MaxVol sampling for agricultural survey

Anna Petrovskaia

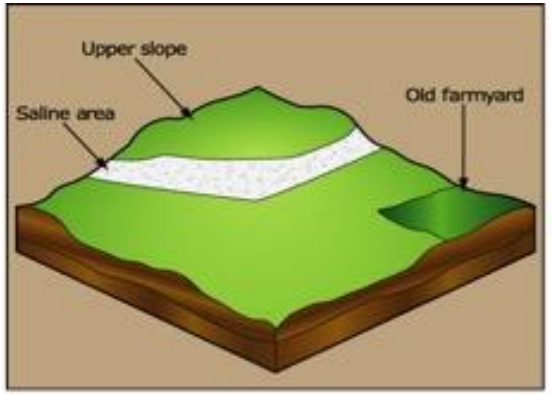
Skoltech, 2019

Soil sampling

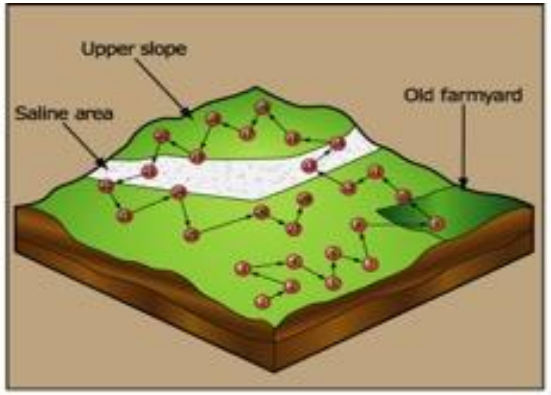
- **Purpose** - obtain data that enable the estimation of some statistical parameter, or spatial predictions of some properties over an area

- **Constrains** - the financial and available resources

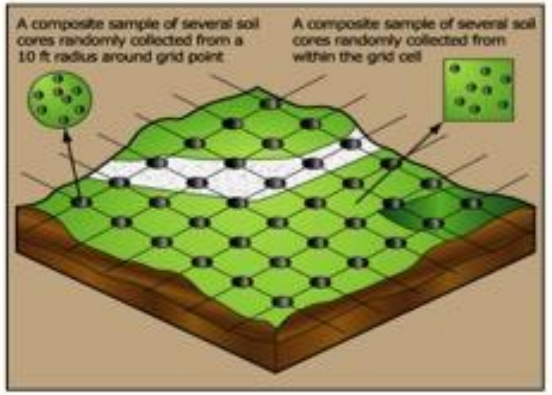
Thus an efficient sampling strategy is sought.



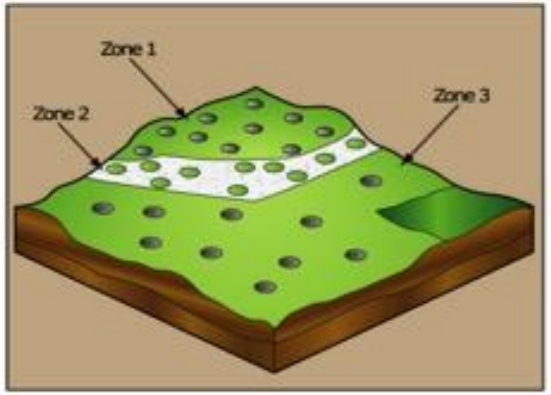
The field



Conventional sampling



Grid Sampling

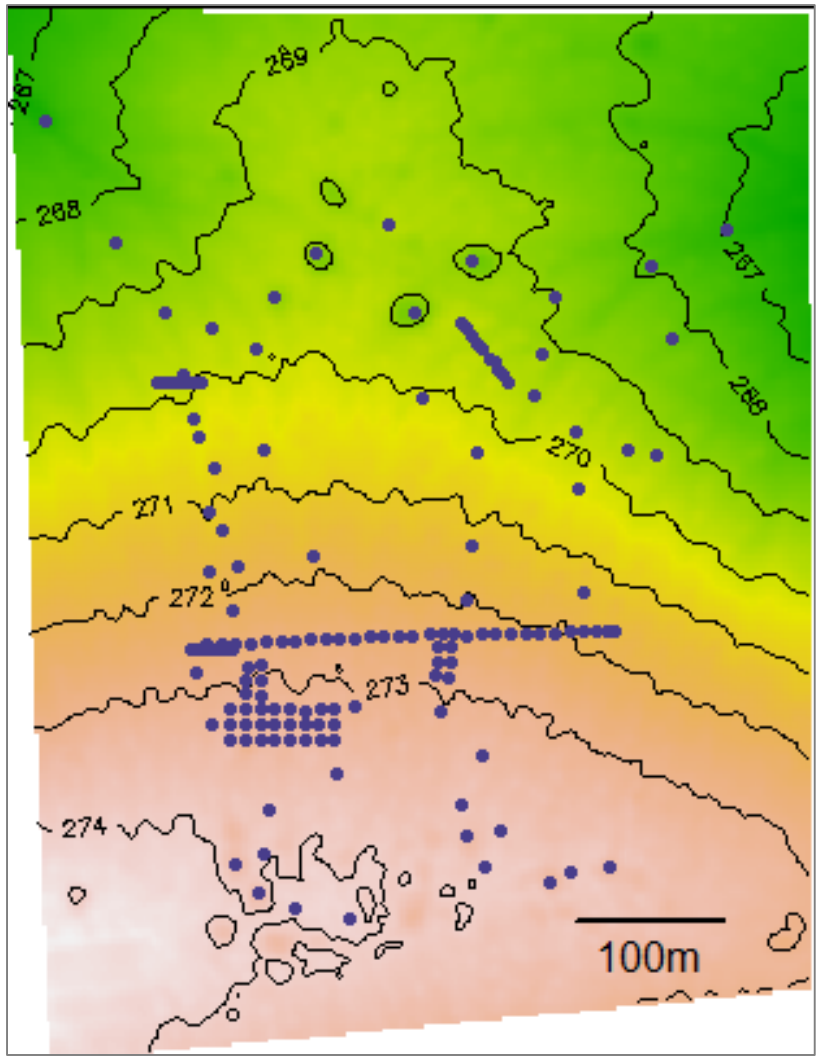


Sampling by management zone

Study area

location of study area

photo of study area



location of soil sampling points

Data

Features

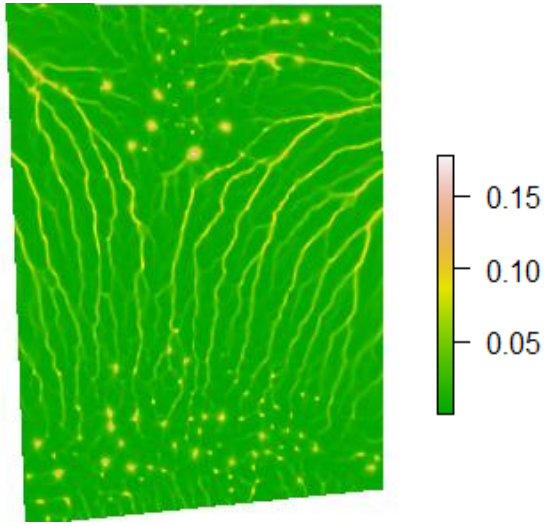
- Digital elevation model
- Topographic wetness index
- Elevation difference in 10 m area
- Slope steepness
- Overland flow hydrologic simulation

Derived from digital elevation model

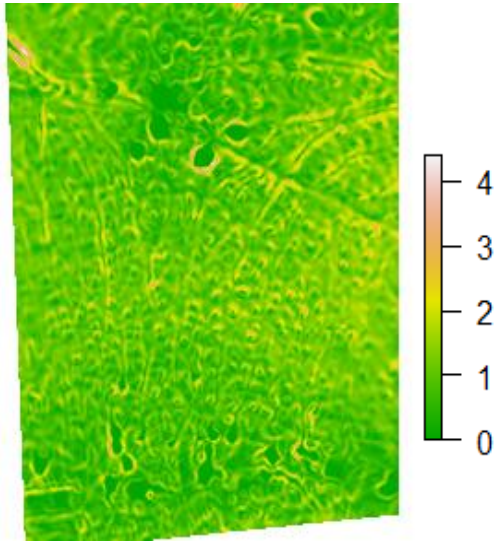
Predicted value

Soil type in every pixel

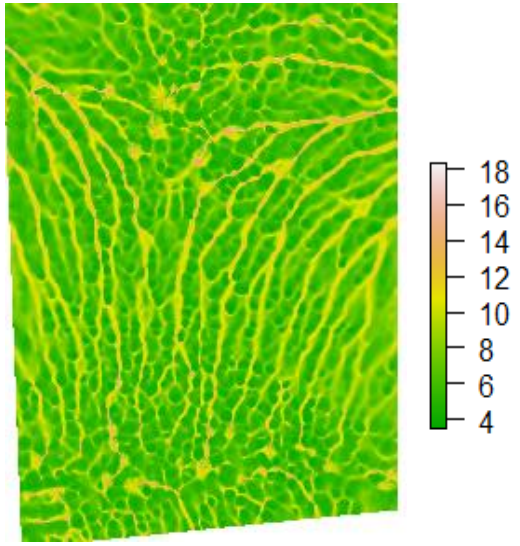
Data



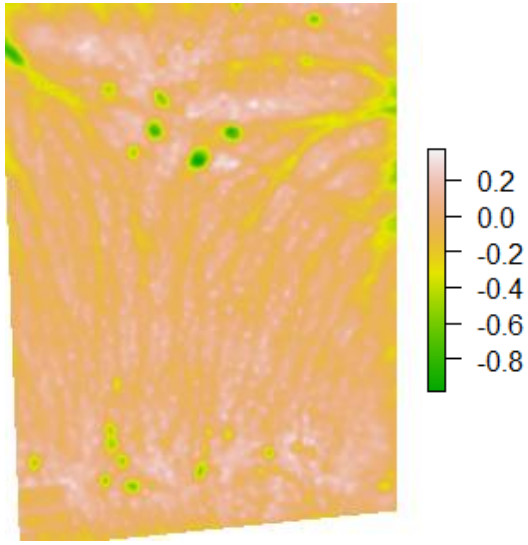
Topographic wetness index



Slope steepness



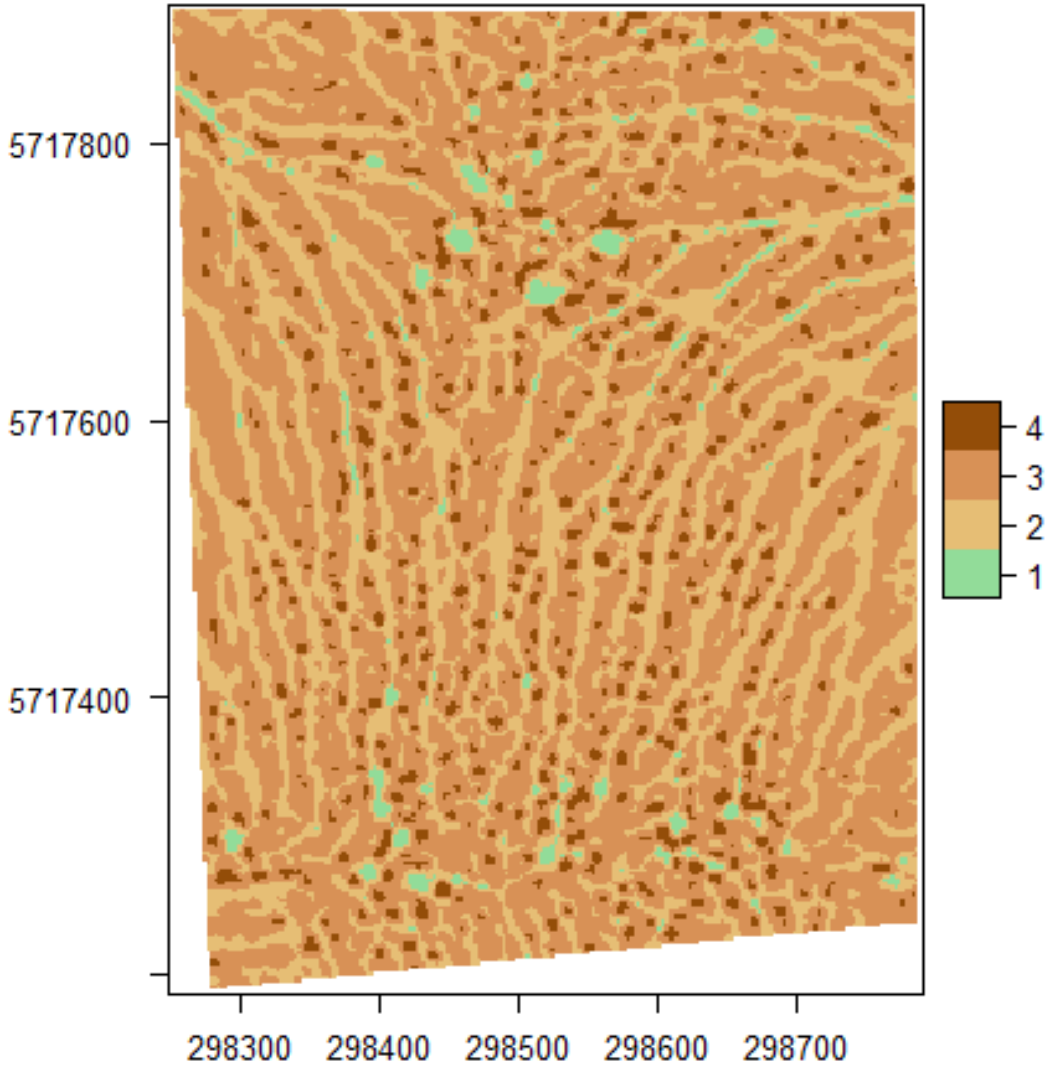
Overland flow hydrologic simulation



Elevation difference in 10 m area

Size of images = 285 x 217 pixels

Data



Soil map

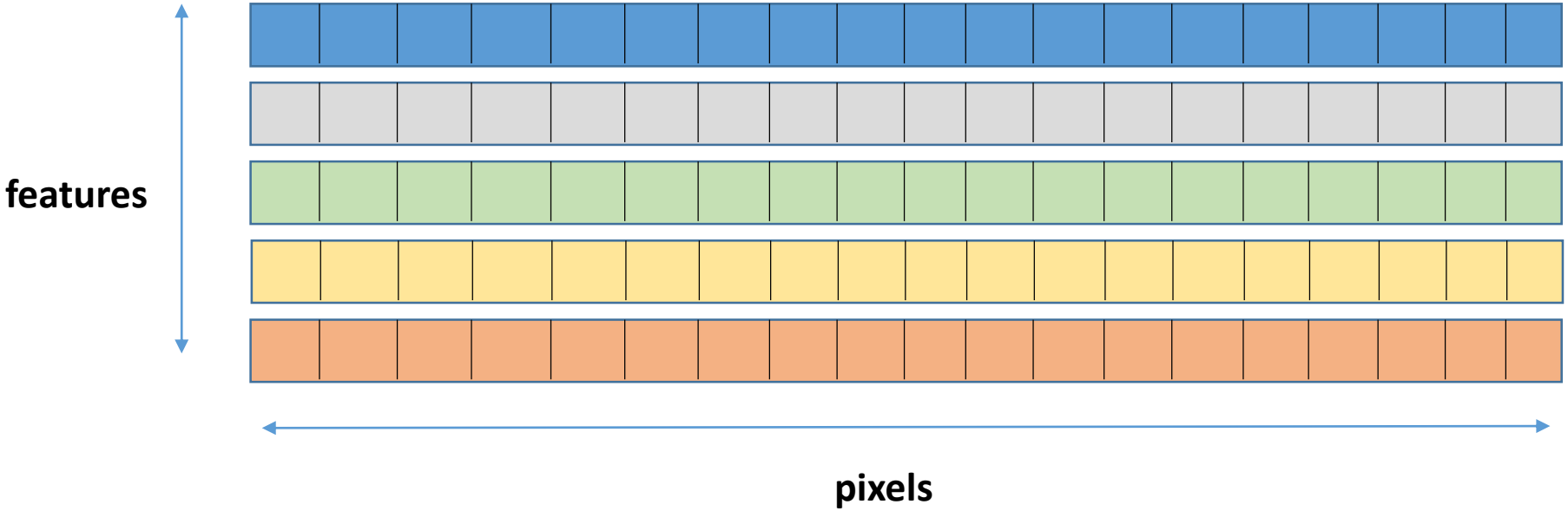
4 soil types:

Chernozems typical calcareous with bioturbations
Chernozems typical
Chernozems leached
Meadow-chernozemics

Predicted by Naïve Bayes classification
on 157 soil samples

Matrix of training data

Every image of feature is flattened



Shape of the matrix = 5 x 61 845

Clustering layer

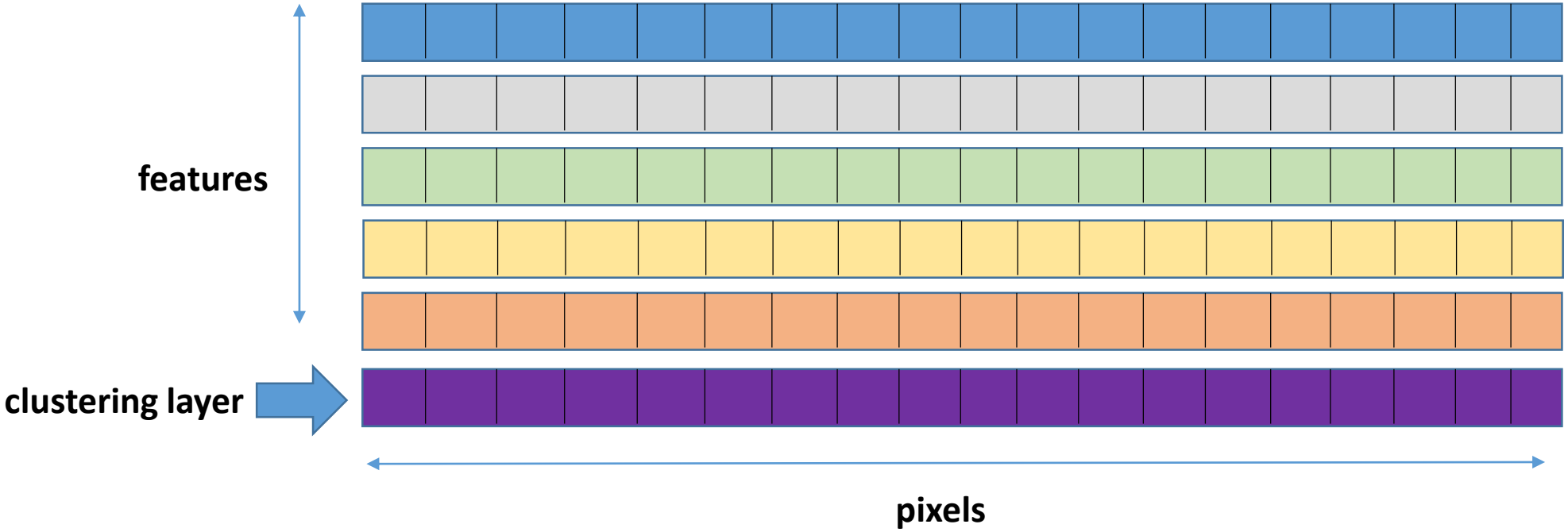
Two steps:

1. Dimensionality reduction
Method – t-SNE
Distributed Stochastic Neighbor Embedding

2. Clustering
Method – DBSCAN
Density-based spatial clustering with noise

Parameters for this methods were obtained by MCMC

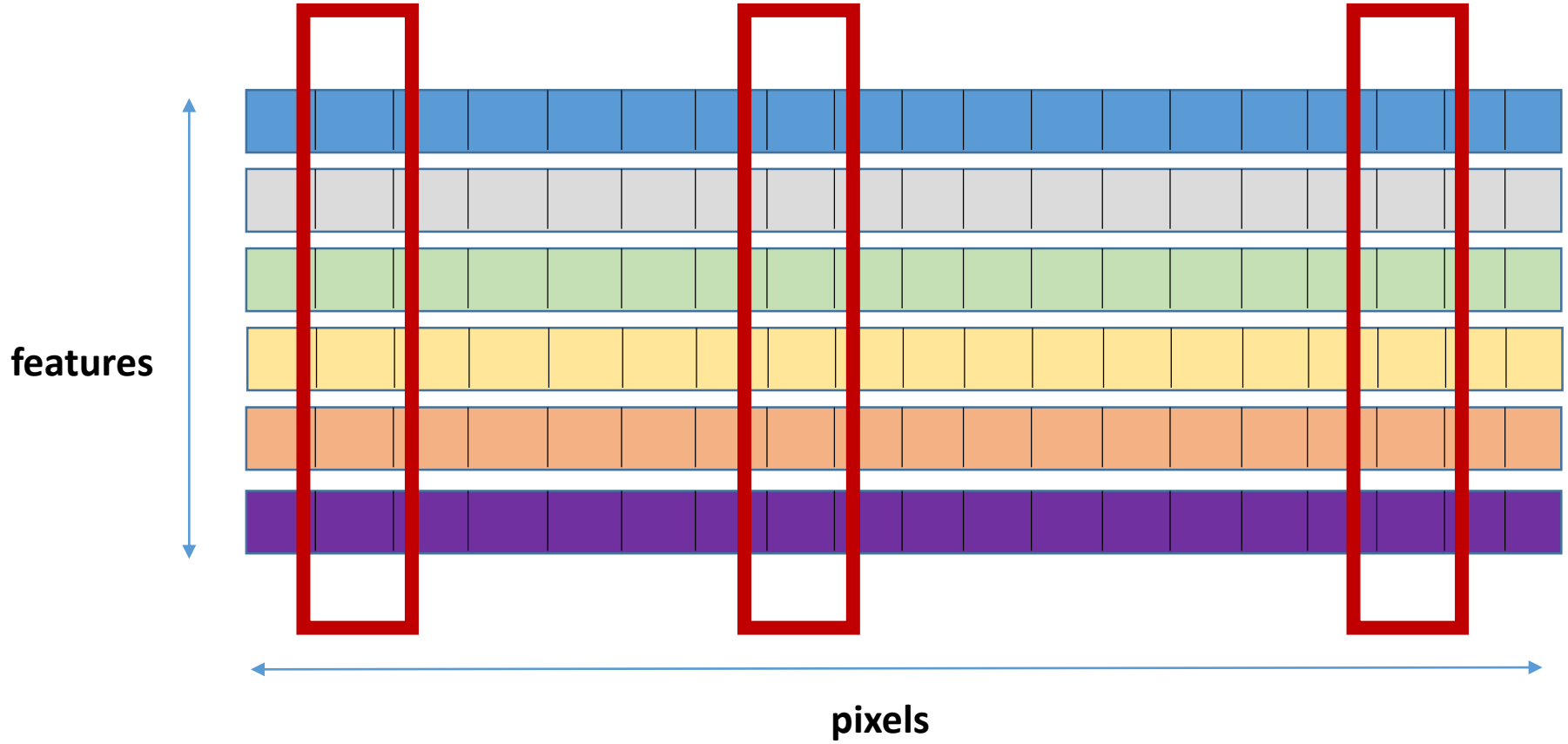
Matrix of training data with clustering layer



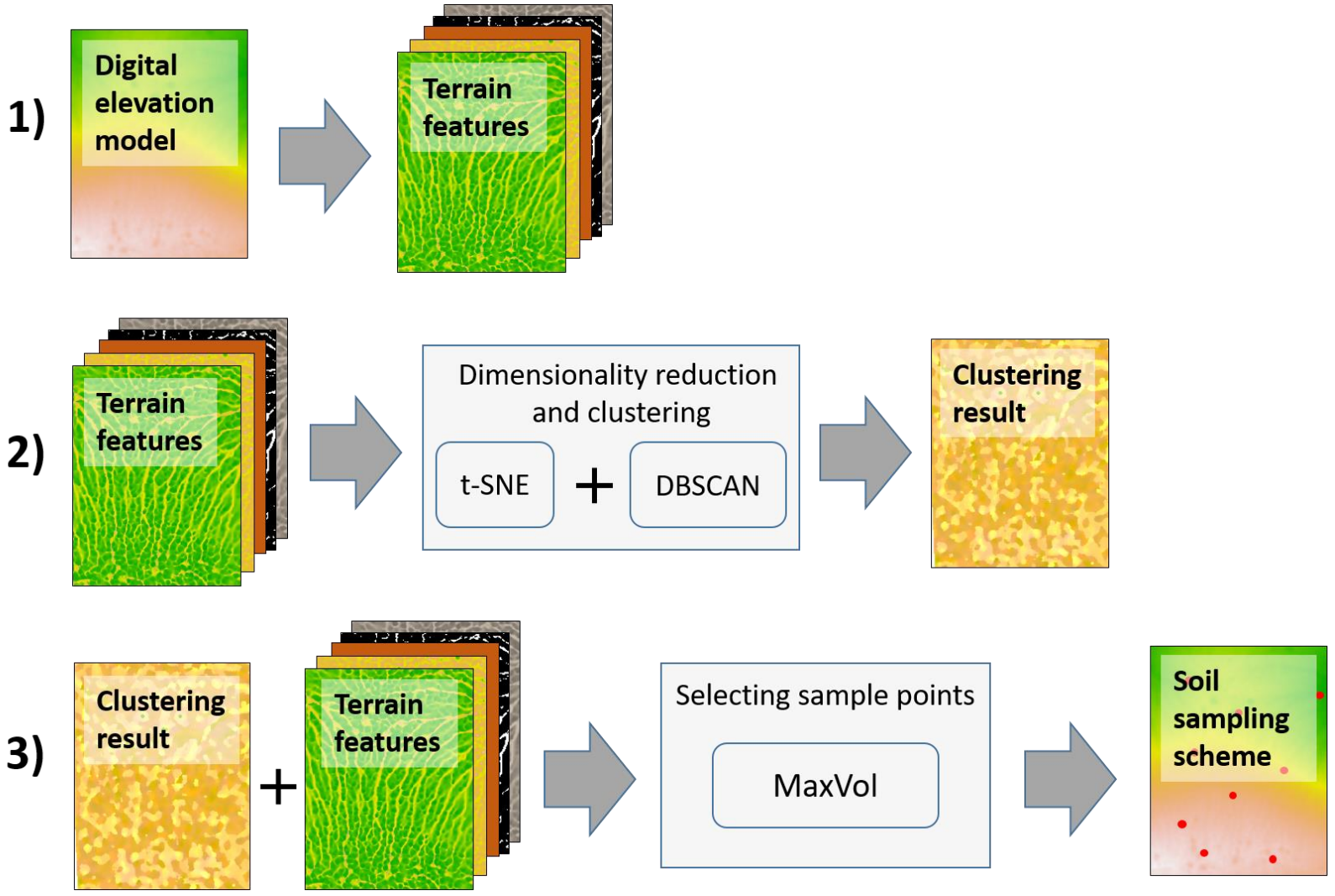
Shape of the matrix = 6 x 61 845

Application of MaxVol algorithm

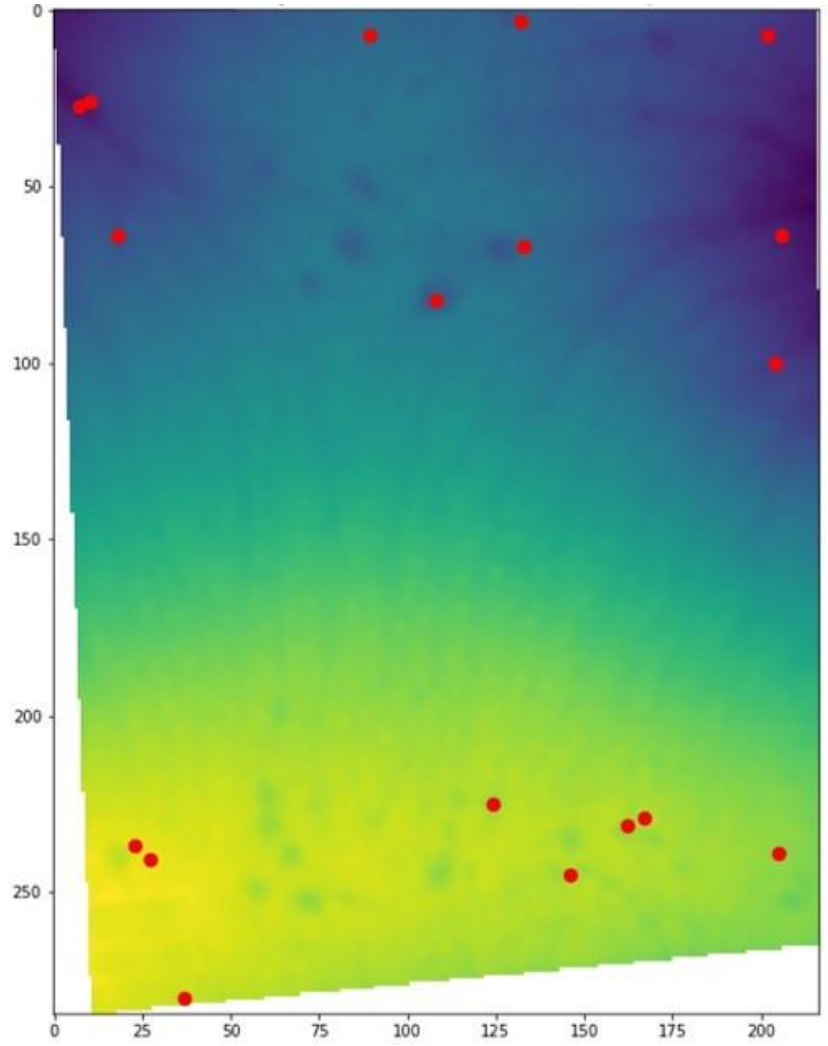
Maxvol is an algorithm for obtaining submatrices of maximum volume



General scheme



Results



Accuracy of prediction

MaxVol with clustering layer = 87%

Conditional Latin Hypercube = 76%

Thank you for your attention!