Macadamias are recognised as a difficult crop for research – a number of industry workshops and forums have struggled to define the key influences on production. As exemplified in McFadyen et al. (2004, 2005), even mature and well-managed orchards display varying yield patterns. Our statistical models provide evidence of the more important influences, but these have varied somewhat over the years. Obviously, as each year of this project goes by, a very valuable extra observation and degree of freedom for the models is generated – so the models should be getting more accurate. For future years, we can hope that the suite of climate adjustment models will tend to perform more similarly – certainly more confidence could be placed on these predictions if the range is smaller. In particular, this is likely to occur if the observed year aligns (climatically) with one or more of the historical years.

The growers and pest scouts forecasts have thus far been disappointing. Faced with these results, the surveyed personnel are very keen to do better. As this exercise involves comparatively little time commitment, and does keep the growers involved, it will continue into the foreseeable future.

In predicting the macadamia crop each year, we have an underlying trend which gives an overall expectation of an increase in crop size, and then a climate-adjustment to this value. In 2001–2003, the forecast direction (from the long-term expectation) was correct, but the actual forecast amounts were always optimistic. In 2004, which proved to be a year of record production, our forecast was well short. Perhaps the first three years of overestimates guided us to make more conservative assumptions at each step in the modelling process, such as minimal scale-ups for the non-AMS proportion of the industry, and the adoption of a 'positive-only' signal for bienniality of bearing. In 2005 the climate-adjustment models were the most accurate. However, it should be recognised that this was an 'easier' year to forecast – the record crop of 2004 re-set the benchmark of the potential of these trees, and we were fairly confident that the 2005 crop would be lower due to the strong influence of biennial bearing. The forecast for the 2006 crop will prove to be more challenging.

Acknowledgements

We are grateful to the Australian Macadamia Society Limited and Horticulture Australia Limited for funding, to the macadamia growers for access to their data, to J. Owens and G. Fraser for parameterisation of the soil water balance model, and to R. Evans for conducting the surveys of the growers and pest scouts.

References

Allan, P., De Jagar, J., 1979. Net photosynthesis in macadamia and papaw and the possible alleviation of heat stress. Californian Macadamia Society Yearbook 25, 150.

Chatfield, C., 2005. Time-series forecasting. Significance 2, 131–133.

Deng, X., Luo, Y., Dong, S., Yang, X., 2005. Impact of resources and technology on farm production in northwest China. Agricultural Systems 84, 155–169.

Garcia-Paredes, J.D., Olsen, K.R., Lang, J.M., 2000. Predicting corn and soybean productivity for Illinois soils. Agricultural Systems 64, 151–170.

GenStat, 2000. GenStat for Windows, Release 6.1, Sixth ed. VSN International Ltd., Oxford.

Jeffrey, S.J., Carter, J.O., Moodie, K.M., Beswick, A.R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environmental Modelling and Software 16, 309–330.

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

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

McFadyen, L.M., Morris, S.G., Oldham, M.A., Huett, D.O., Meyers, N.M., Wood, J., McConchie, C.A., 2004. The relationship between orchard crowding, light interception, and productivity in macadamia. Australian Journal of Agricultural Research 55, 1029–1038.

McFadyen, L.M., Morris, S.G., McConchie, C.A., Oldham, M.A., 2005. Effect of hedging and tree removal on productivity of crowding macadamia orchards. Australian Journal of Experimental Agriculture 45, 725–730.

McKeon, G.M., Day, K.A., Howden, S.M., Mott, J.J., Orr, D.M., Scattini, W.J., Weston, E.J., 1990. Northern Australian savannas: management for pastoral production. Journal of Biogeography 17, 355–372.

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

Stephenson, R.A., Cull, B.W., Mayer, D.G., 1986. 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.

Stone, R., Auliciems, A., 1992. SOI phase relationships with rainfall in eastern Australia. International Journal of Climatology 12, 625–636.

Stone, R.C., Hammer, G.L., Marcussen, T., 1996. Prediction of global rainfall probabilities using phases of the Southern Oscillation Index. Nature 384, 252–255.