Making Sense of Sensed Data Using Geostatistics

Margaret A. Oliver

Department of Soil Science
The University of Reading
United Kingdom

READING
SOIL SCIENCE

Acknowledgements

I thank Ruth Kerry for the use of her data, and the USA Army for the use of their SPOT and DEM data.

Overview

- What do we mean by sensed data?
- Why geostatistics might be of use with such data?
- Nested variation and how it can be investigated.
- Case studies:
  - » using three types of sensed data.
  - » using satellite imagery
  - » taking account of trend in data
- Conclusions

Sensed data

- A major problem for farmers in the precision agriculture context is to obtain enough data about soil and crop properties to show the variation accurately for management
- Methods of sensing have the advantage of producing large amounts of spatially referenced data relatively cheaply and quickly
- Some sensors can be linked with GPS and farm equipment e.g. tractor mounted or on combines
- In many cases the process is non-destructive and non-invasive which avoids damage to the soil and crop roots.

Sensed data

Sensed data can measure soil and crop properties - such information is likely to change agricultural management in the context of site specific farming

Examples:

- radiometers to detect weeds
- yield monitoring
- satellite imagery, hyperspectral data, aerial photographs, microwave radiation
- chlorophyll sensor - measures short wave radiation
- soil conductivity - by electromagenetic induction or direct contact

Sensed data - difficulties

- Sensors generally produce large amounts of data that can be both difficult to process and interpret.
- Interference during recording
- There are differences in the resolution of different sources of sensed data.
  - » what is a suitable resolution for precision agriculture, for example?
  - » does more data necessarily mean better information?

Sensed data - difficulties (cont.)

- Too much detail and also noise from measurement errors can obscure the structures of interest in the variation.
- Often more than one scale of variation is present (nested variation).
- Can be difficult to relate sensor information to ground information, such as soil conditions and vegetation types.

How can geostatistics help?

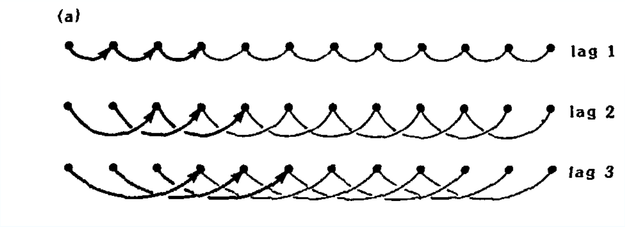
- Geostatistics provides tools to explore the variation of both sparse and large sets of data.
- The variogram can be used to detect the presence of nested variation.
- Nested variogram models can decompose the variation to the spatial scales of variation present.

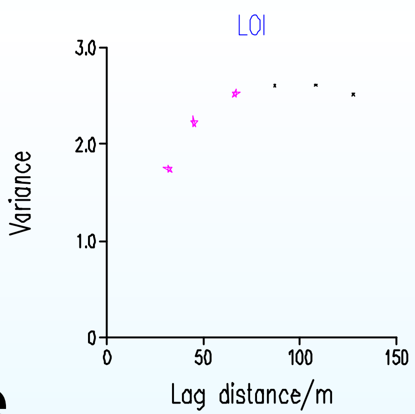
How can geostatistics help? (cont.)

- Ordinary kriging can smooth the variation so that the main structures in the variation can be observed.
- Kriging analysis or factorial kriging can filter the components of variation of interest.
- An important aim is to understand the factors that are controlling the variation

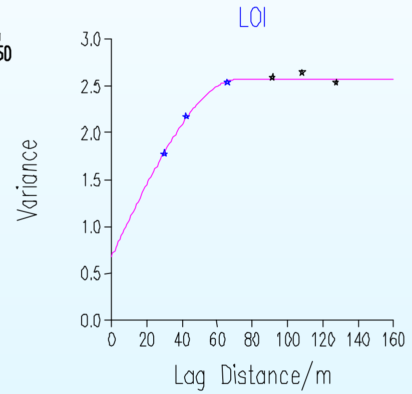
The variogram

- Describes how a property varies with distance and direction
- Computed by: $$\hat{\gamma}(\mathbf{h}) = \frac{1}{2M(\mathbf{h})} \sum_{i=1}^{M(h)} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2$$





- Models can be fitted to the experimental semi-variances



Kriging: geostatistical estimation

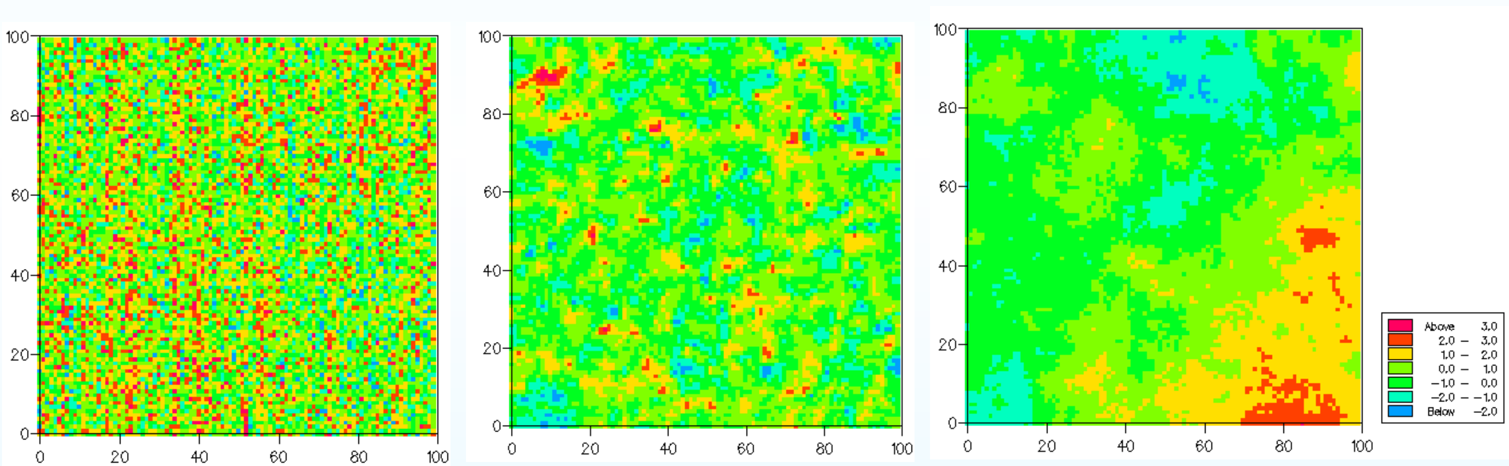
- A *local weighted moving average* calculated for points or blocks
- Weights depend on the structure of spatial variation and configuration of sampling points
- The weights are derived from the variogram
- *Kriging* differs from other interpolators - it uses a model of the spatial variation

Nested Variation

- Variation in many environmental properties arises from processes that operate and interact at different spatial scales
  - » climate, geology, relief, hydrology, trees, earthworms, microbiota and so on
- Each factor might result in several scales of variation
- Structure at one scale is ‘noise’ at another

Nested Variation

- A random process can be several independent processes nested within one another.
- They act at different spatial scales.



- The variogram of $Z(\mathbf{x})$ is then a nested combination of two or more individual variograms.

Nested variogram: linear model of regionalization

- The nested variogram comprises more than one variogram structure.
- Each structure might represent a separate process.
- The individual variograms that comprise it are additive.
- They are uncorrelated with each other and are independent orthogonal functions.

Nested variogram: linear model of regionalization

- Assume that the variogram of $Z(\mathbf{x})$ is a nested combination of S individual variograms:

$$\gamma(\mathbf{h}) = \gamma^{1}(\mathbf{h}) + \gamma^{2}(\mathbf{h}) + \Lambda + \gamma^{s}(\mathbf{h})$$

- Assuming that the processes are uncorrelated, the linear model of regionalization for S basic variograms is:

$$\gamma(\mathbf{h}) = \sum_{k=1}^{S} b^{k} g^{k}(\mathbf{h}),$$

each process has its own variogram:

$$b^{k} g^{k}(\mathbf{h})$$

Kriging Analysis or Factorial Kriging

- The aim is to separate out the components of variation of interest and to krige them
- Devised by Matheron (1982) to estimate the variogram components separately
- This is equivalent to filtering each component from the others
- If the variation is nested it can be explored further by factorial kriging

Kriging Analysis or Factorial Kriging (cont.)

- Kriging analysis is based on the concept that $Z(\mathbf{x})$ can be decomposed into two or more independent processes
- For a property with three spatial components including the nugget, the relation becomes

$$Z(\mathbf{x}) = Z^1(\mathbf{x}) + Z^2(\mathbf{x}) + Z^3(\mathbf{x}) + \mu.$$

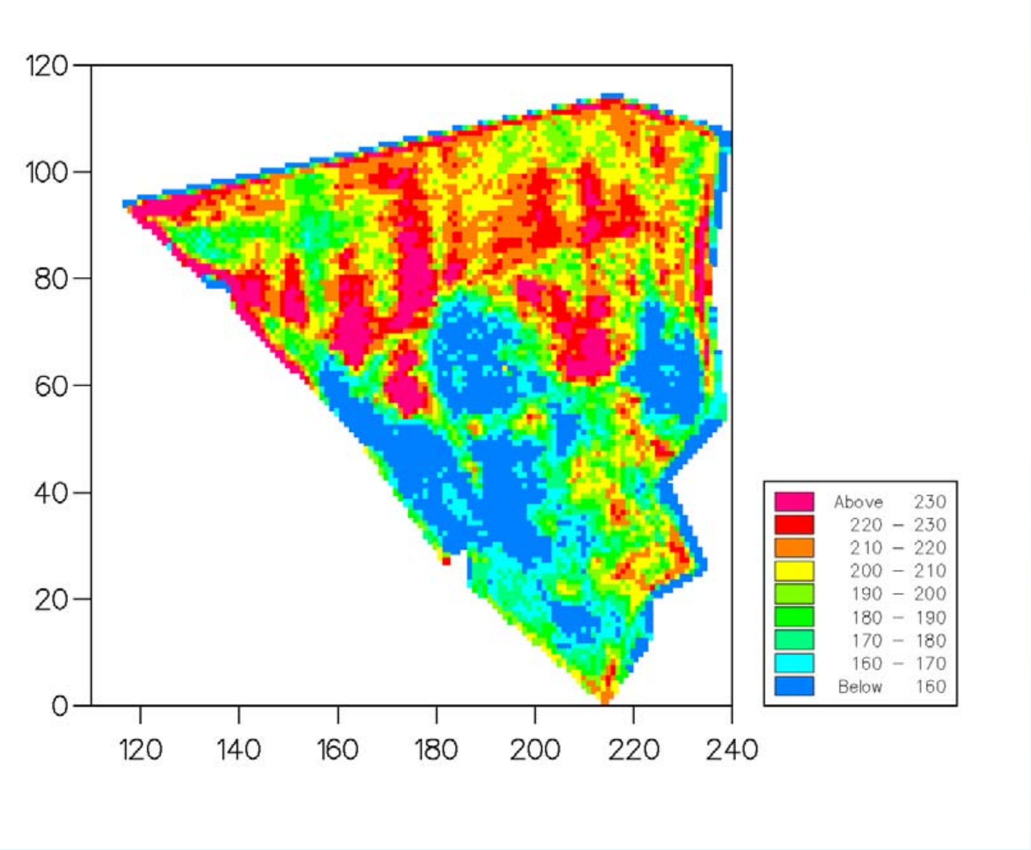
- Each component of the variation is treated as signal in turn.
- Noise at one scale of variation is regarded as information at another.

Case study: Yattendon Estate, Berkshire

This study describes an analysis of three kinds of sensed data:

- Yield data
- Digital information from aerial photographs
- EMI data

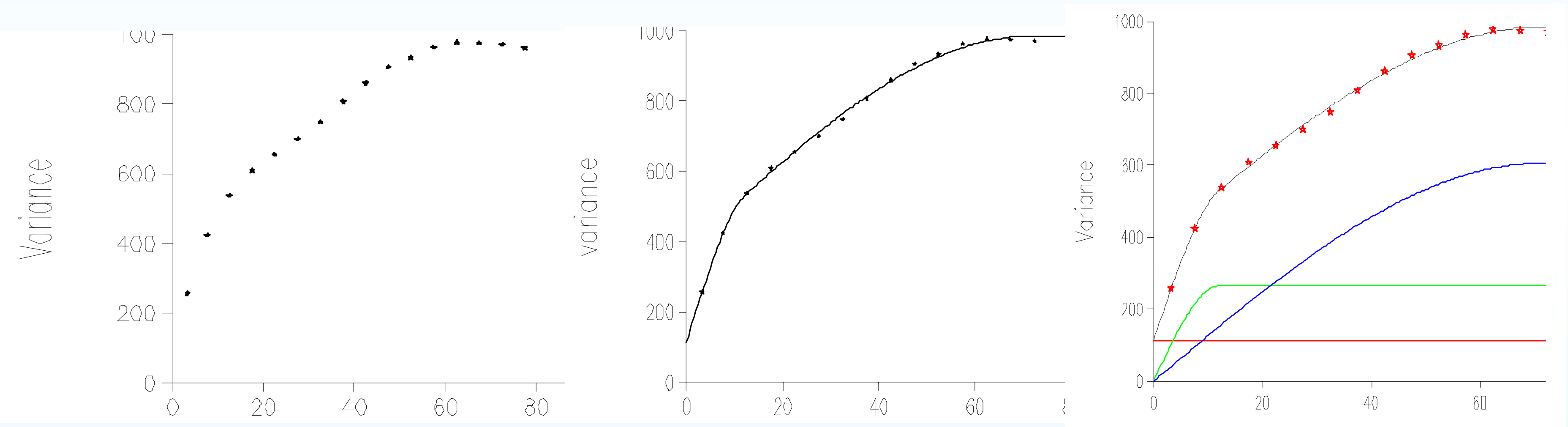
Aerial photograph image for Yattendon 1986



READING
SOIL SCIENCE

Experimental variogram and model for green waveband1986

a) Experimental variogram b) Fitted nested model c) Decomposed variogram



Lag distance/3.4m Lag distance/3.4m Lag distance/3.4m

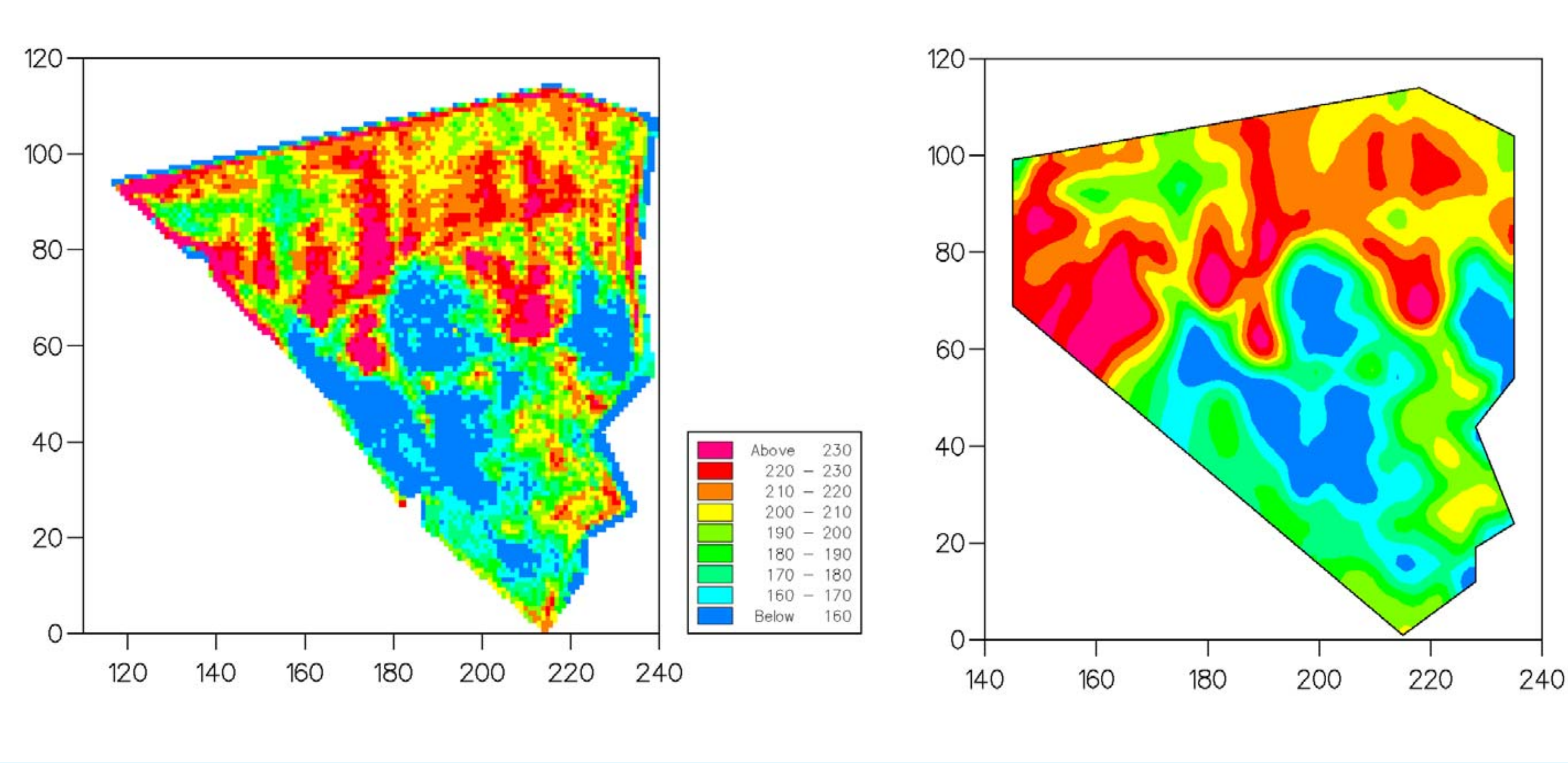
$$\gamma(\mathbf{h}) = 112.9 + 265.5\left\{\frac{3h}{41.51} - \frac{1}{2}\left(\frac{h}{41.51}\right)^3\right\} + 605.5\left\{\frac{3h}{242.2} - \frac{1}{2}\left(\frac{h}{242.2}\right)^3\right\}$$

READING
SOIL SCIENCE

Aerial photograph image for Yattendon 1986

a) Raw data

b) Ordinary kriged predictions

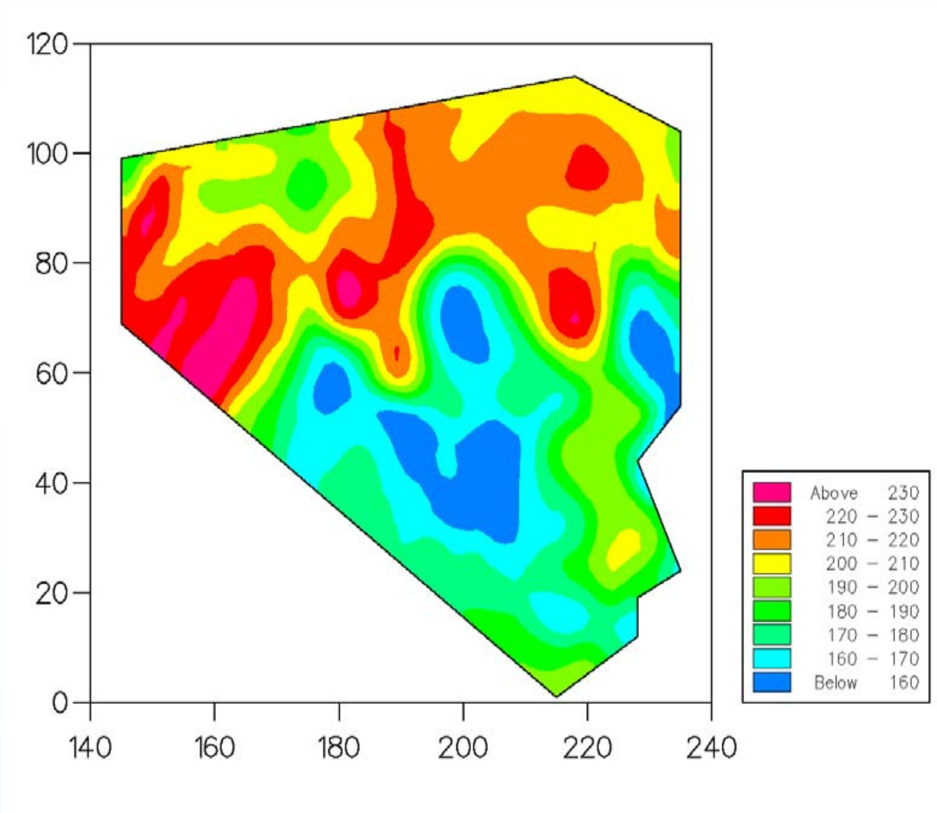


READING
SOIL SCIENCE

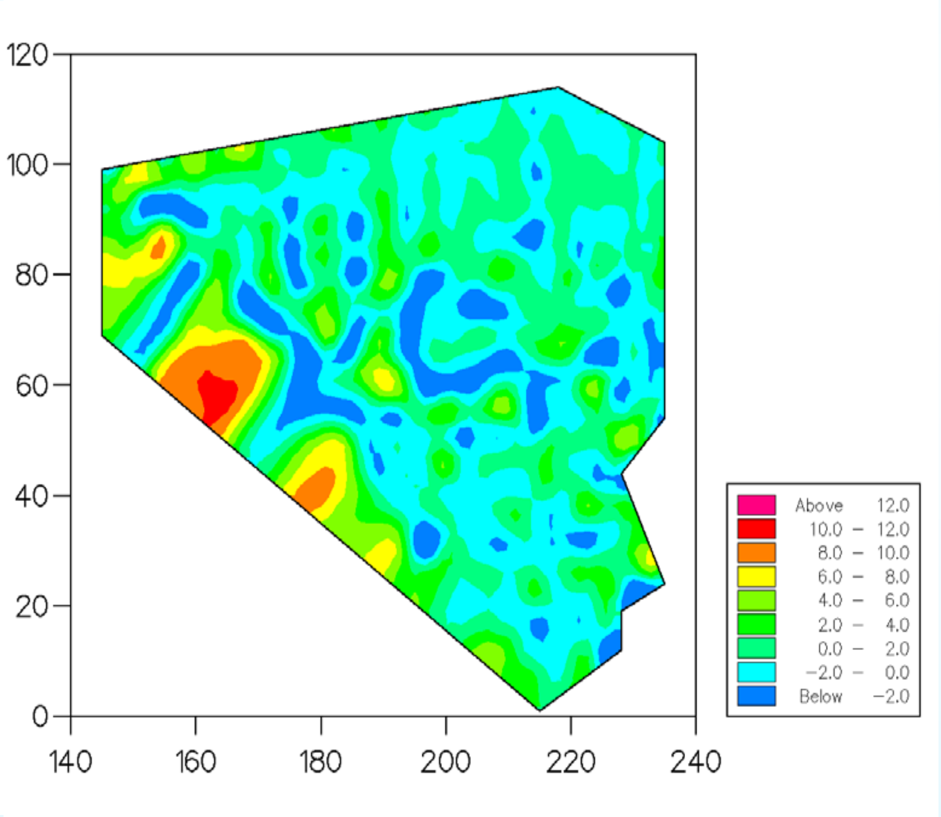
Aerial photograph image for Yattendon 1986

Results of factorial kriging:

a) Long-range estimates

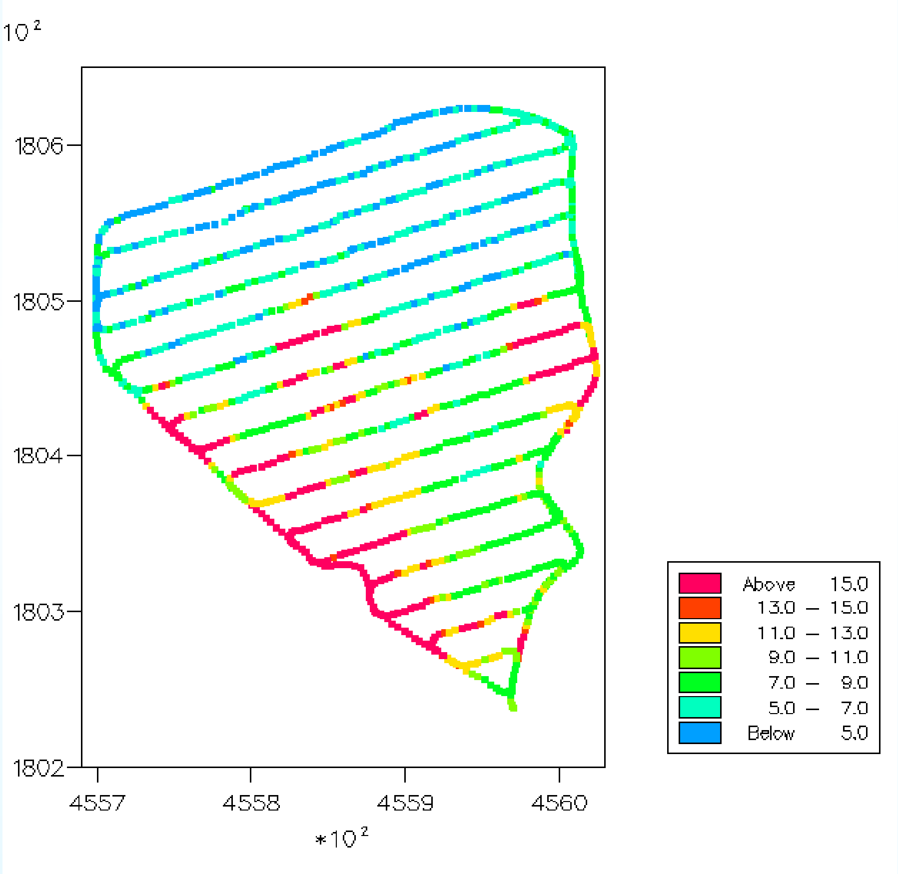


b) Short-range estimates



READING
SOIL SCIENCE

The EMI data



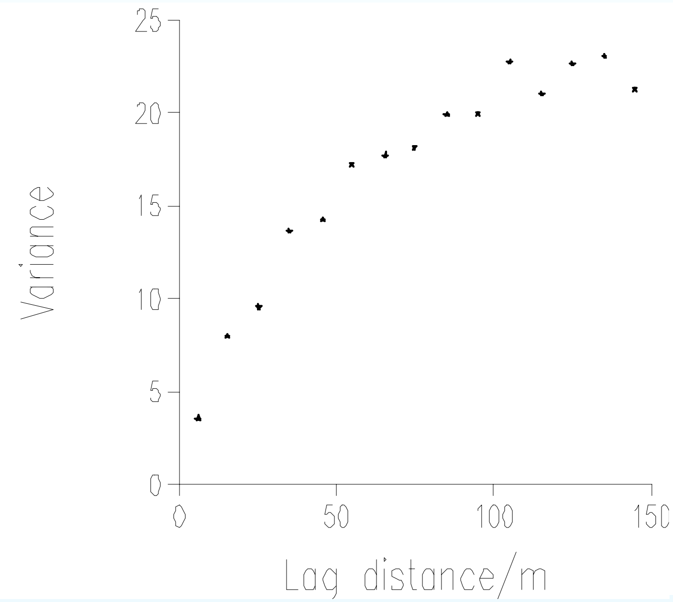
The EMI data contained long-range trend.

This was removed by a linear function.

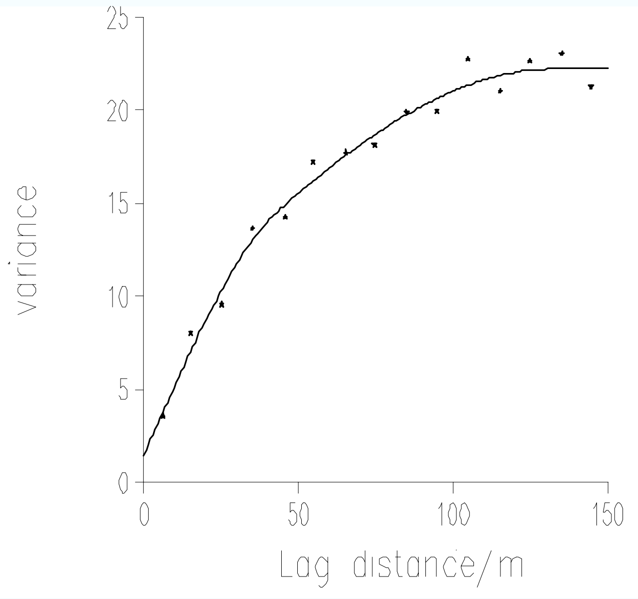
The remaining analyses were done on the residuals from this trend

Experimental variogram and model for residuals from the EMI data

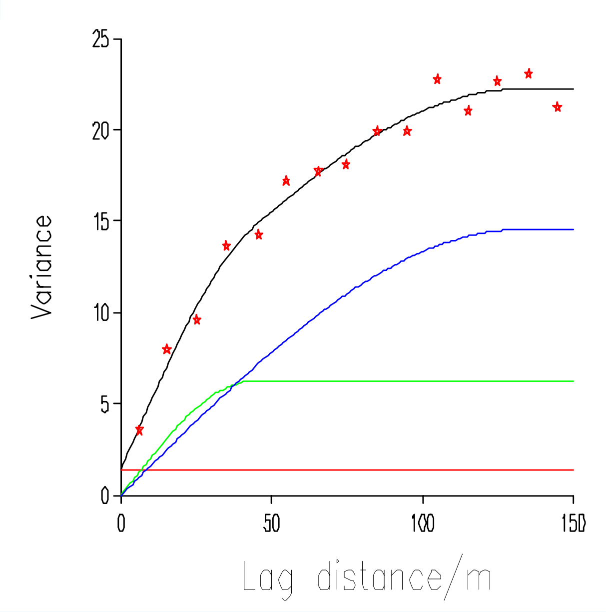
a) Experimental variogram



b) Fitted nested model



c) Decomposed variogram

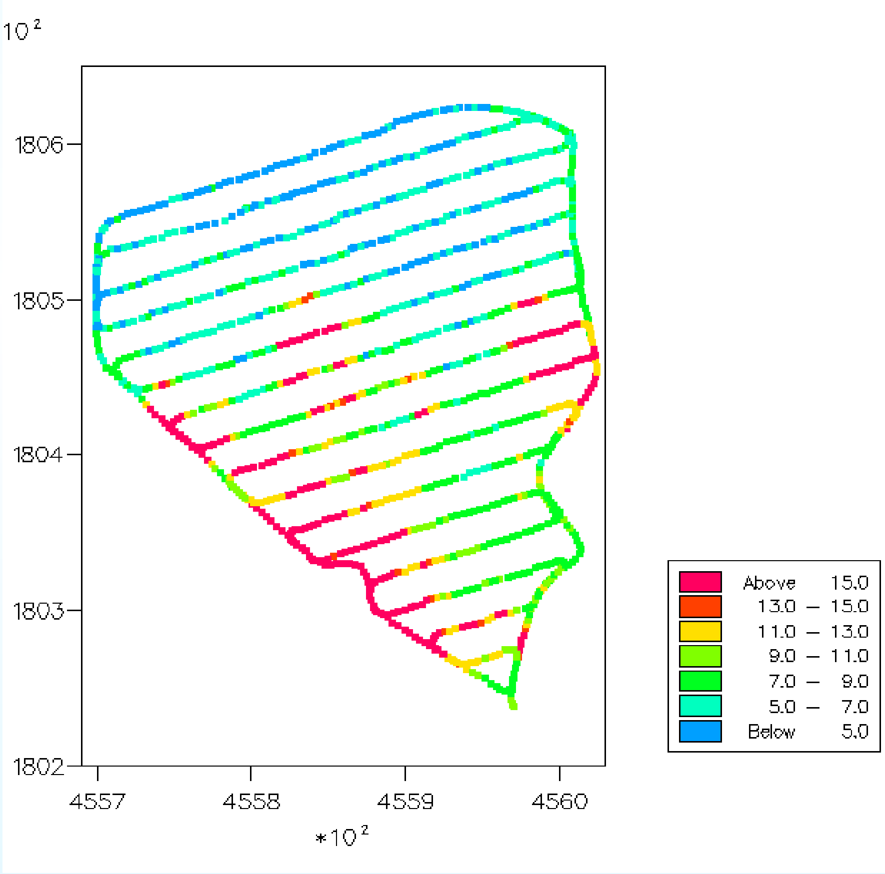


$$\gamma(\mathbf{h}) = 1.4 + 6.24\left\{\frac{3h}{43.57} - \frac{1}{2}\left(\frac{h}{43.57}\right)^3\right\} + 14.53\left\{\frac{3h}{132.3} - \frac{1}{2}\left(\frac{h}{132.3}\right)^3\right\}$$

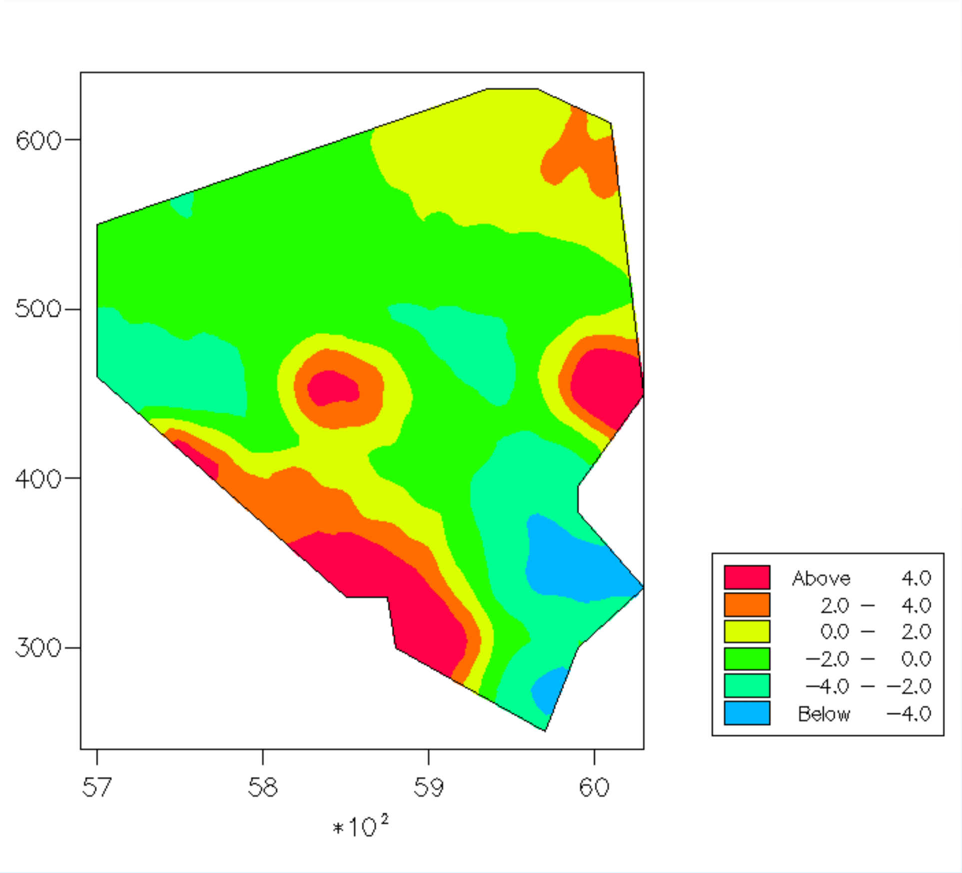
READING
SOIL SCIENCE

Electromagnetic Induction (EMI) data for Yattendon 2000

a) Raw data



b) Ordinary kriged predictions of the residuals

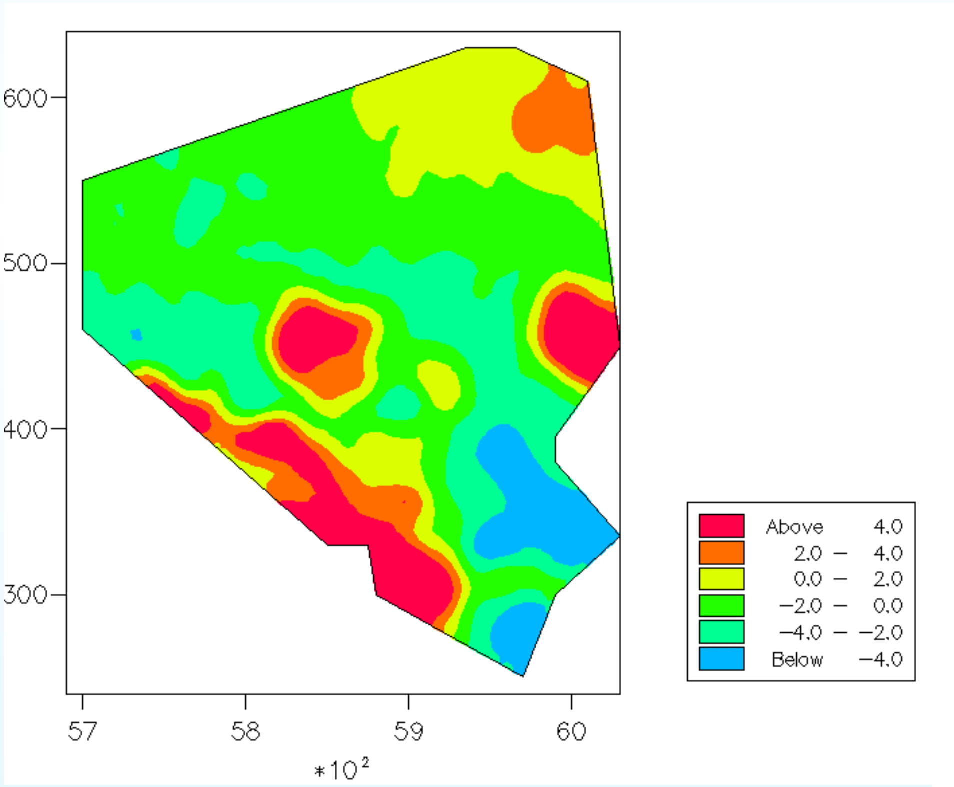


READING SOIL SCIENCE

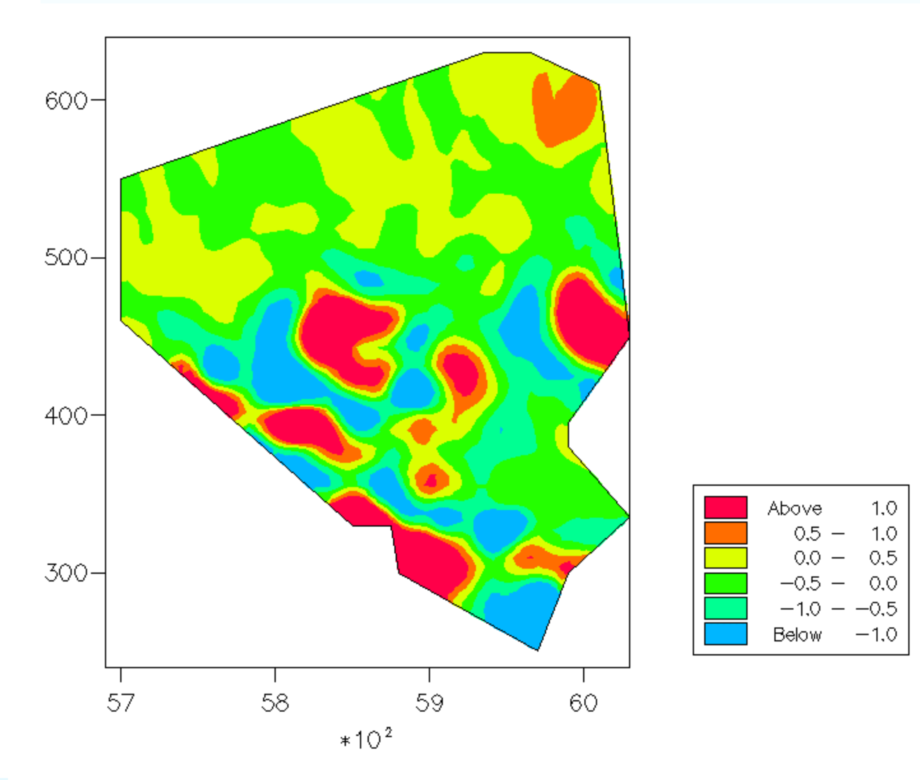
Factorial kriging of (EMI) data for Yattendon 2000

Results of factorial kriging:

c) Long-range component



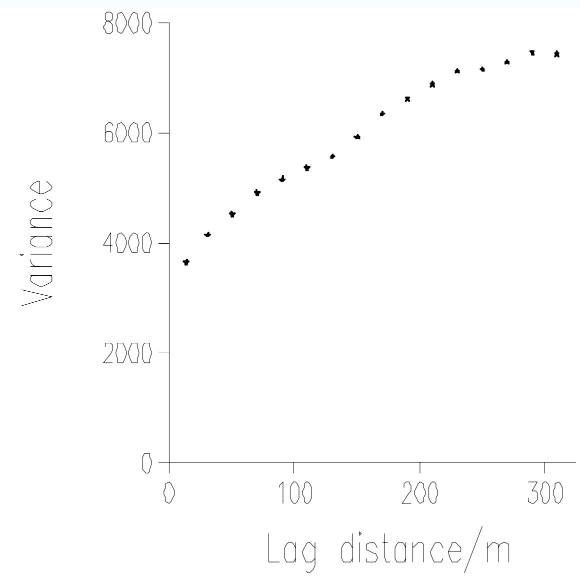
d) Short-range component



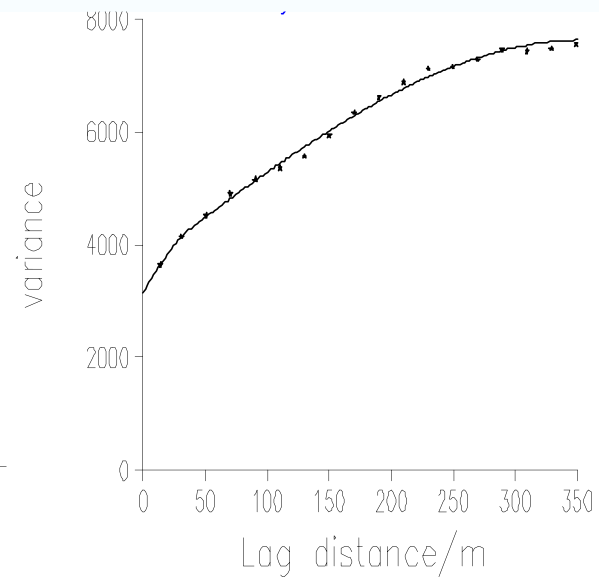
READING
SOIL SCIENCE

Experimental variogram and model for yield 1997

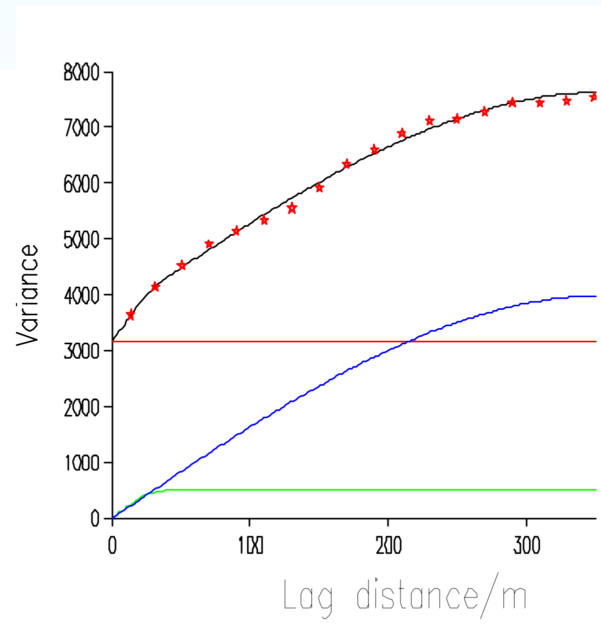
a) Experimental variogram



b) Fitted nested model



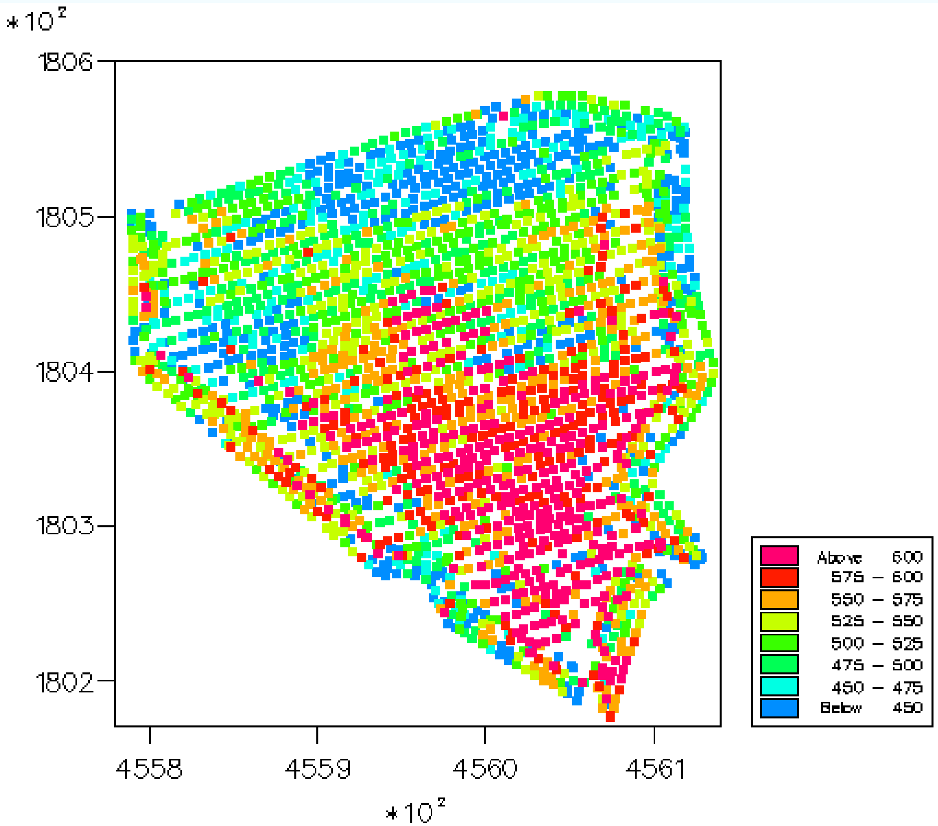
c) Decomposed variogram



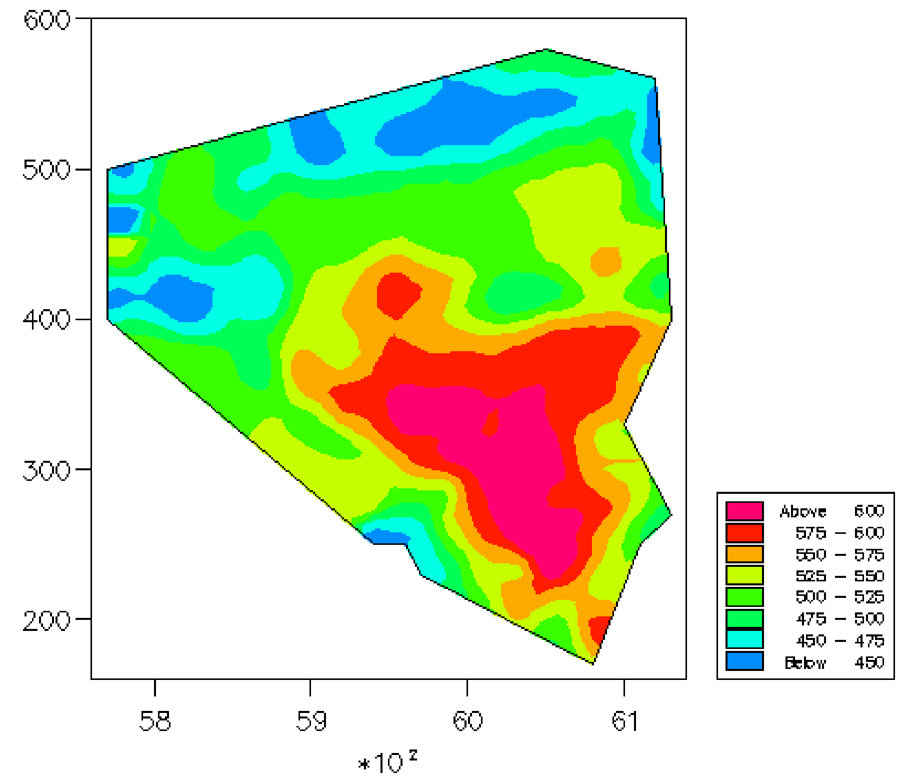
$$\gamma(\mathbf{h}) = 31550 + 491\left\{\frac{3h}{37.19} - \frac{1}{2}\left(\frac{h}{37.19}\right)^3\right\} + 3988\left\{\frac{3h}{355.5} - \frac{1}{2}\left(\frac{h}{355.5}\right)^3\right\}$$

Yield for Yattendon

a) Raw data



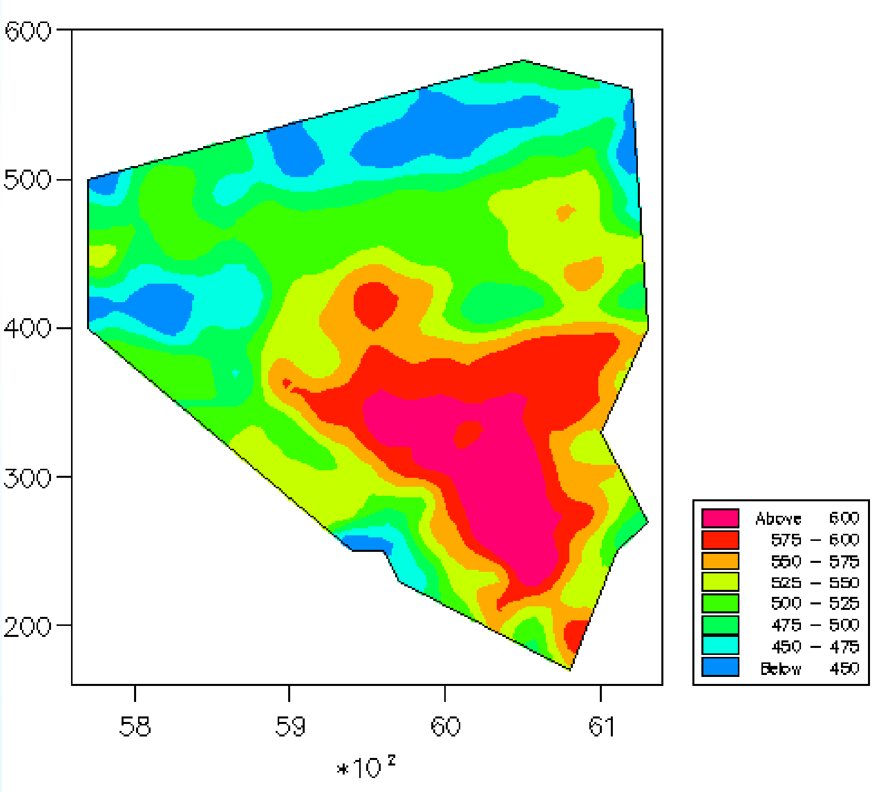
b) Ordinary kriged predictions of the residuals



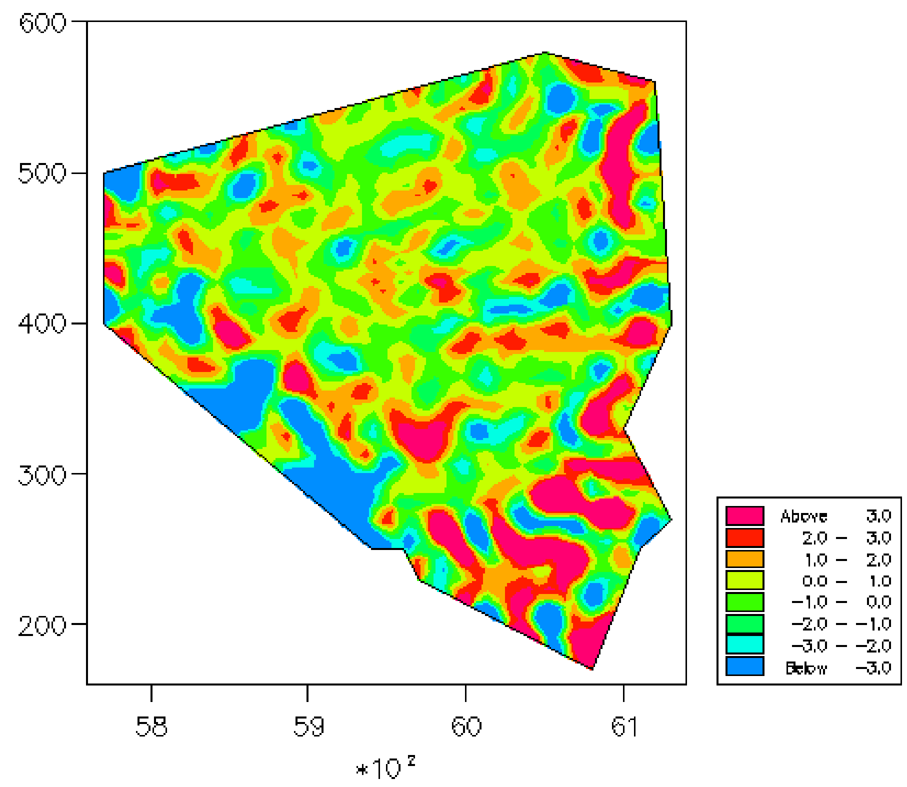
Factorial kriging of Yield for Yattendon

Results of factorial kriging:

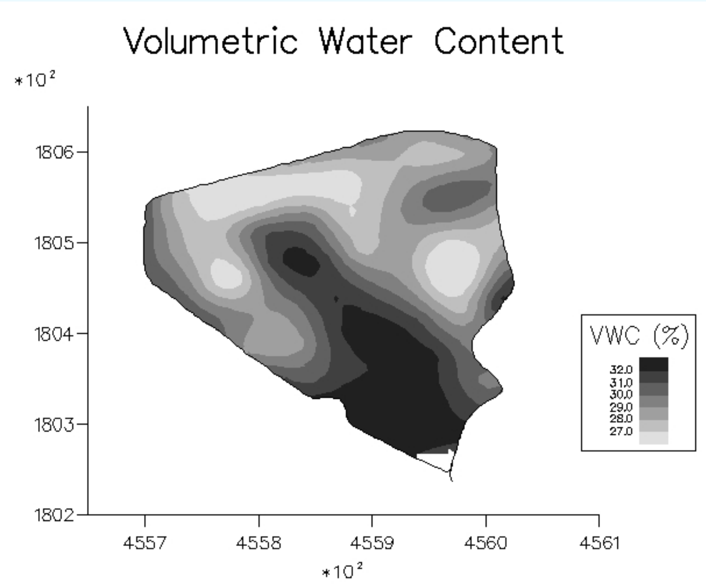
c) Long-range component

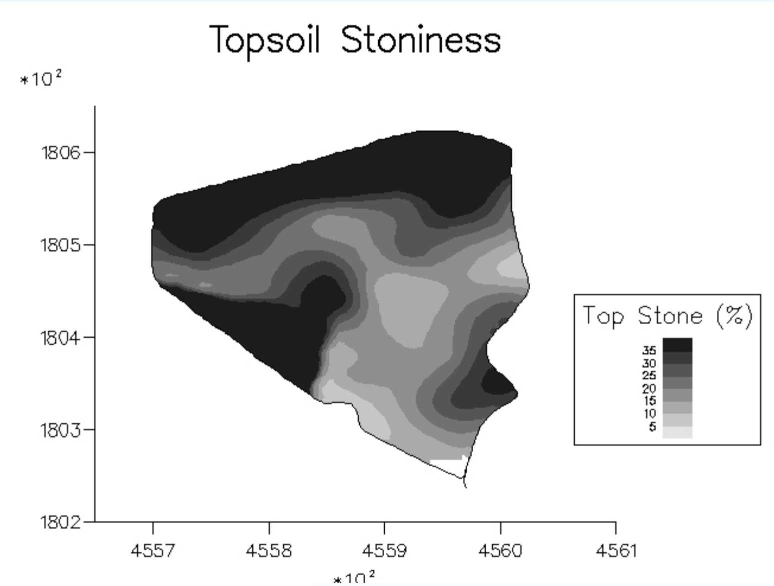


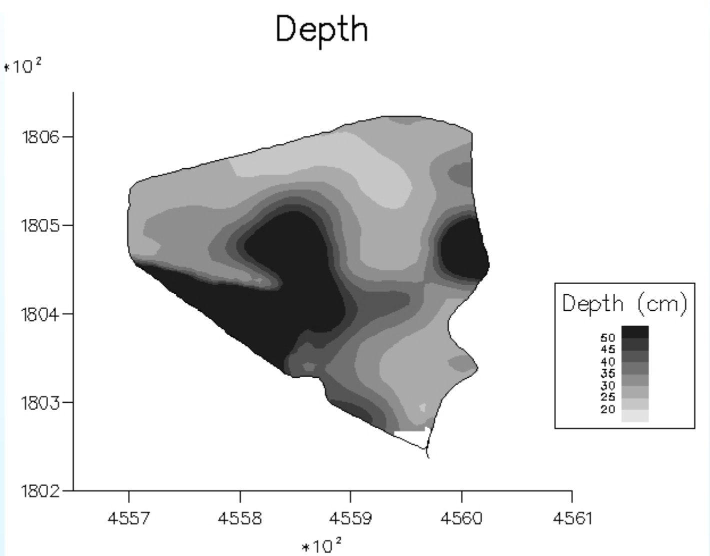
d) Short-range component



Preliminary results for some soil properties







READING
SOIL SCIENCE

Summary

The three kinds of ancillary data show similar nested patterns of variation.

Relations with volumetric water content, topsoil stoniness and loss on ignition are visibly strong.

Suggests that variograms of ancillary data could be used to guide sampling of the soil.

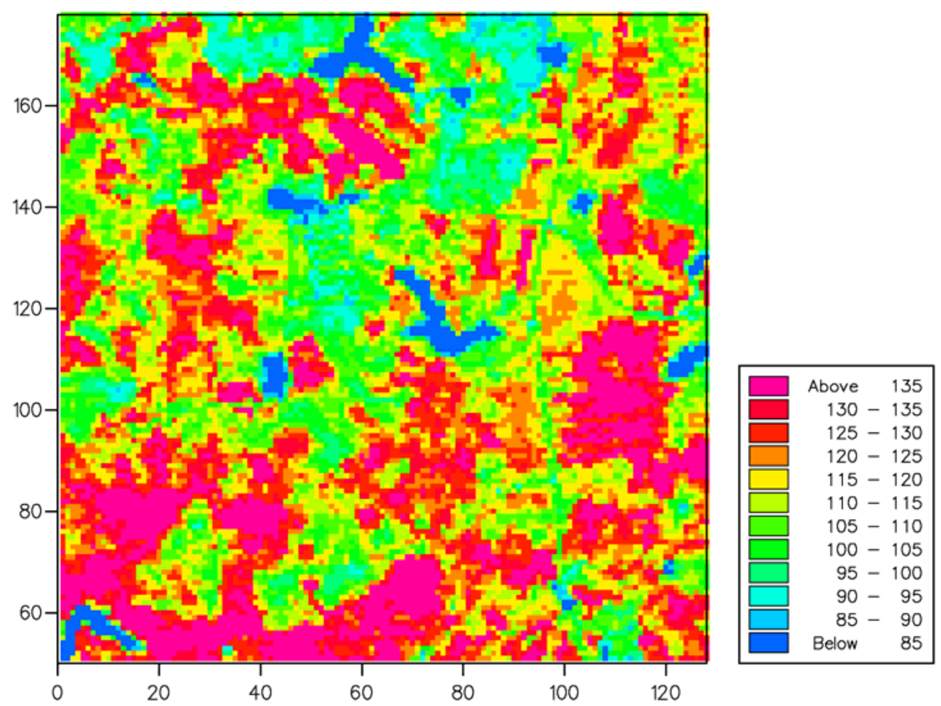
Other soil properties are being analysed at present.

Case Study: SPOT image of Fort A. P. Hill

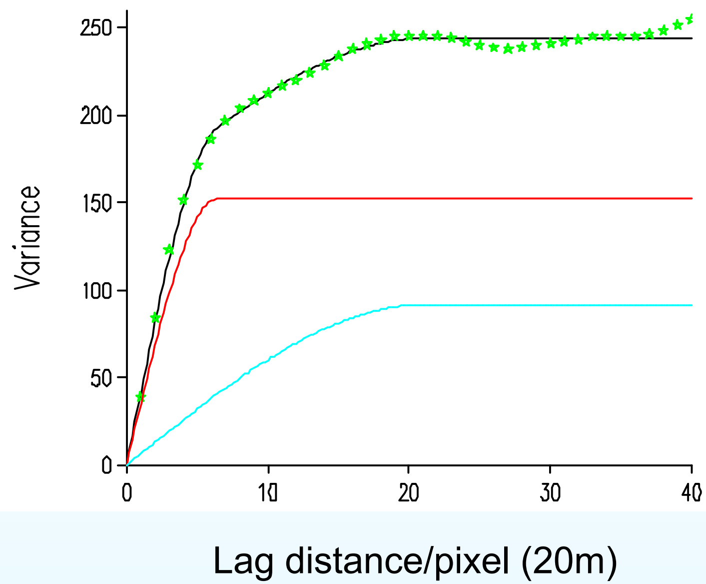
- The part of the scene analysed is of Fort A. P. Hill in Virginia, USA
- 128 by 128 pixels - 16384 in total
- Analysed the near infrared range of the electromagnetic spectrum (NIR)
- Multiresolution analysis has relevance for further sampling and for selecting the level of variation to be retained with data compression.

Near infra red (NIR) for Fort A. P. Hill, USA.

a) Raw data



b) Variogram



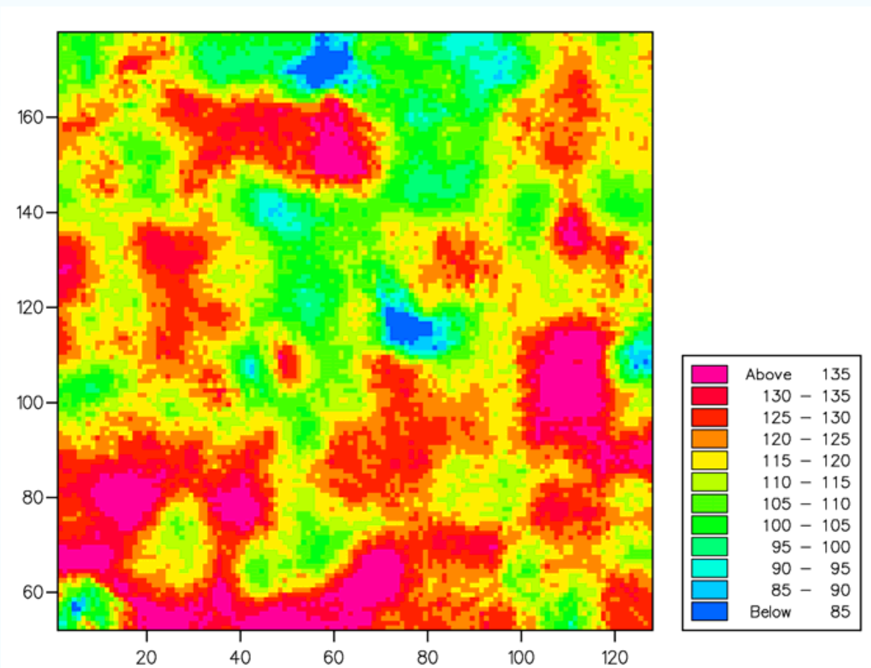
$$\gamma(\mathbf{h}) = 0 + 152\left\{\frac{3h}{129} - \frac{1}{2}\left(\frac{h}{129}\right)^3\right\} + 91.7\left\{\frac{3h}{422} - \frac{1}{2}\left(\frac{h}{422}\right)^3\right\}$$

READING
SOIL SCIENCE

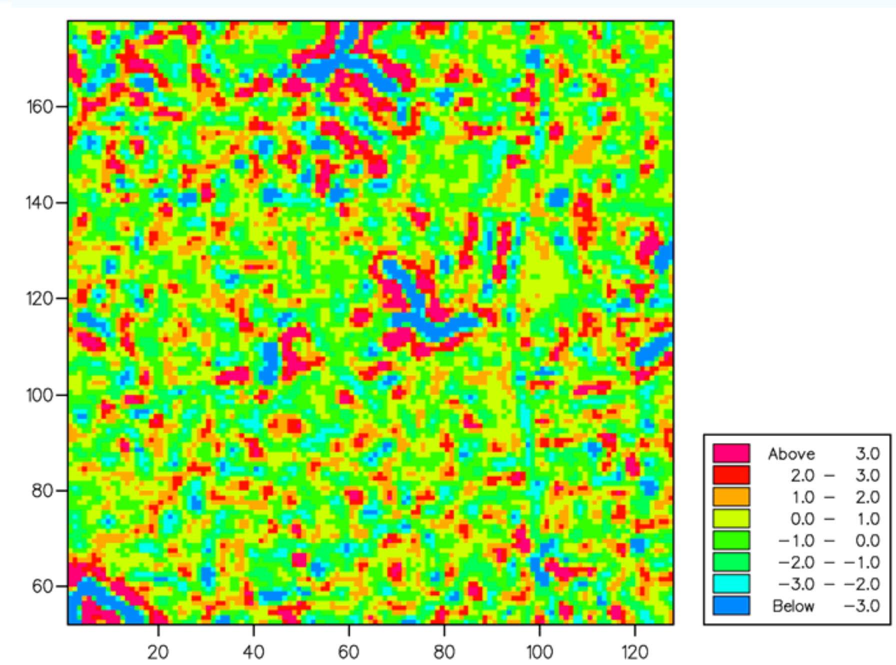
Near infra red (NIR) for Fort A. P. Hill, USA.

Results of factorial kriging:

c) Long-range component



b) Short-range component

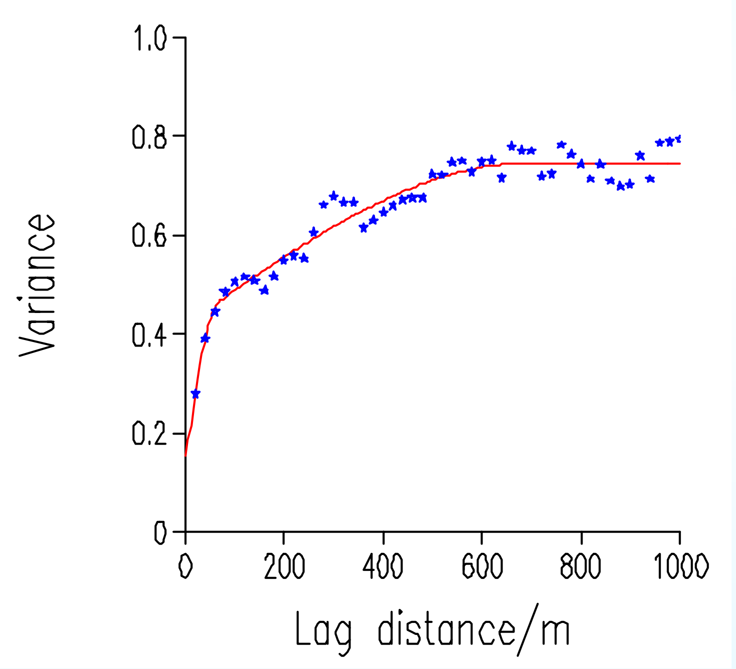


Factorial kriging filtered out effectively the two main scales of spatial variation

READING
SOIL SCIENCE

Ground cover survey of Fort A. P. Hill, USA.

b) Multivariate variogram of ground cover classes

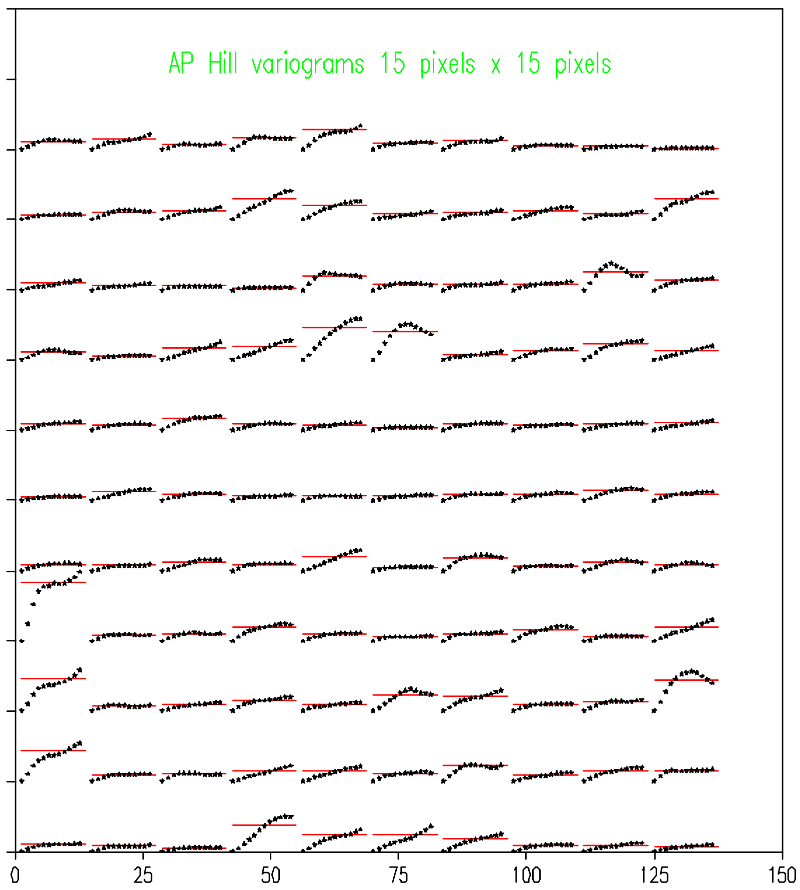


The variograms of the wavebands and NDVI were used to design several surveys of ground cover.

The multivariate variogram computed from seven classes of cover shows a similar form to the variogram of NIR.

READING SOIL SCIENCE

NIR Tiled variograms for Fort A. P. Hill, USA



READING
SOIL SCIENCE

Case Study: DEM of Fort A. P. Hill

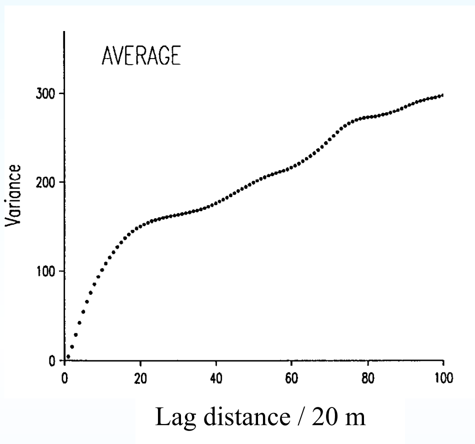
This study examines data that contain trend which violates the assumptions of the random function model that underpins geostatistics.

The data were on a 5 m grid - this was sub-sampled to a 20 m grid to match the SPOT pixel size.

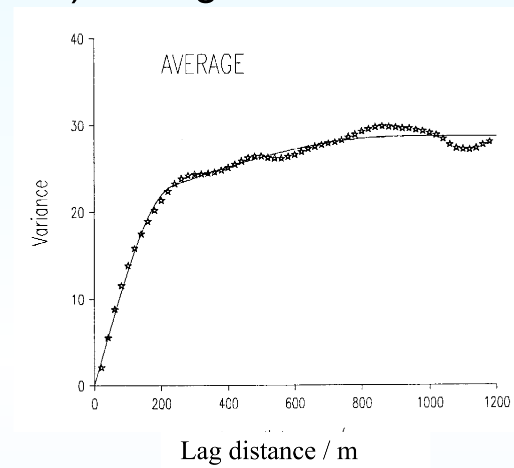
Linear, quadratic and cubic functions were fitted to the coordinates of the data.

Digital elevation data for Fort A. P. Hill, USA: variograms from 20 m grid.

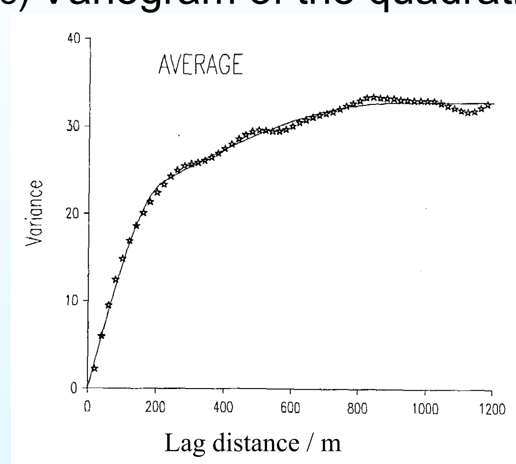
a)Variogram of the raw data



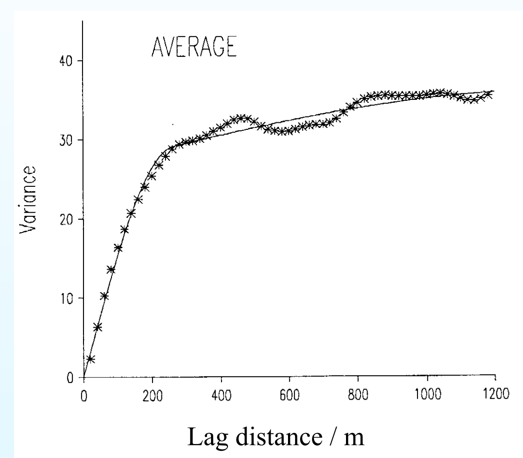
b) Variogram of the linear residuals



c) Variogram of the quadratic residuals

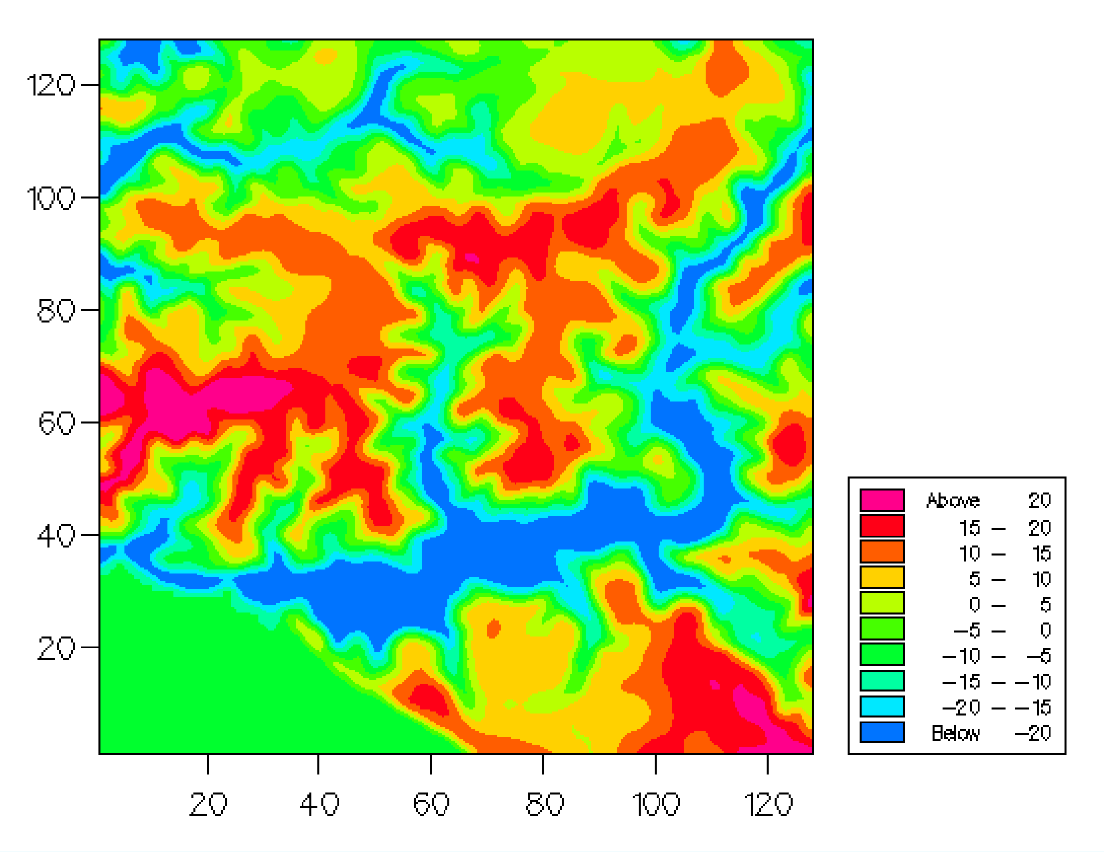


d) Variogram of the cubic residuals



Digital elevation data for Fort A. P. Hill, USA.

a) Ordinary kriged estimates of the quadratic residuals

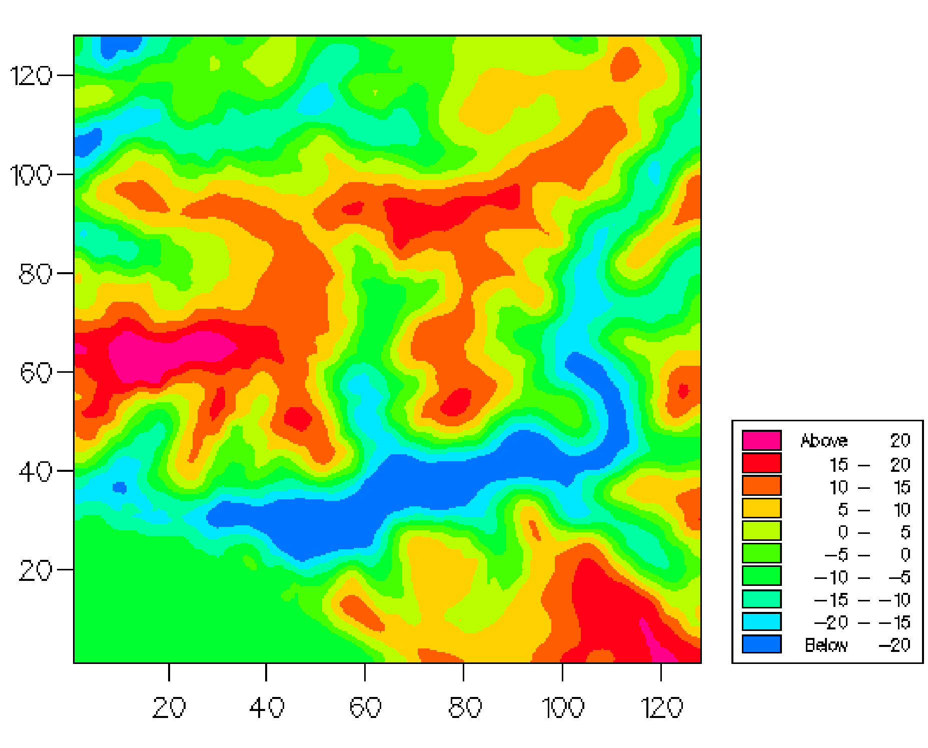


READING SOIL SCIENCE

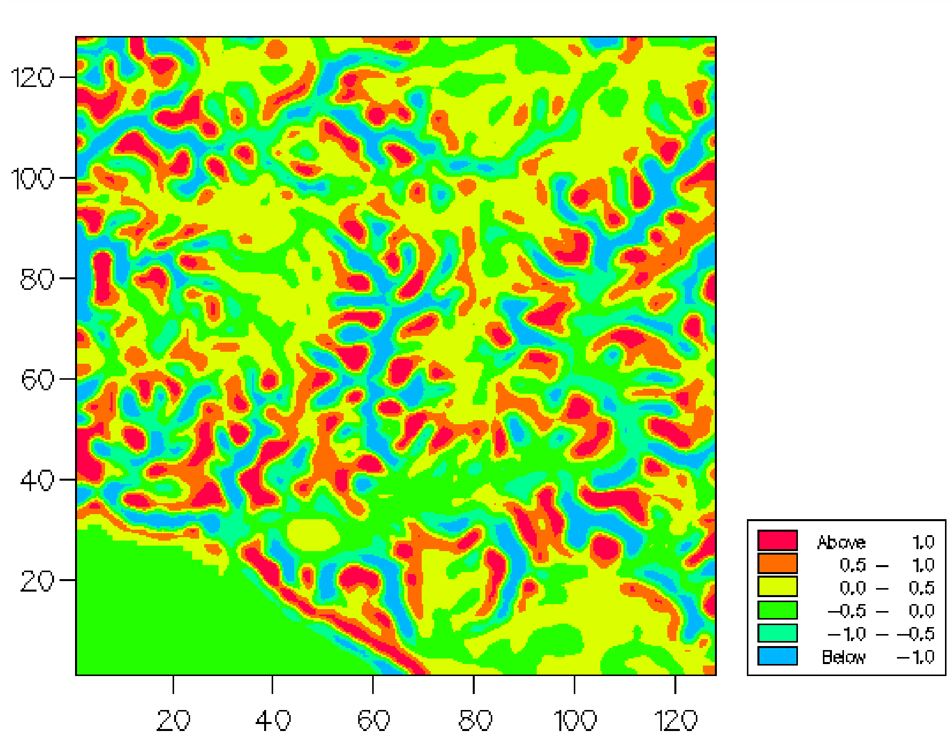
Digital elevation data for Fort A. P. Hill, USA.

Quadratic residuals

a) Long-range estimates

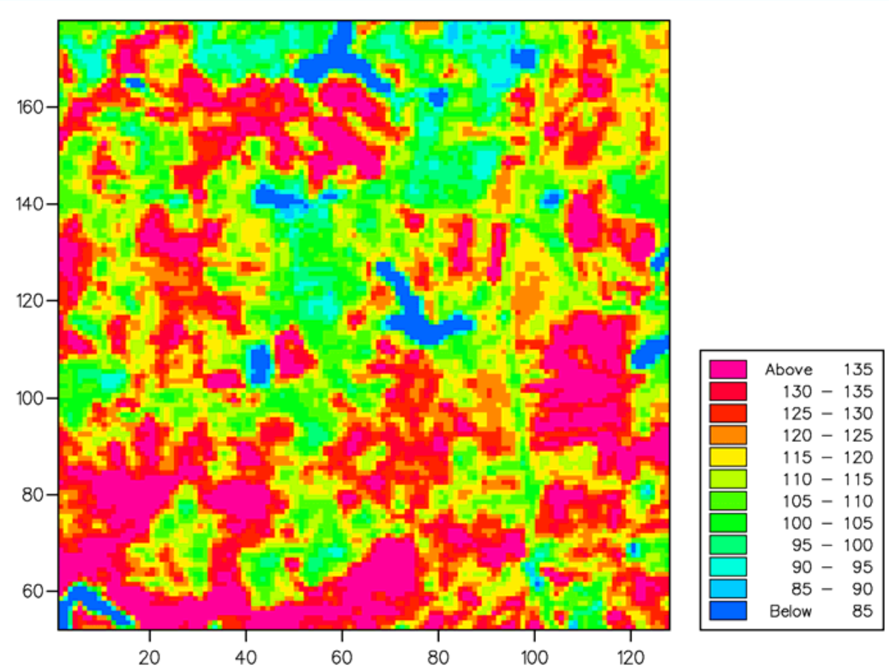


b) Short-range estimates

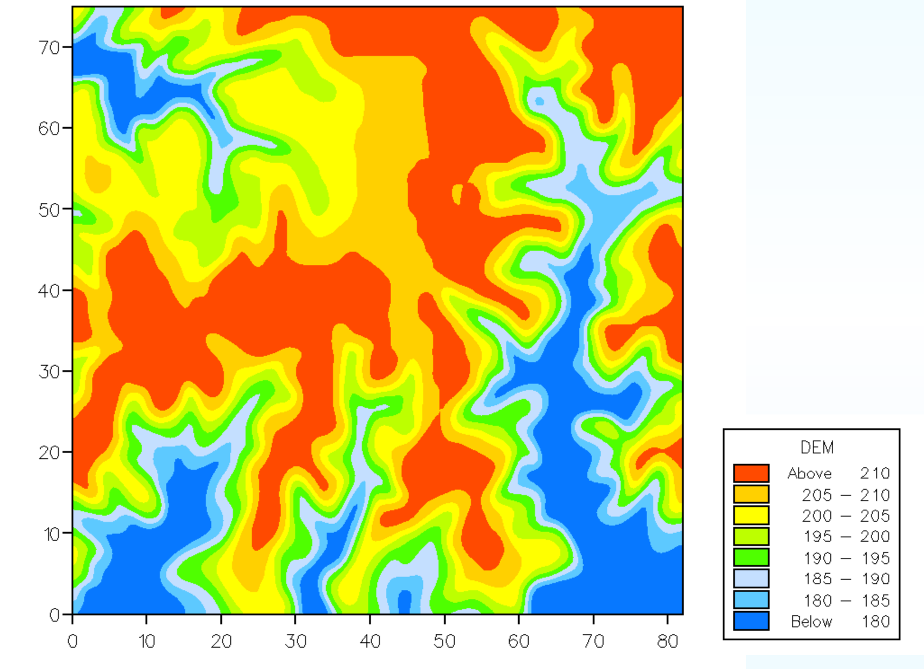


READING
SOIL SCIENCE

Fort A. P. Hill: NIR and DEM



Punctually kriged estimates of NIR



Punctually kriged estimates of DEM

Conclusions

- The richness of data from sensors often obscures the information required for interpretation.
- Nevertheless such information could be the basis for managing many aspects of the environment in the future.
- Geostatistical and other methods, such as the rapidly developing wavelet analyses, provide tools for exploring sensed data in an analytical framework.
- The links with ground information are vital and require detailed fieldwork as a precursor to using these relatively cheap sources of information as a partial substitute.