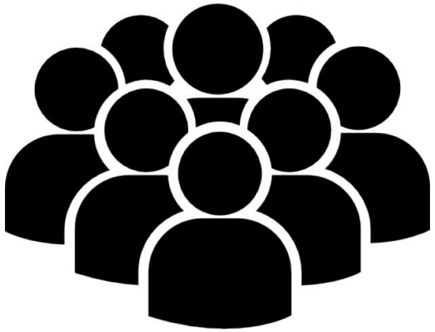
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

A discussion of crowdsourced geographic information initiatives and big Earth observation data architectures for land-use and land-cover change monitoring

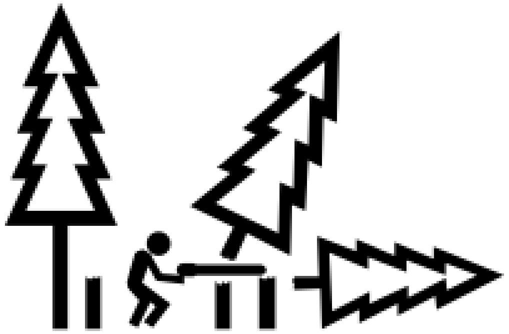
LUIZ FERNANDO F. G. DE ASSIS, TÉSSIO NOVACK, KARINE R. FERREIRA, LUBIA VINHAS AND ALEXANDER ZIPF

Environmental Challenges

GLOBAL WARMING



OVERPOPULATION AND OVERCONSUMPTION



DEFORESTATION

- **Land Use** and **Land Cover** Change Monitoring;

Land Use → "purpose for which and activities by which human use land"

Land Cover → "physical properties of a land surface"

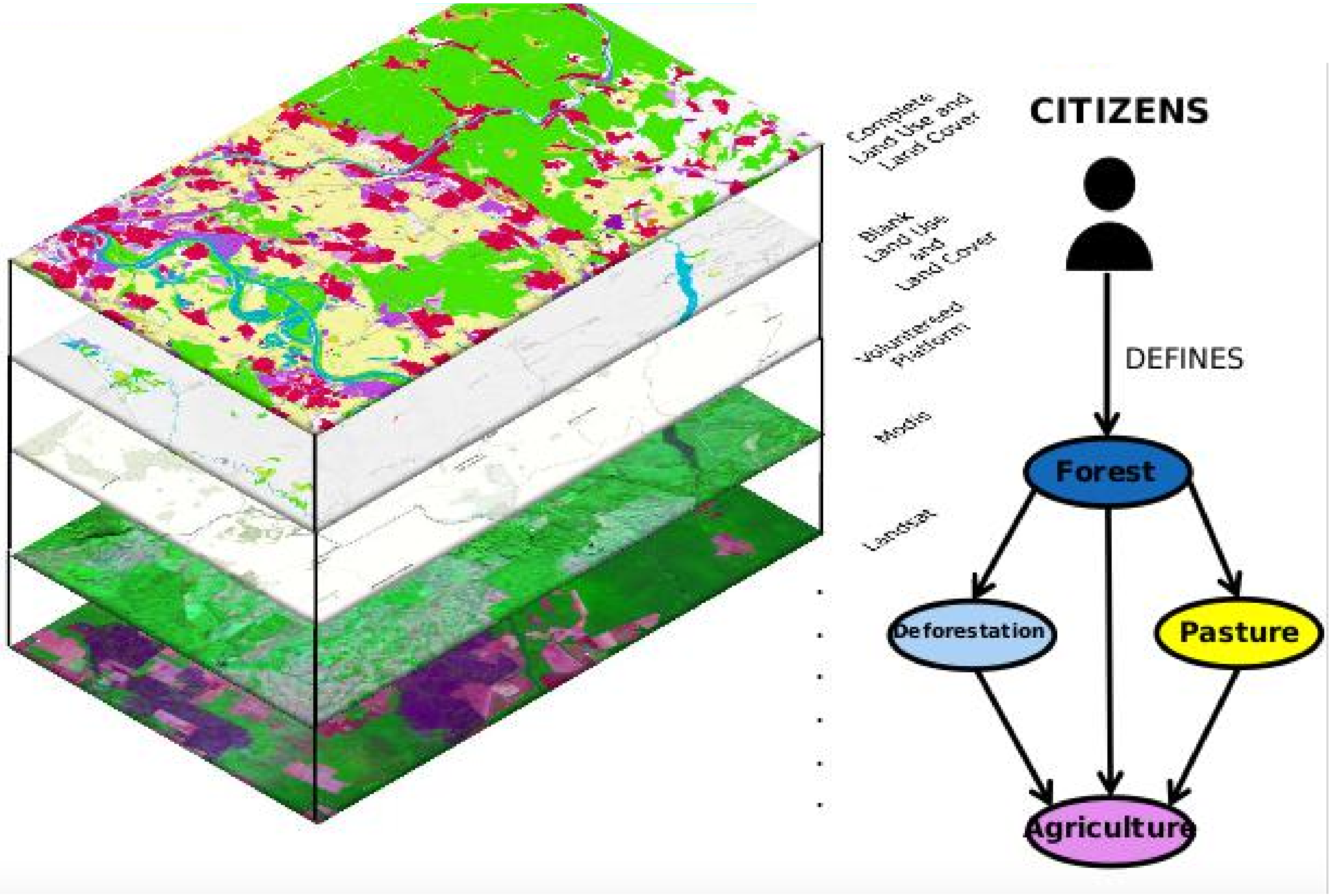
(JOSHI, Neha et al., 2016)

Citizen Science for Land-Use and Land-Cover Monitoring

1. Providing up-to -date and detailed information for land use and land cover monitoring;

2. Incorporating the society judgement into the analysis of datasets;

3. Analyze existing information about land use and land cover change.



Participatory Monitoring Platforms

- Provide collectively a map;

- Bring together the community;

- Keep costs low for users;

- Estimulate academic programms;

- Participatory sensing;

- Estimating Accuracy of LULC changes throughout crowd-based validation methods;

- Improve information quality;

Participatory Action Research

- Project Design;

- Infrastructure Technology and Communications;

- Data Management;

- Participants Engagement vs Participants Motivation;

- Participants Preparation;

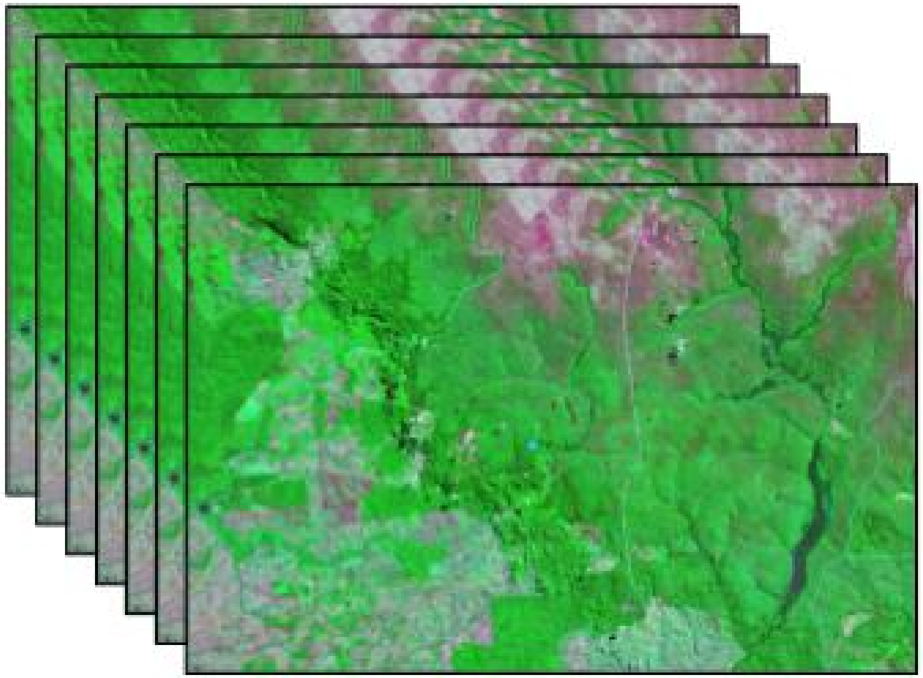
- Data Quality;

- Data Privacy;

- LULC Change Classification Purpose.

Big Earth Observation Data

Earth Observation Programs:
Large and open satellite imagery



Remote sensing applications: meteorology, oceonography, agriculture, etc.

Researchers need **stable and efficient solutions to support the development and validation** of algorithms for big Earth observation data analysis.

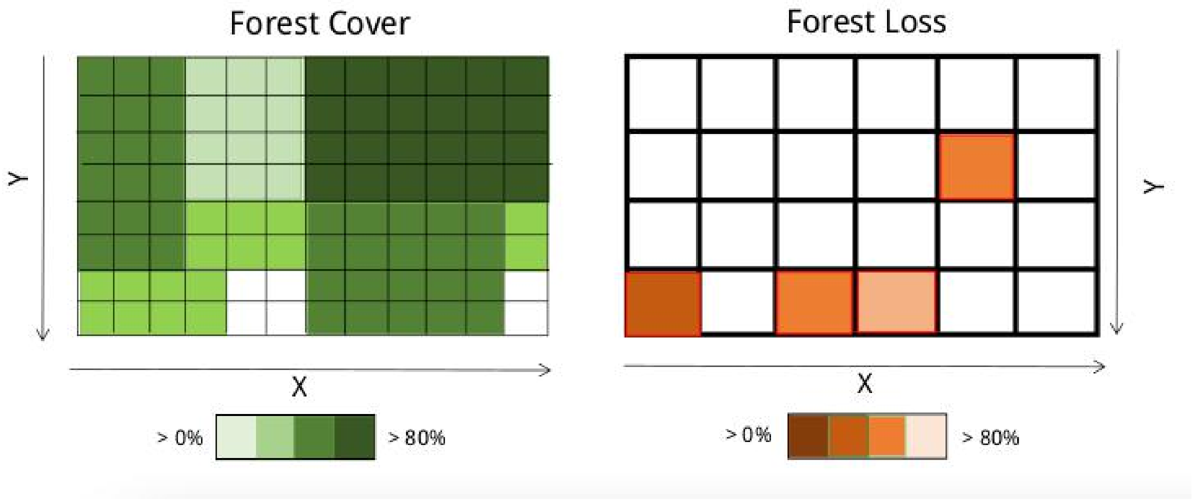
- **Analytical scaling:** allow algorithms developed at the desktop to run on big databases with minor changes.
- **Software reuse:** allow researchers to adapt existing methods for big data with minimal reworking.
- **Collaborative work:** share results with the scientific community.
- **Replication:** encourage research teams to build their own infrastructure.

Research Question

How can citizen science be applied in big Earth observation data analytics to improve land use and cover change monitoring?

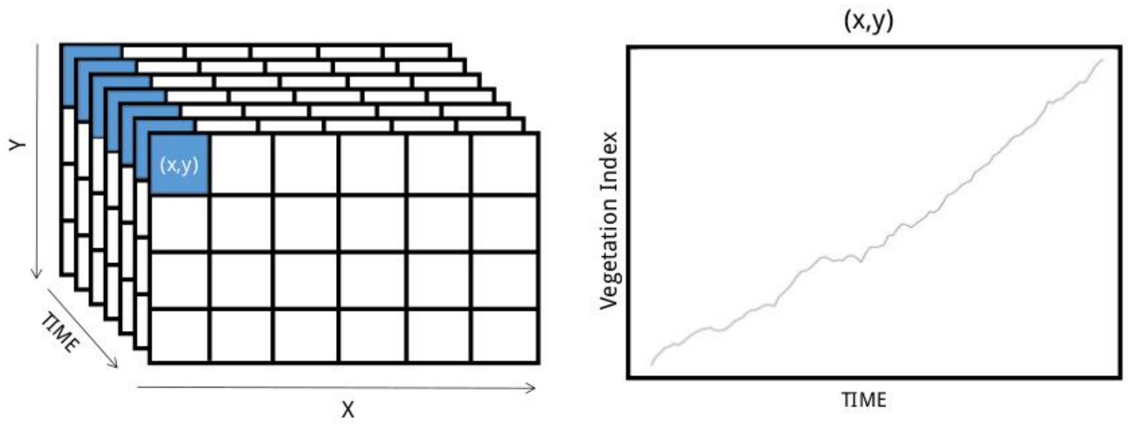
Spatio-Temporal Analysis

Space First Time Later Analysis



Evaluate and compare the results independently for different time instances;

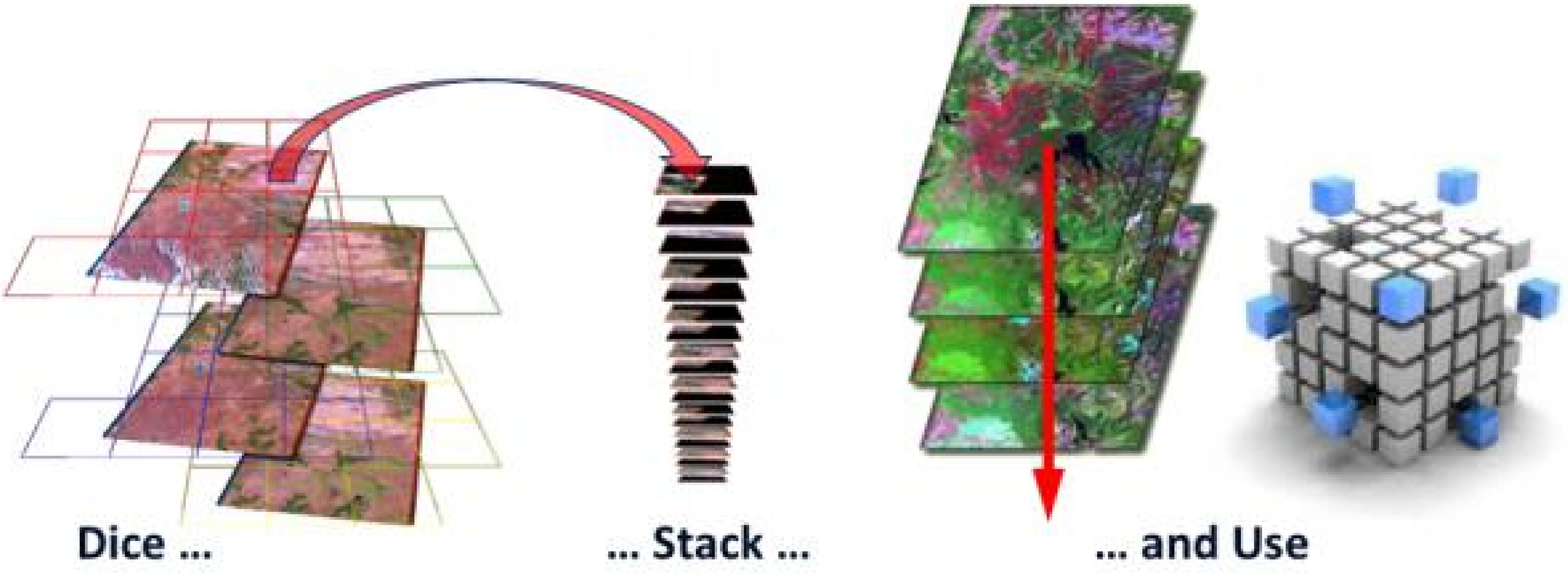
Time First Space Later Analysis



The key is to consider the temporal auto-correlation of the data instead of the spatial auto-correlation.

The importance of a computing infrastructure for analysis

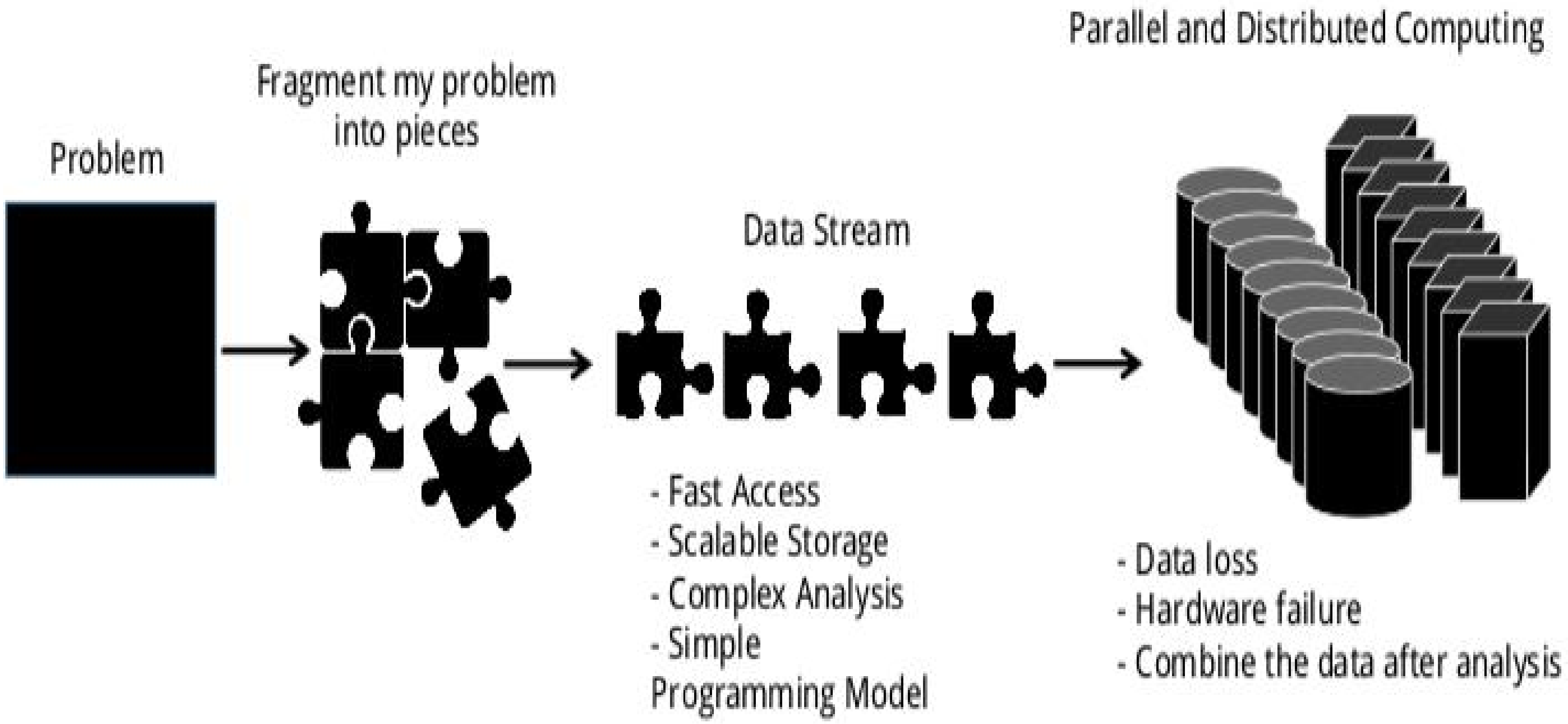
- **Data Cubes worldwide:** Time-series multi-dimensional (space, time, data type) stack of spatially aligned pixels ready for analysis.



The importance of a computing infrastructure for analysis

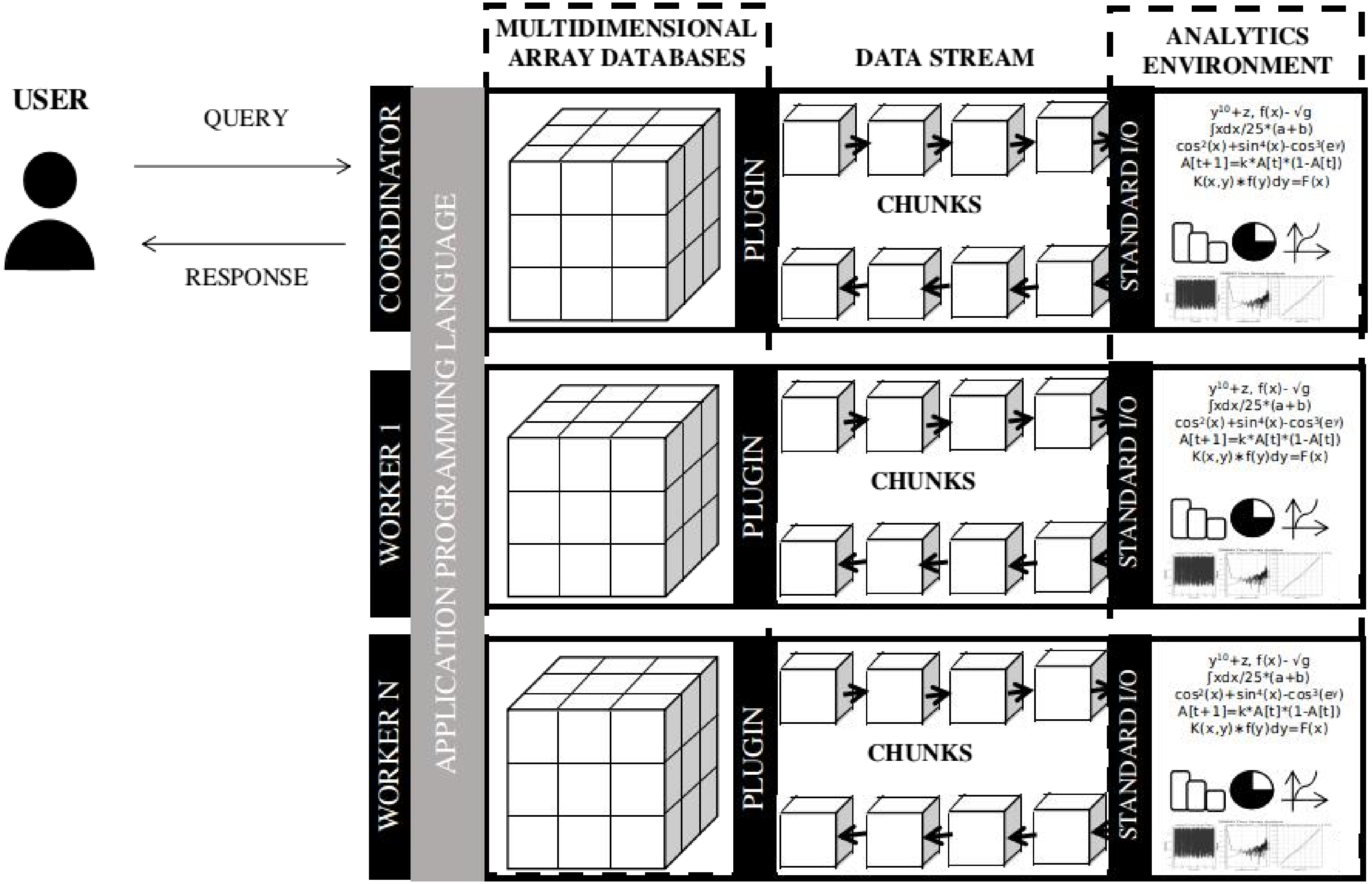
Why server side? It is possible to **move questions** and the **data**. Most of the time it turns out to be **more efficient** to move the questions than to move the data. (Fourth paradigm principles (**Jim Gray, 2005**))

Data streaming processing



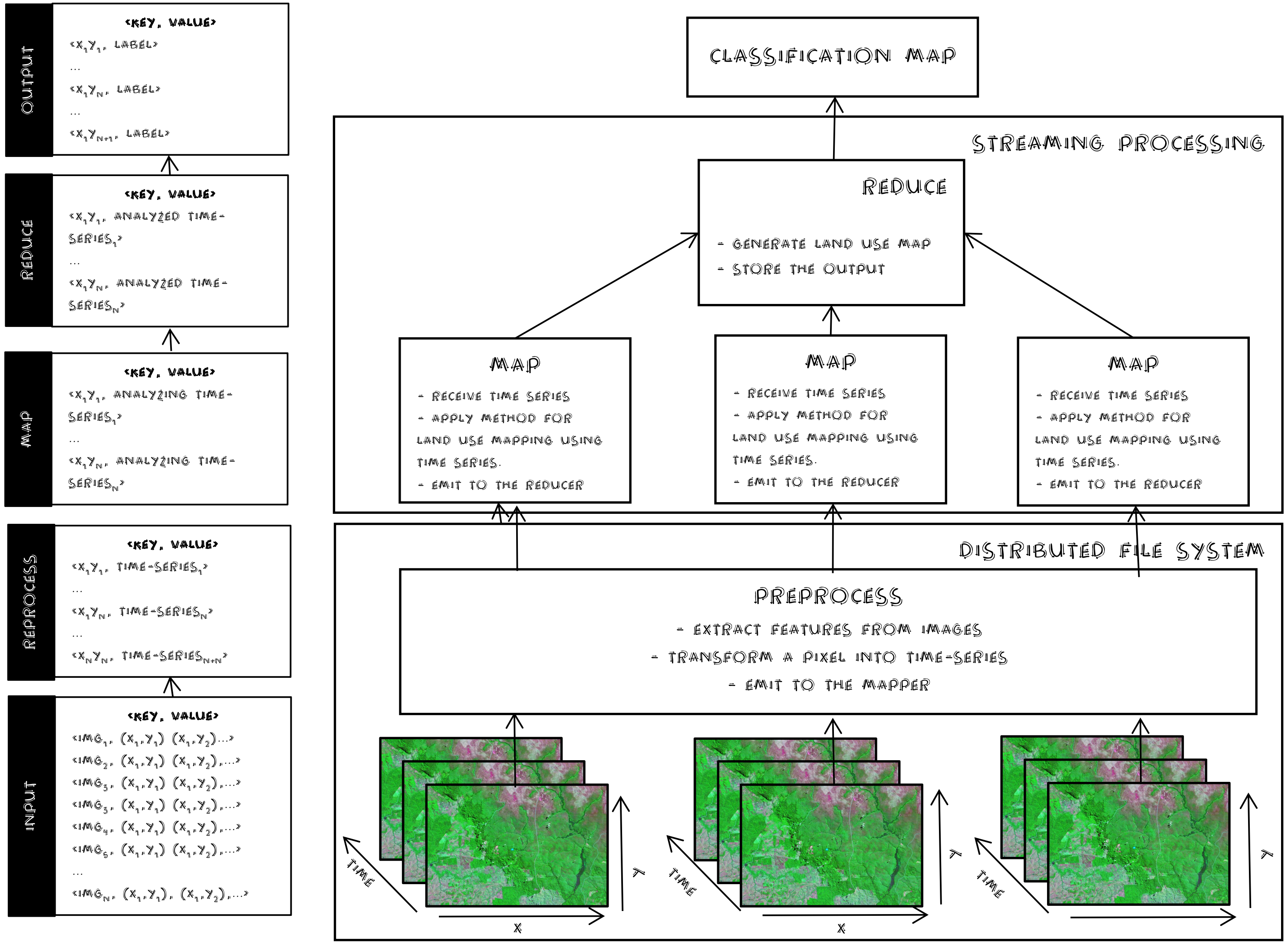
The importance of a computing infrastructure for analysis

Multidimensional Array Databases

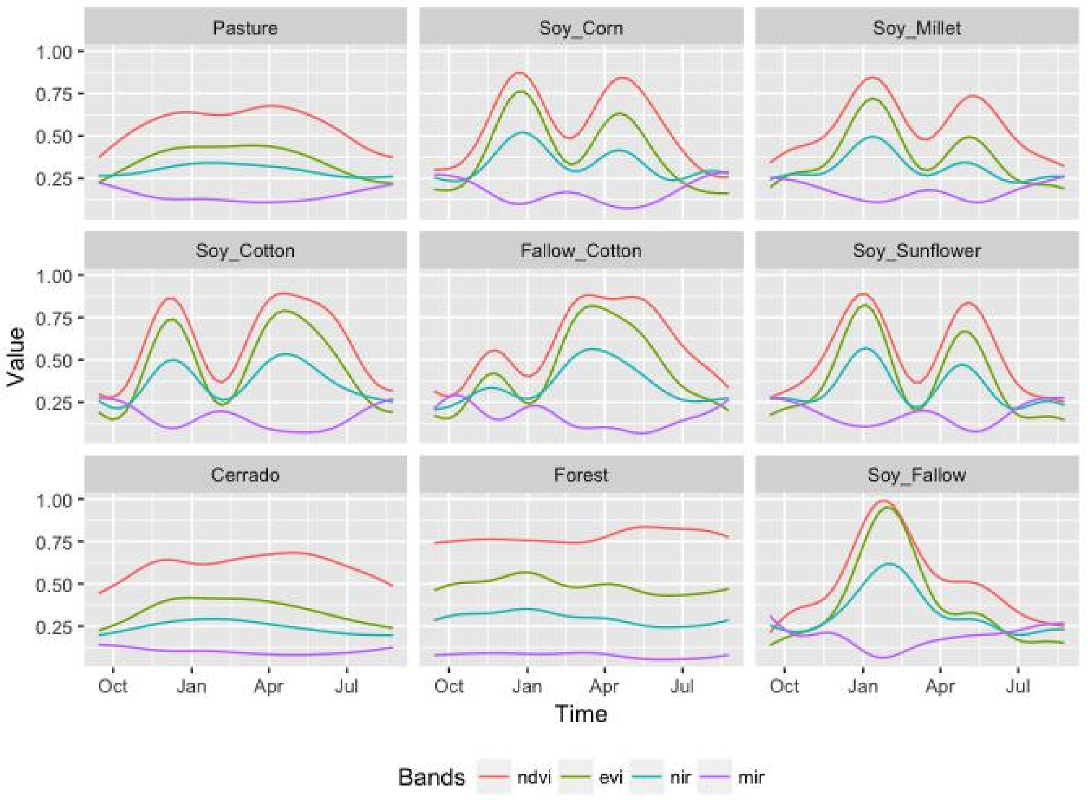


The importance of a computing infrastructure for analysis

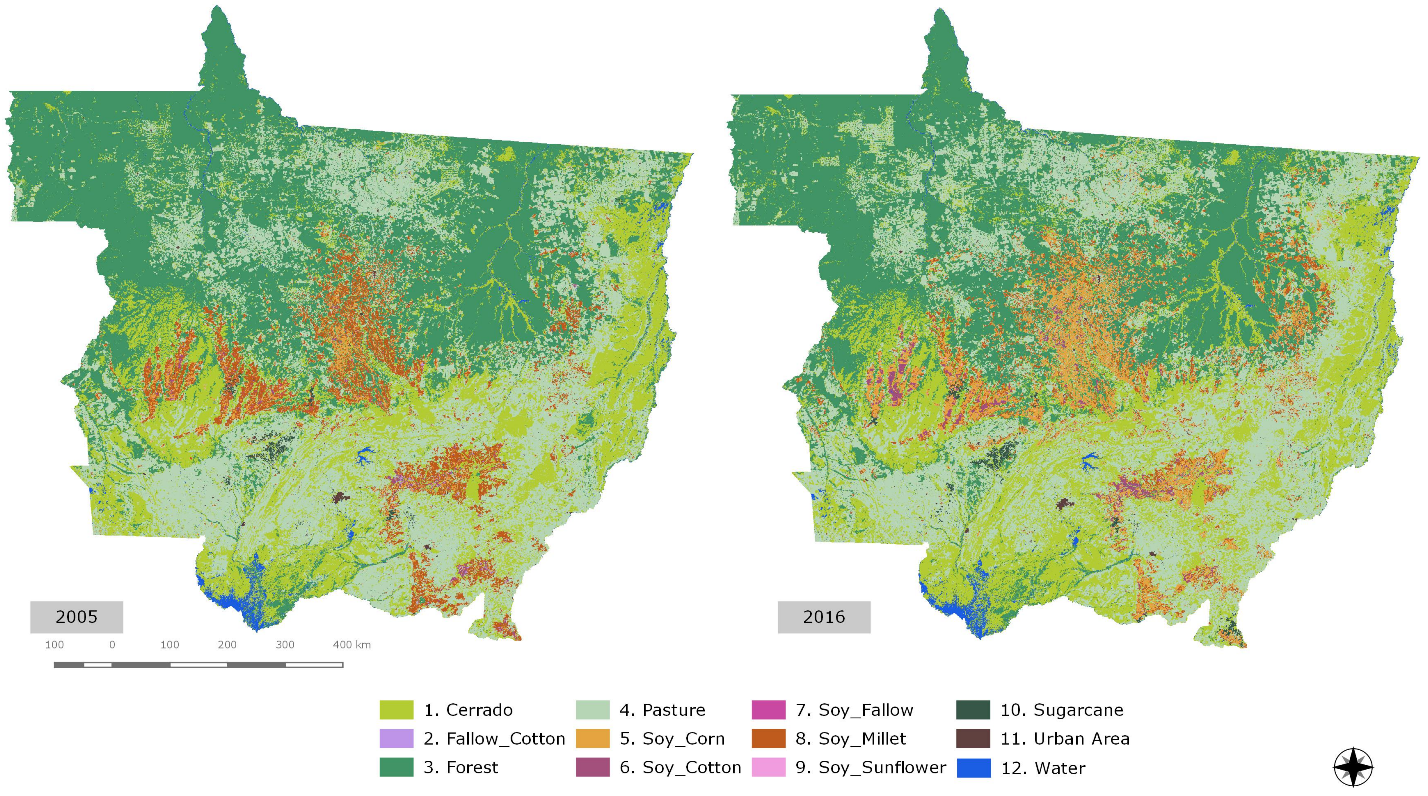
MapReduce Paradigm



Big Earth Observation Data Analytics for Land-Use and Land-Cover Monitoring

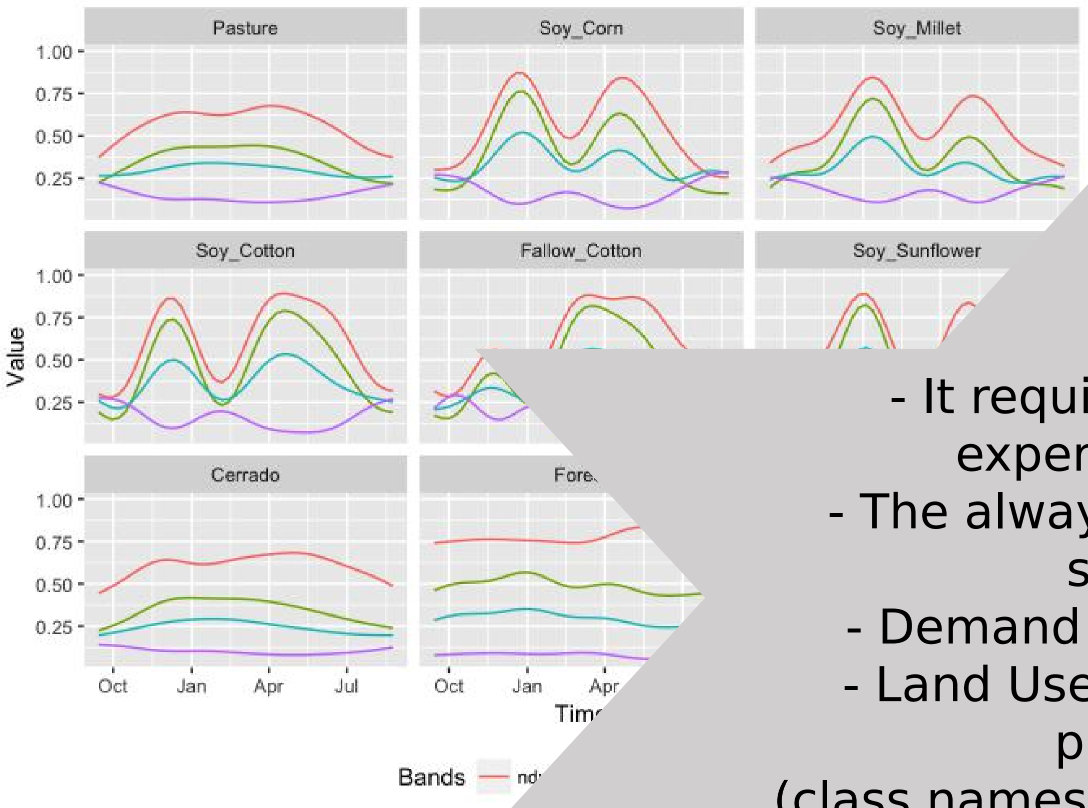


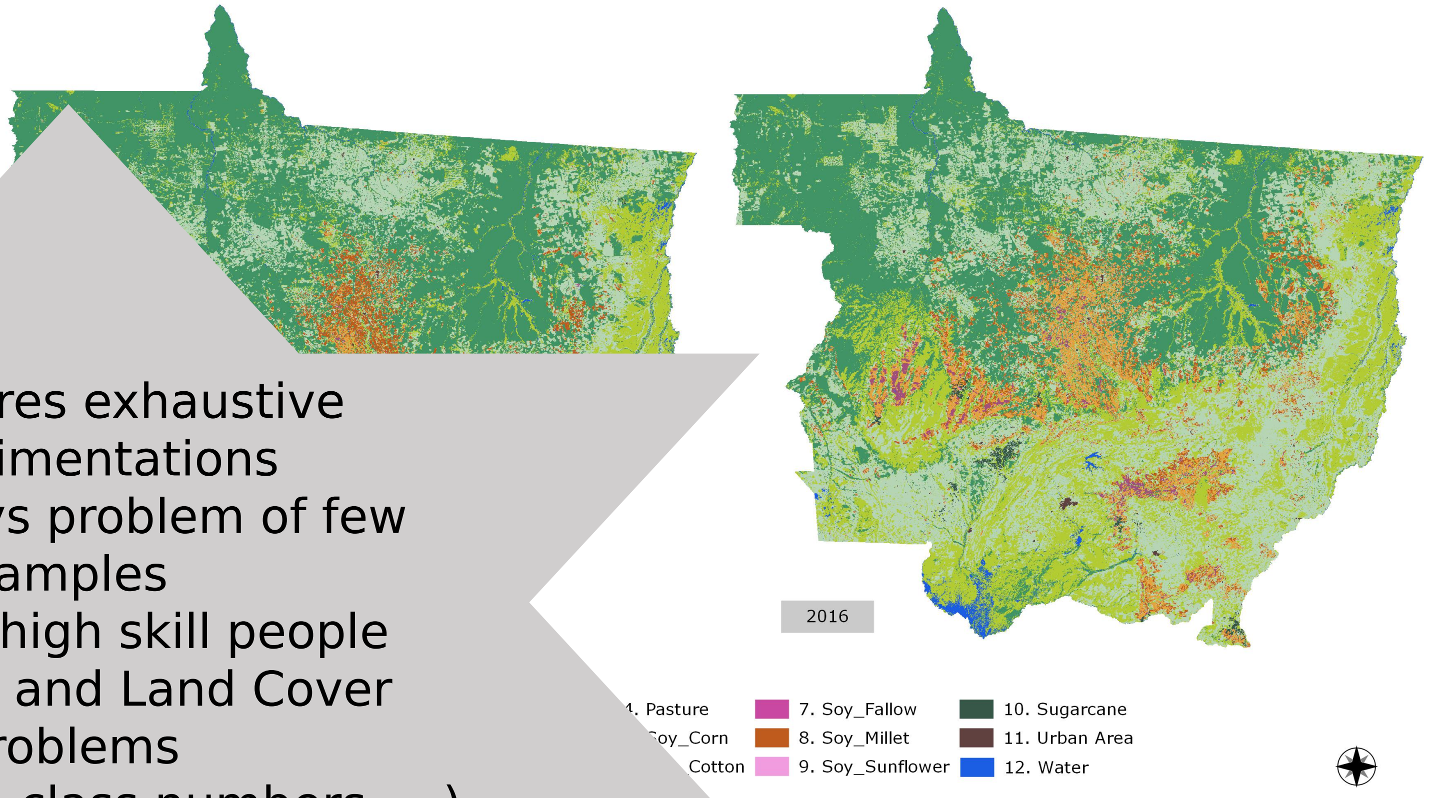
LULCC Time-Series Patterns



Land change dynamics (2001-2017)

Big Earth Observation Data Analytics



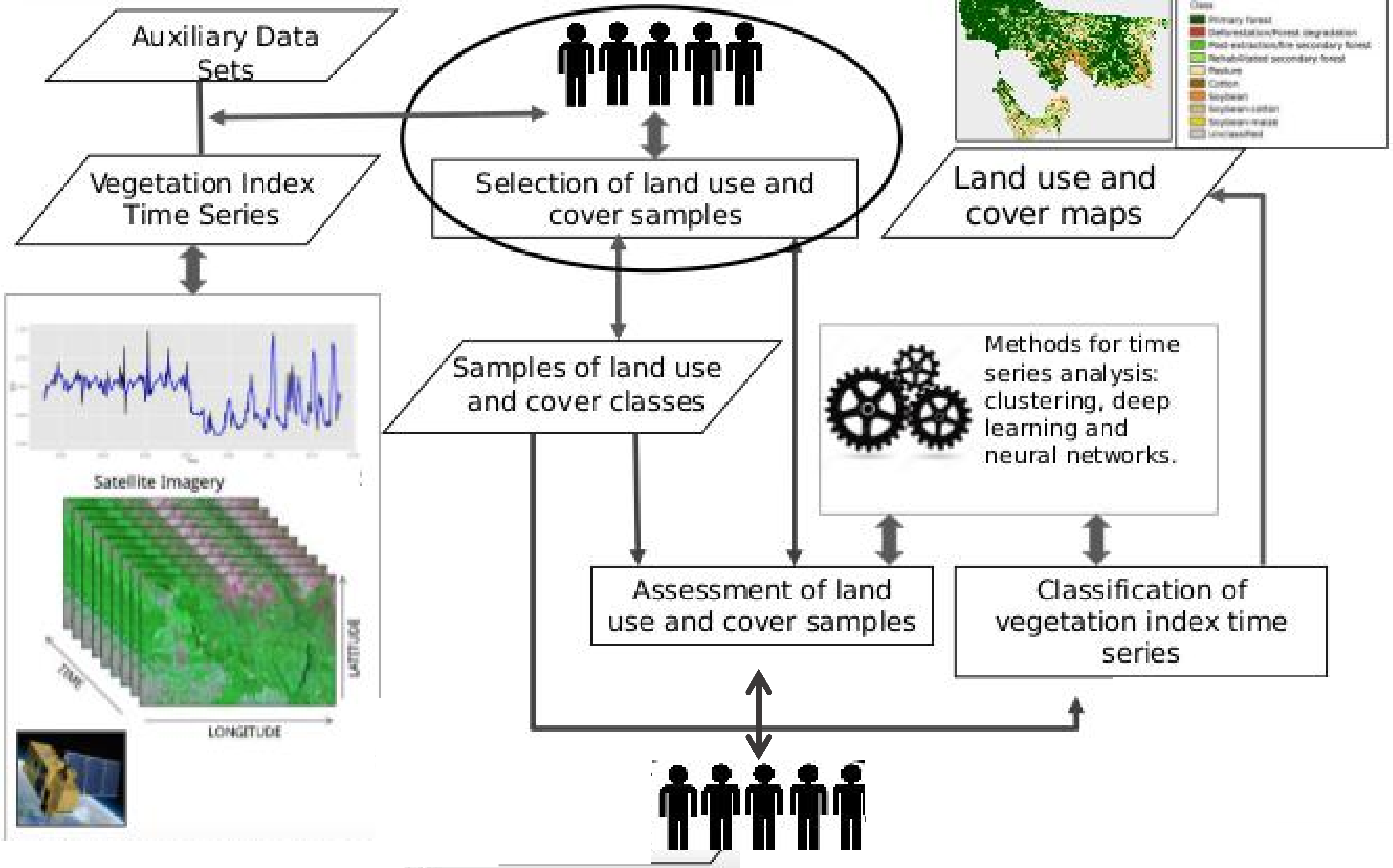


- It requires exhaustive experimentations
- The always problem of few samples
- Demand high skill people
- Land Use and Land Cover problems (class names, class numbers, ...)

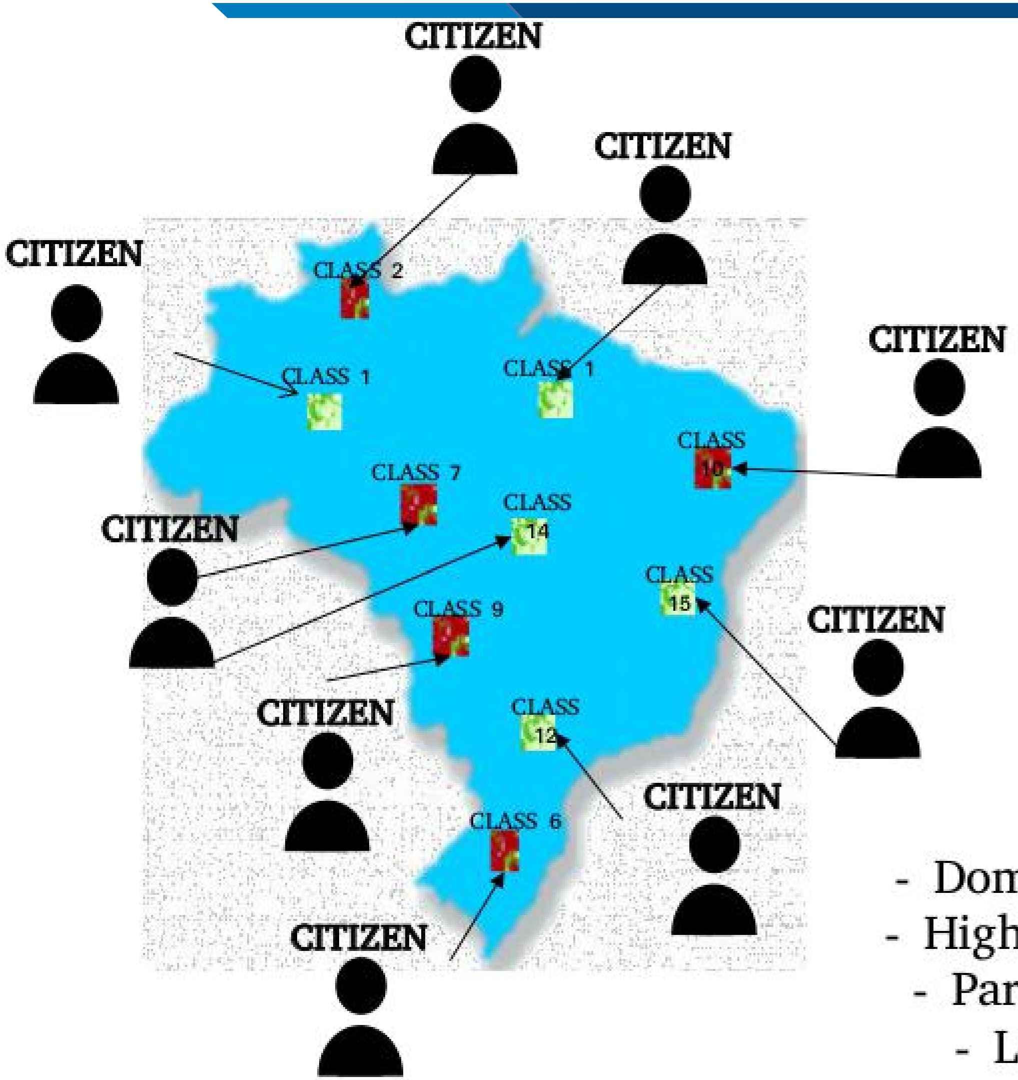
LULCC Time-Series Patterns

Land change dynamics (2001-2017)

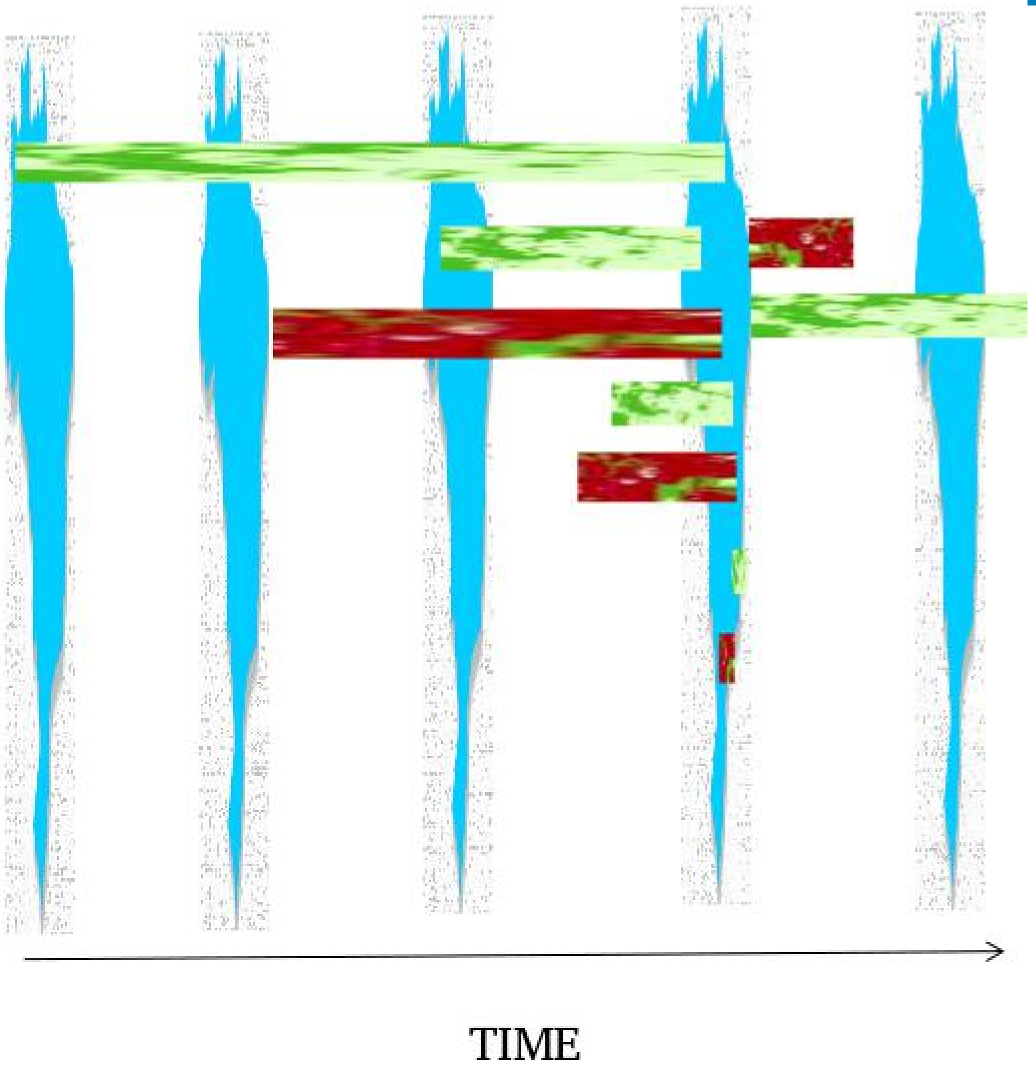
Citizen Science applied in Big Earth Observation Data Analytics



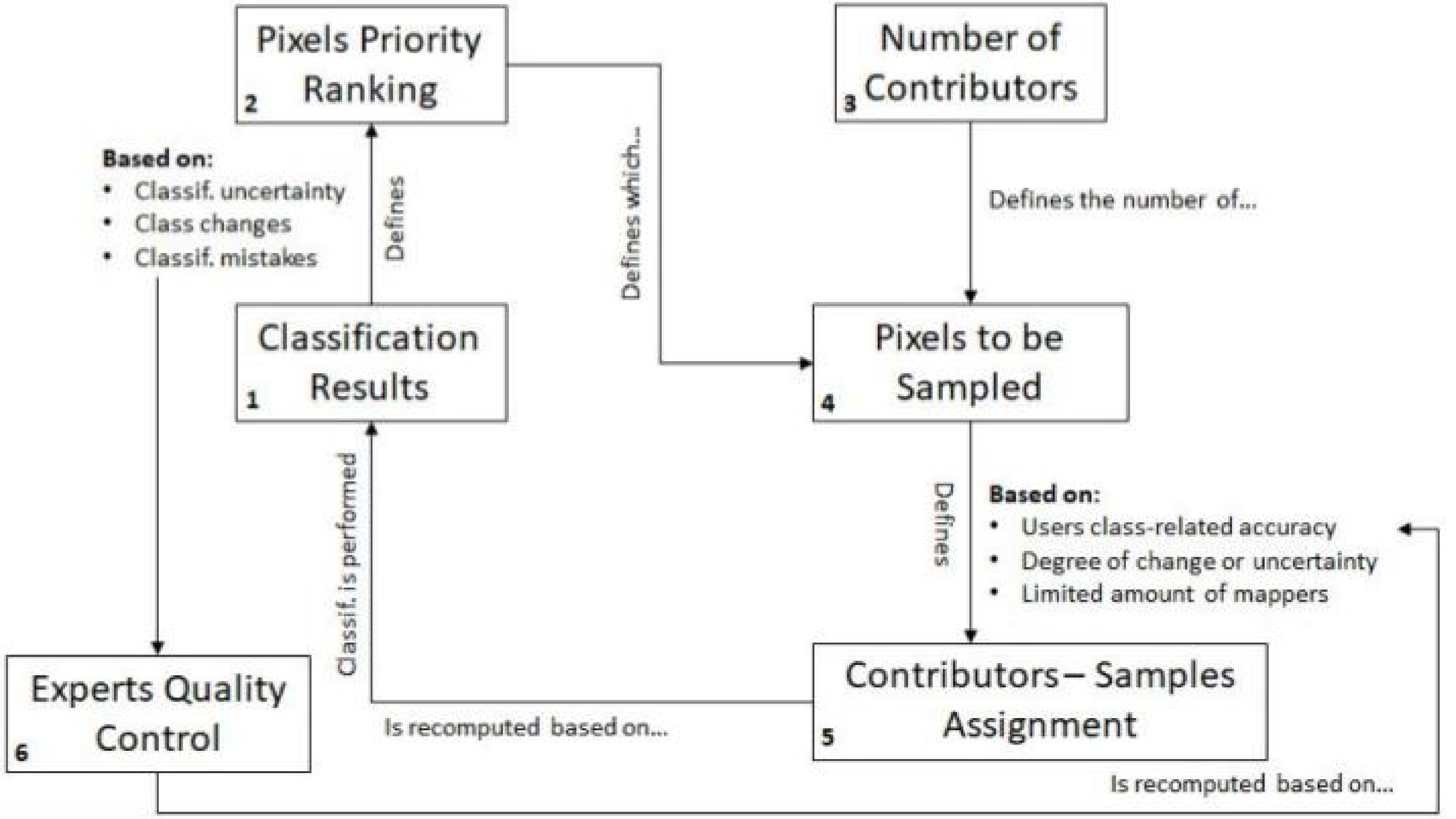
Citizen Science applied in Big Earth Observation Data Analytics - Sampling



- Domain-Aware Citizen
- High cognitive demand
- Participatory process
- Local Knowledge



Citizen Science applied in Big Earth Observation Data Analytics - Validation



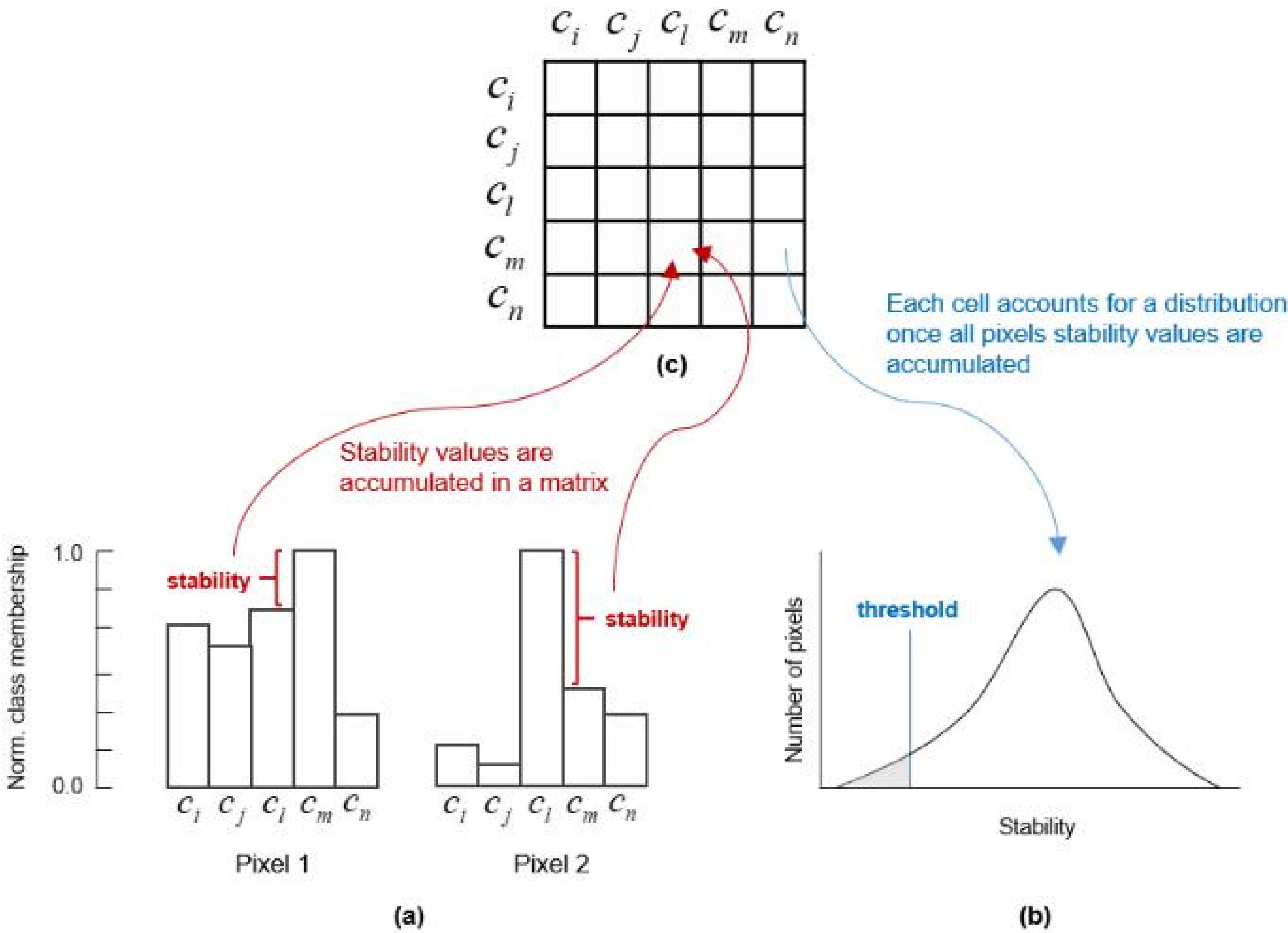
Validation Workflow

Citizen Science applied in Big Earth Observation Data Analytics - Validation

(1) pixels wrongly classified

(2) pixels whose LULC class has changed with respect to the previous classification

(3) pixels whose classification is unstable



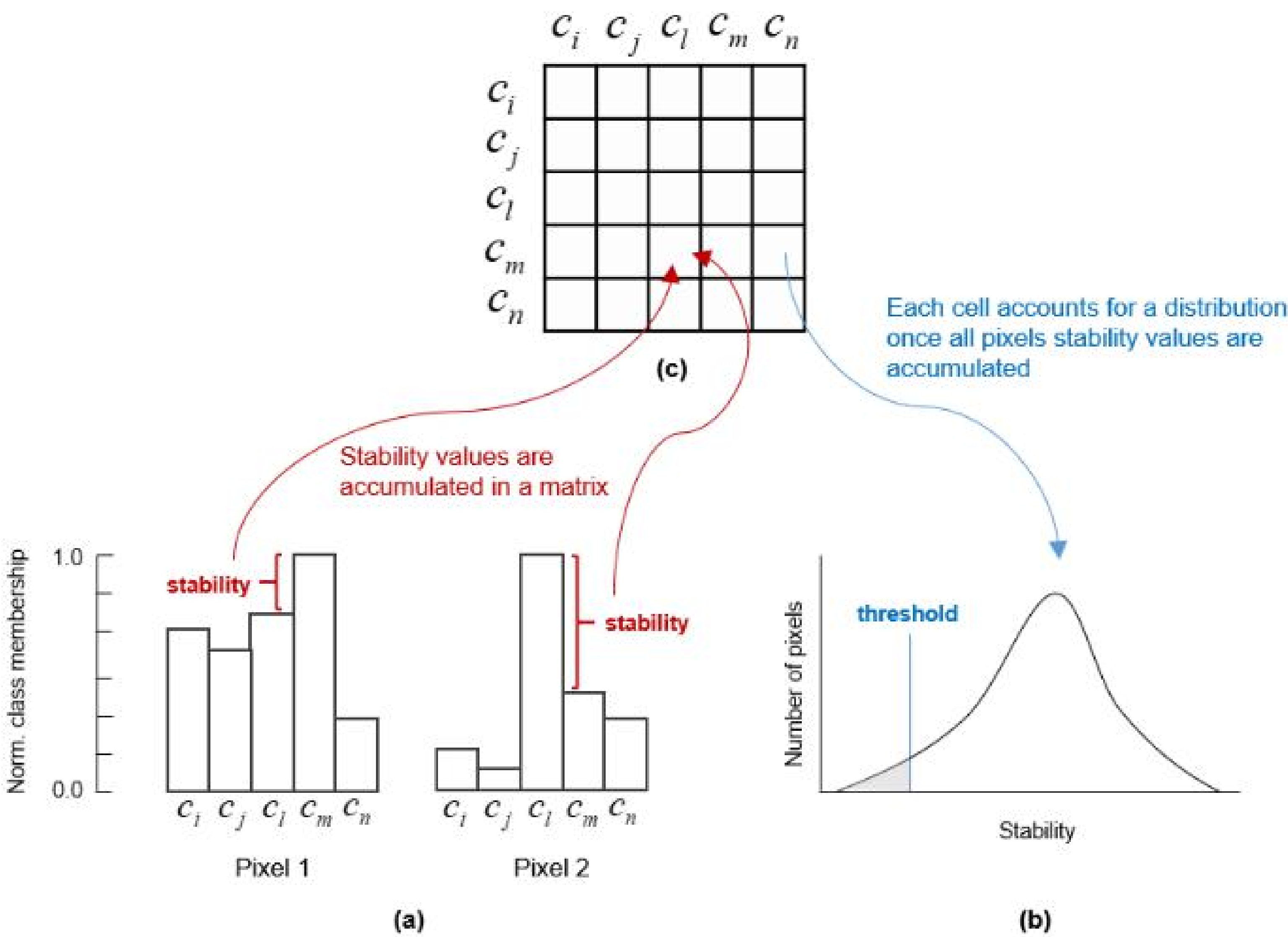
Citizen Science applied in Big Earth Observation Data Analytics - Validation

The third measure refers to pixels whose class membership values, computed by the classification method presented above, are similarly high for two or more classes.

The pixels LULC class stability is a continuous variable.

Normalized class membership distributions for two pixels.

The class stability is computed as the difference between the highest and the second highest normalized class membership values.



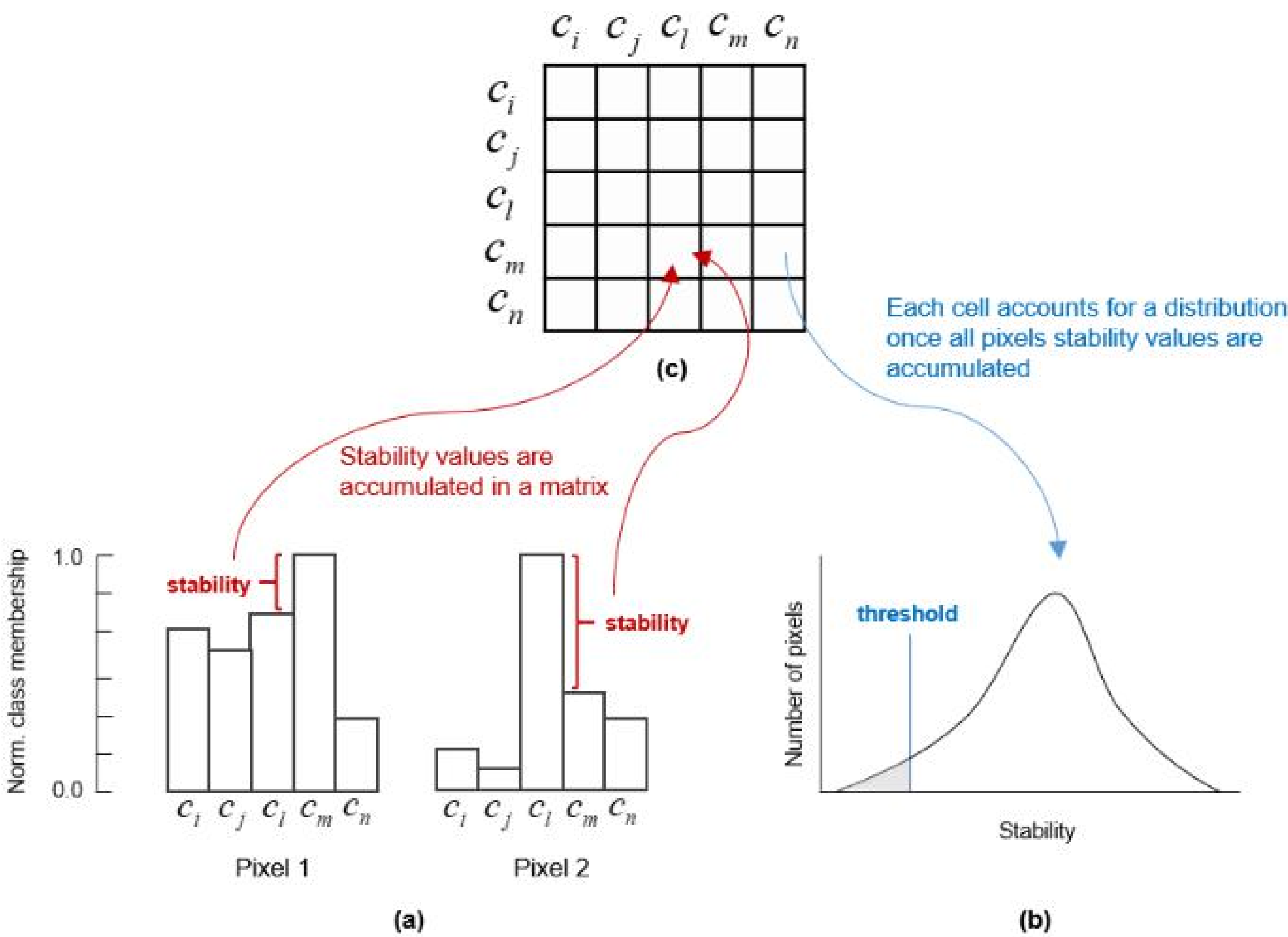
Citizen Science applied in Big Earth Observation Data Analytics - Validation

Thus, the classification stability for pixel 2 in Figure (a) is higher than in pixel 1.

Therefore, priority should be set on inspecting pixel 1 as pixel 2.

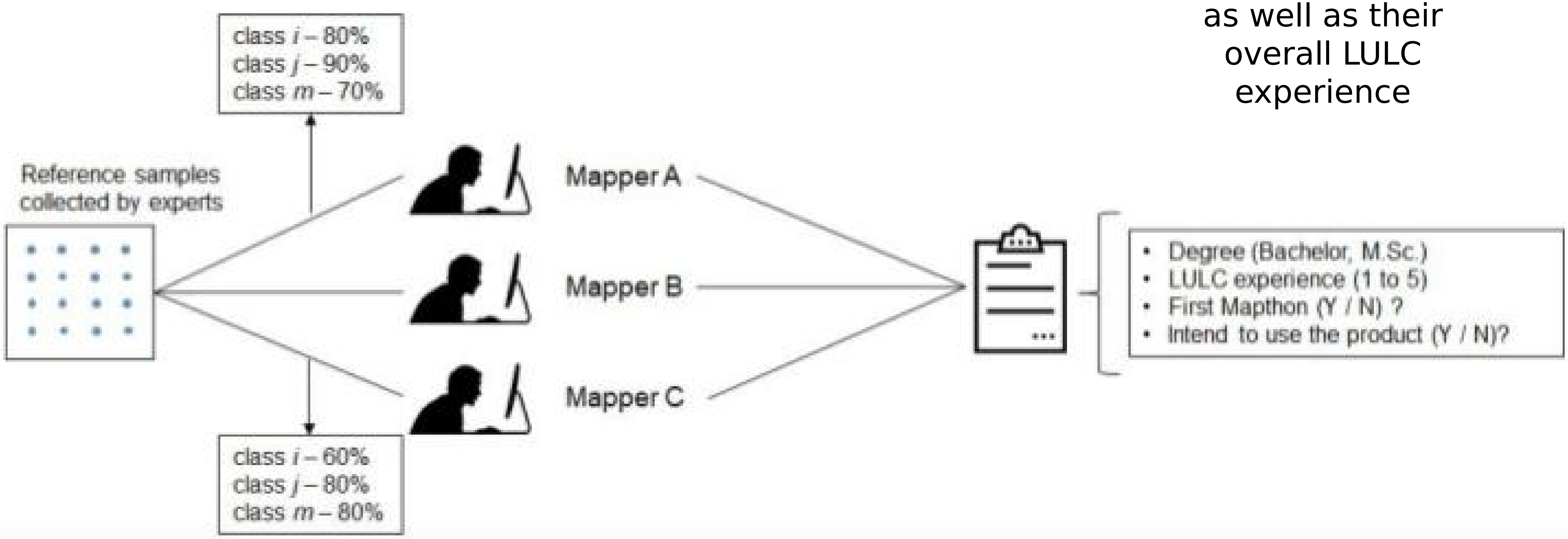
By selecting the n pixels with lowest stability or all pixels underneath a percentile p in the pixels stability distribution (b), we are able detect, according to this measure, pixels that need sampling and further inspection.

The actual values of n and p will vary according to the number of available contributors. This can be class-based also.

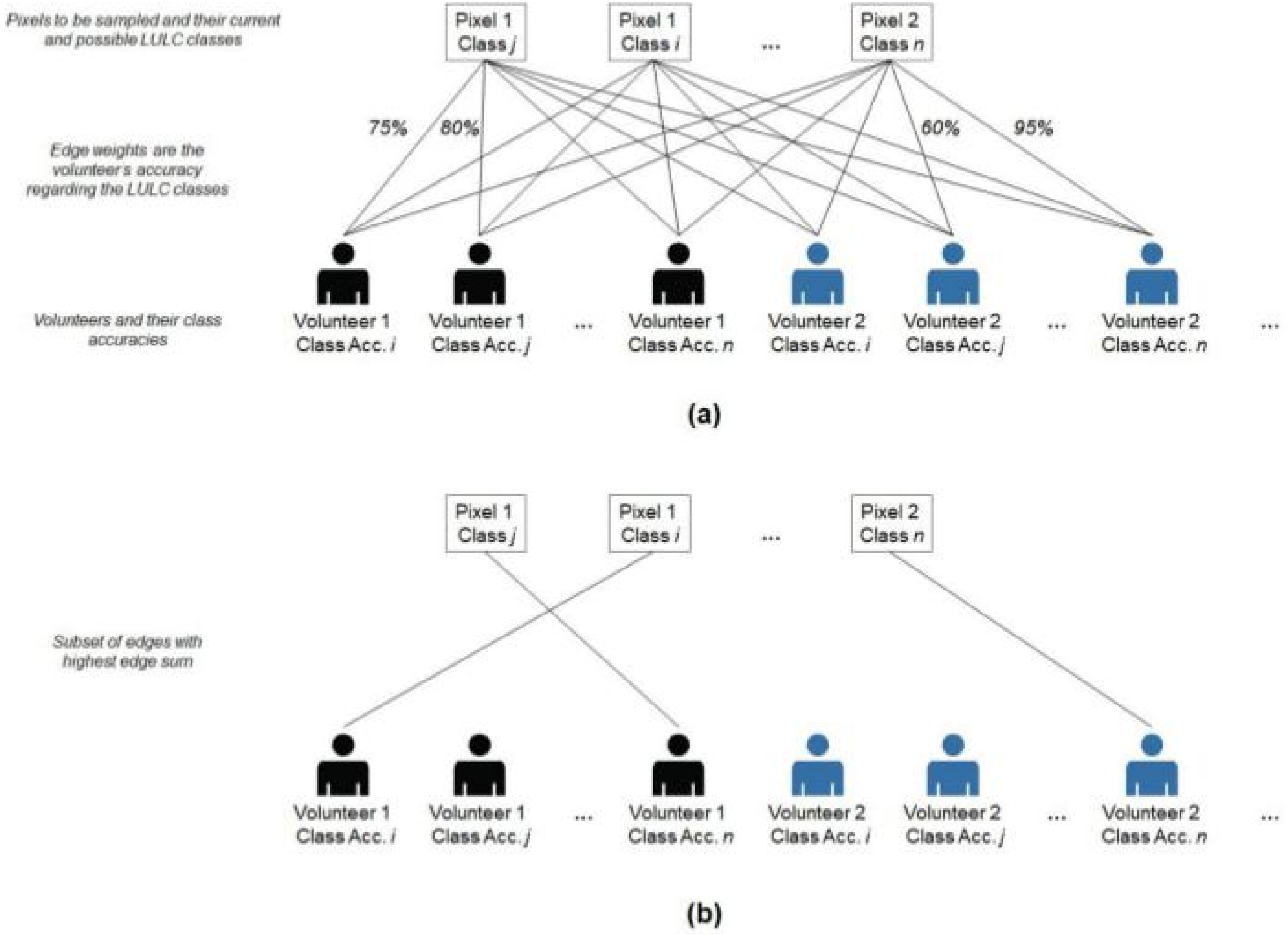


Citizen Science applied in Big Earth Observation Data Analytics - Validation

Considering the contributors class-related accuracies as well as their overall LULC experience



Citizen Science applied in Big Earth Observation Data Analytics



Exemplary Validation Result

Conclusions

The aim of our future work should throw light on questions of very distinct nature:

1. **engaging citizens** (e.g., how to engage and motivate citizens to contribute as a way to ensure the sustainability of the project?);

2. **project sustainability** (e.g., how can citizens be assigned to tasks and their contributions managed in an intelligent manner as a way to optimize the effectiveness/data input relation?);

3. **reliability of information** (e.g., to which extent appropriate architecture designs should help citizens to contribute with reliable LULCC information?);

4. **the use of an specific "technique" such as active learning** (e.g., to what extent crowdsourced data may improve the accuracy of machine learning algorithms within big EO data architectures?);

5. **passing by decision-making support** (e.g., what are the possibilities and constraints in terms of remote sensing and auxiliary data inputs to a crowd-assisted LULCC monitoring system?).

Thank you!

References

ASSIS, Luiz Fernando et al. Big data streaming for remote sensing time series analytics using MapReduce. In: GeoInfo. 2016. p. 228-239.

CAMARA, Gilberto et al. Big earth observation data analytics: matching requirements to system architectures. In: Proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data. ACM, 2016. p. 1-6.

JOSHI, Neha et al. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. Remote Sensing, v. 8, n. 1, p. 70, 2016.

MAUS, V., Câmara, G., Cartaxo, R., Sanchez, A., Ramos, F. M., and de Queiroz, G. R. (2016). A time-weighted dynamic time warping method for land-use and land-cover mapping. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.