Advances in Animal Biosciences: Precision Agriculture (ECPA) 2017, (2017), **8:2**, pp 678–683 © The Animal Consortium 2017
doi:10.1017/S2040470017000851

advances in animal biosciences

Conceptual Spatial Crop Models for Potato Production

H. Chen[1†], I. Leinonen[1], B. Marshall[2] and J.A. Taylor[1]

[1]School of Agriculture, Food and Rural Development, Newcastle University, NE1 7RU, UK; [2]Consultant, 5 Muirloch Farm Cottage, Liff, Dundee, DD2 5NQ, UK

Advances in agricultural machinery, information and sensor technology have led to an increasing amount of data that is available spatially both pre and within season. The case is compelling for the spatialisation of existing, non-spatial (field-scale) crop models that can accommodate this 'big data' and lead to more precise predictions of yield and quality and an improved field management. This study explores the conceptual spatial models based on the potato crop models that simulate crop physical and physiological processes and predict yields and graded yields at a field-scale. Through exploring the possible spatial scales and model application approaches considering spatial variation an optimal and more effective solution is expected. Issues concerning model quality and uncertainty are also discussed.

Keywords: crop model spatialisation, spatial autocorrelation, spatial stratified heterogeneity

Introduction

Precision agriculture includes site specific crop management (SSCM), a varying management approach that considers the variation in terrain, soil, water and other environmental elements as well as management effects within fields (Whelan and Taylor, 2013). In order to adapt to this variation, the corresponding management is required to supply the exact amount of inputs (water, fertiliser, agri-chemical etc.) needed to promote healthy crop growth while simultaneously reducing unnecessary inputs, which may lead to environmental pollution and increase production costs.

SSCM requires site-specific data, which may include data on plant growth and the environment in which the plant grows. These spatial data can be categorised into two groups: relatively stable data, such as soil type and soil depth, and constantly changing data, such as time series of spatial canopy development (leaf area index), soil moisture deficit, soil temperature, solar radiation interception etc. In addition, there will be aspatial data for a field, such as cultivar information, uniform management activities and market price, which are needed for a full analysis of crop management decisions and associated risks.

Due to a lack of fine-scale spatial and spatial-temporal data, the ability of a grower to implement SSCM has not been straightforward. One practical simplified approach to SSCM is the use of a management unit (MU) approach, where several MUs are identified in a field that are assumed to have a relatively homogeneous crop response (Taylor *et al.*, 2007). Analysis and management are then conducted uniformly within individual MUs but differentially between MUs. With the increasing availability of data, it is possible to have smaller units for more precise management. This incremental expansion of the number of units can be regarded as belonging to a spectrum of precision agriculture, ranging from a traditional uniform field management through an interim management unit approach finally to a true SSCM approach potentially down to individual plants or at least sites within a field. Work to date has focussed primarily on empirical analyses of multi-variate field information layers to generate units for sub-field management (Ortega and Santibanez, 2007). The next step is to incorporate these information layers into crop models to improve the predictive power and utility of these models, particularly for tactical (in-season) crop management.

For potatoes, crop growth models, yield (quantity) models and tuber size distribution (quality) models are already commercially applied and are representative of existing research capacity in the sector. However, these are limited to point modelling at field or farm-scales using 'average' response functions and data. This is despite the fact that higher resolution data is available at sub-field scales. Consequently, the need to develop approaches for the spatialisation of these traditional field-scale crop models, by combining them with the emergent availability of 'big data' spatial information layers from various agri-sensor platforms, has become a compelling task for researchers in agriculture.

This study explores conceptual spatial crop modelling based on the non-spatial potato crop models in the Management Advisory Package for Potatoes (MAPP) (MacKerron *et al.*, 2004), which simulate crop physical and physiological processes to predict field-scale total yields and graded (marketable) yields.

† Email: hongyan.chen@newcastle.ac.uk

Crop growth models

There are many crop growth models for crop yield prediction and management (Steduto, 2003; Panday, 2014). In these models, the main physiological and physical processes include the development of leaf area, light interception and dry matter production, partitioning of dry matter, effects of soil water dynamics and, in some, nutrients (Kabat *et al*., 1995).

The potato growth model in MAPP simulates the development and growth process of a single plant, including stages of tuber sprouting, appearance and expansion of leaves, initiation of tubers and senescence. It calculates the fraction of radiation intercepted as a function of leaf area index, and then relates this to the dry matter production, which the soil moisture deficit affects (Jefferies and Heilbronn 1995). In MAPP, the dry matter partitioning is modelled using a thermal-time dependent 'harvest index' and the total yield is calculated by estimating the tuber dry matter concentration as a function of soil moisture deficit and temperature.

The growth model calculates a water balance daily from inputs of precipitation and irrigation and losses by evaporation from both soil and crop, and the total extractable soil water is obtained by defining the maximum rooting depth and multiplying it by the extractable soil moisture volume fraction. Nutrient dynamics and the dependence of plant growth on nutrients are not currently modelled in MAPP.

The tuber size distribution (TSD) model in MAPP is an empirical model, a truncated normal distribution. It estimates the proportion by fresh weight of daughter tubers produced in any specified size range, where tuber size is defined either by square mesh riddle size (mm) or by the fresh weight (g) of the individual tuber. In order to predict a TSD, one needs to know both the total yield and total number of daughter tubers produced which are greater than 15 mm square mesh riddle size. This is usually generated from mid-season manual digs.

The growth and yield model and the tuber size distribution model of the MAPP were developed by the James Hutton Institute (formerly Scottish Crop Research Institute), Dundee, Scotland. The growth and yield model is designed to simulate an individual plant that is representative of the field average. It does not aggregate over multiple individuals nor does it model the interaction between neighbouring plants. However, it is known that there is variability within a field in many of the drivers of the model, particularly the canopy development and size and the availability of soil water supply to the crop (Allaire *et al*., 2014).

Spatialisation - conceptual spatial crop modelling

There are several potential ways that the MAPP model (or any crop model) could be made spatial. It could be simply *spatialised* through running the existing model independently at a finer spatial resolution ignoring neighbourhood effects, or converted into a *true spatial model* that runs at a finer spatial scale and incorporates information on the effect of the neighbourhood in the modelling process. The latter, modifying existing models by integrating spatial interactions, is preferable but more complex and not practical at the current stage.

The possibility for running the model at a finer spatial scale lies in the availability of spatial data. Currently MAPP uses aspatial environmental input data, including weather data and soil profile information, for the simulation of each field/farm. However, fine resolution spatial data such as soil electrical conductivity, crop growth imagery, and sampling data of soil moisture is increasingly available. Table 1 compares the current model input with current available data, highlighting the gap

Table 1 *Key input for the MAPP growth model (as currently specified) with indications of spatial data that could be used as input for a spatialised model.*

Input Requirement	Spatialise or not	Potential replacement with the data available
Husbandry info:		
planting date	yes if necessary	Variable rate planting, variable rate haulm destruction (plant growth regulator)
haulm destruction date		
planting spacing, etc.		
Plant development:		
Emergence date	yes	ground cover of different time points taken from high resolution imagery (UAV, Aerial) or from remote or proximal Vis-NIR sensors
Stem density	yes	
Date of Senescence	yes	
Soil profile:		
Soil type	not yet	Apparent soil electrical conductivity; networked soil moisture probes
Horizon depth		
Weather data:		
Solar radiation	not yet	Modelled based on local Digital Elevation Model and weather parameters
Other variables	not currently	
Test digs:		
Weight of tubers	yes	Mobile phone-based apps for counts and size distribution
Number of tubers	yes	

between the spatial data available and the data required for modelling at a finer spatial scale. However, there exists a potential to make use of the spatial data for spatial modelling. The challenge now lies in which scale the model should be applied and how to replace the input with the spatial data we have.

Possible scales to run the MAPP model at include but are not limited to:

1. Individual plant
2. Bed planting spacing (~3 m*3 m grid)
3. An arbitrary pixel size (e.g. 5 m^2 or 10 m^2)
4. Management units
5. Whole field

To determine which scale is the best to run the model is a complex task, depending on multiple criteria: including the purpose, data availability and model requirements. Generally, the purpose is to provide growers with as accurate information as possible for their decision-making, so the finest scale is preferred. Since the original model simulates individual plants, applying the model at a scale smaller than a single plant scale is not sensible. The whole field, the original scale of output of the MAPP models, is obviously not appropriate for this project either.

Considering the nature of the basic MAPP model, the more reasonable choices would be choices 1 and 2. However, due to a lack of corresponding input data at those scales, these choices would be better regarded as a long-term target for crop production and management. As a simplified interim step, modelling crop development and yield/tuber distribution can be done at a MU level. This MU scale is variable, depending on the number of management units specified. This is normally limited to a maximum of 3 or 4 units, and often only 2 (Pedroso *et al.*, 2010), due to the feasibility of machinery operations.

With advances in agricultural machinery, in the foreseeable future, variable-rate operations are unlikely to be a factor that limits the scale of SSCM. It is more likely that limitations in scale will be imposed by the ability to process, interpret and generate decisions from the multitude of information sources now available to growers/agronomists. In this study, the intent was to identify, without considering the limitation of operation, what management unit level maximally integrates the data available and is the most accurate when applying the models to estimate the yield and graded yields of the field and assist with the field management.

Spatial variability is the driver for model spatialisation. The other challenge (as mentioned above) is to find spatially varying data that reflect different environmental conditions and management activities of different areas within a field as model input to run the model spatially for more accurate prediction and decision-making. In many cases, available spatial data are non-identical to required input variables, so investigating how to transform them becomes the main task. To identify what spatial data to use to replace some certain model inputs and how, their influence on the crop yield needs to be analysed and appropriate methods for their inclusion to replace the original homogeneous variables need to be determined.

With in-season stratified sampling of crops, it is possible to test the spatial variability of their dry matter, tuber weight and number, plant and stem density, etc. Also, with stratified heterogeneity test (Wang *et al.*, 2009), we can see if they are significantly different among the strata, derived from soil ECa. If yes, then the ECa probably is the reason for the heterogeneity. This test could also be done based on strata other than those used for sampling, even on the strata generated with different criteria, e.g. planting date. In this way, an optimum MU level would be found for a criterion (here each unit does not have to be contiguous) and potential influential factors would be identified.

As mentioned above, quantification of influential factors as model input is desirable and challenging, requiring a thorough exploration through lab experiments or data analysis. Once this is achieved, the crop models can be applied to the determined units with varied input. Decisions that can be made using the model for the field as a whole can be made for each unit, for example, when to irrigate, where and how much. Figure 1 illustrates the whole process, in which model spatialisation is realised through several steps: determination of optimal units, preparation of varied input, model application to spatial units and at the end, uncertainty analysis.

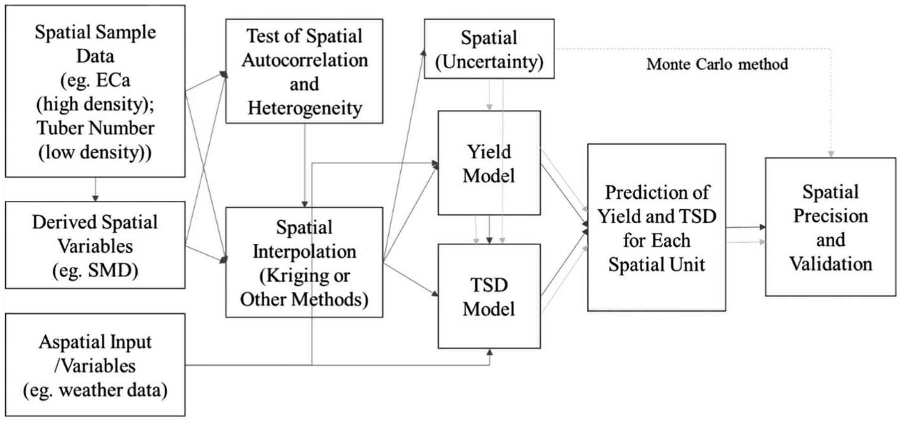

Figure 1 Spatialisation through applying the models to each spatial unit

Table 2 *Autocorrelation test with Moran's I for selected crop parameters mid-season and at harvest (n = 100)*

	Mid-season Dig			Harvest Dig		
	Yield	DM	Stems	Yield	DM	Stems
Moran's I	0.302	0.253	0.165	0.161	0.344	− 0.115
z-score	3.846	3.238	2.161	1.465	4.230	−0.819
p-value	0.0001***	0.0012***	0.0307**	0.1428	0.0008***	0.4130

**5% significance level; ***1% significance level

Table 3 *Stratified heterogeneity test with q statistic for different MUs derived from different spatial criteria.*

Statistic	ECa (3 MUs)	Soil Type (2 MUs)	Planting date (2 MUs)	ECa (3 MUs)	Soil Type (2 MUs)	Planting date (2 MUs)
		Yield			*Canopy cover*	
		Mid-season Dig			18 June 2015	
q statistic	0.057	0.066	0.374	0.054	0.137	0.715
p value	0.212	0.032**	0.000***	0.075*	0.000***	0.000***
		Harvest Dig			6 July 2015	
q statistic	0.061	0.052	0.133	0.082	0.220	0.896
p value	0.826	0.413	0.115	0.020**	0.000***	0.000***

*10% significance level; **5% significance level; ***1% significance level

Some experimental results

A survey was performed in a field in Yorkshire, UK for a preliminary case study and the results are listed in Tables 2 and 3, where sampling data of yield, dry matter, stems and canopy cover have been processed and analysed. In this field, multiple data collections have been conducted at 100 sampling points that were selected based on an existing apparent soil electrical conductivity map (EC_a).

Spatial autocorrelation helps understand the degree to which one object is similar to other nearby objects. Moran's I (Index) (Moran, 1950) is used to measure spatial autocorrelation. As we can see from Table 2, the yield, DM and stems of the first dig have significant spatial autocorrelation, while only the DM variable exhibits autocorrelation at harvest. Through *cluster and outlier analysis* in ArcGIS (ESRI, Redlands CA, USA), which identifies statistically significant hot/cold spots and outliers using the Anselin Local Moran's I statistic, it is easy to find that the high value clusters are all concentrated on the northern section of the field.

The field was then divided into two units, a northern and a southern unit, and tested for stratified heterogeneity with the q statistic (Wang *et al.*, 2016). The results are presented in Table 3, where the greatest q value is 0.896, which means that 89.6% of variance of the canopy cover measured on July 6, 2015 can be explained by the division and is statistically significant. Farmer's planting at the northern part approximately 3 weeks earlier than the southern was identified as a major reason for this variance.

Therefore, the influence factor leading to the difference between these two units is planting date. Since planting date is directly a model input, the process of transforming the influence factor as an input can be omitted and the varied input for planting date are directly formed by the two actual dates. Applying the growth model and graded model to the two units with the varied planting dates, yield and tuber size distribution predictions were generated from the model, and the accuracy can be calculated through comparing with the observed unit averages.

Although the above is a simple case study, in principle, it embodies the essence of the spatialisation process in Figure 1. The situation could be more complex if the influence factors are environmental, not managerial. For example, with soil EC_a, the field can be split into management units (classes) using k-means classification (Taylor *et al.*, 2007) (Fig. 2) in a number of ways. Testing their stratified heterogeneity using the canopy cover of 6 July, 2015, it can be seen that the most statistically significant q value is 0.31, indicating that 5 management units is optimal (see Table 3). Statistically this level is more reasonable than others for transforming soil ECa as varied input for model simulation. Combining varied input (layers), managerial or environmental, could lead to a finer spatial scale for model application.

Discussion

The preliminary results presented above shed a light on where to go to achieve a sound spatial crop model for agricultural practice. However, many hurdles are in the way: where is direct incorporation of a spatial data layer possible? where is surrogate spatial data available and what transformation is required for use in the crop model? In addition, how to integrate higher resolution data, especially imagery

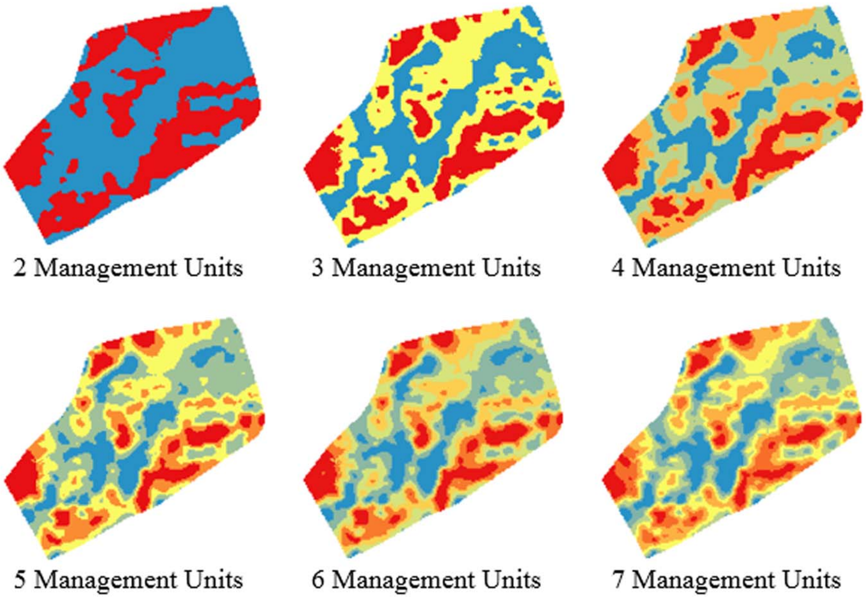

Figure 2 Potential Management Units based on classification for 2–7 classes using the SoilEC$_a$ data

Table 4 *Stratified heterogeneity test with the q statistic for mid-season canopy cover based on management units derived from EC$_a$ classification.*

Statistic	2 MU	3 MU	4 MU	5 MU	6 MU	7 MU
			Canopy Cover *6 July 2015*			
q statistic	0.019	0.099	0.161	0.319	0.262	0.351
p value	0.431	0.123	0.156	0.028**	0.097*	0.116

*10% significance level; **5% significance level

data such as canopy cover, to update and 'reset' the model during the season. All these need to be further explored thoroughly through lab experiments or data analysis before reaching our target, namely, applying the models spatially.

Uncertainty in spatialisation has to be analysed through comparing our model estimations with the field samples for a more informative model application. There are two possible ways to improve the model: 1) The first is to use the measured data for iterative adjustment of the model's initial conditions and to determine cultivar specific parameters, so that the model's predictions agree with periodic remotely-sensed measurements of a modelled variable e.g. LAI; 2) The second approach is to correct the predictions of the model through comparing the real data at a few validation sites with the estimated values from the model and to create a spatial adjustment coefficient to correct model output spatially across the field.

One important point also to consider is whether the grower or modeller know what the cause of diversion between observation and prediction is. If one does and there is a measurement of the independent variable causing this divergence, then the first approach above is possible. If there is more than one known cause then some means of weighting or combining the effects of the various causes is needed. Finally, if the cause is unknown, then only the second approach is an option. Regardless of the approach, there will also be nugget variation (unexplained random variation) in the data, for which the Monte Carlo method can be used to provide an estimation.

Conclusions

This study explored the conceptual spatial models based on non-spatial potato crop models, which simulate crop physical and physiological processes and predict yields and graded yields at a field-scale. Through preliminary experiments that delve into optimal spatial scales, more effective approaches to model application considering spatial variation and dependence in environmental variables and crop development, a favourable scale and approach is appointed and the issues concerning model quality and uncertainty are discussed.

Acknowledgements

This project was funded by Innovate UK and BBSRC.

References

Allaire SE, Cambouris AN, Lafond JA, Lange SF, Pelletier B and Dutilleul P 2014. Spatial Variability of Potato Tuber Yield and Plant Nitrogen Uptake. Agronomy Journal 106, 851–859.

Kabat E, Marshall B, van den Broek BJ, Vas J and van Keulen H 1995. Modelling and Parameterisation of the Soil-Plant-Atmosphere System – A Comparison of Potato Growth Models. Wageningen Pers, Wageningen, The Netherlands.

Jefferies RA and Heilbronn TD 1991. Water stress as a constraint on growth in the potato crop. 1. Model development. Agricultural and Forest Meteorology 53, 185–196.

MacKerron DKL, Marshall B and McNicol JW 2004. MAPP and the underlying functions that it contains. In D.K.L. MacKerron and A.J. Haverkort (eds) Decision support systems in potato production: bringing models to practice. Wageningen Academic Publishers, The Netherlands.

Matheron G 1963. Principles of geostatistics. Economic Geology 58, 1246–1266.

Moran PAP 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17–23.

Ortega R and Santibanez O 2007. Determination of management zones in corn (Zea mays L.) based on soil fertility. Computers and Electronics in Agriculture 58 (1), 49–59.

Panday SG 2014. Use of remote sensing data in a crop growth model to estimate actual crop. http://www.itc.nl/library/papers_2014/msc/nrm/gangarampanday.pdf (retrieved 12/01/2017).

Pedroso M, Taylor JA, Tisseyre B, Charnomordic B and Guillaume S 2010. A segmentation algorithm for the delineation of management zones. Computers and Electronics in Agriculture 70, 199–208.

Pringle MJ, McBratney AB, Whelan BM and Taylor JA 2003. A preliminary approach to assessing the opportunity for site-specific crop management in a field, using a yield monitor. Agricultural Systems 76, 273–292.

Steduto P 2003. Biomass water-productivity. Comparing the growth-engines of crop models. FAO Expert Meeting on Crop Water Productivity under Deficient Water Supply, Rome. February 26–28.

Taylor JA, Whelan BM and McBratney AB 2007. Establishing management classes for broad acre agriculture. Agronomy Journal 99, 1366–1376.

Whelan BM and Taylor JA 2013. Precision Agriculture for Grain Production Systems. CSIRO Publishing: Collingwood, Australia.

Wang JF, Christakos G and Hu MG 2009. Modelling spatial means of surfaces with stratified non-homogeneity. IEEE Transactions on Geoscience and Remote Sensing 47, 4167–4174.

Wang JF, Zhang TL and Fu BJ 2016. A measure of spatial stratified heterogeneity. Ecological Indicators 67, 250–256.