Mining time series mining of satellite images

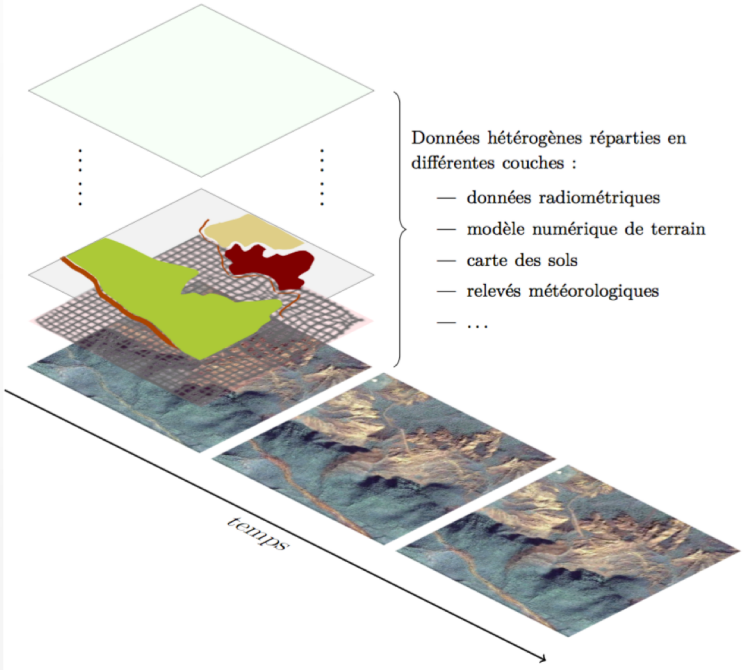
Nazha Selmaoui-Folcher

Data Science
ISEA (Institut des Sciences Exactes et Appliquées)
Université de la Nouvelle Calédonie

Séminaire OSS, 02-04 November 2021, Nouméa

Background and Issues :

- Real Data :
  - Time series of satellite images very high resolution
  - DTM (Digital Elevation Model), meteorological data, …
  - Soil description : nature, vegetation, ….



- Problems:
  - Explosion of the volume stored
  - Complexity of data

(heterogeneous, multiscale, temporal, spatial, etc.)

- for.ex. VHR images, sensor data, DTM, rainfall, socio-economic data, etc.

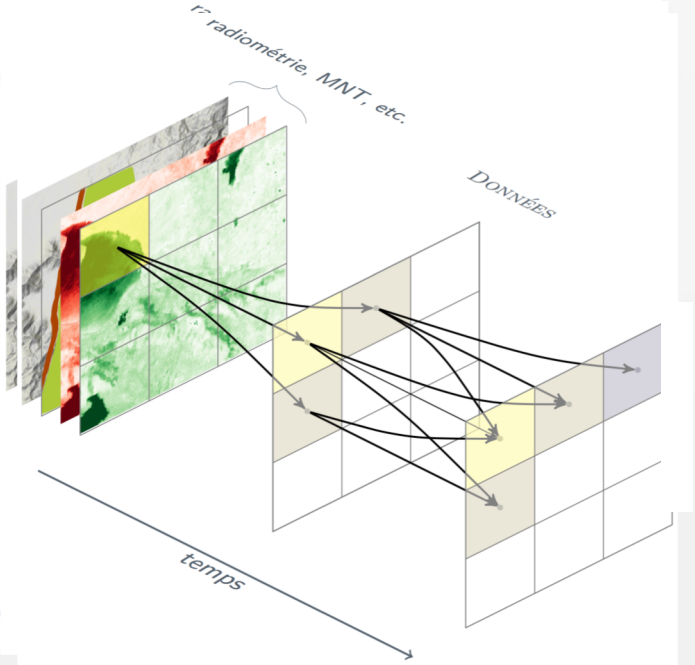
  - Many hidden information in these data (Veracity)
    - for.ex. relationships, behaviors, trends

- How to extract knowledge efficiently and automatically without a priori hypothesis? ➔ Data mining

Dynamics and complex phenomena

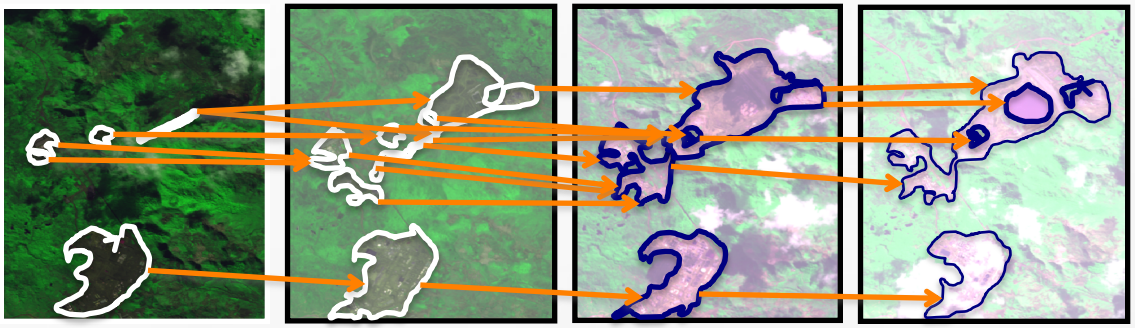
- Study object:
  - Interaction between objects
  - Described by several attributes/properties

- What is evolving between two times (their dynamics)
  - Attributes
  - Position
  - Existence (appearance/disappearance)
  - Structure (merging/splitting)



Towards spatio-temporal patterns

- Study dynamics of objects evolving over time according to their spatial environment:



⇨ Adding spatial and temporal dimensions

⇨ from simple structure (table) to more complex structure: Graphs

Pattern mining in graphs: complex problem

Patterns : definition

- Patterns : detailed descriptions of regularities or trends in the data. We talk about local models
- How to mine patterns :
  - Different algorithm with search strategies (Apriori, FPGrouth, etc.)
- Problem of scalability :
  - Huge number of patterns (n items ⇨ 2^n itemsets)

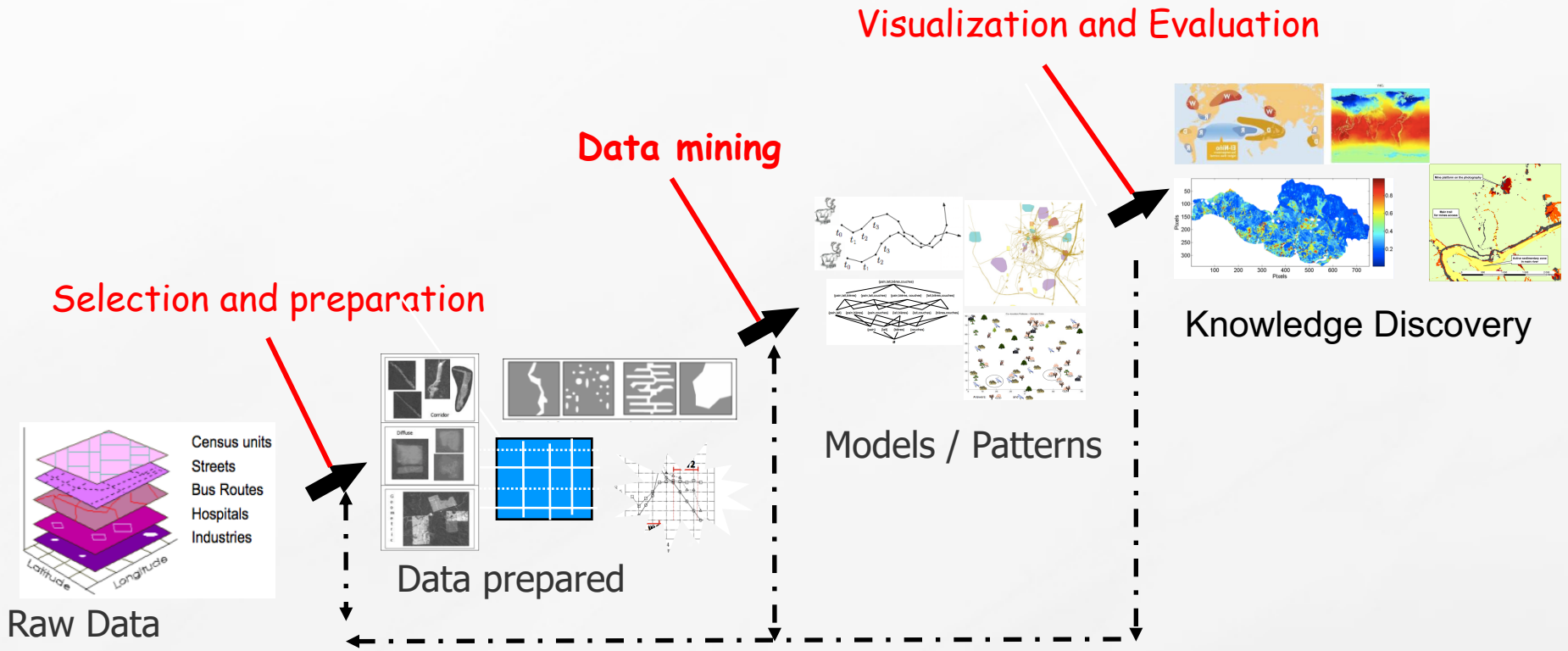
  ➔ Constraint definition :
    - objective constraints: Interestingness measures (frequency, Lift, etc.)
    - subjective constraints: Domain application constraints (expert knowledge)

  We talk about Constrained frequent or interesting pattern mining

« Data Science »

Towards a continuum: data - knowledge – values :

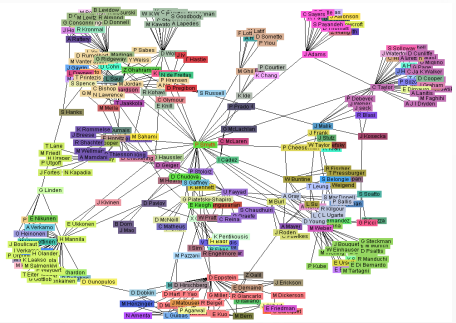
Process of Knowledge Discovery in Databases (KDD)



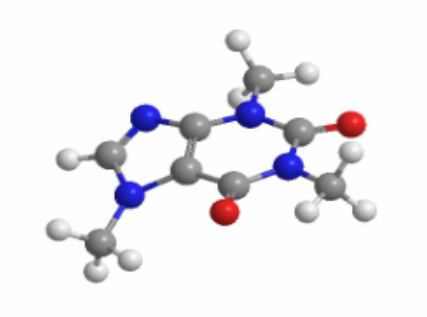
Graph data

- Many complex structured data can be modelled as graphs

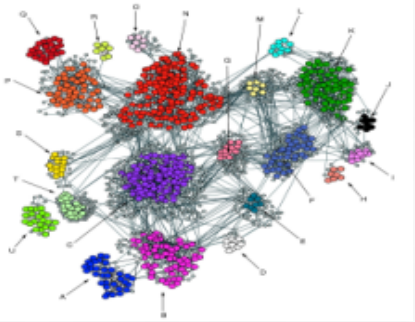
social networks

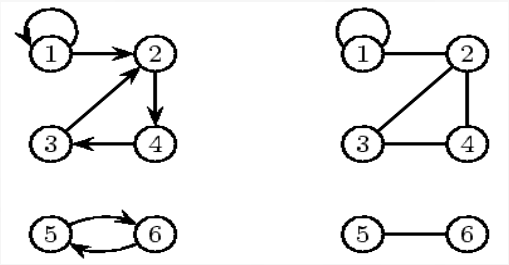


chemical compounds



gene interaction networks



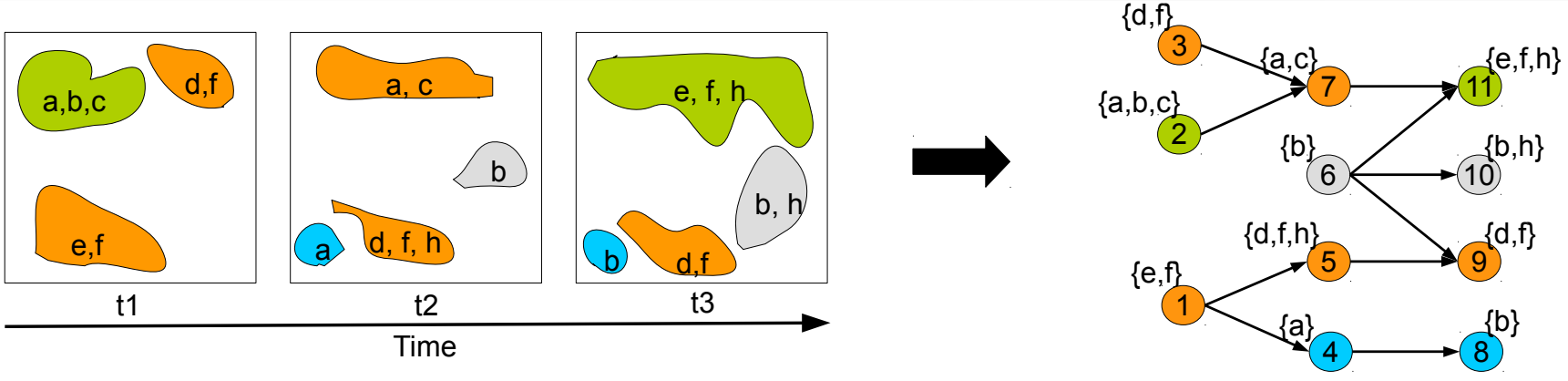


- What is a graph ?
  - a mathematical structure composed of
    - a set of *vertices* (also called *nodes* or *points*)
    - a set of *edges* (also called *arcs* or *lines*) linking two vertices
    - a set of attributes characterizing vertices and/or edges
    - a labelling function that maps each vertex to a subset of attribute values
    - a labelling function that maps each edge to a subset of attribute values

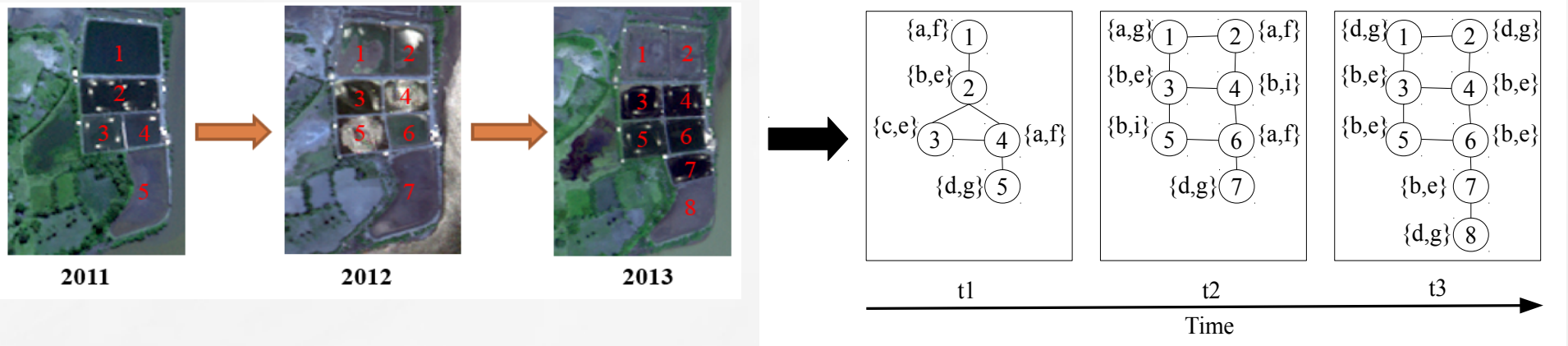
Graphs and Ecological data

- Modelling spatio-temporal relations using graphs

Attributed DAG



Dynamic attributed graph



Graph mining : difficult task

- Most contributions focus on frequent sub-graph mining in a collection of labelled graph
  - Example of algorithm: *gSpan* [X. Yan et al., ICDM'02]
  - Labelled = only one information per vertex/edge
  - Frequency of a pattern = number of input graphs supporting the pattern ≠ number of occurrences
  - ➢ Still a difficult task
    - size of the search space and cost of sub-graph isomorphism
- Few works on attributed graphs and dynamic attributed graphs
  - A more difficult task
    - combinatorial explosion (itemsets + graphs + sequence)
    - complexity of the single-graph setting
  - ➢ Specific pattern domains, strong constraints and hypothesis
    - e.g. mining "communities" (sub clusters of vertices with the same values or trend) [F. Moser, SDM'09] [M. Fukuzaki, PAKDD'10] [E. Desmier, PKDD'13]
    - [Zhi et al, 2018], [Fournier-Viger et al. 2020]

FOSTER Project

Spatio-temporal data mining: application to the understanding and monitoring of soil erosion 2011-2014

For more information about project and partners : http://foster.univ-nc.nc

~900 000€

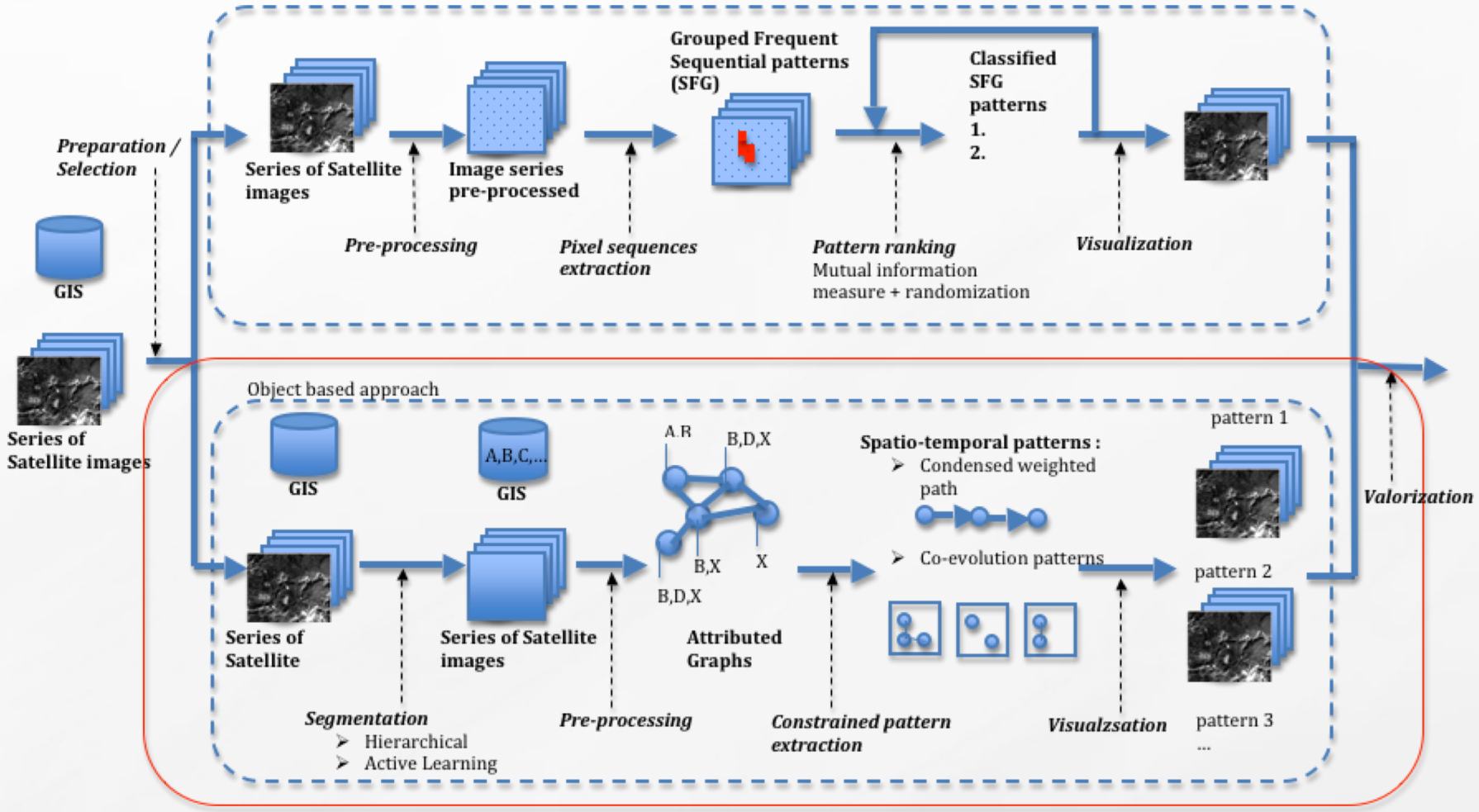
~280 000 € for ISEA/UNC

Foster Project objectives

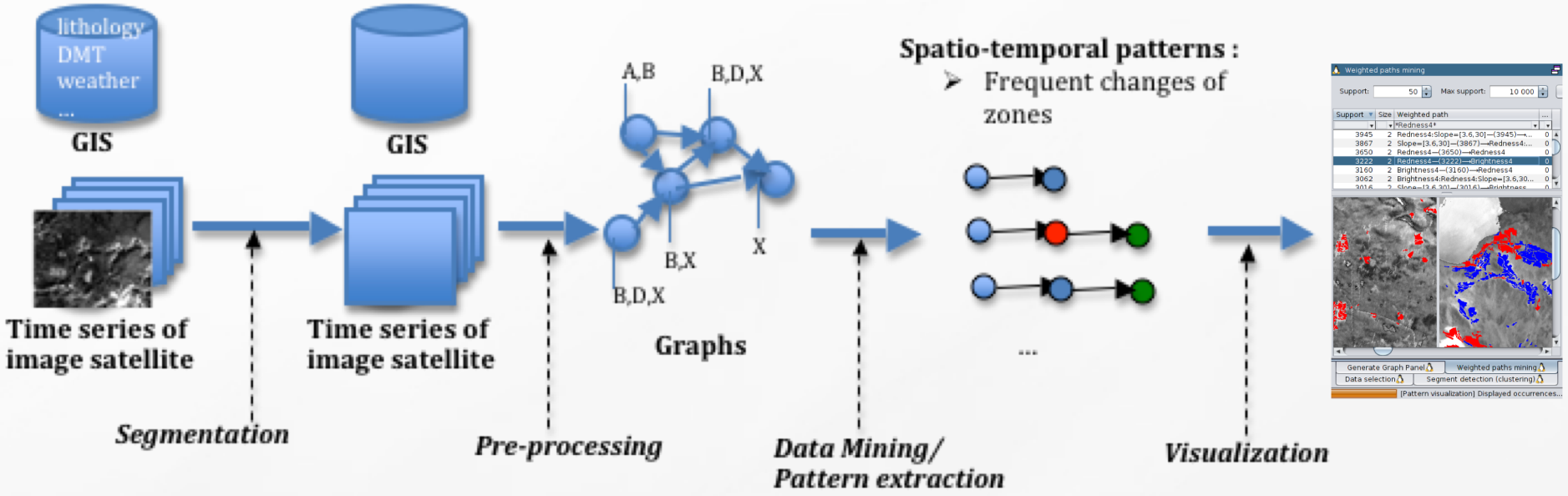
- Developing new data mining methods : Innovative, Efficient, Relevant, Generic
- To build dynamic model to observe and study environmental change
  - Application to erosion
- Construction and validation of methods with experts throughout the project
- Visualization and Integration of results in plate-forme to provide a decision support tool for communities and industrial

Contributions to project

- Development of a semi-automatic process of multi-temporal images used jointly with other data (DTM, vegetation, geological data, ...):



Development of a hole KDD Process
Application to time series of satellite images



- *Segmentation* : identifying interest objects = homogeneous region in images
- *Pre-processing* : compute indices (NDVI, Redness,…) + information merging (indices + GIS + zones) into directed and attributed graph structure (ADAG)
- *Local pattern extraction in ADAG*: analysis and crossing locally the data to identify frequent changes - new domain of patterns condensed weighted path
- *Prototype of visualization* (with possibility to filter and sort results)

Data

- **Time serie of satellite images**
- **Objects with associated properties :**
  - **Radiométriques**
  - **Indices**
  - **GIS Layer :**
    - **Type of vegetations**
    - **Land use**
    - **Lithology**
    - **…**
  - **…**

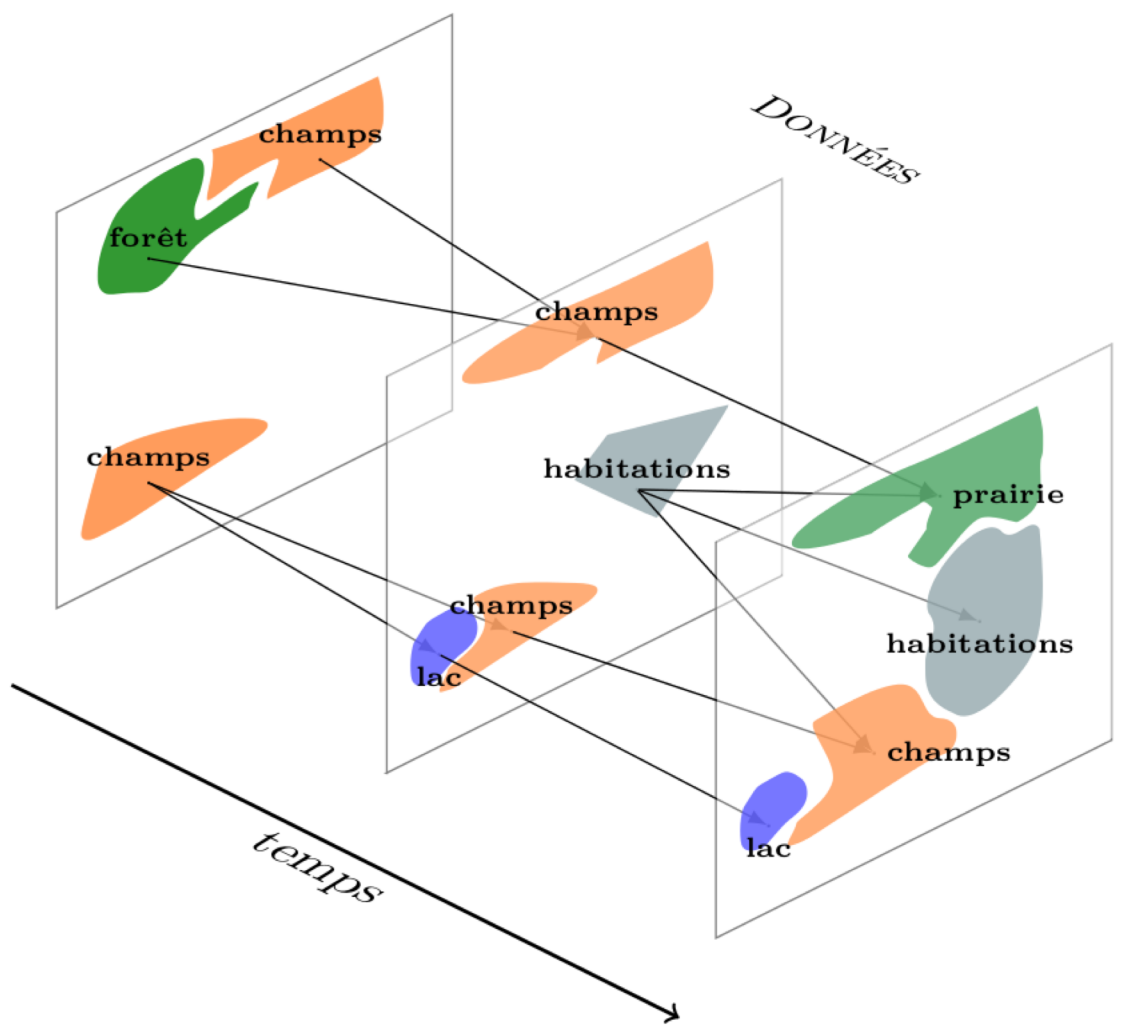
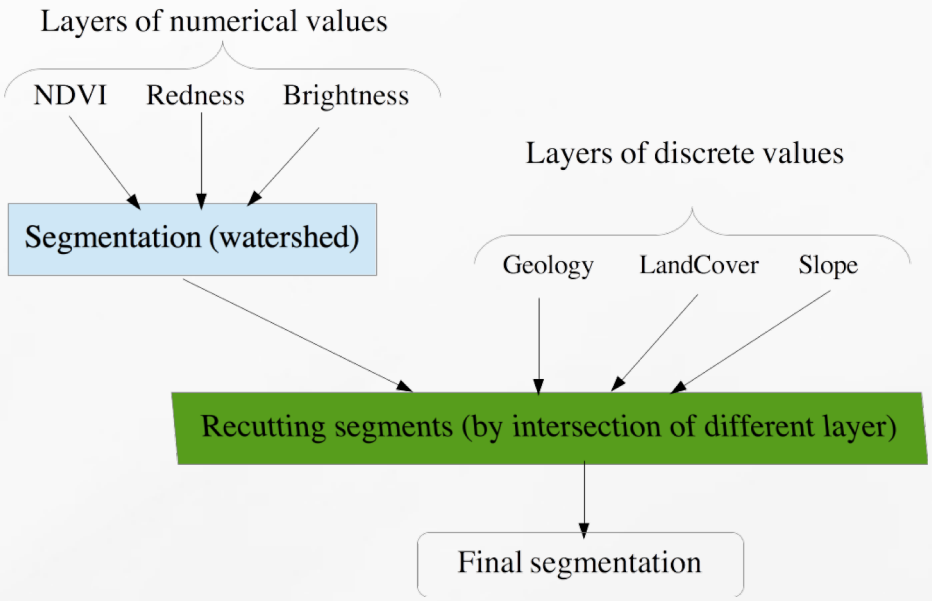
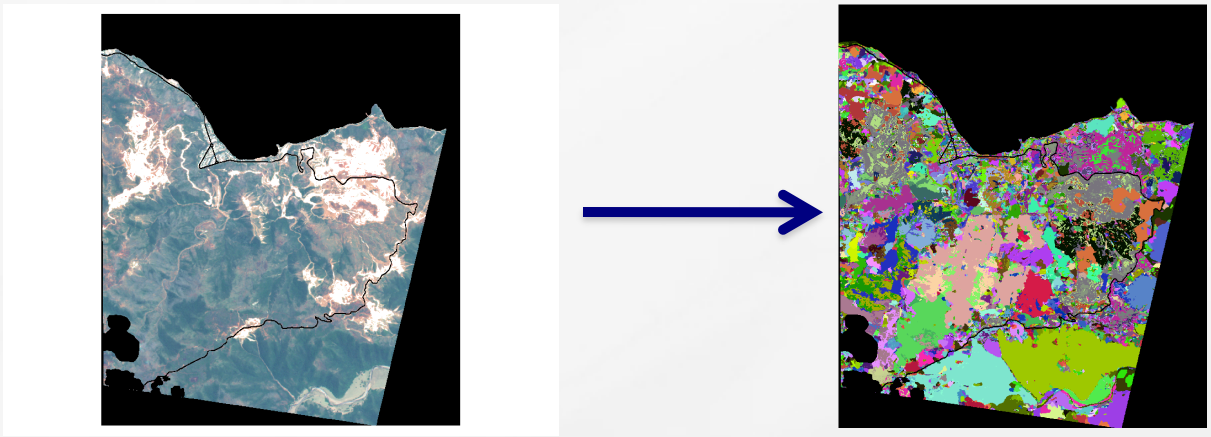


Image segmentation phase : object detection

- Calculation of radiometric indices: NDVI (vegetation), Redness (ground redness), Brightness, IHN (humidity)

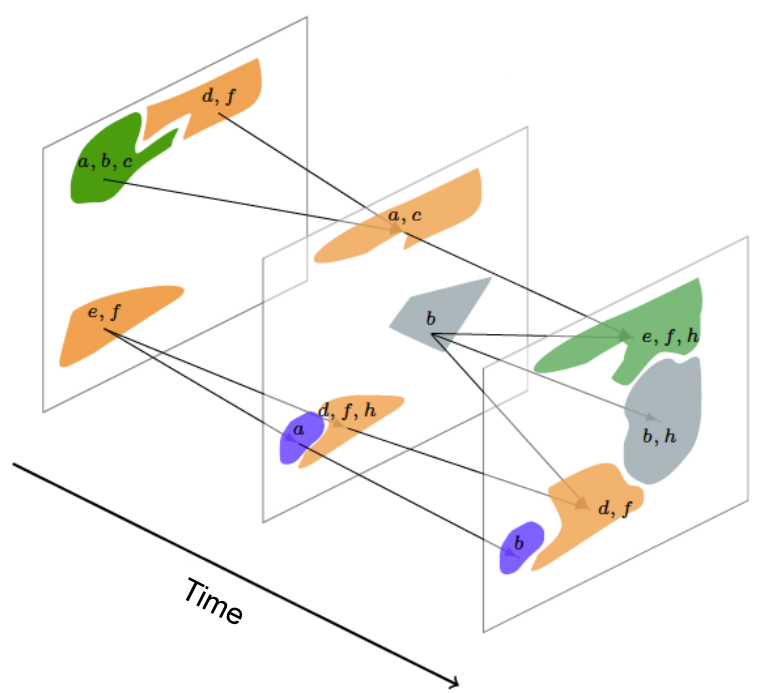


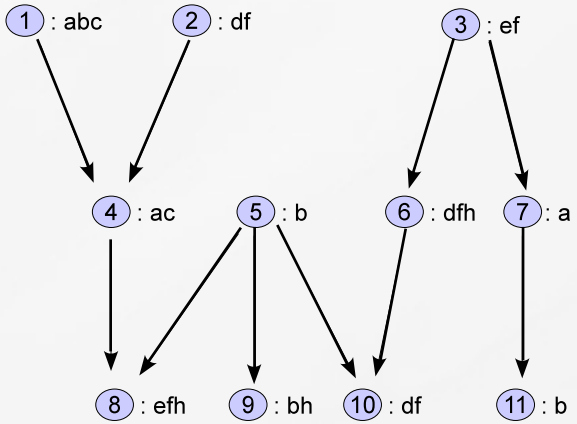
- Example of result:



Data representation: Attributed Directed Acyclic Graph

- Attributed directed acyclic graph (a-DAG)
  - Node: object
  - Node label: set of attributes/properties of object (*itemset*)
  - Edge between nodes: spatial and temporal evolution

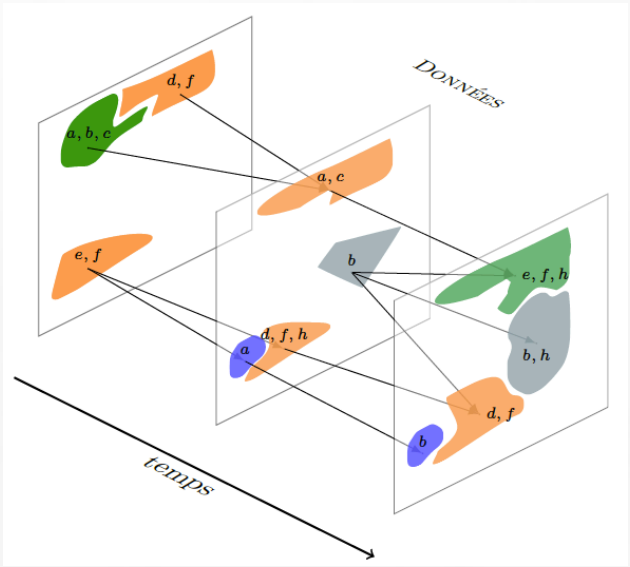


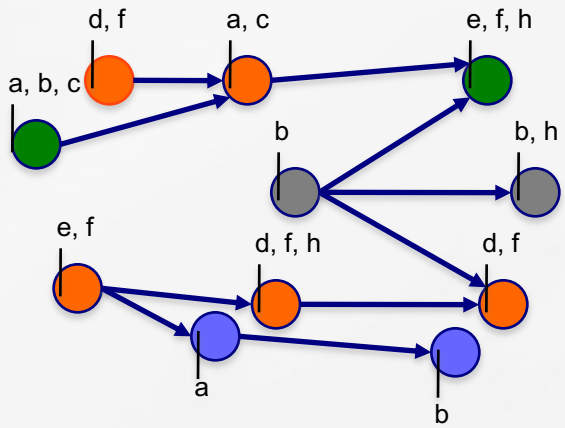


Pre-processing

- Discretization in 5 classes (very weak, weak,…) for image indices
- Merging information (indices + GIS + zones) as an a-DAG

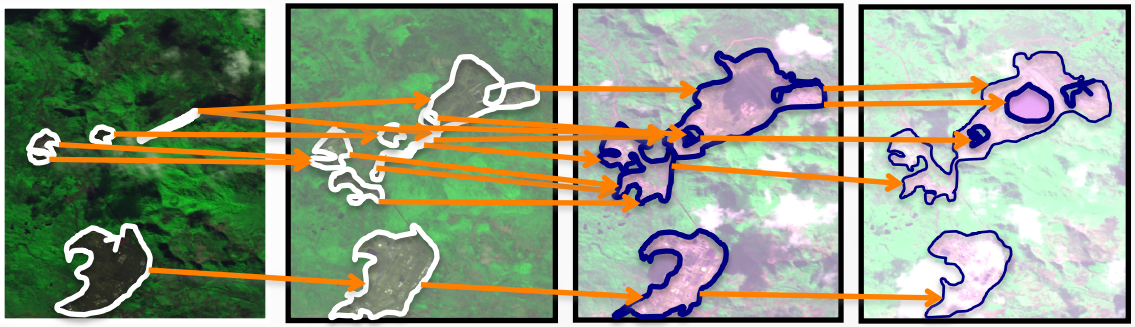
advantage: possibility to represent mergers/divisions of zones

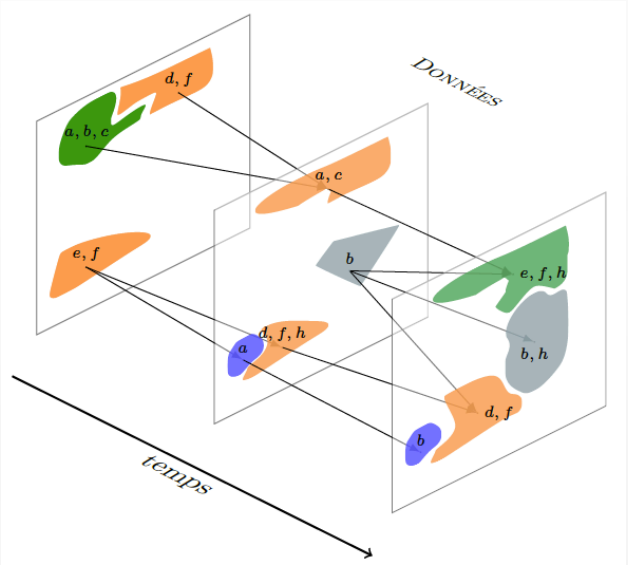


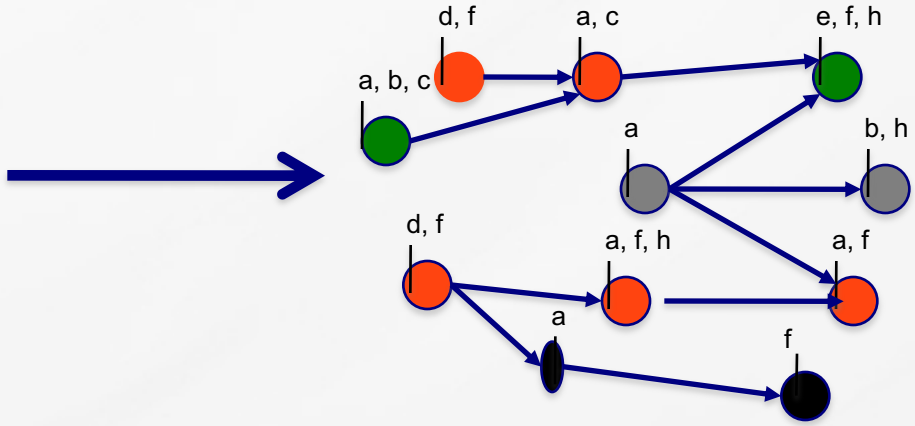


a, b, c, d, e … = weak NDVI, strong NDVI, weak Redness, dry forest, slope 30% …

Extraction of evolutions : mining weighted path







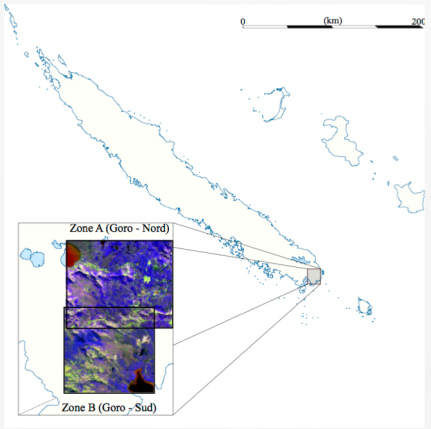
description model / evolution pattern :

$$d,f \xrightarrow{2} a \xrightarrow{2} f \qquad a \xrightarrow{4} f$$

Example of studying the phenomenon of erosion: The data

- 2 zones are tested

- Zone "Poro" (CNRT)
  - zone of 35 km^2
  - dates: 2011 et 2012
  - Series Image of RapidEye (resolution 5m)
  - Land cover



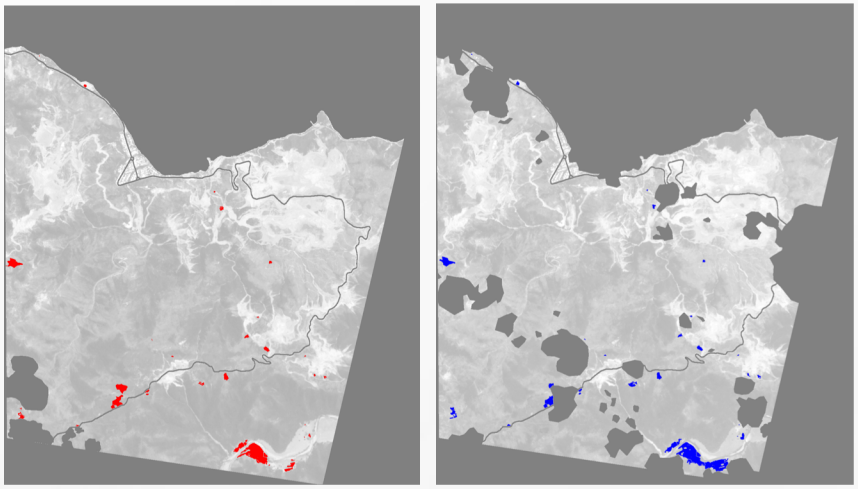
- Zone "Goro" (ANR FO.S.T.ER.)
  - 2 zones: 52.5 km^2 + 41.9 km^2
  - dates: 1999, 2002, 2005, 2008, 2009
  - 2 series images of SPOT4/5 (resolution 10m)
  - DTM, type of soil, land cover (2008)

Example of Local Pattern evolution (1/4)

43

Redness4 ➔ Redness2, NDVI4

➢Redness index decreases, supported by high NDVI (revegetation)

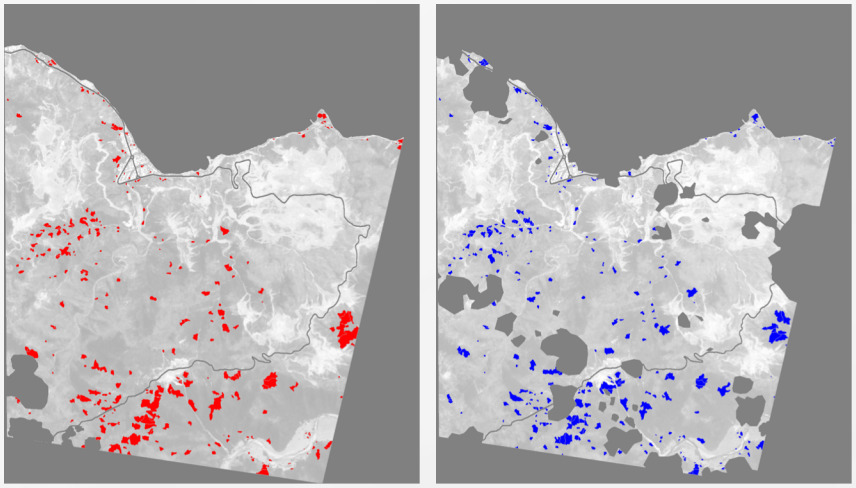


2011 2012

350

Low sparse forest ➔ sparse forest

Vegetation evolution for 350 zones

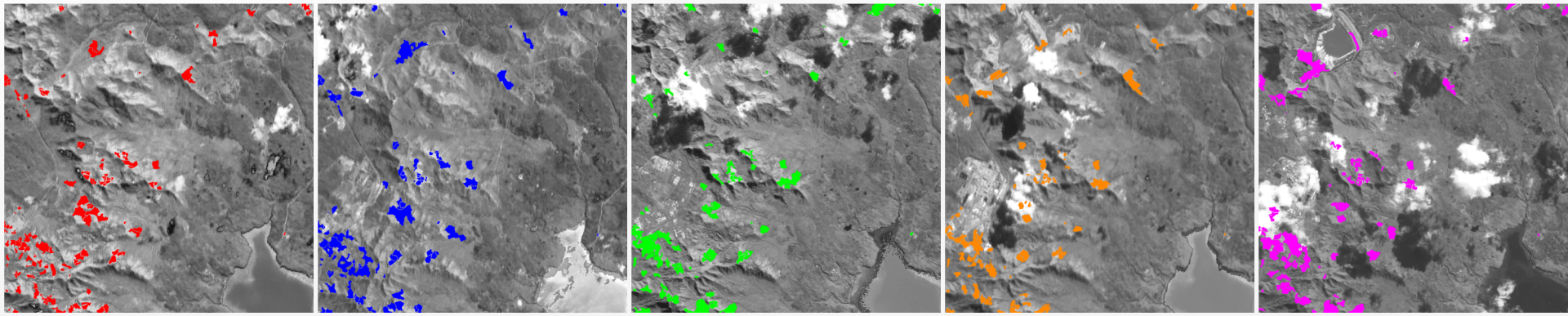


Example of Local Pattern evolution (2/4)

Redness4 $\xrightarrow{159}$ Redness4 $\xrightarrow{116}$ Redness4 $\xrightarrow{112}$ Redness4 $\xrightarrow{110}$ Redness4

No change for Redness over 4 dates

zone "Goro"



1999 2002 2005 2008 2009

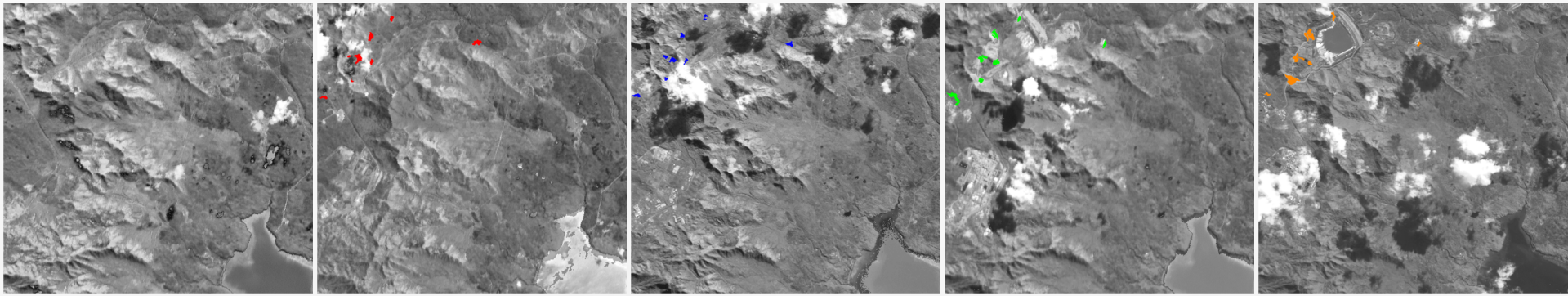
➢ but changes in surface area

Example of Local Pattern evolution (3/4)

NDVI3,Redness2 $\xrightarrow{9}$ Brightness3 $\xrightarrow{10}$ Redness4 $\xrightarrow{10}$ NDVI0,Redness4

Increased redness index together with a decrease of the vegetation index

zone "Goro"



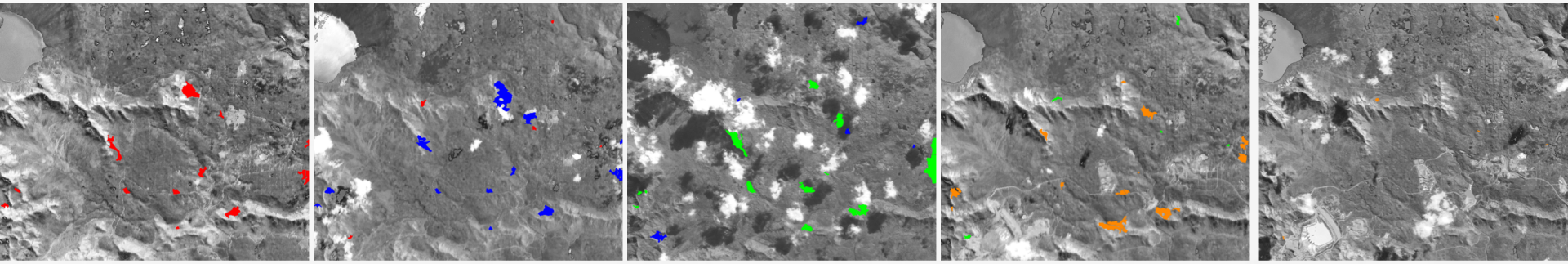
1999 2002 2005 2008 2009

Example of Local Pattern evolution (4/4)

Redness1 $\xrightarrow{20}$ Redness1 $\xrightarrow{21}$ Redness2:Slope[3,6;30] $\xrightarrow{23}$ Redness3

Increased Redness index for zones with high slop

zone "Goro"

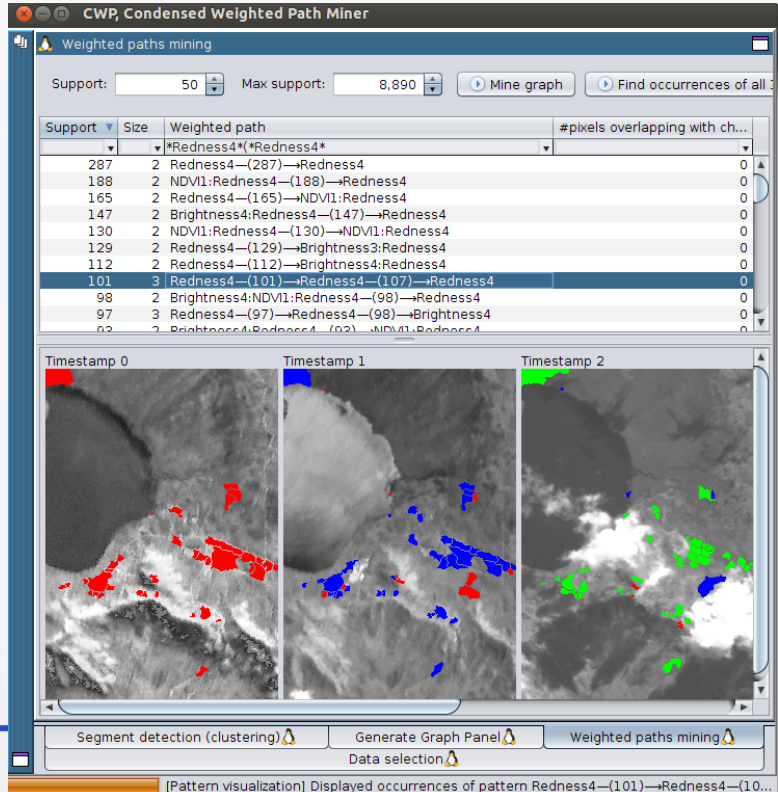


1999 2002 2005 2008 2009

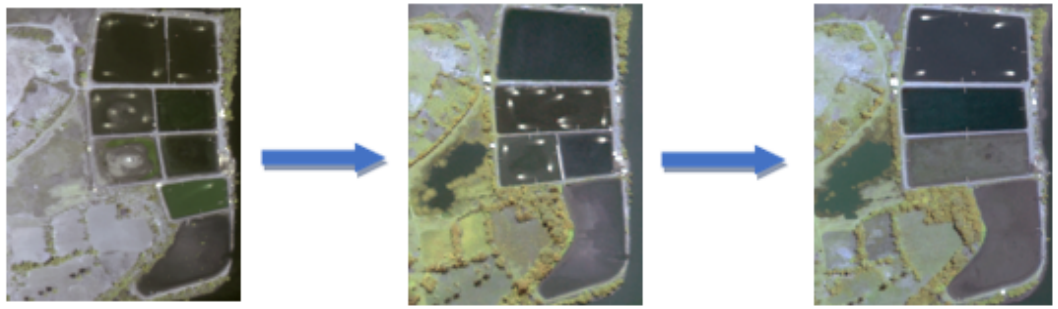
Visualization/Prototype and platform (KNIME)

- Development of software of visualization (prototype)
  - Visualize Evolution/Patterns
  - Sort patterns(frequency, size, content)
  - Filter Patterns through regular expressions
    - For example "*Redness0*NDVI4*Redness4*NDVI0*"

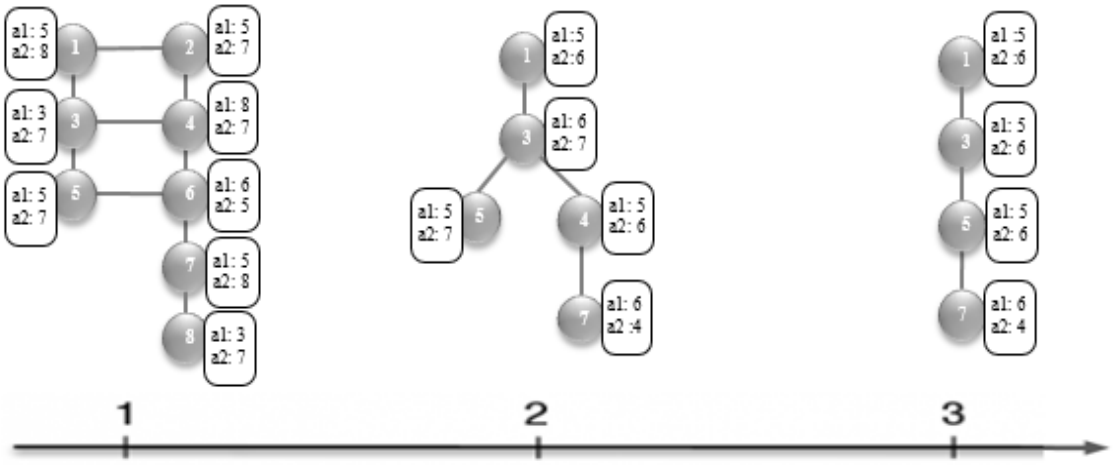
    -> find all patterns with strong increase Redness conjointly with strong decrease NDVI



Other application: Evolution of aquaculture pond



Transformation des données



Aquaculture Ponds

Node: Pond

Edge: Neighbourhood relations

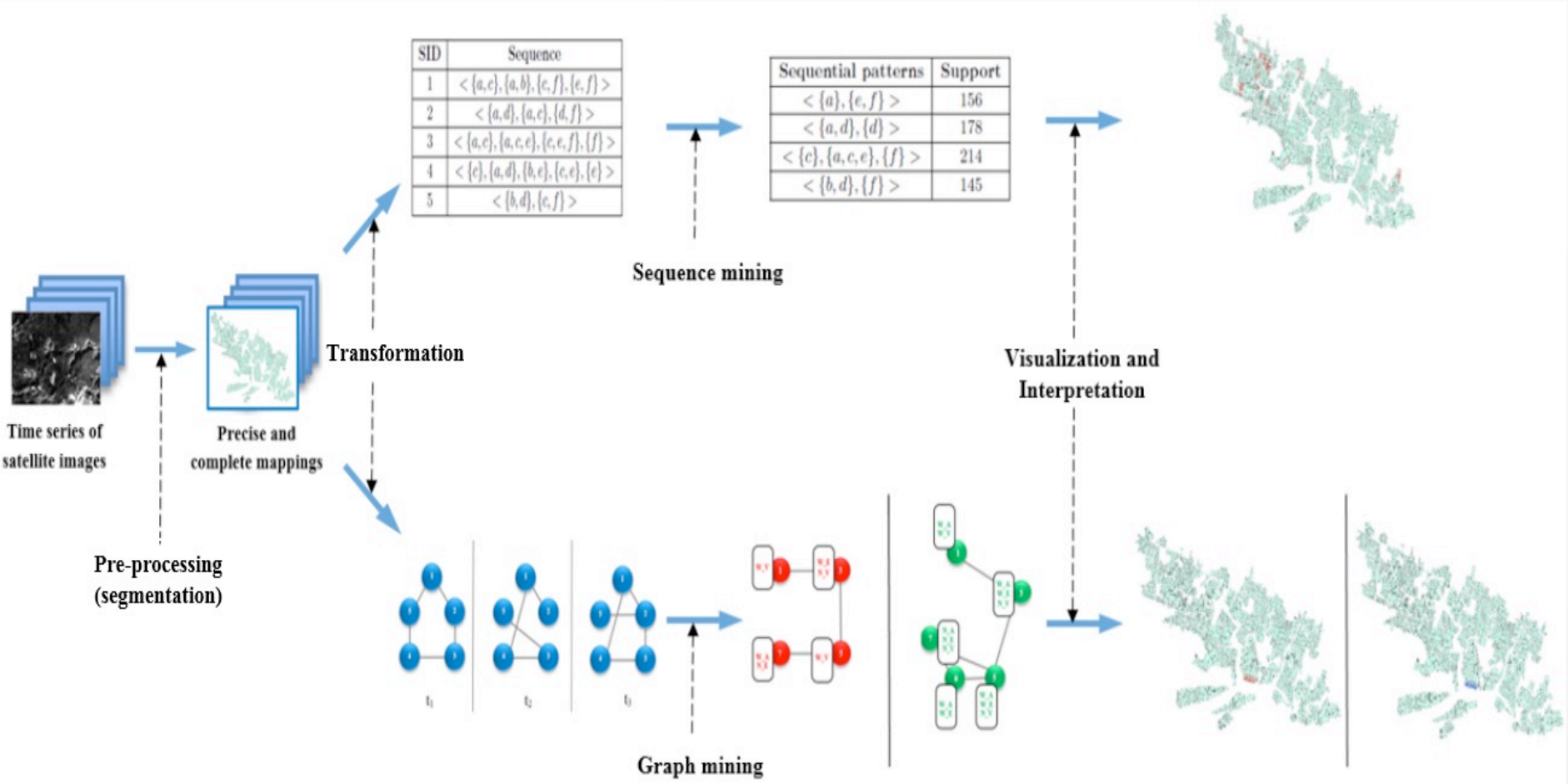
a1 : NDVI

a2 : activity state

a3: surface

…

Other application: Evolution of aquaculture pond



Conclusion

- Analysis of Big Data (and their crossing)
  - New method of data mining (New pattern domain) on VHR image satellites and GIS layers
  - Implementation of whole process
  - Application to environment

Integration the process in Knime software (done)

Now is free available software

Partners

- Jean-François Boulicaut, Pr. au LIRIS Insa de Lyon
- Frédéric Flouvat, MCF Informatique ISEA
- Claude Pasquier, CR Informatique UMR CNRS Nice
- Philippe Fourbier-Viger Pr. shenzeen university, Chine
- Hugues Lemonnier, Entropie / IFREMER
- PhD Student :
  - Hugo Alatrista-Salas
  - Jérémy Sanhes
  - Niken Gusmawati
  - Zhi Cheng
- Current PhD Student
  - Jannai Tokotoko
  - Romane Sherrer
  - Rodrigue Govan

- Thank you for your attention
- http://foster.univ-nc.nc

29

The questions

- Which representation adapted to these data?

Determine what is considered of the studied phenomenon

- Which pattern domain ?

Determine what kind of correlations are analyzed

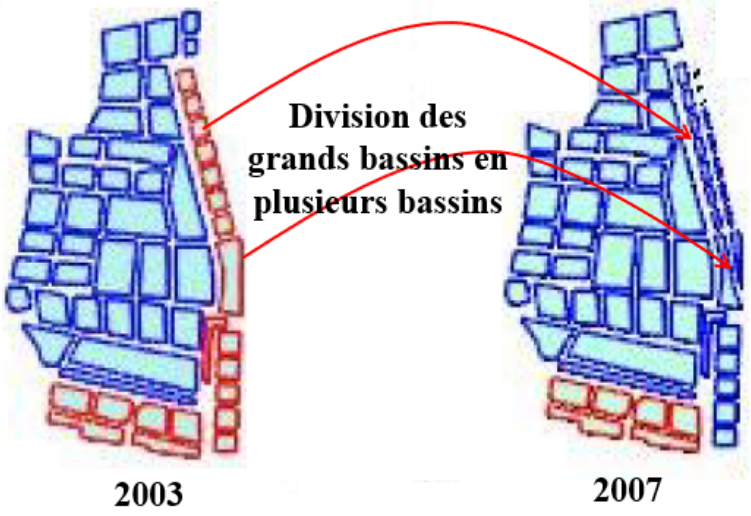
- Which constraints ?
  - Which objective interestingness measure (i.e. statistic) ?
  - How to quantify the interest of the expert?
    - How to take into account the knowledge of the experts?

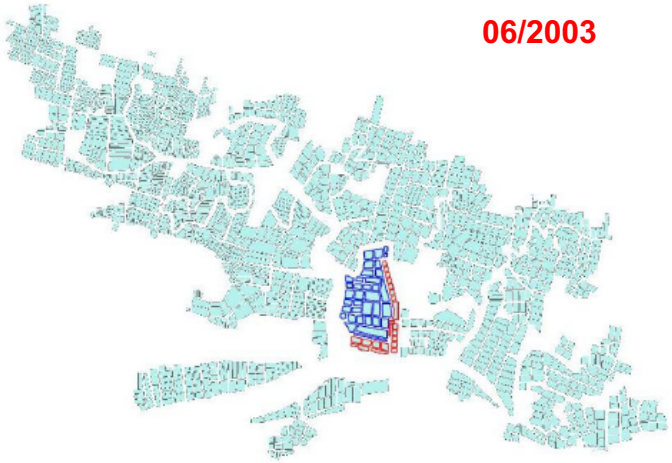
Determine which properties have these correlations

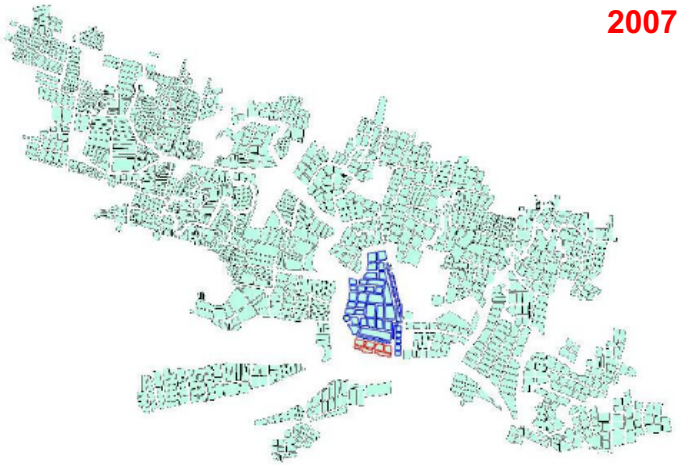
➔ How to efficiently extract these patterns ?

- Which strategy of exploration ?
- How to exploit constraints in order to optimize algorithm?

Evolutions des bassins par RPMiner







- **Impact important de la maladie sur l'activité des bassins**