Precision Agriculture, Food Security and Geostatistics

Margaret A. Oliver

Soil Science Research,
The University of Reading, United Kingdom
Competition
Why the world needs PA

Competition for land, water and resources

PA can make better use of these

Soil and land are finite – provide food, fodder, fuel, housing, infrastructure, industry, recreation

Agriculture is largest consumer of water – Hedley & Yule (2009) showed 20% saving in water use with better irrigation management

Exhaustion of underground aquifers – conservation

Science and technologies of PA are crucial for greater food security
Why the world needs PA

Population to ~9.6 billion by 2050 (UN 2013) - 35%
Last 50 years - population tripled
    - cultivated area increased by 12%
Land per person decreased from ~0.45 — ~0.25 ha
Food production needs to increase 70% globally and 100% in developing countries (FAO)
1 person in 20 still likely to be under-nourished
    - 370 m people mainly in Africa and Asia
Increased agricultural production from more sustainable use of existing land

Sustainable intensification
Food security and malnutrition

“when everyone has access to safe, affordable and nutritious food all of the time and in ways the planet can sustain into the future” (Benton, 2012).

Agriculture is central in providing adequate, good quality, nutritious food that is traceable

Insufficient food leads to malnutrition – protein-energy malnutrition Kwashiorkor and Marasmus

About 1 billion (FAO, 2011) suffer from malnutrition
Under-nutrition (hidden hunger)

Food lacking in essential elements, particularly trace elements - affects about 2 billion worldwide.

I (~30% deficient), Fe (>60%), Se (unknown) and Zn (>30%) play major role in human nutrition and health.

Body cannot synthesize minerals required - must be obtained from food.

Selenium - added to fertilizers in Finland increased dietary intake 4-fold.

Other trace elements also added such as Zn and Fe - opportunity for PA.
Agricultural productivity

Wheat, rice and maize provide 2/3rds of energy for humans

Past 50 years - cereal production almost tripled in line with population growth

Green Revolution - increase in average yield of cereals from 1.35-3.35 t ha\(^{-1}\) between 1961 and 2007

Increase plateaued - 1990s
Green Revolution

Increased yields related to:

1) Improved germplasm
   e.g. dwarfing genes in wheat
2) Fertilizers, especially N
3) Herbicides and pesticides
4) Improved irrigation
Cost to the environment

Green Revolution - adverse effect on environment
- Excessive applications of N and P - pollution
- Accelerated soil erosion - loss of \( \sim 10 \text{M ha yr}^{-1} \)
  (Pimentel, 2006)
- Salinization - severely damages \( \sim 10 \text{M ha yr}^{-1} \)
  (Pimentel & Wilson, 2004)

Decline in crop nutrients, organic matter, CEC, structure and water-holding capacity

Increase in leaching, acidification, compaction, desertification etc.

Loss of yield
Malnutrition, Under-nutrition
Toxicity
Soil and land degradation

PA can help to prevent soil and land degradation (24% of land degraded) and maintain soil fertility and soil condition for agricultural sustainability.

Degradation decreases resilience of ecosystems and may be irreversible.
What can PA do to limit malnutrition, under-nutrition and degradation?

Increased use of sensor information to manage N, OM, soil moisture, etc. to increase yields

Variable-rate applications are feasible at any scale for both macro and trace elements – improve efficiency of fertilizer and pesticide use

This could benefit nutrient status of food and environment

BUT

Determining nutrient concentrations of soil, pH and texture are still stumbling blocks to PA
What can PA do to limit malnutrition, under-nutrition and degradation?

PA has made great strides from being linked only to technological advances.

Needs to be proactive in reducing soil erosion, salinization, desertification and soil infertility.

Use cropping systems appropriate to environmental conditions and integrated pest management.

Detailed and accurate spatial information is vital for precise management and land stewardship.

At present, use what we have available to overcome deficiencies in information—one way is to use developments in geostatistics.
Geostatistics and precision agriculture

Spatial variation is central to both precision agriculture and geostatistics
**Geostatistics and PA**

Geostatistics has been applied in PA since 1980s

The variogram and kriging are the central tools

**Variogram**

\[
\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} \{z(x_i) - z(x_i + h)\}^2
\]

**Kriging** - local weighted moving average for point or block estimates

\[
\hat{Z}(x_0) = \sum_{i=1}^{N} \lambda_i z(x_i) \quad \sum_{i=1}^{N} \lambda_i = 1.
\]

Spatially correlated

Spatially uncorrelated

Sill variance = \(c_0 + c\)

Spatially correlated component, \(c\)

Nugget variance, \(c_0\)
Sampling

Often too costly to obtain enough soil and crop data to estimate an accurate variogram

Matheron’s MoM variogram estimator requires about 100 data

BUT

Residual maximum likelihood (REML) estimator can be used with fewer data 50-<100
Guide to sampling intensity

Sampling interval must relate to spatial scale

Existing variograms from soil or crop properties
Or
Variograms of ancillary information such as ECₐ, elevation, aerial photographs or imagery

Use:
(1) ‘Rule of thumb’ – sample at less than half the variogram range
(2) Kriging errors or variances to identify an optimal sampling interval for a given variable for kriging
Half the variogram range of ancillary data (Kerry et al., 2010)

Variograms computed from digital numbers of two aerial photographs

Based on half range of aerial photograph variograms - sampling interval of 90-120-m (Wallingford field - 44 ha)

\[ c_0 = 1.861, \ c = 1.180, \ a' = 247.7 \text{ m} \]

\[ c_0 = 1.861, \ c = 1.180, \ a' = 247.7 \text{ m} \]

Variogram of B/W photograph:
\[ c_0 = 0, \ c = 538.3, \ a' = 204.6 \text{ m} \]
Variogram and optimal sampling

Determine the kriging errors for a range of grid intervals from existing variogram of LOI (Wallingford field)

Determined kriging errors for blocks of different size

Plot errors against grid spacing

Determine a tolerable error and where this meets the graph - it indicates the optimal sampling interval - 120 m in this case

Topsoil LOI Wallingford
Pentespherical model
\( c_0 = 0.0535, c = 0.4610, a = 226 \text{ m} \)
Sub-sampled data

Loss on ignition (LOI) data on the original 30-m grid in a 31-ha area of Wallingford field were sub-sampled to:

- 30-m, 296 sites
- 60-m, 70 sites
- 90-m, 36 sites
- 120-m, 23 sites
- 120+60-m, 50 sites

Sub-sampled data - computed variograms by the method of moments (MoM) and residual maximum likelihood (REML)
Variograms estimated by MoM and REML

- 296 sites — 30-m
- 70 sites — 60-m
- 36 sites — 90-m
- 23 sites — 120-m
- 50 sites — 60 and 120-m
Kriging LOI data with variograms estimated by MoM and REML

30-m MoM (296 sites)  60-m MoM (70 sites)  90-m MoM (70 sites)
120-m (23 sites)      120+60-m MoM (50 sites)  60-m REML (36 sites)
90-m REML (36 sites)  120-m REML (23 sites)  120+60-m REML (50 sites)
Kriging sparse data with the standardized variogram

Standardized variogram from ECₐ data with the sill scaled to 1
Marchant and Lark (2010) used Bayesian adaptive sampling to estimate topsoil water content with a prediction variance less than 21\%^2.

No variogram to outset - need prior distribution for variogram parameters based on what might be known.

Each sampling phase was optimized iteratively based on an objective function (MSE of an approximation to the kriging variance) and updated prior distribution of variogram parameters.
Optimal adaptive sampling

Study on soil water content using a theta probe - whole survey took <3 hours

Approach is advantageous for surveys with sensors.

1) 30 observations (x)
2) 30 more (x) - 60 sites
3) 10 more (x) - 70 sites
4) 5 more (x) - 75 sites
4 more to complete survey - 79 sites
Ancillary data to guide sampling to delineate management zones (Corwin and Lesch, 2010)

Used intensive $E_{Ca}$ data to direct soil sampling and then used soil information to delineate site-specific management units (SSMUs) for cotton.

**Factors Influencing Cotton Yield**

- **ECe**
  - Depth: 0-1.5 m
  - $E_{Ce}$ (dS/m)
    - 2 - 3
    - 3 - 4
    - 4 - 5
    - 5 - 8
    - 8 - 11
    - 11 - 15

- **LF**
  - Depth: 0-1.5 m
  - LF
    - 0 - 0.2
    - 0.2 - 0.3
    - 0.3 - 0.4
    - 0.4 - 0.5
    - 0.5 - 0.6
    - 0.6 - 0.8

- **H2O**
  - Depth: 0-1.5 m
  - H2O Content
    - 0.18 - 0.23
    - 0.23 - 0.27
    - 0.27 - 0.3
    - 0.3 - 0.32
    - 0.32 - 0.34
    - 0.34 - 0.37

- **pH**
  - Depth: 0-1.5 m
  - pH
    - 7.2 - 7.5
    - 7.5 - 7.6
    - 7.6 - 7.7
    - 7.7 - 7.8
    - 7.8 - 7.9
    - 7.9 - 8.2

**Figure 6.6**

- ECe directed soil sample sites
- Reduce ECe to <7.17 dS m⁻¹
- Reduce leaching fraction (LF) to <0.4
- Reduce pH to <7.9
- Coarse textured soil requires more irrigation

Kriged maps of properties affecting cotton yield.
Goovaerts and Kerry (2010) incorporated secondary information with cokriging (CK), kriging with external drift (KED) and simple kriging with local means (SKlm). CK requires a cross-variogram, and KED and SKlm require secondary data at all prediction points.

Cross-validation results show CK performed best.

<table>
<thead>
<tr>
<th>Method</th>
<th>Ancillary variables</th>
<th>Mean absolute error (MAE)</th>
<th>Mean squared deviation ratio (MSDR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OK</td>
<td>---</td>
<td>2.930</td>
<td>0.996</td>
</tr>
<tr>
<td>SKlm-h</td>
<td>Aerial$^91$, EC$_a$, elevation</td>
<td>2.805</td>
<td>1.078</td>
</tr>
<tr>
<td>SKlm-s</td>
<td>Soil series</td>
<td>2.824</td>
<td>0.865</td>
</tr>
<tr>
<td>KED</td>
<td>Aerial$^91$ and EC$_a$</td>
<td>2.777</td>
<td>0.875</td>
</tr>
<tr>
<td>CK</td>
<td>Aerial$^91$ and EC$_a$</td>
<td><strong>2.674</strong></td>
<td><strong>1.046</strong></td>
</tr>
</tbody>
</table>
Cross-variograms for cokriging

Models fitted using the linear model of coregionalization with a spherical structure, nugget component and range of 85 m.
Ancillary data to improve prediction of sand content (Yattendon Estate 15.3 ha)
Pringle et al. (2010) used geostatistics to analyse local-response experiments for site-specific management (SSM).

Examined fine-scale spatial variation of crop response to variable inputs.

Used standardized cokriging (takes account of means for different treatments) to map yield responses to each treatment.

Zero treatment had the smallest yields, but the other 3 levels of N show similar responses.
Local-response experiments

<table>
<thead>
<tr>
<th>Treatment (kg N ha⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td># 0</td>
</tr>
<tr>
<td>&quot; 100</td>
</tr>
<tr>
<td>D 170</td>
</tr>
<tr>
<td>l 225</td>
</tr>
</tbody>
</table>

8.9 ha

Yield (Mg ha⁻¹)
< 3  
3 – 4  
4 – 5  
5 – 6  
6 – 7  
> 7

(a) 0 kg N ha⁻¹  
(b) 100 kg N ha⁻¹  
(c) 170 kg N ha⁻¹  
(d) 225 kg N ha⁻¹
Space-time (ST) geostatistics

ST geostatistics characterizes variation in both space and time.

Assess change over time to assist management, but not to predict over time.

Heuvelink and van Egmond (2010) applied ST geostatistics to a potato crop.

Aim to fine-tune fertilizer application, and timing and amount of chemicals to remove above-ground biomass.

Also for timely management decisions on irrigation and pesticide applications.
Space and time: NDVI of potato

Large spread of values indicates considerable spatial variation.

Red is Innovator (I)  
Blue is Sophista (S)

Day 149, 29 May  
Day 208, 27 July

Late development
**Space-time (ST) geostatistics**

Variogram computed in space and time and both sources of variation must be modelled

\[ \gamma(h,u) = \gamma_S(h) + \gamma_T(u) + \gamma_{ST}\left(\sqrt{h^2 + (\alpha,u)^2}\right) \]

**Sum-metric model** simplifies ST modelling

\( \gamma_{ST} \) is joint ST structure and \( \alpha \) is the geometric anisotropy ratio to match distance in time with that in space

\[ \alpha = 20 \text{ m per day} \]

All points on ellipse have same semivariance
Day 165, 15 June           Day 200, 19 July                 Day 235, 23 August

Spatial range - 100 m

Temporal range - 8 days

ST range 120 m

α 6 m per day

Spatial range - 100 m

ST variogram then used for prediction by ST kriging

Day 165, 15 June          Day 200, 19 July          Day 235, 23 August
Gebbers (2010) showed recently how to apply geostatistical simulation (GS) in PA.

GS incorporates uncertainty into modelling to provide a more realistic impression of variation and potential errors in maps.

Simulation can be unconditional or conditional, which incorporates existing knowledge.
Geostatistical simulation vs kriging

Kriging provides the ‘best’ local estimates, but it smooths: over-estimates smaller values and under-estimates larger ones and loses variance.

Simulation can provide many realizations that will all have the same statistical characteristics.

Simulation retains the original variance in the data and texture of the variation, but these are at the expense of local accuracy.

![Image showing comparison of kriging and simulation transects](attachment:image.png)
Comparison between kriging and conditional simulation of pH

Simulation mean is based on 100 realizations
Probabilities

Probabilities that a value is outside its optimal range can be determined from the cumulative probability distributions from several simulated realizations.

In this example the optimum range is 5.8-6.3.
Where next?

Promote the value of PA in developing countries where fertilizers are scarce and costly

Use available technology and scientific expertise in PA to tackle increasing land and soil degradation

Challenge is to move from diagnostic to prescriptive phase of PA (Hatfield and Kitchen, 2013)

Increase knowledge that can be elucidated from sensor technology (remote and proximal) – relatively cheap to obtain and intensive

Scientific advances in crop physiology

Apply selected trace elements in fertilizer management
Where next?

Geostatistics needs to develop to reflect the changes in data – more sensed data

Automated adaptive sampling based on geostatistics

Spatio-temporal geostatistics needs to be developed and simplified for practitioners

Greater use of soft data and of Bayesian geostatistics

Use of geostatistical simulation to identify uncertainties in mapping
Acknowledgements

I thank contributors to

*Geostatistical Applications for Precision Agriculture*
(M. A. Oliver (Ed.), Springer 2010)

for allowing me to use their examples