

## Inverse meta-modelling to estimate soil available water capacity at high spatial resolution across a farm

M. J. Florin · A. B. McBratney · B. M. Whelan · B. Minasny

Published online: 7 July 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** Geo-referenced information on crop production that is both spatially- and temporally-dense would be useful for management in precision agriculture (PA). Crop yield monitors provide spatially but not temporally dense information. Crop growth simulation modelling can provide temporal density, but traditionally fail on the spatial issue. The research described was motivated by the challenge of satisfying both the spatial and temporal data needs of PA. The methods presented depart from current crop modelling within PA by introducing meta-modelling in combination with inverse modelling to estimate site-specific soil properties. The soil properties are used to predict spatially- and temporally-dense crop yields. An inverse meta-model was derived from the agricultural production simulator (APSIM) using neural networks to estimate soil available water capacity (AWC) from available yield data. Maps of AWC with a resolution of 10 m were produced across a dryland grain farm in Australia. For certain years and fields, the estimates were useful for yield prediction with APSIM and multiple regression, whereas for others the results were disappointing. The estimates contain ‘implicit information’ about climate interactions with soil, crop and landscape that needs to be identified. Improvement of the meta-model with more AWC scenarios, more years of yield data, inclusion of additional variables and accounting for uncertainty are discussed. We concluded that it is worthwhile to pursue this approach as an efficient way of extracting soil physical information that exists within crop yield maps to create spatially- and temporally-dense datasets.

**Keywords** Inverse modelling · Soil available water capacity · Meta-modelling · Precision agriculture · Crop growth simulation modelling

---

M. J. Florin · A. B. McBratney · B. M. Whelan · B. Minasny  
Australian Centre for Precision Agriculture, University of Sydney, Lvl. 2 John Woolley Bldg. A20,  
Manning Rd., University of Sydney, NSW 2006, Australia  
e-mail: Alex.McBratney@sydney.edu.au

M. J. Florin (✉)  
Plant Production Systems, Wageningen University, Droevendaalsesteeg, 1, 6708 PB Wageningen,  
The Netherlands  
e-mail: madeleine.florin@wur.nl

## Introduction

One aspect of precision agriculture (PA) is site-specific crop management (SSCM). To match inputs better with soil and crop requirements as they vary in space and time, SSCM involves management of spatial units that are smaller than fields (often using potential management classes). Spatially-dense, geo-referenced information is an integral component of SSCM. Whelan and McBratney (2003) consider crop yield, topography and apparent soil electrical conductivity to be basic information layers for determining potential management classes in Australia.

Analysis of crop yield in any one place across several years illustrates that temporal variation driven by climate is as important as spatial variation on the potential to undertake SSCM (McBratney and Whelan 1999). This suggests that yield maps from many years would provide important information for PA, but a library of these for a field accumulates slowly. Crop growth simulation models can provide a wealth of temporal data, but are traditionally point-based, and model complexity and data requirements limit the number of feasible simulation points across an area of interest. This produces a mismatch between the spatial resolution of outputs from crop growth simulation models and the spatial resolution required by PA. The challenge is, therefore, to develop methods that simultaneously satisfy spatial and temporal data requirements with crop growth simulation models.

Attempts to address this challenge within the PA literature consist of running crop growth simulation models at individual points on a grid spanning the study area (Link et al. 2006; Braga and Jones 2004; Timlin et al. 2001). These studies acknowledge the intensity of data required for input to their models and consequently take advantage of readily available high-resolution yield information. Inverse use of crop growth simulation models (inverse modelling) has proved to be a useful approach that capitalizes on this yield information to estimate soil properties site-specifically. This type of work is akin to model parameterization or model calibration where a range of estimation procedures can be used to search multivariate space resulting in optimal choices of model parameters.

The literature documents a variety of optimization approaches. Soil water properties are included in all of the examples. Link et al. (2006) calibrated the CERES-Maize crop growth simulation model by varying potential rooting depth and available soil water using simulated annealing. Yield data from 1998 to 2002 were used for calibration across a single field divided into 30 grid cells. It was reported that for this field the calibrated model explained 60% of yield variability in each grid cell over the same 5 years. Braga and Jones (2004) also used simulated annealing to optimize several soil water properties and a root growth factor. Timlin et al. (2001) used a genetic algorithm (an algorithm that attempts to emulate evolution) for the optimization of water holding capacity (WHC) across a single field for 2 separate years.

Irmak et al. (2001) described an innovative and exhaustive search technique to estimate five spatially variable soil properties that involved creating a database with different combinations of these five properties (74 536 model runs). Four different searching scenarios were used to find optimum estimates for each grid. This approach was reported to be computationally efficient, and the resulting yield predictions explained as much as 90% of the variation in soybean yield. Similarly, Morgan et al. (2003) created 'look-up' tables containing yield values as a function of plant available water for several years.

These studies demonstrate the use of high-resolution yield data in conjunction with crop growth simulation models for estimating high-resolution soil data and subsequently for predicting spatial and temporal variation in yield. Nevertheless, each of these studies focus on single fields and the next step for PA is to extend the spatial cover to an entire farm.

This dramatically increases the number of simulation points to thousands, and at the same time raises an additional methodological question regarding the appropriate level of model complexity.

Meta-modelling is a statistical technique that approximates complex simulation models by transforming model inputs into model outputs (Kleijnen and Sargent 2000). Meta-models are used for a variety of purposes such as gaining insight into critical relationships within simulation models, validating simulation models and reducing their size (Ruben and van Ruiven 2001). In the context of this work, a clear advantage of meta-modelling is that complex models can be simplified into their most important relationships thereby reducing input parameter and computer processing requirements, and increasing their potential for use at a higher spatial resolution.

The methods presented in this paper depart from the current state of crop modelling within the PA literature by introducing meta-modelling in combination with inverse modelling to estimate site-specific soil properties that can then be used in crop growth simulation models. In combination, inverse- and meta-modelling should provide a novel method towards better matching of potential outputs from crop growth simulation modelling with the requirements for PA across whole farms. To achieve this goal this research targets three distinct aims: (1) to create an inverse meta-model from the agricultural production simulation model (APSIM) that can estimate soil available water capacity (AWC) from crop yield; (2) to generate AWC maps across a farm and (3) to validate the inversely modelled AWC maps for high-resolution yield prediction. We demonstrate this approach for a 3000 ha dryland grain farm in South Australia.

## Materials and methods

### Study site

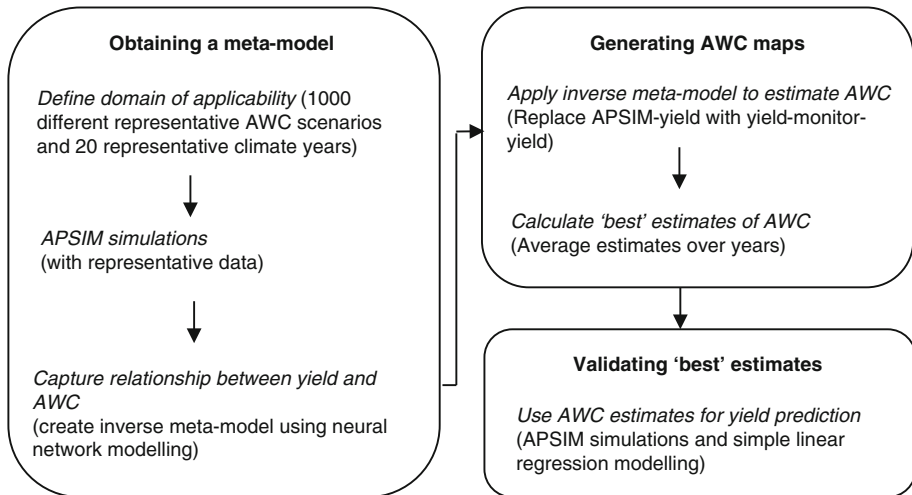
The farm under study, 'BrookPark', is 200 km north of Adelaide near the town of Crystal Brook in South Australia. The climate is characterized by a mean daily maximum temperature of 24.4°C, a mean daily minimum temperature of 12.6°C and a mean annual rainfall of 345 mm that falls during the winter. Dominant soil types are Planosols, Chernozems and Durosols (FAO 1998). The most important crops grown are wheat and barley.

Figure 1 outlines the sequential steps undertaken during this research. An inverse meta-model of APSIM was created to enable AWC to be estimated based on wheat yields. The meta-model was used to estimate AWC at a high spatial resolution across the farm. These estimates were validated on the basis of their representation of AWC derived from soil samples and their value for high-resolution yield prediction.

### APSIM

APSIM is a point-based crop growth simulation model that runs on a daily time-step. It has been used widely, particularly in Australia, and it features in a range of publications (e.g. Asseng et al. 1998; de Voil et al. 2006; Wong and Asseng 2006).

The model incorporates a number of modules that can be added and removed with relative ease. For this study three modules that describe crop growth, soil hydraulic behaviour and management decisions will be outlined further. Numerous other modules are available and explanations can be found on the APSIM website (<http://www.apsim.info/wiki/>).



**Fig. 1** Flowchart depicting the sequential steps of the research methods

The module describing soil hydraulic behaviour in APSIM is SOILWAT. This module describes runoff, evaporation, saturated flow, unsaturated flow and solute movement in one dimension. SOILWAT is a successor of CERES and PERFECT (Godwin and Jones 1991; Littleboy et al. 1999 respectively). Infiltration and runoff are partitioned using the curve number method (Mishra and Singh 2003). Water movement within the profile is described using a cascading water balance model. Saturated and unsaturated water flows are described with separate algorithms. For saturated flow the parameter SWCon is used, which defines the fraction of water above the drained upper limit (DUL) (water content at  $-0.1$  bar) that will flow to the soil layer below. For unsaturated flow, two parameters describing diffusivity are used. Evaporation is treated in two stages. The first depends on atmospheric demand (assuming that the soil profile will meet demand). The second stage is limited by soil moisture and is a function of time from the end of first stage evaporation. The parameter U defines the amount of cumulative evaporation to occur until atmospheric demand exceeds soil moisture supply. The parameter Cona describes the rate of second stage evaporation as a function of the square root of time. It is also necessary to indicate initial moisture conditions. This module interacts with climate, crop growth and management modules.

APSIM contains a generic crop module (GCROP) that describes biological processes common to many crop species (Wang et al. 2002). GCROP simulates seven processes that occur in a daily loop; these are transpiration, phenology, biomass accumulation, leaf area development, senescence, crop N and plant death. These processes interact with climate, management and soil properties. APSIM approaches root growth and water uptake as an extraction front depth meaning the depth of water depletion rather than actual rooting depth is modelled (Wang and Smith 2004). As a result the root water uptake factor (kl) includes root length density (l) and a diffusion constant (k). This factor and a root advancement factor (xf) are defined for each soil layer.

The manager module allows the user to make agronomic decisions relevant to growing a crop. Inputs include cultivar variety, sowing date, sowing depth, row spacing, plant density, fertilizer and irrigation amounts. These management decisions have an effect on crop development to differing degrees. The model is particularly sensitive to cultivar choice.

Cultivars are distinguished in terms of photoperiod sensitivity and vernalization sensitivity, both of which affect the rate at which a crop matures.

### Key assumption

It is assumed that water is the most important limiting factor in the dryland grain farming system examined in this study. This is supported to a substantial degree in the literature; for example, Irmak et al. (2002) showed correlations between root-zone plant available water and soybean yield across a field in Iowa, USA. A practicality of this assumption is that hydraulic properties were the only model inputs considered in detail. Consequently, other parameters that potentially affect crop yield were held constant both spatially and temporally. The hydraulic property considered was AWC, which is defined as the difference between DUL (water content at  $-0.1$  bar) and lower limit (LL) (water content at  $-15$  bar). Previous studies have derived linear relationships between wheat yields simulated by APSIM and AWC of the topsoil (Wong and Asseng 2006). These authors demonstrated the varying importance of AWC depending on interactions with rainfall, initial moisture content and nitrogen application. The AWC was considered to be the same as plant available water capacity (PAWC) (or extractable water). This was justified by the fact that wheat is the only crop considered in this study.

### Soil, crop and landscape data

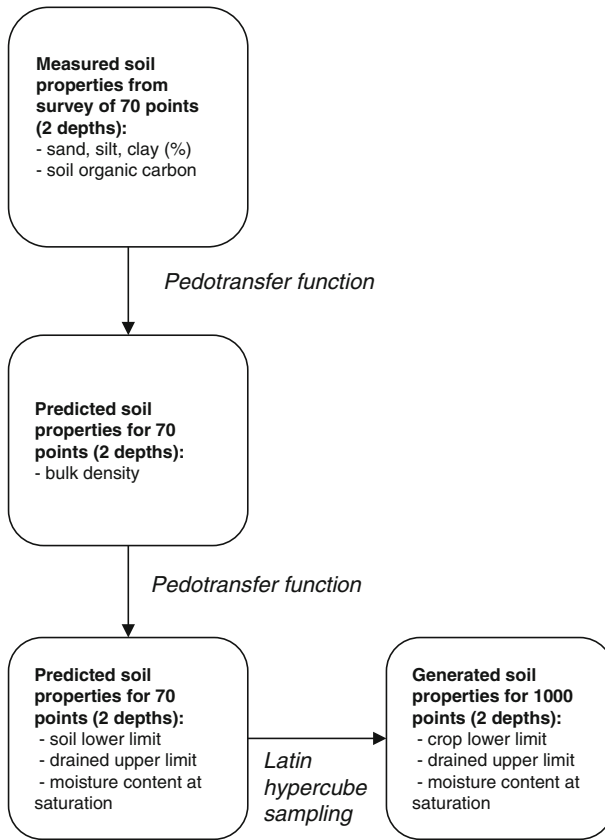
Soil and climate information were the two main data types of data required to create an inverse meta-model. The envisaged domain of applicability for the meta-model is all the possible climate and AWC scenarios that might occur across the study site. This gives meaning to the requirement for a 'representative' collection of AWC and climate data for the derivation of a meta-model.

Figure 2 illustrates the three-step process to generate data on AWC and LL at two depths for 1000 hypothetical soil profiles. These steps involved soil sampling, and the application of pedotransfer functions (PTFs) and of latin hypercube sampling (LHS).

A soil survey across the farm undertaken prior to this research resulted in 140 bulked soil samples (0–30 and 60–90 cm from 70 locations across the farm). A random stratified soil sampling scheme was designed to ensure that the main soil types were sampled. Seven strata were delineated using a combination of elevation, apparent electrical conductivity and gamma radiometrics data (Florin et al. 2005). The particle size analysis (hydrometer method) and organic carbon (Walkley Black) measurements from these samples were used to predict AWC using PTFs.

Bulk density (BD) was first predicted with a PTF that includes total sand%, organic carbon and depth (Tranter et al. 2007). Second, moisture content at saturation (SAT), DUL, LL and air dry moisture content were predicted from the BD, total sand%, silt%, clay% and total carbon content using a neural-network PTF trained with Australian data (Minasny and McBratney 2002). The AWC was calculated by subtracting the LL from the DUL.

Next, LHS (McKay et al. 1979) was used to generate AWC and LL for 1000 hypothetical soil profiles. Latin hypercube sampling is an efficient method for sampling a multi-dimensional statistical space and is often used for uncertainty analysis (e.g. Post et al. 2008). Again each profile is characterized at two depths, 0–30 and 60–90 cm. The minimum and maximum predicted  $AWC_{0-30\text{ cm}}$ ,  $AWC_{60-90\text{ cm}}$ ,  $LL_{0-30\text{ cm}}$  and  $LL_{60-90\text{ cm}}$  for the 70 sampled points and correlations between these properties were used to guide this



**Fig. 2** Flowchart depicting the sequential steps used to estimate and generate the soil data for this study

generation of data. The outcome was 1000 equally probable combinations of these four soil properties.

Daily rainfall and radiation data for the study site were obtained from the SILO Data Drill (Jeffrey et al. 2001). This is a database of continuous daily climate information constructed from the Australian Bureau of Meteorology ground-based observations. Twenty years of weather data were considered adequate to include a number of dry, wet, hot, cold and average years. Twenty consecutive years including 1980 and 1999 were selected.

Spatially referenced yield information was required to apply the inverse meta-model. Eight years of data from wheat yield monitoring with a positional accuracy of 0.1–0.2 m across ‘BrookPark’ between 1999 and 2006 were available. As a result of crop rotation, between 2 and 4 years of wheat data were available for individual fields. Four years of yield data were considered a minimum requirement for this work. A 10-m grid was generated for the farm that excluded the farmhouse, water bodies and natural vegetation. Raw yield data were inspected for distributional outliers within individual fields and erroneous values were removed. Detailed description of this ‘cleaning’ process is in Florin (2008). Next, the wheat yield data were predicted onto the grid using local variograms and block kriging (Walter et al. 2001) with the software Vesper (Minasny et al. 2002).

## Creating the inverse meta-model

APSIM was run for 20 years (1980–1999) of continuous wheat cropping for each of the 1000 soil profiles characterized with AWC. Prior to this simulation, APSIM was calibrated to ensure that it was possible to simulate an adequate range of yield values, i.e. values that included the minimum and maximum recorded yield values for the farm. This calibration involved varying a number of model parameters and observing simulation results for a random subset (100 points) of the 1000 AWCs. The parameters that were varied and the values defined are given in Tables 1 and 2. Some guidance regarding the values assigned to these parameters was obtained from previous studies, from within the APSIM soil database (Dalglish et al. 2006) and from expert opinion. This process refers back to an initial simplifying assumption that these parameters are not spatially or temporally variable across the farm.

The APSIM-predicted yield for the 20 years was plotted against total AWC (summed over the rooting depth of the profile in mm) and the relationships were observed. Linear stepwise regression was used to determine which climate variables (monthly rainfall and radiation, including both pre-season and in-season months) together with AWC best predicted the APSIM yield output.

Results from the stepwise linear regression determined which climate variables were entered into the meta-modelling process, together with APSIM-yield, to predict AWC and LL (at two depths). A neural network model was the proposed meta-model because neural networks can identify non-linear relationships (Tamari et al. 1996). Neural networks with two, three and four nodes were investigated. Two thirds of the data were randomly selected as a subset for data training and the remaining third was used to validate the neural network's predictive ability. The best fitting models were chosen with reference to a cross-validated (CV)  $R^2$ .

## Applying and validating the inverse meta-model

The neural network model was applied across the farm by replacing APSIM-yield with yield-monitor yield. As a result, estimates of AWC and LL for two depths were made at every point across the farm where yield monitor data existed. 'Best' estimates of AWC were calculated by averaging estimates from as many years as there were yield data available for each field.

A three-step validation process was used. First, the 'best' estimates were compared to the hydraulic properties estimated by PTFs at the sampling points that coincided with the locations that were inversely modelled. The 'best' estimate of AWC was plotted against PTF estimated AWC.

**Table 1** APSIM parameters and variables that change with depth and describe water movement, plant root function and soil chemical properties

Soil layer	Depth (cm)	SWCon	Kl	Xf	OC <sup>a</sup>	pH
1	0–30	0.5	0.04	0.5	0.67	7.0
2	30–60	0.5	0.02	0.2	0.67	8.0
3	60–75	0.5	0.02	0.2	0.50	8.5
4	75–90	0.5	0.02	0.2	0.30	9.5

<sup>a</sup> Organic carbon

**Table 2** APSIM parameters and variables that describe soil evaporation, runoff, unsaturated flow and management decisions

APSIM parameter or variable	Value
U (to describe first stage evaporation dependent on atmospheric demand)	2
Cona (to describe second stage evaporation limited by soil moisture)	2
Diffusivity constant and slope (to describe unsaturated water flow as a function of the average water content between two layers)	88 and 35
Curve number 2 (to specify the runoff response curve for average antecedent rainfall conditions)	73
Curve number reduction (to define the reduction in runoff due to crop and residue cover)	20
Sowing window start date	15-May
Sowing window end date	20-May
Sowing density (plants m <sup>-2</sup> )	150
Sowing depth (mm)	30
Cultivar	Kirchauff
Row spacing (mm)	300

Secondly, a year of yield data was excluded from the ‘best’ estimate of AWC. This year was used to validate the inversely modelled hydraulic properties from an APSIM yield prediction potential. One hundred points were chosen randomly from the sample fields and estimates of LL, DUL and SAT were used as input into APSIM to predict yield for the excluded year. Modelling scenarios using the correct sowing date and cultivars were input into APSIM. To obtain an initial moisture content for the simulations, APSIM was run continuously for 3 years prior to the validation year. This validation process was undertaken for 2 years across three example fields on the farm. Plots of yield monitor yield and the APSIM predictions were compared by fitting a linear model to these data.

Third, estimates of LL, DUL and SAT (those used for validation of APSIM) were used to create simple linear yield prediction models across the same fields. Stepwise linear regression was used to predict yield using the hydraulic properties and monthly rainfall data. The  $R^2$  values for these models were recorded.

## Results

### Hydraulic properties generated using LHS

Table 3 gives data on the hydraulic properties that were generated by LHS and data from the PTF predictions. Hydraulic properties measured for similar soil types taken from the APSIM soil database are included in this table for verification of the PTF-estimates. The mean topsoil PTF-estimated AWC is less than the mean AWC measurements in the database (0.10 vs. 0.16). For the subsoil, the mean PTF-estimated AWC is only slightly smaller than the APSIM database measurements (0.10 vs. 0.11).

### The inverse APSIM meta-model

Figure 3a illustrates the relationship between APSIM simulated wheat yield and AWC generated using LHS. The different strands that are visible represent individual years. This



**Table 3** Range and mean values for the hydraulic properties generated for ‘BrookPark’ using LHS and PTFs; also a comparison with a similar soil characterized within the APSIM soil database

	Lower limit (0–30 cm)	Lower limit (60–90 cm)	Drained upper limit (0–30 cm)	Drained upper limit (60–90 cm)	Available water capacity (0–30 cm)	Available water capacity (60–90 cm)
‘BrookPark’ (PTF)	0.03–0.26 (0.13)	0.05–0.30 (0.17)	0.08–0.39 (0.23)	0.11–0.40 (0.26)	0.04–0.17 (0.10)	0.04–0.14 (0.10)
‘BrookPark’ (LHS)	0.07–0.13 (0.10)	0.11–0.15 (0.13)	0.11–0.30 (0.21)	0.15–0.29 (0.22)	0.04–0.17 (0.10)	0.04–0.14 (0.09)
Similar soil from APSIM soil database					0.16	0.11

illustrates that the relationships are year-specific and suggests that climate variables are important inputs for the proposed meta-model.

The climate variables that were identified as important predictors of APSIM-yield using stepwise linear regression were rainfall observations for June, July, August and November. Consequently, rainfall for these months and APSIM-yield formed the predictor variables in the neural network model. The model with three nodes proved slightly better at approximating APSIM than that with two nodes ( $CV R^2$  values were 0.76 and 0.74, respectively). A visual comparison of Fig. 3a and b shows how well the inverse meta-model approximates the relationship between crop yield and profile AWC that is contained within APSIM.

#### Applying the inverse meta-model across a farm

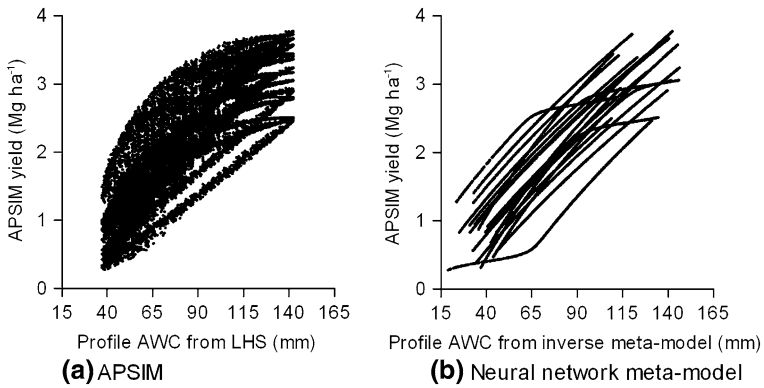
Figure 4 illustrates inversely modelled profile AWC derived from individual years of yield data across the farm. Four of the 8 years are provided as an example. Based on visual comparison, these maps demonstrate the temporal specificity of the inverse modelling method. For some fields the spatial patterns between years are consistent, but the estimated AWC values are different. However, for some fields it is difficult to recognise consistent spatial patterns between the estimates for different years.

The location and names of the fields where estimates were possible for a minimum of 4 years are shown in Fig. 5. In this figure the ‘best’ estimates of profile AWC for these fields are mapped.

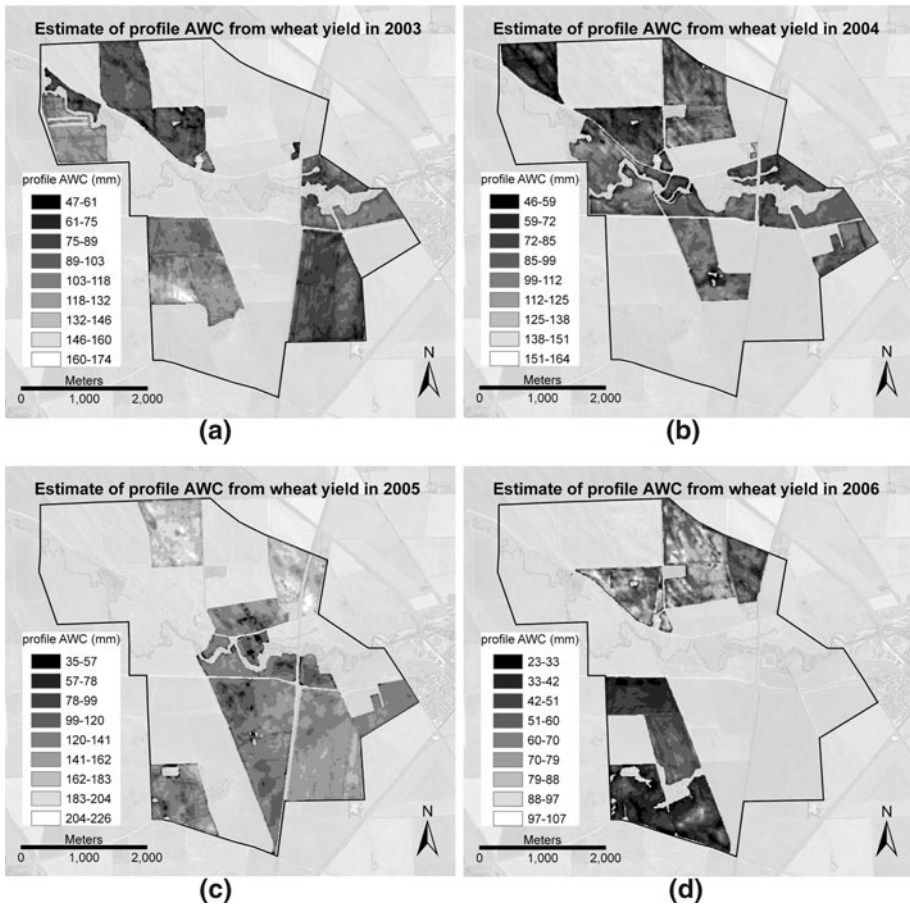
#### Validating the ‘best’ estimates of AWC

##### *Comparing ‘best’ estimates of AWC with PTF-derived AWC*

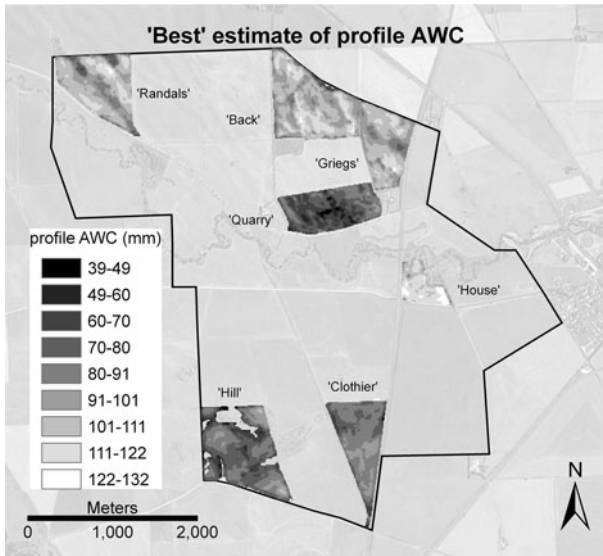
Figure 6 illustrates the relationship between ‘best’ estimates of profile AWC and the profile AWC estimated from a PTF for corresponding points across ‘BrookPark’. This figure shows that there is no correlation between PTF-estimated and inversely-estimated hydraulic properties. Reasons for this disappointing result are discussed in the following sections of this paper.



**Fig. 3** a Simulated yield versus profile AWC for APSIM simulated yield and b inversely meta-modelled AWC

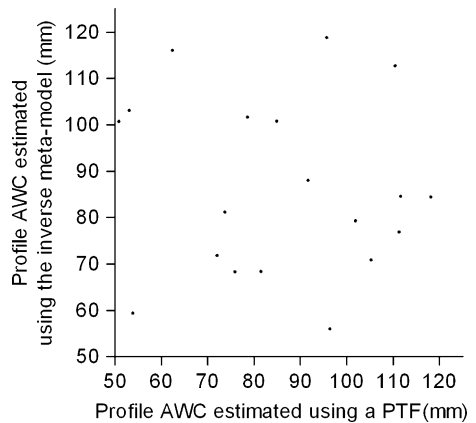


**Fig. 4** Profile AWC estimates using wheat yield from: a 2003, b 2004, c 2005 and d 2006



**Fig. 5** ‘Best’ estimates of profile AWC calculated by averaging all available data

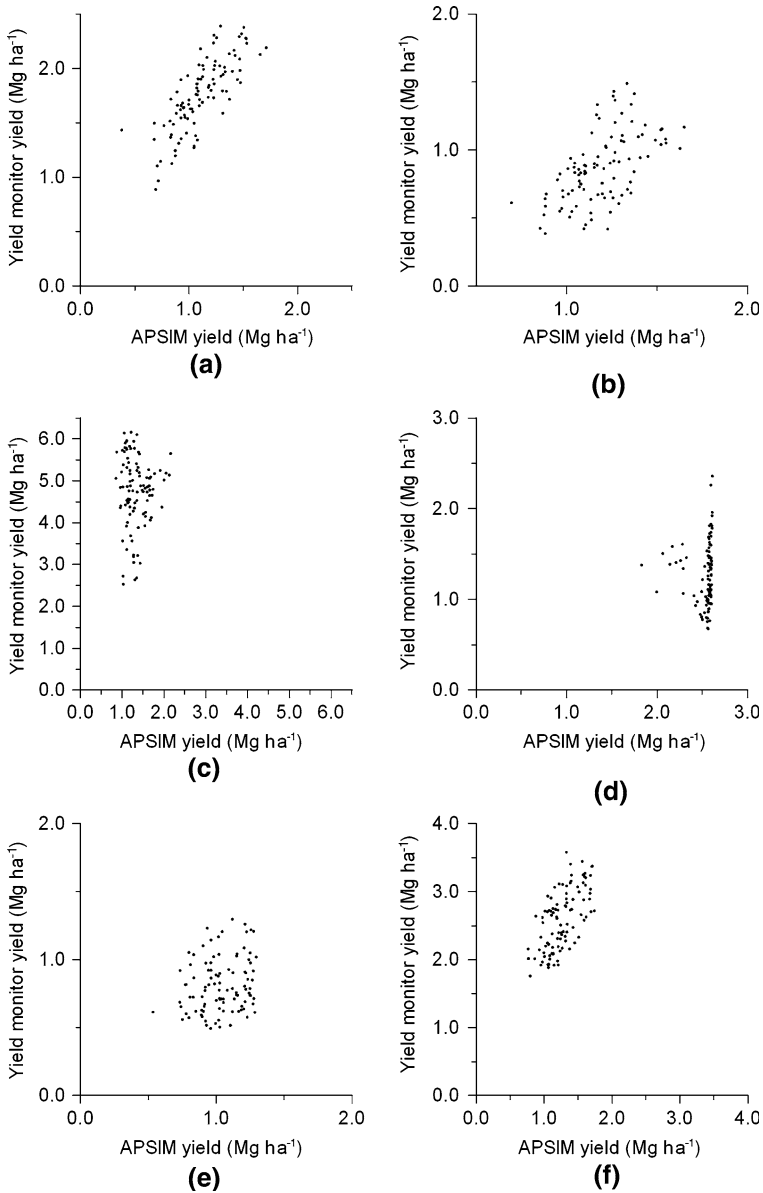
**Fig. 6** Comparison of inversely modelled profile AWC with AWC estimated using pedotransfer functions



*Estimated AWC for yield prediction with APSIM and with linear regression models*

Figure 7 displays plots of yield-monitor yield versus APSIM-predicted yield for the fields ‘Hill’, ‘Quarry’ and ‘Randals’. Table 4 gives the  $R^2$  values for linear models fitted to all years available for these fields. The result is strongly field- and year-specific. For ‘Hill’ (in 1999 and 2006) and for ‘Quarry’ (in 2005), the results are reasonable. The  $R^2$  values suggest that the AWC estimates within APSIM can explain one- to two-thirds of the variation in crop yield. However, the remainder of the results are disappointing and suggest that the estimated AWCs are not useful as input to APSIM for yield prediction.

Table 4 also gives the linear model parameters that were used to predict wheat yield across the three individual fields. The differences in performance between years and fields



**Fig. 7** Plots of APSIM simulated yield versus yield-monitor yield where the APSIM simulations are used for validation of the AWC estimates. **a** 'Hill' 1999, **b** 'Hill' 2006, **c** 'Randals' 2001, **d** 'Randals' 2004, **e** 'Quarry' 2002, **f** 'Quarry' 2005

are consistent with the APSIM predictions. Again, predictions for 'Hill' and 'Quarry' show greater promise than for 'Randals'. On the whole, the linear regression models based on the estimated AWC values appear to out-perform APSIM. Reasons for these results are discussed in the following sections of this paper.

**Table 4** APSIM and linear yield predictions: the predictors (for the linear models) and the  $R^2$  values associated with the linear model

Field	APSIM prediction year— $R^2$	Yield data used for hydraulic property estimation <sup>a</sup>	Predictor variables	Linear prediction year— $R^2$
'Hill'	1999–0.61	Wheat 1999	LL1	1999–0.59
	2006–0.33	Wheat 2002	AWC2	2002–0.62
		Wheat 2005	Rain:	2005–0.45
		Wheat 2006	September	2006–0.59
		Radiation:		
		August		
		September		
'Quarry'	2002–0.04	Wheat 1999	AWC2	1999–0.61
	2005–0.42	Wheat 2000	Rain:	2000–0.59
		Wheat 2002	September	2002–0.06
		Wheat 2005	April–November	2005–0.40
		Radiation:		
		May		
'Randals'	2001–0.00	Wheat 1999	LL1	1999–0.22
	2004–0.00	Wheat 2001	Rain:	2001–0.00
		Wheat 2002	September	2002–0.05
		Wheat 2004	Radiation:	2004–0.14
		August		
		September		

<sup>a</sup> Year of yield data excluded for AWC estimate for that particular years yield prediction

## Discussion

### The meta-model for approximating APSIM

The CV  $R^2$  value obtained from the neural network analysis suggests that the meta-model identifies about 70% of the variation in yield predictions that APSIM produces. There are several possible reasons why 30% of the variation was missed by the meta-model. One possibility is that a sample of 1000 different soil profiles does not describe the whole population. Perhaps, more combinations of hydraulic properties could be simulated with APSIM and be used to derive the meta-model. A second consideration is that the temporal resolution of the climate data was decreased from daily to monthly. This may have reduced the effectiveness of reproducing the APSIM model with a neural network.

One benefit from this type of meta-modelling is that information about important variables within the model is gained. For example, the rainfall variables selected for the meta-model confirm the importance of in-season rainfall for winter cropping across 'BrookPark'. Furthermore, in terms of reducing computer processing requirements, the meta-model shows promise. In this respect, the neural network model is dramatically more efficient than APSIM.

### AWC estimates

The realizations of different AWC estimates for different years (Fig. 4) are consistent with some previous reports in the literature, e.g. Timlin et al. (2001). This explains why 'best' estimates of AWC were obtained by averaging across as many years as were available.

However, this approach requires that temporal variability is considered. Some previous spatio-temporal yield analysis (Florin et al. 2009) suggested that certain fields appear temporally stable, in which case this averaging across years to estimate AWC would be useful. However, it might not be useful in the case of a temporally unstable field where no correlation in yield patterns between years can be detected. Information about different degrees of temporal stability between fields and years might also be gained from this analysis. From a PA management perspective, information about temporally stable yield-AWC relationships might be more valuable than yield alone.

The lack of a significant relationship between inversely estimated AWC and PTF-estimated AWC (Fig. 6) is of concern. This result might indicate that the inverse meta-model has reproduced yield values at the expense of valid AWC values. This phenomenon of over-fitting was discussed by Braga and Jones (2004). These authors demonstrated that prediction of soil water content using yield as the objective function variable leads to unreliable soil water predictions compared to using soil water content as the objective function variable (i.e. a variable that is closely related to the properties being estimated).

The differences in this relationship between the years also suggest that the estimated AWCs contain some information about climatic interactions with the crop, soil and landscape. Given that the only variables considered in the meta-model were AWC and rainfall, it is likely that other crop, soil and landscape variables are affecting the estimated AWC values. Some possible spatial, temporal and spatio-temporal soil, crop or landscape properties that may be interacting with climate variables are: root growth, soil depth, soil structure, soil chemical properties, pests, diseases and topography. Further work is required to understand which 'implicit information' has been incorporated into the inversely estimated AWCs. For example, information on the spatial variation in soil depth and plant rooting depth would be useful. The assumption that root extraction front depths (determined by the 'kl factor') do not vary spatially is a possible source of 'implicit information' in the AWC estimates. The relationship between AWC and yield will change if the extraction front depth of crop roots varies because this soil and crop property affects the rate at which crop roots can take up water. In addition, crop roots might not always reach the maximum rooting depth every year, thereby introducing an interaction with climate.

Further, APSIM yield predictions are sensitive to management variables such as sowing dates, crop varieties and crop rotations. In reality crop varieties and sowing dates are different between fields and between years. The meta-models were derived from APSIM simulations where the crop (type and variety) and sowing date were held constant across all points and years. These assumptions provide further entry points for the inclusion of 'implicit information' in the inversely estimated AWCs.

Finally, this method does not account explicitly for the uncertainty propagated through the process. Sources of uncertainty within the data arise from the PTFs, interpolated crop yields and climate data. Inclusion of an uncertainty factor for the AWC values would be a useful addition to this method.

#### AWC estimates for predicting yield

Yield prediction using APSIM and by linear regression lead to contrasting results between fields and between years (Fig. 7; Table 4). On the whole the outcome is disappointing; however, a few exceptions suggest some promise for this methodology. Therefore, it is worthwhile discussing differences between fields, years and the results generally.

Results were most promising for 'Hill' compared with the other fields. One explanation for this may be that the most dramatic changes in elevation occur within (aptly named)

‘Hill’ and as a result water plays the most important role in the determination of yield. ‘Hill’ includes a relatively flat hilltop in the west surrounded by steeper slopes to the north, east and south which again give way to relatively level sections. The estimated AWC values for ‘Hill’ follow the slope pattern, with larger values on the flatter sections of the field. The result suggests that this method of AWC estimation and yield prediction may be most suitable for particular combinations of terrain and soil types.

It is clear that the quality of the results is specific to the year. For example, results for ‘Quarry’ (predicting yield using both APSIM and linear regression) appear reasonable for 1999 yet poor for 2002 (Table 4), which might be because 2002 was a particularly dry year. As a result, water was an equally limiting factor across the whole field whereas in 1999 the variation in AWC results in spatially variable yield. These differences between years indicates that 3 years of crop yield data to estimate AWC and 1-year’s data for validation are not adequate. Temporally unstable yield patterns and processes certainly demand more data through time.

It should also be considered whether APSIM is too complex a model and if AWC alone is adequate for robust simulation of spatial and temporal variation in yield. A logical point for improvement would be to parameterize APSIM more thoroughly. This raises questions surrounding model complexity versus model simplicity. One could ask: *What degree of model complexity is optimum?* Reynolds and Acock (1985) attempted to answer to this question by separating the total error of model predictions into error due to simplification of the system and error due to uncertainty in parameter estimation. This partition of total error enables an optimal level of model complexity to be identified. Within the literature there are several examples where simple models have proved to be more reliable than more complex models. For example, Bell and Fischer (1994) found that a regression model was superior to CERES for potential wheat yield prediction in Mexico. Some further examples are outlined by Sinclair and Seligman (1996).

Potential errors related to assumptions about interactions between soil, water and crop that were raised in relation to estimates of AWC also have some relevance to the issue of model complexity. This methodology has focused explicitly on AWC with the assumption that this hydraulic property controls soil water supply to the crop, which in turn determines yield response. There is scope to challenge this in terms of other sources or limitations to water supply and other important factors limiting yield. An intuitive source or limitation of water might be lateral surface and subsurface flow. Previous studies have mentioned the inability of point-based crop growth simulation models to account for lateral movement of water (e.g. Ferreyra et al. 2006; Timlin et al. 2001; Fraisse et al. 2001).

Finally, in the discussion of model complexity, it is useful to compare the linear yield predictions with those from APSIM. The differences in performance between fields and years are consistent between approaches. With respect to predictive capabilities, the linear models were either similar to or better than the APSIM predictions. It is apparent that the value of simple linear models, as compared with APSIM, is the greater ease of prediction at a high resolution across fields and ultimately across farms.

## Conclusions

This methodological contribution to the challenge of using crop growth simulation models to identify variation in spatial and temporal crop yield adequately shows promise and raises many avenues for further work.

Inverse modelling is an efficient way of extracting site-specific soil properties from readily available crop yield information. Furthermore, meta-modelling is useful in terms of simplifying a complex model and addressing the challenge to computer power that simulation modelling brings to PA. The combination of inverse- and meta-modelling is a useful concept to be pursued.

However, future research to address the ability of the meta-model to approximate APSIM closely is imperative. The conceptual model underpinning this research requires modification. The AWC and climate variables alone do not adequately explain all of the spatio-temporal variation in crop yield across a farm. Incorporation of additional variables into the meta-model is necessary to provide a better understanding of how climate interacts with the soil, crop and landscape. More years of yield data are required to improve model building in the light of spatio-temporal variation. Finally, inclusion of uncertainty estimates within this approach are necessary to create confidence in the validity of the model and would be useful to explore the trade-off between model complexity and simplicity.

**Acknowledgements** The authors are grateful to the Grains Research and Development Corporation, Australia for funding this research and to Malcom Sargent for providing access to the study site and the necessary data. Thank you to the reviewers and editor for their careful and useful comments.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Asseng, S., Fillery, I. R. P., Anderson, G. C., Dolling, P. J., Dunin, F. X., & Keating, B. A. (1998). Use of APSIM wheat model to predict yield, drainage and NO<sub>3</sub> leaching for a deep sand. *Australian Journal of Agricultural Research*, *49*, 363–377.
- Bell, M. A., & Fischer, R. A. (1994). Using yield prediction models to assess yield gains: A case study for wheat. *Field Crops Research*, *36*, 161–166.
- Braga, R. P., & Jones, J. W. (2004). Using optimisation to estimate soil inputs of crop models for use in site-specific management. *Transactions of the ASAE*, *47*, 1821–1831.
- Dalgliesh, N., Wockner, G., & Peake, A. (2006). Delivering soil and water information to growers and consultants. In N. C. Turner & T. Acuna (Eds.), *Proceedings of the 13th Australian society of agronomy conference*, Perth, Australia. [http://www.regional.org.au/au/asa/2006/poster/water/4552\\_dalgliesh.htm](http://www.regional.org.au/au/asa/2006/poster/water/4552_dalgliesh.htm).
- de Voil, P., Rossing, W. A. H., & Hammer, G. L. (2006). Exploring profit-sustainability trade offs in cropping systems using evolutionary algorithms. *Environmental Modeling and Software*, *21*, 1368–1374.
- FAO. (1998). *World reference base for soil resources*. Rome: Food and Agriculture Organization of the United Nations.
- Ferreira, R. A., Jones, J. W., & Graham, W. D. (2006). Parameterizing spatial crop models with inverse modelling: Sources of error and unexpected results. *Transactions of the ASABE*, *49*, 1547–1561.
- Florin, M. J. (2008). *Towards precision agriculture for whole farms using a combination of simulation modelling and spatially dense soil and crop information*. Ph.D. thesis. <http://hdl.handle.net/2123/3169>.
- Florin, M. J., McBratney, A. B., & Whelan, B. W. (2005). Extending site-specific crop management from individual fields to a whole farm. In J. V. Stafford (Ed.), *Precision agriculture '05* (pp. 857–863). Wageningen: Wageningen Academic Press.
- Florin, M. J., McBratney, A. B., & Whelan, B. W. (2009). Quantification and comparison of wheat yield variation across space and time. *European Journal of Agronomy*, *30*, 212–219.
- Fraisse, C. W., Sudduth, K. A., & Kitchen, N. R. (2001). Calibration of the CERES-maize model for simulating site-specific crop development and yield on claypan soils. *Applied Engineering in Agriculture*, *17*, 547–556.



- Godwin, D. C., & Jones, A. C. (1991). Nitrogen dynamics in soil–plant systems. In S. H. Mickelson, G. A. Peterson, C. W. Stuber, & R. J. Luxmoore (Eds.), *Modeling plant and soil systems* (pp. 287–321). Madison, WI: American Society of Agronomy Inc.
- Irmak, A., Batchelor, W. D., Jones, J. W., Irmak, S., Paz, J. O., Beck, H. W., et al. (2002). Relationship between plant available soil water and yield for explaining soybean yield variability. *Applied Engineering in Agriculture*, 18, 471–482.
- Irmak, A., Jones, J. W., Batchelor, W. D., & Paz, J. O. (2001). Estimating spatially variable soil properties for application of crop models in precision farming. *Transactions of the ASAE*, 44, 1343–1353.
- Jeffrey, S. J., Carter, J. O., Moodie, K. B., & Beswick, A. R. (2001). Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modeling and Software*, 16, 309–330.
- Kleijnen, J. P. C., & Sargent, R. G. (2000). A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research*, 120, 14–29.
- Link, J., Graeff, S., Batchelor, W. D., & Claupein, W. (2006). Evaluating the economic and environmental impact of environmental compensation payment policy under uniform and variable-rate nitrogen management. *Agricultural Systems*, 91, 135–153.
- Littleboy, M., Freebairn, D. M., Silburn, M., Woodruff, D. R., & Hammer, G. L. (1999). *PERFECT version 3.0: A computer simulation model of productivity erosion runoff functions to evaluate conservation techniques*. QLD Department of Natural Resources and QLD Department of Primary Industries, Brisbane.
- McBratney, A. B., & Whelan, B. (1999). The “null hypothesis” of precision agriculture. In J. V. Stafford (Ed.), *2nd European conference on precision agriculture* (pp. 947–957). Sheffield: Sheffield Academic Press.
- McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21, 239–245.
- Minasny, B., & McBratney, A. B. (2002). The neuro-m method for fitting neural network parametric pedotransfer functions. *Soil Science Society of America Journal*, 66, 352–361.
- Minasny, B., McBratney, A. B., & Whelan, B. (2002). *VESPER version 1.62*. Sydney: Australian Centre for Precision Agriculture.
- Mishra, S. K., & Singh, V. P. (2003). *Soil conservation service curve number (SCS-CN) methodology*. Water science and technology library (Vol. 42). Dordrecht: Kluwer.
- Morgan, C. L. S., Norman, J. M., & Lowery, B. (2003). Estimating plant-available water across a field with an inverse yield model. *Soil Science Society of America Journal*, 67, 620–629.
- Post, J., Hattermann, F. F., Krysanova, V., & Suckow, F. (2008). Parameter and input data uncertainty estimation for the assessment of long-term soil organic carbon dynamics. *Environmental Modelling and Software*, 23, 125–138.
- Reynolds, J. F., & Acock, B. (1985). Predicting the response of plants to increasing carbon dioxide: A critique of plant growth models. *Ecological Modelling*, 29, 107–129.
- Ruben, R., & van Ruiven, A. (2001). Technical coefficients for bio-economic farm household models: A meta-modelling approach with applications for Southern Mali. *Ecological Economics*, 36, 427–441.
- Sinclair, T. R., & Seligman, N. G. (1996). Crop modeling: From infancy to maturity. *Agronomy Journal*, 88, 698–704.
- Tamari, S., Wosten, J. H. M., & Ruiz-Suarez, J. C. (1996). Testing and artificial neural network for predicting soil hydraulic conductivity. *Soil Science Society of America Journal*, 60, 1732–1741.
- Timlin, D., Pachepsky, Y., Walthall, C., & Loechel, S. (2001). The use of a water budget model and yield maps to characterise water availability in a landscape. *Soil and Tillage Research*, 58, 219–231.
- Tranter, G. S., Minasny, B., McBratney, A. B., Murphey, B. W., McKenzie, N. J., Grundy, M., et al. (2007). Building and testing conceptual and empirical models for predicting soil bulk density. *Soil Use and Management*, 23, 437–443.
- Walter, C., McBratney, A. B., Douaoui, A., & Minasny, B. (2001). Spatial prediction of topsoil salinity in Chelif Valley Algeria using ordinary kriging with local variograms versus a whole-area variogram. *Australian Journal of Soil Research*, 39, 259–272.
- Wang, E., Robertson, M. J., Hammer, G. L., Carberry, P. S., Holzworth, D., Meinke, H., et al. (2002). Development of a generic crop model template in the cropping system model APSIM. *European Journal of Agronomy*, 18, 121–140.
- Wang, E., & Smith, C. J. (2004). Modelling the growth and water uptake function of plant root systems: A review. *Australian Journal of Agricultural Research*, 55, 501–523.
- Whelan, B., & McBratney, A. B. (2003). Definition and interpretation of potential management zones in Australia. In *Agronomy solutions for a better environment: Proceedings of the 11th Australian*

- agronomy conference*, 2nd–6th February, Geelong, Victoria. Australian Society of Agronomy. <http://www.regional.org.au/au/asa/2003/index.htm>.
- Wong, M. T. F., & Asseng, S. (2006). Determining the causes of spatial and temporal variability of wheat yields at sub-field scale using a new method of upscaling a crop model. *Plant and Soil*, 283, 203–215.