Precision agriculture and geostatistics

How to manage agriculture more exactly

Since the very beginnings of agriculture, farmers have known which corners of their fields were wetter, which drier, which sandier and where the crops grow best. Now, says **Margaret A. Oliver**, technology and statistics are taking this age-old knowledge to a new level. It is called precision agriculture.

Intense sampling of the soil and crops using satellite imagery, GPS, and high-tech sensors; sophisticated analysis of the relationships between them; tractors with drills and sprayers linked into those data that vary the amounts of seed, irrigation water, fertilizers, and pesticides as they go, and that fine-tune those applications to a few square metres and that even steer themselves while they are doing it; these are some of the ingredients of modern precision agriculture. It may seem a long way from the subsistence farmer pacing his fields and feeling the soil with his fingers, but there is an unbroken succession between them.

Precision agriculture in the modern sense uses information and technology to identify and manage the variation within farmed fields. The information can come from any, or frequently all, of the sources we have mentioned. Using it, inputs can be varied to maximise profitability, sustainability and protection of the environment. The aim is to manage variation within fields so that individual areas receive the kinds and the quantities of inputs that are tailored precisely to their needs.

Precision agriculture

It has been described as a new paradigm in agriculture, but if we look back we discover that farmers have practised a form of precision agriculture since the very earliest days. Subsistence farmers have always known their fields intimately. They know which patches of their land are most suited to different crops, and which produce the most. For subsistence farmers precision was and still is about ensuring enough food for the family. Rothamsted Research Centre, near Harpenden in Hertfordshire, has a long history of agricultural research, starting with the work of Gilbert and Lawes in 1843 – which was also about precision farming. They wanted to assess the benefits of different combinations and amounts of crop nutrients in their field experiments. The great R. A. Fisher, perhaps the founder of modern statistics, was eventually employed at Rothamsted in 1919 to develop statistical methods to analyse the results of these experiments. Fisher and

Photograph courtesy of Rothamsted Research

Precision agriculture

Precision agriculture is now a global phenomenon. Some aspects of it – soil sampling, autoguidance, yield mapping – are used in probably more than 10% of food production. A seven-year research programme in Brazil saw maize yields increase by an average of 14% and soybean by 10%. As well as wheat and arable crops, Brazil is actively researching its use in rice, cotton, pasture, sugar cane, oranges, and even eucalyptus trees. Or it can be used to increase quality at the expense of yield, as for example with grapes for wine.

Much data for precision agriculture comes from sensors on satellites or planes. They measure the wavelengths of radiant energy that are absorbed or reflected from the land surface – showing geology and soil conditions – or from the crop itself; measurements in visible, near-infrared, thermal infrared, and microwave wavelengths can show how efficiently plants are photosynthesising and whether they are thriving or are under stress. According to NASA's Landsat 7 Science Team[1], an optimum remote sensing system for precision farming would provide data as often as twice per week for irrigation scheduling and once every two weeks for general crop damage detection. The spatial resolution of the data should be as high as 2–5 square metres per pixel. Perhaps even more importantly, the data must be available to the farmer within 24 hours of being acquired.

No satellite currently fills all these requirements. Instruments aboard the satellites *Landsat 7* and *Terra* provide good spatial and spectral resolution (up to 15 metres per pixel) as well as good positional calibration accuracy, but these satellites overfly each area too infrequently for regular sensing: farmers may prefer more frequent but lower-resolution data to high-resolution but less up-to-date data. Another sensor carried by *Terra*, the Moderate-Resolution Imaging Spectroradiometer (MODIS), sees a given patch of ground almost every day but its highest spatial resolution is only 250 metres per pixel – too coarse for precision farming.

Sensors on aircraft can fill the gap, collecting data in visible and near-infrared channels at resolutions ranging from 0.3 to 1 metre per pixel; commercial plane operators will overfly as frequently as the farmer is willing to pay for. The images here were acquired by a Daedalus sensor aboard a NASA aircraft flying over the Maricopa Agricultural Center in Arizona. The top image shows colour variations determined by crop density, where dark blues and greens indicate lush vegetation and reds show areas of bare soil. The middle image is a map of water deficit, derived from reflectance and temperature measurements. Greens and blues indicate wet soil and reds are dry soil. The bottom image shows where crops are under serious stress, as is particularly the case in fields 120 and 119, indicated by red and yellow pixels. These fields were due to be irrigated the following day.

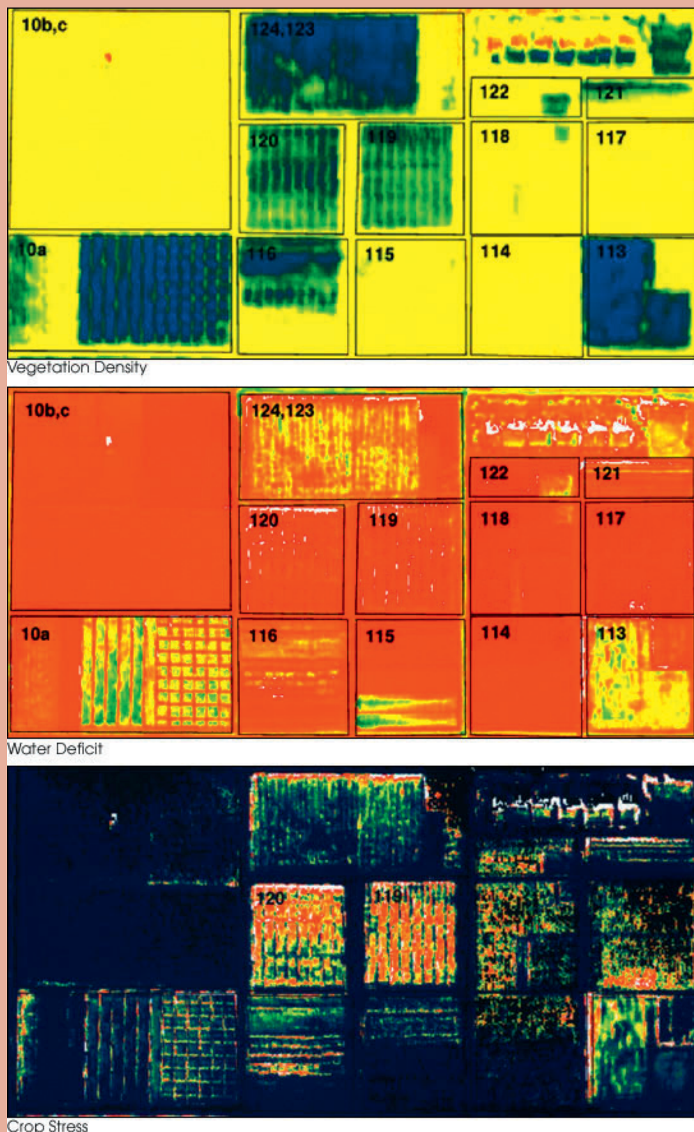


Image courtesy Susan Moran, Landsat 7 Science Team

his colleagues worked on medium-sized plots within fields.

Scales of farming changed as we entered the modern era after the Second World War. The size of fields increased with the growth in farm machinery, but within each farm, management was at the level of the entire field. Soil samples were taken from various parts of the field and often mixed together to form a single bulked sample on which concentrations of crop nutrients and pH were determined, and the amount of additives needed. Fertilizers, lime, seeds so on were applied as evenly as possible over the field. The whole field was managed as a unit and its crop yield was simply the total weight taken from it. Nevertheless, farmers were well aware that there was variation within their fields and that some patches yielded more than others in spite of the uniform treatment.

There were also concerns about the environmental effects of the intensification of agriculture. From the 1970s it was recognised that too much nitrogen and phosphorus in ground and surface waters and losses of pesticides and soil from fields were harmful. Environmental as well as economic pressures are demanding that inputs be minimised without prejudicing yields; farmers are becoming more aware of different approaches to management so that the land remains in a condition to continue farming it for future generations.

The driving force behind modern precision agriculture has been a technological revolution based on information technology[2]. This is what has allowed the change in the scale of management. Before, we considered the field as uniform; now we are treating the variation within it. This new focus enables more precise local management. The aim is to apply nutrients, ameliorants, seeds, pesticides and water at levels appropriate to the soil conditions in terms of fertility, drainage and so on. This much finer-scale management avoids the over- and under-application which inevitably means a loss in farm profit and sometimes means damage to the environment as well.

It needs, of course, technology. One of the most significant steps in the development of precision agriculture was the introduction of a yield meter by Massey Ferguson in 1982. It is a small device, fixed on the grain elevator of a combine harvester; it meant that yield could be recorded continuously for the first time. In 1984 Massey Ferguson carried out a United Kingdom field trial. At the time global positioning systems (GPS) were not available and the position within the field was recorded manually using poles stuck into the ground at ten-metre intervals as a guide; the yield of wheat was found to vary by as much as 10 tonnes per hectare over the field. The first GPS became available on tractors in 1991, and in the 1990s yield mapping became fairly routine. The arrival of microprocessors[3] brought computers that could be placed on farm equipment to process, on the job, the large volumes of spatial data acquired from yield monitors and electronic sensors. We now have equipment that applies fertilisers at variable rates over a field automatically according to the variation in nutrients already in the soil or crop characteristics, or that spray water or pesticides according to water content in the field or where the weed patches are.

Sparse data – and the solution

Micro-managing variation within fields requires sufficient information to describe the variation. Some data are available at high resolution – yield, from yield meters, remotely-sensed data, from satellites or aerial photographs; data obtained proximally – that is, on-site, by mounting a sensor on a tractor or dragging it over the field; digital elevation models. There is, though, an elephant in the room because many other types of data are available only at large spatial resolutions – that is, with large distances between the sampling points. A major stumbling block to the wider spread and adoption of precision agriculture is the sparsity of soil and some crop information. We need detailed maps of soil and crop properties; but the data to make such maps are far from complete.

This is where an essential partner to precision agriculture comes in. It is no accident that precision agriculture and geostatistics are brought together in the title of this article[4]. Before the 1990s maps, other than of the soil and possibly landscape, played little part in agricultural management. Geostatistics can now provide the maps.

Geostatistics

Geostatistics is a technique that can predict values between sample points accurately. If, for reasons of cost or practicality, there are data for only a few points within a field, geostatistics provides the tools to determine what lies between. It had its origins in the 1960s, to evaluate reserves for the mining industry. Drilling core samples is expensive, so miners have to space them far apart; nevertheless, they need to know the variation of the ores in between. But many of its ideas have a longer history. For example, the famous statistician Student (in his other life William Gosset, chief brewer for Guinness in Dublin) in an appendix to a 1911 paper by Rothamsted researchers[5], observed that yields in adjacent plots were more similar to one another than between other plots further away. He suggested that there were two sources of variation: one that was spatially correlated or dependent – plots near to each other resemble each other – and another that was completely random or spatially uncorrelated. This component of variation is unaffected by what the adjacent plot is like. Distinguishing these two components of variation is a central idea in geostatistics.

Geostatistics – the variogram and kriging

The central tool of geostatistics is the variogram. The variogram summarises the way that properties vary from place to place; the usual method to estimate it[7] from data is with the equation

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2,$$

where $z(\mathbf{x}_i)$ and $z(\mathbf{x}_i+\mathbf{h})$ are the actual values of the variable, Z, at places $\mathbf{x}_i$ and $\mathbf{x}_i+\mathbf{h}$, and $m(\mathbf{h})$ is the number of paired comparisons at lag $\mathbf{h}$, where $\mathbf{h}$ is a vector in both direction and distance. The experimental or sample variogram is obtained by changing $\mathbf{h}$.

The experimental variogram must be fitted by a model to describe the spatial variation[6]. Figure 1a shows an annotated example of one of the most commonly fitted models, the spherical function. In this variogram there is a positive intercept on the ordinate known as the nugget variance (c_0), which represents the variation that has not been resolved at the scale of sampling. The monotonic slope of the variogram indicates that as the distance between points increases, the variance increases as the values become increasingly different. This is the spatially correlated variance (c). The slope eventually reaches an upper bound in many models, the sill variance ($c + c_0$), or *a priori* variance, σ^2, of the process. The distance at which the variogram reaches its sill is known as the range of spatial dependence (a); places separated by distances less than the range are autocorrelated and those at distances greater are uncorrelated or spatially independent.

The parameters from the best fitting model can be used for geostatistical prediction or kriging. In precision agriculture kriging is used for interpolation, hence to make maps. Kriging is an optimal method of prediction in the sense that it is unbiased and the predictions have minimum variance.

In addition to the predictions, kriging estimates the errors at each prediction point. It is a method of local weighted moving averaging of the observed values of Z within a neighbourhood near to the point or block to be predicted. There are many types of kriging, but ordinary kriging is the method most widely used. If the variable, Z, has been measured at sampling points, $\mathbf{x}_i$, $i = 1, \ldots, N$, we use this information to estimate its value at the unknown point, $\mathbf{x}_0$, by

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i z(\mathbf{x}_i),$$

where n usually represents the data points within the local neighbourhood, which is much smaller than the total number of data, and λ_i are the weights. The weights are a function of the variogram and the location of the sampling points and target point. To ensure that the estimate is unbiased the weights are made to sum to 1.

A second key concept is the *range of spatial dependence*. This is the distance beyond which there is no longer any correlation between points. Places separated by distances less than the range are correlated; those at distances greater are uncorrelated or spatially independent. Geostatistics uses the spatial correlation between places that are close together to describe the variation quantitatively using the variogram. This is a diagram,which summarizes how properties vary from place to place; it is based on the square of the difference between pairs of sample values.

It was Danie Krige, a mining engineer in the South African gold mines, who in the early 1950s observed that block and core sample grades were correlated. Matheron[4] expanded Krige's empirical ideas, in particular the concept that neighbouring samples could be used to improve prediction, and put them into the theoretical framework of the theory of random processes that underpins geostatistics[6]. Fundamentally, geostatistical prediction – kriging – predicts the value of any property – ore concentration, soil acidity and so on – at any point x in the area from sparse data. Kriging takes all nearby sample points into account, suitably weighted according to the variogram and their positions relative to one another and to the point to be predicted. Maps based on these kriged predictions can be used in precision agriculture. For the mathematics, see the box. Geostatistics has now been applied in many different fields, such as fisheries, hydrology, geology, meteorology, petroleum, epidemiology, remote sensing, soil science – and precision agriculture.

Case study

We can see how these ideas come together in practice with a case study – a field on the Yattendon Estate in Berkshire[7]. The 23-hectare field has a complex geography, with plateau areas in the north and east and a large dry valley in the centre and south of the field. The landscape is generally undulating, a characteristic of the chalk hills in England. The soil has a range of textural types with areas that are more sandy and others that are more clayey.

Farmers can use many types of data to aid the best management of their fields. For example, the form and slopes of the landscape play a crucial role because of their effect on soil water moisture conditions. Sloping areas shed water and may suffer from drought, whereas more level areas store water and the soil is more moist and may even become waterlogged in wet seasons. Soil moisture has a major effect on the crop growth. Digital elevation models of fields are used with other digital maps to aid the farmer in decision-making. Many of the maps are made by geostatistical interpolation – kriging.

Figure 1a shows an annotated diagram of the experimental variogram (symbols) of sand content in the soil; the solid line is the fitted spherical model. The range of spatial variation of sand is about 260 metres – points further apart than this are uncorrelated in their sand content. Figure 1b shows the kriged map of predictions of sand content made using the parameters of the spherical model. The central white feature is a small copse of trees that is excluded from the map. The areas with sand content less than 43% (the blue areas) are where the clay content

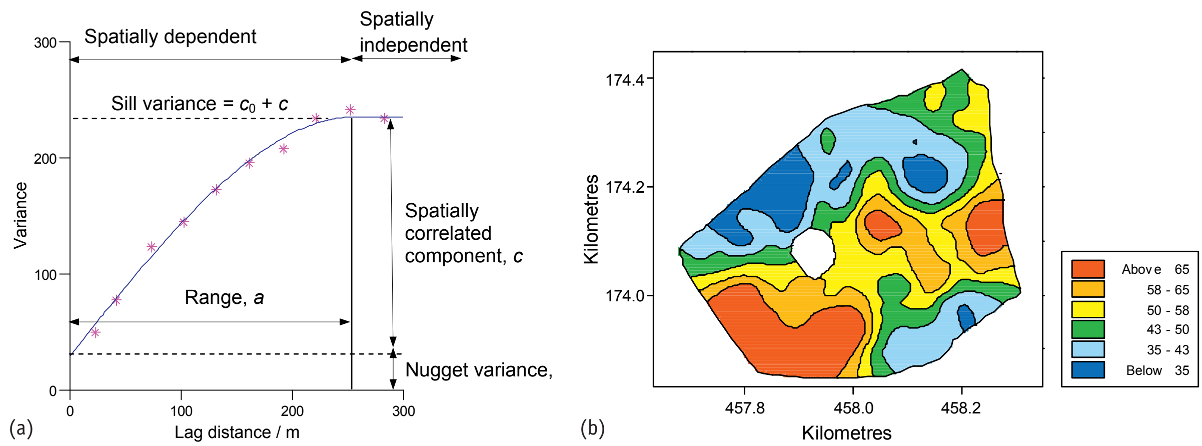


Figure 1. (a) A variogram, the basic tool of geostatistics. (b) A kriged map of sand content in the case study field.

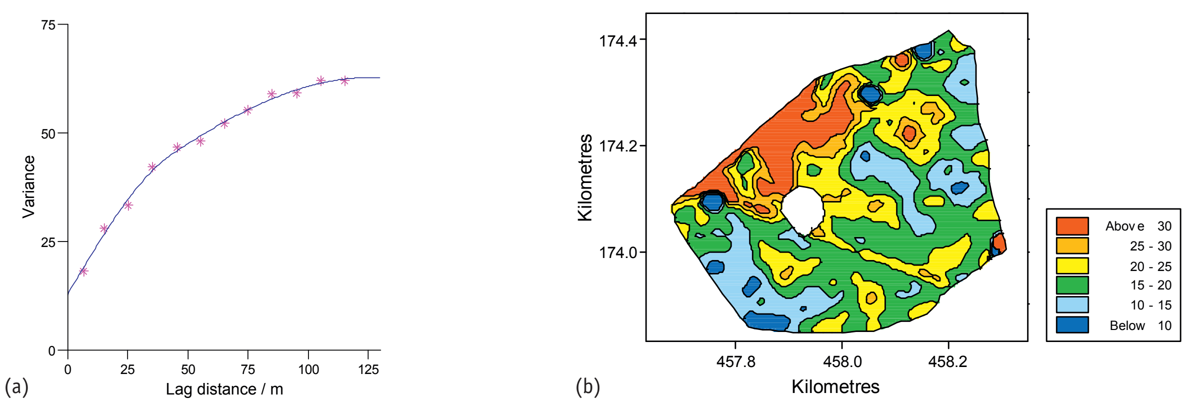
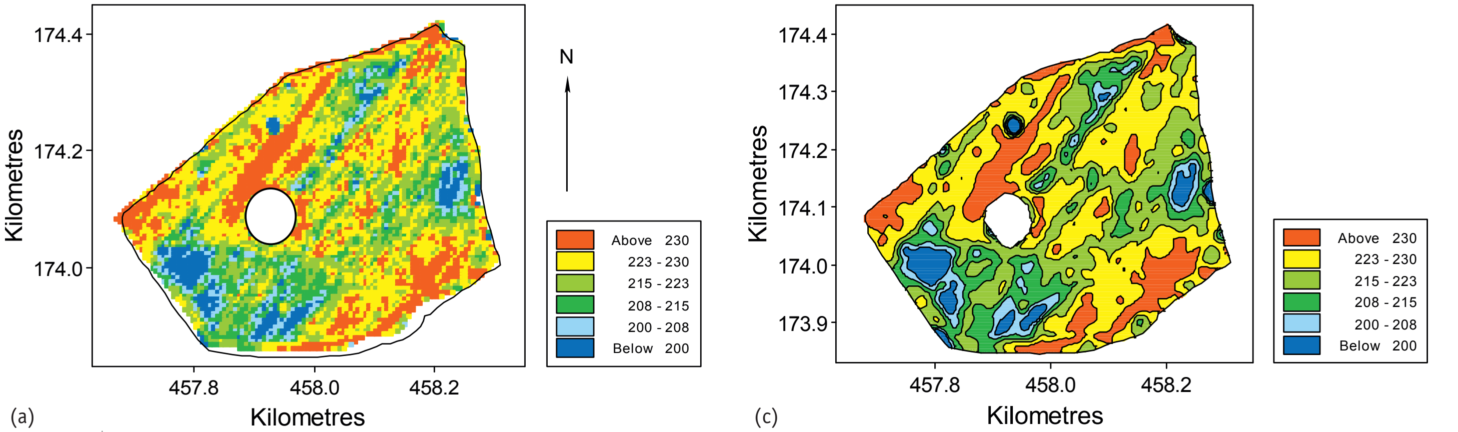


Figure 2. (a) Variogram and (b) kriged map of the electrical conductivity of the field



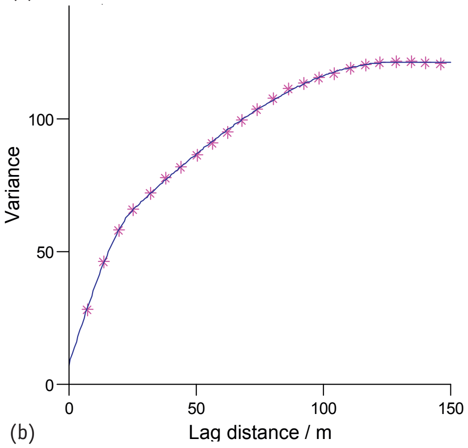


Figure 3. (a) Pixel map, (b) variogram and (c) kriged map of red wavelength of stubble, from data from air photography

is greater than 25%, and they relate closely to the topography of the field. They occur on the plateau top in the northwest of the field and at the bottom of the dry valley in the southeast. The water holding capacity of sandy soil is generally less than that of a clayey soil. The large-scale patterns in the variation of sand content reflect the long scale of variation described by the variogram. If irrigation were required in this field it could be managed in a precise way based on a digital map such as this and more water would be applied to the central (red and orange) band where sand content is greatest.

The map of sand content was based on an intensive sampling on a 30-metre grid supplemented by sampling along randomly selected transects at an interval of 15 m. Most farmers cannot afford this level of investment in sampling, and often rely on data from satellites or aircraft, or from tractor-borne equipment.

An example of the latter is electromagnetic induction (EMI) of the soil, which is measured by a sensor, dragged along behind a tractor, that records the soil's apparent electrical conductivity, EC_a (in millisiemens per metre). Soil EC_a is strongly affected by soil moisture, salinity, texture and organic matter, but is not as yet a precise science. Nevertheless, in our case study field there is a strong relation between the spatial pattern in EC_a and the sand content as shown by the kriged map in Figure 2b – but it is not absolute. Areas with EC_a greater than 30 mS m^{-1} have the largest soil clay content and areas with values less than 15 mS m^{-1} are the most sandy. The variogram for EC_a has two changes of gradient, indicating that this property varies at two spatial scales of about 43 m and 130 m, both shorter than the scale of variation for sand. The variation in EC_a is more complex than that of sand, which supports the fact that relations between soil properties are not straightforward.

The study uses remotely-sensed data as well. Figure 3a shows the pixel map of digital numbers of the red waveband generated from a colour aerial photograph of the field taken in 1991 when the field was in stubble; Figure 3c shows the kriged map of these values. The kriged map is much easier to use than the raw digital image, and shows that kriging is valuable not only when the data are sparse, but also when they provide full cover but are noisy. Figure 3c shows a stronger relation with sand content in the field than does EC_a, although the latter is also related to the features in the red waveband. The variogram of the red waveband also describes two scales of spatial variation of about 25 m and 130 m; the longer-range variation is the same as that for EC_a. Figure 4 shows the variogram and kriged map of available potassium content in the soil.

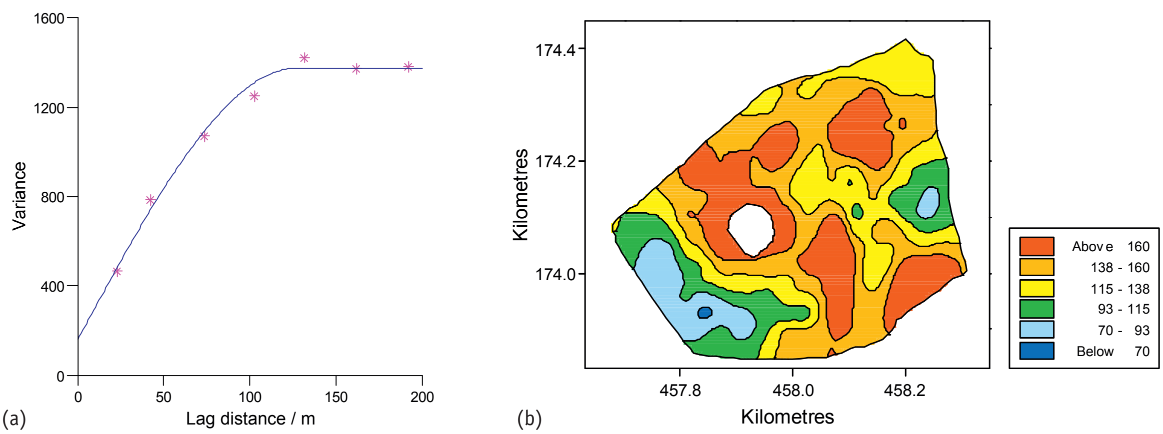


Figure 4. (a) Variogram and (b) kriged map of potassium content in the soil.

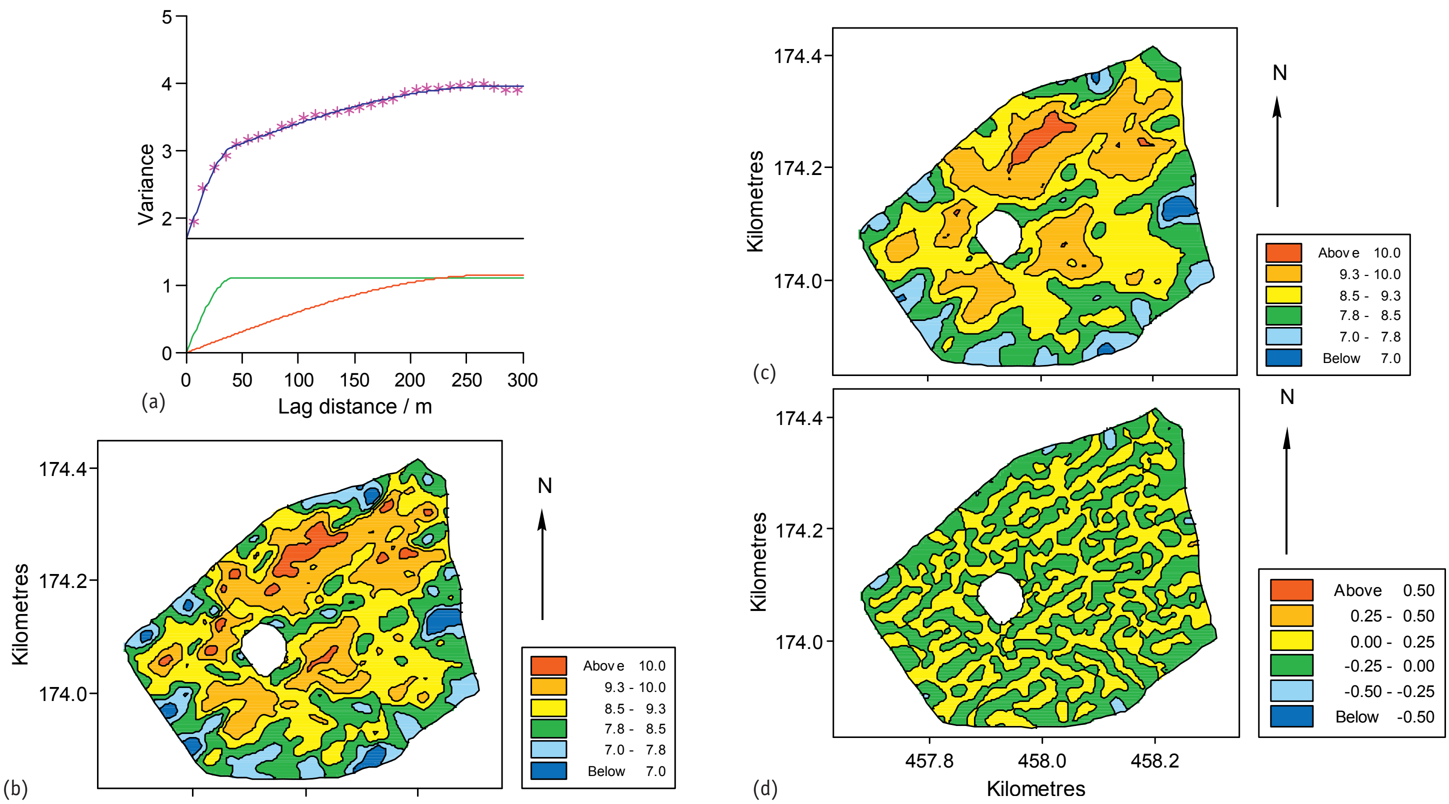


Figure 5. (a) The various components of the variation have been extracted from the full model (blue line): the nugget variance is the horizontal line, the short-range variation is the green line and the long-range variation the red line. The long-range variation in yield relates to the long-range variation in EC_a and the red waveband and also the scale of variation in potassium. (b) The kriged map of yield shows considerable spatial variation with both small and larger features that reflect the two scales of variation present. (c) The spatial patterns of the long-range variation in yield show a relation with elevation and sand and clay content; the plateau area has the largest yields. The small digital numbers in the aerial photograph and the smaller EC_a values (Figures 2b and 3b, respectively) also appear to relate to the lower-yielding areas. (d) The kriged map of the short-range component of the variation, which is also evident in the pixel map of the red waveband (Figure 3a); there is also some evidence of other lines of management at right angles to the dominant lines

Soil conditions are the first half of the story; the denouement is the yield. Figure 5 shows the variogram of wheat yield for 1997, with kriged maps for both the long-range variation (of about 130 m) and the short-range (of about 30 m). The long-range variation (Figure 5c) shows a relation with elevation and sand and clay content; the plateau area has the largest yields. The short-range (Figure 5d) shows the effects of management. The tramlines for vehicle movement in the field have recently been in a NE–SW direction. The green areas reflect areas of smaller yield and the effects of soil compaction from farm traffic over the field. This is especially noticeable at the field margins where the tractors turn.

The farm managers used the results of the study to vary their applications of farmyard manure, putting more on the lower, sandy, regions; and tried increasing the wheat seed rate on the western and northwestern parts of the field – with such success that they are now varying the seed rate automatically on most of their fields. They are using satellite imagery of soil zones to vary their nitrogen applications.

These site-specific management efforts are based on geostatistical analyses of data from sensors both near to the field and distant. They are both economically and environmentally sound., the epitomy of precision agriculture. They are becoming increasingly standard for all kinds of crop production and will be of crucial importance in the near future as the world faces increasing issues of food security.

Acknowledgements

I thank the HGCA for funding the project that provided the data for the Yattendon Estate and also the Estate management for allowing us access to their land.

References

1. Moran, M. S. (2000) Image-based remote sensing for precision crop management – a status report. In S. W. Johnson, K. M. Chua, R. G. Galloway and P. I. Richter (eds), *Space 2000: Proceedings of the Seventh International Conference and Exposition on Engineering, Construction, Operations and Business in Space*, pp. 185–193. New York: American Society of Civil Engineers.

2. Schueller, J. K. (1997) Technology for precision agriculture. In J. V. Stafford (ed.), *Precision Agriculture '97*, pp. 19–33. Oxford: BIOS Scientific.

3. Robert, P. C. (1999) Precision agriculture: research needs and status in the USA. In J. V. Stafford (ed.) *Precision Agriculture '99*, pp. 19–33. Sheffield: Sheffield Academic Press.

4. Oliver, M. A. (2010) (ed) *Geostatistical Applications for Precision Agriculture*. Dordrecht: Springer.

5. Mercer, W. B. and Hall, A. D. (1911) The experimental error of field trials. *Journal of Agricultural Science*, **4**, 107–132.

6. Webster, R. and Oliver, M. A. (2007) *Geostatistics for Environmental Scientists*. Chichester: Wiley.

7. Oliver, M. A. and Carroll, Z. L. (2004) Description of spatial variation in soil to optimize cereal management. Project Report no. 330. Home-Grown Cereals Authority, London.

Dr Margaret Oliver is at the Soil Research Centre, Department of Geography and Environmental Science, University of Reading.