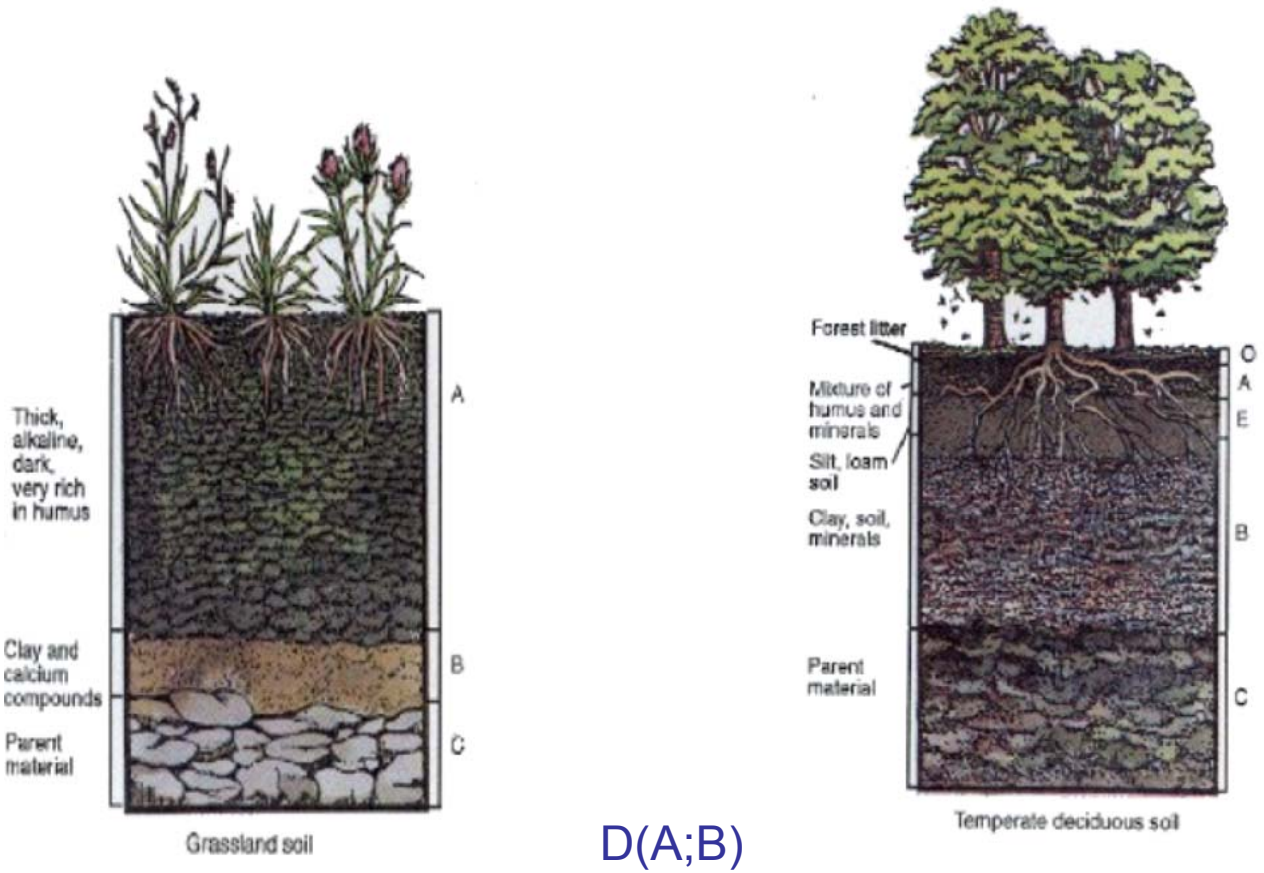
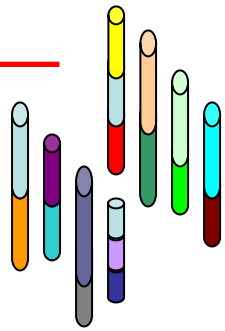
Concepts and applications



OSACA finalities

❶ To classify soil observations into soil types (ST) or clusters (soil classification tool)

Range of ST required



Soil observations

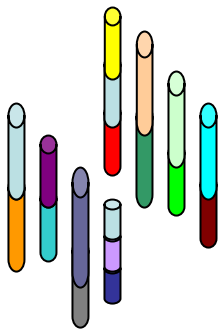
inputs

Soil types

- Description of ST
- Table of distances between observations and ST
- Table of distances between ST
- Quality of the classification

outputs

❷ To allocate soil observations according to soil types provided by the user



Soil observations

\+

User's soil types

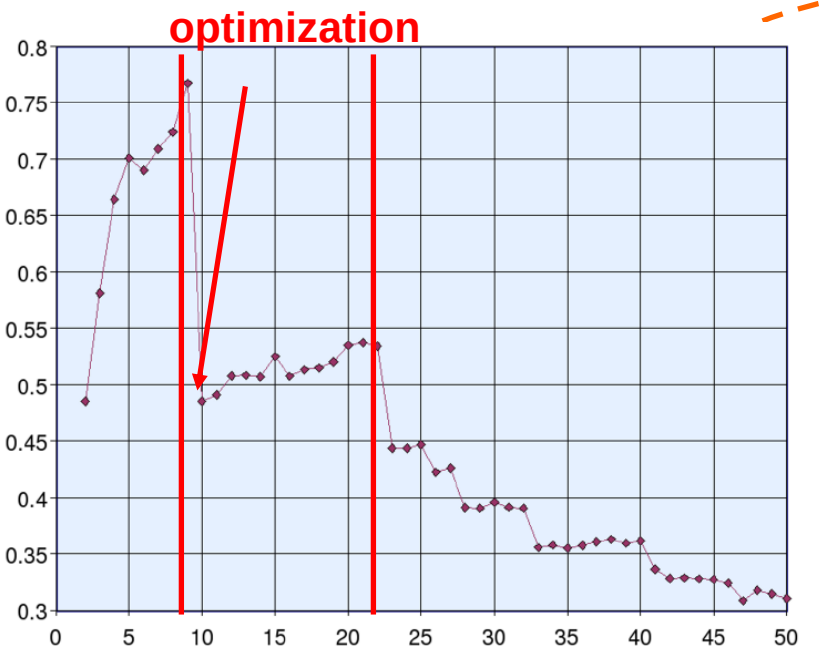
inputs

- Table of distances between observations and ST
- Table of distances between ST

outputs

❸ To classify soil observation horizons into horizon types (id. ❶)

The classification process



❶ Initialisation of cluster centroids

Max. # Soil Type centroids

❷ Observation allocation

Calculation of distances

K-means techniques

❸ Centroid readjustment

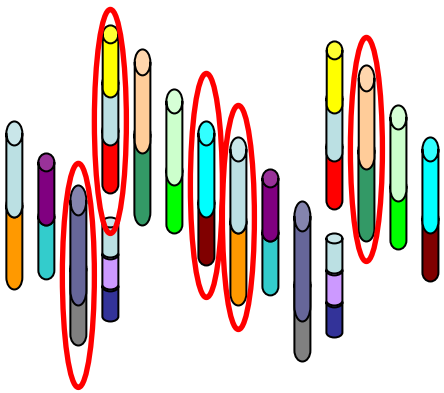
Calculation of new centroids

❹ Optimisation process

D_{intra}/D_{inter} ratio

Initialisation of centroids

❶ *NC* Soil Types or Clusters being required by the user, we fix γ an initial large distance



❷ Pick up randomly 2NC observations

❸ Calculate the distance between the 2NC observations

❹ If distance < γ, choose other observations and go to ❷, otherwise go to ❺

❺ Calculate distance between the whole set of observations and the 2NC observations, and allocate the corresponding observations (those with d_{min})

❻ Sort by decreasing order of population the 2NC clusters

❼ If more than NC clusters are empty, γ = γ/2 and go to ❷, otherwise, choose the NC first ones and consider the corresponding centroids as initial centroids

The classification process

❶ Initialisation of cluster centroids

Max. # Soil Type centroids

❷ Observation allocation

Calculation of distances

K-means techniques

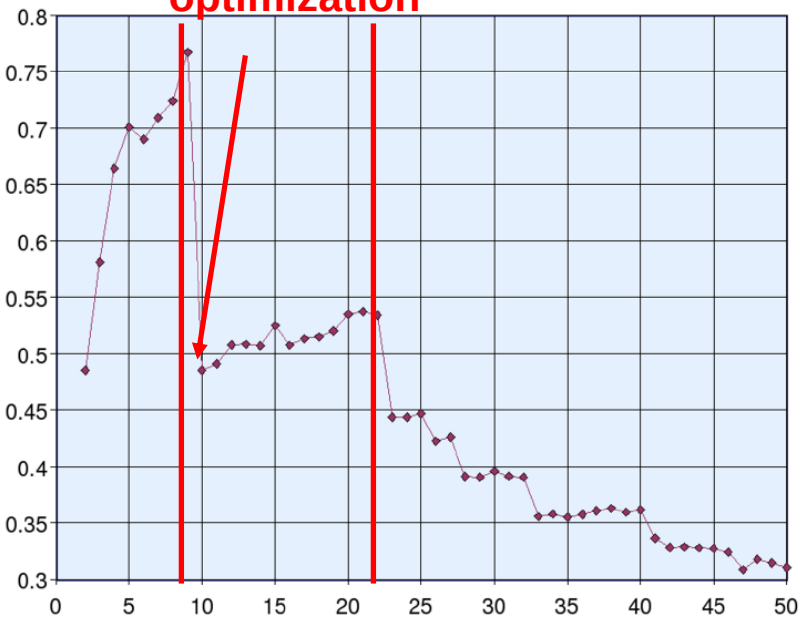
❸ Centroid readjustment

Calculation of new centroids

❹ Optimisation process

D_{intra}/D_{inter} ratio

optimization



Calculation of distances

- OSACA provides 3 kinds of distances useful for different purposes

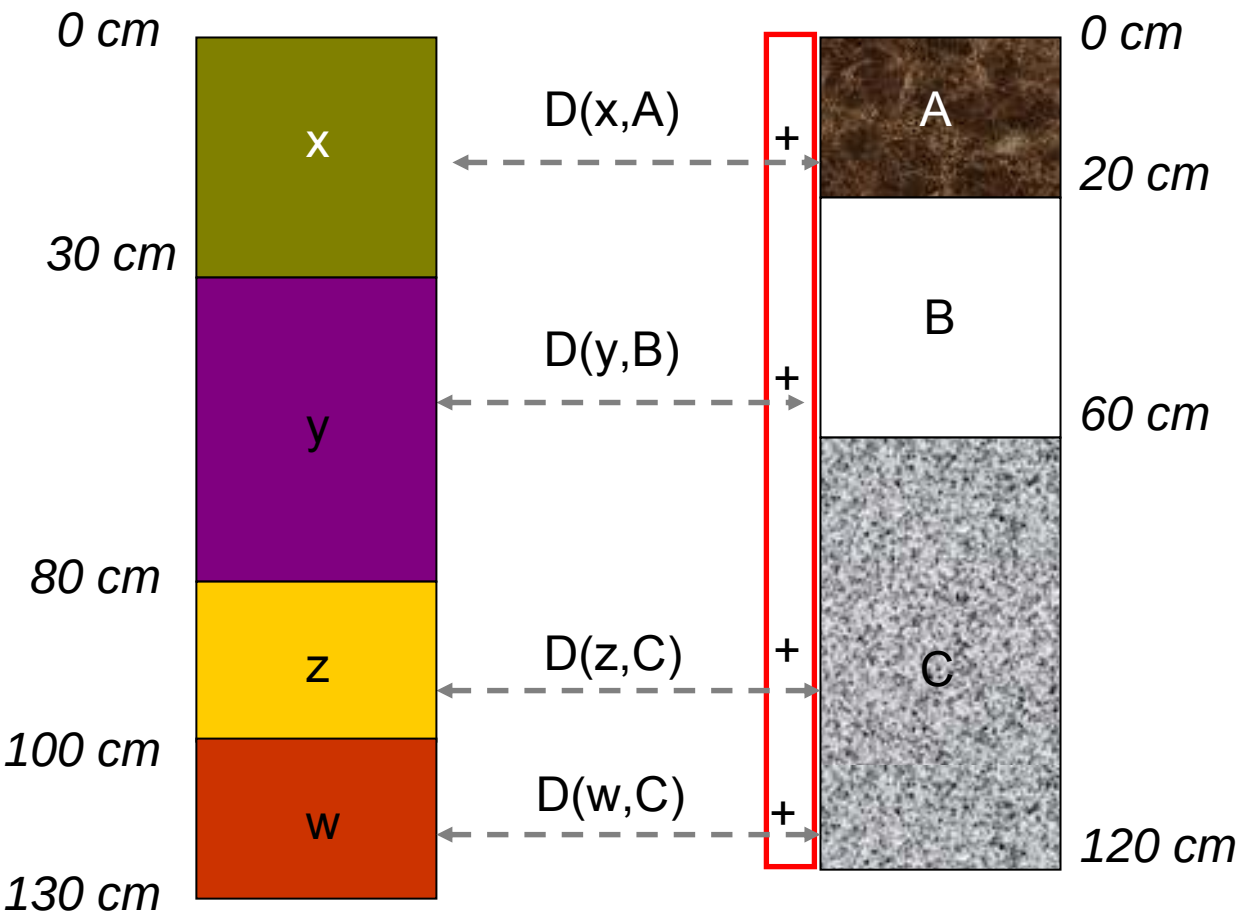
  - **Pedological distance** for emphasizing the different succession of horizons whatever their thickness

  - **Utilitarian** or environmental **distance** for emphasizing the successions of horizons for which the thickness is important

  - **Joint** or normalised **distance** is a mixed between the two previous distances

Pedological distance

$$\frac{\sum D}{\sum (\#\ D)}$$



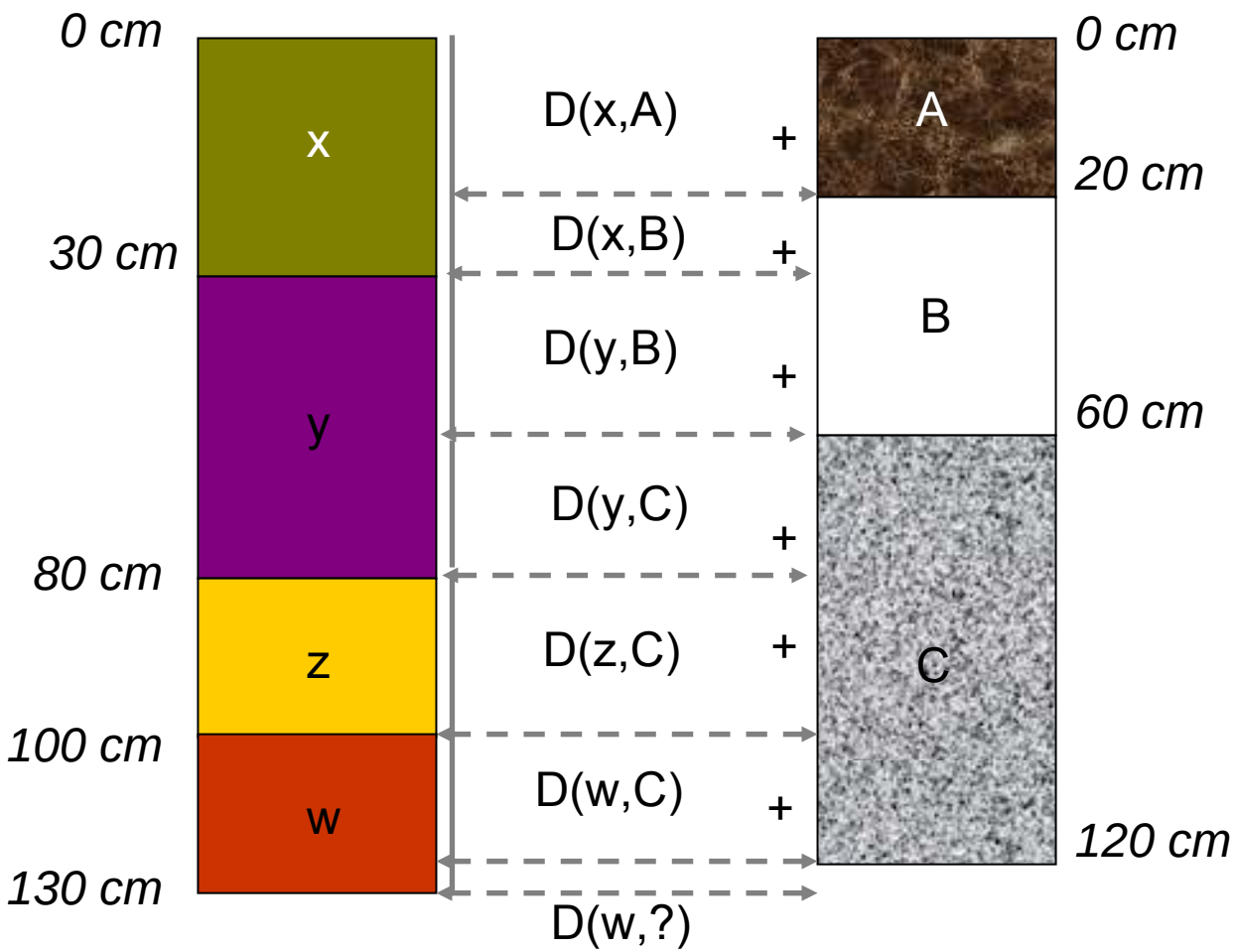
$$D(x, A) = \sqrt{\sum_{k=1}^{n} \left(v_{xk} - v_{Ak} \right)^2}$$

$$M(x, A) = \sum_{k=1}^{n} \left| v_{xk} - v_{Ak} \right|$$

Calculation of distances

Utilitarian distance

$$\frac{\sum D^{*}e}{\text{Max depth}}$$

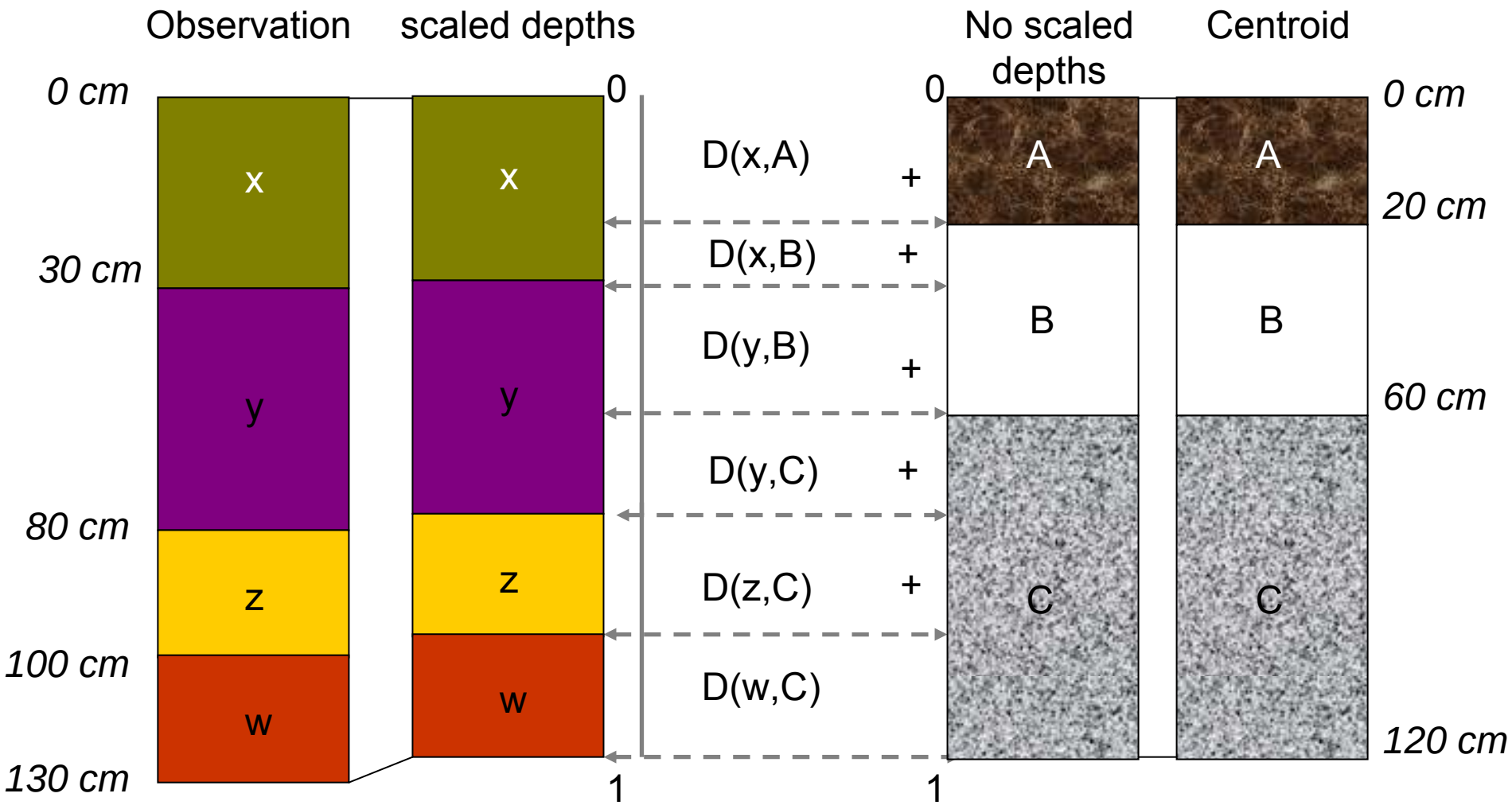


Calculation of distances

Institute for Environment and Sustainability

Joint distance

$$\frac{\sum D^{*}e}{\text{Max depth}}$$



The classification process

❶ Initialisation of cluster centroids

Max. # Soil Type centroids

❷ Observation allocation

Calculation of distances

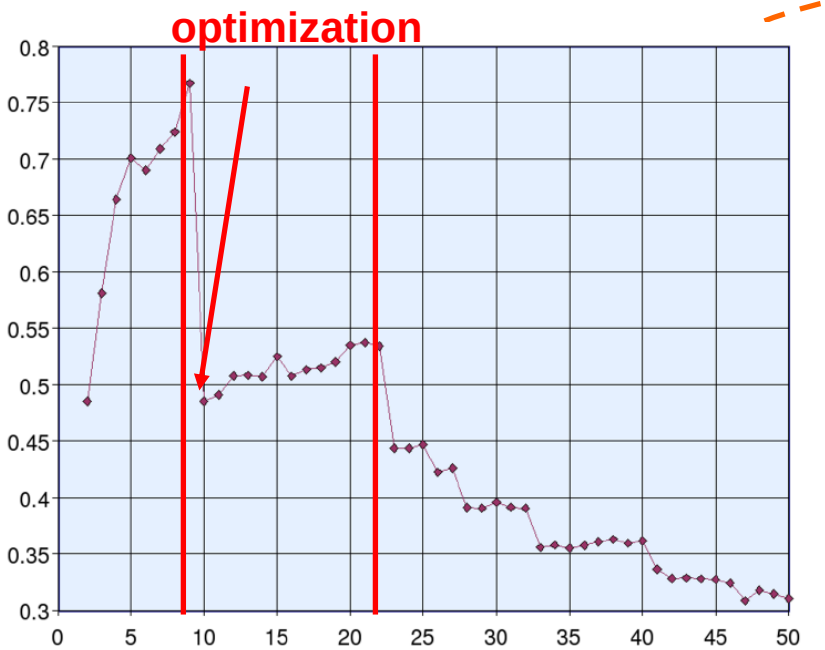
K-means techniques

❸ Centroid readjustment

Calculation of new centroids

❹ Optimisation process

D_{intra}/D_{inter} ratio



Distance table

Distances to centroids — Allocation process

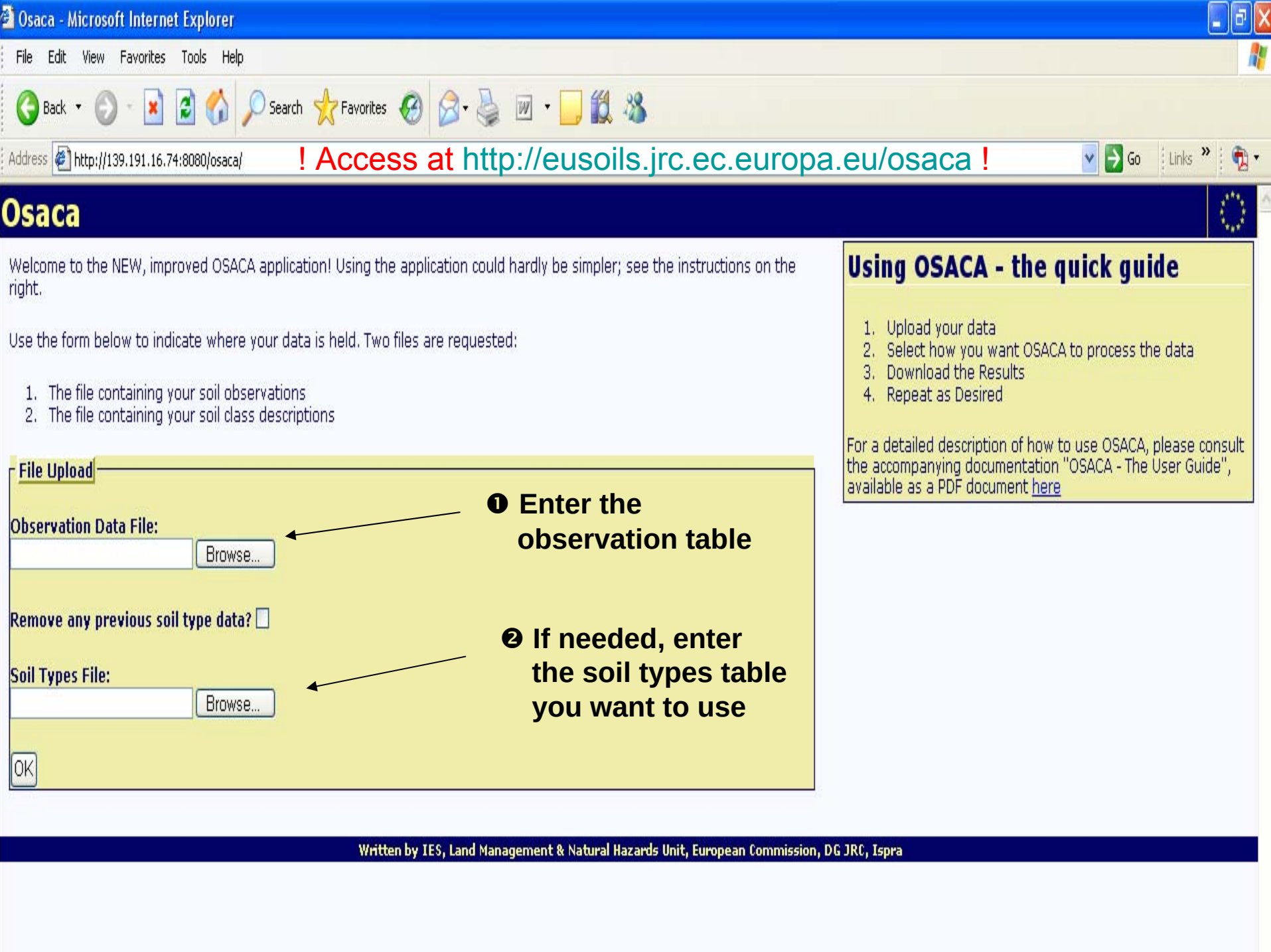
	A	B	C	D	ST	dmin
1	0.7	0.1	0.3	1.3	B	0.1
2	2.5	1.5	0.1	0.6	C	0.1
3	0.6	0.2	1.2	0.4	B	0.2
4	0.8	0.3	1.9	0.4	B	0.3
5	0.1	1.2	0.0	3.0	A	0.1

Soil observations

New centroid for B cluster = observation for which $d_{\text{intra B cluster}}$ minimal

Centroid readjustment process

Distance calculation iteration stopped when centroids *ite* $_{n}$ = centroids *ite* $_{n-1}$


Osaca - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://139.191.16.74:8080/osaca/
! Access at http://eusoils.jrc.ec.europa.eu/osaca !
Go Links
Osaca
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
File Upload
Observation Data File:
Browse...
Remove any previous soil type data?
Soil Types File:
Browse...
OK
❶ Enter the observation table
❷ If needed, enter the soil types table you want to use
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 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites

Address http://139.191.16.74:8080/osaca/upload Go Links

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

❶ Enter the number of horizon classes

Soil Profiles

Distance metric: Pedological

Classification using: Clustering ◉ Not required ○

Minimum Clusters: 20 (or required number of clusters)

Maximum Clusters: 30

❷ Enter the number of soil classes you want

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.

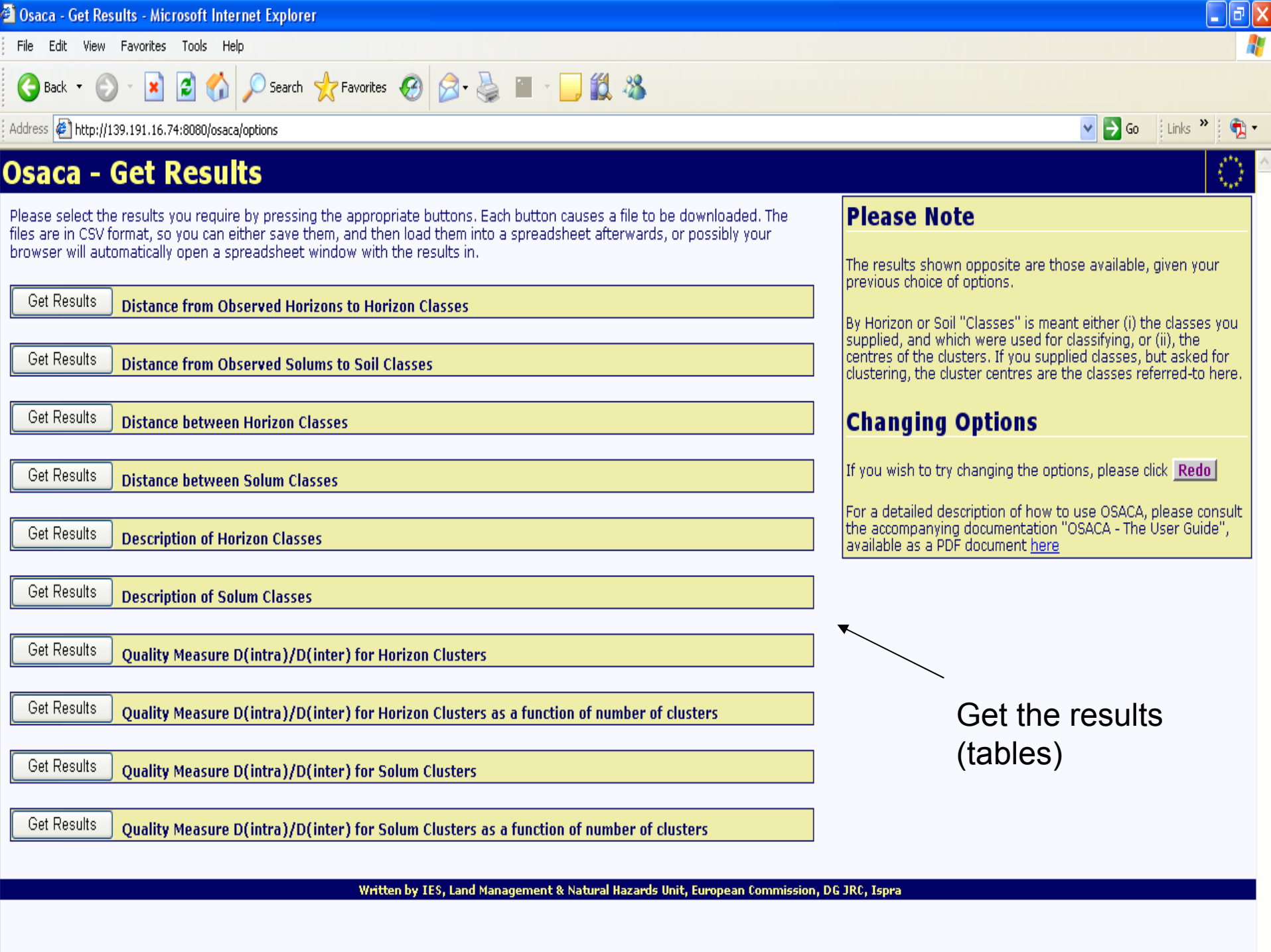
Optimal Clustering

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



Get the results (tables)

OSACA Applications

- OSACA has been used for testing
  - Soil taxonomy purpose
  - Available Water Capacity prediction of soil profiles

Focusing on

  - The number of classes (level of taxonomic detail)
  - The classes robustness (sensitivity of the distances)

Look at:

Carré & Jacobson, 2008. Numerical classification of soil profile data using distance metrics. Geoderma (available on line)