Crop-SI, a yield estimation model based on earth observation and meteorological-driven stress indices

Yang Chen[1], Randall J. Donohue[2], Tim R. McVicar[2], François Waldner[3], Gonzalo Mata[1], Noboru Ota[1], Alireza Houshmandfar[1], Kavina Dayal[4], Roger A. Lawes[1]

[1] CSIRO Agriculture and Food, Underwood Ave, Floreat WA 6014, Australia; y.chen@csiro.au
[2] CSIRO Land and Water, GPO Box 1700, Canberra, ACT 2061, Australia;
[3] CSIRO Agriculture and Food, 306 Carmody Rd, St Lucia QLD 4067, Australia;
[4] CSIRO Agriculture and Food, College Rd, Sandy Bay TAS 7005, Australia.

Abstract

Inter-annual national crop yields fluctuate due to area planted, management strategies, prevailing weather conditions, weeds, diseases, and pests. High temporal and spatial variability in grain production makes nationwide crop yield prediction challenging. The present study developed a model combining semi-physical and empirical approaches to estimate yield of major crops (*i.e.*, canola, wheat and barley) across the Australian dryland wheatbelt using a remote sensing (RS) based radiation use efficiency approach and meteorology-driven Stress Indices (*SI*). Crop specific stress indices (*e.g.*, drought, heat and cold stress) in critical months (*e.g.*, anthesis and grain-filling) were used to explain the impact of highly variable (in both space and time) actual grain yield across the wheatbelt. The present model, Crop-SI, reduces the field-scale prediction error rate by ~20% when compared with existing models for nationwide crop yield simulations. Our finding have improved the predictive capability of RS-based models for crop yield for a wide range of variability in meteorological conditions by incorporating rainfall and temperature into the simulation and provides new insights for the next-generation of nationwide agricultural yield models.

Key Words

Crop yield estimation, radiation use efficiency, stress indices, NDVI, remote sensing, canola, wheat, barley

Introduction

Crop yield can be predicted by physical, empirical, and semi-empirical approaches (Lobell, 2013; Holloway and Mengersen, 2018). The physical models (e.g. the Agricultural Production Systems sIMmulator – APSIM and WOrld FOod Studies – WOFOST; van Diepen et al., 1989; Keating et al., 2003; Holzworth et al., 2014) can explain crop growth at point-level in space on the basis of underlying interactions among genotype, environment, and management strategies but require detailed inputs of data about local soils, crop management, and prevailing weather conditions to predict yield (Holzworth et al., 2014). The parameterisations used in such models have limited validity over large areas, as weather conditions, soils and farmer management can change over short distances (Lobell, 2013).

The empirical models often relate yield observations to remotely-sensed variables like vegetation indices (VIs; Ferencz et al., 2004; de Wit and van Diepen, 2008). These VIs provide an indication of plant biophysical condition and above ground biomass (Pinter Jr et al., 2003). Many of these empirical models are specific to agrology and climate; therefore, they cannot be applied to multiple regions exhibiting high spatial variability in agricultural production.

Semi-empirical models focus on the biochemical mechanism of photosynthesis, dry matter (biomass) accumulation, and water consumption, and use linear relationships (e.g. harvest function) for yield estimates (Mo et al., 2005). The RS-based harvest function cannot be downscaled to the field level because the harvest index (*HI*) needs to be calibrated according to diverse climatic and soil conditions (Gaiser et al., 2010). The present study aims to improve the nationwide semi-empirical RS-based crop yield estimation by relating carbon estimation to drought, cold, and heat stress for canola, wheat, and barley in Australian wheatbelt.

Methods

Data

We used 16-day composite Moderate Resolution Imaging Spectroradiometer (MODIS) Collection 6 satellite data to estimate crop yields at both field and national scales. MODIS provides at best daily observation of canopy-level Normalized Difference Vegetation Index (NDVI; Colwell, 1974), which can be used to estimate photosynthetic activity globally at 250 m resolution.

Yield estimation

Figure 1 describes the framework implemented for the semi-empirical model for crop yield estimation. The grain yield model was developed by relating remotely sensed carbon fixation to actual yield while accounting for meteorology-driven environmental stresses.

$$y = a \cdot C + b \cdot SI + d \tag{1}$$

where y is the estimated yield (t/ha), *C* is a remotely sensed indicator of carbon fixed by plants during the growing season (gC/m^2), and *a*, *b*, and *d* are the model coefficients. *SI* (see Table 1) were selected and integrated from the meteorological variables using Support Vector Machines (SVMs; Weston et al., 2001) to maximize their correlation with the actual yield. The non-*SI* ($b = 0$, $d = 0$) and *SI*-based models are compared to analyse the contribution of *SI* to improvement in the yield prediction.

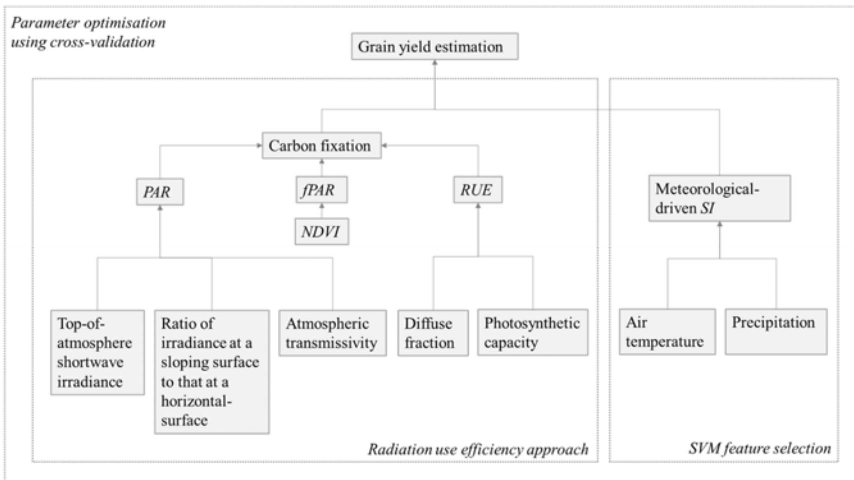

Figure 1. The implementation flowchart of the Crop-SI estimation. The shaded frames indicate the model required parameters and the dashed frames illustrate the links between parameters.

Table 1. Meteorological-driven accountability of crop yield estimation from 2009 to 2015.

Meteorological variables at each calendar month (*m*)	Explanation
$T_{max(m)}$	mean daily maximum temperature (°C) across each month from April to November (*m:4-11)*
$T_{min(m)}$	mean daily minimum temperature (°C) across each month from April to November (*m:4-11)*
$P_{(m)}$	mean monthly precipitation (mm) across each month from April to November (*m:4-11)*
$H_{(m)}$	numbers of days of heat stress: $T_{max} \geq 32.0$ °C (d) from August to September (*m:8-9)*
$F_{(m)}$	numbers of days of frost stress: $T_{min} \leq -2.0$ °C (d) from May to September (*m:5-9)*

Carbon fixation

The carbon fixation (*C*) was derived using a radiation use efficiency (RUE) approach. For unstressed plants, *RUE* and the fraction of photosynthetically active radiation (*fPAR*) are conventionally used to estimate net carbon fixation (Monteith, 1972; Monteith, 1977; Mo et al., 2005). Growing season carbon fixation is given as:

$$C = \int_{i_{min}}^{i_{max}} GPP(i)\, \Delta_i \tag{2}$$

where GPP is gross primary productivity ($gC/m^2/d$). GPP can be estimated by $PAR_i \cdot fPAR_i \cdot RUE_i$, where *i* is the time step of the model (16-days) (Monteith, 1972; Monteith, 1977; Leblon et al., 1991), i_{min} and i_{max} represent the beginning and the end of the growing season.

Calibration and validation

The total of 291 field-year combinations were available, and these observations were randomly split into three independent datasets for calibration (60% of the field-years), test (20% of the field-years), and validation (20% of the field-years). To avoid autocorrelation issues arising from pixels being used for the same field-years in both calibration and validation sets, pixels from one field-year could only be in one of the two datasets. Pixel-level cross-validation was used to optimise the parameters used in Equation 1.

Results and Discussion

The independent validation set was used to compare the results with a widely used harvest function (HF; Whisler et al., 1986) and a carbon turn-over model (C-Crop; Donohue et al., 2018). Figure 2 shows that for pixel-level yield prediction, the HF underestimates yield prediction, the predicted yield is linearly correlated to the observed values with a coefficient of determination *(R^2)* of 0.68, 0.77, and 0.43 for canola, wheat, and barley, respectively. Relative to HF, C-Crop performed better for pixel-level yield prediction and reduced the *Root Mean Square Error* (*RMSE)* from ~1 t/ha to ~0.5 t/ha and from ~2.3 t/ha to ~0.8 t/ha, for canola and barley, respectively (Figure 2). The Crop-SI produced the highest R^2 for canola (0.82) and barley (0.71) and the lowest *RMSE* of ~0.5 t/ha across all three crops. For barley and canola it was able to explain an extra 18% and 12% of variability, respectively, with a reduction of *RMSE* by ~0.3 t/ha for barley.

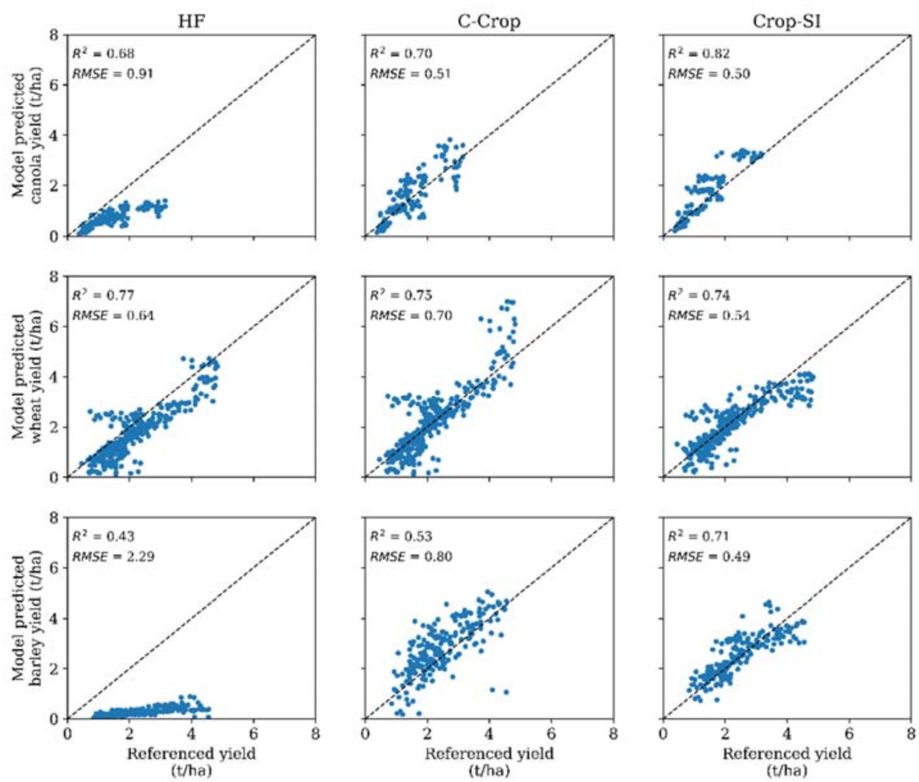

Figure 2. Scatter plots comparing modelled and observed canola (n = 1,606), wheat (n = 12,388), and barley (n = 2,508) yields on 250 m pixel-level. From left to right, the 1st column shows HF predicted yields against the observed values; the 2nd column is the C-Crop estimated yields against the observed values; the 3rd column demonstrates Crop-SI predicted yields against the observed values. From top to bottom, the 1st, 2nd, and 3rd row shows the model comparisons of canola, wheat, and barley yield, respectively.

This improvement in predictive accuracy illustrates that meteorological variability strongly influences the yield variation given the carbon production for cereals (*i.e.*, wheat and barley) across Australia. While yield models such as C-Crop (*i.e.*, $y = a \cdot C$) adequately explain actual yield variation, their performance can be improved as evidenced by lower (higher) values of *RMSE* (R^2) when extended to include the *SI* (See Table 2). For instance, during 'haying off' more aboveground biomass early on in the season leads to a reduction in tiller economy and an inability to complete grain filling (Van Herwaarden et al., 1998). Meteorological variables, therefore, should be used for regional to national scale yield estimation by accounting for spatial and temporal heterogeneities in the plant *RUE* to improve the yield model prediction across multiple climatic zones.

Table 2. Statistical performance of the calibrated yield model with and without *SI* on 250 m pixel-level for canola, wheat, and barley. For each crop the number of 250 m pixels in the validation dataset are reported, in order, after the crop name.

Model validation at pixel-level	Non-*SI* model		*SI* model	
Crop (mean observed yield (t/ha), number of the pixels)	R^2	*RMSE* (t/ha)	R^2	*RMSE* (t/ha)
Canola (1.5 t/ha, 1,606)	0.47	0.59	0.82	0.50
Wheat (2.2 t/ha, 12,388)	0.68	0.65	0.74	0.54
Barley (2.5 t/ha, 2,508)	0.58	0.62	0.71	0.49

Future study should contribute to extensive yield map collection. Crop-SI should be tested using finer spatial resolution data (e.g. Landsat and Sentinel 2) to provide more detailed information for practical and profitable decision making at sub-field scales.

Conclusion

There is considerable demand for regional to national-scale grain yield estimation during the cropping season by growers, grain marketers, grain handlers, agricultural businesses, and market brokers. This study developed a parsimonious, robust and repeatable method with improved predictive accuracy to estimate regional and nationwide crop yield by combining gridded meteorological data with a remotely-sensed description of plant carbon fixation during the growing season following a radiation use efficiency approach.

References

Colwell JE (1974). Vegetation canopy reflectance. Remote Sensing of Environment. 3(3), 175-183.

de Wit AJW and van Diepen CA (2008). Crop growth modelling and crop yield forecasting using satellite-derived meteorological inputs. International Journal of Applied Earth Observation and Geoinformation. 10(4), 414-425.

Donohue RJ, et al. (2018). Towards a national, remote-sensing-based model for predicting field-scale crop yield. Field Crops Research. 227, 79-90.

Ferencz C, et al. (2004). Crop yield estimation by satellite remote sensing. International Journal of Remote Sensing. 25(20), 4113-4149.

Gaiser T, et al. (2010). Validation and reliability of the EPIC model to simulate maize production in small-holder farming systems in tropical sub-humid West Africa and semi-arid Brazil. Agriculture, Ecosystems & Environment. 135(4), 318-327.

Holloway J and Mengersen K (2018). Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review. Remote Sensing. 10(9), 1365.

Holzworth DP, et al. (2014). APSIM–evolution towards a new generation of agricultural systems simulation. Environmental Modelling & Software. 62, 327-350.

Keating BA, et al. (2003). An overview of APSIM, a model designed for farming systems simulation. European Journal of Agronomy. 18(3-4), 267-288.

Leblon B, et al. (1991). The use of remotely sensed data in estimation of PAR use efficiency and biomass production of flooded rice. Remote Sensing of Environment. 38(2), 147-158.

Lobell DB (2013). The use of satellite data for crop yield gap analysis. Field Crops Research. 143, 56-64.

Mo X, et al. (2005). Prediction of crop yield, water consumption and water use efficiency with a SVAT-crop growth model using remotely sensed data on the North China Plain. Ecological Modelling. 183(2), 301-322.

Monteith JL (1972). Solar Radiation and Productivity in Tropical Ecosystems. Journal of Applied Ecology. 9(3), 747-766.

Monteith JL (1977). Climate and the efficiency of crop production in Britain. Philosophical Transactions of the Royal Society of London. B, Biological Sciences. 281(980), 277-294.

Pinter Jr PJ, et al. (2003). Remote sensing for crop management. Photogrammetric Engineering & Remote Sensing. 69(6), 647-664.

van Diepen CA, et al. (1989). WOFOST: a simulation model of crop production. Soil Use and Management. 5(1), 16-24.

Van Herwaarden A, et al. (1998). 'Haying-off', the negative grain yield response of dryland wheat to nitrogen fertiliser. I. Biomass, grain yield, and water use. Australian Journal of Agricultural Research. 49(7), 1067-1082.

Weston J, et al. (2001). Feature selection for SVMs, Advances in neural information processing systems, pp. 668-674.

Whisler F, et al. (1986). Crop simulation models in agronomic systems, Advances in Agronomy. Elsevier, pp. 141-208.