Advances in precision agriculture in south-eastern Australia. I. A regression methodology to simulate spatial variation in cereal yields using farmers' historical paddock yields an...

Article *in* Crop and Pasture Science · September 2009

DOI: 10.1071/CP08347

CITATIONS
27

READS
515

5 authors, including:

P. D. Fisher
Department of Economic Development, Jobs, Transport and Resources
22 PUBLICATIONS **486** CITATIONS
SEE PROFILE

Mohammad Abuzar
Department of Jobs, Precincts and Regions (Government of Victoria)
38 PUBLICATIONS **328** CITATIONS
SEE PROFILE

Abdur Rab
State Government Victoria
54 PUBLICATIONS **1,319** CITATIONS
SEE PROFILE

Subhash Chandra
Agriculture Victoria
285 PUBLICATIONS **3,283** CITATIONS
SEE PROFILE

CSIRO PUBLISHING
www.publish.csiro.au/journals/cp *Crop & Pasture Science*, 2009, **60**, 844–858

Advances in precision agriculture in south-eastern Australia. I. A regression methodology to simulate spatial variation in cereal yields using farmers' historical paddock yields and normalised difference vegetation index

P. D. Fisher[A,D], *M. Abuzar*[B], *M. A. Rab*[A], *F. Best*[C], *and S. Chandra*[A]

[A]Department of Primary Industries, 255 Ferguson Road, Tatura, Vic. 3616, Australia.
[B]Department of Primary Industries, 32 Lincoln Square North, Carlton, Vic. 3053, Australia.
[C]Birchip Cropping Group, 73 Cumming Ave, Birchip, Vic. 3483, Australia.
[D]Corresponding author. Email: peter.fisher@dpi.vic.gov.au

Abstract. Despite considerable interest by Australian farmers in precision agriculture (PA), its uptake has been low. Analysis of the possible financial benefits of alternative management options that are based on the underlying patterns of observed spatial and temporal yield variability in a paddock could increase farmer confidence in adopting PA.

The cost and difficulty in collecting harvester yield maps have meant that spatial yield data are generally not available in Australia. This study proposes a simple, economical and easy to use approach to generate simulated yield maps by using paddock-specific relationships between satellite normalised difference vegetation index (NDVI) and the farmer's average paddock yield records. The concept behind the approach is illustrated using a limited dataset. For each of 12 paddocks in a property where a farmer's paddock-level yield data were available for 3–5 years, the paddock-level yields showed a close to linear relationship with paddock-level NDVI across seasons. This estimated linear relationship for each paddock was used to simulate mean yields for the paddock at the subpaddock level at which NDVI data were available. For one paddock of 167 ha, for which 4 years of harvester yield data and 6 years of NDVI data were available, the map of simulated mean yield was compared with the map of harvester mean yield. The difference between the two maps, expressed as percentage deviation from the observed mean yield, was <20% for 63% of the paddock and <40% for 78% of the paddock area. For 3 seasons when there were both harvester yield data and NDVI data, the individual season simulated yields were within 30% of the observed yields for over 70% of the paddock area in 2 of the seasons, which is comparable with spatial crop modelling results reported elsewhere. For the third season, simulated yields were within 30% of the observed yield in only 22% of the paddock, but poor seasonal conditions meant that 40% of the paddock yielded <100 kg/ha. To illustrate the type of financial analysis of alternative management options that could be undertaken using the simulated yield data, a simple economic analysis comparing uniform *v.* variable rate nitrogen fertiliser is reported. This indicated that the benefits of using variable rate technology varied considerably between paddocks, depending on the degree of spatial yield variability. The proposed simulated yield mapping requires greater validation with larger datasets and a wider range of sites, but potentially offers growers and land managers a rapid and cost-effective tool for the initial estimation of subpaddock yield variability. Such maps could provide growers with the information necessary to carry out on-farm testing of the potential benefits of using variable applications of agronomic inputs, and to evaluate the financial benefits of greater investment in PA technology.

Additional keywords: yield map, variable rate technology, remote sensing, yield prediction, spatial variability, management zone.

Introduction

Introduction to paper series

Crop production is affected by factors that vary both in space and time. Site-specific crop and soil management systems, also described as precision agriculture (PA), apply agronomic science to the management of production practices to address spatial and temporal yield variability (Whelan and McBratney 2000; Brouder *et al*. 2001). Growers are aware that yield differences exist between and within their paddocks and recognise the potential value of using variable rate technology instead of uniform applications of inputs as one method of managing yield variability (Shanahan *et al*. 2004; Jayroe *et al*. 2005). Site-specific management of fertiliser, for example, may improve farm profitability through both increased yields and reduced fertiliser losses (Wilkerson and Moody 2004).

This paper is the first of 5 related papers that address the advances in PA resulting from a 5-year research project in the state of Victoria, Australia. This first paper describes advances

 10.1071/CP08347 1836-0947/09/090844

in rapidly and cost effectively mapping subpaddock crop variability by developing and evaluating a technique that converts normalised difference vegetation index (NDVI) data into simulated yield maps. This is achieved by using growers' historical annual paddock yield records to generate a paddock-specific conversion efficiency of biomass into yield. The approach is used to produce a whole-farm map of subpaddock yield variability, which is compared with harvester yield maps for one of the paddocks (Paddock 17). In contrast to this NDVI approach for estimating subpaddock yield variability, the paper by Robinson *et al*. (2009) explores the possibility of using proximally sensed geophysical data and terrain derivatives to map subpaddock yield variability. Maps of electromagnetic induction (EMI), gamma ray (γ-ray) spectrometry, and other terrain derivatives from Paddock 17 were compared with defined production zones and harvester yield maps.

The papers by Armstrong *et al*. (2009) and Rab *et al*. (2009) explain in detail the relationship of spatial variability in crop yield within Paddock 17 with variation in relevant soil properties and factors. The paper by Armstrong *et al*. (2009) describes how key soil chemicophysical properties interact with water and nitrogen supply, and how these relate to the growth and yield of crops over three consecutive seasons in different management zones. The paper by Rab *et al*. (2009) describes the connection between the spatial variation in soil hydraulic properties (field capacity, permanent wilting point, and plant-available water capacity) and yield variability, and how this helps to explain the phenomenon of yield flip-flop. Rab *et al*. (2009) also determines the optimum number of zones needed to develop variable rate technology applications.

The final paper by Anwar *et al*. (2009) uses the soil data presented in the papers by Armstrong *et al*. (2009) and Rab *et al*. (2009) to investigate the long-term behaviour of Paddock 17. In this paper, Anwar *et al*. (2009) use the APSIM crop model to determine whether the average wheat and barley yields over the past 119 years of weather data match the paddock zones developed using EMI data and recent yield maps.

Introduction to this paper

Many parallel techniques for paddock zone mapping are being developed through the use of technologies such as yield monitors, satellite and aerial imagery, and geophysical tools such as EMI and radiometrics. They are used both individually and combined, and can be applied to different scales of subpaddock management from just a few paddock zones, to applications that vary continuously. However, many Australian grain growers are uncertain whether the level of variability on their property justifies the capital investment in PA technology and it is estimated that only around 3% of Australian grain growers are using some form of PA technology (Price 2004). The cost of PA equipment and concern over the cost-benefit of investing in it are the two major reasons why Australian growers are cautious about adopting PA (Price 2004). It is difficult for growers to know if PA is cost effective until something is known about the subpaddock yield variability, which cannot be obtained without spending money on PA (Robertson *et al*. 2007). There is therefore a need to provide growers, particularly those who have not adopted PA, with reliable but easily and cheaply available information about the spatial yield variability within their paddocks. With this information they can estimate the expected cost-benefit ratio of PA management options, such as applying variable rate technology, and develop an initial level of PA investment to suit their enterprise.

Actual measurement of crop yield, using harvester-mounted yield monitors, provides the most accurate spatial yield data. However, there are several reasons why to date in Australia there are still only small amounts of actual harvester data available. Firstly, retrospectively fitting yield monitors is expensive, and even for harvesters that have them fitted the annual cost of a differential signal can be prohibitive. Secondly, often more than one machine will harvest a single field and therefore all machines need to be similarly equipped. Thirdly, setting up the required electronics, GPS signal, and data retrieval and storage systems is often difficult, time consuming, and unreliable, leading to lost data. Fourthly, the time and skill required to convert, combine, and generate yield maps are often not available to growers. A further problem with harvester-collected yield data is that even when they are available, many seasons of data are required to produce a representative picture of how a paddock responds to different crops and seasonal weather conditions. Crops or pasture rotations mean it could take a decade or more before interpretation of the paddock performance can be made for any particular crop type.

An alternative approach to actually measuring yield is to use the spatial variation in crop biomass at a relevant growth stage as a surrogate for yield. The advantage of this approach is that it is possible to use aerial or satellite images of spectral reflectance indices to estimate crop biomass, and many seasons of historical satellite images are often available. This means that the time lag necessary to collect sufficient seasons of data to describe paddock performance is less of an issue. Several indices of different reflected wavelengths have proved successful in predicting crop biomass, but the NDVI is most commonly used.

Paddock zone maps generated using NDVI have been proposed for many years. However, these do not provide yield data, and therefore cannot be used in yield-based economic calculations that are most useful in making farm management decisions. NDVI maps are traditionally used to provide information on which parts of a paddock, each year, perform badly or well, relative to the rest of the paddock. This is achieved through standardisation of each year's data by dividing by the average NDVI value for the whole paddock for that year (Larscheid and Blackmore 1996). By doing this the inter-year offset (Blackmore *et al*. 2003) and the inter-paddock offset are removed. The consequence of this approach is that high and low NDVI regions in one paddock cannot be compared, in terms of yield production, with high and low regions in any other paddocks or seasons.

An alternative approach is to use NDVI data to estimate yield variability. The agronomic basis for using biomass to forecast cereal yields in an Australian context has been discussed by Smith *et al*. (1995) and various applications for reflectance measurements have been reviewed by Scotford and Miller (2005). Around the world, estimates of biomass using ground-based and air-borne spectral vegetation indices and their

relationship to grain yield variability have been reported for many crop species (e.g. Hatfield 1983; Rudorff and Batista 1990; Elliott and Regan 1993; Bellairs *et al*. 1996; Roy and Ravan 1996; Zhang *et al*. 1998; Aparicio *et al*. 2000; Serrano *et al*. 2000; Staggenborg and Taylor 2000; Basso *et al*. 2001; Jayroe *et al*. 2005).

Generally, good relationships between satellite-derived NDVI and crop yield have also been reported for several decades (Nalepka *et al*. 1976; Bauer *et al*. 1977; MacDonald and Hall 1978; Shanahan *et al*. 2001; Labus *et al*. 2002; Wright *et al*. 2003) and the use of early season NDVI images to assist with current-season crop management has been considered by several authors (Ashcroft *et al*. 1990; Raun *et al*. 1999). In each of these studies, field sampling of yield has been used for calibration with NDVI, thus greatly increasing the cost and complexity of using NDVI data for estimating yield.

A simpler and more cost-effective approach to yield forecasting using NDVI data uses agricultural statistics as the main input for calibration. This has been used at large regional scales to predict annual yields (MacDonald and Hall 1978; MacDonald and Hall 1980; Bullock 1992; Bochenek 2000; Genovese *et al*. 2001; Weissteiner and Kühbauch 2005). In Australia, satellite images at a 100-ha resolution were used to predict wheat yields in 68 Western Australian Local Government Areas (Smith *et al*. 1995). In this study, average NDVI values for each area were correlated with grain measurement recorded at cooperative grain receival areas. They found that correlation between NDVI images captured at different times during the growing season and final yield reached a maximum in late August/early September and continued to remain significant during the period of crop senescence. They also found that a higher proportion of the variance in yield was explained by combining early and late season NDVI. At a finer resolution, Dawbin *et al*. (1980) using Landsat images from 1976, found high correlation between radiance values, although not an NDVI ratio, of the whole paddock and average paddock wheat yields on up to 88 paddocks in New South Wales, Australia ($r^2 = 72\%$). They found that the predicted paddock yields were within an average of 14.5% of the observed yields (range 1.6–4.0 t/ha).

The objective of this research was to extend these empirical modelling approaches for estimating paddock yield to estimate yield variability at a subpaddock level. For this, a simulated yield mapping (SYM) approach based on modelling the empirical relationship between historical paddock-level satellite NDVI data and growers' paddock-level yield records over years was developed. The approach is illustrated on a property comprising 12 paddocks, using a limited dataset of 3–5 years of historical yield data. The SYM approach produces estimated annual, or temporal-averaged, yield maps that incorporate individual paddock efficiencies of converting biomass, measured by NDVI, into final yield at a land-unit level at which NDVI data are available. The simulated yield maps can be used by growers to objectively position on-farm trials, or to economically evaluate variable rate management options for their enterprise, thus minimising their exposure to investment risk and increasing the rate of PA adoption.

Materials and methods

Site description

Twelve paddocks, contiguous as far as possible, were selected for the study from a single farm located near Birchip in the Victorian Mallee. One of these paddocks (Paddock 17; 35°48′S, 142°59′E) was selected for more intensive analysis. These paddocks represent a diversity of soil-landscape systems from the North West Dunefields and Plains geomorphology division (Joyce *et al*. 2007). Soil Orders (Isbell 2002) vary from Vertosols in the gilgaied plains and inter-ridge corridors, to Hypercalcic Calcarosols and Red Sodosols with sandy loam soil surface horizons that cover ridges and predominate on the rise slopes. Deep sandy soils (Red Tenosols and Sodosols) occur as superimposed dunes (Robinson *et al*. 2009). Long-term annual rainfall for Birchip (Woodlands) is 354 mm and the growing-season rainfall (April–October) is 239 mm. Annual and growing-season rainfall figures are presented in Rab *et al*. (2009). In this area, cereal crops are generally sown in May, anthesis is around mid-October, and crops are generally harvested during December.

Grain yield

This study focused on wheat and barley crops. Information on the rotation for each paddock, and the yield of the cereal crops as measured from the weight of grain delivered to the silo, were obtained from the grower (Table 1). Yield maps were collected on Paddock 17 in 1996, 1998, 2004, and 2005 by the grower, using a harvester (Case IH 2188) and an AFS yield monitoring system (Case IH AFS system) linked to a differentially corrected GPS (for details see Rab *et al*. 2009). The harvester yield point data for each season were converted to a raster format and re-sampled to a 25 m by 25 m grid by kriging to make them comparable with the Landsat satellite data.

Biomass estimation

Landsat 5 images were used for the years between 1991 and 2004 when cloud-free images could be obtained close to crop anthesis, and a SPOT-2 image was used in 2005 (Table 1). All images were geometrically and radiometrically corrected before analysis. The location of each paddock boundary was identified on the satellite images, and the NDVI values were calculated within each 25 m by 25 m pixel for each year that a cereal crop was grown, using (after Goward *et al*. 1991):

$$R_{\mathrm{NDV}} = (\lambda_{\mathrm{NIR}} - \lambda_R)/(\lambda_{\mathrm{NIR}} + \lambda_R) \quad (1)$$

where $R_{\mathrm{NDV}} \in [-1,+1]$ is the normalised difference vegetation index, λ_{NIR} is the reflectance in the near-infrared spectral band (760–900 nm), and λ_R is the reflectance in the red spectral band (630–690 nm).

Traditional NDVI paddock zoning

Satellite-derived NDVI data from the different seasons were combined to develop maps of the standardised temporal mean NDVI (Larscheid and Blackmore 1996):

$$\overline{r}_i = \left(\sum_{t=1}^{n} r_{i,t} \right) / n \quad (2)$$

Table 1. Crop rotations and paddock yields (t/ha) for each of the paddocks shown in Fig. 5*a*; and the date of satellite image acquisition
Rotations: W, wheat; PL, pasture legume; B, barley; L, lentils; OM, oats and medic; C, canola; CF, chemical fallow; FP, field peas; PG, pasture grass

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
# 13	FP	W	PL	B	PL	W	–	PL	W	W	L	W	W	OM	W
Yield									2.15	1.65	0.16		1.76		1.5
# 14	PL	PL	W	PL	C	W	–	W	L	W	OM	W	W	OM	W
Yield								1.2	1.2	3.0			2.47		2.3
# 15	PG	OM	CF	W	FP	W	–	W	W	L	W	OM	L	W	OM
Yield				0.6	1.2				2.9	0.9	2.24		1.1	1.0	
# 17	FP	B	PL	B	PL	B	L	W	OM	W	OM	W	OM	W	B
Yield							0.8	2.0		2.8					1.7
# 18	B	PL	B	FP	W	PL	–	B	OM	W	L	W	W	L	W
Yield								0.7		2.98	0.45		2.63		1.4
# 35	B	PG	B	PL	B	PL	–	OM	W	L	W	PL	W	OM	W
Yield									3.0	0.8	2.85		2.64		2.4
# 36	FP	W	PL	W	PL	W	–	W	OM	C	W	L	W	L	W
Yield								1.7		1.5	3.19		2.54	0.3	2.5
# 39	PL	FP	W	FP	W	PL	–	W	PL	W	L	W	W	B	OM
Yield								1.1		2.85	0.12		2.3		
# 40	PL	W	FP	W	FP	W	–	W	OM	W	OM	W	W	OM	W
Yield				0.6	1.2			0.7		2.34			2.03		1.3
# 41	FP	W	CF	W	OM	PL	–	W	L	W	OM	W	W	L	W
Yield								2.0	1.15	3.1			2.7	0.2	2.0
# 42	B	PL	B	FP	L	W	–	W	OM	L	W	OM	L	W	OM
Yield				0.2	1.6			2.4		0.9	3.34		1.9	1.6	
# 44	W	PL	B	FP	W	PL	–	W	L	W	OM	W	W	OM	W
Yield								0.8	0.4	3.0			2.6		2.0
								Date of satellite image acquisition							
	11.x	11.ix	16.x	19.x	–	–	27.x	28.ix	–	25.viii	29.ix	18.x	12.x	27.viii	10.x

where $\overline{r}_i$ is the standardised mean NDVI value at point *i* over *n* years, and $r_{i,t}$ is the standardised NDVI reflectance value (%) at point *i* in year *t*:

$$r_{i,t} = (R_{i,t}/\overline{R}_t) \times 100 \quad (3)$$

where $R_{i,t}$ is the NDVI value at point *i* for year *t*, and $\overline{R}_t$ is the average NDVI value for the whole paddock for that year (= (1/m) $\sum_i R_{i,t}$ $i = 1, \ldots, m$).

Satellite-derived NDVI data from different seasons are combined to develop maps of the standardised temporal variability in NDVI (Larscheid and Blackmore 1996):

$$\sigma_i = \left(\left(\sum_{t=1}^{n} (r_{i,t} - \overline{r}_i)^2 \right) / n \right)^{0.5} \quad (4)$$

where σ_i is the standard deviation of the standardised data at point *i* over *n* years.

Simulated yield mapping data standardisation

For simulated yield mapping, it is necessary to consider all the causes of paddock variation, including those from 'good' or 'bad' overall weather conditions in different seasons. This variation can be defined as the difference between the average yield values between 2 years in the same paddock and is referred to as the inter-year offset (Blackmore *et al*. 2003). However, when the data are standardised according to Eqn 3, this inter-year offset is removed. To retain this offset in the simulated yield mapping, non-standardised NDVI values can be used. However, we chose to use the following globally standardised NDVI (GS-NDVI):

$$g_{i,t} = (R_{i,t}/\overline{G}) \times 100 \quad (5)$$

where $g_{i,t}$ is the GS-NDVI value at point *i* in year *t*, $R_{i,t}$ is the NDVI value at point *i* for a particular year *t*, and $\overline{G}$ is a paddock's global (spatio-temporal) average NDVI estimated as:

$$\overline{G} = \left(\sum_{t=1}^{n} \sum_{i=1}^{m} R_{i,t} \right) / (n \times m) \quad (6)$$

We used the standardisation in Eqn 5 for two reasons. The first is that the GS-NDVI values retain the inter-annual patterns (offset) that are present in the original NDVI values. The second is that scaling a paddock's raw NDVI values by its global average NDVI allows the NDVI values to be expressed relative to a point in the centre of their paddock-specific distribution. This is referred to in imagery analysis as the 'departure from average vegetation greenness' (e.g. Burgan *et al*. 1996), and enables one to assess how much an individual year's NDVI deviates from this 'typical' average NDVI value for a paddock. The use of GS-NDVI in subsequently deriving its relationship with yield (see next section) allows interpretation of the rate of biomass conversion to yield (the regression coefficient β) as resulting from 1% change in NDVI relative to this 'typical' global average for the paddock, rather than from one unit of change in the raw NDVI. This, without affecting the model fit, enables the value

of β (rate of biomass conversion to yield) to be more objectively compared across different paddocks relative to their individual 'typical' global average NDVI.

Simulated yield analysis

We suggest that the globally standardised NDVI (GS-NDVI) data $g_{i,t}$ can be converted into simulated yield data by modelling the relationship between the farmer's average paddock yield and the corresponding average paddock GS-NDVI value over a sufficiently large number of years. This empirically derived relationship, which we observed to be close to linear in each of the 12 paddocks we studied, provides an estimate of an individual paddock's expected efficiency of converting biomass to yield.

The relationship between the farmer's observed average paddock yield Y_t (t/ha) in year t and the corresponding paddock-level average GS-NDVI g_t, was modelled for each paddock using the linear regression equation:

$$Y_t = \beta g_t + \alpha + \varepsilon_t \; t = 1, \ldots, n \tag{7}$$

where β and α are, respectively, the slope and the intercept of the underlying linear regression relationship for the paddock in question, ε_t is the deviation of Y_t from the assumed linear relationship ($\beta g_t + \alpha$), and n is the number of years for which (Y, g) data are available for the paddock. The slope β denotes the expected increase in yield for a unit increase in GS-NDVI (biomass), or the rate at which biomass gets converted into yield. The value of n in this study varied from 2 to 5 for different paddocks. Equation 7 is fitted by ordinary least-squares (LS) to obtain LS estimates 'b' and 'a' of β and α, respectively. The coefficient of determination (R^2) was used to assess the adequacy of the fitted linear regression model.

The fitted linear regression model:

$$y = b\,g + a \tag{8}$$

is then used to obtain the simulated yield y_{it} at an appropriately defined land-unit i of certain size in the paddock in season t, from the available knowledge of the corresponding GS-NDVI value $g_{i,t}$ for that unit in the paddock.

This approach makes several assumptions, including the following.

(*a*) The relationship between a farmer's average paddock-level yield and GS-NDVI is linear.
(*b*) The ε_t are uncorrelated with each other and have a normal distribution with a constant variance.
(*c*) A sufficiently long time series of data on farmer's paddock-level annual average yield and GS-NDVI is available.
(*d*) The temporally derived annual-averages based relationship (Eqn 8) will also hold true spatially across the entire paddock for any individual year. In other words, had we fitted model (Eqn 7) spatially for the paddock in question for each year separately using grid-level (Y, g) data, then the resulting spatially derived estimates 'b' and 'a' for that paddock would not only be close to each other from year to year but also close to those as obtained from the temporally derived model (8) for that paddock.

For assumption (*a*), our data as well as other published findings (Smith *et al*. 1995) indicate that a linear relationship generally holds reasonably well. Significant departures from linearity, if they exist, can, however, be easily accommodated by including additional polynomial terms of appropriate order in Eqn 7 to obtain better prediction of yield.

For assumption (*b*), any correlation among residuals, arising from dependence in yield from successive years on a paddock, can be accounted for by using a generalised least-squares (GLS) approach. An appropriate transformation of data may help in situations where the residuals do not follow a normal distribution with a constant variance. Assumption (*c*) is important to obtain a realistic estimate of the 'typical' global average NDVI for the paddock in question. A minimum of ~15 years of data would be ideal not only to achieve this, but also to obtain reliable estimates of regression coefficients, their precision, and of the coefficient of determination R^2 to assess the adequacy of the fitted regression model. The limited datasets we had access to of measured paddock yield that spanned only 3–5 years, were not sufficient to test these assumptions, including the assumption made in (*d*). We have used this limited dataset only to illustrate the methodology. Consequently, the results presented later for the 12 paddocks may not be very accurate. However, given the success of a similar approach at the regional scale, this is more a reflection of the lack of sufficient data than a result of the inadequacy of the methodology. The results still provide some idea of the inherent spatial and temporal variability in the 12 paddocks using the proposed SYM approach.

Simulated temporal mean yield and temporal yield variability mapping

Using the simulated yield data $y_{i,t}$, the simulated temporal mean yield, as in the traditional NDVI paddock mapping approach, was derived as:

$$\overline{y}_i = \left(\sum_{t=1}^{n} y_{i,t} \right) / n \tag{9}$$

where $\overline{y}_i$ is the mean simulated yield (t/ha) at point i over n years. Likewise, the temporal variability in simulated yield was calculated as:

$$\sigma_{y_i} = \left(\left(\sum_{t=1}^{n} (y_{i,t} - \overline{y}_i)^2 \right) / n \right)^{0.5} \tag{10}$$

where σ_{y_i} is the standard deviation (t/ha) of the simulated yield at point i over n years.

Economic evaluation

If nitrogen fertiliser (N) is applied at a uniform rate to a paddock, then economic losses occur from two sources: (*i*) wasted N in areas where the application rate is greater than crop requirement, unless it remains available for a subsequent crop; and (*ii*) reduced yield where the application rate is less than crop requirement. As an example of how growers might use a whole-farm simulated yield map, the economic benefits of using variable rate technology (VRT) have been estimated based on the variability in yield found in the whole-farm simulated yield

map. The amount of wasted N at a point has been calculated as the amount of N applied at the uniform rate that is in excess of the rate required to match the yield on the simulated yield map at that point. The N required is calculated using a simple nitrogen budget equation:

$$N_i = \left(\frac{\bar{y}_i \times P \times P_N}{G} - S_N \right) \times E \qquad (11)$$

where N_i is the crop nitrogen fertiliser requirement at point i (kg/ha), $\bar{y}_i$ is the mean crop yield from the simulated yield map at point i (kg/ha), P is the proportion of grain protein (0.11), P_N is the proportion of N in protein (0.172), G is the proportion of N taken up by the plant that is transferred to the grain protein (0.75), S_N is the amount of soil N available to the crop that has been assumed spatially constant at 30 kg/ha, and E is to allow for efficiency of applied N (1.25). The cost of wasted N has been calculated using the current value of A$2/kg N.

In locations where the uniform N rate is less than crop requirement, the yield reduction has been estimated as the difference between the yield at a point from the simulated yield map, and the potential yield obtainable from the uniform rate of N applied, using the same equation as N requirement:

$$Y_{\text{pot}} = \left(\left(\frac{N_{\text{rate}}}{E} \right) + S_N \right) \times \frac{G}{P \times P_N} \qquad (12)$$

where Y_{pot} is the potential yield obtainable (kg/ha) when nitrogen is applied at a rate of N_{rate} (kg/ha), and the other parameters are defined above. The cost of this yield loss has been calculated using a grain price of A$350/t but has not been adjusted for any variability in protein value. However, since more N than the uniform rate would be required to realise this extra yield, the total cost of the yield penalty has been reduced by the cost of the extra N requirement.

In reality, growers do not apply the same N-rate to every paddock. Therefore, to make this economic analysis more realistic at a whole-farm scale, and conservative, it has been assumed that the uniform N-rate applied to each paddock is the rate required to match the whole paddock's average yield from the simulated yield map. Therefore the economic differences calculated here between uniform and variable rate N applications are only the result of the spatial variation within each of the paddocks.

Results

Comparison of actual and simulated yield maps in paddock 17

Observed cereal yields (wheat and barley) were mapped by the grower in Paddock 17 in 1996, 1998, 2004, and 2005. The map of temporal mean grain yield, measured over these seasons, is shown in Fig. 1*a*. The mean simulated yield map for Paddock 17, which has been calculated for the years 1992, 1994, 1998, 2000, 2004, and 2005 when cereals were grown and cloud-free satellite images could be obtained, is shown in Fig. 1*b*. Both of these maps have been segregated into the same yield classes of 0.25 t/ha.

The accuracy of the simulated yield mapping was evaluated by the magnitude of difference between the 4 years' mean observed yield map and the 6 years' mean simulated yield map. This difference, D = Observed yield – Simulated yield, was calculated on a 25-m grid and ranged from –2.3 to 1.4 t/ha across this paddock. This range was divided into 3 categories: accurate prediction when $|D| < 0.3$ t/ha, moderate prediction when $0.3 < |D| < 0.8$ t/ha, and poor prediction when $|D| > 0.8$ t/ha. The accurate prediction category accounted for 56% of the paddock area (93 ha), while 33% (55 ha) and 11% (18 ha) were in the moderate and poor categories, respectively (Fig. 1*c*). Most of the area of moderate and poor prediction category was found along the southern edge of the paddock where actual mean yields were lowest. With the difference, expressed as percentage $D_{\%}$ of the observed yield, O, $[D_{\%} = (|D|/O) \times 100]$, 63% of the area had <20% difference from the observed mean yield, and 15%, 10%, 7%, and 6% of the area had 20–40, 40–60, 60–80, and >80% difference, respectively (Fig. 2*a*). In the majority of the paddock (65% or 108 ha), yields were over predicted by the simulated yield mapping (Fig. 2*b*).

For 1998, 2004, and 2005, both observed yield maps (Fig. 3*a*) and satellite imagery were available. The individual

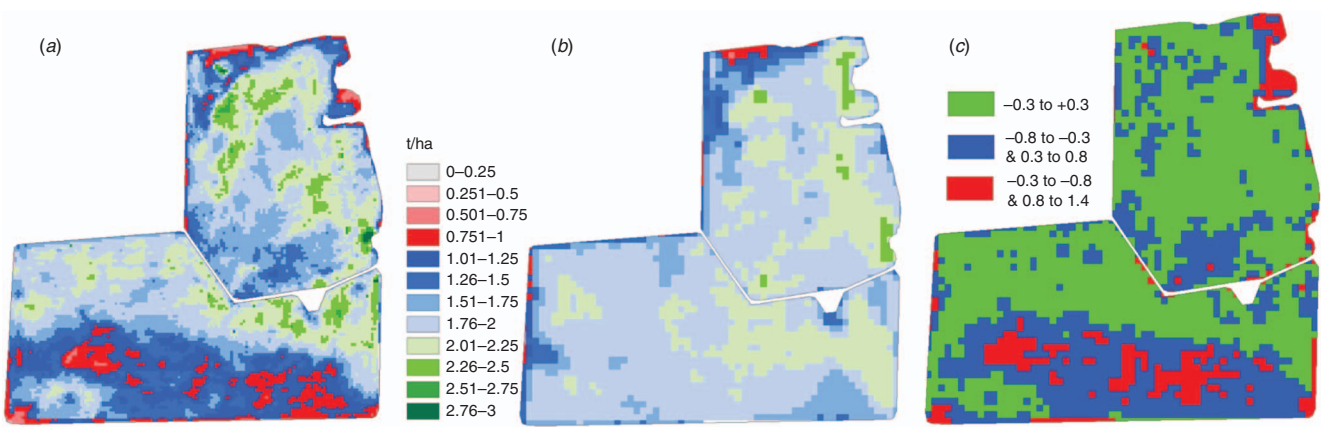

Fig. 1. Map of Paddock 17 showing spatial variation in cereal yield using the same legend (t/ha) for (*a*) mean observed yield (1996, '98, '04, '05) and (*b*) mean simulated yield (1992, '94, '98, '00, '04, '05). The magnitude of yield difference at each point between these two maps, grouped into 3 categories (t/ha), is shown in (*c*) the residual map.

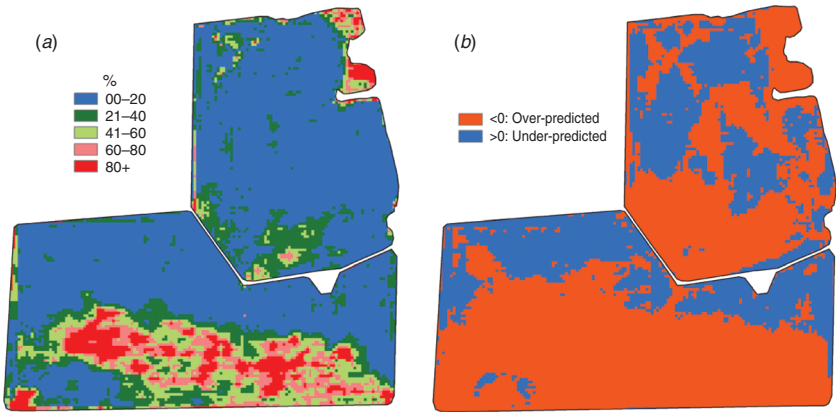

Fig. 2. Maps of Paddock 17 showing the spatial variation in the residual between the mean observed yield map (1996, '98, '04, '05) and the mean simulated yield map (1992, '94, '98, '00, '04, '05), expressed as: (*a*) a percentage of the observed yield map, and (*b*) an over or under prediction by the simulated yield map.

satellite images were converted into annual simulated yield maps (Fig. 3*b*) and the map of the values of D (kg/ha) between these and the corresponding observed yield maps is shown in Fig. 3*c*. The relationships between observed yield and simulated yield at each pixel for 1998 and 2005 are shown in Fig. 4. The linear regression relationships for the points in the accurate prediction category ($|D| < 0.3$ t/ha) are for 1998: intercept = 0.12 (s.e. = 0.079), slope = 0.94 (s.e. = 0.035), adjusted $R^2 = 82\%$; and for 2005: intercept = 0.06 (s.e. = 0.056), slope = 0.97 (s.e. = 0.033), adjusted $R^2 = 82\%$. In both years the percentage difference, $D_{\%}$, in most of the paddock (72% in 1998 and 73% in 2005) has simulated yields that are within 30% of the observed yield.

The residual map for 2004 shows a very different pattern, with most of the paddock falling into the category of D <–0.8 t/ha. The cause of this is that at the time of NDVI image acquisition (approx. anthesis), the crop was growing well and yields in the range of 1–2 t/ha were expected across most of the paddock. But in mid-October, after NDVI image acquisition, there were 2 days at temperatures of ~40°C, followed by a frost. This had the effect of reducing the yield across most of the paddock to <0.5 t/ha and 40% of the paddock yielded <0.1 t/ha.

Excluding the unpredictable results of 2004 from both the mean observed yield map (i.e. leaving 1996, 1998, and 2005) and the mean simulated yield map (i.e. considering only 1992, 1994, 1998, 2000, and 2005) reduced the 0–20% error category from 63% to 53% of the paddock area, but increased the 20–40% error category from 15% to 41% of the area. This means that 94% of the total paddock area had an error of <40%. The paddock area in the 40–60, 60–80, and >80% error categories was 4%, 1%, and 1%, respectively (Fig. 5*a*). Not surprisingly, removing 2004 from the analysis also changed the proportion of the paddock that was over predicted by the simulated yield mapping from 65% to 28% (Fig. 5*b*).

Whole-farm analysis using traditional NDVI paddock zoning

The whole farm covered 12 paddocks (Fig. 6*a*) that ranged in size from 57 to 167 ha (Table 2), and made a combined cropped area of 1348 ha. The map of temporal mean NDVI, calculated from Eqn 2, for the 12 paddocks, is presented in Fig. 6*b*. Thresholds have been set using quantiles of one-third and two-thirds for the frequency distribution of $\bar{r}_i$ for each paddock. Therefore, each paddock is divided into 3 equal area zones of high, medium, and low mean NDVI values.

Similarly the map of temporal stability of the mean NDVI at each point in the 12 paddocks, calculated from Eqn 4, is presented in Fig. 6*c*. Thresholds for temporal stability are defined as above or below the median of the frequency distribution of σ_i for each paddock. Therefore, each paddock is divided into 2 equal area zones of temporal stability: variable or stable.

Whole-farm analysis using simulated yield maps (SYM)

The relationship between the farmer's average paddock yields and the corresponding average paddock GS-NDVI values for each paddock (Eqn 7) is presented in Table 3. The number of data points available to model this relationship in each paddock varied from 2 to 5, depending on the number of seasons for which both the average paddock cereal yields and a cloud-free Landsat image could be obtained. The slope and intercept varied from 0.004 to 0.055, and –3.17 to 1.16, respectively (Table 3). The R^2 values, excluding Paddock 35 that had only 2 points, varied between 0.64 and 0.99. Since these estimates are based on a very small number of data points they are likely to be highly biased and need to be treated only as rough approximations to their true values.

These individual paddock relationships were applied to GS-NDVI data for each season. The mean of each point over the available seasons has been calculated to produce a whole farm map of temporal mean simulated yield (Fig. 7*a*). This map has been categorised into intervals set at 0.25 t/ha, as this provides a reasonable level of difference that would be of economic importance to growers, although any interval could be chosen. Across the whole farm, constituting the 12 paddocks, the mean simulated yields varied from <0.25 up to 3 t/ha. The amount of spatial variability in mean simulated yield is quite different across the paddocks, from almost uniform paddocks to quite spatially variable paddocks.

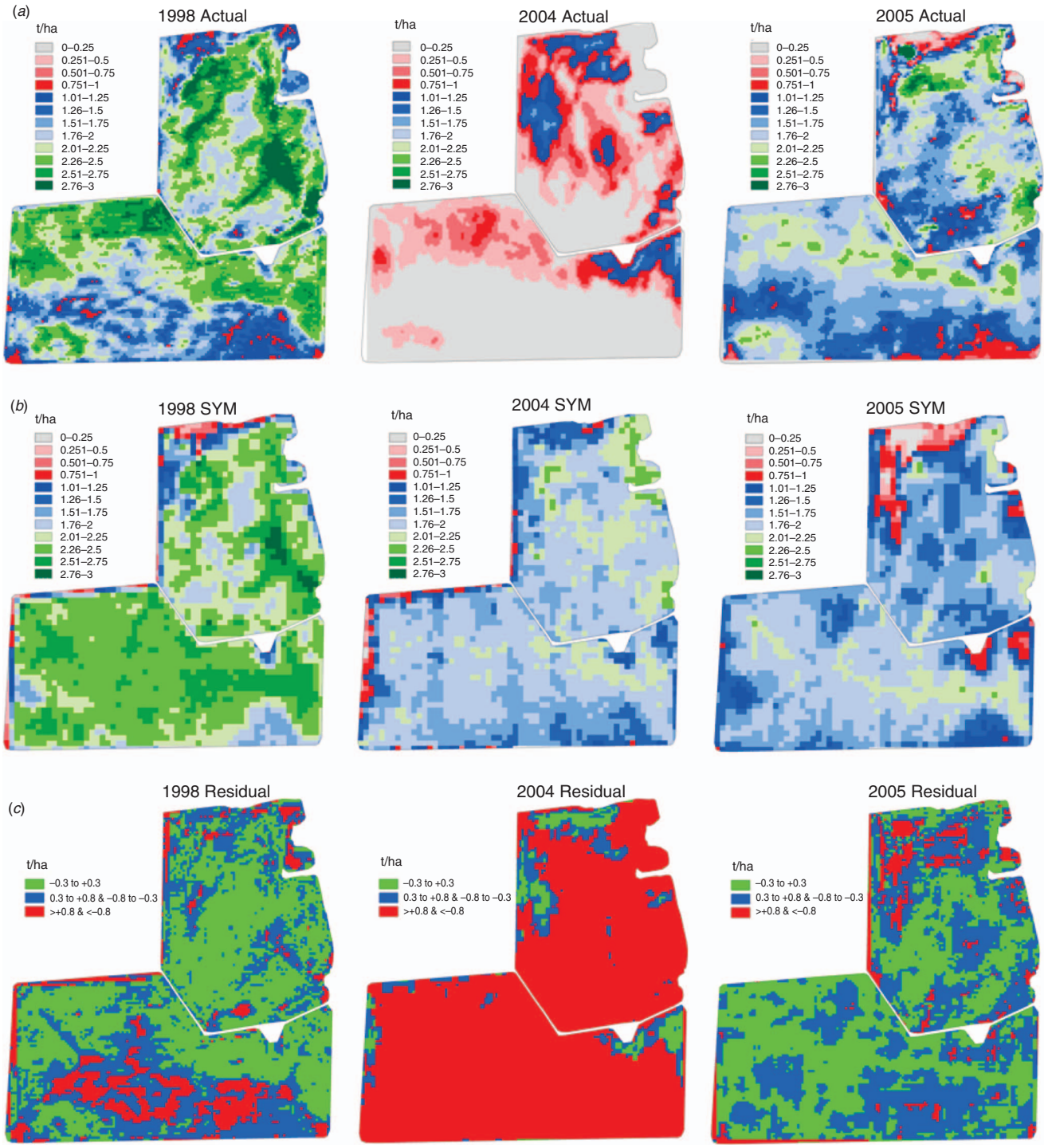

Fig. 3. Maps of Paddock 17 showing three individual years of data (using the same categories as Fig. 1) showing the spatial variability in: (*a*) observed yield, (*b*) simulated yield, and (*c*) differences between observed and simulated yields.

The variability between the individual seasons that make up the mean simulated yield at each point is measured by the temporal variability (Eqn 10). This has been mapped across the 12 paddocks (Fig. 7*b*) and is also divided into 0.25 t/ha classes. Temporal variability values ranged from <0.25 to 1.83 t/ha. The temporal variability for the major proportion of individual paddocks lies within one or two colour bands, i.e. within 0.5 t/ha difference.

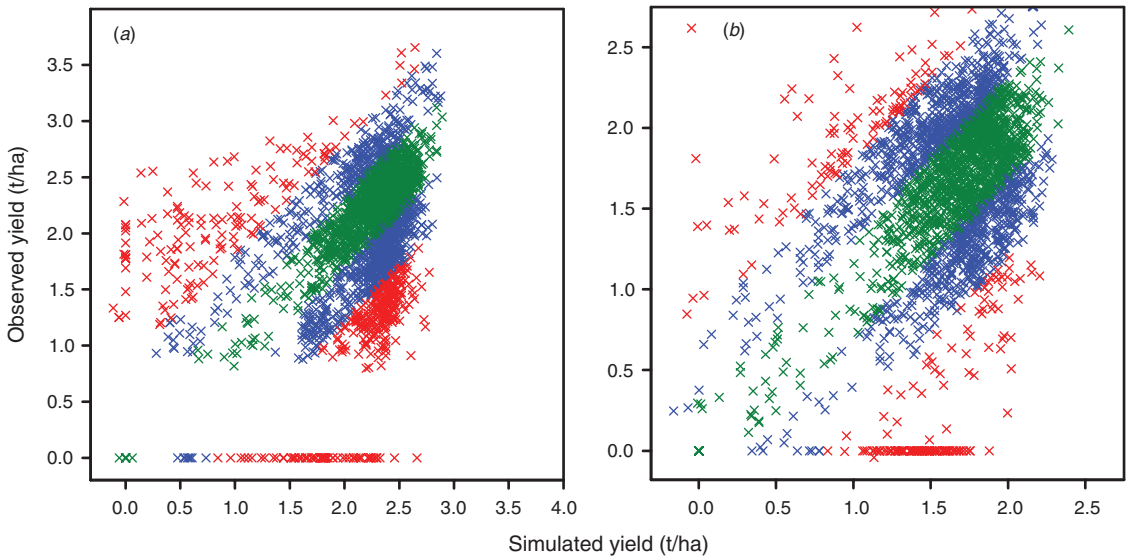

Fig. 4. Relationship for all 25-m^2 pixels in Paddock 17 between the observed yield and simulate yield for (*a*) 1998 and (*b*) 2005. Points where the observed-simulated yield, D, have accurate prediction when $|D| < 0.3$ t/ha (green), moderate prediction when $0.3 < |D| < 0.8$ t/ha (blue), and poor prediction when $|D| > 0.8$ t/ha (red).

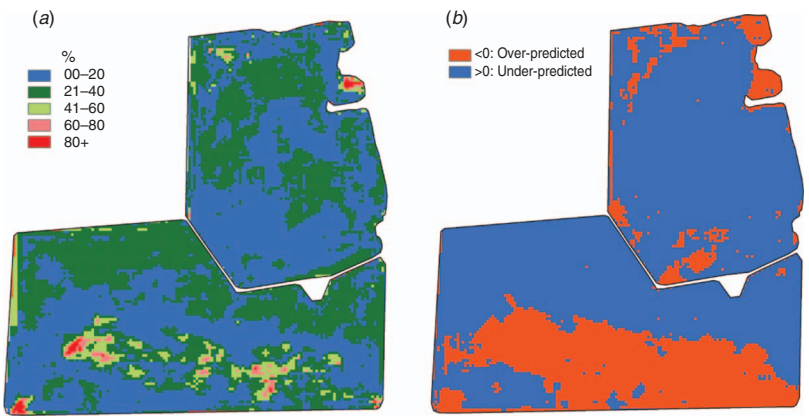

Fig. 5. Maps of Paddock 17 excluding data from 2004 showing the spatial variation in the residual between the mean observed yield map (1996, '98, '05) and the mean simulated yield map (1992, '94, '98, '00, '05), expressed as: (*a*) a percentage of the observed yield map, and (*b*) an over or under prediction by the simulated yield map.

Whole-farm economic analysis

The rate of N applied, and the estimated economic losses across the 12 paddocks as a result of over or under applying N when using a uniform N application rate, are shown in Table 2. The uniform rates for each paddock have been set to match each paddock's average yield less 30 kg N/ha contribution from soil N. Across the 1348 ha making up the 12 paddocks, the total financial loss due to applying the uniform N rates is estimated to be A$28 686 per year. Of this, A$4381 is due to wasted N and A$24 305 is due to the estimated loss of potential yield. In individual paddocks the loss ranges from A$684 to A$5097 per year. Across all the paddocks the loss per ha per year is estimated to be A$21, which ranges in different paddocks between A$6 and A$56/ha.year. This range in potential economic benefit is similar to the benefits derived by others nationally and internationally (Godwin *et al*. 2003; Brennan *et al*. 2007; Robertson *et al*. 2007).

Discussion

Comparison of actual and simulated yield maps in Paddock 17

The underlying assumption of all approaches that use remotely sensed data to estimate crop yield variability is that the remotely sensed data maps at the time of collection have a spatial distribution similar to the actual yield. Since the approach proposed in this paper converts the NDVI data into a simulated yield map, the validity of this assumption for

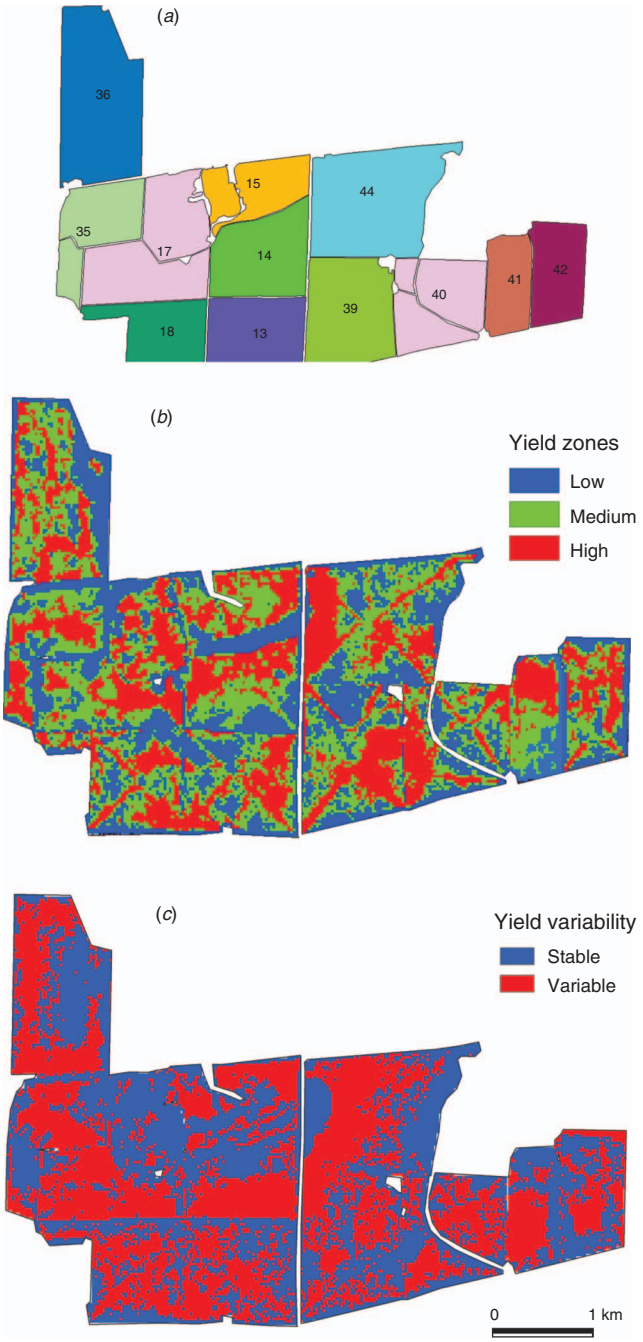

Fig. 6. Whole-farm maps showing (*a*) the paddock locations, and the results of the traditional NDVI analysis approach classified for: (*b*) the standardised temporal mean NDVI and (*c*) the standardised temporal variability in NDVI.

simulated yield mapping cannot only be compared qualitatively but also quantitatively. How well NDVI represents the actual yield differs from season to season and is particularly dependent on how environmental conditions after anthesis alter the final yield in part, or the entire paddock, compared with what was expected at the time of satellite image acquisition. This has been tested in the paper for one paddock (Paddock 17) by measuring the difference between harvester yield maps and the maps of simulated yield generated from the GS-NDVI data.

There was reasonable agreement when all available data were used, with 63% and 78% of the paddock area having >20% and 40% deviation from the observed mean yield, respectively. Many factors can cause the differences between the mean simulated yield map and the observed mean yield map. One factor is the limited availability of actual yield maps for comparison. In this analysis, the mean map included four seasons, one of which (2004) had negligible yield. The corresponding SYM analysis included six seasons, three of which corresponded to seasons when actual yield maps were collected. We therefore expect that more realistic and precise estimates could be obtained if more years of data and a great number of concurrent years with yield and NDVI maps were available.

Comparing the two seasons that had normal weather conditions and both satellite and harvester data available (1998 and 2005), clearly the model does not account for all of the observed spatial variation. As expected, the regression in both years for the points that are in the accurate prediction category ($|D| < 0.3$ t/ha) has slopes that are close to 1, and lines that pass close to the origin. For the points in the moderate prediction category ($0.3 < |D| < 0.8$ t/ha), if divided into areas of over and under prediction, the regression analysis still has a slope close to 1 with a positive or negative offset according to whether it is over or under predicting. The physical causes of these offsets require further research. However, the simulated yields are within 30% of the observed yield in 72% of the paddock area in 1998, and 73% of the paddock in 2005. These results compare favourably with those obtained from simulating paddock yield variability from crop modelling. In an Australian study the yield in a 100-ha paddock was estimated at 13 523 points using the APSIM crop model, with soil depth measured at 375 points as the spatial variable (Brennan *et al*. 2007). Using a 30% error threshold, Brennan *et al*. (2007) found that 78% and 63% of the paddock area in the 1997 and 2000 sorghum crops respectively, were estimated within this error. However, unlike the simulating yield mapping from NDVI data, a modelling approach requires detailed on-ground measurements to be made of the spatial variable, as well as daily weather records. Brennan *et al*. (2007), report that the farmer was comfortable with the results from the analysis of spatially variable nitrogen management options based on these data.

Other factors that can cause differences between the mean simulated yield map and the observed mean yield map include environmental events that occur after NDVI images are acquired, such as frost or haying off due to lack of finishing rainfall. These factors are only likely to be partially, or not at all, accounted for by the regression procedure used in the SYM analysis. This occurred on Paddock 17 in 2004 when virtually the whole paddock was wiped out after anthesis due to poor weather conditions. Excluding seasons where yield results are clearly outliers should improve the precision in converting NDVI data to simulated yield. In Paddock 17 the simulated yield mapping tended to over-predict the yield for most of the paddock when including the 2004 data, but under-predicted the

Table 2. Estimated economic losses as a result of over or under application of nitrogen fertiliser when using a uniform application rate to match the paddock average yield

Paddock	Paddock size (ha)	Applied N rate (kg/ha)	Paddock wasted N (A$)	Paddock yield penalty (A$)	Total paddock loss (A$)	Loss per ha (A$/ha)
13	124	9	52	705	757	6
14	100	22	505	2325	2830	28
15	98	0	0	1224	1224	12
17	167	19	827	4270	5097	31
18	109	9	341	2421	2763	25
35	90	37	371	1688	2059	23
36	165	42	555	2525	3080	19
39	105	12	309	1521	1830	17
40	102	0	0	684	684	7
41	57	23	326	1537	1863	33
42	57	36	505	2685	3190	56
44	174	23	590	2720	3309	19
Total	1348		4381	24 305	28 686	21

Table 3. Estimates of slope, intercept, and R^2 from fitting linear regression of grower's average paddock yields on average paddock GS-NDVI (Eqn 7)

Paddock	Years	Slope	Intercept	R^2
13	3	0.004	1.156	0.99
14	4	0.026	–0.647	0.69
15	3	0.026	–1.561	0.86
17	3	0.030	–1.151	0.94
18	4	0.041	–2.604	0.83
35	2	0.020	0.390	1.00
36	3	0.016	1.005	0.64
39	3	0.029	–1.310	0.90
40	5	0.020	–0.755	0.83
41	4	0.026	–0.350	0.78
42	3	0.055	–3.169	0.79
44	4	0.032	–1.303	0.83

yield in most of the paddock when 2004 was excluded from the analysis. However, the simulated yield mapping consistently over-predicted the southern edge of the paddock. This area corresponds precisely to a region of sodic duplex soil with subsoil constraints. This suggests that with greater development of the technique and with paddock soil maps, it may be possible to adjust the conversion relationships in areas that are known to have difficulty converting biomass at anthesis into yield. The cause and effect of soil constraints on yield production in this paddock have also been investigated and reported in this issue by Armstrong *et al*. (2009), Rab *et al*. (2009), and Robinson *et al*. (2009).

Whole-farm analysis using traditional NDVI paddock zoning

In the whole-farm map developed by the traditional NDVI approach (Fig. 6*b*), results in each paddock were divided into areas of high, medium, or low performance relative to each season's paddock average. This is achieved by standardising the data to remove the inter-year offset (Blackmore *et al*. 2003). Thus, areas in the high mean NDVI category are expected to have greater biomass than the paddock average in most seasons, and *vice versa*. This is useful information for identifying the best and worst producing areas in each paddock. Once these areas have been identified, an experienced grower or agronomist can investigate the causes of this variability, undertake on-farm experiments, and adjust management practices accordingly.

However, since every paddock is divided into equal high, medium, and low areas it is not possible to know whether a high area in one paddock has more or less biomass (and thus potential yield) than an area marked high in another paddock, and *vice versa*. This means that in one paddock the high, medium, and low mean NDVI areas could all be quite high yielding and possibly without much yield difference between the areas, while another paddock could be generally low yielding but with a much greater yield difference between the high and low areas.

This is a considerable drawback to interpreting traditional NDVI paddock zoning as it means that when time and resources are limited it is not possible to prioritise which paddocks should be the focus for developing subpaddock management. More importantly it is not possible to estimate the economic difference between subpaddock areas, and therefore not possible to compare potential remedies for their cost effectiveness without more detailed investigation.

Whole-farm analysis using simulated yield maps (SYM)

In the proposed approach, subpaddock simulated yield maps have been produced by calibrating NDVI data with the growers' observed paddock average yield. Growers know the average paddock yield for each season quite accurately from the weighbridge dockets received before unloading grain at a silo. An important benefit of this method is that it objectively accounts for differences between paddocks based on their 'typical' average GS-NDVI and yield potential. These differences in the efficiency of converting biomass into yield can be caused by different microclimate or soil characteristics as a result of changes in the underlying geological or terrain properties.

Visually comparing the whole-farm map developed by the proposed SYM approach (Fig. 7*a*) with that produced using

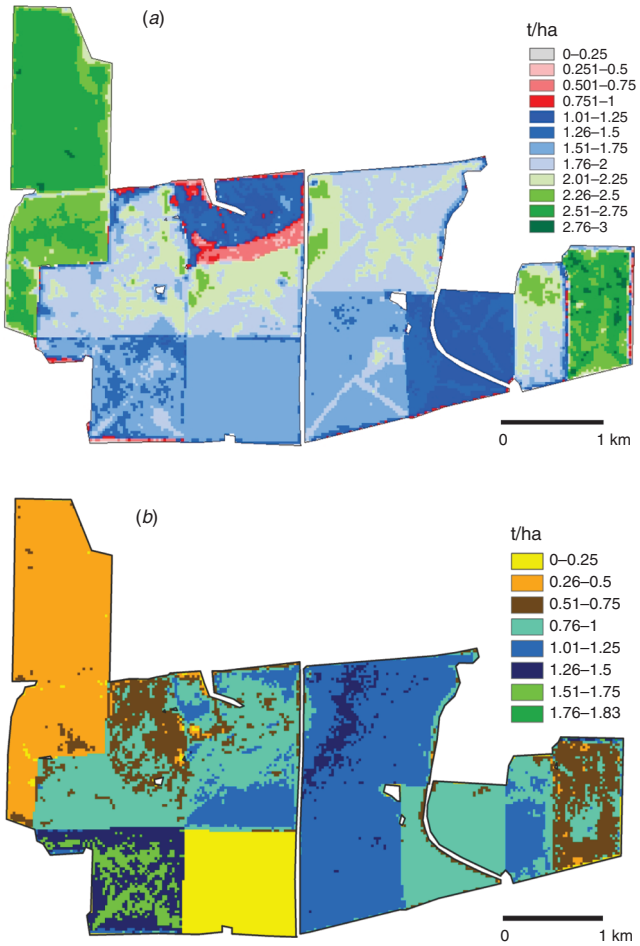

Fig. 7. Whole-farm map showing (*a*) areas of each paddock classified for temporal mean simulated yield, and (*b*) areas of each paddock classified for temporal variability in simulated yield.

traditional NDVI mapping (Fig. 6*b*), it can be seen that in some paddocks a similar spatial pattern can be detected in both maps (e.g. Paddock 35), while in other paddocks there is no resemblance between the two maps (e.g. Paddock 13). The difference between the two maps is the consequence of the SYM approach developing an individual paddock relationship to account for variation in the conversion of NDVI into yield. The use of individual paddock relationships is important as it has generally been found that algorithms linking remotely sensed data to actual biophysical parameters seem only to work at the single-paddock level rather than across entire regions (Lamb 2000).

In this whole-farm study, cereal crops were grown in rotation with pulses and pastures, and paddock yield records were not available before 1998. The calibration functions were therefore based only on 2–5 points, with R^2 ranging from 0.64 to 0.99. More data points for these regression analyses would be desirable to obtain reliable estimates of regression coefficients, their precision, and of the coefficient of determination, R^2. Paddocks 36, 35, and 42 have the highest mean simulated yields, with yields across these paddocks generally being >2.25 t/ha, while the lowest yields were found in Paddocks 15 and 40. This has been confirmed by the grower who considers, for example, Paddock 36 as 'consistently high yielding mainly consisting of sandy loam', Paddock 42 as 'the best paddock with a mixture of sandy loam and clay loam', and Paddock 40 as 'average to low yielding'.

The individual paddock calibration used in the SYM approach means that some paddocks show very little variability in mean yield (i.e. mostly a single colour). This is in contrast to the map developed from the traditional NDVI approach in which every paddock has an equal area of high, medium, and low. When a paddock in the simulated yield map shows no apparent spatial variation in mean yield, this reflects a <0.25 t/ha mean yield variation (the category spacing selected for this map) across the paddock. However, the information on where higher or lower yielding areas are located within a paddock is not lost and can be obtained by decreasing the category size. Paddock 13 had the least spatial variability and this has also been confirmed by the grower who considers Paddock 13 to be 'a very consistent clay loam across the entire paddock resulting in very uniform production'.

It was anticipated that when whole-farm yield maps were available, it would be possible to identify landscape features that relate to yield and run across paddock boundaries. However, such features are not very apparent on the whole-farm simulated yield map. The high yield found in Paddock 36 continues through Paddock 35, and the higher yielding area in the top of Paddock 14 also continues into Paddock 44. The low-yielding area in the bottom right-hand corner of Paddock 17 continues into the top of Paddock 18, and this is known to be associated with an area of gilgai soil feature.

The individual season maps are also used to produce maps of the temporal stability in the mean results at each point. The SYM approach enables each paddock to be divided into standard deviation values using a 0.25 t/ha class range. It can be observed that although there are marked differences between individual paddocks, the amount of estimated variability for most points is not great when measured in units of t/ha. This suggests that where differences in mean simulated yield are economically significant, variable management is likely to be possible. Paddocks that had the lowest spatial variability in mean simulated yield (e.g. Paddocks 13 and 36) also had the lowest spatial variability in temporal variation across the paddock. Although it is hard to see similarity in the spatial patterns in the mean simulated yield and the temporal variance, it would appear that in paddocks that have greater spatial variation in the mean simulated yield, higher yields are associated with lower temporal variation.

Whole-farm economic analysis

The simple example of uniform *v.* variable rate nitrogen application reported here appears sufficient to illustrate the potential for economic analysis using simulated yield maps. The economic losses from using uniform N applications have been estimated for both the wasted N in areas where it is applied above crop requirement, and the reduced yields where it is applied at less than crop requirement. With the assumptions used in this analysis, particularly the cost of N, the potential yield loss due to under supply of N is over 5 times the financial loss

due to wasted N. It is possible to reduce the total financial loss with a uniform application by applying higher N rates, thus decreasing any potential yield losses, but increasing the amount lost as unused N. This is why the environmental benefits can be significant when using a PA system in which the right amount of input is applied in each location. The example also illustrates that the financial losses due to uniform N management are very different in each paddock. Clearly, paddocks with less spatial variability have less financial loss when managed uniformly. The two most variable paddocks account for 30% of the total financial loss across the farm due to applying a uniform N rate, while the two most uniform paddocks accounted for only 5%. These estimates have been based on the uniform rate for each paddock being matched to the average simulated yield for the whole paddock. If a single uniform N rate is used across all the paddocks, the estimated losses due to uniform management are A$74/ha.year for a blanket rate to match a 3 t/ha yield and A$52/ha.year for a blanket rate to match a 2 t/ha yield. This financial analysis is included to demonstrate how simulated yield data can be used by growers to assist with financial planning. In a simple spreadsheet, growers can undertake a sensitivity analysis of the potential benefits of using variable or uniform application rates, taking into account different paddock requirements, and contributions from soil N. There is a potential to make such financial analysis more accurate if extra data exist. For example, with a farm soil map and field testing it may be possible to spatially vary the contribution of soil N, or nitrogen tie-up.

An assumption made in estimating the wasted N and the potential yield loss is that the yield map still represents the yield when N inputs are altered using VRT. However, yield maps themselves are initially generated from data during a period when the paddocks are managed under the traditional practice, usually uniform nutrient application over the whole or large part of the paddock. It is therefore possible that actual yields could be different when using VRT. This is of course true whether the maps are generated from header yield maps or simulated yield maps. This emphasises the need for growers to establish trial plots to monitor yield responses, as recommended for other zoning techniques. A greater discussion of N dynamics in Paddock 17 is found in Armstrong *et al*. (2009).

Although historic satellite data provide a longer record, and thus greater climatic differences, than current harvester yield maps, the results always need to be put in perspective of the prevailing weather conditions during the collection of those data. Greater discussion of modelling results for 119 years of weather data are found in Anwar *et al*. (2009).

The benefit of having whole-farm information on estimated yield variability needs to be contrasted with the approximate cost required to obtain a whole-farm simulated yield map. This is difficult to estimate as it depends on the cost of the required imagery, the size of the area to be examined, and the cost of skilled people to collate and interpret the data. However, we can estimate the satellite image costs for a particular grower take-up rate. For example, a Landsat image of 185 by 185 km costs approximately A$1300 and if 15 years of images, at or just before anthesis, were purchased to have sufficient seasons of data to cover different crop rotations, this would cost A$19 500. If we assumed a take-up rate for simulated yield mapping of 10% of the total 3 422 500 ha making up each Landsat scene, the raw image cost would equate to 5.7 cents/ha for the 10% take-up rate. Doubling the take-up rate would halve this cost. Clearly it could not be sold to growers by a commercial service at that price, as time for image processing and data collection needs to be added, as well as the other costs and a profit margin. However, it would appear that it could be offered as an economic proposition compared with the current commercial cost for an EMI survey of approximately A$4/ha plus travel, soil sampling, and mapping expenses (A. Rab, unpublished).

Uses for simulated yield maps

As well as enabling growers to evaluate the economic costs of spatial yield variability, a comprehensive set of yield maps is a useful resource to be used in combination with crop models. In Australia there is increasing interest by growers in using crop models to provide yield probability predictions based on historical weather data. These yield predictions can be used as a management tool to guide decisions about post-emergence N application. However, spatial application of crop models is severely limited by the amount of soil data available, particularly soil hydraulic properties. Areas of significant yield difference identified from simulated yield maps enable selection of the best locations and number of soil samples required to adequately characterise each paddock. Extending this concept further, a relatively unexplored use of yield maps is for inverse yield modelling. This technique uses multiple seasons of yield and weather data as inputs to an inverse crop model, which then predicts soil property maps, such as plant-available water. A model that could use data on spatial yield variability to map plant-available water would be a step towards providing the required information to run more detailed agronomic models over landscapes at high spatial resolution (Morgan *et al*. 2003).

Conclusions

Agricultural statistics have previously been used as the input for modelling the conversion of NDVI data into crop yield, although the smallest scale at which this has been reported was to predict average paddock yields (Dawbin *et al*. 1980). In this paper, agricultural statistics, in the form of growers' average paddock yields, have been used to develop the relationship between NDVI and subpaddock yield variability. The use of satellite data and agricultural statistics instead of field sampling to develop a relationship between remotely sensed data and yield variability achieves a considerable saving in effort and cost, making the technique commercially feasible. The simulated yield mapping approach taken here develops an individual relationship for each paddock for the conversion of biomass, measured by NDVI, into final yield. This is particularly important as good correlation between remotely sensed data and biophysical parameters seems only to work for single or localised paddocks (Lamb 2000). The proposed technique can therefore account for differences between paddocks due to general soil and site differences, and usual management practices. In this study it has been possible to construct a whole-farm (1348 ha) map of mean subpaddock yield variability. Because 15 years of historical satellite data are

available in Australia, a range of weather and rotation effects can also be accounted for in the simulated yield mapping technique, provided growers also have historical paddock yield records for these years. The accuracy of the technique for estimating the spatial distribution in cereal yields, evaluated on a paddock in north-western Victoria, is very similar to those found in predicting the spatial distribution of sorghum yields in a paddock in northern New South Wales using a crop model and soil depth measurements made at 375 points across the paddock (Brennan *et al*. 2007).

Wider evaluation of simulated yield mapping across different soil types and climate conditions is required. However, because the approach maps spatial variability as simulated yield it can enable growers to economically evaluate the potential benefits for managing spatial variability and significantly increase the rate of PA adoption. For those growers already convinced of the benefits of PA, simulated yield maps could provide useful information on where best to position on-farm trials or the number of detailed soil samples required for crop modelling tools designed to assist with in-season N application.

Future research should consider whether the simulated yield mapping technique can produce useful yield predictions using NDVI images taken earlier in the season to optimise current season nitrogen application rates, and the use of simulated yield maps for inversely predicting the spatial distribution of soil parameters for crop modelling.

Acknowledgments

This research was supported by funding from the Grains Research and Development Corporation through its Precision Agriculture Initiative (SIP09), and the Victorian Department of Primary Industries. The authors would like to thank members of the Birchip Cropping Group, Colin Aumann, Grant Boyle, Tony Fay, Janine Fitzpatrick, Ian Maling, Nick O'Halloran, Cressida Savige, and Cherie Reilly for their support and contributions to this paper. We would particularly like to thank Ian and Warrick McCelland for their support and enthusiasm throughout this project, and without whose generosity in time, equipment, and access to paddocks this project would not have been possible.

References

Anwar MR, O'Leary GJ, Rab MA, Fisher PD, Armstrong RD (2009) Advances in precision agriculture in south-eastern Australia. V. Effect of seasonal conditions on wheat and barley yield response to applied nitrogen across management zones. *Crop & Pasture Science* **60**, 901–911.

Aparicio N, Villegas D, Casadesus J, Araus JL, Royo C (2000) Spectral vegetation indices as nondestructive tools for determining durum wheat yield. *Agronomy Journal* **92**, 83–91.

Armstrong RD, Fitzpatrick J, Rab MA, Abuzar M, Fisher PD, O'Leary G (2009) Advances in precision agriculture in south-eastern Australia. III. Interactions between soil properties and water use help explain spatial variability of crop production in the Victorian Mallee. *Crop & Pasture Science* **60**, 870–884.

Ashcroft PM, Catt JA, Curran PJ, Munden J, Webster R (1990) The relation between reflected radiation and yield on the Broadbalk winter wheat experiment. *International Journal of Remote Sensing* **11**, 1821–1836. doi: 10.1080/01431169008955132

Basso B, Ritchie JT, Pierce FJ, Braga RP, Jones JW (2001) Spatial validation of crop models for precision agriculture. *Agricultural Systems* **68**, 97–112. doi: 10.1016/S0308-521X(00)00063-9

Bauer ME, Hixson MM, Davis BJ, Etheridge JB (1977) Crop identification and area estimation by computer-aided analysis of Landsat data. In 'Fourth Annual Symposium on Machine Processing of Remotely Sensed Data'. (Eds DB Morrison, DJ Scherer) pp. 102–112. (The Institute of Electrical and Electronics Engineers, Inc., Purdue University: West Lafayette, IN)

Bellairs SM, Turner NC, Hick PT, Smith CG (1996) Plant and soil influences on estimating biomass of wheat in plant breeding plots using field spectral radiometers. *Australian Journal of Agricultural Research* **47**, 1017–1034. doi: 10.1071/AR9961017

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. doi: 10.1016/S1537-5110(03)00038-2

Bochenek Z (2000) Operational use of NOAA data for crop condition assessment in Poland. In 'Remote sensing in the 21st Century: economic and environmental applications'. (Ed. JL Casanova) pp. 387–392. (A.A. Balkema: Rotterdam)

Brennan L, Robertson M, Brown S, Dalgliesh N, Keating B (2007) Economic and environmental benefits/risks of precision agriculture and mosaic farming. Rural Industries Research and Development Corporation, Publication No. 06/018, Barton, ACT.

Brouder S, Hofmann B, Reetz HF Jr (2001) Evaluating spatial variability of soil parameters for input management. *Better Crops* **85**, 8–11.

Bullock PR (1992) Operational estimates of western Canadian grain production using NOAA AVHRR LAC data. *Canadian Journal of Remote Sensing* **18**, 23–29.

Burgan RE, Hartford RA, Eidenshink JC (1996) Using NDVI to assess departure from average greenness and its relation to fire business. Gen. Tech. Report INT-GTR-333.Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Dawbin KW, Evans JC, Duggin MJ, Leggett EK (1980) Classification of wheat areas and prediction of yields in North-western New South Wales by repetitive Landsat data. *Australian Journal of Agricultural Research* **31**, 449–453. doi: 10.1071/AR9800449

Elliott GA, Regan KL (1993) Use of reflectance measurements to estimate early cereal biomass production on sandplain soils. *Australian Journal of Experimental Agriculture* **33**, 179–183. doi: 10.1071/EA9930179

Genovese G, Vignolles C, Negre T, Passera G (2001) A methodology for a combined use of normalised difference vegetation index and CORINE land cover data for crop yield monitoring and forecasting. A case study on Spain. *Agronomie* **21**, 91–111. doi: 10.1051/agro:2001111

Godwin RJ, Richards TE, Wood GA, Welsh JP, Knight SM (2003) An economic analysis of the potential for precision farming in UK cereal production. *Biosystems Engineering* **84**, 533–545. doi: 10.1016/S1537-5110(02)00282-9

Goward SN, Markham B, Dye DG, Dulaney W, Yang J (1991) Normalised difference vegetation index measurements from the advanced very high resolution radiometer. *Remote Sensing of Environment* **35**, 257–277. doi: 10.1016/0034-4257(91)90017-Z

Hatfield JL (1983) Remote sensing estimators of potential and actual crop yield. *Remote Sensing of Environment* **13**, 301–311. doi: 10.1016/0034-4257(83)90032-9

Isbell RF (2002) 'The Australian soil classification.' (CSRIO Publishing: Collingwood, Vic.)

Jayroe CW, Baker WH, Greenwait AB (2005) Using multispectral aerial imagery to evaluate crop productivity. *Plant Management Network.* Available at: www.plant managementnetwork.org (accessed 27 July 2008).

Joyce EB, Rosengren NJ, Neilson JL, Rowan JN, Sargeant IJ, Martin JJ, Dahlhaus PG, Rowe RK, Rees DB, Robinson NJ, Imhof MP, MacEwan RJ (2007) A Geomorphology Framework for Victorian land and soil data. Department of Primary Industries, Victoria.

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. doi: 10.1080/01431160110107653

Lamb DW (2000) The use of qualitative airborne multispectral imaging for managing agricultural crops—a case study in south-eastern Australia. *Australian Journal of Experimental Agriculture* **40**, 725–738. doi: 10.1071/EA99086

Larscheid G, Blackmore BS (1996) Interaction between farm managers and information systems with respect to yield mapping. In 'Proceedings of 3rd International Conference on Precision Agriculture'. pp. 1153–1163. (ASA, CSSA, SSSA: Maison, WI)

MacDonald RB, Hall FG (1978) LACIE: an experiment in global crop forecasting. In 'Proceedings of Plenary Session, LACIE Symposium'. October 1978. pp. 17–48. (NASA Johnson Space Center: Houston, TX)

MacDonald RB, Hall FG (1980) Global crop forecasting. *Science* **208**, 670–679. doi: 10.1126/science.208.4445.670

Morgan CLS, Norman JM, Lowery B (2003) Estimating plant-available water across a field with inverse yield model. *Soil Science Society of America Journal* **67**, 620–629.

Nalepka RF, Colwell J, Rice DP (1976) Wheat productivity estimates using Landsat data. ERIM Rep. 114800–12-L. Environ. Research Institute, Michigan, Ann Arbor, MI.

Price P (2004) 'Spreading the PA message. Ground Cover, Issue 51.' (Grains Research and Development Corporation: Canberra, ACT)

Rab MA, Fisher PD, Armstrong RD, Abuzar M, Robinson NJ, Chandra S (2009) Advances in precision agriculture in south-eastern Australia. IV. Spatial variability in plant-available water capacity of soil and its relationship with yield in site-specific management zones. *Crop & Pasture Science* **60**, 885–900.

Raun WR, Johnson GV, Stone ML, Solie JB, Thomson WE, Lukins EV (1999) In-season prediction of yield potential in winter wheat. *Better Crops* **83**, 24–25.

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. doi: 10.1016/j.fcr.2006.12.013

Robinson NJ, Rampant PC, Callinan LAP, Rab MA, Fisher PD (2009) Advances in precision agriculture in south-eastern Australia. II. Spatio-temporal prediction of crop yield using terrain derivatives and proximally sensed data. *Crop & Pasture Science* **60**, 859–869.

Roy PS, Ravan SA (1996) Biomass estimation using satellite remote sensing data—an investigation on possible approaches for natural forest. *Journal of Biosciences* **21**, 535–561. doi: 10.1007/BF02703218

Rudorff BFT, Batista GT (1990) Spectral response of wheat and its relationship to agronomic variables in the tropical region. *Remote Sensing of Environment* **31**, 53–63. doi: 10.1016/0034-4257(90)90076-X

Scotford IM, Miller PCH (2005) Applications of spectral reflectance techniques in Northern European cereal production: a review. *Biosystems Engineering* **90**, 235–250.

Serrano L, Filella I, Penuelas J (2000) Remote sensing of biomass and yield of winter wheat under different nitrogen supplies. *Crop Science* **40**, 723–731.

Shanahan JF, Doerge TA, Johnson JJ, Vigil MF (2004) Feasibility of site-specific management of corn hybrids and plant densities in the great plains. *Precision Agriculture* **5**, 207–225. doi: 10.1023/B:PRAG.0000032762.72510.10

Shanahan JF, Schepersa JS, Francisa DD, Varvela GE, Wilhelma WW, Tringea JM, Schlemmera MR, Majorb DJ (2001) Use of remote-sensing imagery to estimate corn grain yield. *Agronomy Journal* **93**, 583–589.

Smith RCG, Adams J, Stephens DJ, Hick PT (1995) Forecasting wheat yield in a Mediterranean-type environment from the NOAA Satellite. *Australian Journal of Agricultural Research* **46**, 113–125. doi: 10.1071/AR9950113

Staggenborg SA, Taylor RK (2000) Predicting grain yield variability with infrared images. In 'Proceedings of the 5th Annual Precision Agricultural Meeting'. Bloomington, MN. (Eds PC Robert, RH Rust, WE Larson) (American Society of Agronomy: Madison, WI)

Weissteiner CJ, Kühbauch W (2005) Regional yield forecasts of malting barley (*Hordeum vulgare* L.) by NOAA-AVHRR remote sensing data and ancillary data. *Journal of Agronomy & Crop Science* **191**, 308–320. doi: 10.1111/j.1439-037X.2005.00154.x

Whelan BM, McBratney AB (2000) The 'null hypothesis' of precision agriculture management. *Precision Agriculture* **2**, 265–279. doi: 10.1023/A:1011838806489

Wilkerson J, Moody H (2004) Variable rate plant population and nitrogen. In 'Southern Plant Nutrition Management Conference Proceedings'. Ardmore, OK.

Wright DL Jr, Ramsey RD, Baker DJ, Rasmussen VP Jr (2003) A comparison of two geospatial technologies in non-uniform wheat fields: yield monitors and remote sensing. In 'ASPRS Annual Conference'. Anchorage, Alaska. Available at: www.gis.usu.edu/ArcWebpage/inside_table/2003Presentations/0225.pdf

Zhang MM, O'Neill M, Hendley D, Drost D, Ustin SL (1998) Corn and soybean yield indicators using remotely sensed vegetation index. In 'Proceedings of the 4th International Conference on Precision Agriculture. 2'. (Eds PC Robert, RH Rust, WE Larson) pp. 1475–1481. (ASA, CSSA, SSSA: Madison, WI) Available at: http://agis.ucdavis.edu/research/PF98_MN3.pdf

Manuscript received 9 October 2008, accepted 29 July 2009