SIMULATING SPATIAL VARIABILITY OF CEREAL YIELDS FROM HISTORICAL YIELD MAPS AND SATELLITE IMAGERY

YP Dang[1], A Apan[2], RC Dalal[3], S Darr[1], M Schmidt[3], M Pringle[3]

[1]Department of Environment and Resource Management, Toowoomba, Qld 4350;
Yash.Dang@derm.qld.gov.au
[2]The University of Southern Queensland, Toowoomba, Qld 4350
[3]Department of Environment and Resource Management, Indooroopilly, Qld 4068

ABSTRACT

The management of spatial variability of crop yields relies on the availability of affordable and accurate spatial data. Yield maps are a direct measure of the crop yields, however, costs and difficulties in collection and processing to generate yield maps results in poor availability of such data in Australia. In this study, we used historical mid-season normalised difference vegetation index (NDVI), generated from Landsat imagery over 4 years. Using linear regression model, the NDVI was compared to the actual yield map from a 257 ha paddock. The difference between actual and predicted yield showed that 77% and 93% of the paddock area had an error of <20% and <30%, respectively. The linear model obtained in the paddock was used to simulate crop yield for an adjoining paddock of 162 ha. On an average of 4 years, the difference between actual and simulated yield showed that 87% of the paddock had an error of <20%. However, this error varied from season to season. Paddock area with <20% error increased exponentially with decreasing in-crop rainfall between anthesis and crop maturity. Furthermore, the error in simulating crop yield also varied with the soil constraints. Paddock zones with high concentrations of subsoil chloride and surface soil exchangeable sodium percentage generally had higher percent of error in simulating crop yields. Satellite imagery consistently over-predicted cereal yields in areas with subsoil constraints, possibly due to chloride-induced water stress during grain filling. The simulated yield mapping methodology offers an opportunity to identify within-field spatial variability using satellite imagery as a surrogate measure of biomass. However, the ability to successfully simulate crop yields at farm scale or regional scale requires wider evaluation across different soil types and climatic conditions.

Dang, Y.P., Apan, A., Dalal, R.C. ,Darr, S. Schmidt, M. and M. Pringle (2009).Simulating spatial variability of cereal yields from historical yield maps and satellite imagery. In: Ostendorf B., Baldock, P., Bruce, D., Burdett, M. and P. Corcoran (eds.), Proceedings of the Surveying & Spatial Sciences Institute Biennial International Conference, Adelaide 2009, Surveying & Spatial Sciences Institute, pp. 1205-1216. ISBN: 978-0-9581366-8-6.

INTRODUCTION

Recent developments in remote sensing have shown promise for quantifying crop yield variations both within and between agricultural fields (Fisher *et al.* 2009). Remotely sensed images show spatial and spectral variations resulting from soil and crop characteristics. Australia has more than 25 years of historical satellite data available thus there is great potential to increase the quantity and quality of spatial data needed to identify causes of spatial and temporal variability in cropping areas. The potential advantages of remote sensing images are: the ability to bypass field measurements of yield, estimate yield at a range of spatial scales thus eliminating sampling error within field variability, and the availability of archived imagery thus enabling analysis of past growing seasons that may not have recorded yield (Lobell *et al.* 2003).

Measurements of reflected light have often been used for the assessment of green vegetation, biomass or physiological stresses in agricultural plants (Tucker, 1979). A number of vegetation indices have been developed to estimate or represent crop yield. The most commonly used ratio is the normalised difference vegetation index (NDVI) defined as:

$$NDVI = (NIR-RED)/(NIR+RED)$$

where NIR is reflectance in near infrared spectral band, 790-900 nm, and RED is reflectance in red spectral band, 620-680 nm. The interactions of RED and NIR light with green plant material are very different; however, their reactions to most soil types are similar (Carlson & Ripley, 1997). Further, NDVI has been shown to be closely related to grain yield in terms of delineating a spatial pattern (Abuzar *et al.* 2004).

Generally good relationships between satellite image derived NDVI and crop yield have been reported for a number of crop species (Hatfield, 1983; Labus *et al.* 2002; Scotford & Miller, 2005). Historical series of mid-season NDVI maps averaged for seasons of common crop type are traditionally used to determine zones of consistently average, below-average and above-average NDVI (Adams & Maling, 2005). Also growers readily recognise NDVI and easily relate it to their knowledge of the field (Robertson *et al.* 2007). However, paddock zone maps obtained using NDVI do not provide spatial variability of yield for making economic calculations and farm management decisions. We therefore, attempted to develop an empirical-statistical model to estimate paddock yield variability at sub-paddock scale and to predict yield variability at farm scale.

MATERIALS AND METHODS

Site description

Two adjoining paddocks were selected on a farm near Goondiwindi in southern Queensland. The Goondiwindi district is typical of much of Australia's northern subtropical grains region. The climate of the region is semi-arid with high potential evapotranspiration (1300-2200 mm per annum), low (550-800 mm of average annual rainfall) and variable (coefficient of variation 27%) rainfall, most of which falls during summer (Webb *et al.* 1997). The common soil was grey cracking clay (Vertosol) (Isbell, 1996). One of these two paddocks (15A; 257 ha; 28° 19' S and 150° 30' E) was selected for developing the empirical-statistical model to estimate grain yield. In general, the

western side of both the paddocks had higher levels of subsoil constraints particularly high concentration of subsoil chloride and high exchangeable sodium percent (Fig. 1).

(a)15A (b)15B

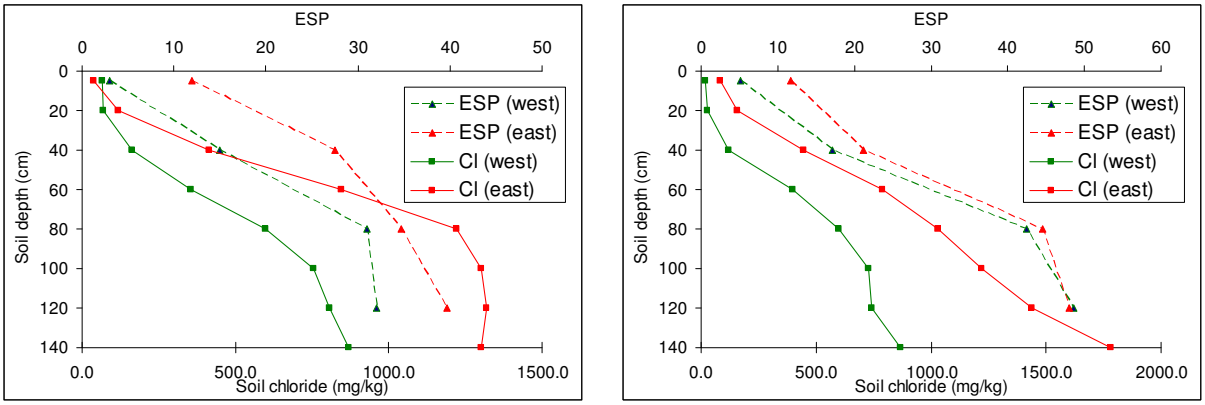


Fig. 1. Distribution of subsoil constraints in the profiles of western (high yielding) and eastern (low yielding) sides of paddocks (a) 15A, and (b) 15B.

In this paddock, cereal crops are generally sown in May. The anthesis is around mid-September, and crops are harvested during October. Long-term average annual rainfall for this area is 617 mm and the growing season rainfall (May-October) is 225 mm. The second paddock (15B; 162 ha) was selected to validate the empirical-statistical model and predict grain yield using relationship developed from paddock 15A. The study focussed on wheat and barley crops. Information on the rotation for each paddock and average yield of the cereal crops as reported by growers are given in Table 1.

Table 1. Crop rotations, paddock yields (t ha^{-1}) for paddocks 15A and 15B and the date of satellite image acquisition

Paddock	Year	Rotation	Sowing	Harvesting	Image acquisition	Av. Yield (t/ha)
15A	2001	Wheat	19th June	31st Oct	24th Sept	1.78
	2002	Wheat	18th June	13th Nov	10th Sept	0.81
	2003	Chickpea	17th May	15th Oct		0.97
	2004	Wheat	12th May	1st Nov	22nd Aug	2.42
	2005	Barley	29th May	5th Nov	27th Sept	2.10
	2006	Sorghum	5th Sept	10th April		1.27
	2007	Wheat	6th May	28th Oct	31st Aug	1.25
15B	2001	Wheat	19th June	31st Oct	24th Sept	2.35
	2002	Wheat	18th June	13th Nov	10th Sept	1.09
	2003	Chickpea	17th May	15th Oct		0.95
	2004	Wheat	12th May	1st Nov	22nd Aug	2.40
	2005	Wheat	23th May	5th Nov	27th Sept	1.92
	2006	Sorghum	5th Sept	10th April		1.95
	2007	Wheat	6th May	28th Oct	31st Aug	1.88

Grain yield

Site-specific yield data for the cereal crops since 2001 were accessed from a grower who collected yield data at harvest using AgLeader yield-monitoring equipment fitted in

the grain harvester linked to a differentially corrected GPS. Point data were collected at 1-s intervals along the path of the harvester. The raw yield monitor data were extracted from yield mapping software format and pre-processed passed through a routine to provide a consistent data set for analysis. Positional data were transformed into Map Grid of Australia 1994 (MGA 94, Zone 56) co-ordinate set. Yield monitor data was cleaned by deleting outliers (0 t/ha) and above a specific improbable value for the region (>10 t/ha). The average yield and standard deviation for the whole dataset was calculated, and then the normalised yield was obtained for each individual data point:

Normalised yield = (Yield – average yield)/standard deviation

Data was sorted on the basis of normalised yield and all records outside ±3 t/ha were deleted. The clean yield data was used for statistical analysis and map production. Grain yield from different seasons were combined to develop standardised temporal mean yield (Larscheid and Blackmore, 1996), thereby removing inter-year offset (Blackmore *et al* 2003).

Actual grain yield data and standardised yield data for each season was spatially interpolated using kriging to the 25x25 m grid using Vesper software (Whelan *et al.* 2001). With all seasons' yield data on a common grid, multivariate *k*-means clustering analysis based upon a process of fitting data iteratively to a specified number of clusters was used to define the potential management zones. The method creates disjoint classes by estimating cluster means which maximises the variation within the cluster groupings (Whelan & McBratney, 2003). The number of clusters was obtained by calculating confidence interval (Taylor *et al.* 2007) using the mean kriging variance to determine if yield response in classes is statistically different from each other.

Satellite images

Cloud-free images of Landsat 5 TM (Thematic Mapper) and Landsat TM 7 ETM+ (Enhanced Thematic Mapper) satellite sensors were acquired close to the crop anthesis (Table 1). All images were geometrically and radiometrically corrected before analysis using National Mapping 9 second DEM, and Supplemental Control Points, derived from controlled full length Landsat 7 passes of the Australian continent. The effects of relief displacement are corrected using DEM (Geoscience Australia 2009). The location of both the paddocks' boundary was identified on the satellite image using ENVI software (ITT Visual Information Solutions, 2008), and the NDVI transformations were obtained for each crop year that a cereal crop was grown. Similar to grain yield data, standardised temporal mean NDVI (SNDVI) for each individual data point was obtained by dividing each NDVI value with mean NDVI for the season. For each node of 25x25 m grain yield grid, the NDVI values and standardised NDVI (SNDVI) values were obtained using nearest neighbour interpolation.

All interpolated files for grain yield and NDVI were then imported into the SAS JMP statistical analysis software (SAS Institute Inc., 2007). A linear relationship was obtained between temporal standardised mean grain yields and corresponding temporal standardised mean NDVI (Fig. 2).

The coefficients of the linear model were used to simulate the spatio-temporal variability of grain yield, based on NDVI. Using simulated yield data, simulated potential management zones were created with *k*-mean clustering to compare with

management zones created using actual yield maps. The coefficients of the linear model were further used to predict the spatial and temporal variability in the grain yield of adjoining paddock (15B) and compared with actual yield maps obtained by the grower. The accuracy of the simulated yield mapping was evaluated by calculating the difference between actual yield map and the simulated yield map.

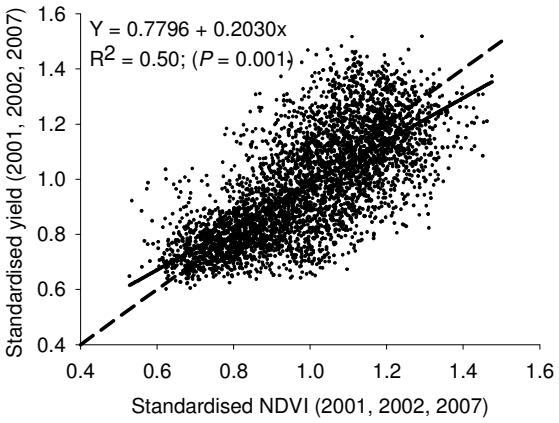


Fig. 2. Relationship between standardised temporal mean actual yield and satellite derived standardised temporal mean NDVI for Paddock 15A.

RESULTS AND DISCUSSION

Actual cereal yields were mapped by the grower for paddock 15A in 2001, 2002 and 2007. The mean spatial and temporal variability map for paddock 15A over these years is shown in Fig. 3a. The mean simulated yield map for Paddock 15A, obtained from satellite images acquired in 2001, 2002, 2004 and 2007, is shown in Fig. 3b. The residual yield map was obtained by calculating the difference between the three years of mean actual yield map and the four years of mean simulated yield map (Fig. 3c). The difference between actual and simulated yield showed that 77% of the paddock area (198 ha) had yield predictions within ±0.2 t/ha, 20% of the paddock area (52 ha) had yield predictions between > ±0.2 but < ±0.4 t/ha and only 3% of the paddock area (7 ha) had > ±0.4 t/ha error of prediction.

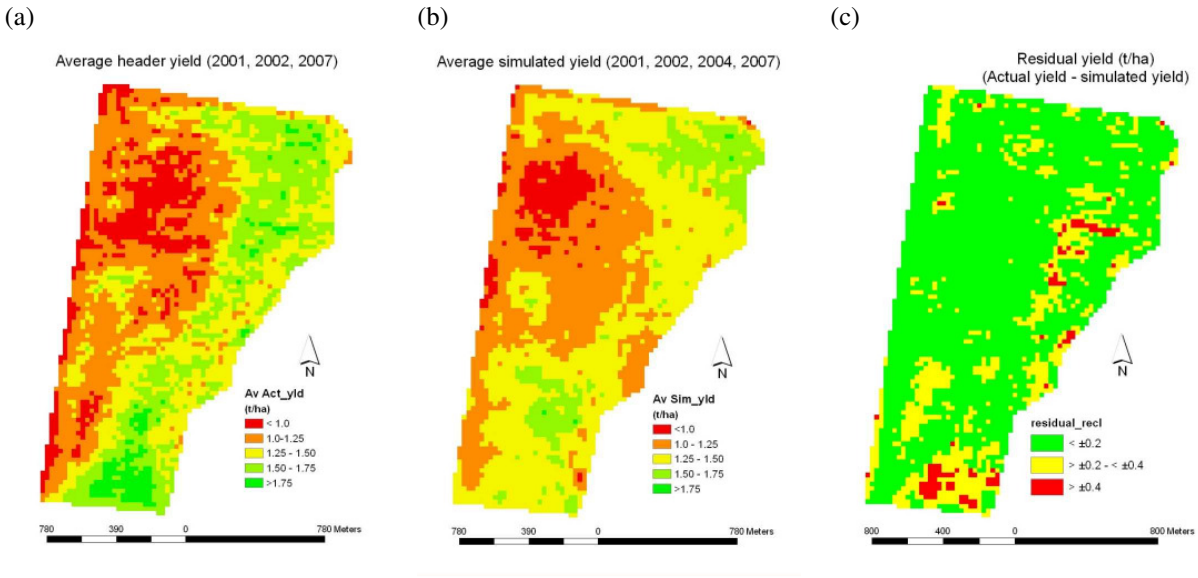


Fig. 3. Maps of Paddock 15A showing spatial variation in cereal yield for (a) mean actual yield (b) mean simulated yield and (c) difference between actual and simulated yield.

In terms of expressing residual yield as a percentage of actual yield, 88% of the area had an error of <20% from the actual mean yield, and 97% of the area had an error of <30% (Figure 4a). The majority of the paddock (55%) yields were over-predicted by simulated yield mapping (Figure 4b). The yields in the eastern side of the paddock with low levels of subsoil constraints were under-predicted whereas the western side of the paddock with relatively high levels of subsoil constraints had over-predicted yields.

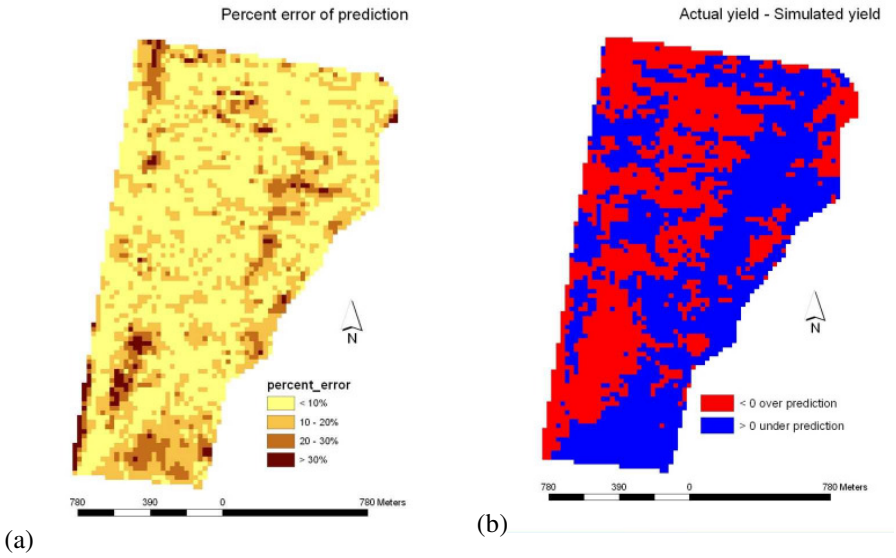


Fig. 4. Map of Paddock 15A showing spatial variability in the difference between actual and simulated yield expressed as (a) percentage of actual yield, and (b) an over or under prediction by the simulated yield map.

Productivity maps prepared either using actual grain yield data for 2001, 2002 and 2007 (Fig. 5a) or using simulated yield maps derived from NDVI data obtained from satellite

images in 2001, 2002, 2004 and 2007 (Fig. 5b) showed an overall classification accuracy of 90%.

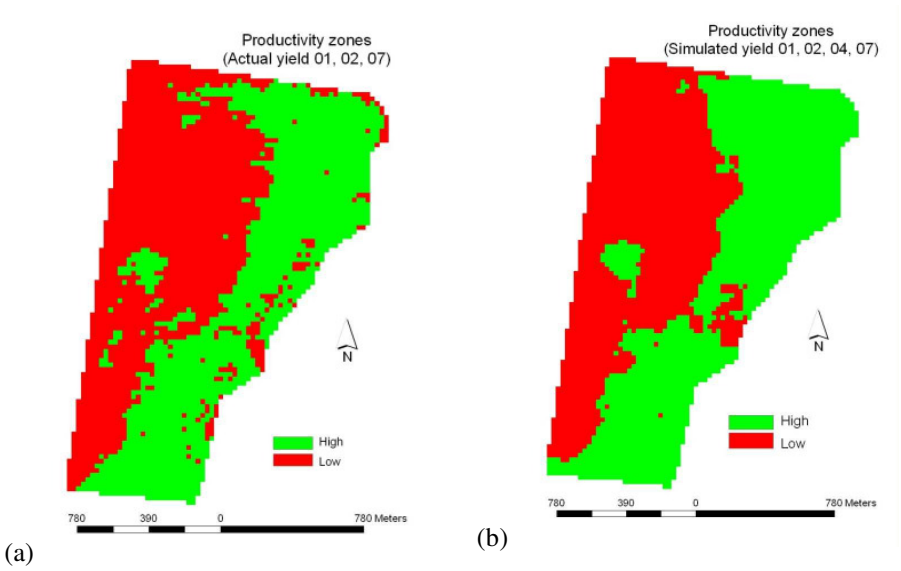


Fig. 5. Map of Paddock 15A showing potential management zones based on crop productivity obtained using (a) actual yield maps, and (b) simulated yield maps.

The linear model coefficients obtained in the paddock 15A was used to simulate crop yield for an adjoining paddock 15B (162 ha). On this paddock, cereal crops were grown in 2001, 2002, 2005 and 2007. The mean spatial and temporal variability of actual yield and simulated yield are shown in Fig. 6a and Fig. 6b, respectively.

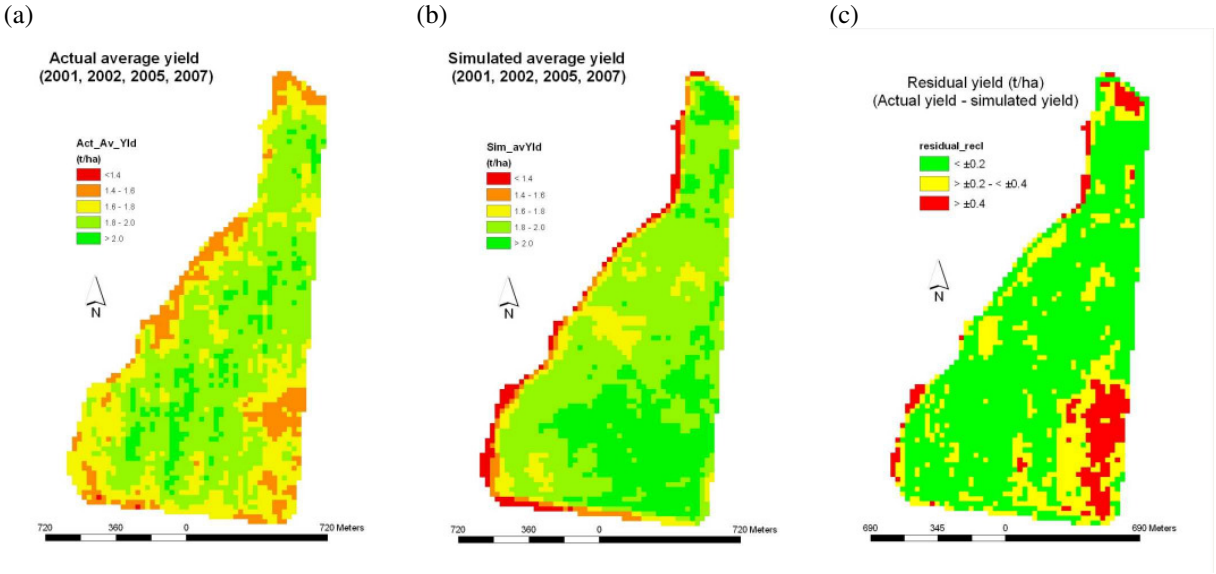
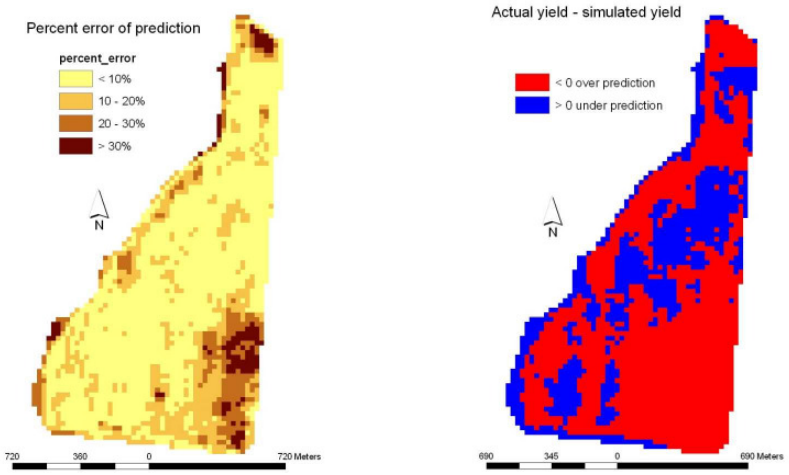


Fig. 6. Maps of Paddock 15B showing spatial variation in cereal yield for (a) mean actual yield (b) mean simulated yield and (c) difference between actual and simulated yield.

On an average of 4 years, the difference between actual and simulated yield showed that 67% of the paddock area (108 ha) had yield predictions within ±0.2 t/ha, 24% of the

paddock area (40 ha) had yield predictions between > ±0.2 but < ±0.4 t/ha and 9% of the paddock area (14 ha) had >±0.4 t/ha error of prediction (Fig. 6c). In terms of percent of actual grain yield, 87% of the paddock area had an error of <20% and 95% of the paddock area had an error of <30% (Fig. 7a).

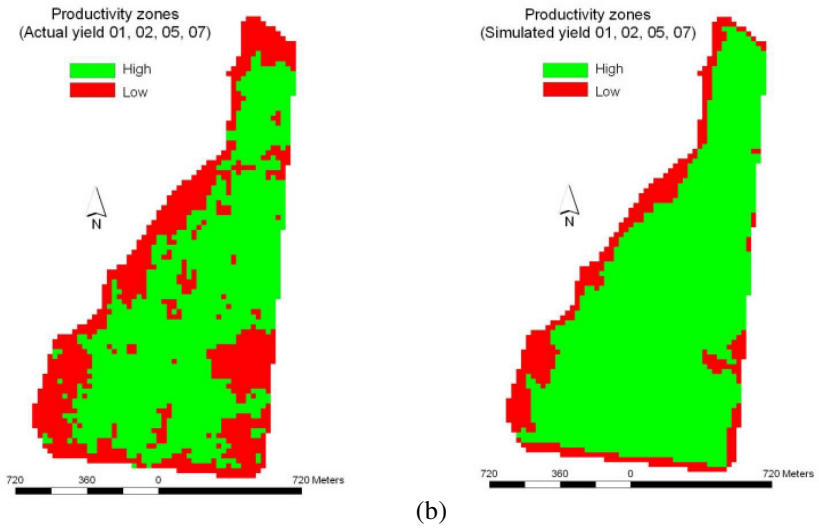


(a) (b)

Fig. 7. Map of Paddock 15B showing spatial variability in the difference between actual and simulated yield expressed as (a) percentage of actual yield, and (b) an over- or under-prediction by the simulated yield map.

The western end of the paddock generally had high concentration of subsoil chloride and high levels of exchangeable sodium percentage was over-predicted as compared to the eastern end with relatively low levels of subsoil chloride and exchangeable sodium percent (Fig. 7b).

Potential management zones obtained, using *k*-clustering, from simulated yield data showed overall classification accuracy of 80% with actual yield data (Fig. 8a, 8b).



(a) (b)

Fig. 8. Map of Paddock 15B showing potential management zones based on crop productivity obtained using (a) actual yield maps, and (b) simulated yield maps.

Individual year NDVI data was converted into annual simulated maps (figure not shown) and compared with actual yield maps (figure not shown). The maps of percent error of prediction showed that this error varied from season to season (Fig. 9).

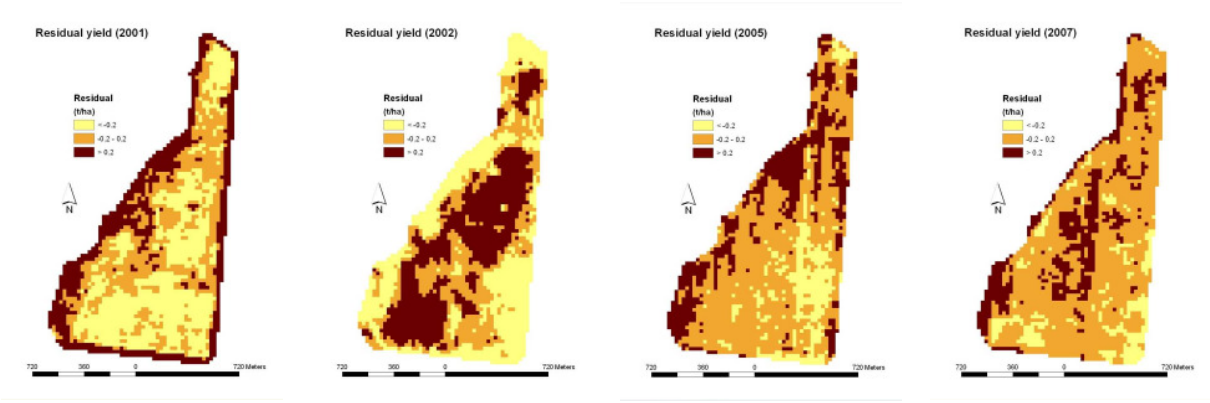


Fig. 9. Maps of paddock 15B showing difference between actual and simulated yield for individual years.

The paddock area with <20% error increased exponentially with decreasing in-crop rainfall between anthesis and crop maturity (Fig. 10).

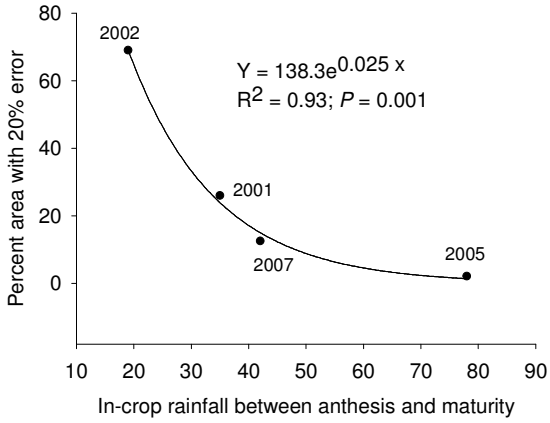


Fig. 10. Relationship between percent area with <20% prediction error and in-crop rainfall.

Given that only a small number of Australian growers have yield monitoring, the methodology proposed in this paper suggests that grain yield can be adequately predicted using NDVI derived from satellite images acquired at anthesis, allowing comparisons between simulated yield maps and actual yield maps. The usefulness of this technique relies on a strong positive correlation between NDVI and yield. Generally, there was a good agreement between harvested yield and NDVI in this study. A number of studies have shown good relationship between mid-season satellite-derived NDVI and grain yield (Hatfield, 1983; Labus et al. 2002; Scotford & Miller, 2005). However, a number of factors can cause differences between the mean simulated yield map and mean actual yield map. The relationship between NDVI and actual yield can differ from season to season, and is particularly dependent on the in-crop rainfall

especially after anthesis. Other environmental factors such as frost that usually occurs after anthesis can cause differences between simulated yield and actual yield.

In this study, the presence of subsoil constraints which are commonly present in the region (Dang *et al.* 2006) can influence the prediction, due possibly to water stress at grain filling from high subsoil Cl concentrations (Dang *et al.* 2008). The simulated yield mapping consistently over-predicted the western side of both the paddocks which correspond to relatively high concentration of chloride in the subsoil and high levels of exchangeable sodium (Fig. 1). The effect of subsoil constraints have been investigated (Dang *et al.* 2008) and shown to result in reduced uptake of subsoil water and nutrients.

The simulated yield mapping methodology offers an opportunity to identify within-field spatial variability using satellite imagery as a surrogate measure of biomass. Moreover, the technique developed here it provides an opportunity to identify areas suspected of subsoil constraints using multi-year satellite imagery on farm or regional scale at low cost. However, the ability to successfully simulate crop yields at farm scale or regional scale requires wider evaluation across different soil types and climatic conditions.

CONCLUSIONS

This study demonstrated that historical mid-season NDVI generated from Landsat imagery can predict crop yield with high level of accuracy. The location and magnitude of prediction errors varied from season to season primarily due to in-crop rainfall between anthesis and crop maturity. However, in this study, the error in simulating crop yield was also found to vary with the soil constraints prevailing in the area. Paddock zones with high concentrations of subsoil chloride and surface soil exchangeable sodium percentage generally had higher percent of error in simulating crop yields. Satellite imagery consistently over-predicted cereal yields in areas with subsoil constraints. The techniques developed in this work offer an opportunity to identify within-field spatial variability using satellite imagery as a surrogate measure of biomass. However, the ability to successfully simulate crop yields at farm scale or regional scale requires wider evaluation across different soil types and climatic conditions.

REFERENCES

Abuzar M, Rampant P, Fisher PD (2004) Measuring spatial variability of crops and soils at sub-paddock scale using remote sensing technologies. In 'Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS2004), 20 –24 September 2004, Anchorage, Alaska, Volume III, 1633 – 1636.

Adams ML, Maling I (2005) Simplifying management zones – a pragmatic approach to the development and interpretation of management zones in Australia. In Proceedings of the 7th International Conference on Precision Agriculture (Ed. DJ Mulla), 25-28 July 2004, Precision Agriculture Centre, University of Minnesota.

Blackmore S, Godwin RJ, Fountas S (2003) The analysis of spatial and temporal trends in yield map data over six years. *Biosystems Engineering* **84**, 455-466.

Bramley R (2002) Some tips on map production for precision Agriculture. *Southern Precision Agriculture Newsletter* **1(2)**: 9-11.

Dang YP, Dalal RC, Mayer DG, McDonald M, Routley R, Schwenke GD, Buck SR, Daniells IG, Singh DK, Manning W, Ferguson N (2008) High subsoil chloride reduces soil water extraction and crop yield from Vertosols. *Australian Journal of Agricultural Research* **59(5)**, 321-330.

Dang YP, Dalal RC, Routley R, Schwenke GD, Daniells I (2006) Subsoil constraints to grain production in the cropping soils of the north-eastern region of Australia: an overview. *Australian Journal of Experimental Agriculture* **46**, 19-35.

Fisher PD, Abuzar M, Best F, Rab MA (2009) Advances in precision agriculture in south-eastern Australia, Part I: A methodology for the combined use of historical paddock yields and Normalised Difference Vegetation Index to simulate spatial variation in cereal yields. *Australian Journal of Agricultural Research* (submitted)

Geoscience Australia (2009) Landsat Geometric and Radiometric Specifications. http://www.ga.gov.au/remote-sensing/get-satellite-imagery-data/technical-information/processing/landsat-geometric-radiometric.jsp

Hatfield JL (1983) Remote sensing estimators of potential and actual crop yield. *Remote Sensing of Environment* **13**, 301-311.

Isbell RF (2002) The Australian Soil Classification. Revised Edition. CSIRO Publishing, Collingwoood, Victoria.

ITT Visual Information Solutions (2008), ENVI 4.4, Boulder, CO, 80301, ITT Visual Information Solutions.

Labus MP, Nielsen GA, Lawrence RL, Engel R (2002) Wheat yield estimates using multi-temporal NDVI satellite imagery. *International Journal of Remote Sensing* **23**, 4169-4180.

Larscheid G, Blackmore BS (1996) Interaction between Farm Managers and Information Systems with respect to Yield Mapping, Proceedings of 3rd International Conference on Precision Agriculture, MN. ASA, CSSA, SSSA, pp. 1153 - 1163.

Lobell DB, Asner GP, Ortiz-Monasterio JI, Benning TL (2003) Remote sensing of regional crop production in the Yaqui Valley, Mexico: Estimates and uncertainties. *Agriculture Ecosystem and Environment* **94**: 205-220.

Robertson M, Isbister B, Maling I, Oliver Y, Wong M, Adams M, Bowden B, Tozer P (2007) Opportunities and constraints for managing within-field spatial variability in Western Australian grain production. *Field Crops Research* **104**, 60-67.

SAS Institute Inc. (2007) JMP User's Guide Version 7, Cary, NC, SAS Institute.

Scotford IM, Miller PCH (2005) Applications of spectral reflectance techniques in Northern European cereal production: A Review. *Biosystems Engineering* **90**, 235-25.

Taylor JA, McBratney AB, Whelan BM (2007) Establishing management classes for broadacre agricultural production. *Agronomy Journal* **99**: 1366-1376.

Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* **8**: 127-150.

Webb AA, Grundy MJ, Powell B, Littleboy M (1997) The Australian sub-tropical cereal belt: soils, climate and agriculture. In 'Sustainable crop production in the sub-tropics'. (Eds AL Clarke, PB Wylie) pp. 8-26. (Queensland Department of Primary Industries, Brisbane QI 97035).

Whelan BM, McBratney AB, Minasny B (2001) Vesper spatial prediction software for precision agriculture. In Proceeding 3rd European Conference on Precision

Agriculture (Eds G Grenier, S Blackmore), Montpellier, France, 18-20 June 2001. pp. 139-144.

Whelan BM, McBratney AB (2003) Definition and interpretation of potential management zones in Australia. In Proceeding 11th Australian Agronomy Conference (Eds M Unkovich, G O'Leary) 2-6 Feb. 2003, Horsham, Australia.

BRIEF BIOGRAPHY OF PRESENTER

I graduated with PhD in soil Science and Plant Nutrition from the University of Queensland in 1992. Currently, I am working as a Soil Scientist on utilisation of advanced techniques to identify and manage subsoil constraints in northern grains region of north-eastern Australia.