INTRODUCTION TO
PEDOMETRICS

A COURSE OF LECTURES

Alex. B. McBratney, Cunningham Laboratory,
St Lucia

INTRODUCTION

This Technical Memorandum consists of a set of student hand-outs (also
used as overhead transparencies) for an introductory lecture course in
pedometrics given by the author at the Department of Soil Science,
University of Sydney from October 21st to October 25th 1986. These
lectures form part of a larger course in Advanced Pedology for fourth year
undergraduates majoring in Soil Science.

Although some of the material presented here is not readily
comprehensible without further explanation, it is hoped that the
references, notes and many of the illustrations will be of interest and
assistance to those interested in this new and exciting area of pedological
and pedagogical development.

Much of the material contained herein is the author’s unpublished work
and is currently being written up.
COURSE OUTLINE

LECTURES

1 PEDOMETRICS

Why do we want to quantify?
Types of quantitative model
Three parallel streams of pedometrics

2 SOIL INFORMATION SYSTEMS

Outline of an information system
Relational databases
Geographical information systems
Expert systems

3 NUMERICAL CLASSIFICATION AND ALLOCATION OF SOIL

Overview of methodology
Ordination
Numerical classification
Allocation (pattern recognition)
4 SOIL GEOSTATISTICS

Types of spatial process
Assumptions- stationarity and isotropy
The semi-varioagram
Kriging
Non stationary prediction
Non linear prediction

5 QUANTITATIVE SOIL SURVEY METHODS

Conventional soil survey
Methods for quantifying conventional soil survey
Geostatistical soil survey
Real-time soil survey

TUTORIAL

Discussion of quantitative soil survey methods.
Introduction to Pedometrics

LECTURE 1

PEDOMETRICS
"In view of the considerable variations of soil and of plants I emphasized the need for proper statistical control of the design of the experiments and of the deductions drawn from them."

Sir E.J. Russell on his visit to Australia in 1928 published in
Definition

Pedometrics is that area of science concerned with the description, classification, formation and distribution of soil by quantitative mathematical and statistical techniques.

General texts


Why do we want to quantify?

1) Increased precision

2) Reproducibility
   use the same technique on the same data
gives the same answer.

3) Comparability
   use exactly the same technique on
different sets of data and compare the
results legitimately.

4) Objectivity?
   don't believe it; difficult to define and
   justify, generally personal experience
   will be involved and should not be denied.

   but we must go deeper
Scientific Methodology

**Inspiration, Imagination, Creativity**

```
<table>
<thead>
<tr>
<th>Particular observations</th>
<th>induction</th>
<th>abduction</th>
<th>General statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA</td>
<td></td>
<td></td>
<td>conjectures</td>
</tr>
<tr>
<td></td>
<td>deduction</td>
<td></td>
<td>&quot;laws&quot;</td>
</tr>
</tbody>
</table>
```

Role of quantitative methods

1) **Induction** - exploratory data analysis

2) **Deduction** of logical corollaries of the conjecture (mathematics)

3) **Falsification** of conjectures (improved - greater precision)
## Two-Way Classification of Quantitative Models

<table>
<thead>
<tr>
<th></th>
<th>Deterministic</th>
<th>Stochastic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Empirical</strong></td>
<td>Regression Generalised linear</td>
<td>Time series Spatial processes</td>
</tr>
<tr>
<td></td>
<td>Jenny’s functional relations</td>
<td>Temporal and spatial variation</td>
</tr>
<tr>
<td><strong>Mechanistic</strong></td>
<td>Flow and diffusion Profile and</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>landscape development</td>
<td></td>
</tr>
</tbody>
</table>
A Qn b}an

Test for "ARGILLIC"

mode < 2μm and $\frac{X^2_{\text{data}} - X^2_{\text{model}}}{4 \text{df} \cdot \text{var}} > 9.488$

Ali 0-8
Bl 40-60
C 120-40
Stochastic process

"in the first place some possible actual, e.g. physical, process in the real world, that has some random or stochastic element in its structure. It is convenient, however, also to use the same phrase for the mathematical representation as well as the physical concept."


Probability theory

"is the only mathematical model available to the scientist who seeks to map the unknown or controllable; fortunately, it is at the same time extraordinarily powerful and convenient"

\[ S(t) = 0.5 \, S(t-1) + e(t), \quad e(t) \sim N(0,1) \]
Simulation of soil and slope development

Three most important processes:
soil creep, fluvial erosion and weathering

\[ C_S = K_S \sin \alpha \]
where \( C_S \) is the rate of soil creep, \( K_S \) is a constant and \( \alpha \) is the slope angle.

\[ C_r = K_r Q S \]
where \( C_r \) is the fluvial transportation rate, \( K_r \) is a constant representing the efficiency of the fluvial system in transporting sediment, \( Q \) is the discharge and \( S \) is the slope of the river.

\[ W = W_{pot} \exp (-K_d D) \]
where \( W \) is the rate of weathering, \( W_{pot} \) is the potential rate of weathering of a bare rock surface, \( K_d \) is a constant and \( D \) is soil depth.

Using finite difference approach divide landscape up into cells for which:

Change in height = Inflow - Outflow
Change in soil depth = Inflow - Outflow + Weathering
Soil Landscape systems modelling

References


Schematic diagram of the simulated landform.

Height of the top and bottom of a single side slope against time. Superimposed on the plot are the slope profiles that constitute the two ends of the plotted points (Comparable with DAVIS's diagram of the evolution of the geographical cycle.)

(from Armstrong)
Mean, maximum, and minimum heights of the simulated basin, mean erosion rate and the mean soil depths against simulated time.

(from Armstrong)
Sequence of block diagrams illustrating the global evolution of the soil depth distributions. The initial distribution, of a completely uniform soil depth at all locations is not shown. Time $t$ is in iterations.

(from Aspeny)
Typical sequence of soil depth and slope form profiles developed by the simulation. Profile a is the initial form, b the state after 500 iterations, c after 1000 iterations, and d after 2000 iterations. Subsequent profiles in the sequence (d to m) are plotted at increments of 2000 iterations.
THREE PARALLEL STREAMS OF PEDOMETRICS (PEDOLOGY)

- Soil survey and classification
- Genetic soil classification
- Dynamic soil-landscape modelling
2 SOIL INFORMATION SYSTEMS

Outline of an information system

Relational databases

Geographical information systems

Expert systems
SCHEMATIC DIAGRAM OF AN INFORMATION SYSTEM

DATA GENERATION

DATA INPUT

DATABASE MANAGEMENT

DATA ANALYSIS & SYNTHESIS

DATA DISPLAY

DATA MANAGEMENT

DECISION MAKING
Relational databases

Soil scientists naturally store and manipulate data in tables with rows and columns.

In database jargon tables are called relations with columns called fields or attributes and rows called records.

Each table or relation should have a unique identifier, with a fixed number of fields (which may be varied) and a variable number of records.

Each record must have a unique identifier.

Linking data from separate tables is achieved via a common field.

The form in which the data is stored is immaterial to the user, although it is important that it is stored efficiently and can be retrieved quickly.
Soil Information Systems

STAG Soil Relational Database
7 Oct 1986 - 08:41
System Menu

1. Schema Maintenance
2. Schema Listing
3. Create Data Base
4. SFORM Menu
5. ENTER Screen Registration
6. SQL - Query/DML Language
7. SQL Screen Registration
8. Listing Processor

9. Data Base Test Driver
10. MENUH Screen Menu
11. MENUH Report Menu
12. Reconfigure Data Base
13. Write Data Base Backup
14. Read Data Base Backup
15. Data Base Maintenance Menu

SELECTION: 1

Soil Information Systems

STAG Soil Relational Database
7 Oct 1986 - 08:42
Schema Maintenance

DATABASE ID: 1

<table>
<thead>
<tr>
<th>LN</th>
<th>CMD</th>
<th>RECORD</th>
<th>EXPECTED</th>
<th>LONG NAME</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>site</td>
<td>150</td>
<td>site</td>
<td>site</td>
<td>Samford STAG site descr</td>
</tr>
<tr>
<td>2</td>
<td>horizon</td>
<td>1000</td>
<td>horizon</td>
<td>STAG horizon description</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>substrat</td>
<td>150</td>
<td>substrat</td>
<td>Substratum description</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>chem0_10</td>
<td>150</td>
<td>chem0to10cm</td>
<td>Chemistry 0 to 10cm</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>chem070</td>
<td>150</td>
<td>chem0to70cm</td>
<td>Chemistry 0 to 70cm</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>phys5_85</td>
<td>1000</td>
<td>phys0to90cm</td>
<td>Physical 0 to 90cm</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>psa0_10</td>
<td>150</td>
<td>psa0to10cm</td>
<td>Part. size 0-10cm wh.soil</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>psa070</td>
<td>150</td>
<td>psa0to70cm</td>
<td>P. size 0-70cm wh.soil</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>infiltrn</td>
<td>150</td>
<td>infiltration</td>
<td>Surface sorptivity etc.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>tcolour</td>
<td>300</td>
<td>transcolour</td>
<td>Translate Munsell/CIELab</td>
<td></td>
</tr>
</tbody>
</table>
### Soil Information Systems

**STAG Soil Relational Database**  
7 Oct 1986 - 08:42  
Schema Maintenance

**RECORD: site**

<table>
<thead>
<tr>
<th>LN</th>
<th>CMD</th>
<th>FIELD</th>
<th>KEY REF</th>
<th>TYPE</th>
<th>LEN</th>
<th>LONG NAME</th>
<th>COMB. FIELD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>ident</td>
<td>*</td>
<td>NUMERIC</td>
<td>3</td>
<td>sid</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>north</td>
<td></td>
<td>FLOAT</td>
<td>91</td>
<td>northing</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>east</td>
<td></td>
<td>FLOAT</td>
<td>81</td>
<td>easting</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>elev</td>
<td></td>
<td>FLOAT</td>
<td>52</td>
<td>elev</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>pfp</td>
<td></td>
<td>STRING</td>
<td>6</td>
<td>pfp</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>gsg</td>
<td></td>
<td>STRING</td>
<td>4</td>
<td>gsg</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>person</td>
<td></td>
<td>STRING</td>
<td>4</td>
<td>pers</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>slope</td>
<td></td>
<td>NUMERIC</td>
<td>2</td>
<td>sl</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>azimuth</td>
<td></td>
<td>NUMERIC</td>
<td>3</td>
<td>azi</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>dated</td>
<td></td>
<td>STRING</td>
<td>6</td>
<td>dated</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>x</td>
<td></td>
<td>NUMERIC</td>
<td>2</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

[Next page, [P]rev page, [A]dd line, or number: n]

### Soil Information Systems

**STAG Soil Relational Database**  
7 Oct 1986 - 08:42  
Schema Maintenance

**RECORD: site**

<table>
<thead>
<tr>
<th>LN</th>
<th>CMD</th>
<th>FIELD</th>
<th>KEY REF</th>
<th>TYPE</th>
<th>LEN</th>
<th>LONG NAME</th>
<th>COMB. FIELD</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td></td>
<td>y</td>
<td></td>
<td>NUMERIC</td>
<td>2</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>column</td>
<td></td>
<td>NUMERIC</td>
<td>2</td>
<td>co</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>row</td>
<td></td>
<td>NUMERIC</td>
<td>2</td>
<td>ro</td>
<td></td>
</tr>
</tbody>
</table>

[Next page, [P]rev page, [A]dd line, or number: ]
1. Paint Screen
2. Screen Entry
3. Test Screen
4. Process Screen
5. Screen Reports
6. Restore Screen
7. List Screens
8. Create Default Screen Form

SELECTION: site
Soil Information Systems

STAG Soil Relational Database
7 Oct 1986 - 11:19
Test Screen

cid: x

hue: x
value: x
chroma: x

CIE-L*: x
CIE-a*: x
CIE-b*: x

SCREEN: tcolour
Soil Information Systems

sql> help
-- Structured Query Language (SQL)

SQL is a query language based on an English keyword syntax. A query is composed of a series of keyword clauses. The keywords that introduce each clause are as follows:

select
from
where
group by
having
order by
into

Help about any keyword can be obtained by typing "help <keyword>".
The complete list of SQL keywords is as follows:

and desc help min separator where
asc edit in not set write
avg end insert or start
between fields into order sum
by from is records unique
count group lines restart unlock
delete having max select update

sql>

Soil Information Systems

UNIFY SQL -- VERSION 3.1
Copyright Unify Corporation 1983

sql> records
site horizon substrat chem0_10 chem6070 phys5_85 psa0_10 psa6070
infiltrm tcolour
sql> fields phys5_85
NAME TYPE LENGTH
plid LONG 5
sid INTEGER 3
adcm INTEGER 2
lbdws FLOAT 42
sbdws FLOAT 42
lgws FLOAT 42
sgws FLOAT 42
smwps FLOAT 42
sgwps FLOAT 42
swpe FLOAT 42
swgps FLOAT 42
sql>
### Soil Information Systems

```sql
sql> fields tcolour
NAME | TYPE       | LENGTH
-----+-----------+--------
cid  | INTEGER   | 3      
hue_ | STRING    | 5      
va   | INTEGER   | 1      
chr  | INTEGER   | 1      
CIE_L| FLOAT     | 52     
CIE_a| FLOAT     | 52     
CIE_b| FLOAT     | 52     
```

```sql
sql> select * from tcolour where hue_ = '10yr'
recognized query!

```
cid | hue_  | va | chr | CIE_L | CIE_a | CIE_b
-----+-------+----+-----+-------+-------+-------
300  | 10yr  | 2  | 1   | 28.54000000000000000 | 1.97000000000000000 | 5.44000000000000000
301  | 10yr  | 2  | 2   | 28.54000000000000000 | 3.62000000000000000 | 11.08000000000000000
302  | 10yr  | 2  | 3   | 28.54000000000000000 | 5.55000000000000000 | 17.08000000000000000
303  | 10yr  | 3  | 1   | 30.77000000000000000 | 1.87000000000000000 | 6.02000000000000000
304  | 10yr  | 3  | 2   | 30.77000000000000000 | 3.47000000000000000 | 12.21000000000000000
305  | 10yr  | 3  | 3   | 30.77000000000000000 | 5.13000000000000000 | 17.87000000000000000
306  | 10yr  | 3  | 4   | 30.77000000000000000 | 6.61000000000000000 | 24.06000000000000000
307  | 10yr  | 4  | 1   | 41.22000000000000000 | 1.72000000000000000 | 6.82000000000000000
311  | 10yr  | 4  | 6   | 41.22000000000000000 | 9.32000000000000000 | 37.80000000000000000
```

```sql
sql> icat Atwos
select count(*)
from site, horizon
where gsg_ = '*rp*' and hdes = '*A2*'
and site.sid= horizon.sid
/
sql> start Atwos
recognized query!

```
```
```
count(*)
65
```

```sql
sql> select count(*)
sql> from site
sql> where gsg_ = '*rp*'
sql> /
recognized query!

```
```
count(*)
72
```
```
```
sql> !cat sqlpH

SELECT horid, horizon.sid, hdes, upb, lowb, pHwat50
FROM horizon, chem6070
WHERE horizon.sid=chem6070.sid
AND ((upb<=0.6 AND lowb<=0.6) OR (upb<=0.7 AND lowb>=0.7)
OR (upb>=0.6 AND lowb<=0.7))
INTO phqresult
/

sql> start sqlpH

recognized query!
sql> !more phqresult

<table>
<thead>
<tr>
<th>horid</th>
<th>horizon.sid</th>
<th>hdes</th>
<th>upb</th>
<th>lowb</th>
<th>pHwat50</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>2</td>
<td>B2</td>
<td></td>
<td>0.35000</td>
<td>0.75000</td>
</tr>
<tr>
<td>35</td>
<td>3</td>
<td>B2</td>
<td></td>
<td>0.50000</td>
<td>0.90000</td>
</tr>
<tr>
<td>43</td>
<td>4</td>
<td>B2</td>
<td></td>
<td>0.50000</td>
<td>0.90000</td>
</tr>
<tr>
<td>55</td>
<td>5</td>
<td>B22</td>
<td></td>
<td>0.50000</td>
<td>0.64000</td>
</tr>
<tr>
<td>56</td>
<td>5</td>
<td>B3</td>
<td></td>
<td>0.64000</td>
<td>0.75000</td>
</tr>
<tr>
<td>94</td>
<td>9</td>
<td>B1</td>
<td></td>
<td>0.44000</td>
<td>0.61000</td>
</tr>
<tr>
<td>104</td>
<td>1</td>
<td>B2</td>
<td></td>
<td>0.46000</td>
<td>0.69000</td>
</tr>
<tr>
<td>105</td>
<td>1</td>
<td>B3</td>
<td></td>
<td>0.69000</td>
<td>0.81000</td>
</tr>
<tr>
<td>123</td>
<td>1</td>
<td>B2</td>
<td></td>
<td>0.35000</td>
<td>0.74000</td>
</tr>
<tr>
<td>154</td>
<td>1</td>
<td>B2</td>
<td></td>
<td>0.50000</td>
<td>0.70000</td>
</tr>
</tbody>
</table>
```
### Soil Information Systems

<table>
<thead>
<tr>
<th>NAME</th>
<th>TYPE</th>
<th>LENGTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>sid</td>
<td>INTEGER</td>
<td>3</td>
</tr>
<tr>
<td>cs</td>
<td>FLOAT</td>
<td>41</td>
</tr>
<tr>
<td>fs</td>
<td>FLOAT</td>
<td>41</td>
</tr>
<tr>
<td>zi</td>
<td>FLOAT</td>
<td>41</td>
</tr>
<tr>
<td>cl</td>
<td>FLOAT</td>
<td>41</td>
</tr>
<tr>
<td>gr</td>
<td>FLOAT</td>
<td>41</td>
</tr>
<tr>
<td>om</td>
<td>FLOAT</td>
<td>41</td>
</tr>
</tbody>
</table>

```sql
sql> select sid, cl, gr, om from pso0_10
sql> into clgrom/
recognized query!
sql> !head clgrom
```

<table>
<thead>
<tr>
<th>sid</th>
<th>cl</th>
<th>gr</th>
<th>om</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.91700</td>
<td>36.10000</td>
<td>3.19500</td>
</tr>
<tr>
<td>2</td>
<td>7.70000</td>
<td>23.00000</td>
<td>5.39000</td>
</tr>
<tr>
<td>3</td>
<td>6.57000</td>
<td>34.30000</td>
<td>4.59900</td>
</tr>
<tr>
<td>4</td>
<td>12.65149</td>
<td>28.40000</td>
<td>4.25346</td>
</tr>
<tr>
<td>5</td>
<td>9.40606</td>
<td>22.40000</td>
<td>5.48687</td>
</tr>
<tr>
<td>6</td>
<td>9.50784</td>
<td>25.40000</td>
<td>5.85098</td>
</tr>
<tr>
<td>7</td>
<td>6.27900</td>
<td>51.70000</td>
<td>3.86400</td>
</tr>
</tbody>
</table>
Geographical Information Systems

Geographical information systems are information systems which are based on data referenced by geographical co-ordinates.

References


Diagrammatic representation of a 12 level geographic information system

<table>
<thead>
<tr>
<th>Objective</th>
<th>Example</th>
<th>Data elements (Figure 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>Retrieval of one or more data elements</td>
<td>4 Precipitation</td>
</tr>
<tr>
<td>ii</td>
<td>Transform, manipulate or combine the values of one or more data elements</td>
<td>1 Topography</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 Land cover</td>
</tr>
<tr>
<td>iii</td>
<td>Store new data elements</td>
<td>12 Spare</td>
</tr>
<tr>
<td>iv</td>
<td>Search and identify particular combinations of data elements</td>
<td>5 Population</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 Land cover</td>
</tr>
<tr>
<td>v</td>
<td>Perform statistical analysis on one or more data elements</td>
<td>8 Crop type</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 Landsat MSS 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11 Landsat MSS 7</td>
</tr>
<tr>
<td>vi</td>
<td>Measure area and distance of one or more data elements</td>
<td>2 Geology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 Land cover</td>
</tr>
<tr>
<td>vii</td>
<td>Produce composite images of several data elements</td>
<td>1 Topography</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Precipitation</td>
</tr>
<tr>
<td>viii</td>
<td>Be capable of modelling using several data elements</td>
<td>1 Topography</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 Soils</td>
</tr>
<tr>
<td></td>
<td>If it is suspected that the majority of the archaeological sites are</td>
<td>5 Population</td>
</tr>
<tr>
<td></td>
<td>located in unpopulated clay valleys display an image of unpopulated,</td>
<td>6 Archaeology</td>
</tr>
<tr>
<td></td>
<td>clay valleys and compare with known archaeological sites</td>
<td></td>
</tr>
</tbody>
</table>
Schematic representations of how ARC/INFO holds topology, geometry and attributes of geographical data

Options available in computing the overlay of two geographical data sets
THREE CONCEPTUAL COMPONENTS OF A GEOGRAPHIC BASED INFORMATION SYSTEM

(from Mulca et al.)
<table>
<thead>
<tr>
<th>Feature Data</th>
<th>Points</th>
<th>Lines</th>
<th>Polygons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point Feature (Archaeological Site)</td>
<td>Linear Features (Roads)</td>
<td>Homogeneous Polygons (Soils)</td>
<td></td>
</tr>
<tr>
<td>Aerial Units</td>
<td>Polygon Centroids</td>
<td>Administrative Polygon Boundaries</td>
<td>Aerial Unit (Census Tract)</td>
</tr>
<tr>
<td>Network Topology</td>
<td>Nodes (Intersections)</td>
<td>Links (Streets)</td>
<td>Polygons (Blocks)</td>
</tr>
<tr>
<td>Sampling Records</td>
<td>+09</td>
<td>+405</td>
<td>73</td>
</tr>
<tr>
<td>Surface Data</td>
<td>Weather Station</td>
<td>Flight Lines</td>
<td>Field Test Plots</td>
</tr>
<tr>
<td>+206</td>
<td>+207</td>
<td>+206</td>
<td>+207</td>
</tr>
<tr>
<td>Labeled/Text Data</td>
<td>+Redlands</td>
<td>Santa Ana River</td>
<td>Commercial</td>
</tr>
<tr>
<td>+Fontana</td>
<td>+Cotton</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Place Names</td>
<td>Linear Feature</td>
<td>Polygon Naming</td>
<td></td>
</tr>
<tr>
<td>Graphic Symbol Data</td>
<td>Point Symbols</td>
<td>Line Symbols</td>
<td>Polygon Shading</td>
</tr>
</tbody>
</table>

Breakdown of Geographic Data Types & Methods of Representation

(from Marble et al.)
### X, Y Coordinate File

A map can be expressed in Cartesian Coordinates by describing the locations of various features. Here is a table showing how different features are represented:

<table>
<thead>
<tr>
<th>Feature Number</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point</td>
<td>$X,Y$ (Single Point)</td>
</tr>
<tr>
<td>Line</td>
<td>$X_1Y_1, X_2Y_2, \ldots X_nY_n$ (string)</td>
</tr>
<tr>
<td>Polygon</td>
<td>$X_1Y_1, X_2Y_2, \ldots X_nY_n$ (closed loop)</td>
</tr>
</tbody>
</table>
### Basic Readings in Geographic Information Systems

![Coded Network Map](image)

<table>
<thead>
<tr>
<th>Link #</th>
<th>Right Polygon</th>
<th>Left Polygon</th>
<th>Node 1</th>
<th>Node 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>0</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

### Topologically Coded Network & Polygon File

<table>
<thead>
<tr>
<th>Node #</th>
<th>X Coordinate</th>
<th>Y Coordinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
<td>38</td>
</tr>
</tbody>
</table>

### X, Y Coordinate Node File

**X, Y COORDINATE NODE FILE**

*From Markle et al.*
GRID FILE

(from Middle et al.)
Basic Readings in Geographic Information Systems

FIG. 7 MANUAL CELL ENCODING

MANUAL DIGITIZER

(source: McIlv et al.)
DATA RETRIEVAL

MAP GENERALIZATION
MAP SHEET MANIPULATION

SCALE CHANGE
DISTORTION REMOVAL (LINEAR AND RUBBER SHEETING)
PROJECTION CHANGE
COORDINATE ROTATION/TRANSLATION
BUFFER GENERATION

OVERLAY FOR MAP COMPOSITE

POLYGON OVERLAY/DISSOLVE
Basic Readings in Geographic Information Systems

(Boolean Map Overlay)

(Area Calculations Overlay)

(Search Radius Aggregations)

(Distance Calculation (Access))

(Optimum Corridor Selection)

(Grid Cell Technique)

(from Meule et al)
Basic Readings in Geographic Information Systems

![Diagram of measurement methods](from Mcle et al)
Basic Readings in Geographic Information Systems

DIGITAL TERRAIN ANALYSIS

(from Marble et al.)
Basic Readings in Geographic Information Systems

OUTPUT FORMAT

FILES

CRT DISPLAY

LISTINGS

MAPS

(from Meeks et al.)
Expert systems

An expert system can be defined as a computer system that solves problems normally requiring a human "expert".

References


According to McCracken and Cate expert systems can advise, analyse, categorize, consult, design, diagnose, explain, explore, forecast, form concepts, identify, interpret, justify, learn, manage, monitor, plan, present, retrieve, schedule, test, and tutor.

If only they could.
Two extreme types of computerised knowledge systems. The upper one uses a human expert and a knowledge engineer to derive a knowledge base. The lower one induces knowledge using numerical taxonomic methods from a set of data. Many possible systems between these two extremes exist. Note the expert or decision system is identical for both extremes.
<table>
<thead>
<tr>
<th>Source of existing soil classificatory knowledge</th>
<th>Organisation of existing soil classificatory knowledge and related data</th>
<th>Soil class assignment and related information</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEDOLOGISTS</td>
<td>KNOWLEDGE ENGINEER</td>
<td>KNOWLEDGE BASE</td>
</tr>
<tr>
<td>EXISTING SOIL CLASSIFICATION SYSTEM</td>
<td></td>
<td>Soil class assignment rules in specified order</td>
</tr>
<tr>
<td></td>
<td>test self consistency of assignment rules</td>
<td>INFEERENCE ENGINE</td>
</tr>
<tr>
<td></td>
<td>Soil Class Database</td>
<td>Assignment using deductive logic on rules given</td>
</tr>
<tr>
<td></td>
<td>soil and related land use data and summary statistics</td>
<td>some data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>USER INTERFACE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>interactive questions &amp; answers</td>
</tr>
<tr>
<td></td>
<td>&quot;What kind of soil is this?&quot;</td>
<td>Soil scientist, agronomist, land manager, land</td>
</tr>
<tr>
<td></td>
<td>&quot;How well will this soil grow a certain crop?&quot;</td>
<td>user</td>
</tr>
<tr>
<td></td>
<td>&quot;Does this soil have a high erosion hazard?&quot;</td>
<td>SOIL DATA</td>
</tr>
<tr>
<td></td>
<td>&quot;What management practices are suggested for this soil?&quot;</td>
<td>(usually limited)</td>
</tr>
</tbody>
</table>
Introduction to Pedometrics

LECTURE 3

NUMERICAL CLASSIFICATION AND ALLOCATION OF SOIL
3 NUMERICAL CLASSIFICATION AND ALLOCATION OF SOIL

Overview - multivariate methods, matrices

Ordination - dimension reduction

Numerical classification

Allocation (identification)
Multivariate methods

Many soil data sets \( \mathbf{X} \) can be regarded as tables or matrices with \( n \) rows each representing an individual and \( p \) columns each representing a soil property, attribute or variate.

Methods which deal with such multi-column matrices are called multivariate methods.

Some soil data, such as profile description data, are not easy to describe as \( n \times p \) matrices and require transformation first.
> print(exdat, rowlab=exrownname, collab=excolname)

Array:
8 by 5

<table>
<thead>
<tr>
<th></th>
<th>CEC</th>
<th>%Cl</th>
<th>pH1</th>
<th>pH5</th>
<th>%OC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0981</td>
<td>5</td>
<td>9</td>
<td>6.3</td>
<td>6.3</td>
<td>0.9</td>
</tr>
<tr>
<td>A0982</td>
<td>18</td>
<td>28</td>
<td>6.1</td>
<td>6.5</td>
<td>2.0</td>
</tr>
<tr>
<td>A0983</td>
<td>20</td>
<td>18</td>
<td>8.3</td>
<td>8.8</td>
<td>1.3</td>
</tr>
<tr>
<td>A0984</td>
<td>16</td>
<td>23</td>
<td>8.0</td>
<td>8.8</td>
<td>0.7</td>
</tr>
<tr>
<td>A0985</td>
<td>10</td>
<td>13</td>
<td>6.1</td>
<td>8.2</td>
<td>0.5</td>
</tr>
<tr>
<td>A0986</td>
<td>6</td>
<td>18</td>
<td>7.8</td>
<td>8.8</td>
<td>0.3</td>
</tr>
<tr>
<td>A0987</td>
<td>13</td>
<td>13</td>
<td>7.5</td>
<td>7.6</td>
<td>1.0</td>
</tr>
<tr>
<td>A0988</td>
<td>6</td>
<td>12</td>
<td>7.3</td>
<td>7.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>

> smatrix(exdat, rowlab=exrownname, collab=excolname)

exdat

A0981 . . . . .
A0982 * * . . . *
A0983 * + * * * +
A0984 * * * * *
A0985 + . * * *
A0986 . . * * .
A0987 + . + + +
A0988 . . + + .

C % p p p %
E C H H O
C I I 5 C

> # symbols . small, + medium, * large
Example physiognomic classification

A0981

A0984

A0987

A0982

A0985

A0988

A0983

A0986

A0989
Metrics, Dissimilarity of individuals and Dissimilarity Matrices

**Euclidean metric**

Diagonal metric

Mahalanobis metric

Pythagorean distance

**Manhattan metric**

Sum of absolute deviations

**Maximum metric**

Sum of maximum differences

**Gower metric**

Deals with continuous, binary and multistate attributes

**Similarity, Dissimilarity or Distance Matrix**

The similarities or dissimilarities between all individuals j and k form a n row by n column symmetric matrix S or D with elements $S_{j,k}$ or $D_{j,k}$. 
> # distance matrix Euclidean metric
> print(dist2full(dist(scaledexdat,"euclidean")),rowlab=exrownname,collab=excolnname)
>
<table>
<thead>
<tr>
<th></th>
<th>A0981</th>
<th>A0982</th>
<th>A0983</th>
<th>A0984</th>
<th>A0985</th>
<th>A0986</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0981</td>
<td>0.0</td>
<td>4.201850</td>
<td>4.668290</td>
<td>4.285674</td>
<td>3.121144</td>
<td>2.223284</td>
</tr>
<tr>
<td>A0982</td>
<td>4.201850</td>
<td>0.0</td>
<td>3.982339</td>
<td>4.625358</td>
<td>4.741124</td>
<td>5.445993</td>
</tr>
<tr>
<td>A0983</td>
<td>4.668290</td>
<td>3.982339</td>
<td>0.0</td>
<td>1.516429</td>
<td>2.431228</td>
<td>3.264931</td>
</tr>
<tr>
<td>A0984</td>
<td>4.285674</td>
<td>4.625358</td>
<td>1.516429</td>
<td>0.0</td>
<td>1.939597</td>
<td>2.686547</td>
</tr>
<tr>
<td>A0985</td>
<td>3.121144</td>
<td>4.741124</td>
<td>2.431228</td>
<td>1.939597</td>
<td>0.0</td>
<td>1.132066</td>
</tr>
<tr>
<td>A0986</td>
<td>2.223284</td>
<td>5.445993</td>
<td>3.264931</td>
<td>2.686547</td>
<td>1.132066</td>
<td>0.0</td>
</tr>
<tr>
<td>A0987</td>
<td>2.478865</td>
<td>3.577211</td>
<td>2.136369</td>
<td>2.136369</td>
<td>1.395655</td>
<td>2.178463</td>
</tr>
<tr>
<td>A0988</td>
<td>1.849028</td>
<td>4.558748</td>
<td>3.533812</td>
<td>2.922504</td>
<td>1.448967</td>
<td>1.569984</td>
</tr>
<tr>
<td>A0981</td>
<td>0.0</td>
<td>2.839108</td>
<td>2.553776</td>
<td>2.491340</td>
<td>2.185576</td>
<td>2.491340</td>
</tr>
<tr>
<td>A0982</td>
<td>2.839108</td>
<td>0.0</td>
<td>2.674924</td>
<td>2.321534</td>
<td>2.678963</td>
<td>3.835852</td>
</tr>
<tr>
<td>A0983</td>
<td>2.553776</td>
<td>2.674924</td>
<td>0.0</td>
<td>1.871477</td>
<td>1.782514</td>
<td>2.385319</td>
</tr>
<tr>
<td>A0984</td>
<td>2.491340</td>
<td>2.321534</td>
<td>1.871477</td>
<td>0.0</td>
<td>1.489531</td>
<td>1.936399</td>
</tr>
<tr>
<td>A0985</td>
<td>2.185576</td>
<td>2.678963</td>
<td>1.782514</td>
<td>1.489531</td>
<td>0.0</td>
<td>0.681065</td>
</tr>
<tr>
<td>A0986</td>
<td>2.491340</td>
<td>3.835852</td>
<td>2.385319</td>
<td>1.936399</td>
<td>0.681065</td>
<td>0.0</td>
</tr>
<tr>
<td>A0987</td>
<td>1.459820</td>
<td>2.234296</td>
<td>1.959462</td>
<td>1.489531</td>
<td>0.6928975</td>
<td>1.250856</td>
</tr>
<tr>
<td>A0988</td>
<td>1.215875</td>
<td>2.857273</td>
<td>2.383518</td>
<td>1.782513</td>
<td>0.972668</td>
<td>1.395152</td>
</tr>
<tr>
<td>A0981</td>
<td>1.215875</td>
<td>2.857273</td>
<td>2.553776</td>
<td>2.491340</td>
<td>2.185576</td>
<td>2.491340</td>
</tr>
<tr>
<td>A0982</td>
<td>2.857273</td>
<td>0.0</td>
<td>2.674924</td>
<td>2.321534</td>
<td>2.678963</td>
<td>3.835852</td>
</tr>
<tr>
<td>A0983</td>
<td>2.553776</td>
<td>2.674924</td>
<td>0.0</td>
<td>1.871477</td>
<td>1.782514</td>
<td>2.385319</td>
</tr>
<tr>
<td>A0984</td>
<td>2.491340</td>
<td>2.321534</td>
<td>1.871477</td>
<td>0.0</td>
<td>1.489531</td>
<td>1.936399</td>
</tr>
<tr>
<td>A0985</td>
<td>2.185576</td>
<td>2.678963</td>
<td>1.782514</td>
<td>1.489531</td>
<td>0.0</td>
<td>0.681065</td>
</tr>
<tr>
<td>A0986</td>
<td>2.491340</td>
<td>3.835852</td>
<td>2.385319</td>
<td>1.936399</td>
<td>0.681065</td>
<td>0.0</td>
</tr>
<tr>
<td>A0987</td>
<td>1.459820</td>
<td>2.234296</td>
<td>1.959462</td>
<td>1.489531</td>
<td>0.6928975</td>
<td>1.250856</td>
</tr>
<tr>
<td>A0988</td>
<td>1.215875</td>
<td>2.857273</td>
<td>2.383518</td>
<td>1.782513</td>
<td>0.972668</td>
<td>1.395152</td>
</tr>
</tbody>
</table>

> # distances from maximum metric above
> edit

<table>
<thead>
<tr>
<th></th>
<th>A0981</th>
<th>A0982</th>
<th>A0983</th>
<th>A0984</th>
<th>A0985</th>
<th>A0986</th>
<th>A0987</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0981</td>
<td>0.0</td>
<td>7.728465</td>
<td>9.716335</td>
<td>8.695014</td>
<td>6.664862</td>
<td>5.527256</td>
<td>5.869529</td>
</tr>
<tr>
<td>A0982</td>
<td>7.728465</td>
<td>0.0</td>
<td>7.947394</td>
<td>7.909334</td>
<td>10.39128</td>
<td>12.01938</td>
<td>7.578169</td>
</tr>
<tr>
<td>A0983</td>
<td>9.716335</td>
<td>7.947394</td>
<td>0.0</td>
<td>2.862010</td>
<td>4.717012</td>
<td>5.968875</td>
<td>4.648887</td>
</tr>
<tr>
<td>A0984</td>
<td>8.695014</td>
<td>7.909334</td>
<td>2.862010</td>
<td>0.0</td>
<td>3.587708</td>
<td>4.596396</td>
<td>4.339803</td>
</tr>
<tr>
<td>A0985</td>
<td>6.664862</td>
<td>10.39128</td>
<td>4.717012</td>
<td>3.587708</td>
<td>0.0</td>
<td>2.447787</td>
<td>2.731100</td>
</tr>
<tr>
<td>A0986</td>
<td>5.527256</td>
<td>12.01938</td>
<td>5.968875</td>
<td>4.596396</td>
<td>2.447787</td>
<td>0.0</td>
<td>4.449279</td>
</tr>
<tr>
<td>A0987</td>
<td>5.869529</td>
<td>4.648887</td>
<td>4.339803</td>
<td>2.731100</td>
<td>4.449279</td>
<td>0.0</td>
<td>8.01489</td>
</tr>
<tr>
<td>A0988</td>
<td>3.643494</td>
<td>9.539814</td>
<td>7.495476</td>
<td>6.122999</td>
<td>2.778467</td>
<td>2.479574</td>
<td>0.0</td>
</tr>
</tbody>
</table>

> # distances from Manhattan metric (sum of absolute differences) above
> edit
```R
> print(exdat, rowlab = exrownname, collab = excolnname)

Array:
8 by 5

<table>
<thead>
<tr>
<th></th>
<th>CEC</th>
<th>%Cl</th>
<th>pH1</th>
<th>pH5</th>
<th>%OC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0981</td>
<td>5</td>
<td>9</td>
<td>6.3</td>
<td>6.3</td>
<td>0.8</td>
</tr>
<tr>
<td>A0982</td>
<td>18</td>
<td>28</td>
<td>6.1</td>
<td>6.6</td>
<td>2.0</td>
</tr>
<tr>
<td>A0983</td>
<td>20</td>
<td>18</td>
<td>8.3</td>
<td>8.8</td>
<td>1.3</td>
</tr>
<tr>
<td>A0984</td>
<td>16</td>
<td>23</td>
<td>8.0</td>
<td>8.8</td>
<td>0.7</td>
</tr>
<tr>
<td>A0985</td>
<td>10</td>
<td>13</td>
<td>8.1</td>
<td>8.2</td>
<td>0.5</td>
</tr>
<tr>
<td>A0986</td>
<td>6</td>
<td>10</td>
<td>7.8</td>
<td>8.8</td>
<td>0.3</td>
</tr>
<tr>
<td>A0987</td>
<td>13</td>
<td>13</td>
<td>7.5</td>
<td>7.6</td>
<td>1.0</td>
</tr>
<tr>
<td>A0988</td>
<td>6</td>
<td>12</td>
<td>7.3</td>
<td>7.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>

# scale columns by dividing by standard deviation

> print(scaledexdat, rowlab = exrownname, collab = excolnname)

Array:
8 by 5

<table>
<thead>
<tr>
<th></th>
<th>CEC</th>
<th>%Cl</th>
<th>pH1</th>
<th>pH5</th>
<th>%OC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0981</td>
<td>0.851257</td>
<td>1.348578</td>
<td>7.660816</td>
<td>6.278177</td>
<td>1.428637</td>
</tr>
<tr>
<td>A0982</td>
<td>3.064524</td>
<td>4.178585</td>
<td>7.416841</td>
<td>6.577137</td>
<td>3.571591</td>
</tr>
<tr>
<td>A0983</td>
<td>3.465027</td>
<td>2.681155</td>
<td>10.09176</td>
<td>8.769516</td>
<td>2.321535</td>
</tr>
<tr>
<td>A0984</td>
<td>2.724821</td>
<td>3.425926</td>
<td>9.727004</td>
<td>8.769516</td>
<td>1.250058</td>
</tr>
<tr>
<td>A0985</td>
<td>1.782513</td>
<td>1.936389</td>
<td>9.846591</td>
<td>8.175955</td>
<td>0.992898</td>
</tr>
<tr>
<td>A0986</td>
<td>1.021509</td>
<td>1.498531</td>
<td>9.483928</td>
<td>8.769516</td>
<td>0.535739</td>
</tr>
<tr>
<td>A0987</td>
<td>2.213268</td>
<td>1.936389</td>
<td>9.119866</td>
<td>7.573674</td>
<td>1.785796</td>
</tr>
<tr>
<td>A0988</td>
<td>1.021509</td>
<td>1.787437</td>
<td>8.875891</td>
<td>7.374365</td>
<td>0.714318</td>
</tr>
</tbody>
</table>
```
Soil profile metrics

Levenshtein metric
Sequence or string distance


Weighted soil profile metric

\[
S_{j,k} = \sum_{l=1}^{2} \int \sum_{i=1}^{p} w_{i,j,k,l} e^{-cz} \frac{d}{dz} i=1
\]

s refers to similarity
j and k refer to individuals, there are n of these,
l refers to no of layers up to a maximum of L,
z1 and z2 are the boundary depth of layer 1,
a and c are constants which control the amount of depth (z) weighting,
i refers to the p attributes,
for continuous variables \( q_{i,j,k,l} = 1 - \frac{\vert x_{j,i,l} - x_{k,i,l} \vert}{r_i} \) where the x is an observation and \( r_i \) is the range of the ith attribute,
for qualitative attributes \( q_{i,j,k,l} = 1 \) if \( x_{j,i,l} = x_{k,i,l} \) and zero otherwise,
the weights \( w_{i,j,k,l} \) is set to 1 when a comparison can be made for the ith attribute and to 0 if \( x_{j,i,l} \) or \( x_{k,i,l} \) are both unknown or inapplicable.

Finally the distance is given by \( \Delta_{j,k} = \sqrt{2(1-S_{j,k})} \)
Association of attributes - linear correlation matrix

The association between attributes can be expressed as a symmetric p by p matrix \( R \) with elements

\[
R_{h,i} = \frac{1}{(n-1)} \sum_{j=1}^{n} (x_{j,h} - \bar{x}_h)(x_{j,i} - \bar{x}_i)
\]

\[
\sqrt{ \frac{1}{(n-1)} \sum_{j=1}^{n} (x_{j,h} - \bar{x}_h)^2 \cdot \frac{1}{(n-1)} \sum_{j=1}^{n} (x_{j,i} - \bar{x}_i)^2 }^{1/2}
\]

where the \( h \) and \( i \) refer to attributes, the \( j \) refer to the \( n \) individuals and the \( x \) are the observations.

This can also be written as

\[
R_{h,i} = \frac{\text{covariance}(h,i)}{\text{square root(variance}(h) \times \text{variance}(i))}
\]
```
> print(var(exdat), collab=excolname, rowlab=excolname)

Array:
5 by 5

<table>
<thead>
<tr>
<th></th>
<th>CEC</th>
<th>%Cl</th>
<th>pH1</th>
<th>pH5</th>
<th>%OC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEC</td>
<td>34.50000</td>
<td>32.50000</td>
<td>0.9214282</td>
<td>1.532142</td>
<td>2.450000</td>
</tr>
<tr>
<td>%Cl</td>
<td>32.50000</td>
<td>45.07143</td>
<td>-0.8928579</td>
<td>0.0178569</td>
<td>2.821428</td>
</tr>
<tr>
<td>pH1</td>
<td>0.921428</td>
<td>-0.8928579</td>
<td>0.6764287</td>
<td>0.7725001</td>
<td>-0.2192058</td>
</tr>
<tr>
<td>pH5</td>
<td>1.532143</td>
<td>0.0178569</td>
<td>0.7725001</td>
<td>1.000964</td>
<td>-0.2216715</td>
</tr>
<tr>
<td>%OC</td>
<td>2.450000</td>
<td>2.821428</td>
<td>-0.2192058</td>
<td>-0.2216715</td>
<td>0.3135714</td>
</tr>
</tbody>
</table>

> * above is the variance-covariance matrix

> print(cor(exdat), collab=excolname, rowlab=excolname)

Array:
5 by 5

<table>
<thead>
<tr>
<th></th>
<th>CEC</th>
<th>%Cl</th>
<th>pH1</th>
<th>pH5</th>
<th>%OC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEC</td>
<td>1.00000</td>
<td>0.8041820</td>
<td>0.1907395</td>
<td>0.2599457</td>
<td>0.744883</td>
</tr>
<tr>
<td>%Cl</td>
<td>0.8041820</td>
<td>1.000000</td>
<td>-0.1617039</td>
<td>0.00255626</td>
<td>0.750499</td>
</tr>
<tr>
<td>pH1</td>
<td>0.1907395</td>
<td>-0.1617039</td>
<td>0.9999999</td>
<td>0.9999999</td>
<td>-0.476136</td>
</tr>
<tr>
<td>pH5</td>
<td>0.2599457</td>
<td>0.00255626</td>
<td>0.9999999</td>
<td>0.9999999</td>
<td>-0.393421</td>
</tr>
<tr>
<td>%OC</td>
<td>0.744883</td>
<td>0.750499</td>
<td>-0.476136</td>
<td>-0.393421</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

> * above is the correlation matrix
```
Ordination methods

Dimension reducing techniques which allow us to look at the essentials of the multivariate scatter of individuals in a few dimensions (e.g. 2).

1) Linear dimension reducers (projections)
   - principal components analysis
   - correspondence analysis
   - principal factor analysis
   - projection pursuit

2) Nonlinear dimension reducers (global)
   - non-linear principal components analysis

3) Nonlinear representations only defined for the points given (usually start with distance matrix)
   - metric multidimensional scaling (principal coordinate analysis or classical scaling)
   - multivariate lining and planing (based on minimal spanning tree)


Classical multi-dimensional scaling on Manhattan distance
Classical multi-dimensional scaling on Euclidean distance
Minimal spanning tree and multivariate planning
Biplot

\[ X_2 = G H' \]

\( X_2 \) is the best rank two approximation of the \( n \) by \( p \) data matrix \( X \), \( G \) is \( n \) by 2, \( H' \) is 2 by \( p \).

Minimal spanning tree and multivariate planning
NUMERICAL CLASSIFICATION

Hierarchical methods
(classical numerical taxonomy) work on similarity or distance matrix.

Single linkage
Average linkage

Non-hierarchical methods
(more statistical in nature) work on data matrix.

**hard c (k) means algorithm**

\[
\begin{align*}
    & \text{minimise } J(v) = \sum_{k=1}^{n} \sum_{i=1}^{c} (x_k - v_i)^\prime A (x_k - v_i) \\
\end{align*}
\]

**fuzzy c means algorithm**

\[
\begin{align*}
    & \text{minimise } J(U,v) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{i,k})^m (x_k - v_i)^\prime A (x_k - v_i) \\
\end{align*}
\]
Handbook Surface Data Single linkage Euclidean metric
Handbook Surface Data Average Clustering Euclidean metric
Conventional soil classification -
discrete classes usually based on
sharply defined cut-offs.
Reality is more continuous.

**FUZZY SETS**

encompass continuity by describing
*degree of belongingness* to a class.

**FUZZY \(c\) means**
given \(n\) soil individuals and \(p\) soil properties
finds \(c\) centroids described in
terms of the \(p\) properties and
membership of each individual in
each of the \(c\) classes.
Location of the 70 Australian stations in attribute space showing centroic crosses and membership isopleths for the 4-group solution. Station names are given in Table I.
<table>
<thead>
<tr>
<th>Station number/Name</th>
<th>Group A (hot, wet)</th>
<th>Group B (hot, dry)</th>
<th>Group C (warm, dry)</th>
<th>Group D (cool, moist)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Bourke</td>
<td>0.007</td>
<td>0.927</td>
<td>0.044</td>
<td>0.022</td>
</tr>
<tr>
<td>2 Canberra</td>
<td>0.024</td>
<td>0.121</td>
<td>0.041</td>
<td>0.815</td>
</tr>
<tr>
<td>3 Dubbo</td>
<td>0.032</td>
<td>0.482</td>
<td>0.090</td>
<td>0.396</td>
</tr>
<tr>
<td>4 Hay</td>
<td>0.025</td>
<td>0.621</td>
<td>0.076</td>
<td>0.277</td>
</tr>
<tr>
<td>5 Port Macquarie</td>
<td>0.513</td>
<td>0.147</td>
<td>0.129</td>
<td>0.401</td>
</tr>
<tr>
<td>6 Sydney</td>
<td>0.123</td>
<td>0.138</td>
<td>0.103</td>
<td>0.637</td>
</tr>
<tr>
<td>7 Walgett</td>
<td>0.017</td>
<td>0.831</td>
<td>0.090</td>
<td>0.073</td>
</tr>
<tr>
<td>8 Wilcannia</td>
<td>0.003</td>
<td>0.970</td>
<td>0.013</td>
<td>0.014</td>
</tr>
<tr>
<td>9 Alice Springs</td>
<td>0.004</td>
<td>0.955</td>
<td>0.027</td>
<td>0.014</td>
</tr>
<tr>
<td>10 Charlotte Waters</td>
<td>0.014</td>
<td>0.843</td>
<td>0.103</td>
<td>0.040</td>
</tr>
<tr>
<td>11 Daly Waters</td>
<td>0.045</td>
<td>0.056</td>
<td>0.874</td>
<td>0.026</td>
</tr>
<tr>
<td>12 Katherine</td>
<td>0.214</td>
<td>0.113</td>
<td>0.603</td>
<td>0.070</td>
</tr>
<tr>
<td>13 Port Darwin</td>
<td>0.909</td>
<td>0.022</td>
<td>0.047</td>
<td>0.022</td>
</tr>
<tr>
<td>14 Tennant’s Creek</td>
<td>0.011</td>
<td>0.053</td>
<td>0.923</td>
<td>0.013</td>
</tr>
<tr>
<td>15 Wave Hill</td>
<td>0.015</td>
<td>0.033</td>
<td>0.949</td>
<td>0.012</td>
</tr>
<tr>
<td>16 Bouris</td>
<td>0.027</td>
<td>0.192</td>
<td>0.742</td>
<td>0.038</td>
</tr>
<tr>
<td>17 Brisbane</td>
<td>0.272</td>
<td>0.189</td>
<td>0.216</td>
<td>0.320</td>
</tr>
<tr>
<td>18 Burketown</td>
<td>0.048</td>
<td>0.060</td>
<td>0.863</td>
<td>0.028</td>
</tr>
<tr>
<td>19 Cairns</td>
<td>0.717</td>
<td>0.076</td>
<td>0.103</td>
<td>0.103</td>
</tr>
<tr>
<td>20 Charleville</td>
<td>0.031</td>
<td>0.648</td>
<td>0.230</td>
<td>0.091</td>
</tr>
<tr>
<td>21 Cloncurry</td>
<td>0.002</td>
<td>0.005</td>
<td>0.992</td>
<td>0.002</td>
</tr>
<tr>
<td>22 Cooktown</td>
<td>0.956</td>
<td>0.011</td>
<td>0.018</td>
<td>0.013</td>
</tr>
<tr>
<td>23 Longreach</td>
<td>0.026</td>
<td>0.220</td>
<td>0.713</td>
<td>0.041</td>
</tr>
<tr>
<td>24 Mackay</td>
<td>0.777</td>
<td>0.057</td>
<td>0.076</td>
<td>0.088</td>
</tr>
<tr>
<td>25 Mapoon</td>
<td>0.935</td>
<td>0.016</td>
<td>0.032</td>
<td>0.015</td>
</tr>
<tr>
<td>26 Normanton</td>
<td>0.287</td>
<td>0.115</td>
<td>0.523</td>
<td>0.075</td>
</tr>
<tr>
<td>27 Rockhampton</td>
<td>0.255</td>
<td>0.191</td>
<td>0.352</td>
<td>0.172</td>
</tr>
<tr>
<td>28 Thargomindah</td>
<td>0.016</td>
<td>0.771</td>
<td>0.167</td>
<td>0.044</td>
</tr>
<tr>
<td>29 Thurunday Island</td>
<td>0.954</td>
<td>0.012</td>
<td>0.021</td>
<td>0.013</td>
</tr>
<tr>
<td>30 Toowoomba</td>
<td>0.028</td>
<td>0.072</td>
<td>0.036</td>
<td>0.563</td>
</tr>
<tr>
<td>31 Townsville</td>
<td>0.530</td>
<td>0.109</td>
<td>0.258</td>
<td>0.103</td>
</tr>
<tr>
<td>32 Windorah</td>
<td>0.029</td>
<td>0.478</td>
<td>0.427</td>
<td>0.056</td>
</tr>
<tr>
<td>33 Adelaide</td>
<td>0.030</td>
<td>0.425</td>
<td>0.079</td>
<td>0.466</td>
</tr>
<tr>
<td>34 Cape Borda</td>
<td>0.013</td>
<td>0.069</td>
<td>0.022</td>
<td>0.896</td>
</tr>
<tr>
<td>35 Cape Northumberland</td>
<td>0.011</td>
<td>0.050</td>
<td>0.016</td>
<td>0.921</td>
</tr>
<tr>
<td>36 Cook</td>
<td>0.027</td>
<td>0.713</td>
<td>0.053</td>
<td>0.177</td>
</tr>
<tr>
<td>37 Ermabellia</td>
<td>0.003</td>
<td>0.971</td>
<td>0.012</td>
<td>0.014</td>
</tr>
<tr>
<td>38 Farina</td>
<td>0.006</td>
<td>0.937</td>
<td>0.035</td>
<td>0.022</td>
</tr>
<tr>
<td>39 Port Augusta</td>
<td>0.003</td>
<td>0.569</td>
<td>0.013</td>
<td>0.015</td>
</tr>
<tr>
<td>40 Streaky Bay</td>
<td>0.023</td>
<td>0.679</td>
<td>0.071</td>
<td>0.227</td>
</tr>
<tr>
<td>41 Tarcoola</td>
<td>0.006</td>
<td>0.940</td>
<td>0.027</td>
<td>0.027</td>
</tr>
<tr>
<td>42 William Creek</td>
<td>0.010</td>
<td>0.894</td>
<td>0.064</td>
<td>0.032</td>
</tr>
<tr>
<td>43 Currie</td>
<td>0.014</td>
<td>0.041</td>
<td>0.016</td>
<td>0.927</td>
</tr>
<tr>
<td>44 Hobart</td>
<td>0.032</td>
<td>0.134</td>
<td>0.051</td>
<td>0.783</td>
</tr>
<tr>
<td>45 Launceston</td>
<td>0.021</td>
<td>0.077</td>
<td>0.031</td>
<td>0.871</td>
</tr>
<tr>
<td>46 Stanley</td>
<td>0.024</td>
<td>0.062</td>
<td>0.030</td>
<td>0.854</td>
</tr>
<tr>
<td>47 Gabo Island</td>
<td>0.010</td>
<td>0.025</td>
<td>0.013</td>
<td>0.952</td>
</tr>
<tr>
<td>48 Melbourne</td>
<td>0.011</td>
<td>0.055</td>
<td>0.018</td>
<td>0.916</td>
</tr>
<tr>
<td>49 Albany</td>
<td>0.026</td>
<td>0.053</td>
<td>0.029</td>
<td>0.892</td>
</tr>
<tr>
<td>50 Broome</td>
<td>0.022</td>
<td>0.036</td>
<td>0.926</td>
<td>0.015</td>
</tr>
<tr>
<td>51 Cape Leewin</td>
<td>0.040</td>
<td>0.085</td>
<td>0.048</td>
<td>0.827</td>
</tr>
<tr>
<td>52 Carnarvon</td>
<td>0.017</td>
<td>0.779</td>
<td>0.160</td>
<td>0.044</td>
</tr>
<tr>
<td>53 Condon</td>
<td>0.023</td>
<td>0.150</td>
<td>0.796</td>
<td>0.031</td>
</tr>
<tr>
<td>54 Eclipse Island</td>
<td>0.003</td>
<td>0.010</td>
<td>0.004</td>
<td>0.383</td>
</tr>
<tr>
<td>55 Esperance</td>
<td>0.014</td>
<td>0.090</td>
<td>0.028</td>
<td>0.688</td>
</tr>
<tr>
<td>56 Eucla</td>
<td>0.013</td>
<td>0.852</td>
<td>0.048</td>
<td>0.087</td>
</tr>
<tr>
<td>57 Eyre</td>
<td>0.025</td>
<td>0.667</td>
<td>0.076</td>
<td>0.232</td>
</tr>
<tr>
<td>58 Halls Creek</td>
<td>0.001</td>
<td>0.002</td>
<td>0.996</td>
<td>0.001</td>
</tr>
<tr>
<td>59 Kalgoorlie</td>
<td>0.003</td>
<td>0.970</td>
<td>0.013</td>
<td>0.014</td>
</tr>
<tr>
<td>60 Laverton</td>
<td>0.001</td>
<td>0.586</td>
<td>0.008</td>
<td>0.005</td>
</tr>
<tr>
<td>61 Mundhwindi</td>
<td>0.029</td>
<td>0.505</td>
<td>0.409</td>
<td>0.057</td>
</tr>
<tr>
<td>62 Nullagine</td>
<td>0.019</td>
<td>0.118</td>
<td>0.838</td>
<td>0.025</td>
</tr>
<tr>
<td>63 Onslow</td>
<td>0.030</td>
<td>0.218</td>
<td>0.709</td>
<td>0.043</td>
</tr>
<tr>
<td>64 Peak Hill</td>
<td>0.024</td>
<td>0.644</td>
<td>0.278</td>
<td>0.054</td>
</tr>
<tr>
<td>65 Perth</td>
<td>0.052</td>
<td>0.155</td>
<td>0.081</td>
<td>0.712</td>
</tr>
<tr>
<td>66 Port George IV</td>
<td>0.748</td>
<td>0.057</td>
<td>0.144</td>
<td>0.051</td>
</tr>
<tr>
<td>67 Rawlinna</td>
<td>0.015</td>
<td>0.844</td>
<td>0.056</td>
<td>0.085</td>
</tr>
<tr>
<td>68 Warburton Range</td>
<td>0.010</td>
<td>0.879</td>
<td>0.081</td>
<td>0.030</td>
</tr>
<tr>
<td>69 Wyndham</td>
<td>0.118</td>
<td>0.107</td>
<td>0.721</td>
<td>0.054</td>
</tr>
<tr>
<td>70 Yalgoo</td>
<td>0.004</td>
<td>0.955</td>
<td>0.027</td>
<td>0.013</td>
</tr>
</tbody>
</table>
Lawes transect pH centroids ($m = 1.5$)
Lawes transect pH centroids (c = 4, m = 1.5)
SOIL CLASSIFICATION USING
FUZZY SETS
- TWO APPROACHES

1) a) CLASSIFY ALL RECOGNISED HORIZONS.

b) CLASSIFY PROFILES DESCRIBED IN TERMS OF SEQUENCE OF HORIZONS.
   - potentially more useful for pedological and stratigraphic studies

2) a) CLASSIFY DEPTH FUNCTIONS OF SUITES OF SOIL PROPERTIES e.g. texture and related, chemical etc.

b) CLASSIFY PROFILES DESCRIBED IN TERMS OF A SERIES OF DEPTH FUNCTION CLASSES
   - potentially more useful for agronomic studies

MORE FLEXIBILITY
Allocation of unknown soil entities to an existing classification system

(Also known as identification, placement, assignment, pattern recognition but not classification)

Discriminant analysis

good when,

a) plenty of individuals
b) not too many attributes
c) few groups
d) plenty of computing time

Diagnostic keys

expert systems?

Centroid radius method

\[ u_i = \frac{1}{c} \left( \sum_{j=1}^{c} \frac{d_i}{d_j} \right)^{2/(m-1)} \]

> discrexdat

\$cor
     9.99978e-1  9.22297e-1  7.37827e-7
\$vars:
Array:
  3 by 3

[1,]  51.19246  -1.852363 -0.8459430 -0.5488213 -0.1331055
[2,]  27.32754   0.4184926  0.3519876  0.1466383  0.2803361
[3,] 135.98911   8.087022  7.051726  0.1955801  0.2711332
[4,]  27.50194  -1.955857 -2.3558922  2.140227  1.053181

\$groups:
Array:
  3 by 3

   [,1]     [,2]     [,3]
[1,] 0.9144675   0.1968476 -0.3535536
[2,] 0.00696594  -0.8659883  -0.4999996
[3,] 0.40460885   0.4596717   0.7908596

> exdat

Array:
  8 by 5

[1,]  5   9   6.3   6.3   0.8
[2,]  18  28   6.1   6.6   2.0
[3,]  20  18   8.3   8.8   1.3
[4,]  16  23   8.0   8.8   0.7
[5,]  10  13   8.1   8.2   0.5
[6,]   6  10   7.8   8.8   0.3
[7,]  13  13   7.5   7.6   1.8
[8,]   6  12   7.3   7.4   0.4

> newdat <- c(17,25,6,6,1.9)    # new unknown
> newdat * discrexdat * vars
     1743.236  42.65552  35.19568  15.98468  8.98784

>
Introduction to Pedometrics

LECTURE 4

SOIL GEOSTATISTICS
4 SOIL GEOSTATISTICS

Assumptions – stationarity and isotropy

The semi-variogram

Kriging

Non-stationary methods

Non-linear methods

Clark, I. 1979. Practical geostatistics. Pion?


SPATIAL PREDICTION OF SOIL PROPERTIES
(essence of soil survey)

GEOSTATISTICS

Principal method - K R I G I N G

provides:

• estimate at unmeasured location (point or block)

/depends on (nearby) data values and semi-varioogram/

• standard error of that estimate

/depends on semi-varioogram only/
Assumptions - stationarity and isotropy

Intrinsic stationarity

\[ E[z(x)] = \mu \]
\[ \text{var}[z(x) - z(x+h)] = \text{E}[(z(x)-z(x+h))^2] = 2 \gamma(h) \]

A slightly more general model

\[ z(x) = \mu_R + \epsilon(x) \]
\[ E[z(x)] = \mu_R \]
\[ \text{var}[\epsilon(x) - \epsilon(x+h)] = \text{E}[(\epsilon(x)-\epsilon(x+h))^2] = 2 \gamma(h) \]

Isotropy

\[ \gamma(h) = \gamma(x) = \gamma(y) \]

where \( h, x \) and \( y \) are vectors in different directions but of the same length.
Estimation of the semi-variogram

Empirical semi-variogram

1 dimension

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

2 dimensions

Cartesian \( \gamma^*(p,q) \)

Polar \( \gamma^*(h,\theta) \)

Fitting models

only authorised models (CNSD)

use e.g. weighted non-linear least squares

minimise for a particular vector of parameters \( \theta \), \( m \)

\[ \sum w_i \gamma(h_i) - \gamma^*(h_i)^2 \]

where \( w_i = n(h_i) / (\gamma(h_i)^2) \)

N.B. \( n(h_i) \geq 30 \)
Semi-vigrogram models

Pure nugget variance
\[ \gamma(h) = c_0 \delta(h), \quad \delta(h) = 1 \text{ if } h > 0, = 0 \text{ otherwise.} \]

Linear model
\[ \gamma(h) = ah \]

Spherical model
\[ \gamma(h) = \begin{cases} c_1 \{1.5 \left(\frac{h}{a}\right) - 0.5 \left(\frac{h}{a}\right)^3\} & h < a \\ c_1 & h \geq a \end{cases} \]

Exponential model
\[ \gamma(h) = c_1 \{1 - \exp(-h/r)\} \quad h > 0 \]

Fractional Brownian model
\[ \gamma(h) = kh^\omega, \quad D = 2 - (\omega/2) \]

Anisotropic models
\[ d(\theta, h) = A^2 \cos^2(\theta - \phi) + B^2 \sin^2(\theta - \phi))^{0.5} \]

Nested models
authorised models can be combined
Properties of semi-variograms

Dispersion variance

\[ D^2(v|v) = \bar{\gamma}(v, v) - \gamma(v, v) \quad \text{v in V} \]
\[ D^2(0|v) = \bar{\gamma}(v, v) \]

\[ D^2(v|v) = D^2(v|v) + D^2(v|v) \quad \text{v in V in R} \]

Regularisation

\[ \gamma_v(h) = \gamma(v, v_h) - \bar{\gamma}(v, v) \]
\[ \gamma_v(h) = \gamma(h) - \bar{\gamma}(v, v) \text{ for } h \gg v \]

Estimation variance

Estimation of average of V from equally weighted observations v,

\[ \sigma_E^2 = 2 \bar{\gamma}(V, V) - \bar{\gamma}(V, V) - \bar{\gamma}(v, v) \]

with unequal weights \( \lambda_i \)

\[ \sigma_E^2 = 2 \sum \lambda_i \gamma(x_i, V) - \bar{\gamma}(V, V) - \sum \lambda_i \lambda_j \gamma(x_i - x_j) \]

now minimise \( \sigma_E^2 = E[(z(x) - z*(x))^2] \) with respect to the \( \lambda_i \) subject to \( \Sigma \lambda_i = 1 \) this gives the kriging equations.
\[ \sum_{i=1}^{n} \lambda_i \gamma(x_i, x_j) + \psi = \bar{\gamma}(x_j, B) \quad \text{for } j = 1, 2, \ldots n. \]

\[ A \begin{bmatrix} \lambda \\ \psi \end{bmatrix} = b, \]

where

\[
A = \begin{bmatrix}
\gamma(x_1, x_1) & \gamma(x_1, x_2) & \cdots & \gamma(x_1, x_n) \\
\gamma(x_2, x_1) & \gamma(x_2, x_2) & \cdots & \gamma(x_2, x_n) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(x_n, x_1) & \gamma(x_n, x_2) & \cdots & \gamma(x_n, x_n)
\end{bmatrix},
\]

\[
\begin{bmatrix} \lambda \\ \psi \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \psi \end{bmatrix} \quad \text{and } b = \begin{bmatrix} \gamma(x_1, B) \\ \gamma(x_2, B) \\ \vdots \\ \lambda(x_n, B) \\ 1 \end{bmatrix}.
\]

\[ \sigma^2_k = b^T \begin{bmatrix} \lambda \\ \psi \end{bmatrix} - \bar{\gamma}(B, B). \]
Elements of one-dimensional semi-variogram.

Semi-variogram of exchangeable potassium in topsoil, measured in µg/g and transformed to common logarithms, at Broom’s Barn with linear model fitted (see Webster, 1981).
Semi-variogram of organic carbon in the subsoil, measured as percentage and transformed to reciprocal, at Tillycorthie, Aberdeenshire (see McBratney and Webster, 1981a).

Semi-variogram of available copper in the topsoil of southeast Scotland. The three component semi-variograms of the nested spherical model are shown by the dashed lines (see McBratney et al., 1982).
Semi-variogram of subsoil silt and sand at Woburn. Here and in Figures 18 and 19 the symbols for the semi-variances show the different directions and the oblique lines the enclosing envelope of the fitted model (see McBratney and Webster, 1983).
Semi-variogram of silt in the topsoil and that in the subsoil at Woburn.
Semi-variogram of topsoil silt and subsoil sand at Woburn.

Weights for kriging stone content at a point, P, at Plas Gogerddan.
Isarithmic map of stone content made by punctual kriging at Plas Gogerddan.

Block diagram of stone content estimated by punctual kriging.
Isarithmic map of stone content made by block kriging.
Punctual kriging

\[
\begin{array}{cccc}
\bullet & 0 & 12.89 & 0 \\
0 & 12.89 & 12.89 & 0 \\
12.89 & 12.87 & 12.87 & 12.89 \\
0 & 12.89 & 12.87 & 0 \\
0 & 12.89 & 12.89 & 0 \\
\end{array}
\]

Block kriging

\[
\begin{array}{cccc}
\bullet & 0 & 0 & 0 \\
0 & 0 & 0 & 0.929 \\
0.929 & 0.935 & 0.935 & 0.929 \\
0 & 0 & 0 & 0.935 \\
0.935 & 0.940 & 0.940 & 0.935 \\
0 & 0 & 0 & 0.935 \\
0.935 & 0.940 & 0.940 & 0.953 \\
\bullet & 0 & 0 & 0 \\
0.929 & 0.935 & 0.935 & 0.929 \\
\end{array}
\]

Variances for punctual and block kriging of stone content. Values are in \((\text{percent})^2\). For punctual kriging the block discs are sampling points and open circles the interpolated ones. In the lower figure they represent the centers of the estimated blocks.
Diagram of the block kriging variance.

Graphs of maximum variance for punctual kriging of thickness of cover loam against sample spacing at Hole Farm, Norfolk, on square and triangular grids.
Comparisons for estimating semi-variances on linear transects at lags of 1, 2, and 3 sampling intervals, (a) for complete data and (b) where some observations are missing the open circles.

Sample semi-variograms of stone content in the topsoil at Plas Gogerddan in four principal directions.
Isarithmic representation and sample semi-variogram of stone content at Plas Gogerddan. Border scales are in sampling intervals.
Perspective representation of sample semi-variogram of stone content at Plas Gogerddan viewed from a position above and to the left.

Cylindrical projection of two-dimensional sample semi-variogram of stone content with separate symbols for eight directions. The oblique lines form the envelope of the fitted linear model.
Graphs of maximum variance for kriging cover loam thickness over 40 × 40-m and 100 × 100-m blocks at Hole Farm. The solid lines are for square grids with the blocks centered over the grid nodes, the broken lines are for blocks centered in the centers of the grid squares.

The relation between punctual kriging variance of stone content at Plas Gogerddan and sample spacing in the directions of maximum and minimum
pH in water

Akima + NNI raw + + G. Med. + G. Mean

Av. (4x4)

Splines + An. K. (all) + + Av. (4)

An. K. (4x4) + NNI smoothed

I.K. (all) + I.K. (4x4) +

Inv. Sq.
Non-stationary methods

model

\[ z(x) = m(x) + \epsilon(x) \]

where \( m(x) \) is usually described by a low order polynomial and \( \epsilon(x) \) is described by semi-variogram. Problem \( m(x) \) and \( \epsilon(x) \) are both unknown.

Method based on above is called universal kriging.

Maximum likelihood methods can be used but are computationally expensive.


An alternative method is based on intrinsic random functions of order \( k \). Basic idea is to filter out low order polynomial drifts by generalised differences.
Ordinary kriging is linear in the data

Lognormal kriging.
\[\ln(z^*(x)) = \sum_{i=1}^{n} \lambda_i \ln(z(x_i))\]
\[\lambda_i\text{'s are not linear in the } z(x_i)\]
\[z^*(x) = \exp(\ln(z^*(x)) + \sigma^2/2)\]

Disjunctive kriging.

Provides better (smaller prediction error) than ordinary kriging, also provides conditional probability of a value occurring above a fixed cut-off.
\[z^*(x) = \sum_{i=1}^{n} \sum_{k=0}^{\infty} f_{ik} H_k [z(x_i)]\]

A SUITE OF GEOSTATISTICAL
TECHNIQUES

1) Stationary situations
   (mean more or less constant)
   ORDINARY
   KRIGING

2) Non-stationary situations
   (large and consistent changes
   in the mean over area of
   interest)
   UNIVERSAL
   KRIGING

3) Non linear prediction
   (estimation of proportion
   of area between two
   cut-offs)
   DISJUNCTIVE
   KRIGING

4) Simultaneous prediction of
   several soil properties
   (also undersampling)
   CO-KRIGING

ALL WILL BE USED
Introduction to Pedometrics

LECTURE 5

QUANTITATIVE
SOIL SURVEY
METHODS
5 QUANTITATIVE SOIL SURVEY METHODS

Conventional soil survey

Methods for quantifying conventional soil survey

Geostatistical soil survey

Geostatistical-conventional soil survey

Real-time soil survey
Conventional soil survey

Questions
Unclear

Decision 1
Create intuitive classes

Field-work 1
Reconnaissance sampling

Decision 2
Improved intuitive classes,
incorporation of exotic classification,
creation of mapping legend

Field-work 2
Mapping soil boundaries

Product
Soil map and legend

PROBLEMS
lack of specific questions
no independent estimate of quality
IMPROVED METHODS OF CONVENTIONAL SOIL SURVEY

1) Specific questions (helps optimisation)

2) Improved classification – numerical classification both a reconnaissance and main mapping stage.

3) Improved sampling design at all field-work stages

Reconnaissance – transects to find distribution of boundary spacings and transition probabilities.

Main stage – systematic random sampling to obtain good cover or biassed sampling to obtain low risks of missing boundaries (may also be a grid)

Post mapping – random transects to obtain quality estimates


RANDOM TRANSECTS: HOW?

Two-stage sampling with strict control of probabilities of selection.

Stage 1: selection of delineations.

Stage 2: selection of pairs of perpendicular transects in delineations.

(from de Snijder pers. comm.)
RANDOM TRANSECTS: STATISTICAL ANALYSIS

Point data

average

Transect means

average

Delineation means

average

Mean of mapping unit

standard deviation

Standard error of mean

(from de Gruijter et al., 2009)
Map showing sites sampled for copper. Soil from almost all of these sites were analysed for cobalt as well as a few additional sites. The principal towns are indicated on this and subsequent maps by their capital letters.

C Coldstream  J Jedburgh
D Duns        K Kelso
E Eyemouth   M Melrose
G Galashiels  P Peebles
H Hawick      S Selkirk
I Innerleithen
Map showing the extent of the region sampled and the more extensive soil associations recognised within it.
Map of cobalt in the soil estimated over 1 km$^2$ blocks and contoured from figure field with 400m interval.
Map of kriging variance for cobalt associated with Figure
Geostatistical Soil Survey

Question  e.g. What is the value of soil attribute \( z \) at any location or block \( z \) within the area of interest?

Assumptions  Assume intrinsic stationarity or some other form of stationarity appropriate to the type of spatial prediction (kriging) to be used.

Decision  Choose the geometric support, the maximum prediction error allowed and the block size.

Field-work 1  Find the scale of variation and possibly reconnaissance semi-variogram using a staggered, unbalanced nested design. Some supports should be contiguous.

Estimates 1  Estimate semi-variograms from nested design. Estimate nugget variance from contiguous support.

Estimates 2  Find grid spacing or spacings appropriate to the maximum prediction error.

Field-work 2  Sample grid plus some random points for corroboration of prediction.

Estimates 3  Re-estimate semi-variograms. Krig figure field and produce contour maps of
estimates and their kriging errors. Check quality of predictions with extra points.

Discontinuities and co-kriging will require a slightly different strategy.

**COMPARISON OF CONVENTIONAL AND GEOSTATISTICAL SOIL SURVEYS**

<table>
<thead>
<tr>
<th></th>
<th>CONVENTIONAL</th>
<th>GEOSTATISTICAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SCALE</strong></td>
<td>1000km-1km small-medium</td>
<td>100km-10m medium-large</td>
</tr>
<tr>
<td><strong>PREDICTION/INFORMATION TRANSFER</strong></td>
<td>soil class</td>
<td>soil semi-variogram or cross semi-variogram</td>
</tr>
<tr>
<td><strong>ASSUMED CONTINUITY</strong></td>
<td>discontinuous</td>
<td>continuous</td>
</tr>
<tr>
<td><strong>NUMBER OF PROPERTIES</strong></td>
<td>many</td>
<td>a few</td>
</tr>
<tr>
<td><strong>PRECISION</strong></td>
<td>unknown or low</td>
<td>medium-high</td>
</tr>
<tr>
<td><strong>COST</strong></td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
INTRODUCTION TO PEDOMETRICS

Tutorial

The three attached letters A, B and C appeared recently in an Australian soil science periodical. Given the three letters and using material contained in the reference list (and your own knowledge of soil science), prepare a fourth letter of not more than 500 words presenting your view of the role of quantitative statistical and numerical methods in soil survey.

The essays will be read and discussed at a tutorial to be held during the week beginning October 20.

References


made of the soil situation by relating profile features and the photo
tone to landform, vegetation, drainage and parent material.

Soil surveying became more efficient and much more meaningful,
and the surveyor could generalise to higher and higher levels. In
all but the most broad-scale studies it came to be the accepted
practice that "no aerial photos meant no soil survey!", but, even in
these, aerial photomosaics, Landsat images etc. depicted patterns
which had some meaning on the ground.

The practice of locating the site for examination with grid
mapping is the reverse of that employed in "free survey". Instead of
selecting the site on the basis of surface characteristics, aerial photo
pattern, and accumulated experience then plotting it on the photo-
graph, the point to be examined is pre-determined by the grid.
There remains the problem of identifying the site in the field
exactly.

Some of the experience obtained recently by a pedologist,
engaged in "grid-cell" mapping, is not encouraging. He
encountered at least two problems:

i. there were difficulties in locating the specified site precisely in
heavy wooded country on steep slopes (the decision had to be
right or the results could be suspect);

ii. there was duplication of the records in areas where there was no
sensible soil change, as on lower slopes where the same soil
could extend over several elements of the grid.

With the fixed grid, the problems of dealing effectively with
complexes are not completely resolved, e.g. those associated with
gilgai or with mixed sedimentary systems. With the free survey
system one continually increases one's understanding, improves
one's judgement and has a heightened perception of the signifi-
cance of what is being observed. It is mandatory to describe each
profile thoroughly in the field, to group it with its fellows to the best
of one's ability and to identify sensible boundaries which can be
extended later by the use of the stereoscope.

Our accumulated experience in free survey, viz. in pattern recog-
nition and interpretation, has been duplicated in countless other
soil survey organizations around the world. Most of these would
not have the time, nor the money, to indulge in computer-based
mapping exercises over broad or small areas, yet they are still
making intelligent assessments which can be interpreted by trained
people in other areas.

The reason that I am concerned at this apparent trend is that our
thinking about soils in the field could become circumscribed by the
limited experience of soils of those who espouse the computer-
oriented approach. In time they may come to appreciate that the
human mind can recognise and sort forms and can analyse patterns
representing landscapes, by integrating stereoscope data with
contour and geological maps. There are many things we could do
to advance our field pedological skills but following the computer
pathway, as would appear to be implied, could put us into a loop
which will re-explore routes we have already grown past. Should
we spend our time "re-inventing the wheel"?

Perhaps we could test the potential usefulness of the method by
selecting two similar sites of about the same area, say 500 to
1000 ha, and simultaneously carry out surveys of each of the two
types. This would enable us to undertake a thorough "cost/
benefit" analysis of each method to see which is the more efficient
in terms of all the inputs, in the field and subsequently in the office,
including amortised cost of equipment, such as computers,
stereoscopes, plotters, and photos, but leaving out skills of
operators which would be hard to quantify. We might then see not
only which is more efficient in terms of time of producing a
meaningful map, but which has the greater potential for
extrapolation to other areas and which is more useful in relating
field data on soils and distribution patterns to other features of the
environments, to agricultural experience and problems.

Computers doubtless have much to offer in terms of speed and
capacity for data handling and processing, but we should not
encourage them whole heartedly to the exclusion of continual
thinking and imaginative correlations, without first attempting to
assess whether the "new" techniques really have something to offer.
In 1985 we have learnt something about patterns of soil
distribution, on which we can build and which we can continually
test. We should find the best way to advance, without accepting
what seems to be a facile approach.

I would welcome any comments.

—

"Soil Survey and Technology"

From some things I have been hearing lately about future trends
in soil survey techniques, I feel there is some cause for concern
about the outcome. From what I can gather it seems that we will
(may) be reverting to grid mapping, because the younger gener-
ation of pedologists as they are currently being trained are familiar
with computing and sample techniques but unfamiliar with pattern
recognition as displayed by aerial photographs, Landsat imagery,
etc. and have perforce to work within the confines of what
computers can deal with.

In the early days of soil surveying, about 50 years ago, soil
surveyors on the plains of the Riverina had to utilise a grid system
because there were no aerial photographs available to them and
they had to have some control over the location of the hole.
Recognising and plotting boundaries depended on describing the
hole adequately, referring to earlier descriptions and interpolating
between holes, as well as on surface features.

With the advent of aerial photographs, even the classical black-
and-whites, the whole practice of soil surveying took quantum
leaps forward. Patterns were revealed about which sensible things
could be said, when sites were judiciously selected for examination
and surface features were intelligently correlated with the shadings
(or tonings) on the photographs.

Early generalisations of the dominant soil within any pattern
(unit) could be readily checked by repeated examinations within
the pattern at other occurrences of it. Interpretations could be
SOIL SURVEY: ANOTHER OPINION

Recent letters to the Editor, A and B, have raised the issue of computers in soil survey and appear to have set the views of the older generation of soil scientists against those of the 'Young Turks'. I find this particularly unfortunate because, at the age of 41, I am unsure where my loyalties should lie!

But the letters express more than differences in age, training, and experience of independent professionals. They present contrasting attitudes to soil survey in particular and to science in general. This might be expected, for, as Sir Peter Medawar has written ('Limits of Science', 1985, Oxford University Press, p. 11), 'Science itself is various enough to satisfy all temperaments. Among scientists are collectors, illustrators, and compulsive tiders-up; many are detectives by temperament and many are explorers; some are artists and others artisans. There are post-scientists and philosophers-scientists and even a few mystics . . .'.

I see in B's letter a plea from a philosopher-scientist for soil survey to use their most powerful tool, the human mind, to recognize, analyze, classify, integrate, and interpret soils data. I see also a concern expressed that these skills may not be fully utilized in surveys in which soil investigations are restricted to sites located by pre-determined sampling grid that is maintained irrespective of aerial photographs, and landform, bedrock or other site characteristics.

In B's response to some of the points raised by A, I find a criticism of the subjective interpretations upon which traditional soil surveys have relied. I fear that B's requirement for more objective methods of 'correlation analysis used in soil survey' may not follow from recent developments in computing and statistics. In these areas so-called 'objective' analysis, the scientist is still constrained by his prejudices. An appropriate numerical strategy and a suitable computer program must be chosen for the analysis, in simple numerical classification work, with which I am familiar, similarity measures and clustering methods have to be assigned. More complex programs undoubtedly require additional subjective decisions from the investigator.

But we know that very different results can arise from different subjective choices in the numerical analysis of soils information (see, for example, R. J. Coventry and W. T. Williams, 1983 'Evaluation of several strategies of pattern analysis of soil profile data', CSIRO Aust. Div. Soils, Div. Rept. No. 66). Like the soil surveyor, the modern numerical analyst is also constrained by his experience and personal judgement. His results must be considered to be no less subjective than those of the traditional field worker.

In his conclusion B has broadened the discussion to developments over the past 30 years in these aspects of pedology relating to soil survey. It should be recognized that these aspects have evolved, but along lines different from those than can be readily handled by a computing approach. They are nonetheless impressive. Pedological research over the past 30 years has supplied the soil surveyor with much better knowledge of:

- geographic models of a range of concepts varying from regional landscape development to surface processes at particular sites;
- stratigraphic concepts and dating methods;
- weathering processes in soils and rocks;
- soil fabrics and processes, through micromorphology;
- applications of new concepts from soil physics, chemistry, mineralogy, and biology.

In some of these areas Australian pedologists (e.g. Roy Brewer and Bruce Butler) have led the world. Clearly, these aspects of pedology relating to soil survey have not stagnated. They have provided a lively, modern appreciation of soil and landscape systems.

The computer may help the pedologist check on interrelationships of some aspects of the system, and may provide the predictions of soils information suggested by B. But as A stressed, the computer should not specify how we devise a project, nor how we interpret the field situation. These are, and must remain, the province of the pedologists' minds, idiosyncratic though they may be.