Soil profile distance measures and classifications

Application to an Australian soil dataset

D(A;B)
Objectives

ências

To make a quantitative tool for soil profile classification

- Which provides different types of distances according to the soil classification purpose
- Which is able to make supervised classification (principle of soil taxonomies) or unsupervised classification

Application

To test the quantitative tool relatively to

- Soil taxonomy purpose
- Available Water Capacity prediction

Focusing on

- The input content of the classification (soil against terron)
- The number of classes (level of taxonomic detail)
- The classes robustness (sensitivity of the distances)
OSACA principles

User requirements
1. Range of horizon classes
2. Range of solum classes
## OSACA principles

### Soil observations

<table>
<thead>
<tr>
<th>Soil types</th>
<th>Soil observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>D(obs, ref)</td>
<td></td>
</tr>
</tbody>
</table>

### Result table

<table>
<thead>
<tr>
<th>Soil types</th>
<th>dmin</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>REF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.7</td>
<td>0.1</td>
<td>0.3</td>
<td>1.3</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>2.5</td>
<td>1.5</td>
<td>0.1</td>
<td>0.6</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>0.1</td>
<td>0.6</td>
<td>0.1</td>
<td>1.2</td>
<td>0.4</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>0.1</td>
<td>0.8</td>
<td>0.1</td>
<td>1.9</td>
<td>0.2</td>
<td>B</td>
</tr>
<tr>
<td>5</td>
<td>0.1</td>
<td>0.1</td>
<td>1.2</td>
<td>0.0</td>
<td>3.0</td>
<td>A</td>
</tr>
</tbody>
</table>

If user has references, only one iteration
Horizon distances between horizon $i$ and horizon $k$ having $n$ soil variables

**Euclidian distance**

$$D_{ij} = \sqrt{\sum_{k=1}^{n} (v_{ik} - v_{jk})^2}$$

**Manhattan distance**

$$M_{ij} = \sum_{k=1}^{n} |v_{ik} - v_{jk}|$$
OSACA distances

\[ \frac{\sum D}{\sum (\# D)} \]

Solum distances

Pedological distance

Observation

Reference

<table>
<thead>
<tr>
<th>Observation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 cm</td>
<td>0 cm</td>
</tr>
<tr>
<td>30 cm</td>
<td>20 cm</td>
</tr>
<tr>
<td>80 cm</td>
<td>60 cm</td>
</tr>
<tr>
<td>100 cm</td>
<td>120 cm</td>
</tr>
</tbody>
</table>

\[ D(x, A) \]
\[ D(x, B) \]
\[ D(y, B) \]
\[ D(y, C) \]
\[ D(z, C) \]
\[ D(?; C) \]
Solum distances

Utilitarian distance

Observation

Reference

0 cm

30 cm

80 cm

100 cm

0 cm

20 cm

60 cm

120 cm

D(x,A) +

D(x,B) +

D(y,B) +

D(y,C) +

D(z,C) +

D(?,C)

A

B

C

∑ D*e

Max depth

Utilitarian distance

OSACA distances

Carré & Jacobson
OSACA distances

\[ \frac{\sum D}{\sum (#D)} \]
Edgeroi area

- Altitude
- Slope
- Aspect
- Plan curvature
- Profile curvature
- CTI

Landsat ETM 7c
- LS Panchromatic
- NDVI
- Clay Index (5/7)

SPOT 4
- NDVI
- RS 20m

RS 25m

K
U
Th
γ 2m

Soil variables
- AWC
- 9 order names/20 subgroup names

341 soil profiles
Profile description

- Sand (%) / Silt (%) / Clay (%)
- pH$_{H_2O}$
- CaCO$_3$
- EC
- C
- Ca, Mg, Na
- RGB (0-255, Hue, Value, Chroma)

Illuminant C two degrees observer (C 1931)

http://WalkillColor.com

Standardization of soil variables (320 soil profiles)

- Solum depth
### Testing OSACA classifications against soil taxonomy

**1st question:** Do we have enough data to speak about pedogenesis?

<table>
<thead>
<tr>
<th>Pool of soil variables</th>
<th>Texture</th>
<th>pH$_{H_2O}$</th>
<th>CaCO$_3$</th>
<th>EC</th>
<th>C</th>
<th>Ca,Mg,Na</th>
<th>Munsell Colour</th>
<th>Solum depth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Pool of soil variables**
- Texture
- pH$_{H_2O}$
- CaCO$_3$
- EC
- C
- Ca,Mg,Na
- Munsell Colour
- Solum depth

#### Sufficient pool of variables

<table>
<thead>
<tr>
<th>Clay&gt;35% throughout</th>
<th>Rudo</th>
<th>Teno</th>
<th>Vert</th>
<th>Kuro</th>
<th>Sodo</th>
<th>Chromo</th>
<th>Calcaro</th>
<th>Dermo</th>
<th>Kando</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cracks &amp; slickenside</th>
<th>Rudo</th>
<th>Teno</th>
<th>Vert</th>
<th>Kuro</th>
<th>Sodo</th>
<th>Chromo</th>
<th>Calcaro</th>
<th>Dermo</th>
<th>Kando</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Texture contrast</th>
<th>Rudo</th>
<th>Teno</th>
<th>Vert</th>
<th>Kuro</th>
<th>Sodo</th>
<th>Chromo</th>
<th>Calcaro</th>
<th>Dermo</th>
<th>Kando</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sodic subsoil</th>
<th>Rudo</th>
<th>Teno</th>
<th>Vert</th>
<th>Kuro</th>
<th>Sodo</th>
<th>Chromo</th>
<th>Calcaro</th>
<th>Dermo</th>
<th>Kando</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B2 pH&lt;5.5</th>
<th>Rudo</th>
<th>Teno</th>
<th>Vert</th>
<th>Kuro</th>
<th>Sodo</th>
<th>Chromo</th>
<th>Calcaro</th>
<th>Dermo</th>
<th>Kando</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B2 pH&gt;5.5</th>
<th>Rudo</th>
<th>Teno</th>
<th>Vert</th>
<th>Kuro</th>
<th>Sodo</th>
<th>Chromo</th>
<th>Calcaro</th>
<th>Dermo</th>
<th>Kando</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Calcareous</th>
<th>Rudo</th>
<th>Teno</th>
<th>Vert</th>
<th>Kuro</th>
<th>Sodo</th>
<th>Chromo</th>
<th>Calcaro</th>
<th>Dermo</th>
<th>Kando</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lack texture</th>
<th>Rudo</th>
<th>Teno</th>
<th>Vert</th>
<th>Kuro</th>
<th>Sodo</th>
<th>Chromo</th>
<th>Calcaro</th>
<th>Dermo</th>
<th>Kando</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structured B2</th>
<th>Rudo</th>
<th>Teno</th>
<th>Vert</th>
<th>Kuro</th>
<th>Sodo</th>
<th>Chromo</th>
<th>Calcaro</th>
<th>Dermo</th>
<th>Kando</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Massive B2</th>
<th>Rudo</th>
<th>Teno</th>
<th>Vert</th>
<th>Kuro</th>
<th>Sodo</th>
<th>Chromo</th>
<th>Calcaro</th>
<th>Dermo</th>
<th>Kando</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rudiment B</th>
<th>Rudo</th>
<th>Teno</th>
<th>Vert</th>
<th>Kuro</th>
<th>Sodo</th>
<th>Chromo</th>
<th>Calcaro</th>
<th>Dermo</th>
<th>Kando</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Isbell et al. (1997) from Minasny & McBratney (to be published)*
Testing OSACA classifications against soil taxonomy

Testing OSACA results against the 20 taxonomy suborders

18-25 Horizon classes
13 Soil variables
15 Landscape attributes

15-25 soil classes
13 PC Landscape attributes

35-45 soil classes
15-25 terron classes

OSACA run
19 classes
R = 40%
R = 41%

20 classes
R = 30%
R = 31%

20 classes
R = 29%
R = 30%

45 classes
R = 51%
R = 52%

43 classes
R = 49%
R = 51%

45 classes
R = 62%
R = 63%

Pedological distance
18 classes
R = 38%
R = 38%

R²(U) likelihood-ratio

Utilitarian distance
24 classes
R = 33%
R = 34%

R = 48%
R = 49%

Joint distance
19 classes
R = 29%
R = 30%

R = 50%
R = 52%

confusions (d_min + 10%)
Testing OSACA classifications against Available Water Capacity (AWC)

$AWC_{1m}$ predicted by Minasny (2007) from:

- Sand, Org C $\rightarrow$ Bulk Density (BD)
- Sand, Clay, BD $\rightarrow$ Field Capacity (FC) & Permanent Wilting Point (PWP)
- $\sum_{\text{hor}}^{1m} \text{FC} \& \text{PWP}$ $\rightarrow$ AWC

$AWC_i = K + \sum_{t=1}^{C} \alpha_t d_{it} + \varepsilon_i$

Pedological distance

- 19 classes $\quad R^2_{adj}= 77\%$
- 45 classes $\quad R^2_{adj}= 82\%$

Utilitarian distance

- 20 classes $\quad R^2_{adj}= 57\%$
- 43 classes $\quad R^2_{adj}= 71\%$

Joint distance

- 20 classes $\quad R^2_{adj}= 57\%$
- 45 classes $\quad R^2_{adj}= 70\%$

$R^2_{adj}$ prediction

- 18 classes $\quad R^2_{adj}= 72\%$
- 45 classes $\quad R^2_{adj}= 83\%$

- 24 classes $\quad R^2_{adj}= 58\%$
- 43 classes $\quad R^2_{adj}= 69\%$

- 19 classes $\quad R^2_{adj}= 58\%$
- 44 classes $\quad R^2_{adj}= 69\%$
Conclusions

- OSACA is a good tool for transforming soil observations into quantitative classes. As a WebApplication, it is easy to use and as an open source, it can be modified.

- The quantitative soil and terron classes formed by OSACA are significantly correlated with soil taxonomy (if enough soil variables to describe the pedogenesis) and secondary soil variables.

- For dealing with soil taxonomy, the soil classes are better predictors than the terron classes. Joint distance and pedological distances give better correlations between soil classes and Australian soil subgroups.

- For dealing with secondary soil variable, the pedological distance is the best predictor. With low and high number of classes, soil seems to be the best predictor. The terron does not increase so much the prediction.

- The allocation confusion due to close distances changes 1% of the correlations.

- The distances can be afterwards used for mapping purposes and for deriving uncertainties associated to predictions (DSM purposes).
Welcome to the NEW, improved OSACA application! Using the application could hardly be simpler; see the instructions on the right.

Use the form below to indicate where your data is held. Two files are requested:

1. The file containing your soil observations
2. The file containing your soil class descriptions

Enter your observations

File Upload

Observation Data File: Browse...

Remove any previous soil type data? 

Soil Types File: Browse...

Enter your references (if you have some)

Using OSACA - the quick guide

1. Upload your data
2. Select how you want OSACA to process the data
3. Download the Results
4. Repeat as Desired

For a detailed description of how to use OSACA, please consult the accompanying documentation "OSACA - The User Guide", available as a PDF document here

Written by IES, Land Management & Natural Hazards Unit, European Commission, DG JRC, Ispra
Osaca - Process Options

Process Options

The form below allows you to make certain choices about how OSACA is to process your data.

Horizons

Distance metric: Euclidean

Classification using: Clustering ○ Not required ○

Minimum Clusters: 18 (or required number of clusters)

Maximum Clusters: 25

Soil Profiles

Distance metric: Pedological

Classification using: Clustering ○ Not required ○

Minimum Clusters: 20 (or required number of clusters)

Maximum Clusters: 30

Submit your Options

OK

Classification vs. Clustering

Classification is the process of associating an observation with a representative of the class. Here, the association is done on the basis of the distance between the observation and the classes: the observation is classified according to the class to which it is closest.

Clustering is a process very similar to classification, except that the classes are derived from an iterative statistical calculation on the observed data.

Optimal Clustering

There is no reliable way of automatically determining the optimum number of clusters to describe the observed data. A common metric is the ratio D(intra)/D(inter). In other words, the mean distance of clustered observations from the centre of the cluster, divided by the mean distance between clusters. The overall trend is generally towards 0 with increasing numbers of clusters, but there are often local minima. If a range of values is asked for, OSACA will calculate the ratio D(intra)/D(inter) over that range, and choose the minimum.

Warning!

Depending on the size of your datasets, and the options you have chosen, there can be a significant delay before the next page is displayed. This is normal.

For a detailed description of how to use OSACA, please consult the accompanying documentation "OSACA - The User Guide", available as a PDF document here.

Written by IES, Land Management & Natural Hazards Unit, European Commission, DG JRC, Ispra
Osaca - Get Results

Please select the results you require by pressing the appropriate buttons. Each button causes a file to be downloaded. The files are in CSV format, so you can either save them, and then load them into a spreadsheet afterwards, or possibly your browser will automatically open a spreadsheet window with the results in.

- Get Results - Distance from Observed Horizons to Horizon Classes
- Get Results - Distance from Observed Solums to Soil Classes
- Get Results - Distance between Horizon Classes
- Get Results - Distance between Solum Classes
- Get Results - Description of Horizon Classes
- Get Results - Description of Solum Classes
- Get Results - Quality Measure D(intra)/D(inter) for Horizon Clusters
- Get Results - Quality Measure D(intra)/D(inter) for Horizon Clusters as a function of number of clusters
- Get Results - Quality Measure D(intra)/D(inter) for Solum Clusters
- Get Results - Quality Measure D(intra)/D(inter) for Solum Clusters as a function of number of clusters

Please Note

The results shown opposite are those available, given your previous choice of options.

By Horizon or Soil "Classes" is meant either (i) the classes you supplied, and which were used for classifying, or (ii), the centres of the clusters. If you supplied classes, but asked for clustering, the cluster centres are the classes referred to here.

Changing Options

If you wish to try changing the options, please click on Redo.

For a detailed description of how to use OSACA, please consult the accompanying documentation "OSACA - The User Guide", available as a PDF document here.

Get the results (tables)