OPEN SPAT

Erasmus+

INSTITUTO SUPERIOR D AGRONOMIA
Hinc Patriam Sustinet
Universidade de Lisboa

Pattern recognition on spatial data

Introduction

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

OpenSpat, ULisboa

June 2019

LIÈGE université Gembloux Agro-Bio Tech

Definitions

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)

LIÈGE université Gembloux Agro-Bio Tech

Machine learning

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

- data selection
- methods selection
- parameters selection

$\Rightarrow$ as most of data modelling, pattern recognition is between **art** and **science**

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*

LIÈGE université Gembloux Agro-Bio Tech

Modifiable areal unit problem

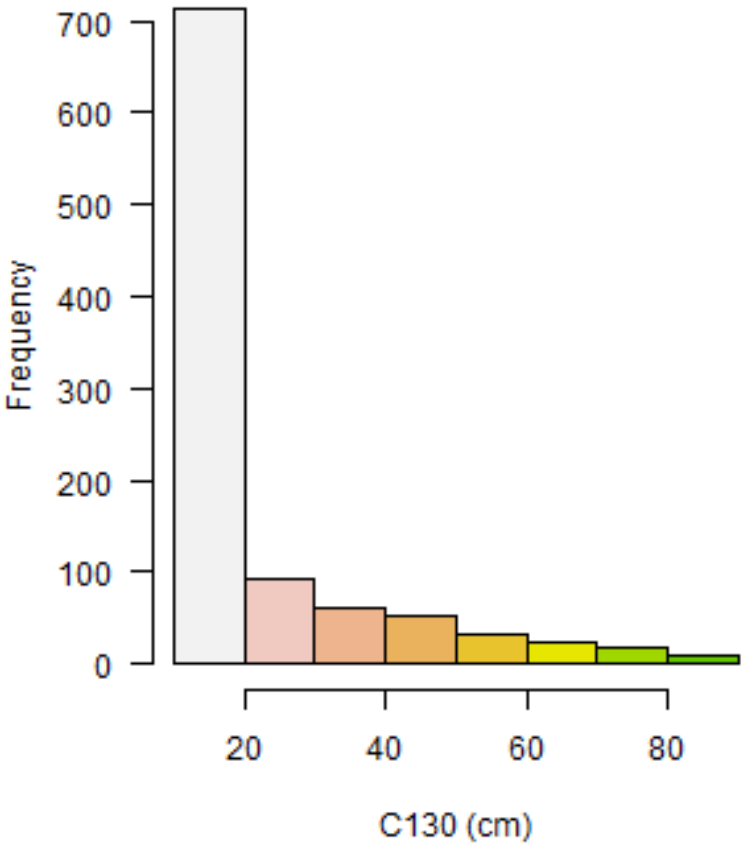
Classical pattern recognition tools search for structures among **individuals**.

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

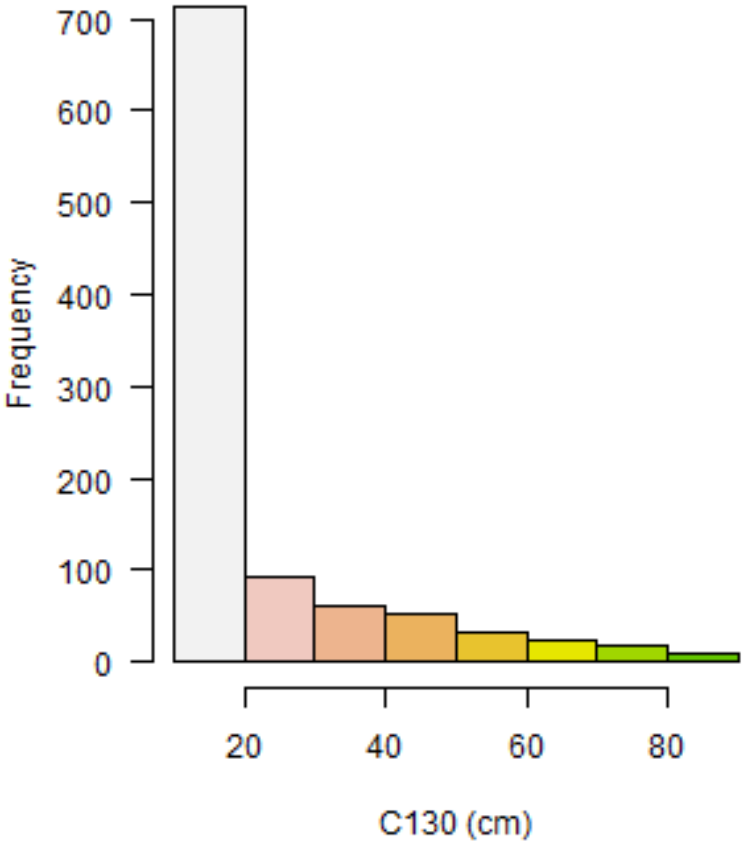
And the linked information change with the chosen unit.

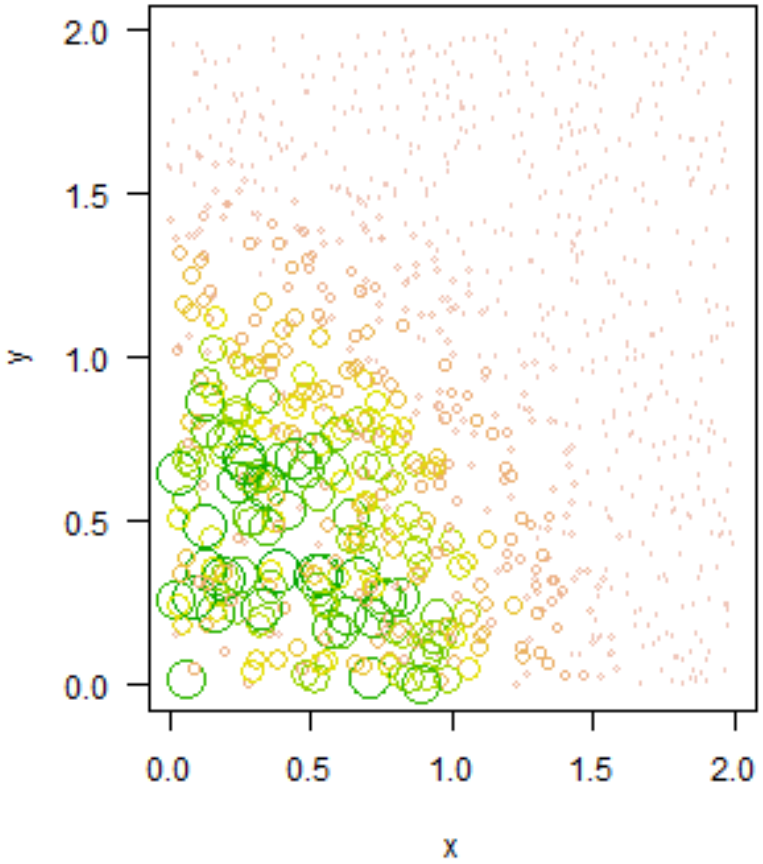
LIÈGE université Gembloux Agro-Bio Tech

Zonation

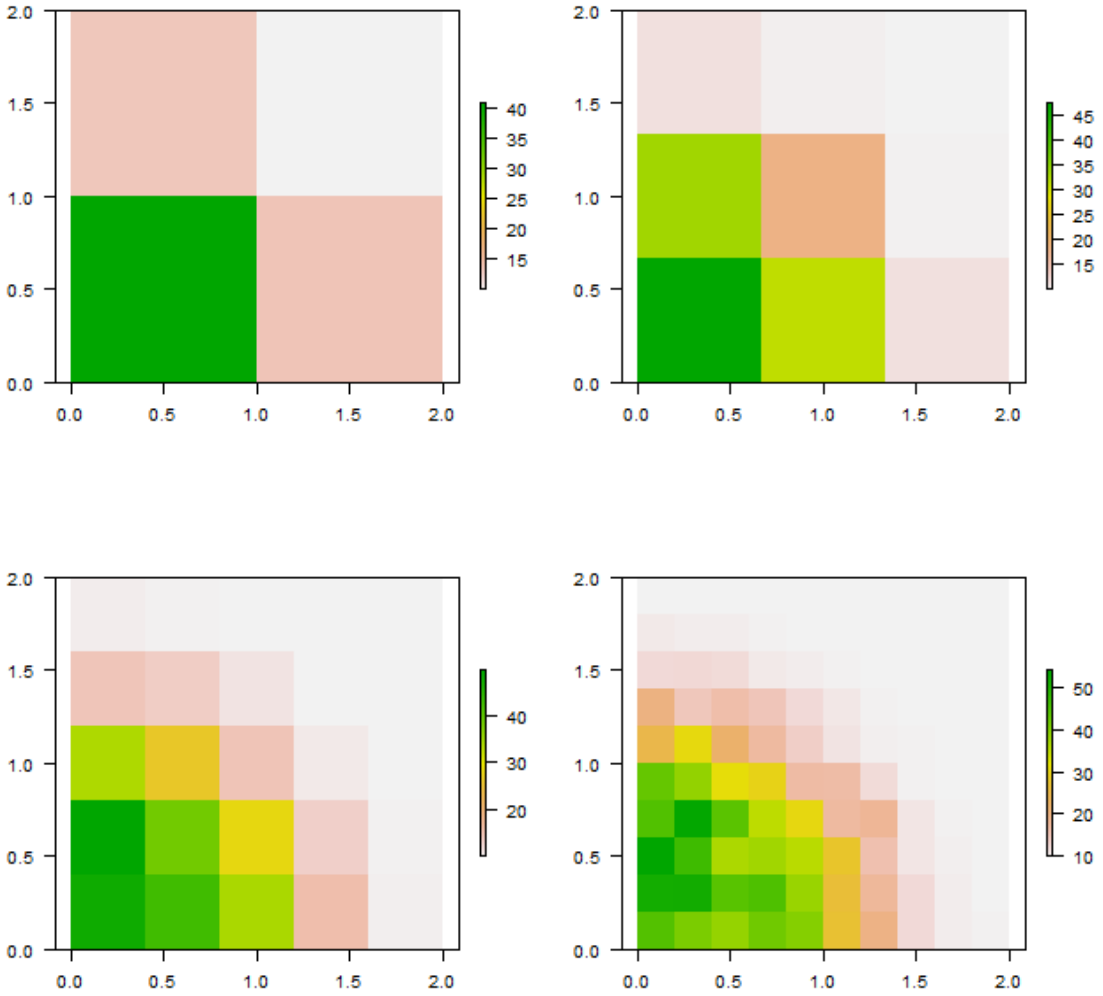


Zonation

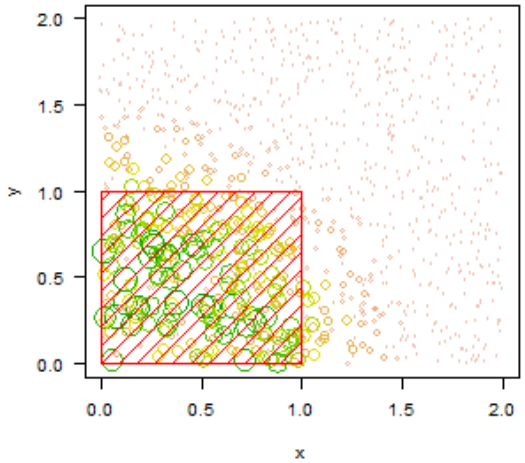

Frequency
700
600
500
400
300
200
100
0
20
40
60
80
C130 (cm)

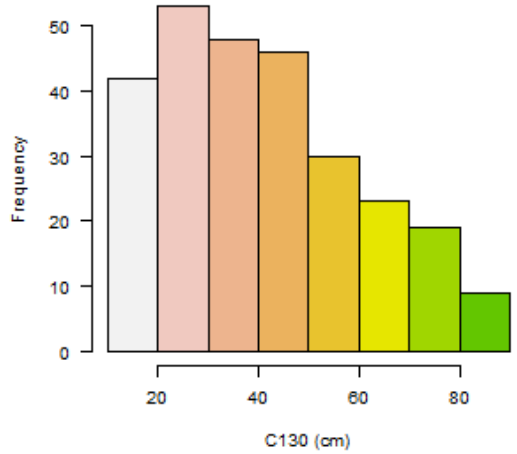

y
2.0
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
x

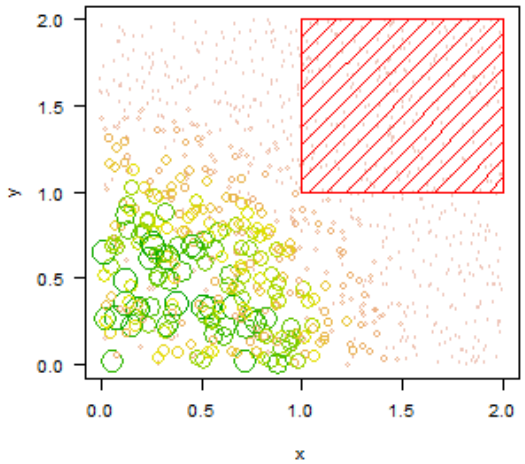
Zonation

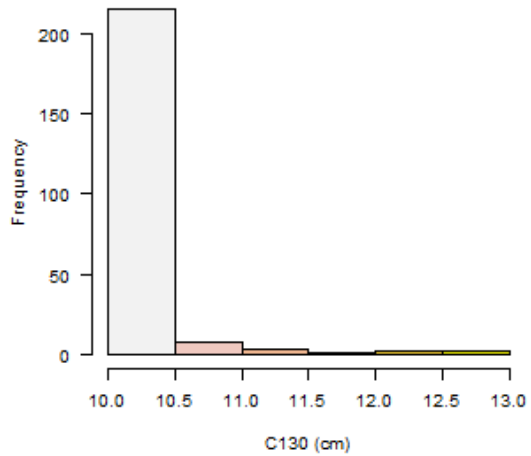

2.0
1.5
1.0
0.5
0.0
0.0 0.5 1.0 1.5 2.0
40 35 30 25 20 15
45 40 35 30 25 20 15
40 30 20
50 40 30 20 10

Zonation









LIÈGE université
Gembloux
Agro-Bio Tech

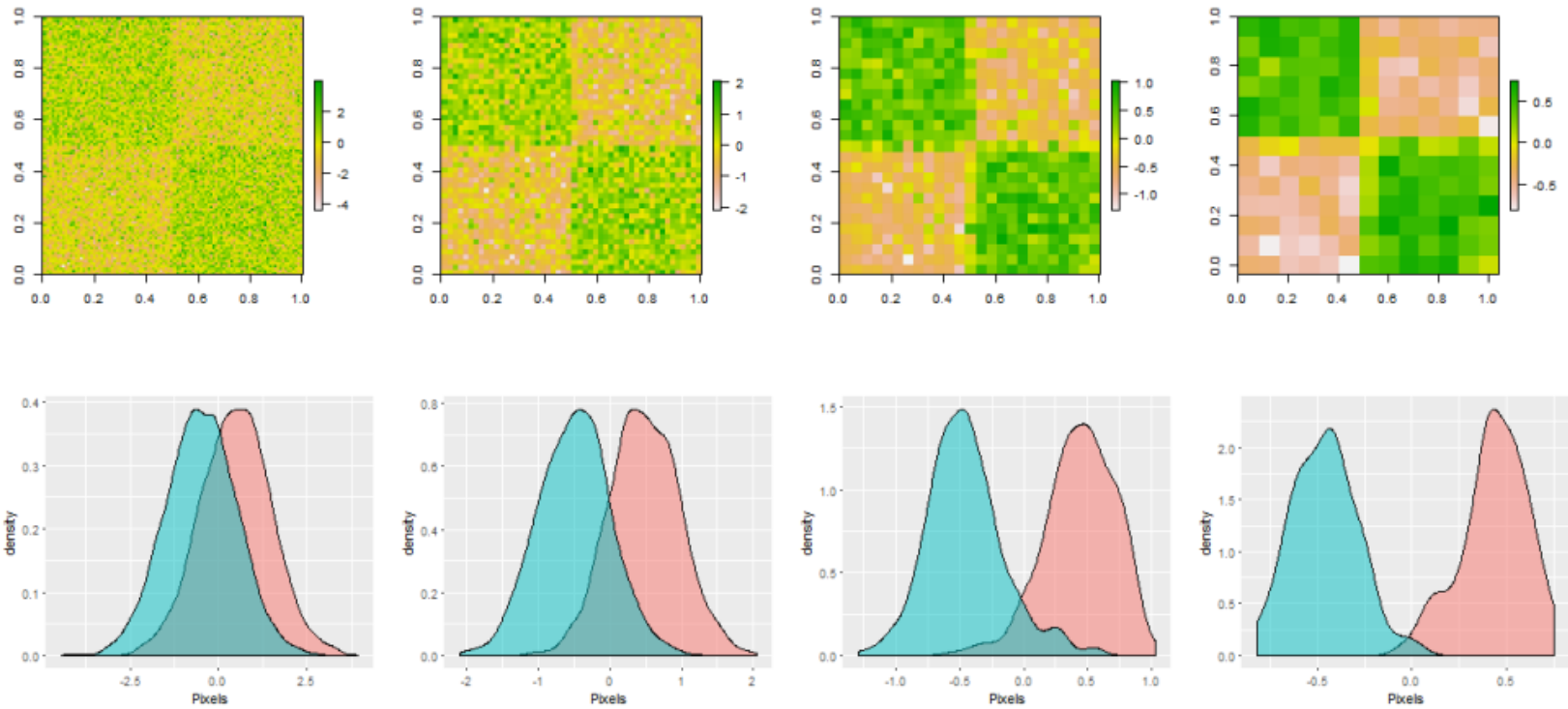
Resolution

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

- on the computing time (resolution x2 $\Rightarrow$ number of pixels x4)
- 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



Modifiable areal unit problem

The choice of the data type is all but trivial and will strongly affect avalaible 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.

Contents

1. Visualise and explore
   $\Rightarrow$ **Principal Component Analysis**
2. Find structures
   $\Rightarrow$ **Numerical classification methods**
3. Predict structures
   $\Rightarrow$ **Discriminant analysis**

LIÈGE université Gembloux Agro-Bio Tech