10.1071/AR06081_ AC

Accessory Publication: Aust. J. Agric. Res., 2007, 58(2), 87-94.

Accessory publication. Derivation and Validation of the APSIM Regional Cumulative Biomass Indices

APSIM Scenario Procedure

Growing conditions within the Ingham, Ayr, Mackay and Bundaberg regions are quite diverse. To address the challenge of using APSIM to derive a regional proxy index of cumulative biomass, a cumulative biomass index was derived for many different scenarios. Though not exhaustive, these scenarios represented a large range of environmental and management conditions found in the region. Table A1.1 details how these scenarios were generated for each region.

In the case of Ingham, scenarios were generated from two irrigation scenarios (rainfed and irrigated), two soils (a Red Ferrosol and a Yellow Chromosol, (Inman-Bamber *et al.* (2000)), two cane lodging settings (no lodging and lodging) as detailed in Inman-Bamber et al. (2004), two flood settings (flooding and no flooding) and three climate stations (Macknade, Ingham and Bambaroo). Flooding damage was simulated by decreasing radiation use efficiency by 30% when more than 1000 mm was recorded during a period of continuous rain. Allocations of 2 ML/ha and 4 ML/ha were considered for the irrigated Ingham scenarios with irrigation occurring at two stress trigger points of 80% and 110%. This means that irrigation was applied when the loss in biomass gain between irrigations or after rainfall, was 80% on the one hand or no loss (110%) on the other. Planting dates from April through to October were also considered. Collectively, these settings generated 672 irrigated and 168 rainfed scenarios.

The scenarios for the remaining locations were generated similarly though the soil, planting, water allocation and stress triggers varied between regions. In the case of Ayr, a 30% reduction in radiation use efficiency (RUE) was considered after 50 t ha^{-1} biomass was reached in the simulation. This was done to account for the observation that RUE is often reduced in ageing crops possibly by lodging, stalk loss and reduced leaf N (Park *et al.* 2005). Table A1.2 gives the locations of all climate stations considered in this analysis. Details of other soil properties used in the simulations were similar to those used by Inman-Bamber et al. (2000).

In total 840 scenarios were considered for Ingham and Ayr, 960 scenarios for Mackay and 384 scenarios for Bundaberg. Similar though less intensive combinatorial type approaches for identifying suitable model input parameters have been considered for modelling the production of soy beans and sorghum (Hansen and Jones 2000, Potgieter et al. 2005).

Cross-validated Scenario Selection Procedure

Each scenario in Table A1.1 generated a biomass index that represented cumulative crop growth immediately preceding the forecast date. Owing to the large number of biomass indices generated, a selection procedure was required to identify suitable scenarios and the corresponding biomass indices for predicting regional sugarcane yields. To assist this task, a cross-validated correlation coefficient (Myers 1990)

$$r_{cv} = \frac{\sum_{i=1976}^{2003} y_i \hat{y}_{-i} - n\bar{y} \times \text{mean}(\hat{y}_{-i})}{(n-1)s_y s_{\hat{y}_{-i}}} \tag{A1.1}$$

was computed between the regional yield y_i for year $i \in (1976, 1977, ...,2003)$ and the predicted yield $\hat{y}_{-i}$, generated from a regression against yield with each simulated biomass index. The subscript '-*i*' in $\hat{y}_{-i}$ signifies the predicted yield was computed from a regression model that was built in isolation of data from year *i*. As an example, $\hat{y}_{-1976}$ represents the predicted yield for 1976 that was generated from a regression model built from actual yields (dependent variable) and a simulated biomass index (independent variable) using data from 1977 to 2003. The process was repeated until each data point has been omitted and predicted. The cross-validated correlation coefficient therefore provides an indication of the predictive capability of the model. The denominator in Eqn A1.1 is a function of the standard deviation of actual yields (s_y) and the standard deviation of the leave-one-out cross-validated predicted yields ($s_{\hat{y}_{-i}}$). Cumulative biomass indices that produced higher cross-validated correlation coefficients across the range of forecast dates were 'shortlisted'. The selected or 'shortlisted' indices were averaged to produce a predictive biomass index (X_B) that was used to produce operational forecasts via a simple linear regression $\hat{y} = a + b X_B$.

Assessing the Significance of the Predictive Biomass Indices

For each region, a large search space containing biomass indices from each scenario from Table A1.1 was generated owing to the broad range of environmental and management conditions. Whilst the cross-validation approach within the scenario selection procedure will make it challenging for an index to correlate well with actual yields purely by chance, there remains no guarantee that a chance correlation has not occurred. To quantify this risk a Monte Carlo procedure (Potgieter *et al.* 2005; Everingham *et al.* 2003; Good, 1997) was implemented. This involved computing the area A_0 under a piecewise linear function. The vertical axis of this function was generated by plotting the cross-validated correlation coefficient (r_{cv}) between the final predictive biomass index (X_B) and actual regional yield (y). This was repeated for each forecast date (horziontal axis). Figure A1.1 for example shows the r_{cv} between actual Ingham yields and the Ingham predictive biomass index. To compare how well the forecasting approach compared against a chance forecasting system, the regional yields were randomly permuted (jumbled) and correlated with each biomass index in the search space at each forecast date. The maximum area obtained from the randomised yields with each scenario was computed. This process was repeated 1000 times to generate the areas A_1, A_2,..., A_{1000}. The number of areas A_i for i = 1,2,...,1000 greater than or equal to A_0 were counted. If the approach taken to generate the operational predictive biomass indices is sound, only a small proportion of areas should exceed or equal A_0. This proportion is equivalent to a P-value or significance level of the predictive index.

Results

The biomass indices that gave the highest r_{cv} with regional yields across the range of forecast dates were extracted from the search space. Eight of the 840 biomass indices were selected Ingham, 17 out of the 840 biomass indices were selected for Ayr, 6 from the 960 indices for Mackay and 10 from the 384 indices for Bundaberg. The environmental and management settings that these indices span are listed in Table A1.3.

The significance level of each regional predictive biomass indices was computed by the Monte Carlo procedure. Significance levels of 0.007, 0.001, 0.072 and 0.000 were obtained for Ingham, Ayr, Mackay and Bundaberg, respectively. These small significance levels indicate the performance of the predictive biomass indices is unlikely a consequence of chance.

Summary

This appendix has described what can be considered a statistical agro-meteorological approach for deriving and validating proxy indicators of regional crop growth. Owing to the large spatial scale that the indices represent, statistical methods were needed to identify the most suitable APSIM outputs for predicting crop size and to quantify the soundness of this selection procedure. Although the discriminant analysis paper centers on yield forecasts produced in December, the biomass input variables to the discriminant procedure were selected on performance measures of these indices across a range of lead-times from December through to April. In December knowledge about crop category is adequate for marketers, but at later lead-times regression approaches (e.g. see Everingham *et al.* 2005) are needed to give more refined estimates of crop size.

References

Everingham YL, Inman-Bamber NG, Ticehurst C, Barrett D, Lowe K, McNeill T (2005) Yield forecasting for marketers. *Proceedings Australian Society Sugar Technologists Association* **27**, 51-60.

Everingham YL, Muchow RC, Stone RC, Coomans DH (2003) Enhancing sugarcane Yield forecasting capability using SOI phases: a case study for north eastern Australia. *International Journal of Climatology* **23**,1195-1210.

Good PI (1997) 'Permutation tests: a practical guide to resampling methods for testing hypotheses (3rd edition).' (Springer Verlag New York Inc: New York)

Hansen JW, Jones JW (2000) Scaling-up crop models for climate variability applications. *Agricultural Systems* **65**, 43-72.

Inman-Bamber NG, Zund PR and Muchow RC (2000) Water use efficiency and soil water availability for sugarcane. *Proceedings Australian Society Sugar Technologists Association* **22**, 264-269.

Inman-Bamber NG, Everingham YL, Muchow RC (2001) Modelling water stress response in sugarcane: Validation and application of the APSIM-Sugarcane model. 10th Australian Agronomy Conference, Hobart, 28 Jan. to 1 Feb., 2001.

Inman-Bamber NG, Attard, SJ and Spillman MF (2004) Can lodging be controlled through irrigation? *Proceedings Australian Society Sugar Technologists Association* **26**, CD-ROM.

Myers RH (1990) ‘Classical and modern regression with applications.’ (Duxbury Press: Beaumont CA)

Park SE, Robertson MJ and Inman-Bamber NG (2005) Decline in the growth of a sugarcane crop with age under high input conditions. *Field Crops Research* **92**, 305-320.

Potgieter AB, Hammer GL, Doherty A, de Voil P (2005) A simple regional-scale model for forecasting sorghum yield across north-eastern Australia. *Agricultural and Forest Meteorology* **132**, *143-153.*

Table A1.1 Crop simulation scenarios

Treatment factors and levels applied to the APSIM-Sugarcane model to represent the range of growing conditions in Ingham (irrigated and rainfed), Ayr, Mackay and Bundaberg. Soils are abbreviated as follows: Rferro = Red Ferrosol soil; Rderm = Red Dermosol soil; Ychrom = Yellow Chromosol (Inman-Bamber *et al.* 2000).

Location	Factor	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Total
Ingham	Plant date	Apr	May	Jun	Jul	Aug	Sep	Oct	7
(irrigated)	Soil	Rferro	Ychrom						2
	Lodging	No lodge	Lodge						2
	Allocation	2 ML/ha	4 ML/ha						2
	Flooding	Yes	No						2
	Climate	Macknade	Ingham	Bambaroo					3
	Irrig cycle	7 days							1
	Stress trigger %	80	110						2
							Total Number of Scenarios		672
Ingham	Plant date	Apr	May	Jun	Jul	Aug	Sep	Oct	7
(rainfed)	Soil	Rferro	Ychrom						2
	Lodging	No lodge	Lodge						2
	Allocation	0							1
	Flooding	Yes	No						2
	Climate	Macknade	Ingham	Bambaroo					3
							Total Number of Scenarios		168
Ayr	Plant date	Apr	May	Jun	Jul	Aug	Sep	Nov	7
	Soil	Rderm							1
	Lodging	Nolodge	Lodge						2
	Allocation	No limit							1
	Flooding	Yes	No						2
	Climate	Ayr	Millaroo	Clare	Kalamia	Shirbourne			5
	Irrig cycle	0 days							1
	Stress trigger %	100	90	80					3
	Late RUE reduction	Yes	No						2
							Total Number of Scenarios		840
Mackay	Plant date	May	Jul	Sep	Nov				4
	Soil	Rferro							1
	Lodging	No lodge	Lodge						2
	Allocation	0 ML/ha	2 ML/ha	4 ML/ha	6 ML/ha				4
	Flooding	Yes	No						2
	Climate	Gargett	Pleystowe	Farleigh	Proserpine	Sarina			5
	Irrig cycle	7 days							1
	Stress trigger %	70	90	100					3
							Total Number of Scenarios		960
Bundaberg	Plant date	Jun	Jul	Aug	Sep	Oct	Nov		6
	Soil	Rferro							1
	Lodging	No lodge	Lodge						2
	Allocation	Variable							1
	Flooding	Yes	No						2
	Climate	Bundaberg	Childers	Fairymead	Maryborough				4
	Irrig cycle	7 days							1
	Stress trigger %	70	80	90	100				4
							Total Number of Scenarios		384

Table A1.2 Weather Stations

Locations of official climate recording stations from which daily climate data were supplied to the APSIM sugarcane crop model.

	Weather Station		
Region	Name	Coordinates	
Ingham	Macknade	146°15’	-18°36’
	Ingham	146°10’	-18°39’
	Bambaroo	146°11’	-18°53’
Ayr	Ayr	147°24’	-19°34’
	Millaroo	147°16’	-20°03’
	Clare	147°13’	-19°47’
	Kalamia	147°25’	-19°31’
	Shirbourne	147°06’	-19°36’
Mackay	Gargett	148°45’	-21°09’
	Pleystowe	149°03’	-21°09’
	Farleigh	149°06’	-21°06’
	Prosperpine	148°32’	-20°30’
	Sarina	149°13’	-21°25’
Bundaberg	Bundaberg	152°23’	-24°51’
	Childers	152°17’	-25°24’
	Fairymead	152°24’	-24°48’
	Maryborough	152°41’	-25°33’

Table A1.3 Scenarios used in operational forecasts

Scenarios that were averaged to predict regional yields for Ingham, Ayr, Mackay and Bundaberg.

Location	Climate	Soil	Plant date	Lodging	Allocation (ML/ha)	Flooding	Stress Trigger (%)	Irrigation cycle (days)	Late RUE reduction
Ingham	Bambaroo	Rferro	Oct	No	2	Yes	80	7d	No
	Bambaroo	Rferro	Oct	No	2	No	80	7d	No
	Bambaroo	Rferro	Oct	No	4	Yes	80	7d	No
	Bambaroo	Rferro	Oct	No	4	No	80	7d	No
	Bambaroo	Rferro	Oct	Yes	2	Yes	80	7d	No
	Bambaroo	Rferro	Oct	Yes	2	No	80	7d	No
	Bambaroo	Rferro	Oct	Yes	4	Yes	80	7d	No
	Bambaroo	Rferro	Oct	Yes	4	No	80	7d	No
Ayr	Ayr	Rderm	Jun	No	No limit	Yes	90	0d	Yes
	Ayr	Rderm	Jun	No	No limit	Yes	90	0d	No
	Ayr	Rderm	Jun	No	No limit	No	90	0d	Yes
	Kalamia	Rderm	Aug	No	No limit	Yes	90	0d	Yes
	Kalamia	Rderm	Aug	No	No limit	No	90	0d	Yes
	Kalamia	Rderm	Sep	No	No limit	Yes	90	0d	Yes
	Ayr	Rderm	Sep	No	No limit	Yes	90	0d	No
	Kalamia	Rderm	Sep	No	No limit	Yes	90	0d	No
	Kalamia	Rderm	Sep	No	No limit	No	90	0d	Yes
	Ayr	Rderm	Sep	No	No limit	No	90	0d	No
	Kalamia	Rderm	Sep	No	No limit	No	90	0d	No
	Kalamia	Rderm	Sep	Yes	No limit	Yes	90	0d	Yes
	Ayr	Rderm	Sep	Yes	No limit	Yes	90	0d	No
	Kalamia	Rderm	Sep	Yes	No limit	Yes	90	0d	No
	Kalamia	Rderm	Sep	Yes	No limit	No	90	0d	Yes
	Ayr	Rderm	Sep	Yes	No limit	No	90	0d	No
	Kalamia	Rderm	Sep	Yes	No limit	No	90	0d	No
Mackay	Gargett	Rferro	Sep	No	4	Yes	100	7d	No
	Gargett	Rferro	Sep	No	4	No	100	7d	No
	Gargett	Rferro	Sep	Yes	4	Yes	100	7d	No
	Gargett	Rferro	Sep	Yes	4	No	100	7d	No
	Gargett	Rferro	Sep	Yes	6	Yes	90	7d	No
	Gargett	Rferro	Sep	Yes	6	No	90	7d	No
Bundaberg	Bundaberg	Rferro	Sep	No	Variable	Yes	70	7d	No
	Bundaberg	Rferro	Sep	No	Variable	Yes	80	7d	No
	Bundaberg	Rferro	Sep	No	Variable	No	70	7d	No
	Bundaberg	Rferro	Sep	No	Variable	No	80	7d	No
	Bundaberg	Rferro	Sep	Yes	Variable	Yes	70	7d	No
	Bundaberg	Rferro	Sep	Yes	Variable	Yes	80	7d	No
	Bundaberg	Rferro	Sep	Yes	Variable	No	70	7d	No
	Bundaberg	Rferro	Sep	Yes	Variable	No	80	7d	No
	Bundaberg	Rferro	Oct	Yes	Variable	Yes	80	7d	No
	Bundaberg	Rferro	Oct	Yes	Variable	No	80	7d	No

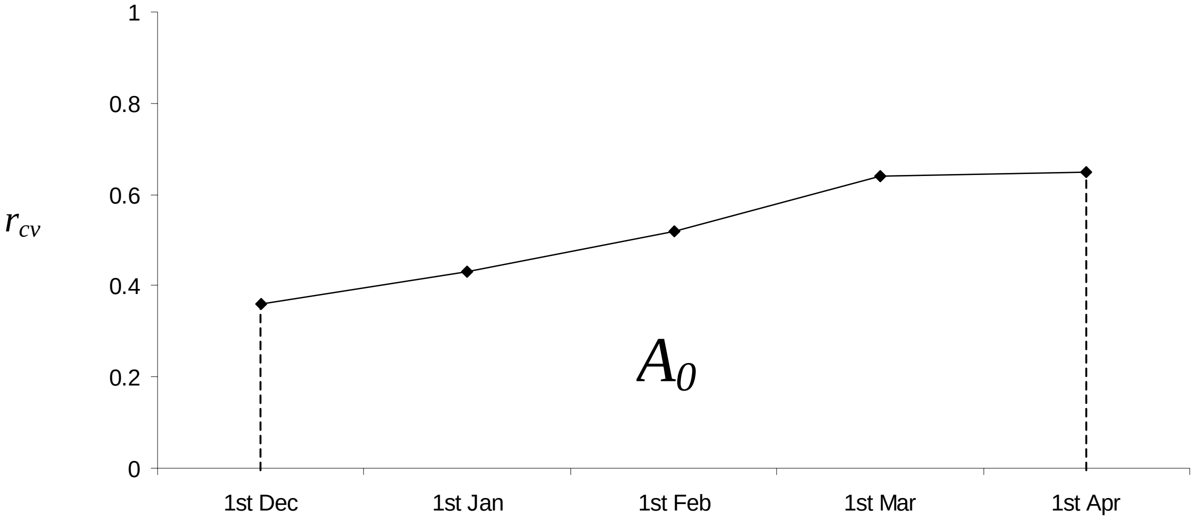

Fig. A1.1 The cross-validated correlation coefficient between Ingham yields and the predictive biomass index used operationally at different forecast dates. The area beneath this line is denoted by A_0.