Pattern-based approach to geospatial datasets – methodology and software

J. Jasiewicz[1], P. Netzel[2], T. F. Stepinski[3]

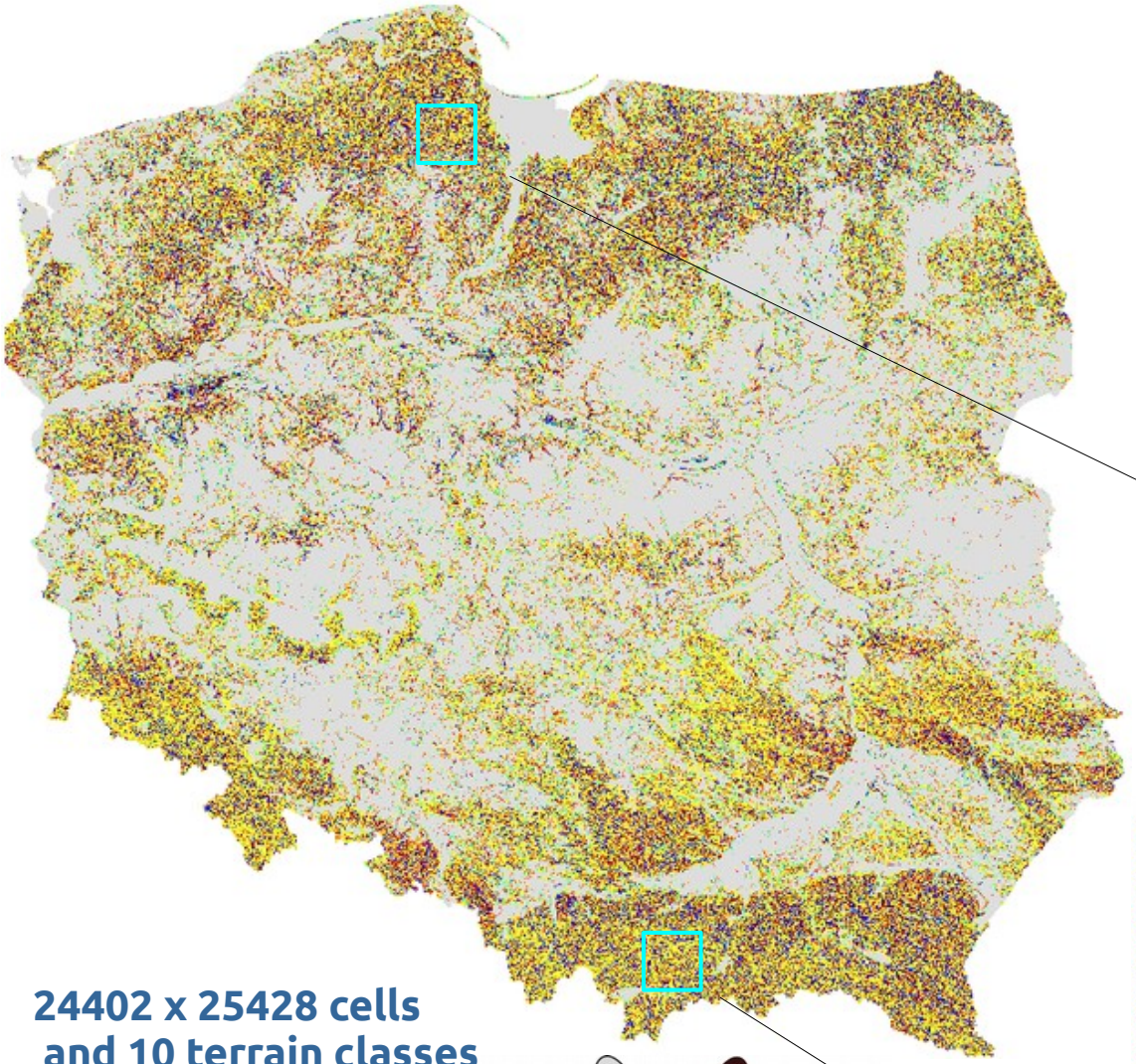
[1]Institute of Geoecology and Geoinformation, Adam Mickiewicz University, Dziegeilowa 27, 60-680 Poznań
Email: jarekj@amu.edu.pl

[2]Department of Climatology and Atmosphere Protection, University of Wroclaw, Kosiby 6/8, 51-621 Wrocław, Poland
Email: pawel.netzel@uni.wroc.pl

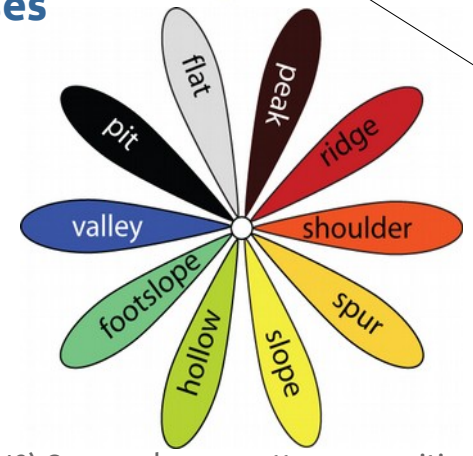
[3]Space Information Lab, University of Cincinnati 401 Braunstein Hall, 45221 Cincinnati, US
Email: stepintz@uc.edu

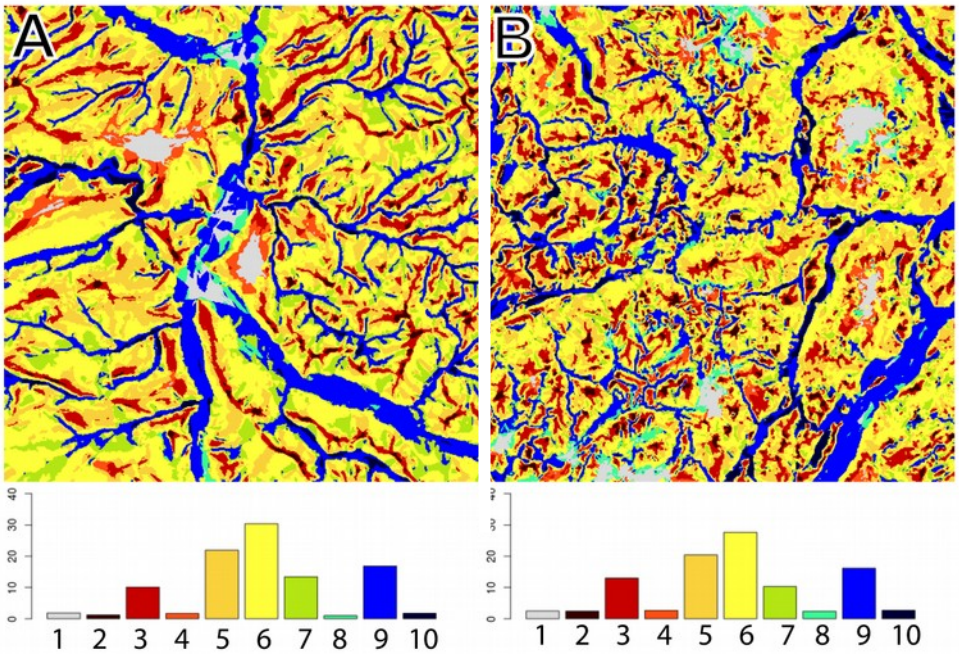
What is a pattern?

Why patterns?



24402 x 25428 cells and 10 terrain classes

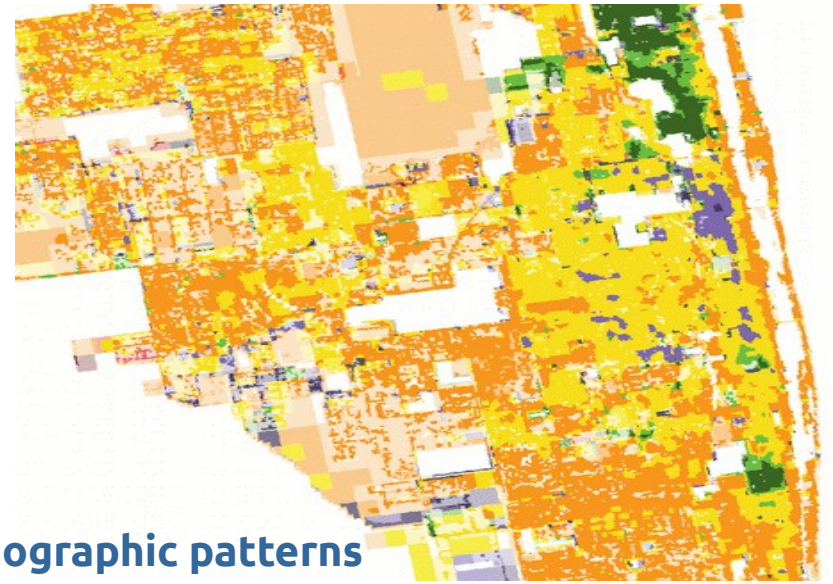




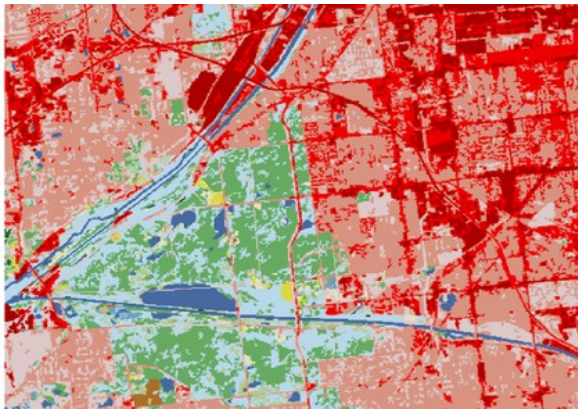
J. Jasiewicz and T.F Stepinski (2013) Geomorphons -a pattern recognition approach to classification and mapping of lanforms. Geomorphology 182, pp. 147-15

When patterns are important?

Patterns in remotely sensed RGB images



Geodemographic patterns



Patterns in Landcover maps

Patterns take into account not only asssamblage of categories but also its spatial arrangement

How to do GIS with Patterns?

- How to describe?

  (Present in a form available for further processing)
- How to attribute?

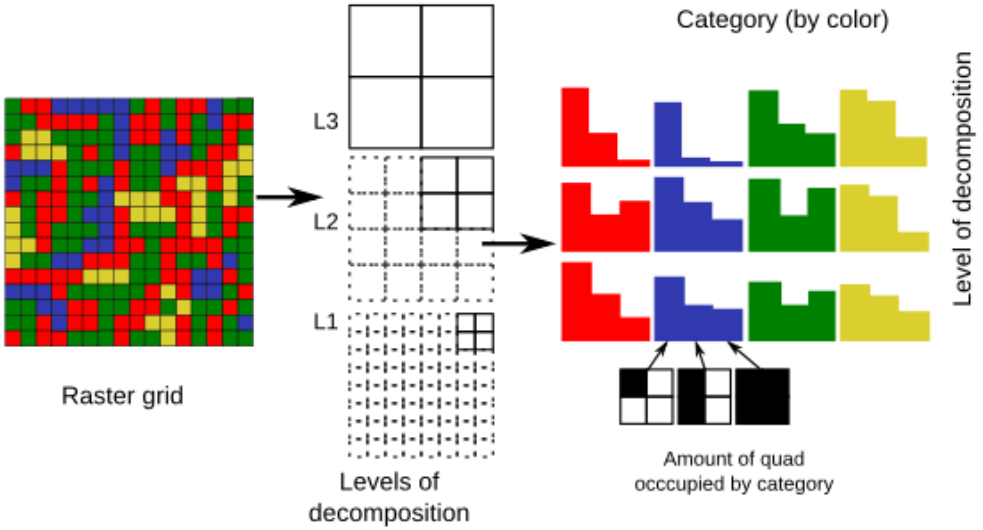
  (Locate pattern's signature in geographical space)
- How to process?

  (Include patterns into geoprocessing routines)

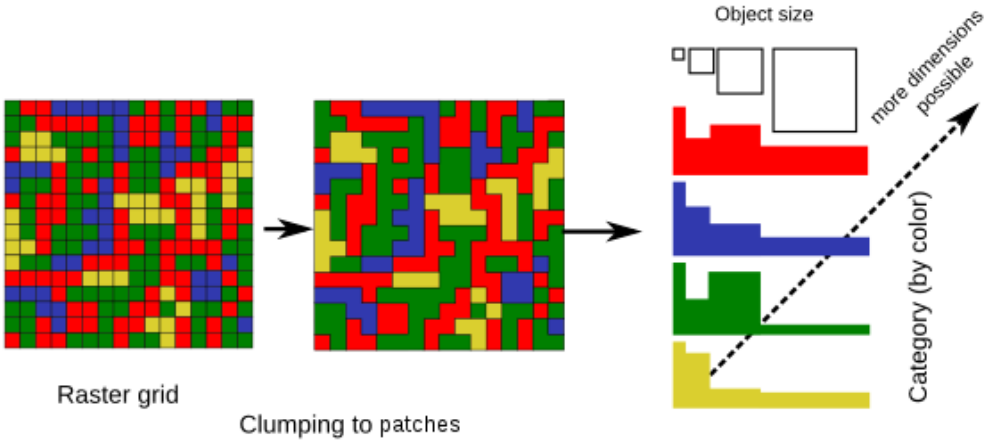
- Do we have appropriate software? **(YES!)**
- Does it really good for anything? **(Examples!)**

Pattern description. What is a signature?

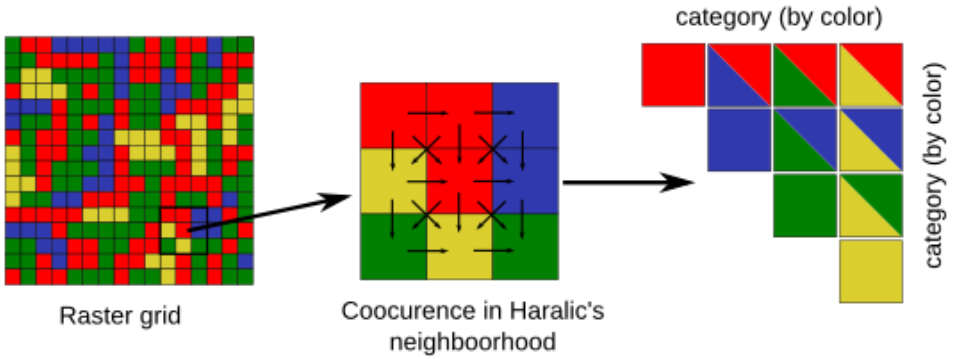
Hierarchical decomposition

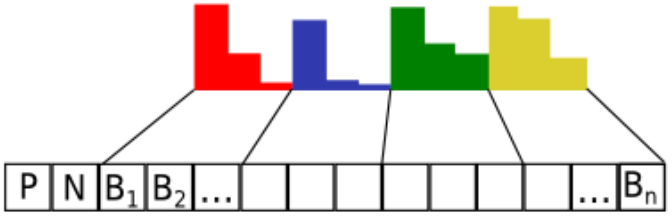


Crossproduct of categories and size



Co-occurrence of categories

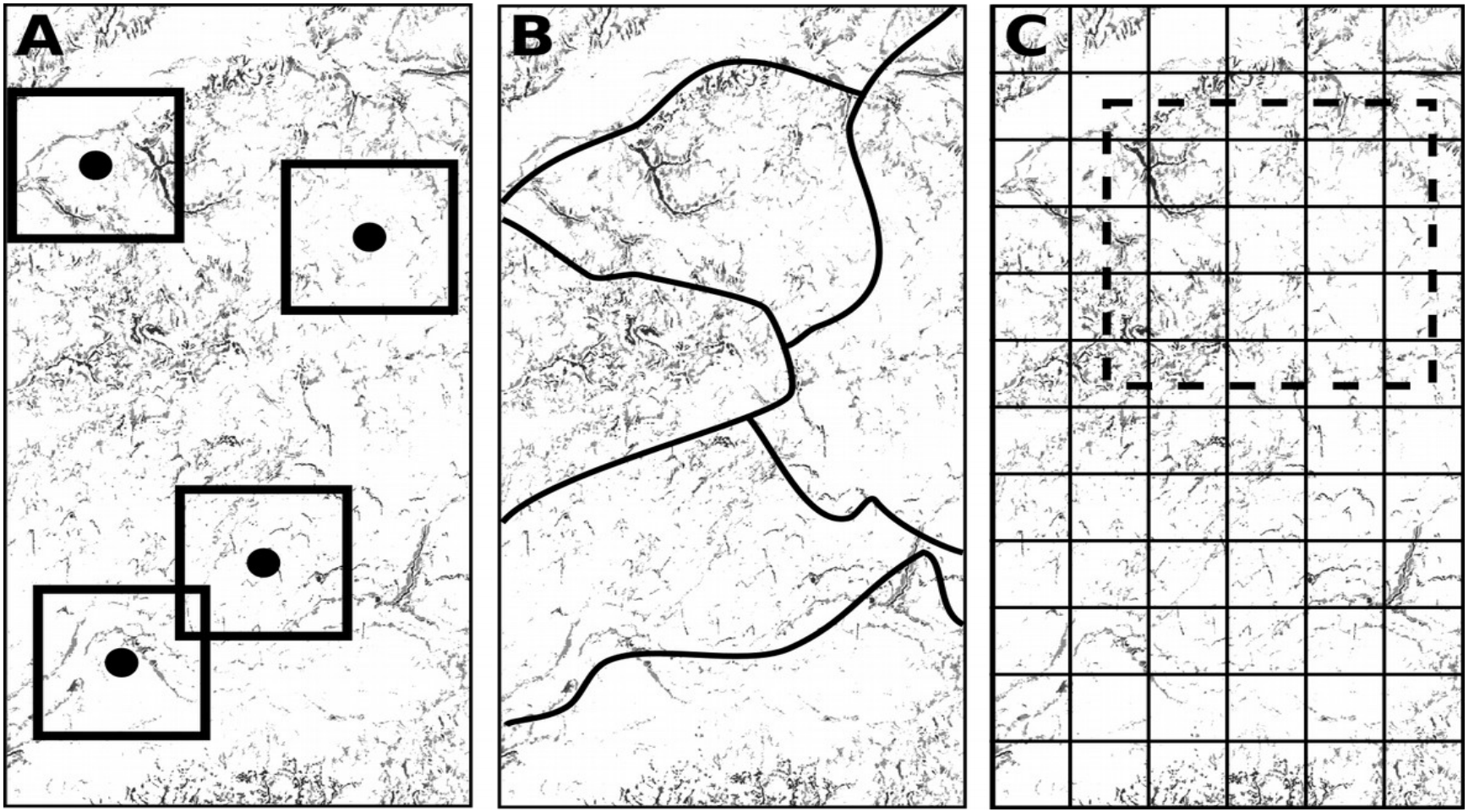




Signature is a histogram transformed to PDF so sum(bins)=1

Signatures are rotationally invariant !

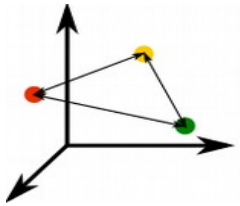
How we put signatures into geographical space?



Points **Polygons** **Grid**

Signatures are calculated for scenes. Scene is our basic data structure

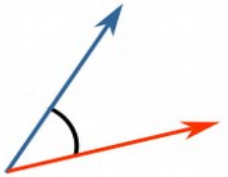
similarity/distance measures



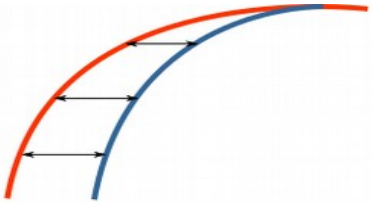
1

2

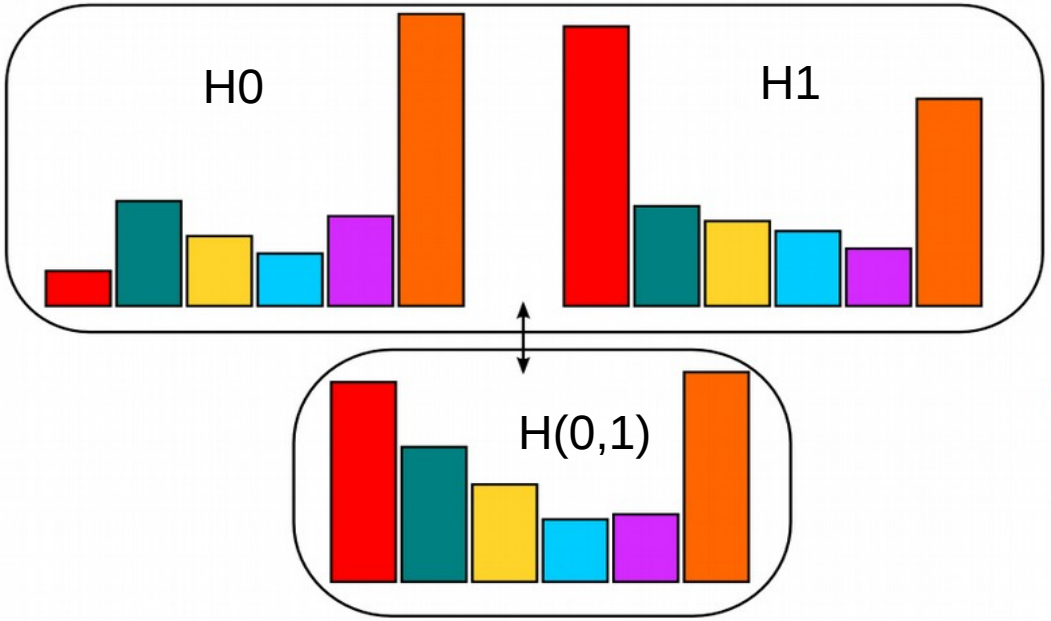
3



4

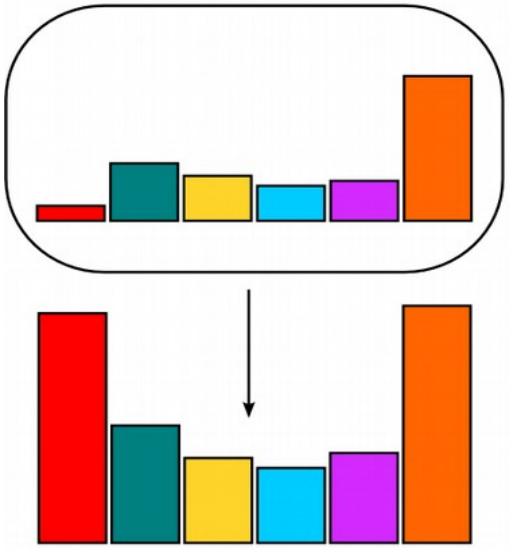


5



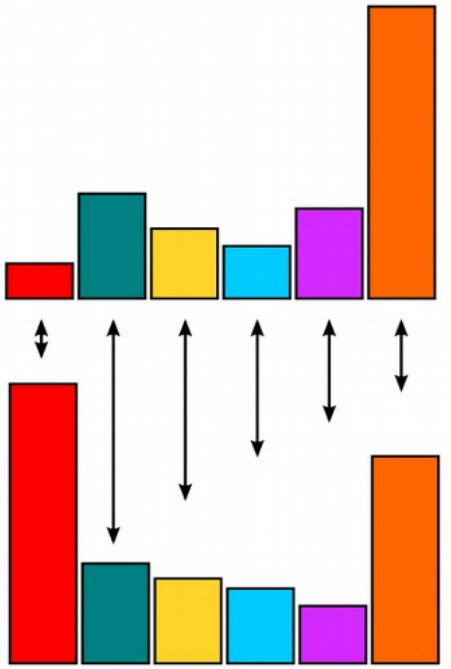
Jensen-Shannon Divergence

$$d_{JSD} = \sqrt{\sum_{i=1}^{d}\left[\frac{P_i log_2 P_i + Q_i log_2 Q_i}{2} - \left(\frac{P_i + Q_i}{2}\right) log_2\left(\frac{P_i + Q_i}{2}\right)\right]}$$



Jaccard

$$d_J = 1 - \frac{\sum_{i=1}^{d} P_i Q_i}{\sum_{i=1}^{d} P_i^2 + \sum_{i=1}^{d} Q_i^2 - \sum_{i=1}^{d} P_i Q_i}$$

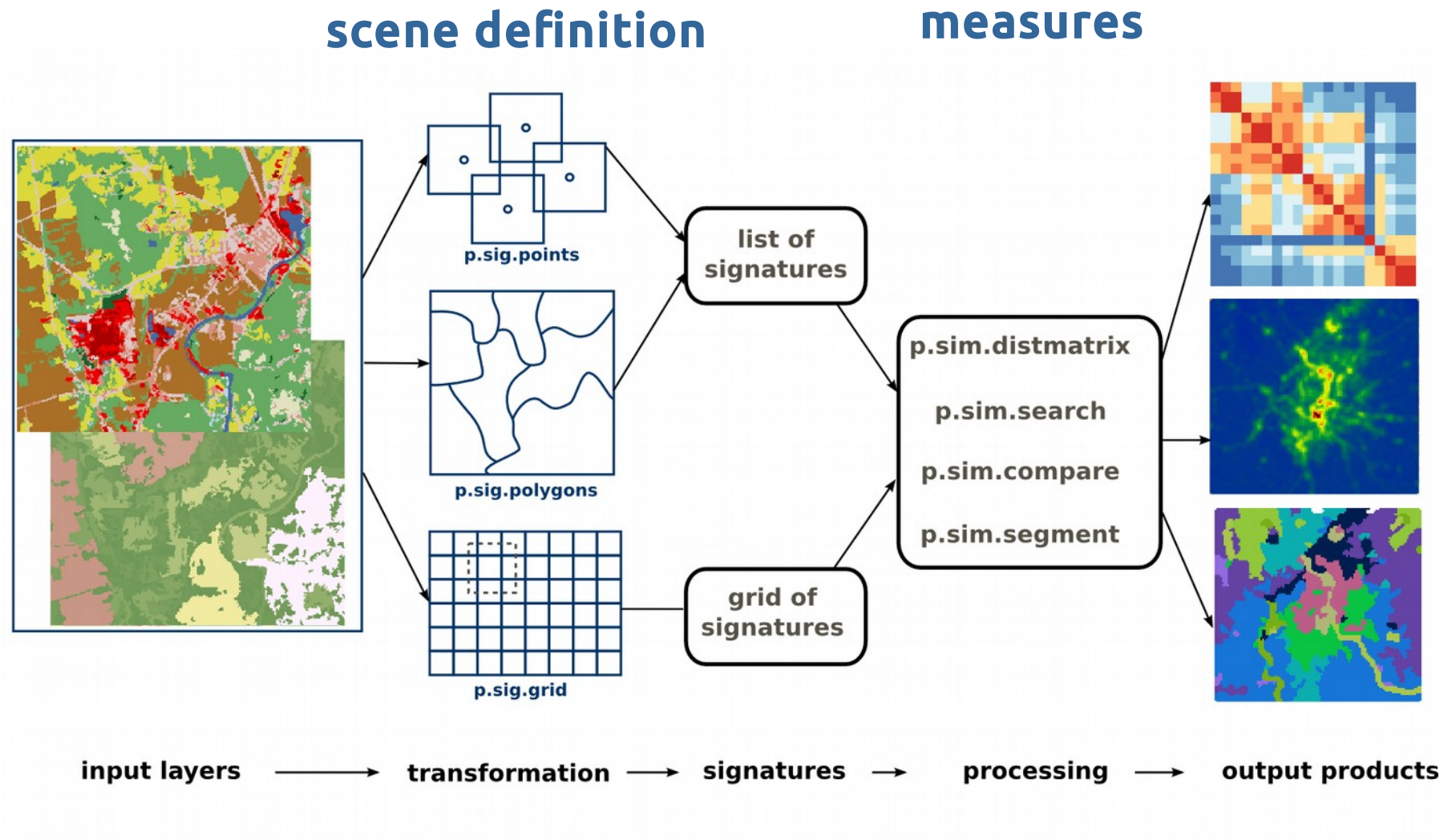


Wave Hedges

$$d_{WH} = \sum_{i=0}^{d} e_i \frac{|P_i - Q_i|}{max(P_i, Q_i)}$$

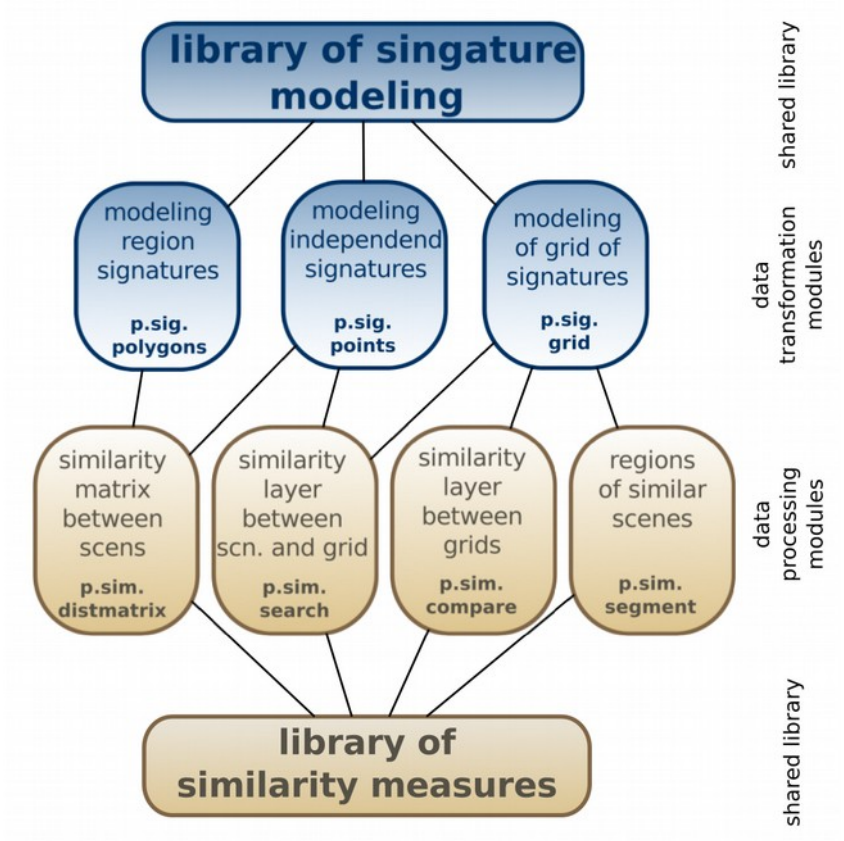
similarity=1-distance

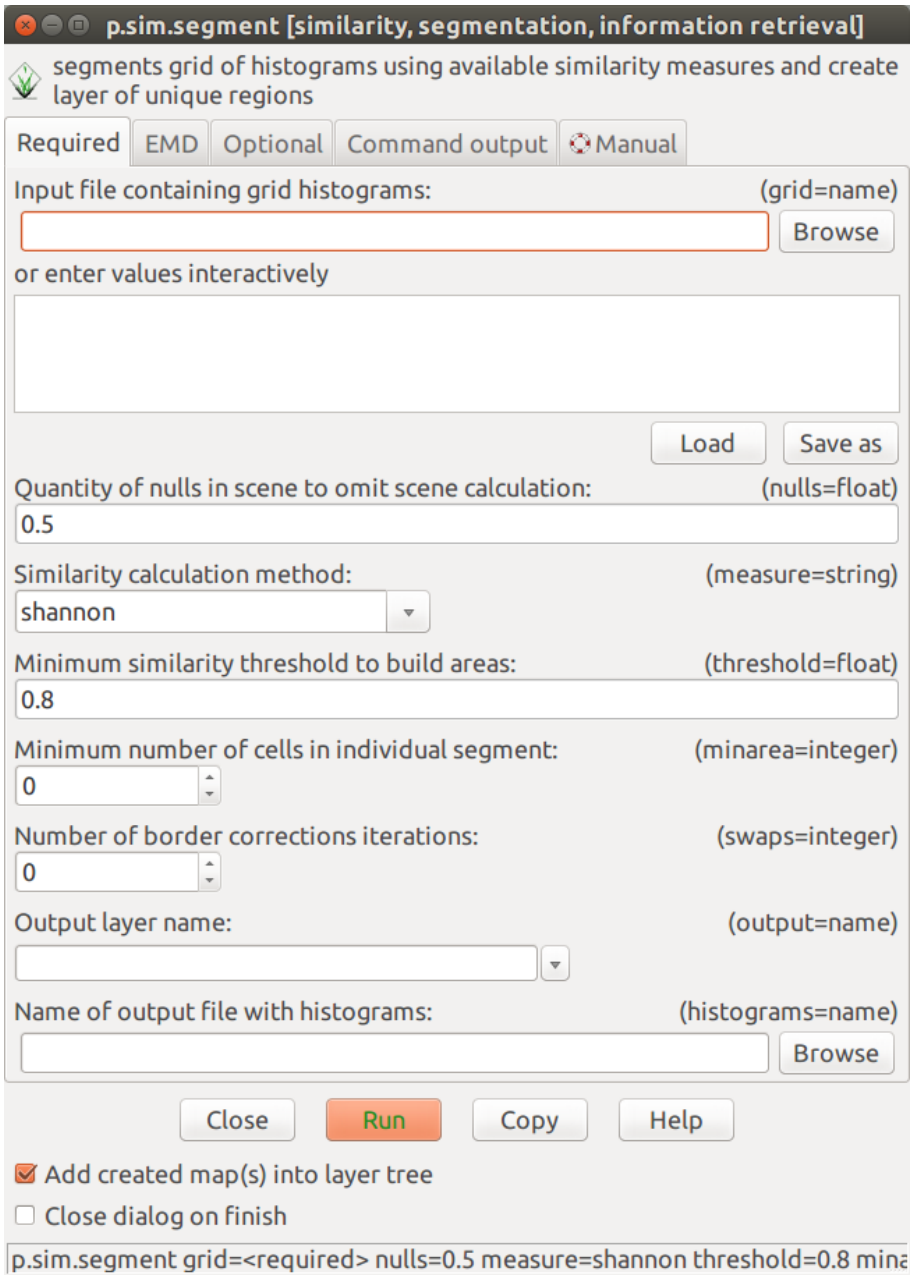
Data processing: raster, scene, process, results



If signature is an attribute – similarity measure becomes essential for the query!

The software



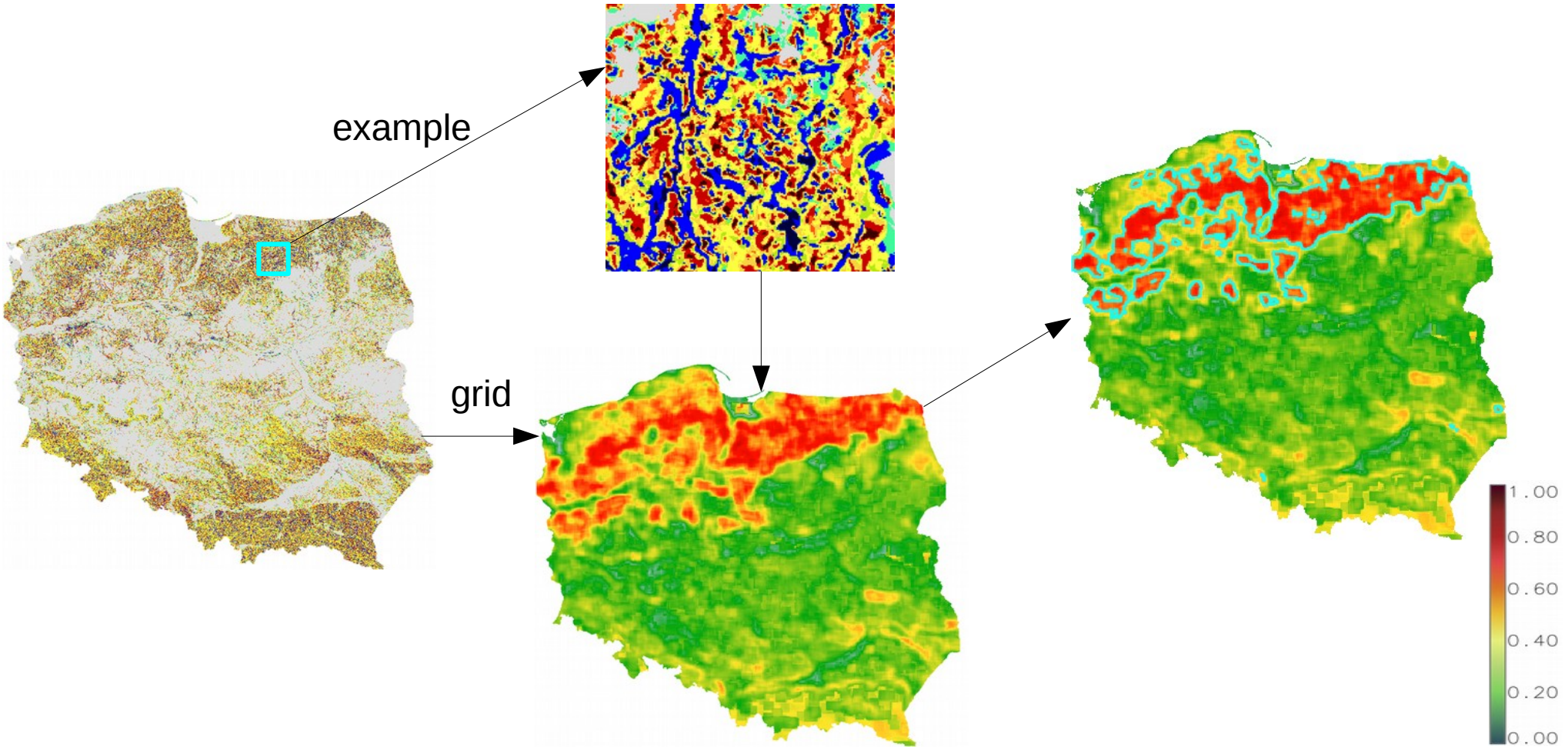


GRASS GIS

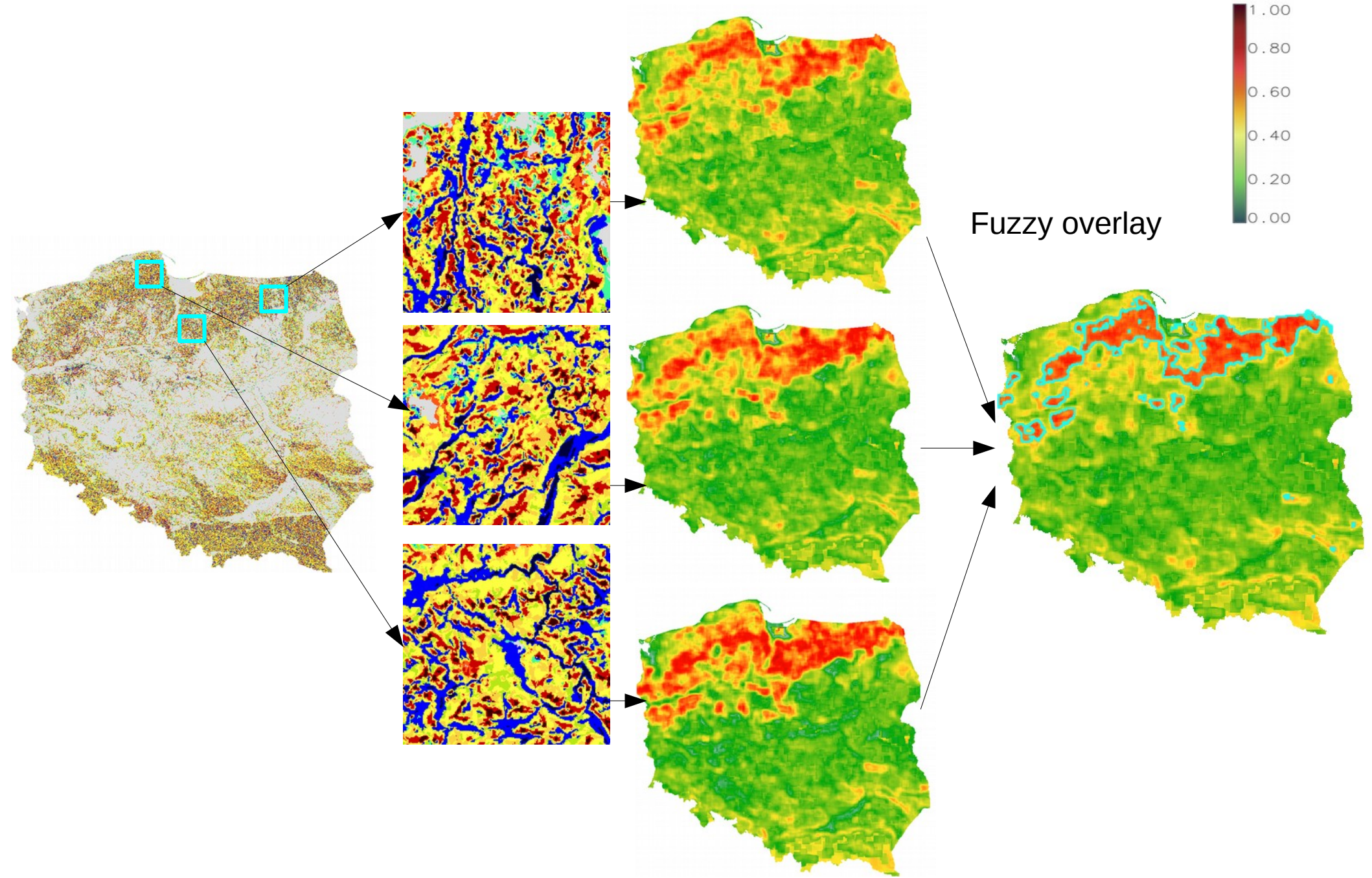
The world's leading Free GIS software

Searching (Query)

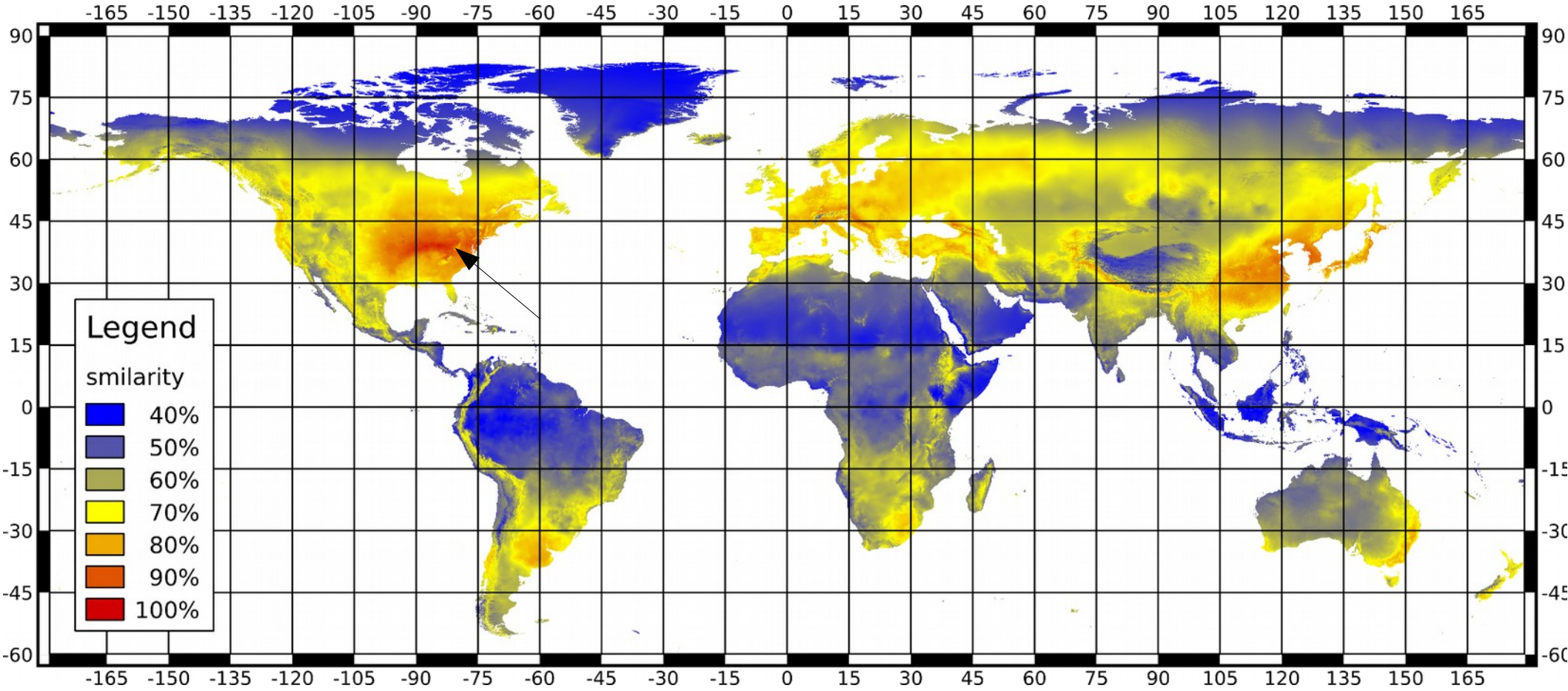
Query-By-Visual-Example (Computer Vision)



Fuzzy search



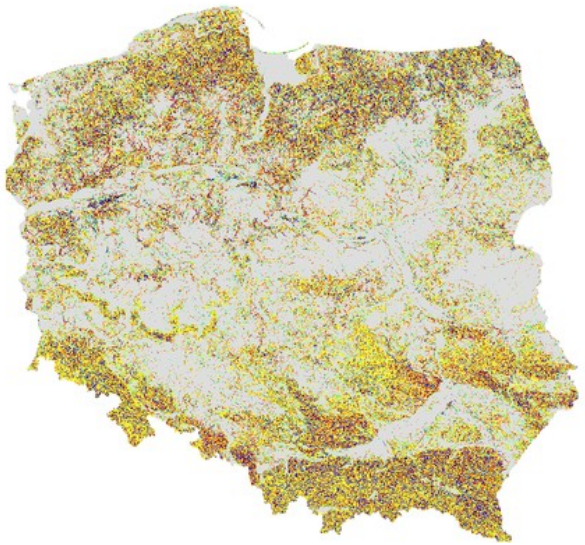
World-wide climatic data: search



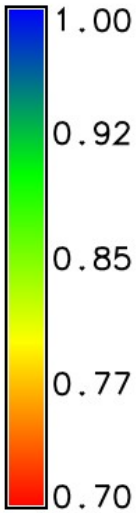
Grid of 43000 x 18000 scenes each containing temporal pattern

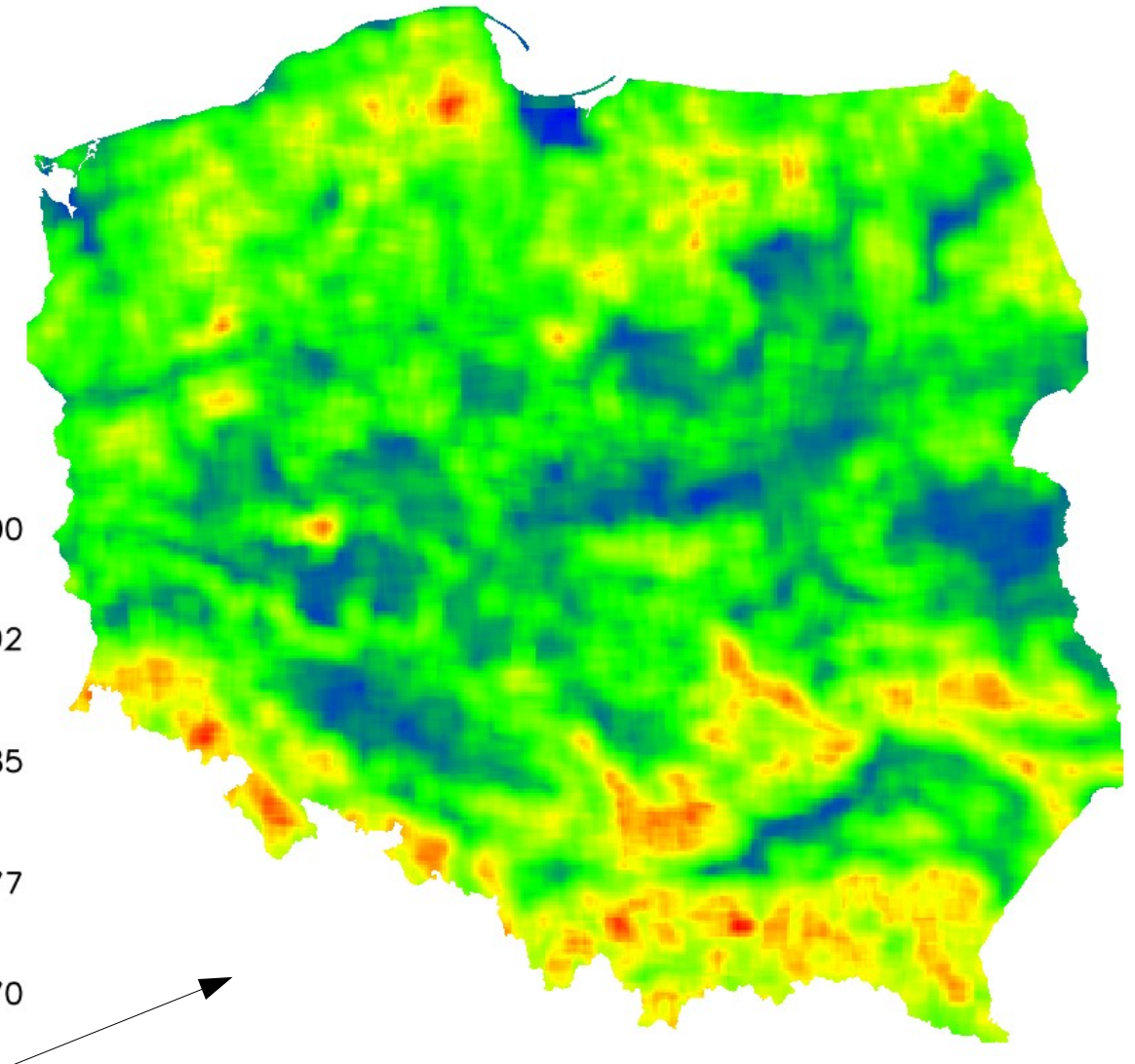
Comparison: differences between two models

smaller search radius

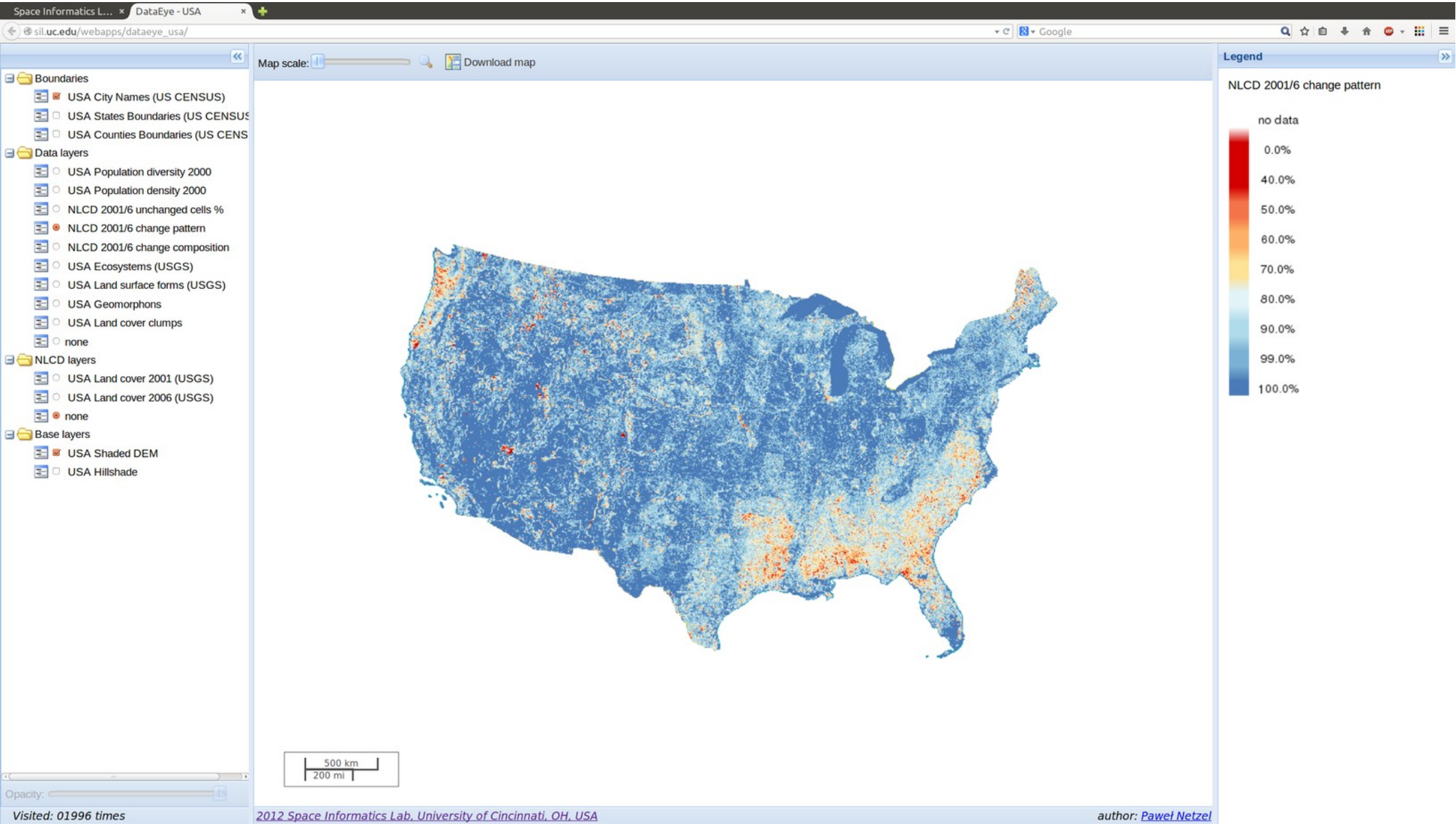


bigger search radius





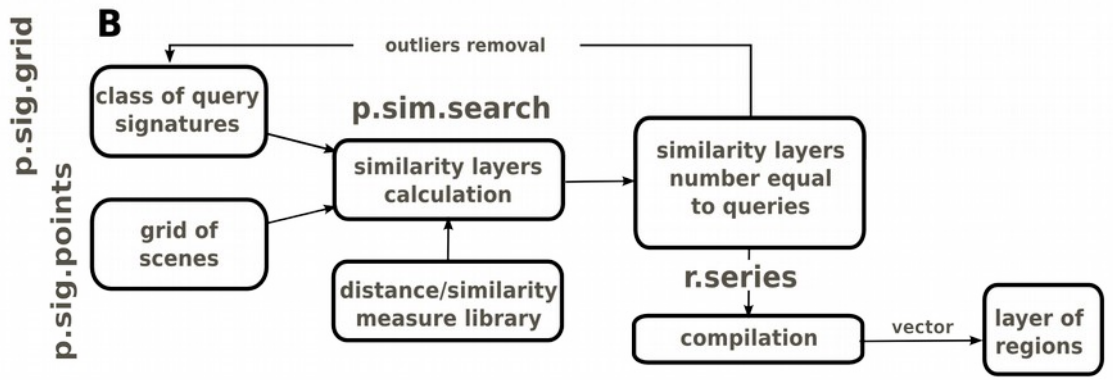
Comparison: continental scale change detection



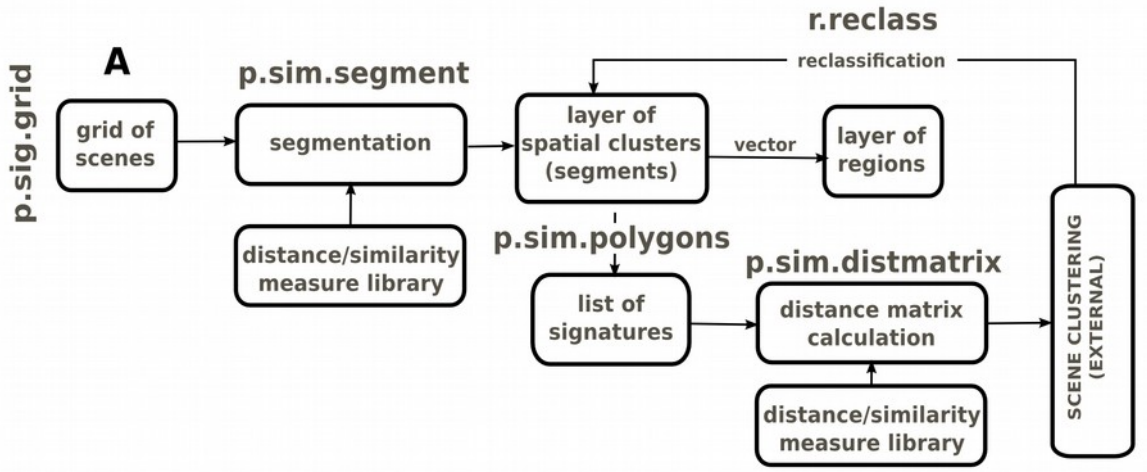
pattern change between NLCD 2001 and NLCD 2006

P. Netzel and T.F. Stepinski (2014) Pattern-based assessment of land cover change on continental scale with application to NLCD 2001-2006. IEEE TGRS

Supervised and unsupervised classifications

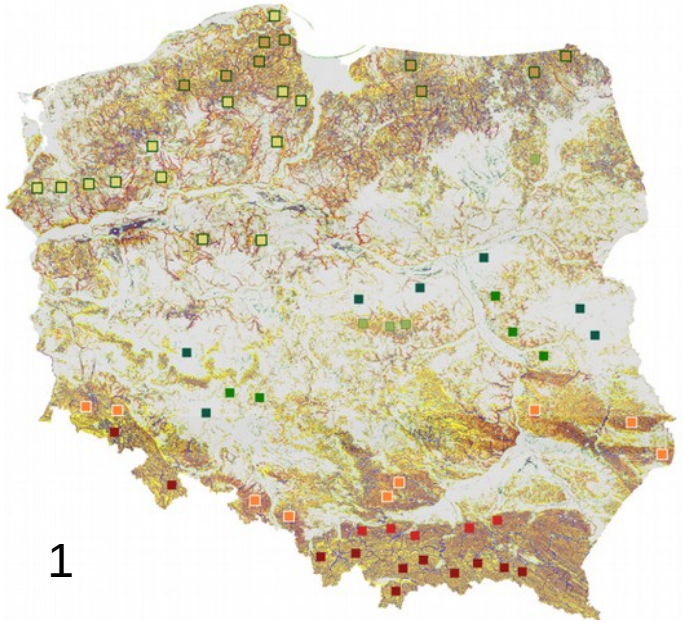


Sample-based supervised classification



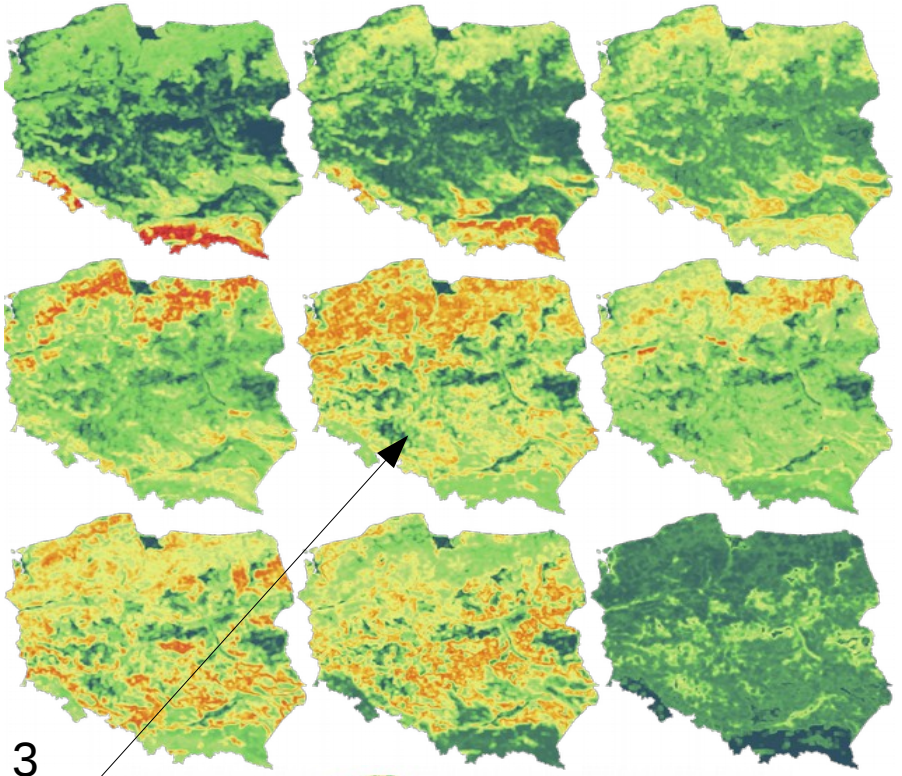
Segmentation-based unsupervised classification

Supervised classification – fuzzy search and class overlay

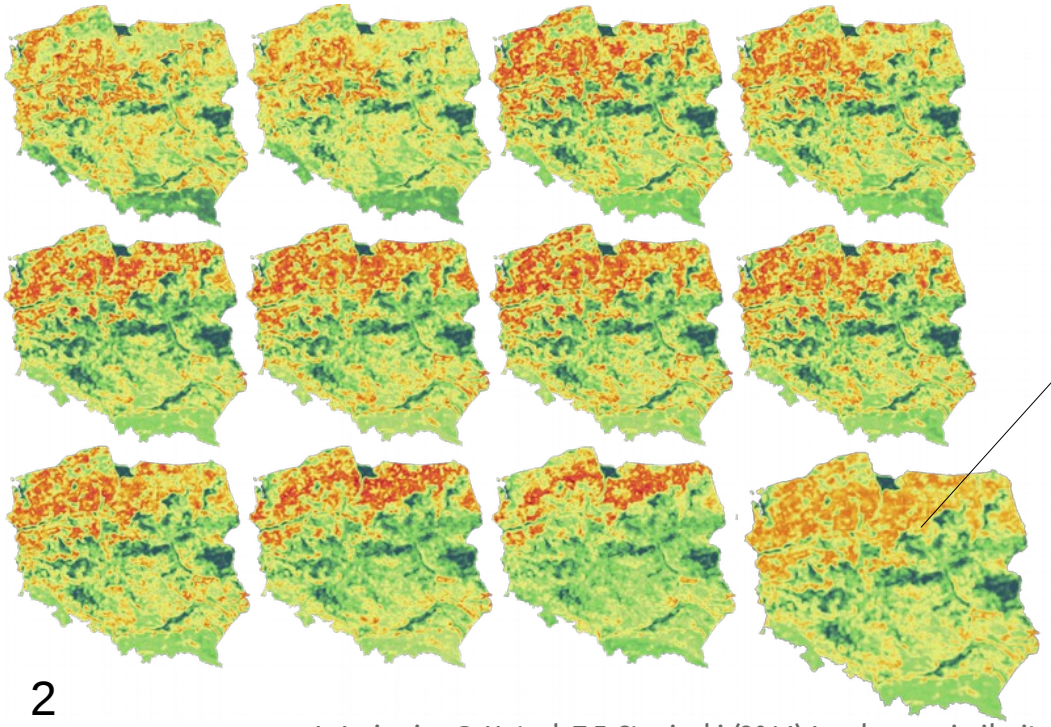


1

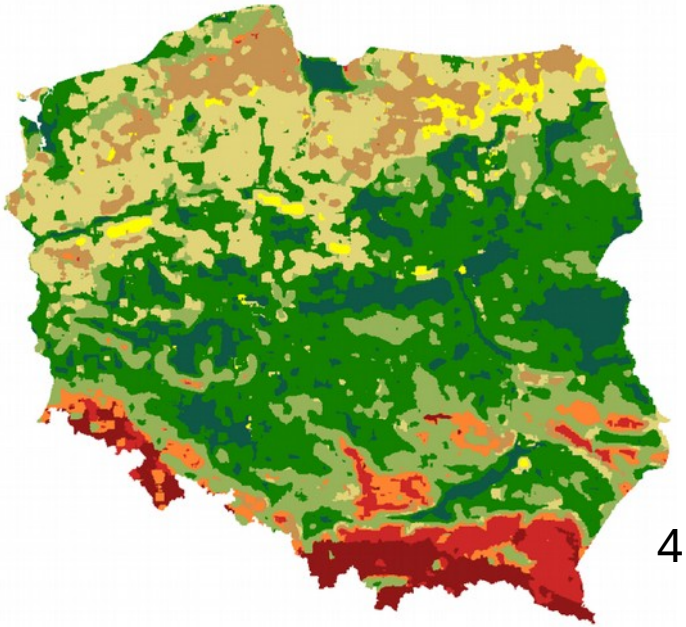
FLP	Flats plains
OMP	Old morainic plains
OMPH	Old morainic plateaus and hills
YMP	Young morainic plateaus and plains
DN	Inland dunes
YMH	Young morainic hills
HL	Highlands and low mountains
LM	Low mountains
MM	Medium mountains



3



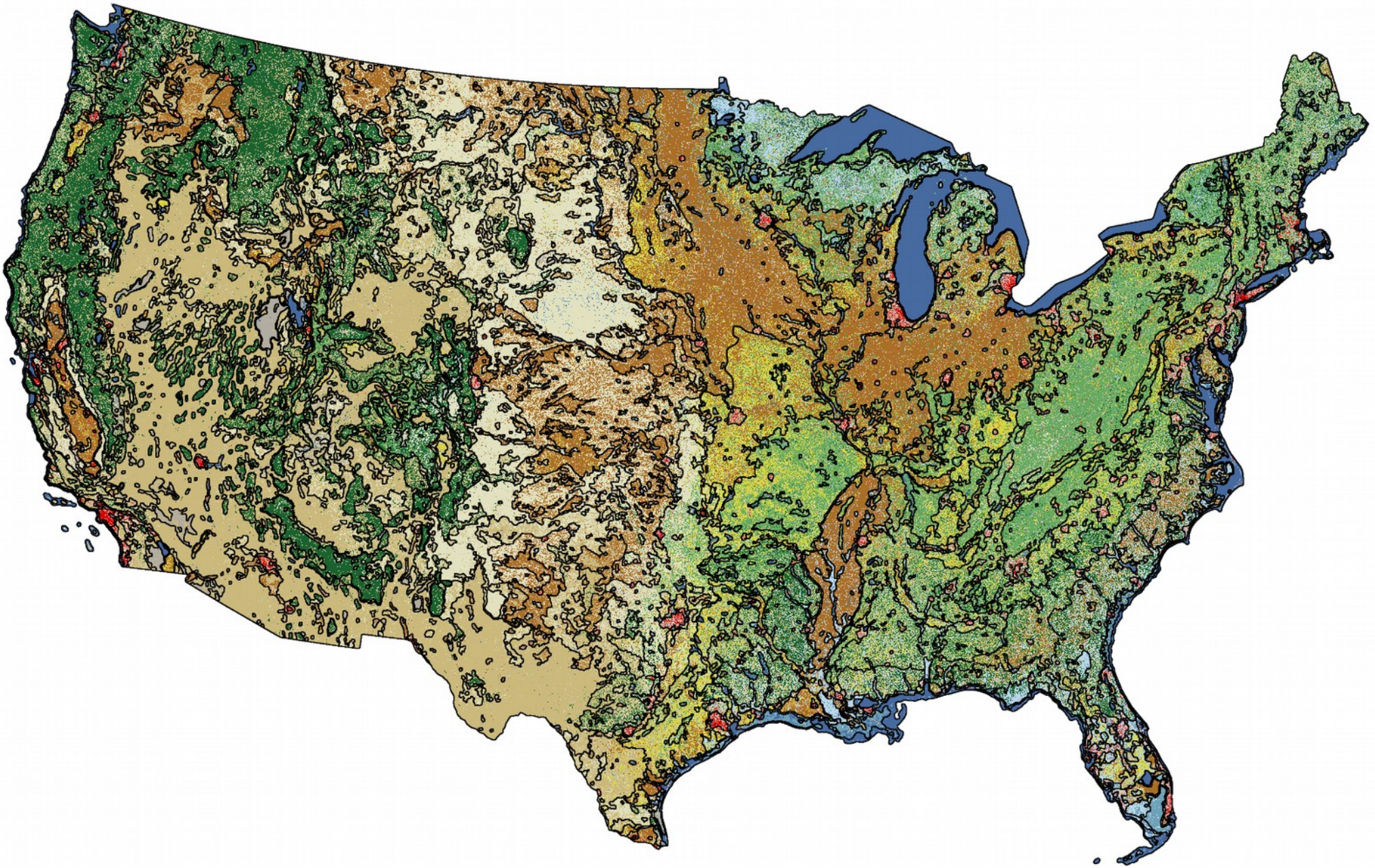
2



4

J. Jasiewicz, P. Netzel, T.F. Stepinski (2014) Landscape similarity, retrieval, and machine mapping of physiographic units. Geomorphology 221, pp.104-112

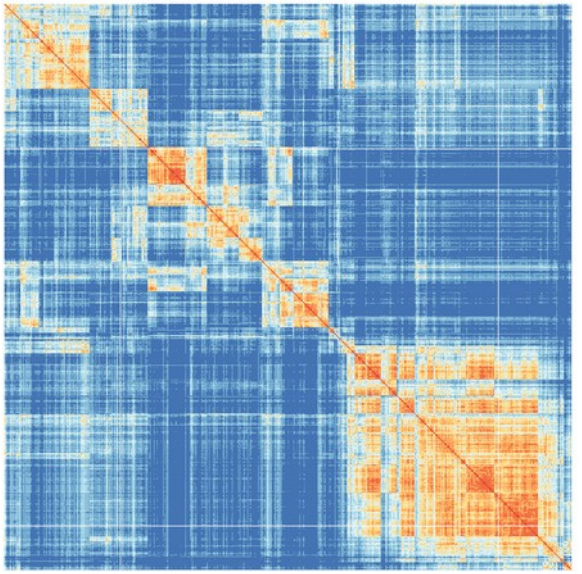
Segmentation of entire US



164000 x 104000 (30 m) US land-cover segmented into 15362 areas with similarity threshold at 0.65

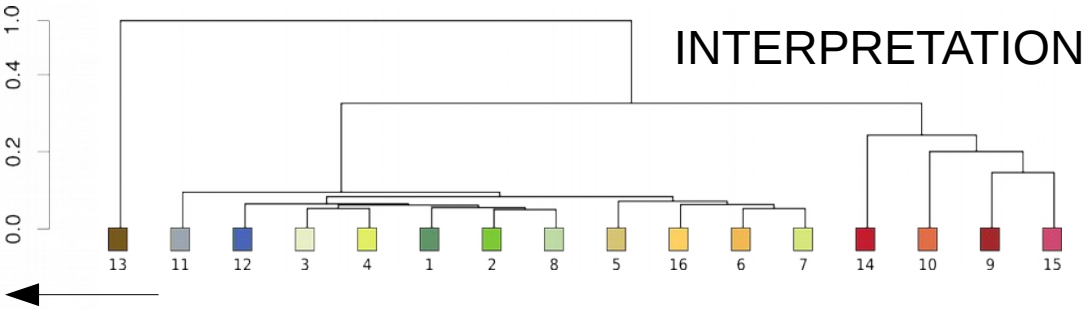
Unsupervised classification

HISTOGRAMS
FOR SEGMENTS



DISTANCE
MATRIX

15362 segments...

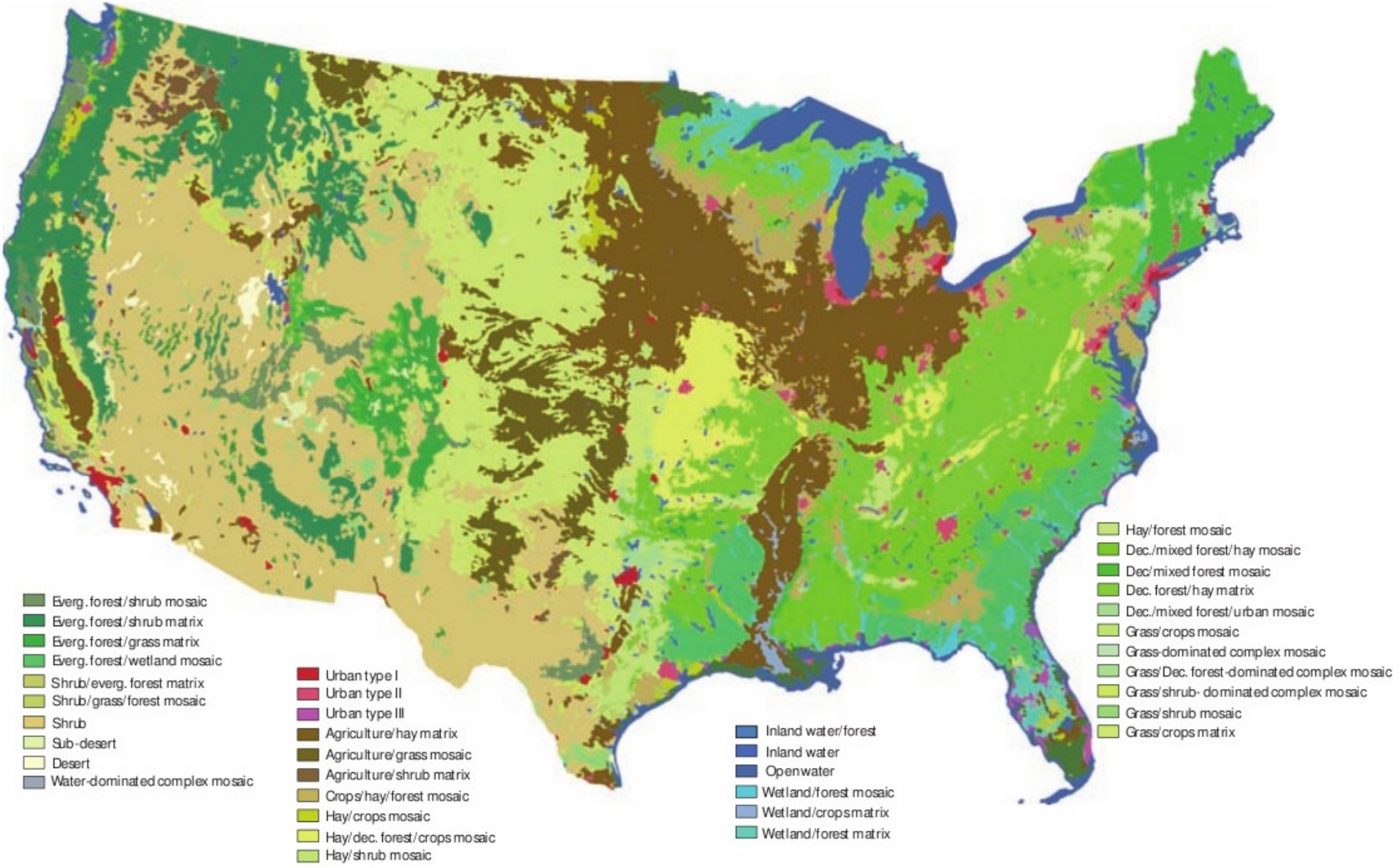


INTERPRETATION

(Size of the tree is reduced)

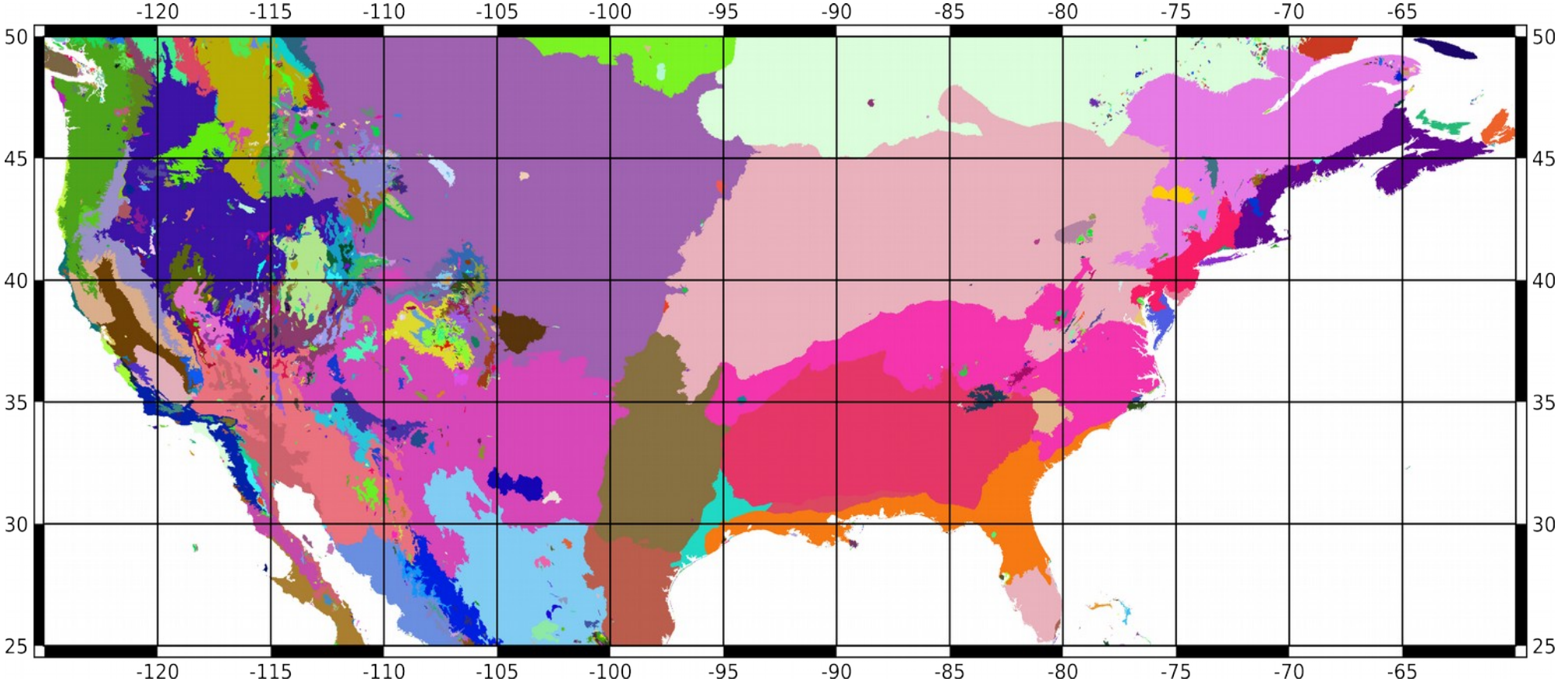
APPLICATION

Unsupervised classification: result

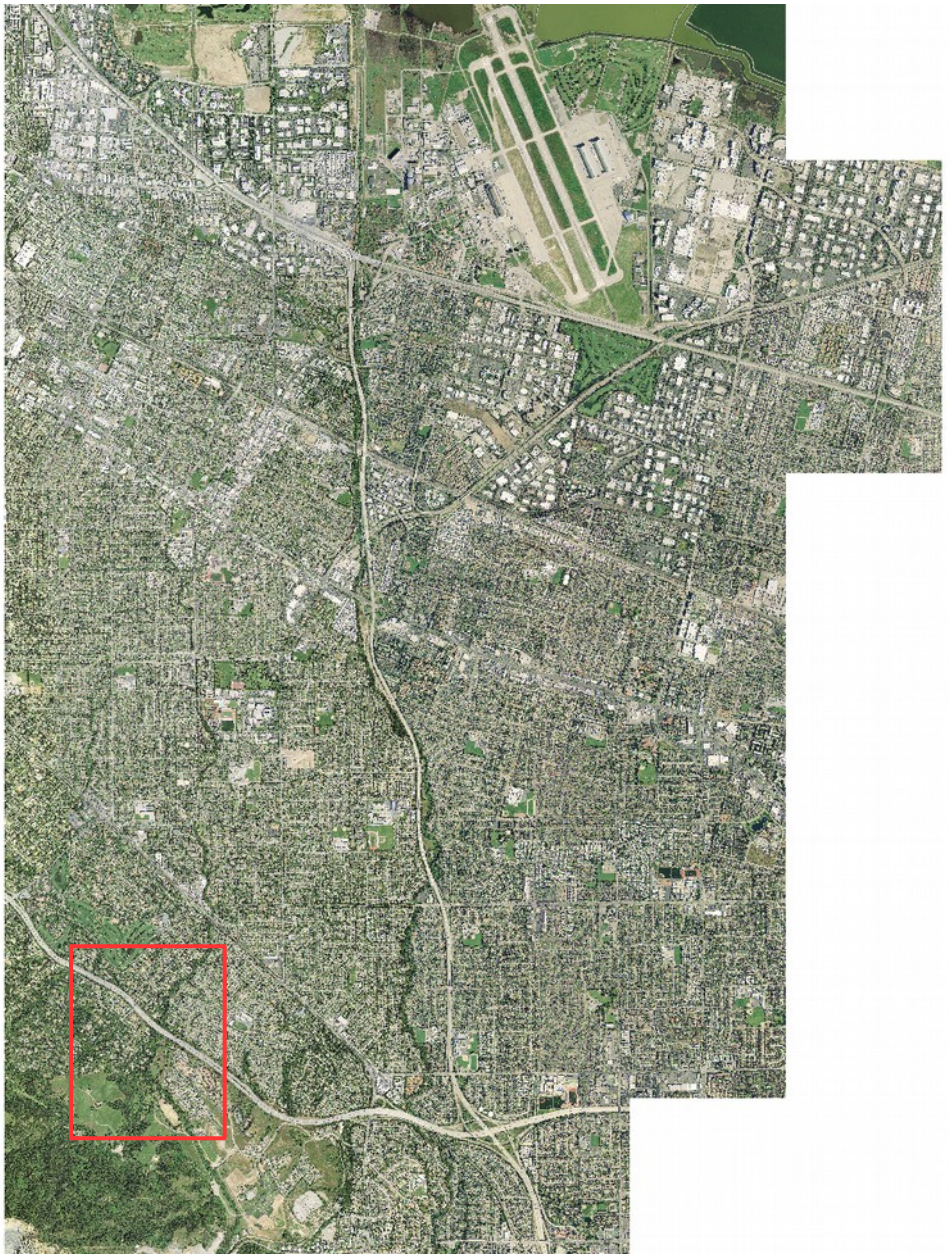


... reduced to 38 most popular landscapes

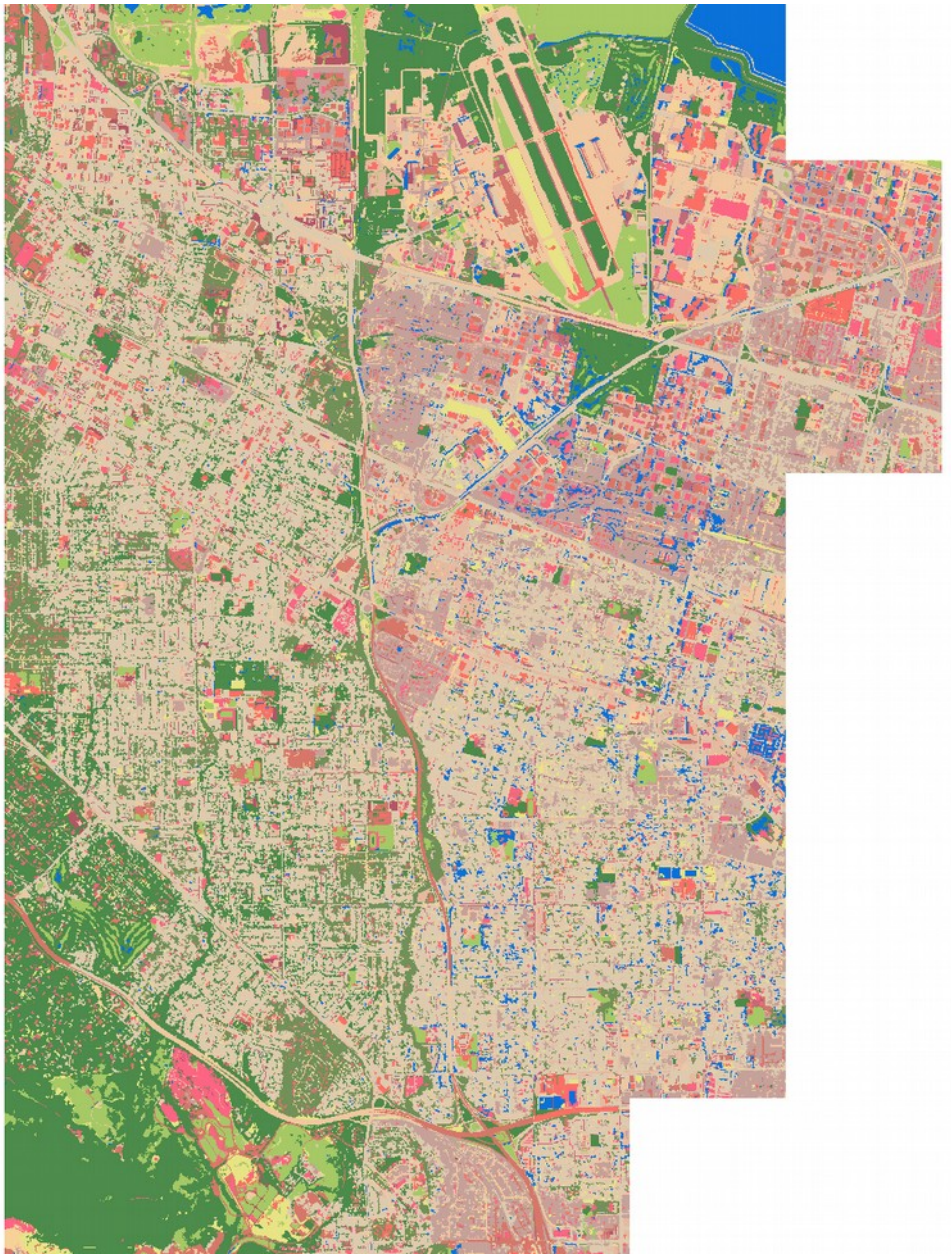
World-wide climatic data: segmetation



First attempt to segmentation of HR remotely sensed images (AVIRIS)

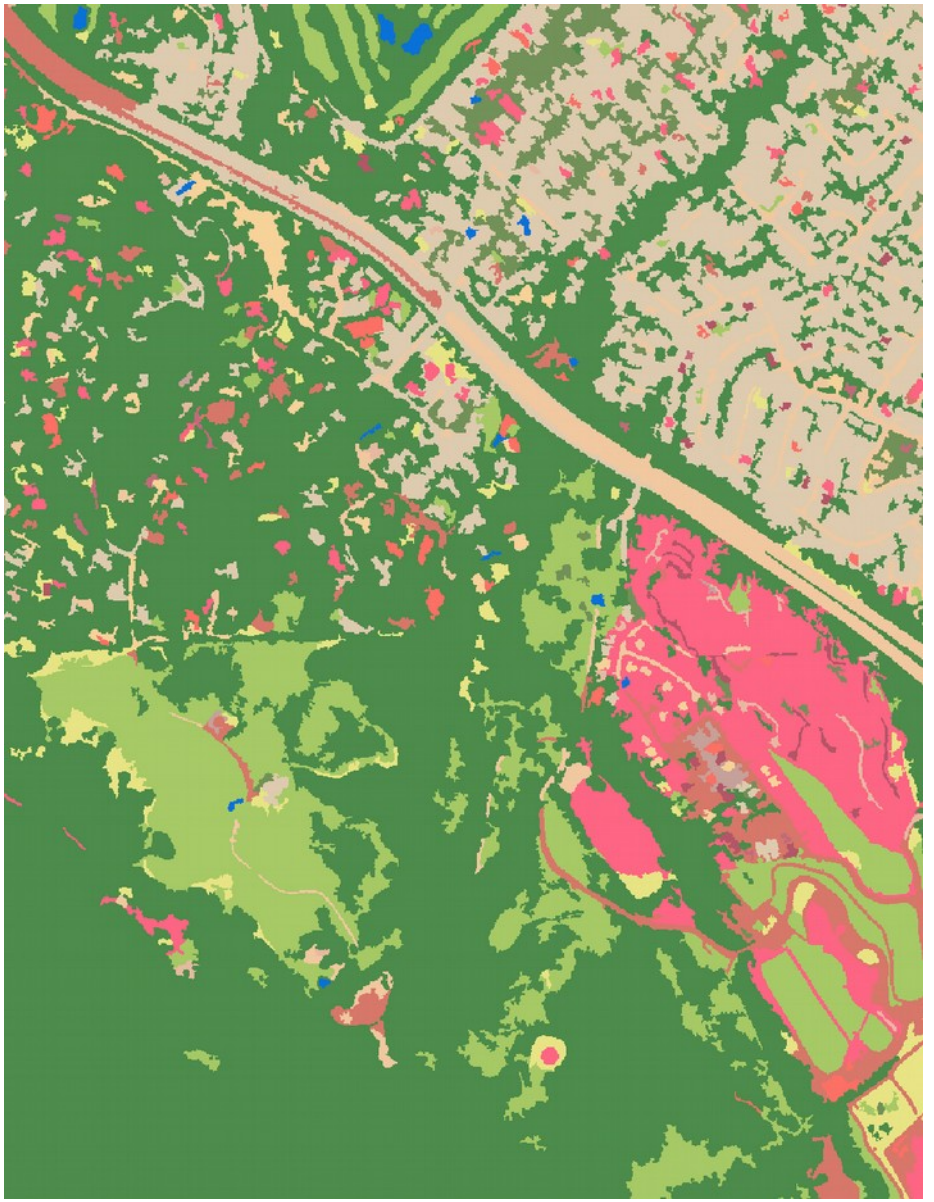


30000 x 40000 pixels RGB image



5000 x 3750 scenes classified into 15 classes

AVIRIS – ZOOM IN



Performance and conclusions

TASK:	TIME (i7, 16 cores, 20GB RAM)
Raster (164000x104000) to P-GRID (NLCD)	~ 3 min (multi-thread)
Raster (40000x3000) to P-GRID (AVIRIS)	~ 40 min classification + 1 min HGRID
P-GRID to SEGMENTS (5125x3250) (NLCD)	~ 1 min 20 sec.
P-GRID to SEGMENTS (10000x7500) (AVIRIS)	~ 2 min 36 sec.
P-GRID to SEGMENTS (911x822) (POLAND)	~ 40 sec.
SEARCHING (NLCD)	~ 1 sec.
COMPARING (NLCD)	~ 1 sec.

A new software which can work transparently with patterns in GIS routines

Software can work fast with huge datasets (“Big Data”)

We can extract information unavailable using standard GIS routines

Authors would like to thank to
Jacek Niesterowicz
and
Anna Domowska
For their tremendous help

http://sil.uc.edu

Landex, Terraex, Socscape, dataeye, geomorphons

This work was done under NCN GRANT
2012/07/B/ST6/01206
and
University of Cincinnati Space Information Lab
Founds