Thrun, M. C., Lerch, F., Lötsch, J., Ultsch, A.

Visualization and 3D Printing of Multivariate Data of Biomarkers

Conference: International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)

Philipps Universität Marburg

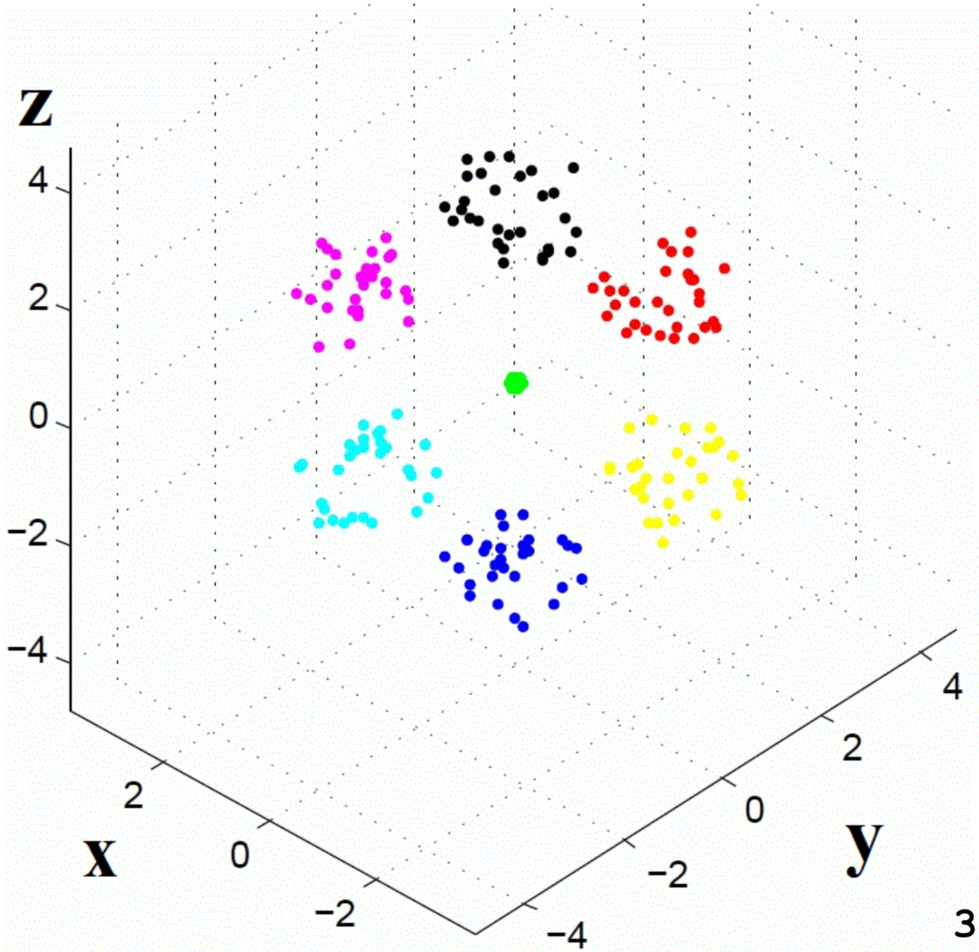
Overview

- Finding structures in high-dimensional data
- Visualization of structures
- How to explain the visualization to topical experts?

Discovering patterns in data

- Searching for similarities
- Clustering: process of finding groups of similar objects in high-dimensional data
  - e.g. labelling data points with colors
- Example: N=3 dimensions in Hepta
  - Equidistant clusters
  - One Cluster has higher density
  - Small distances in each cluster

Hepta data set

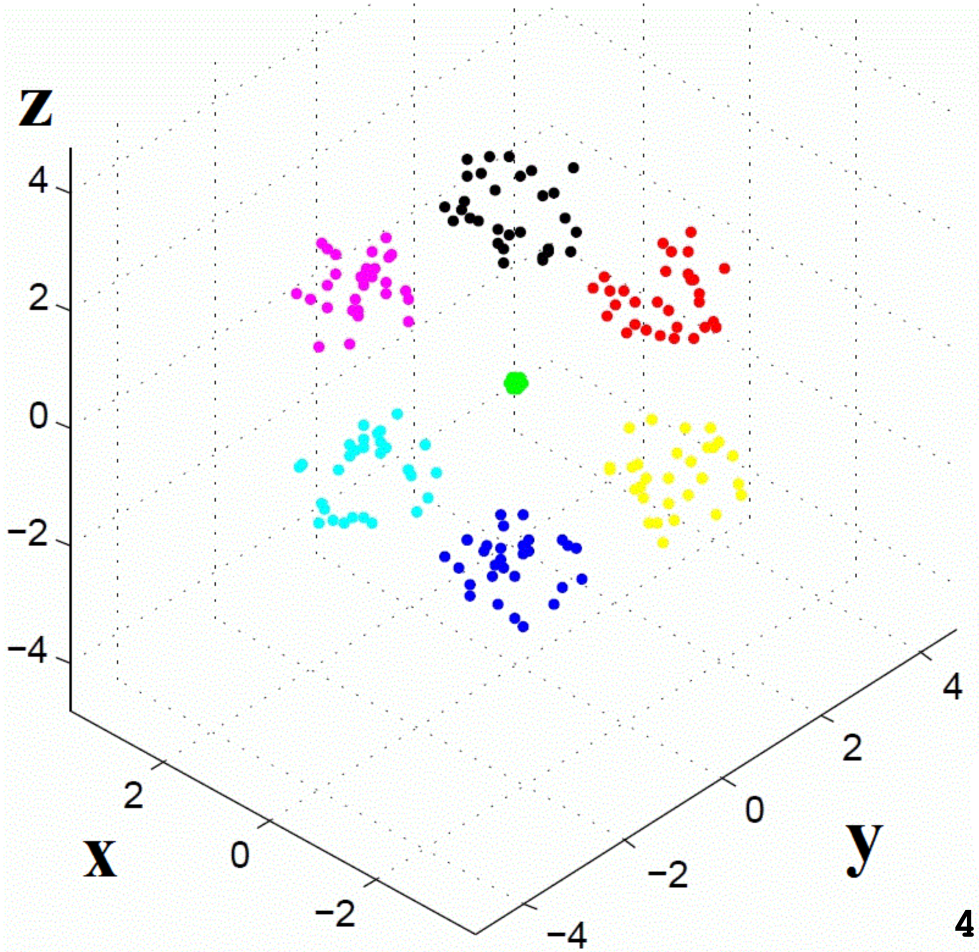


Discovering patterns in data

- Searching for similarities
- Clustering: process of finding groups of similar objects in high-dimensional data
- Problem: high-dimensional: N>>3

How are we able to find similarities?

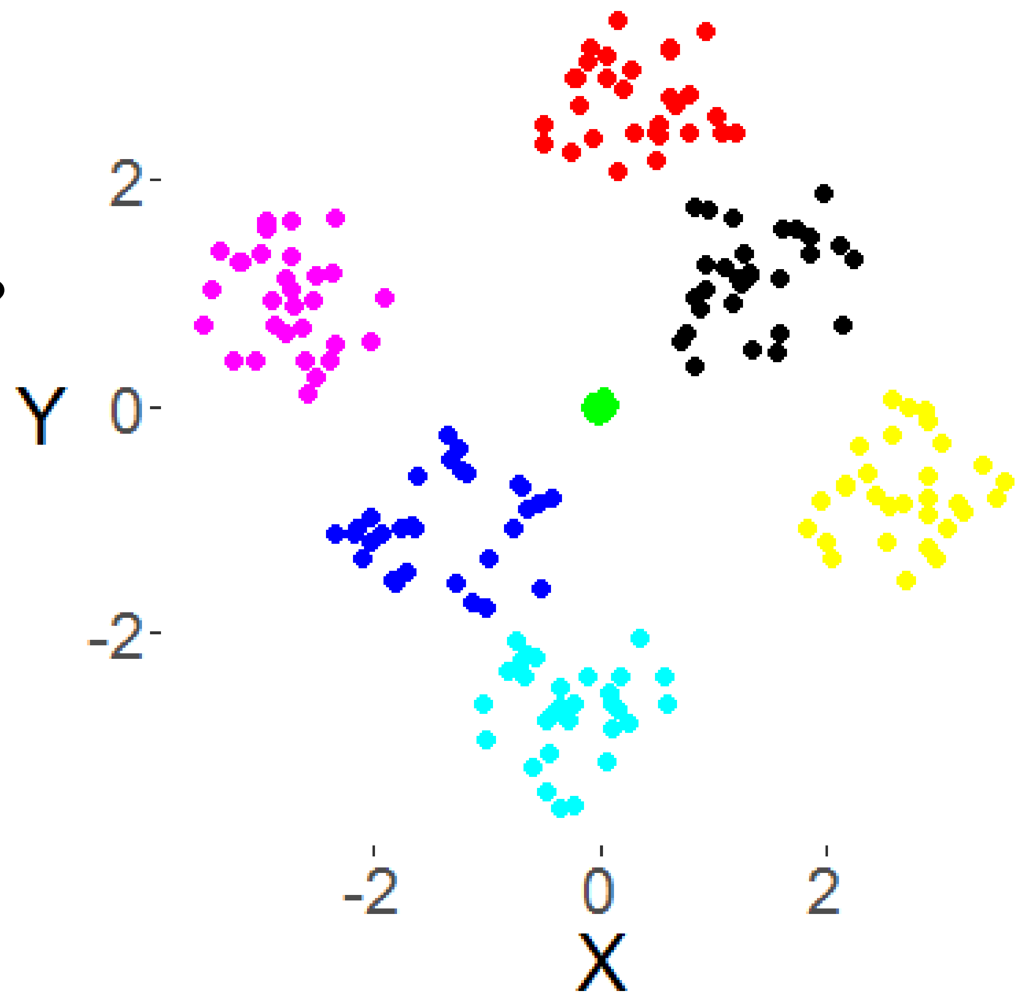
Hepta data set



Finding similarities

1. Use a clustering algorithm
   - Every algorithm has a geometric model for a cluster
   - How to define a cluster? -> application-specific
2. Dimensionality Reduction (DR)
   - In general, *manifold learning,* see [Lee, Verleyson 2007]
   - Special case: ***projections*** into two dimensions

PCA of Hepta

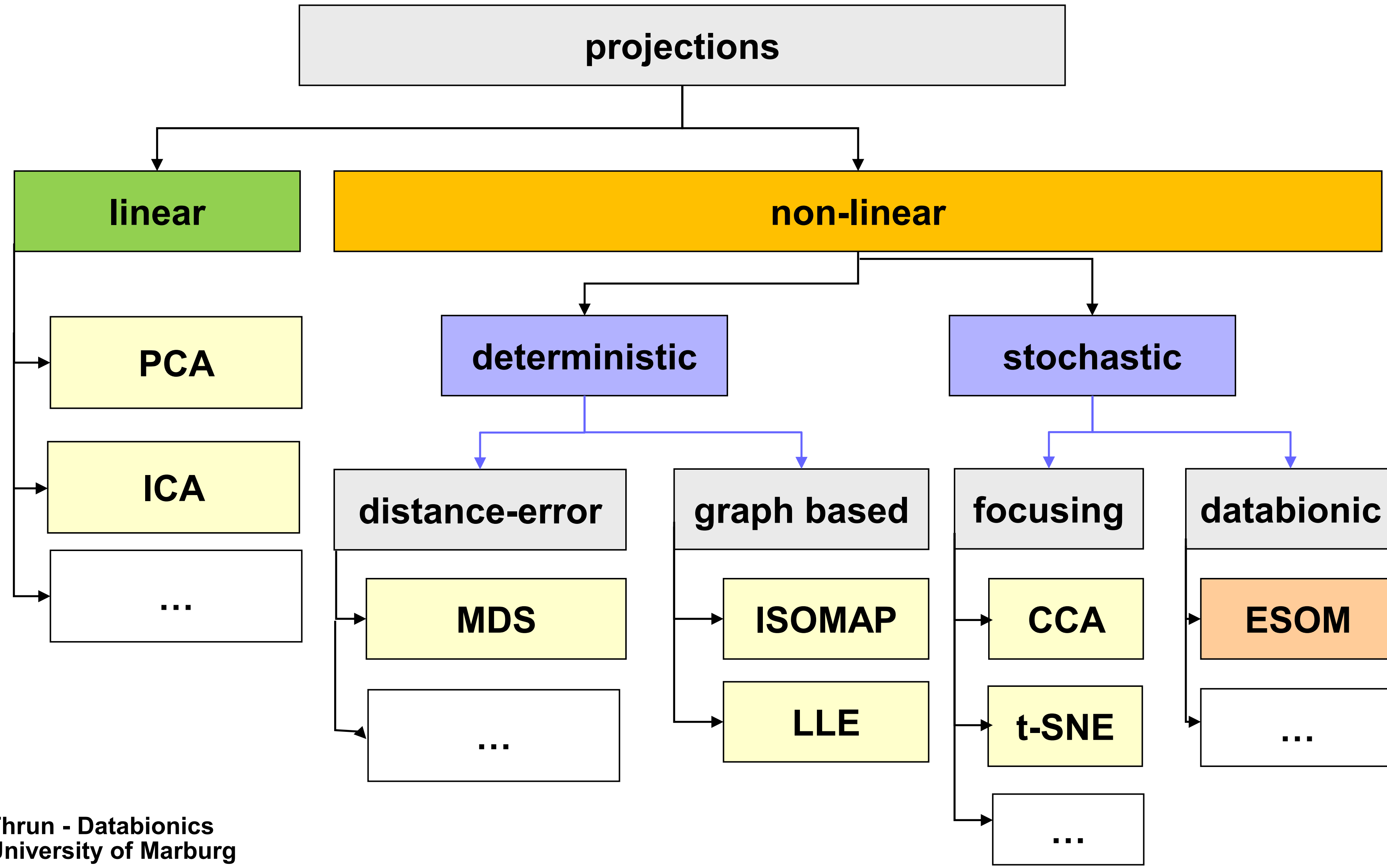


Problems

- Problems of high-dimensionality
  - Empty space and concetration of measure phenomen, curse of dimensionality
- Choice of dissimilarity (distance)

- Projection R^N ->R^2
  - Various methods
  - Choice of parameters
  - Problem of structure preservation: no projection method is able to preserve all distances

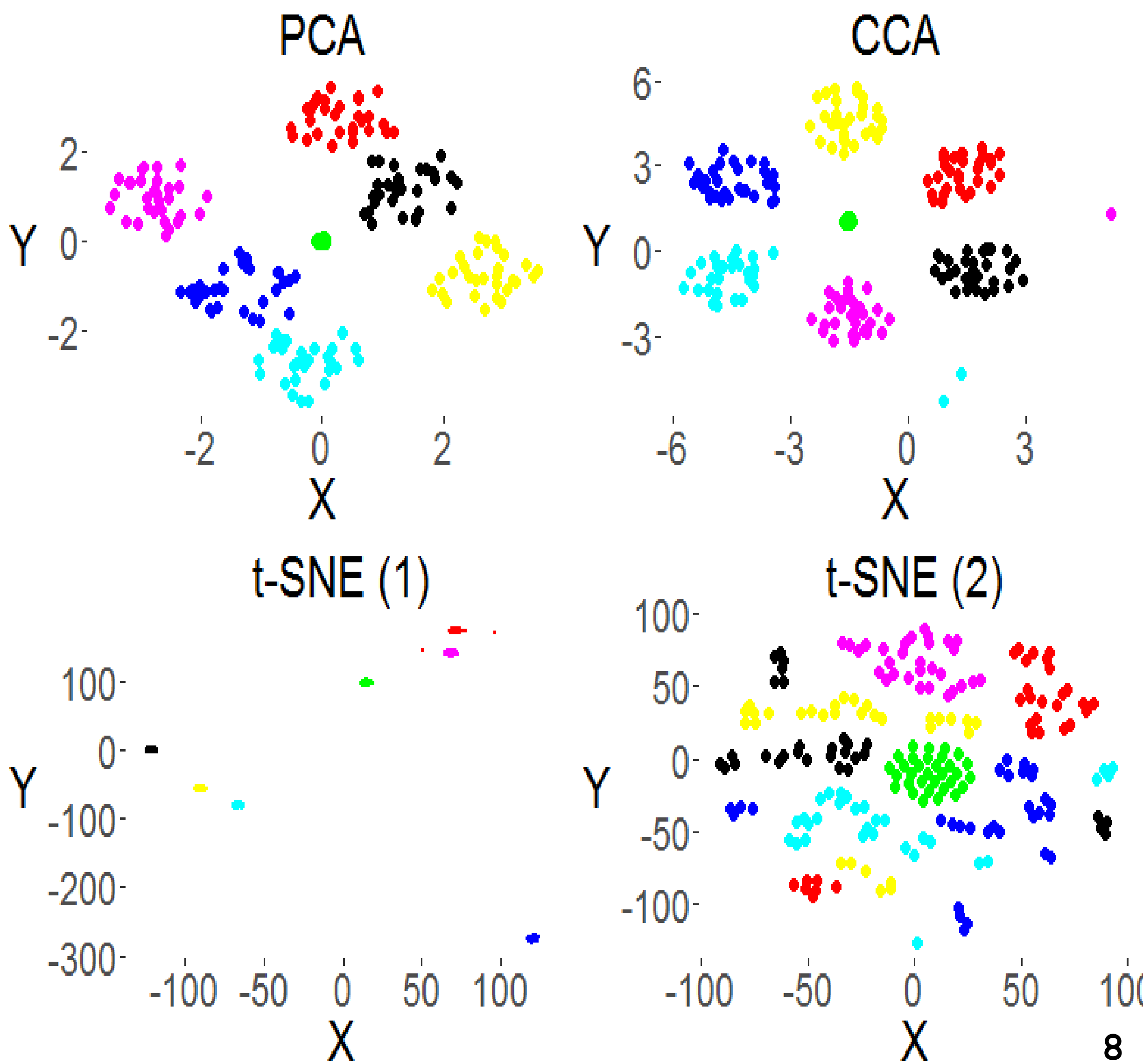
Various projection methods



Example:

- Choice of projection method
- Choice of parameters for the projection

Hepta projections



Structure preservation:

projection

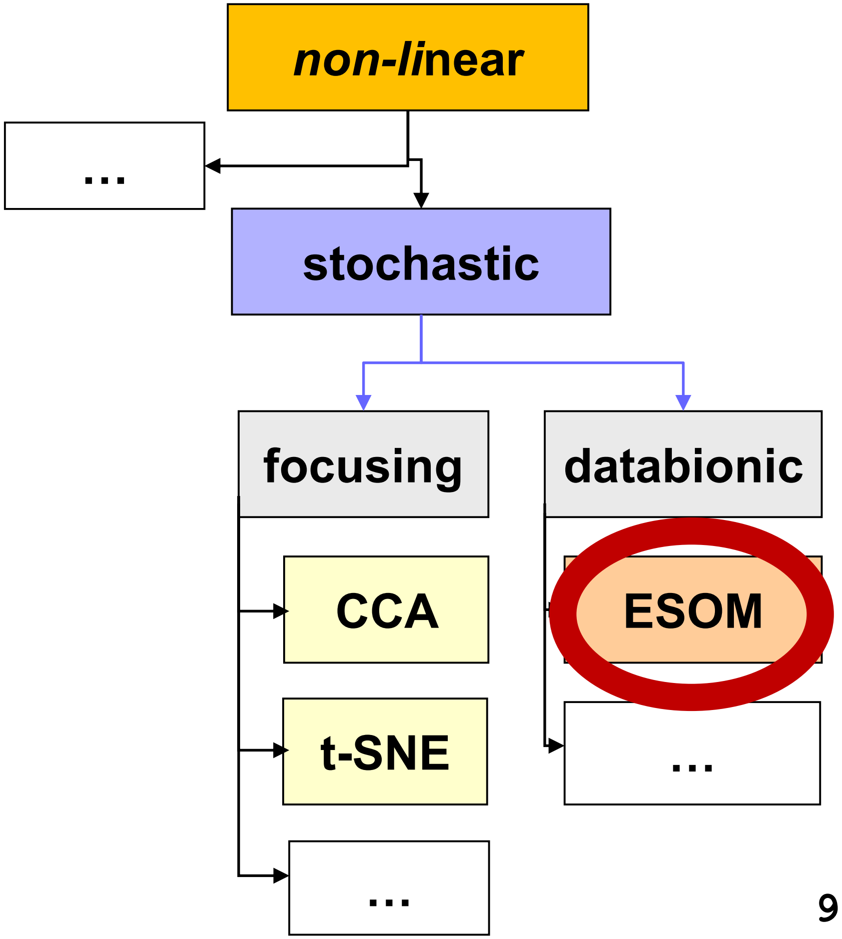
$R^N \; -> R^m$ with $N > M$ cannot preserve all distances!

Emergent Self-Organizing Map (ESOM)

- Unsupervised neural learning algorithm
- Many neurons => projection method with emergent properties
  - Emergent: novel and unforseen properties of a multi-agent system

Why ESOM?

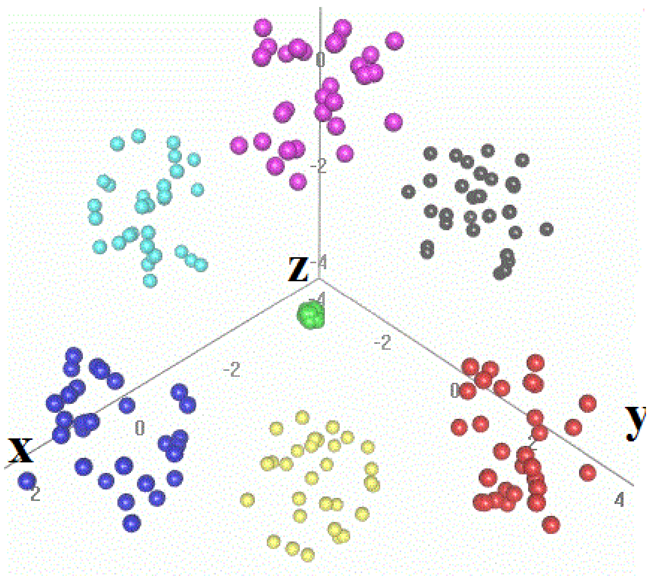
- For high-dimensional data, the SOM remains a reference tool for 2D visualizations. [Lee/Verleysen, 2007, p. 227]

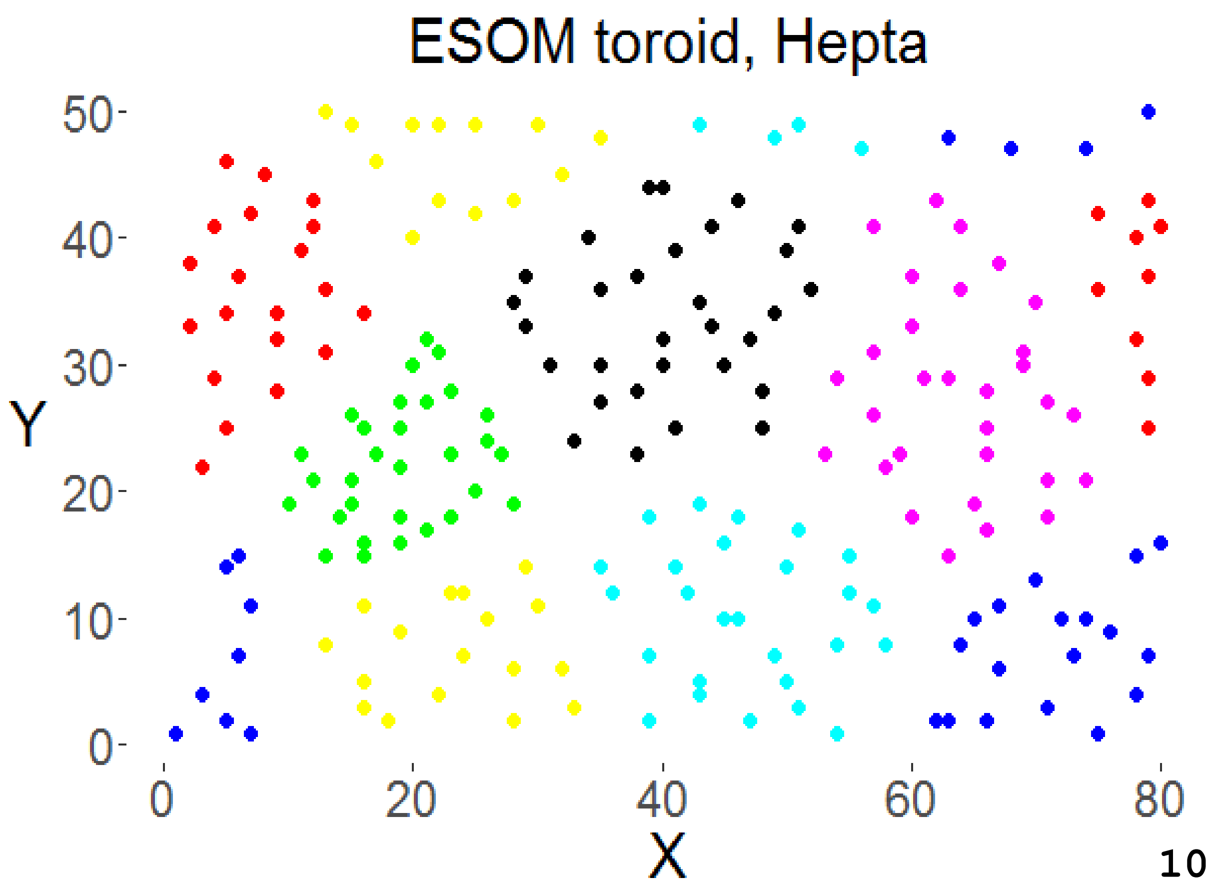


ESOM result of Hepta

- **(-)** Projected points uniformly distributed
- **(-)** toroidal grid of neurons
- **(-)** structure preservation hidden

Toroid: borders are cyclically connected



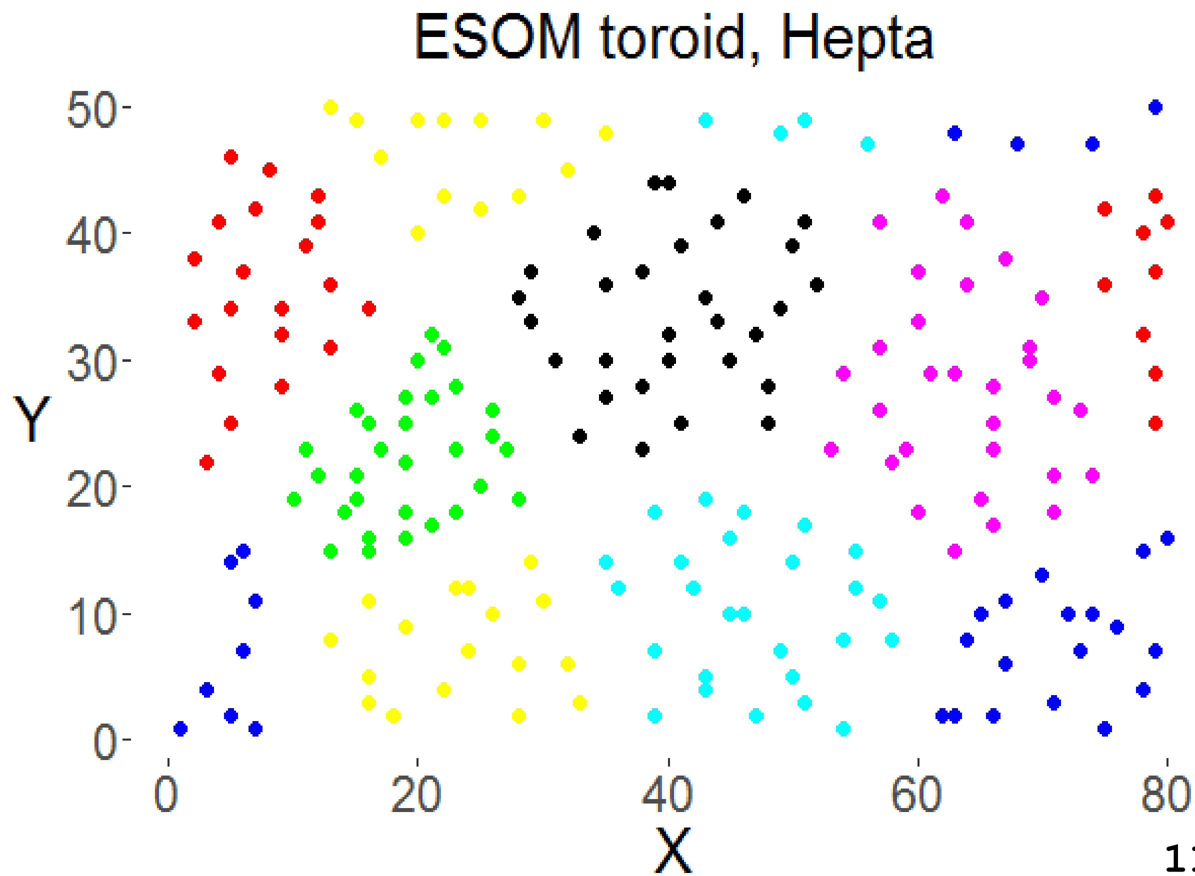


ESOM result of Hepta

(-) Projected Points uniformly distributed
(-) toroidal grid of neurons
(-) structure preservation hidden

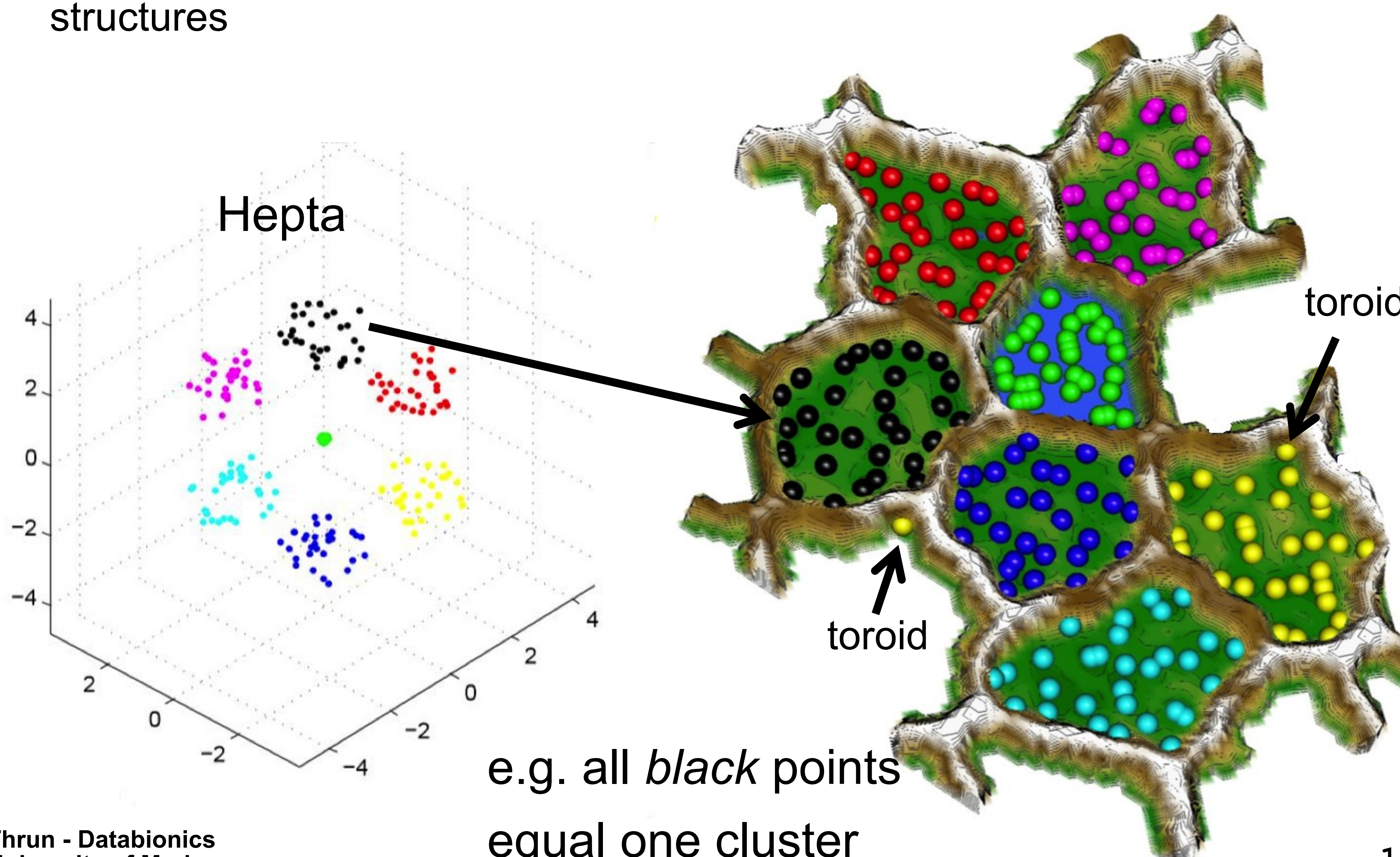
Toroid:
borders are cyclically connected

How can we solve this?



Solution: 3D visualization

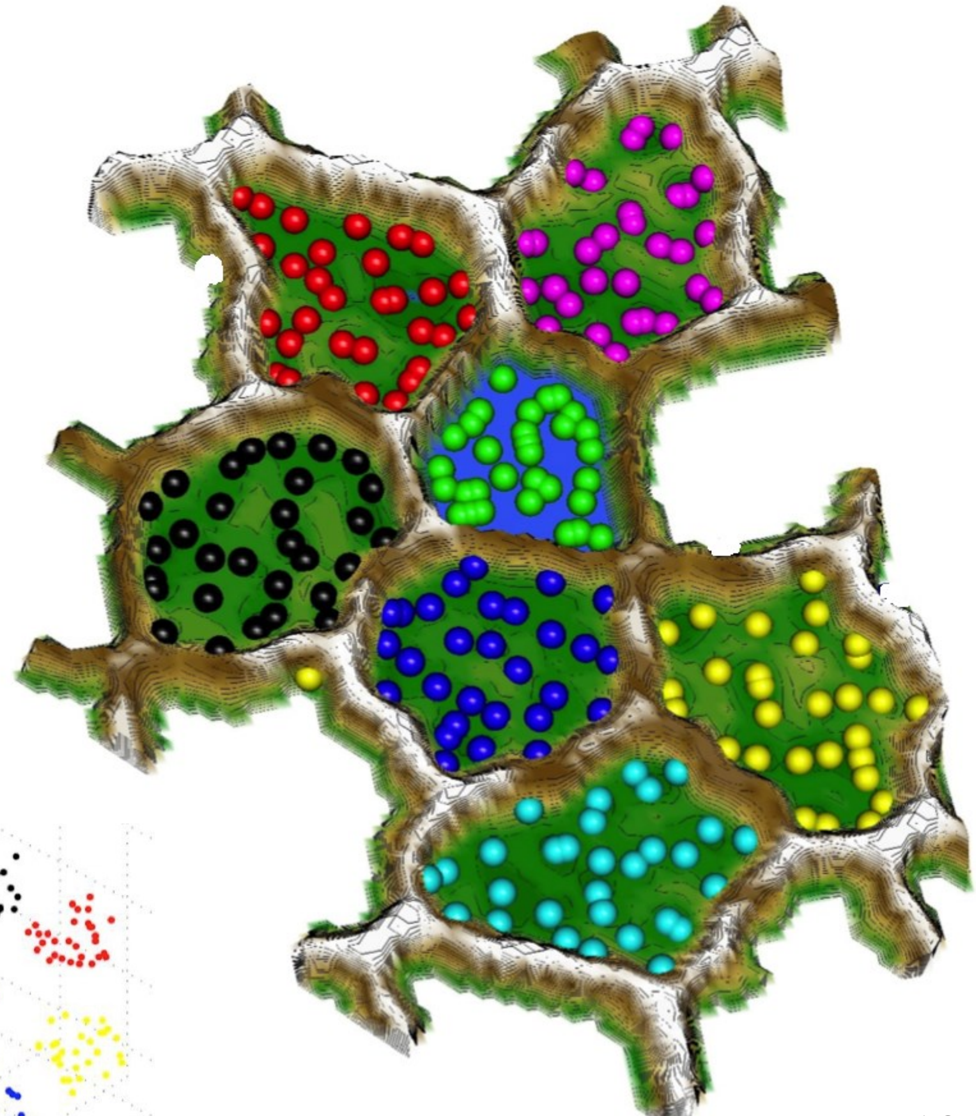
3D landscape allows the human eye to detect meaningful cluster structures

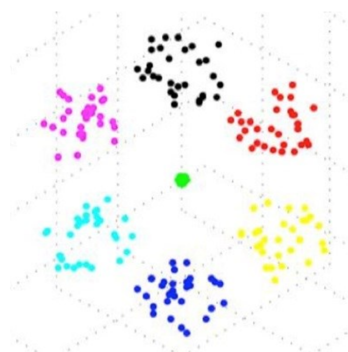


e.g. all *black* points equal one cluster

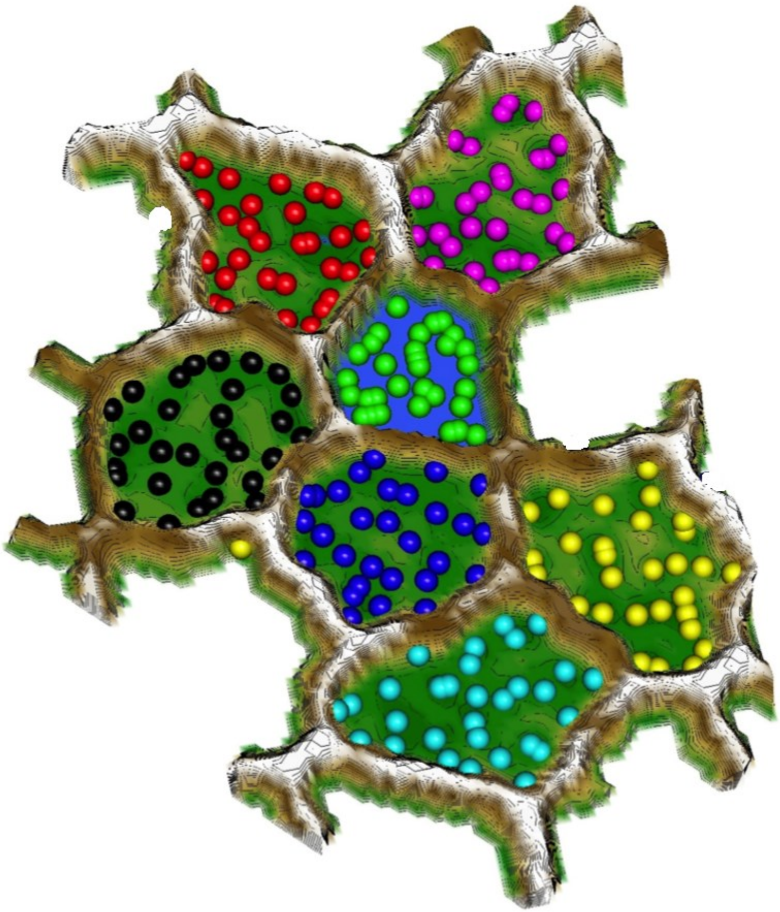
Display information in an easily understandable way

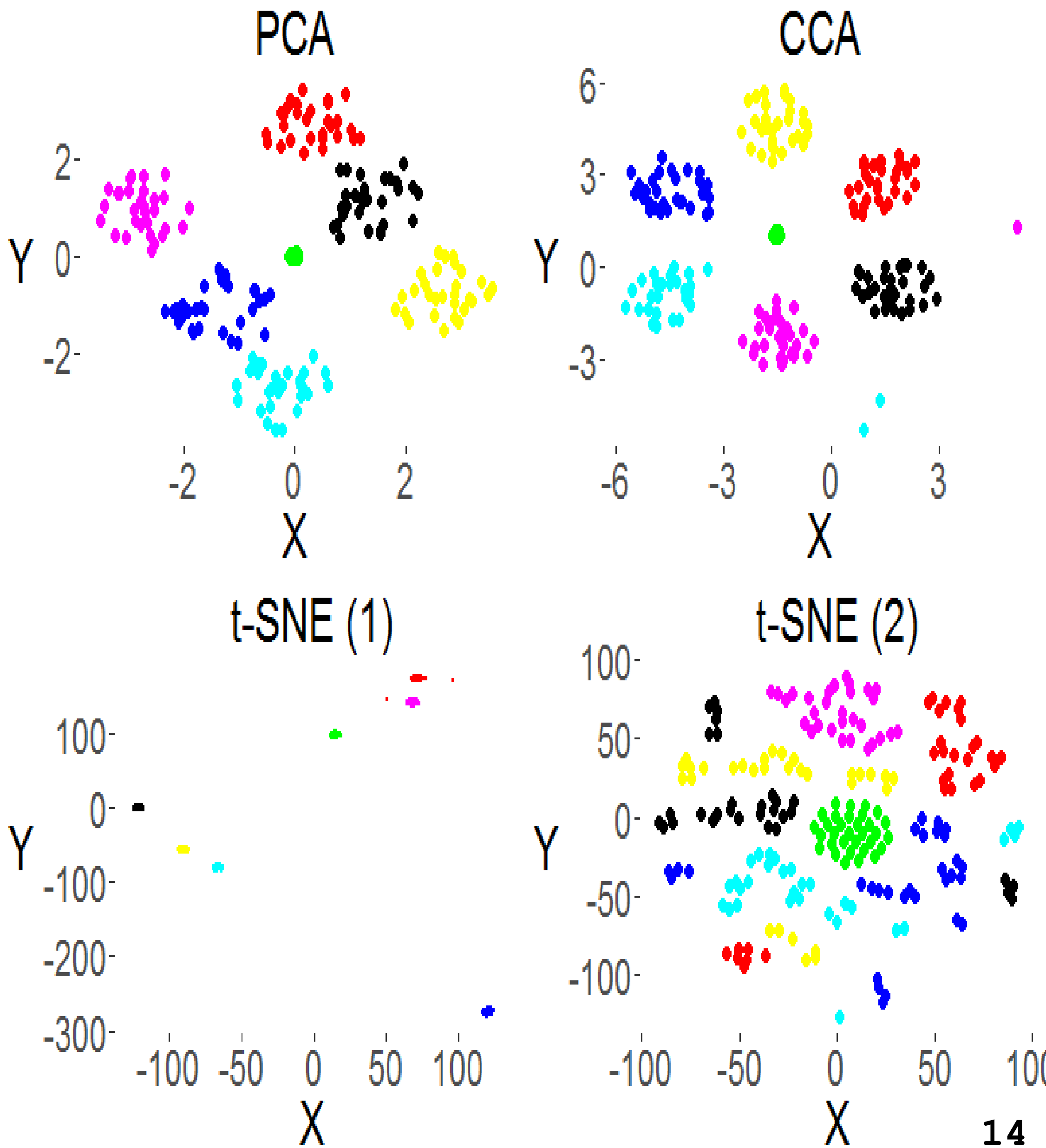
- Blue colors indicate small distances (sea level)
- Green and brown colors indicate middle distances (small hilly country)
- White colors indicate high distances (snow and ice of high mountains).
- The valleys and basins indicate clusters
- Watersheds of hills and mountains indicate borderlines of clusters





Comparison




PCA
2
0
-2
Y
-2
0
2
X
CCA
6
3
0
-3
Y
-6
-3
0
3
X
t-SNE (1)
100
0
-100
-200
-300
Y
-100
-50
0
50
100
X
t-SNE (2)
100
50
0
-50
-100
Y
-100
-50
0
50
100
X

U*matrix

- 3D landscape defined by **U*matrix** [Ultsch et al., 2016]
  - Combines Umatrix and Pmatrix
- Umatrix: folding of high dimensional space [Ultsch/Siemon, 1990]

  (-) In literature cited as grey-scaled 2D vizualization (e.g. [Kadim Tasdemir/Merényi, 2012])

  -> Precise colored definition required
- Pmatrix: high dimensional density estimation technique [Ultsch, 2003]

  (-) Estimation for hypersphere of radius is trying

  -> use ABCanalsys [Ultsch/Lötsch, 2015]
- => **U*matrix represents distance and density based structures!**

U*matrix -> 3D landscape

- use colors proposed in [Ultsch, 2003]

=> hypsometric tints: surface colors which depict ranges of elevation

- Calculate contour lines
  - Normalization of U*heights, define height intervalls, ...
- Combine specific color scale with contour lines
- Create island of toroid map (rectangular)

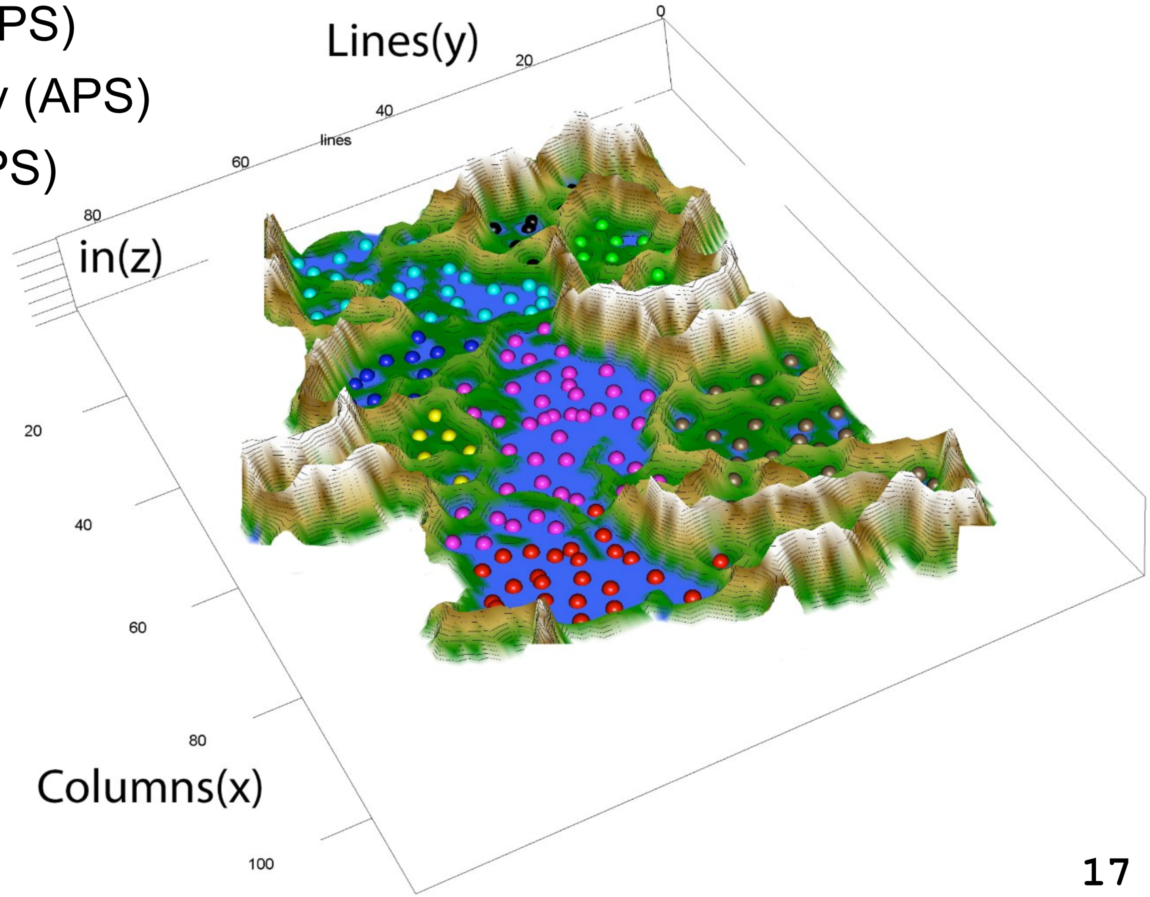
=> topographic map with hypsometric tints

Example: pain biomarkers

- Data: responses to different types of nociceptive stimuli [Flühr et al., 2009]

ESOM -> U*matrix -> 3D vizualization -> Clustering

- Result: three main pain sensitivity groups [Lötsch/Ultsch, 2013]
  - High-pain sensitivity (HPS)
  - Average pain sensitivity (APS)
  - Low-pain sensitivity (LPS)
  - Subclusters...

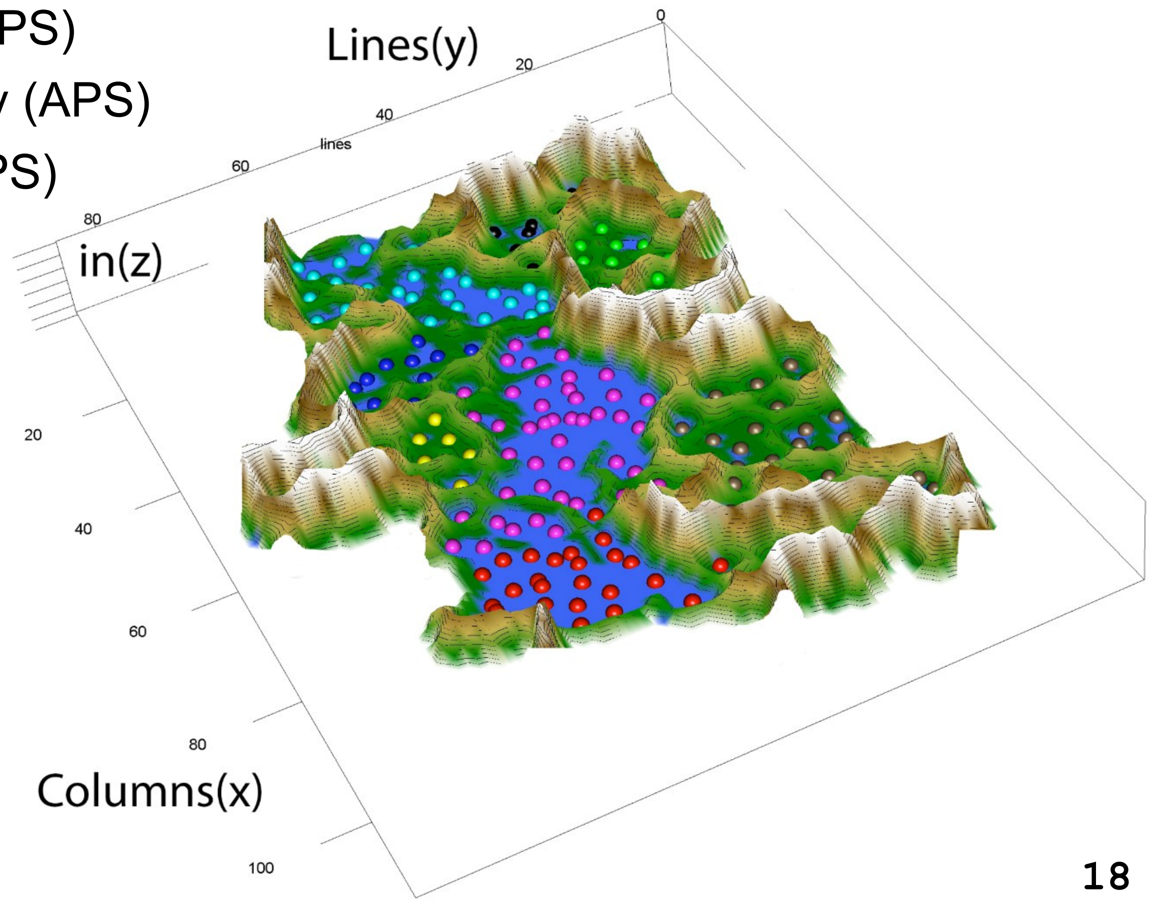


Example: pain biomarkers

- Data: responses to different types of nociceptive stimuli [Flühr et al., 2009]

ESOM -> U*matrix -> 3D vizualization -> Clustering

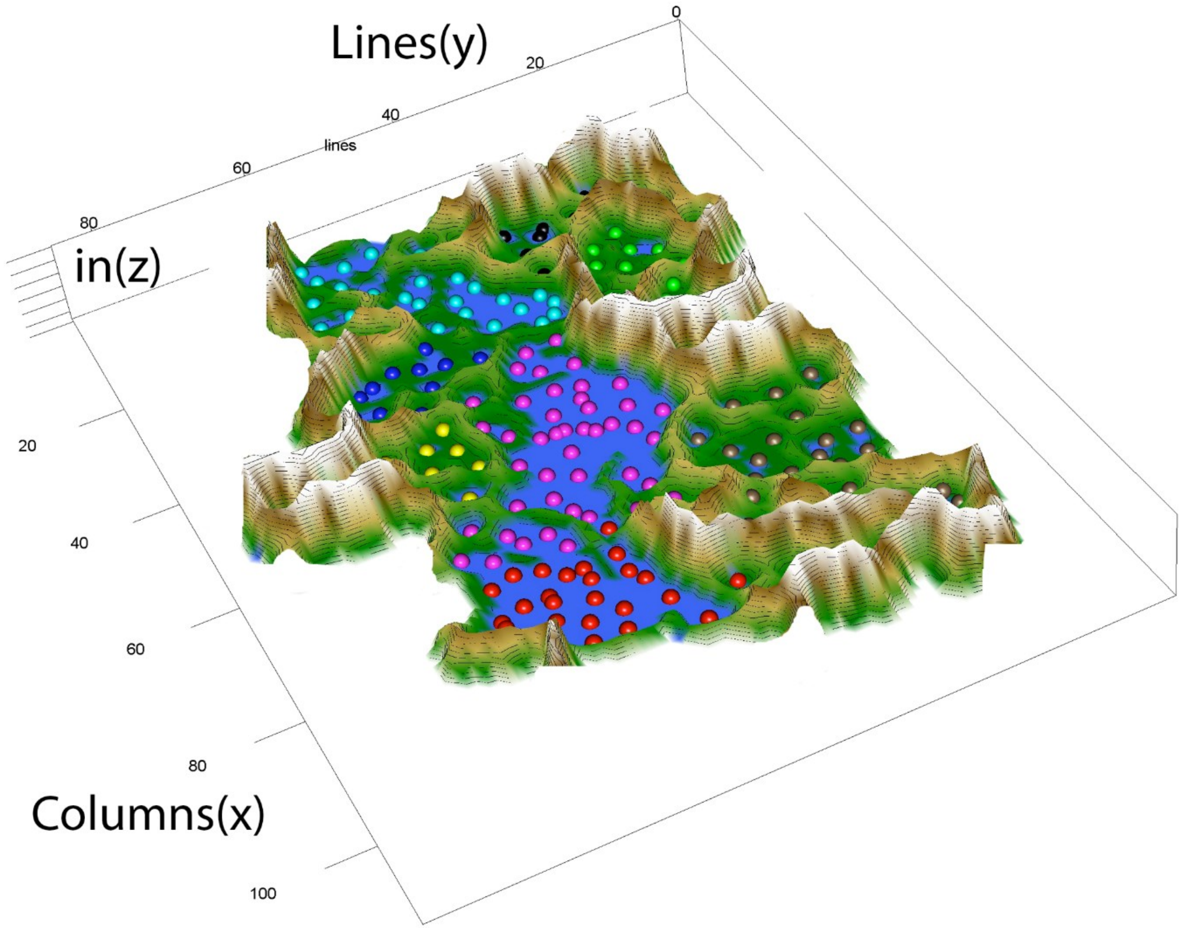
- Result: three main pain sensitivity groups [Lötsch/Ultsch, 2013]
  - High-pain sensitivity (HPS)
  - Average pain sensitivity (APS)
  - Low-pain sensitivity (LPS)
  - Subclusters...



Problem: complexity for domain experts

Complexity for domain experts

- 3D vizualizations [Jansen et al., 2013] cites [Shneiderman, 2003]
  - Have to be viewed from multiple viewpoints
  - Are often subject to serious occlusion and navigation issues



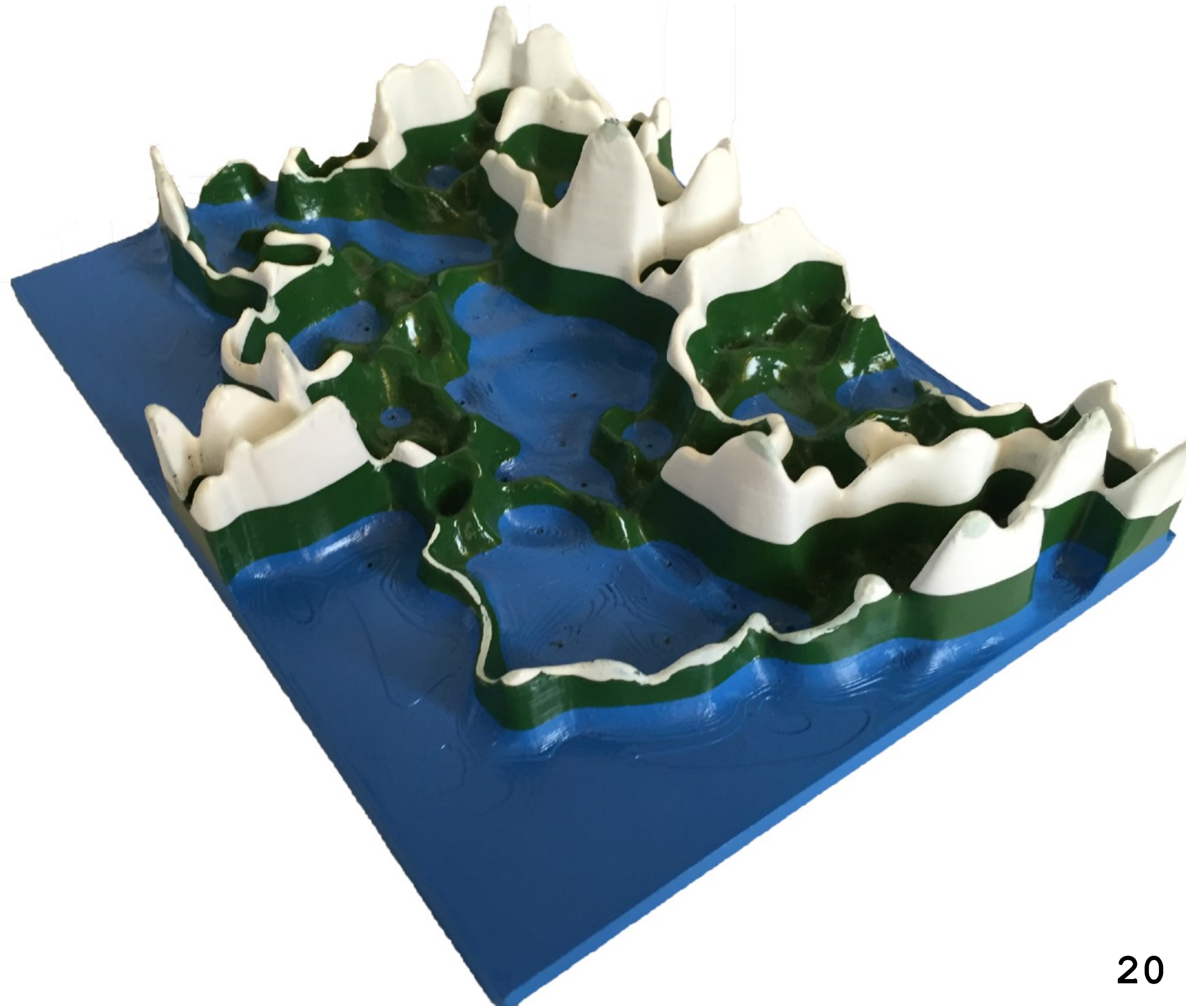
Solution: colored 3D printing

To our knowledge, first application of 3D printing techniques used directly for data mining and knowledge discovery

(+) Provide topical experts a haptic grasp of high-dimensional structures

(-) Technical limitation of height-dependent colors in 3D printers

(-) Automatically cutting a non-rectangular island defined by curved borders remains a problem

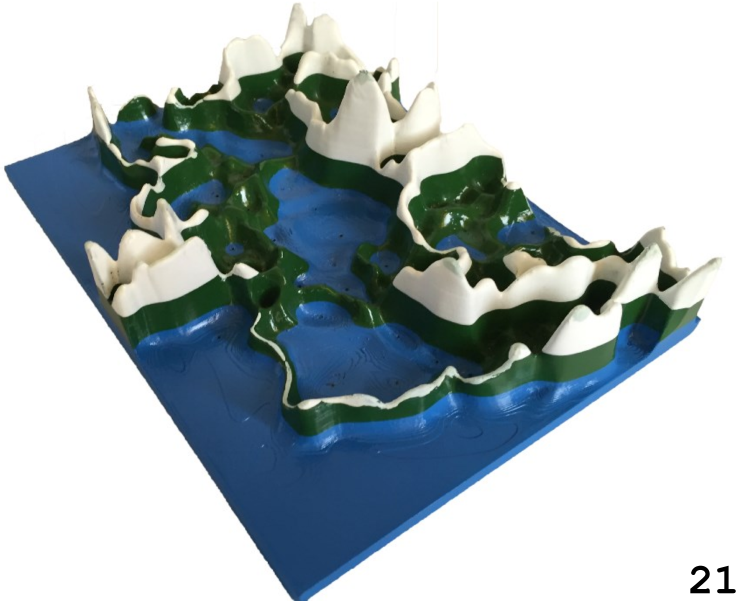


Summary

- Structure preservation: projections do not preserve all distances
  1. Calculate U*matrix to represent distance and density based structures
  2. Generate 3D landscape: topographic map with hypsometric tints
- Solve complexity for domain experts by colored 3D printing

For (1) and (2) use our freely availible R package Umatrix

[Version 2.0.0, www.uni-marburg.de/fb12/datenbionik/software]



Thank you for listening! Any questions?

Sources

Lee, J. A., & Verleysen, M.: *Nonlinear dimensionality reduction*, New York, USA, Springer, **2007**.

Ultsch, A.: *Maps for the visualization of high-dimensional data spaces,* Paper presented at the Workshop on Self organizing Maps (WSOM), Kyushu, Japan, **2003**

Ultsch, A., Behnisch, M., & Lötsch, J.: *ESOM Visualizations for Quality Assessment in Clustering*, Paper presented at the WSOM, Huston, USA, **2016.**

Ultsch, A., & Lötsch, J.: Computed ABC Analysis for Rational Selection of Most Informative Variables in Multivariate Data, *PloS one, 10*(6), e0129767. **2015.**

Ultsch, A., & Siemon, H. P.: *Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis*, Paper presented at the International Neural Network Conference, Paris, France, **1990.**

Shneiderman, B.: Why not make interfaces better than 3D reality?, *Computer Graphics and Applications, IEEE, 23*(6), 12-15. **2003**

Jansen, Y., Dragicevic, P., & Fekete, J.-D.: *Evaluating the efficiency of physical visualizations,* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, **2013**.

Tasdemir, K., & Merényi, E.: SOM-based topology visualisation for interactive analysis of high-dimensional large datasets, *Machine Learning Reports*. **2012.**

Lötsch, J., & Ultsch, A.: A machine-learned knowledge discovery method for associating complex phenotypes with complex genotypes. Application to pain, *Journal of biomedical informatics, 46*(5), 921-928. **2013.**

Flühr, K., Neddermeyer, T. J., & Lötsch, J.: Capsaicin or menthol sensitization induces quantitative but no qualitative changes to thermal and mechanical pain thresholds, *The Clinical journal of pain, 25*(2), 128-131. **2009.**