Final Report

Better Macadamia Crop Forecasting Part 2

Dr Russ Stephenson
The Department of Agriculture and Fisheries (DAF)

Project Number: MC09016

MC09016

This project has been funded by Horticulture Innovation Australia Limited using the macadamia industry levy and funds from the Australian Government.

ISBN 0 7341 3549 1

Published and distributed by:
Horticulture Innovation Australia Limited
Level 8, 1 Chifley Square
Sydney NSW 2000
Tel: (02) 8295 2300
Fax: (02) 8295 2399

Contents

Summary

This project ('Better Macadamia Forecasting Part 2') ran from August 2009 until April 2015. It followed on from the earlier macadamia forecasting projects, which commenced in 2000. This report integrates and covers methods and results for the overall project, rather than focusing specifically on 'Part 2'.

Accurate crop forecasts for the Australian macadamia industry are required each year, to facilitate planning and marketing. This project has produced crop predictions for the whole industry since 2001, and more recently for each of the six separate production regions. These forecasts have been forwarded to the Australian Macadamia Society (AMS) in March each year, forecasting that year's crop (which is then harvested throughout the remainder of the year).

There are two complementary levels of forecasting.

Firstly, the overall longer-term forecast is based on tree census data of growers in the AMS, scaled up to include the non-AMS orchards. Expected yields are based on historical data, with a nonlinear regression model incorporating tree age, variety, year, region and tree spacing. This long-term model forecasts expected production for six to 10 years into the future.

The second level of crop prediction is an annual climate-based adjustment of these overall long-term estimates, taking into account the expected effects of the previous year's climate on production. The dominant climatic variables are observed temperature, rainfall and solar radiation, and modelled water stress. Based on the proven forecasting success of boosted regression trees and 'random forests', the average forecast from an ensemble of regression models is adopted (rather than using a single best-fit model). Exploratory multivariate analyses and nearest-neighbour methods are also used to investigate the patterns in the data.

In parallel, a survey of growers and pest scouts is also conducted early each year, with their estimates of the coming crop being integrated into regional and then overall totals.

Real-world problems, including flooding rain during harvest and the destructive winds of ex-tropical cyclone Oswald, have obviously affected the accuracy of the forecasts that are made early each year. There were also major problems between 2008 and 2011, when industry yields in some regions were well below the levels previously achieved, and this was attributed mainly to management problems caused by lower prices.

These forecasting methods have been evolving over the past decade, and the recent years have shown average absolute error rates of 6.8% for the growers forecast, and 8.6% for the climate forecasts. These are within the targeted ±10%, and also compare well with other crop forecasting applications around the world.

The resources required to continue the growers' forecasts are quite minimal, and it is recommended that these continue. The long-term forecasts are based on the now-somewhat-dated census data of the AMS, and these (or valid alternatives) would need to be updated and revised for this aspect to continue.

Keywords

macadamia; forecast; model; climate; yield; regression; ensemble; statistics

Introduction

As new areas are planted and existing trees age, the production of macadamia nuts in Australia has been generally increasing, from around 30,000 tonnes nut-in-shell at 10% moisture content (NIS) per annum at the start of this century to over 40,000 tonnes in recent years. However, this overall trend is distorted by a high degree of year-to-year variability, with crops ranging from 28,500 tonnes in 2011 to almost 44,000 tonnes in 2004, 2006 and 2014. 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 efficient handling and processing demands, and to plan for future marketing and export contracts, the macadamia industry needs to anticipate and manage both future production increases and this inherent annual variability.

Agricultural production systems can be affected by many 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).

Two main stages were utilised for this macadamia forecasting project.

Firstly, the longer-term ‘expected’ yields were estimated from existing tree numbers, estimated yields and assumed new plantings. Because of the considerable delay in achieving significant levels of production after planting, reasonable predictions out to about six years are possible (Scott, 1992). Beyond this time frame, the effect of (unknown) future plantings starts to impact on the accuracy of forecasts.

Our second stage was to take these estimates and ‘fine-tune’ them each year, by considering the effects of the climate during the 13 months prior to harvest. As a parallel to this second stage, crop estimates made by surveyed growers and pest scouts were integrated into an alternate annual forecast, which we term the growers forecast.

The objective of this project was to produce forecasts for the industry’s total production each year, with a target of ±10% deviance.

Methodology and Results

An informal steering committee for the project was formed, involving a range of industry personnel and experts. Membership of this group evolved somewhat over the years as alternate expertise was incorporated. It met annually, to discuss the forecasts along with the project's methodologies and developments, and to consider possible alternate approaches and improvements.

The AMS has a good working relationship with the Australian macadamia processors. Each year the processors provide to the AMS their confidential data of intakes (tonnes NIS at 10% moisture) from the farms, by defined production regions (Fig. 1). Inland NSW covers the major production area around and north from Lismore, extending east to Alstonville. These intake data form the basis for all forecasts. For each region, these actual production amounts are then standardised to an annual percentage deviance by comparing them with expected production (Mayer and Stephenson, 2000). The expected production for each historical year is derived by formulating a long-term model and hind-casting this.

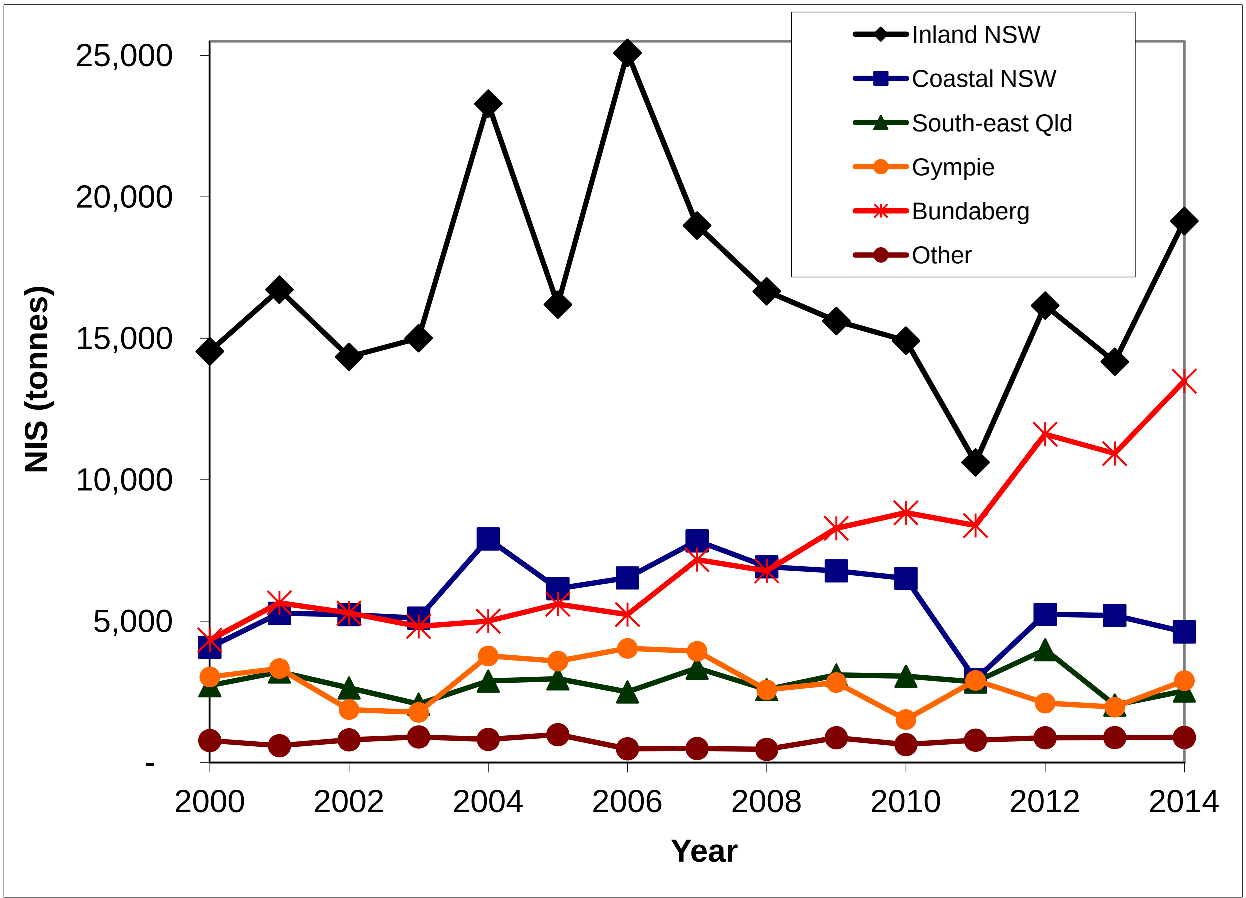

Figure 1. Annual macadamia production by regions.

Long-term model

The long-term model, or expected production for each region in each year, integrates tree numbers with expected yield-per-tree, with both parts being based on data from the now somewhat-dated 2010 AMS survey of producers. The yield-per-tree model (as detailed in Mayer et al. 2006) incorporates the effects of region, variety, and the interaction between age and planting density. The actual tree yields currently utilised to estimate this multiplicative equation are from 2007 to 2010, as these data were felt by the project steering committee to be ‘more representative of the current yields in the industry’.

Subsequent to the initial development of the yield-per-tree models of Mayer et al. (2006), ‘orchard decline’ has become evident amongst the older orchards in the industry. Data from the AMS ‘Benchmark report – The Australian macadamia industry 2009 to 2012 seasons’ (Project MC09001) was used to estimate this effect. This benchmark surveyed 250 farms, covering approximately 55% of both the area planted and production of the industry. Considering both the cross-years averages for tree yields by age groups and the individual-years data, we have adopted a 2.5% decline per year from age 20, plateauing at a 20% decline after age 27. Under this assumption, our model has the ‘25+’ year class (nominally taken as 25 to 35 inclusive) averaging 13% lower yields than the 20-24 year old group. This approximately agrees with the benchmark report, where the cross-years average for this decline (for their ‘25+’ trees *vs*. the 20-24 year old group) was 12%, with the individual years being 7%, 11%, 14% and 12% for 2009 to 2012 respectively.

For the long-term model, new plantings are added in with the existing tree numbers from 2010. These assumed new plantings, as listed in Table 1, have been based on known sales from nurseries, anecdotal information and the opinions of key industry experts. Unfortunately we have no data on tree removals, and it is known that some producers have been removing older less-productive blocks, and mostly replacing these with newer varieties. Our assumptions regarding new plantings are thus ‘nett gains’ in tree numbers; in particular none are assumed for two regions.

Table 1. Assumed nett new plantings (trees/year) by regions.

Year	Inland NSW	Coastal NSW	SE Qld	Gympie	Bundaberg	Other	Total
2010	2,000	3,000	0	0	39,000	14,000	58,000
2011	2,000	3,000	0	0	25,000	4,000	34,000
2012	2,000	3,000	0	0	20,000	7,900	32,900
2013	2,000	3,000	0	0	25,000	10,000	40,000
2014	2,000	3,000	0	0	25,000	10,000	40,000
2015	2,000	3,000	0	0	25,000	10,000	40,000
2016+	2,000	3,000	0	0	25,000	10,000	40,000

The expected production amounts for each year are then calculated for each region, both forecasts and hindcasts. The latter are scaled-up to the actual amounts (see Fig. 2) to account for the farms that were not included in the AMS census. This scale-up factor is similarly applied to the forecasts, to obtain ‘whole-of-industry’ totals. The revised long-term forecasts are for 46,000 tonnes this year (2015), rising to an expectation of just over 50,000 tonnes in 2020.

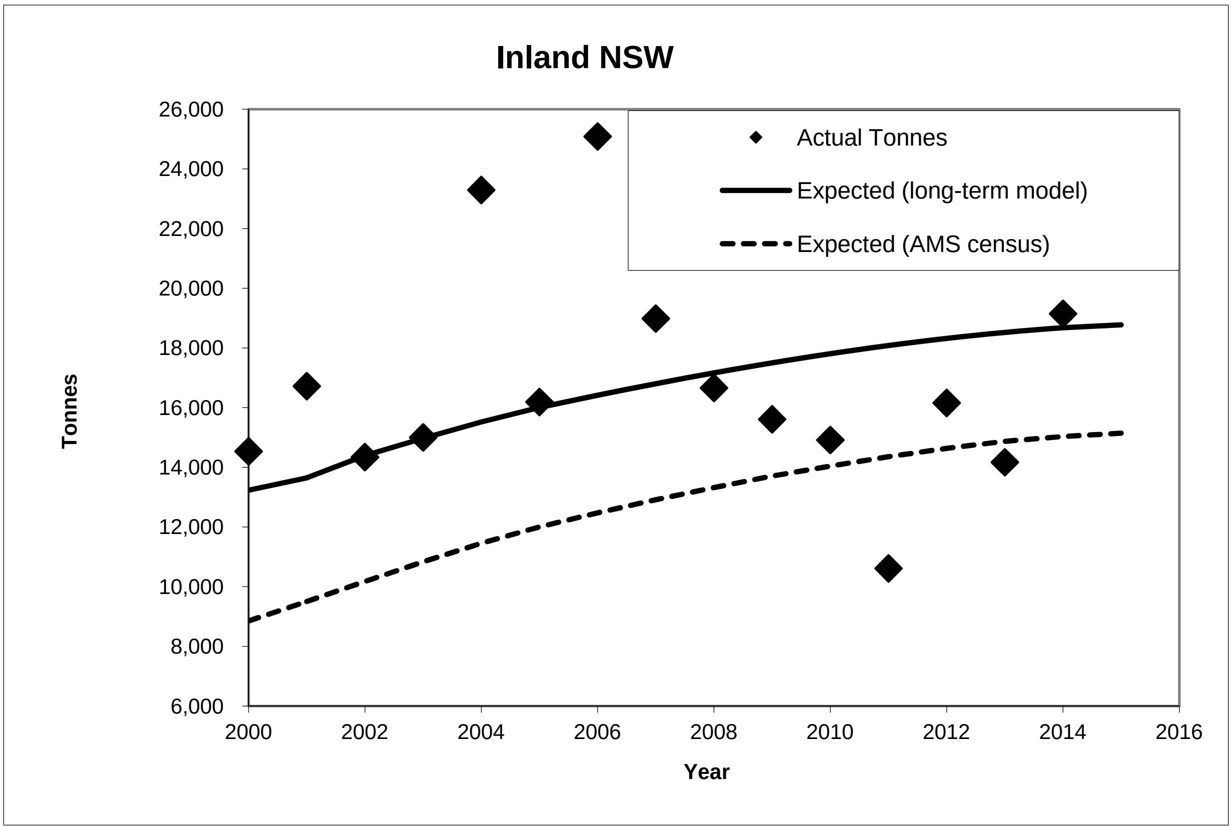

Figure 2. Actual annual production (tonnes Nut In Shell or NIS, at 10% moisture) for the 'inland NSW' region, and expected production (from the AMS-census trees, and overall).

Climate-adjusted model

This process uses an ensemble of regression models to adjust the long-term or expected annual production for the effects of the previous year's climate, plus some additional non-climatic influences. The dependent (Y) variable is the percentage deviance (positive or negative) from the expected production for each historical year.

Monthly meteorological data are extracted on a regional basis (Jeffrey et al., 2001), from the approximate centroid for each defined region – Hinkler Park for the Bundaberg region, Woolvi for the Gympie region, Beerwah for south-east Queensland, and Newrybar for coastal NSW. For inland NSW, we take the average from Alstonville, Clunes, Dunoon and Lismore. No meteorological data (nor climate adjustment) is used for the minor 'other' region, as it ranges widely from tropical Queensland to Western Australia.

The key variables used for the regression ensembles include maximum and minimum temperatures, monthly rainfall ($\log_{10}$ transformed), adjusted monthly rainfall (capped at monthly evaporation), pan evaporation rate, solar radiation, and cumulative day-degrees either side of 26° C (the optimal temperature for photosynthesis in macadamias; Allan and De Jagar, 1979). In addition, the monthly averages for some modelled climate indices are included. These are calculated from a calibrated soil-water-balance model (McKeon et al. 1990), based on 'an average' macadamia orchard (in terms of soil type and depth, and tree age and planting density). A number of agronomic indices were also considered over the years of this project, and following discussion with industry experts, we adopted the average monthly transpiration-efficiency index, the number of water-stress days per month (days with plant-available-water-capacity <15%), and the soil-water-index. Each year these data are scrutinized for each region, looking for extreme periods and possible stressors. Figure 3 shows how the soil-water levels in south-east Queensland were notably low for the latter part (October to December) of 2014 and this is expected to have a negative influence on the 2015 crop.

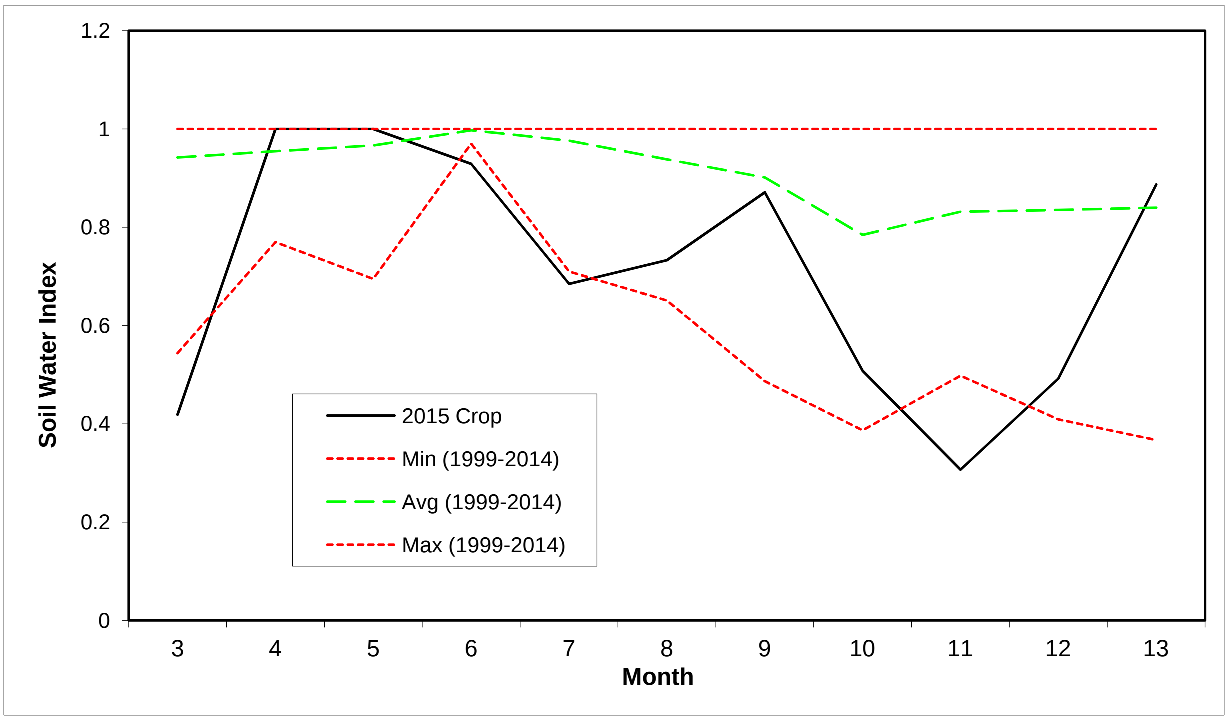

Figure 3. Modelled soil-water index levels for the south-east Queensland region, for historical years and the 2015 crop. Month 3 is the previous March, month 12 the previous December, and month 13 is January of the actual production year.

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 (Dormann et al. 2013), 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 integrating data into key macadamia physiological periods. These are 'last summer' (the previous January), 'floral initiation' (April and May of the previous year), 'winter' (June to August), 'flowering/nut set' (September and October), 'premature nut fall' (November), 'nut growth' (December), and 'oil accumulation' (January of the current year).

For each region, the important 'non-climatic' effects are also screened, namely the biennial-bearing effect (where a large crop suppresses the crop of the following year, and *vice-versa*), and CPI-adjusted nut prices (direct, plus lagged by one or two years; this being a proxy for a 'management intensity' effect). Of the price variables, lag-2 consistently had the best degree of fit, being significant for the regions of inland NSW, coastal NSW and Gympie. Notably, no price variable had an effect for the regions of south-east Queensland or Bundaberg, with this being a consistent result for the ensemble models over the past few years. Current prices then had no additional effect for any region, possibly because these are reasonably-correlated with prices two years ago. Hence lag-2 price effectively captured the ‘price signal’. The biennial-bearing term was generally significant for inland NSW for all the years of this project, and recently has also become significant for Bundaberg.

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 held during this project have struggled to define the key influences on production. As exemplified in McFadyen et al. (2004, 2005, 2013), 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 and between regions. Linear regression models have previously been used to screen for correlations between yields and meteorological effects, for data from Hawaii (Liang et al., 1983) and Australia (Stephenson et al., 1986). In these studies, temperature, rainfall and stress-days proved important.

Ensemble regression methods are based on the relatively new boosted regression techniques (Elith et al. 2008, Hastie et al. 2009), or ‘random forests’, which come from the data-mining and machine-learning sciences. These methods develop self-tuning ensembles of regression tree models, and these have recently shown improved predictive behaviour in a number of areas (Hoerl et al. 2014). Song et al. (2013) shows how the successful ‘multiple models’ concept of these tree-based regressions can effectively be extended into the general linear modelling (GLM) framework. In climate forecasting, the average of multiple models has repeatedly been shown to outperform any of the individual models.

These ensemble predictors ‘are known to lead to highly accurate predictions’ (Song et al. 2013). We are yet to investigate and incorporate the more complex operational parameters of these techniques available in the R language (R Core Team 2013), however our climate-adjustment models have used a baseline implementation of GLM ensembles in GenStat (VSN 2014). For each region, around 30 to 50 step-forward multiple regression models are formed, each with different combinations of ‘the best’ climate terms (and 2nd and 3rd best) at each step in the model-building process. A maximum of four climate terms was imposed for each model, to prevent over-fitting. The models in these ensembles usually agree reasonably well, for example 32 of the 35 forecasts for the 2015 crop in Bundaberg were negative (a lower crop than expected from the long-term model). The overall average forecast from the ensemble of models is adopted for each region.

To assist with the interpretation of the forecasts from the climate adjustment model ensembles, the meteorological data are also subjected to principal components analysis. The dominant two vectors are used to determine which historical years are ‘closest’ to the climate for year being forecast, as shown in Fig. 4 for the south-east Queensland region. The mid-regions of this graph contain generally-positive years, with two poorer (very-negative) years in the lower-right area. Here, the overall climate for the crop of 2015 is in a ‘quite sparse’ region, indicating that the climate pattern (of January 2014 to January 2015) was quite different to those experienced in previous years. These annual climate patterns were also used to investigate nearest-neighbour methods, using a two to eight-dimensional representation of the Euclidean distances, and a range of different weighting schemes and numbers of neighbours. However, these exploratory nearest-neighbour forecasts have proven to be disappointing (Mayer and Stephenson 2008), and are now used more as confirmation of the climate-adjustment models.

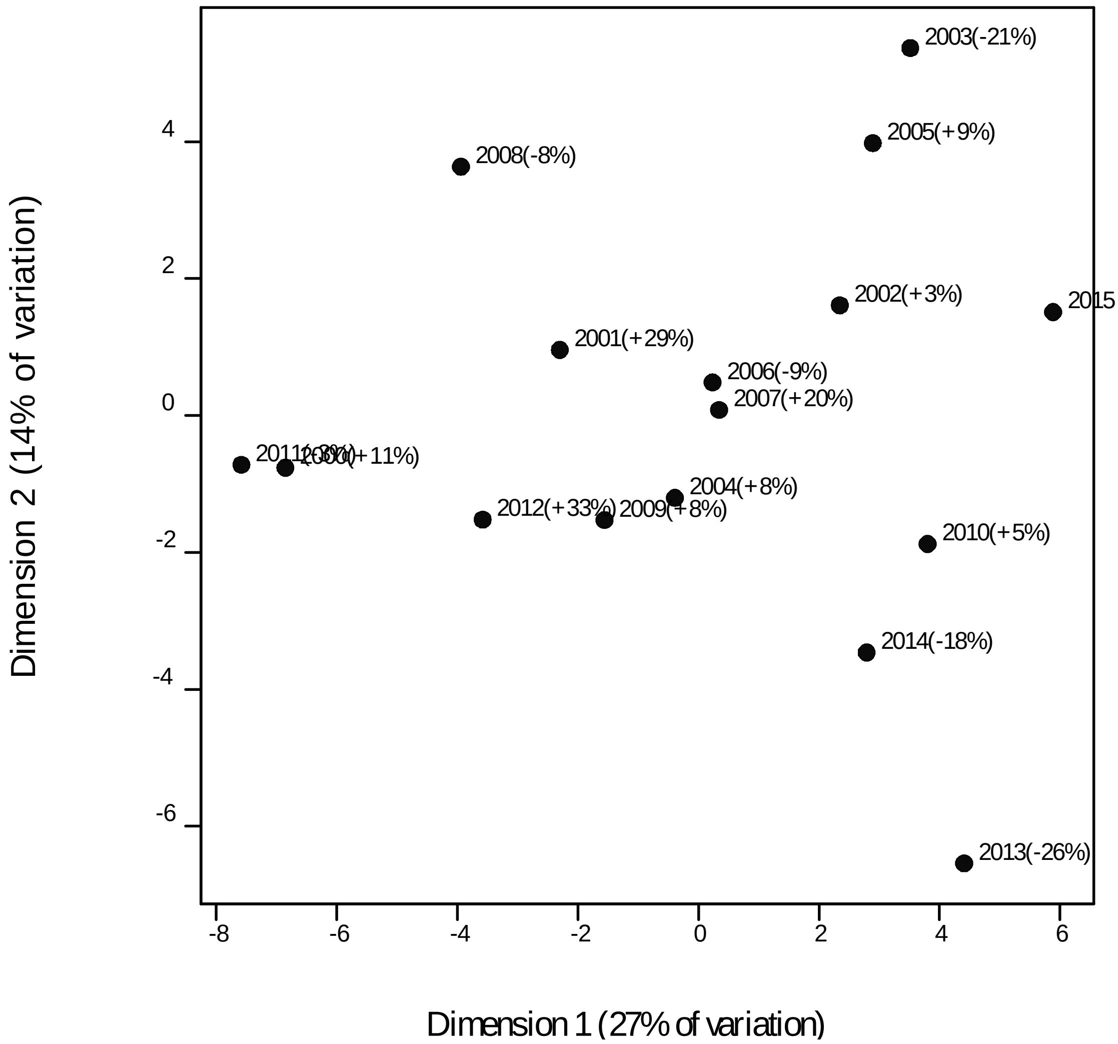

Figure 4. Principal-components representation of the overall climatic effects for the south-east Queensland region for each year, with the percentage crop sizes in brackets.

Growers forecasts

As a complementary and parallel exercise to the statistical climate-adjustment models, an annual survey of key industry growers and pest scouts was also instigated. This form, with an example reply (actual name omitted), is included as the Appendix to this report. Between 2004 and 2010 the replies from the pest-scouts were tabulated separately. However as these showed no real advantage over the growers, and as the separated replies were sometimes ‘quite sparse’ for some regions, these two groups were subsequently pooled.

Early each year the growers and pest-scouts are surveyed to supply estimates of how that year’s crop compares with the crop of the previous year. Their replies are averaged regionally and applied to our estimated regional production breakdown, with these forecasts for each region then 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 the climate-adjustment model forecasts are made.

Outputs

The long-term model produces forecasts of the expected production amounts (by regions) for the next six years. Table 2 lists the final values from this project, resulting from the retuning of this model to include the actual 2014 crop. The first four regions are ‘approaching maturity’, where the increases from the new plantings and younger trees are expected to be approximately counter-balanced by orchard decline in the older farms. There are however continuing expected increases in the remaining two regions (Bundaberg and other), from the higher proportions of younger trees which are expected to increase yields as the age.

Table 2. Long-term forecasts (tonnes NIS at 10% moisture) for the Australian macadamia crop, by production regions.

Year	Inland NSW	Coastal NSW	SE Qld	Gympie	Bundaberg	Other	Total
2015	18,800	6,000	3,100	2,900	14,300	900	46,000
2016	18,889	6,103	3,168	2,913	14,990	940	47,002
2017	18,931	6,181	3,229	2,928	15,642	985	47,896
2018	18,948	6,241	3,286	2,945	16,326	1,035	48,780
2019	18,933	6,280	3,327	2,963	16,975	1,095	49,572
2020	18,894	6,306	3,352	2,978	17,552	1,157	50,238

Table 3 lists the relevant forecasts and error rates for the duration of this project. Out of fourteen years, the forecasts were within the targeted ±10% on only seven years for the climate forecasts, and six for the growers forecasts.

Table 3. Results for the project forecasts.

Year	Actual crop	Climate forecast	% error	Growers forecast	% error	Pest-scouts forecast	% error	Long-term forecast	% error
2001					–				–
	34,800	36,000	3.4	33,400	4.0			33,100	4.9
2002									
	30,200	32,600	7.9	32,300	7.0			34,850	15.4
2003									
	29,700	34,200	15.2	33,800	13.8			36,900	24.2
2004	43,700	35,065	–19.8	33,400	–23.6	34,000	–22.2	38,000	–13.0
2005			–						
	35,500	35,200	0.8	38,650	8.9	40,330	13.6	38,800	9.3
2006	43,900	41,800	–4.8	39,300	–10.5	38,000	–13.4	40,500	–7.7
2007	41,800	39,400	–5.7	36,300	–13.2	37,600	–10.0	42,000	0.5
2008									
	36,000	45,600	26.7	40,100	11.4	43,300	20.3	43,500	20.8
2009	37,500		26.9				22.1		20.3
		47,600		47,500	26.7	45,800		45,100	
2010	35,500		17.2				0.3		24.2
		41,600		40,800	14.9	35,600		44,100	
2011	28,500	38,900	36.5	33,000				40,500	42.1

					15.8		
2012			–7.3		–		–5.6
	40,000	37,070		38,280	4.3	37,760	
2013	35,200	39,180	11.3	38,173	8.4	39,620	12.6
2014	43,600	40,500	–7.1	40,293	–7.6	40,600	–6.9

Considering these data by periods, the first seven years saw a period of ‘generally good crops’. Price/kg NIS for macadamias increased steadily from $2.45 in 2001 to $3.60 in 2005, before falling back to $2.60 in 2006. The absolute error rates for 2001 to 2007 averaged an acceptable 8.2% for the climate forecasts, and 11.6% for the growers forecasts. During this period the regression ensembles worked quite well, despite these statistical methods only being ‘in their infancy’ at this time. The growers did not perform all that well, but were learning from the feedback, and were taking some pride in trying to improve their estimates over time.

The next four years (2008 to 2011) had notably poor crops, and all of the forecasts were too high. These poor crops were associated with lower macadamia prices – crashing to a very low $1.50 in 2007, and staying low at $1.90 in 2009. We note here the lag effect on actual production (as was found in the regression models). During these years the mean error rates were 26.8% (climate) and 17.2% (growers), showing that the growers were more ‘attuned’ to what was really going on in their orchards and in the overall industry. The climate-adjusted forecast models had optimistically assumed ‘about the same production patterns as before’, but these yields were clearly not being achieved. There were anecdotal reports of many producers not fertilizing or pruning their trees, and even not doing their final crop pick-up of the year, because ‘the prices being obtained simply did not justify this effort’. Prices started picking up again in 2010, rising from $2.65 in that year to $3.10 in 2011, and then increasing steadily to $3.60 in 2014.

For the past three years (2012 to 2014), the absolute error rates have averaged 8.6% for the climate models and 6.8% for the growers forecasts. Whilst being worse than the climate forecasts in the initial years, over the time of this project the growers estimates have shown steady improvement. Overall, these growers forecasts have been better than the climate forecasts in 7½ of the 14 years (considering 2009 as effectively a tie).

Extension

All extension of the project results was conducted through the Australian Macadamia Society. In March or April of each year, the annual forecasting report was provided to the CEO of the AMS. In most years, a meeting of the steering committee was held shortly thereafter, and following this the AMS usually issued their ‘official’ crop forecast via a media release.

Outcomes

In each year from 2001 to 2015, the long-term, climate-adjusted and growers forecasts have all been produced for, and communicated on-time to, the CEO of the Australian Macadamia Society. These forecasts 'have become widely accepted as the most accurate within the global macadamia industry' (Jolyon Burnett, pers. comm., 2015). They have been variously used by the key industry bodies and companies for planning and marketing, and for more accurate management of the levy program.

These forecasts have helped the AMS drive the development of a global supply and demand reporting system for the industry. Each year the major macadamia countries now submit data on supply and sales, allowing a reasonably accurate global accounting of the trade. 'This has helped give the market confidence and helped stabilize the price and reduce discounting. This has probably saved the Australian industry hundreds of thousands of dollars.' (Jolyon Burnett, pers. comm., 2015).

Presentations have been prepared and made to a number of national and international conferences, as specified in the 'Scientific Publications' list following. Also, at the request of the AMS, contributions were made at a number of macadamia industry working groups and research forums. In the continual quest for improved forecasts, cross-organisational collaborations have been formed (including CSIRO, NSW Department of Agriculture, and Queensland Department of Natural Resources), and we have also consulted with external experts (e.g., robotics and satellite-data providers at the 2014 International Horticultural Congress, and earlier with a consultancy forecasting company).

Evaluation and Discussion

Whilst appearing somewhat disappointing, the relative accuracies of the more recent macadamia forecasts (8.6% for the climate models and 6.8% for the growers forecasts) are reasonably similar to other tree-crop forecasting exercises around the world. The USDA annually issue forecasts for their almond crop, based on extensive crop-sampling and with only a two-month lead-time. From 2001 to 2014 these forecasts had an average absolute error rate of 7.8%, with the worst being 13.8% (in both 2002 and 2011). Peiris et al. (2008) predicted coconut production in Sri Lanka using seasonal climate information (primarily rainfall), with an average absolute error rate of 6.8% (for the two years of the study only).

The Australian macadamia industry has proven to be quite difficult to forecast, for a number of reasons, including –

- Incomplete data on tree numbers, ages and densities, requiring the use of scale-up factors for the long-term model.
- Possibly incomplete production totals (mainly concerning exports) which were not included in the processors' data, particularly in the earlier years of the project.

- Varietal differences – anecdotal information suggests short-term climatic influences on pollination success and nut-set which can be quite variable across varieties and regions, and we simply do not have the detailed data to capture these effects.

- Variable management, both across and within regions and years. Our models assumed ‘approximately constant management’ which would have resulted in steadier yields. However during the period of low prices it was quite obvious that management was less rigorous, as the trees did not deliver anywhere near their potential (as was evident in the preceding and following years).

- Real-world problems affecting the harvest amounts after the forecasts have been made, including for example flooding rain during harvest, and the destructive winds of ex-tropical cyclone Oswald.

- The macadamia tree being notoriously ‘difficult to quantify’, and known to respond to many influences. As a key industry figure initially told us, ‘There are many ways to ruin a potentially good crop’ (John Wilkie Snr., pers. comm., 2000).

The continued provision of macadamia crop forecasts, at alternate levels, would require differing amounts of investment and future resources, as follows –

- The growers forecasts are the easiest to produce. These involve the AMS continuing both their annual collection of regional data from the processors, and their survey of key growers and pest-scouts. These are quite-easily integrated and updated (in an Excel spreadsheet) to produce the growers forecasts, which in recent years have demonstrated an acceptable degree of accuracy.

- For continuing long-term forecasts to be made, new industry data would need to be sourced. The 2010 tree census is now considered ‘too dated’ for practical use, as anecdotal information suggests a reasonable level of tree removals and replacements as producers try to maintain competitiveness. These data would need to be updated via a new tree census, which would also collect the necessary yield data. Alternately, newer methods including aerial photography or satellite imagery could possibly be used to obtain the necessary estimates of planted areas and likely yields. The long-term model does not change much each year, so would only need to be revised every three or four years.

- If the revised ‘expected cropping amounts’ from the long-term model are available, the same statistical models (or newer developments) could again be used to produce the climate-adjusted forecasts, for a budget of around $8,000 p.a. This however may not be warranted, given their similar degree of accuracy to the simpler growers forecasts.

In summary, this project has led to a greater understanding of the mechanisms and climatic influences on macadamia production. A range of alternate statistical developments were investigated, including nearest-neighbour methods (not so successful) and regression ensembles (somewhat successful). The ‘simpler’ and low-cost growers forecasts have shown steady improvement over time, as these personnel become better educated and ‘take pride and ownership’ of this process.

Recommendations

1. That the AMS continue to coordinate the annual collection of regional data from the processors, and their survey of key growers and pest-scouts, so that the annual growers forecasts can continue to be made. This task would only require quite-minimal financial support.
2. If continuing long-term forecasts are required by the industry, new data would need to be sourced. These could be obtained from an updated tree census, or alternately estimates of planted areas and yields utilizing aerial photography or satellite imagery. These newer techniques would first need to be benchmarked against the current forecasting methods and results.

Scientific Publications (for the successive macadamia forecasting projects)

Journal articles (refereed)

Mayer, D.G., Stephenson, R.A., Jones, K.H., Wilson, K.J., Bell, D.J.D., Wilkie, J., Lovatt, J.L., Delaney, K.E. 2006. Annual forecasting of the Australian macadamia crop – integrating tree census data with statistical climate-adjustment models. *Agricultural Systems* **91**, 159-170.

Stephenson, R.A., Mayer, D.G. 2008. Forecasting the Australian macadamia crop via mechanistic and statistical climate models. *Acta Horticulturae* **773**, 165-172.

Mayer, D.G., Stephenson, R.A. 2015. Statistical forecasting of the Australian macadamia crop. *Acta Horticulturae* (in press).

Conference proceedings (refereed)

Mayer, D.G., Stephenson, R.A., Jones, K.H., Yee-Yet, J.S., Dunstan, A.W., Bell, D.J.D., Delaney, K.E., Wilson, K.J., Wilke, J. 2001. Mechanistic and statistical models to forecast the Australian macadamia crop. Proceedings International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand Inc., 10-13 December 2001, Australian National University, Canberra.

Mayer, D.G., Stephenson, R.A. 2015. Statistical ensemble models to forecast the Australian macadamia crop. Proceedings International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand Inc., 29 November - 4 December 2015, Gold Coast, (in press).

Conference proceedings (not refereed)

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

Mayer, D.G., Stephenson, R.A., Delaney, K.E. 2002. Forecasting for the market. Proceedings of the Macadamia and Australian Nut Industry Council Conference, Australian Macadamia Society Ltd, 9-12 October 2002, Coffs Harbour, 16-22.

Mayer, D.G., Stephenson, R.A., Jones, K.H., Wilson, K.J., Bell, D.J.D., Owens, J.S., Delaney, K.E. 2003. Forecasting the Australian macadamia crop. Proceedings of the Second International Macadamia Symposium, Australian Macadamia Society Ltd, 29 September - 4 October 2003, Tweed Heads, 30-33.

Mayer, D.G., Stephenson, R.A. 2005. Historical yield patterns in the AMS census data. Proceedings of the Australian Macadamia Society Conference 2005, Australian Macadamia Society Ltd, 2 - 5 November 2005, Noosaville, 77-79.

Mayer, D.G., Stephenson, R.A. 2008. Comparison of regression suite and nearest-neighbour forecasts for the national macadamia crop. Proceedings Australasian GenStat Conference, 2 – 5 December 2008, Marysville, 46.

Jones, Kim, Mayer, David 2009. The Australian macadamia industry: the past and the future. Proceedings 4th International Macadamia Symposium, 6 - 9 September 2009, Kruger National Park, South Africa.

Intellectual Property/Commercialisation

No commercial IP generated.

References

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.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., Garcia Marquez, J.R., Gruber, B., Lafourcade, B., Leitao, P.J., Munkemuller, T., McClean, C., Osborne, P.E., Reineking, B., Schroder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27-46.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* **77**, 802-813.

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

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference and prediction, Second edition, Springer.

Hoerl, R.W., Snee, R.D., De Veaux, R.D., 2014. Applying statistical thinking to 'Big Data' problems. *WIREs Computational Statistics* **6**, 222–232.

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.

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.

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., Robertson, D., Sedgley, M., Kristiansen, P., Olesen, T., 2013. Production trends in mature macadamia orchards and the effects of selective limb removal, side-hedging, and topping on yield, nut characteristics, tree size, and economics. *HortTechnology* **23**, 64-73.

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.

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

Mayer, D.G., Stephenson, R.A. 2008. Comparison of regression suite and nearest-neighbour forecasts for the national macadamia crop. Proceedings Australasian GenStat Conference, 2 – 5 December 2008, Marysville, 46.

Mayer, D.G., Stephenson, R.A., Jones, K.H., Wilson, K.J., Bell, D.J.D., Wilkie, J., Lovatt, J.L., Delaney, K.E. 2006. Annual forecasting of the Australian macadamia crop – integrating tree census data with statistical climate-adjustment models. *Agricultural Systems* **91**, 159-170.

Peiris, T. S. G., Hansen, J. W., Zubair, L., 2008. Use of seasonal climate information to predict coconut production in Sri Lanka. *International Journal of Climatology*, **28**, 103–110.

R Core Team 2013. R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

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, 30–37.

Song, L., Langfelder, P., Horvath, S., 2013. Random generalized linear model: a highly accurate and interpretable ensemble predictor. *BMC Bioinformatics* **14**, no. 5, URL http://www.biomedcentral.com/1471-2105/14/5

Acknowledgements

We thank the staff and members of the Australian Macadamia Society, and key macadamia-industry personnel, who have assisted us greatly - in particular Kim Jones, Jolyon Burnett, Robbie Commens, Kim Wilson, David Bell, Lyndsay Bryen, Kevin Quinlan, Chris Searle, Jerry Lovatt and Paul O'Hare. We are grateful to the macadamia processors who volunteered their intake data.

Appendix. Growers Forecast Form

(The replies from the grower are indicated in red.)

AMS Crop Forecasting Feedback Form for 2015 crop *(Jan 2015)*

Name	**(Confidential)**
Date	**28.1.15**
Growing region (Bundaberg, Gympie, Sunshine Coast, Nth Rivers or Nambucca)	**Bundaberg**

Average Pest Levels *(**Please record data on a local region basis with 0=Absent, 1 =low, 2 = medium [spray decision], 3 = high or 4 = very high**)*

Flower Caterpillar	**0**
FSB	**0**
Nut Borer	**0**
Phytophthora	**1**
Other Pests *(for eg - Botrytis, feltid coccid, Banana Fruit Caterpillar)*	**1**
Beneficial activity	**2-3**

How do you rate the amount of crop lost to Husk Spot is: *(Place an "X" below)*

Low	Moderate	Average	High	Very High
0-1				

Additional Information

What tonnage of nut has been lost to storms, hail and other environmental conditions?	
Are there any varietal or regional variances that you are aware of?	**Same as normal**
Do you have any additional comments or concerns regarding the size of the crop?	

At this stage, what is your prediction for the upcoming season's crop on the farms that you monitor? (*Place an "X" below)*

Down 50%	Down 25%	Down 10%	Same as last year	Up 10%	Up 25%	Up 50%
				X		

At this stage what is your estimate for the total industry production *(tonnes of NIS)*-	**50,000t**