

# OPEN SPAT



Erasmus+



INSTITUTO  
SUPERIOR DE  
AGRONOMIA  
*Hinc  
Patriam  
Sustinet*  
Universidade de Lisboa

SupAgro Montpellier

# Pattern recognition on spatial data

## Introduction

Pr Yves Brostaux, GxABT, University of Liege (Belgium)

OpenSpat, Gembloux

June 2017



Gembloux Agro-Bio Tech  
Université de Liège

## Pattern recognition

= recognition of patterns, regularities and hidden structures in data

- different goals
  - formalize, explain and visualize the pattern
  - recognize and predict a pattern
- linked to *machine learning*
  - supervised learning (pre existing structure)
  - unsupervised learning (no pre existing structure)



# Machine learning

So called "Machine learning" needs an heavy human expertise input :

- data selection
- methods selection
- parameters selection

⇒ as most of data modelling, pattern recognition is between **art** and **science**



Gembloux Agro-Bio Tech  
Université de Liège

# Tools and methods

Useful data in pattern recognition are often **highly multidimensionnal**, so that we need tools to

- Explore and visualize
- Summarize
- Regroup
- Predict

Available tools for pattern recognition of spatial data are mostly the same than for "classical data".

**But** spatial information help supporting human decisions and interpretations **and** raise some new questions about *individuals*



Gembloux Agro-Bio Tech  
Université de Liège

# Modifiable areal unit problem

Classical pattern recognition tools search for structures among **individuals**.

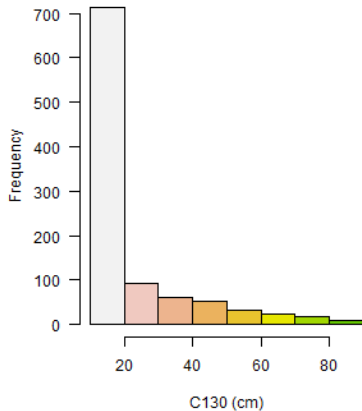
But the definition of an individual is particular with spatial data, as spatial data can be aggregated to an arbitrary level (district, town, region) or resolution.

And the linked information change with the chosen unit.

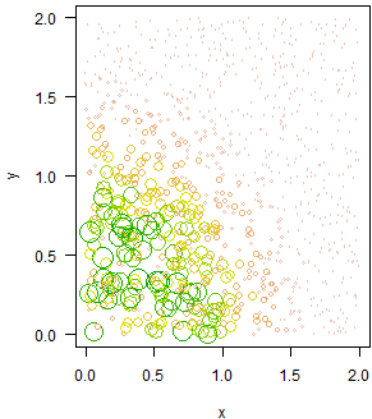
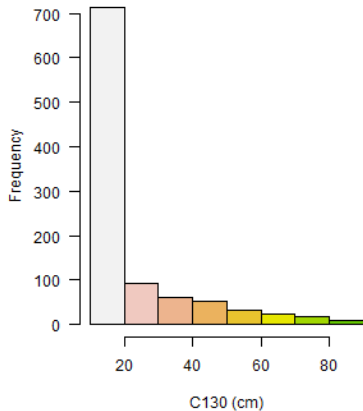


Gembloux Agro-Bio Tech  
Université de Liège

# Zonation

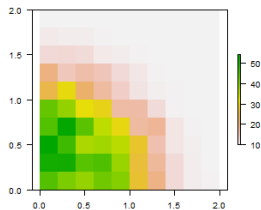
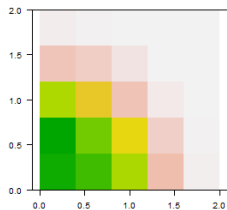
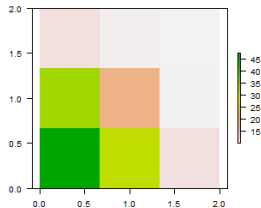
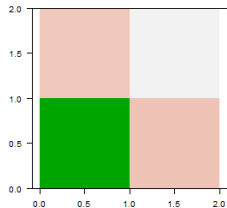


# Zonation



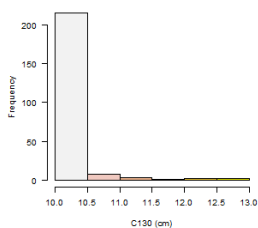
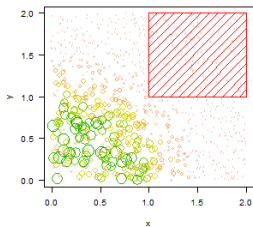
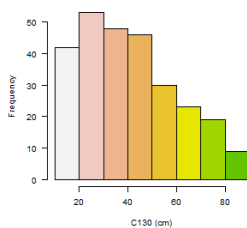
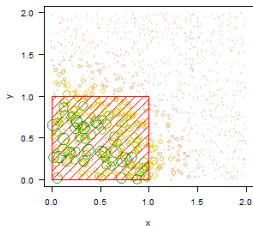


# Zonation



Gembloux Agro-Bio Tech  
Université de Liège

# Zonation



Gembloux Agro-Bio Tech  
Université de Liège

# Resolution

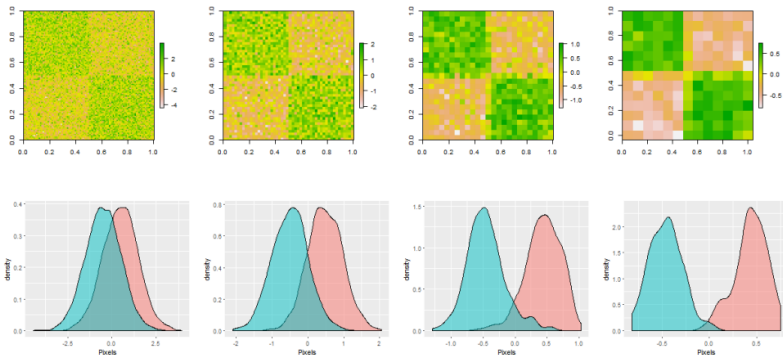
When working with raster data, resolution of the data has two consequences

- on the computing time (resolution  $\times 2 \Rightarrow$  number of pixels  $\times 4$ )
- on the information itself (smoothing)

The choice of a particular resolution (and the smoothing effect it generates) will strongly affect the results.



# Resolution and classification



Gembloux Agro-Bio Tech  
Université de Liège

# Modifiable areal unit problem

The choice of the data type is all but trivial and will strongly affect available data and the corresponding results.

- vector objects or raster pixels ?
- which zonation/resolution ?

Once this choice has been made, data can be extracted to fuel the pattern recognition methods.



- ① Visualise and explore  
⇒ **Principal Component Analysis**
- ② Find structures  
⇒ **Numerical classification methods**
- ③ Predict structures  
⇒ **Discriminant analysis**

