Fusion, mining and unmixing of EO data: challenges and perspectives

Paolo Gamba

University of Pavia

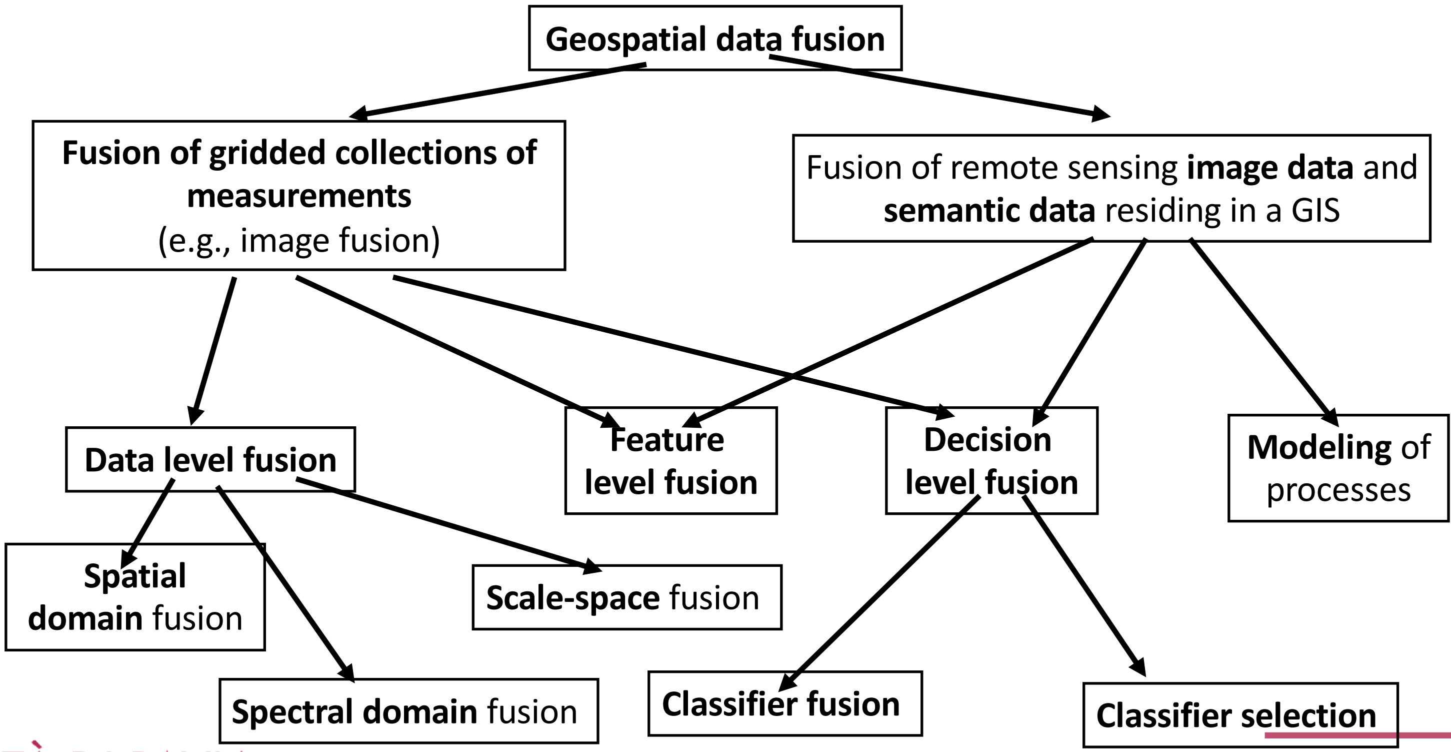
UNIVERSITÀ DI PAVIA

EO data ←→ EO challenges

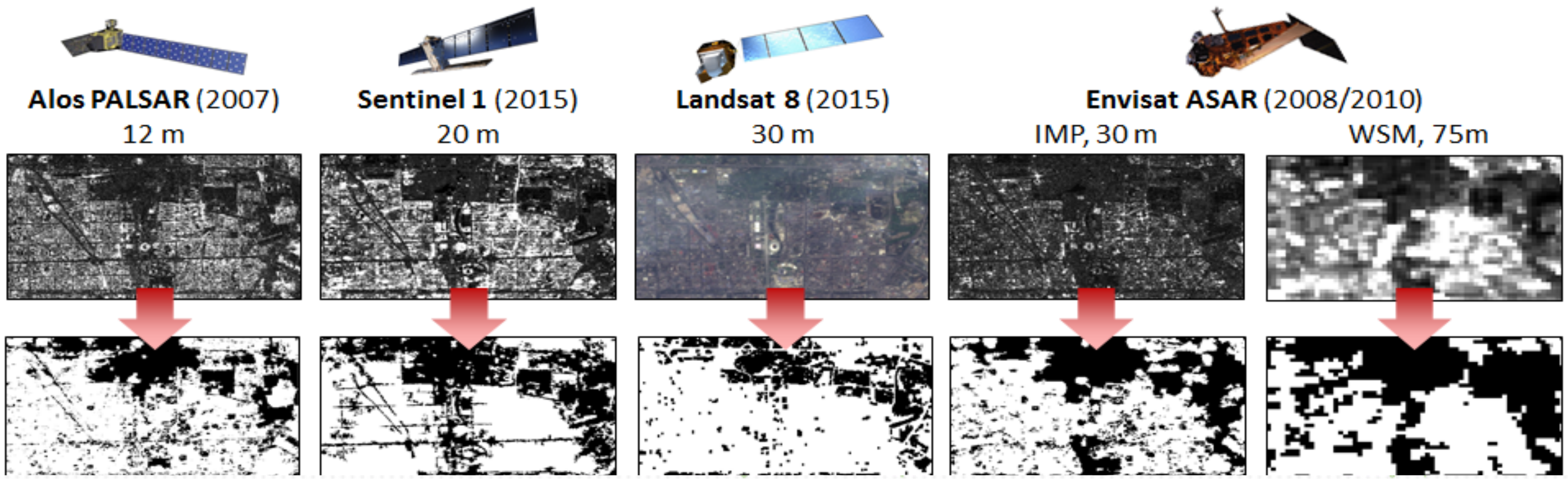
- Earth Observation data represent a huge amount of information with multiple dimensions (spatial, spectral, temporal).
- They also represent the ultimate result of the interaction between electromagnetic waves and the materials of the Earth surface.
- Inverting the process and extracting biophysical parameters is the biggest challenge for EO data, and this where tools such as Quantum Computing are extremely valuable.

Geospatial data fusion

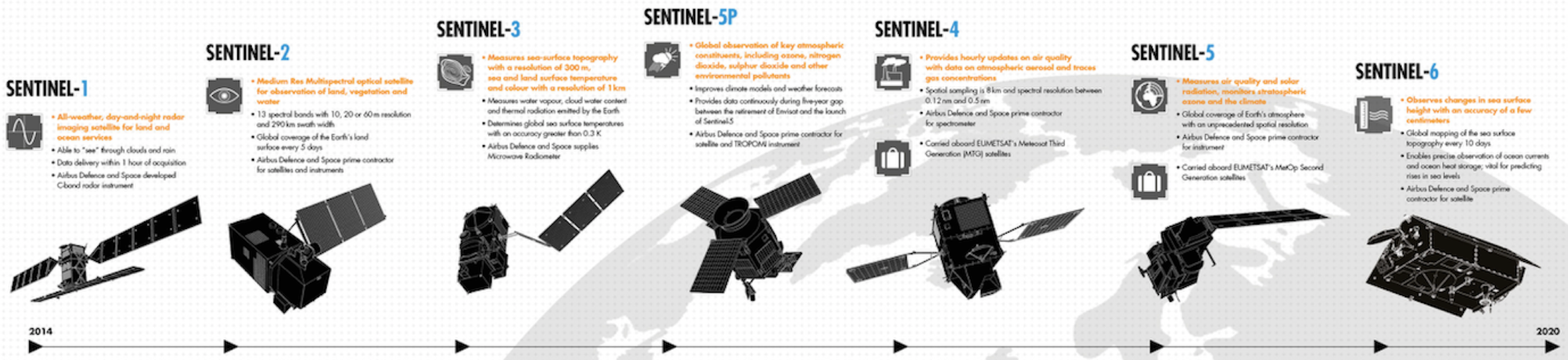
- There is no «one fits all» solution



SAR, optical, ...

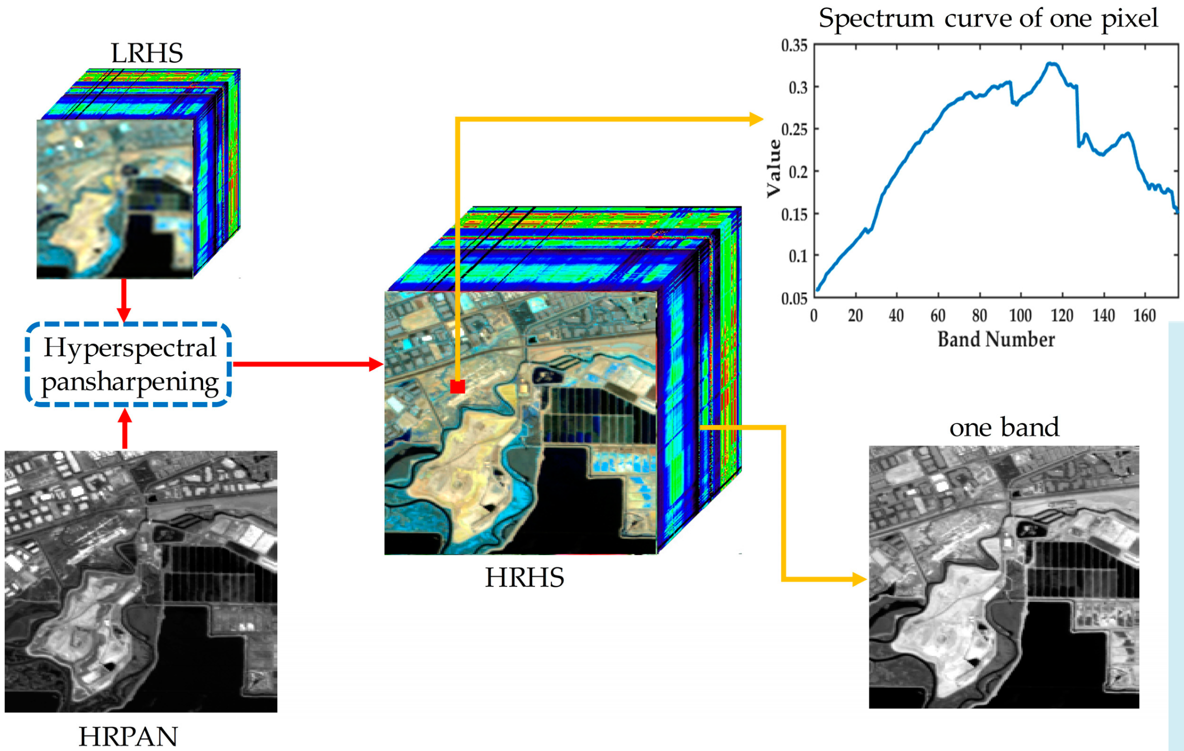


Alos PALSAR (2007)
12 m
Sentinel 1 (2015)
20 m
Landsat 8 (2015)
30 m
Envisat ASAR (2008/2010)
IMP, 30 m
WSM, 75m

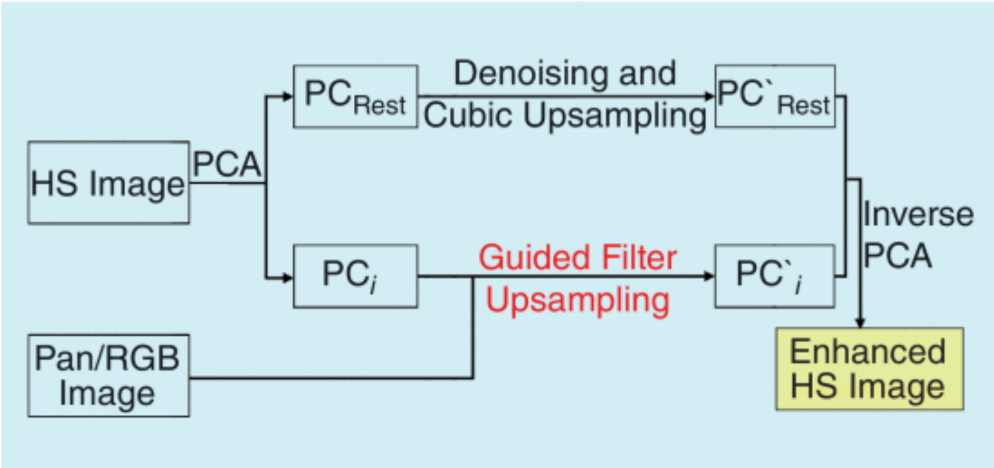


SENTINEL-1
• All-weather, day-and-night radar imaging satellite for land and ocean services
• Able to "see" through clouds and rain
• Data delivery within 1 hour of acquisition
• Airbus Defence and Space developed C-band radar instrument
SENTINEL-2
• Medium Res Multispectral optical satellite for observation of land, vegetation and water
• 13 spectral bands with 10, 20 or 60 m resolution and 290 km swath width
• Global coverage of the Earth's land surface every 5 days
• Airbus Defence and Space prime contractor for satellites and instruments
SENTINEL-3
• Measures sea-surface topography with a resolution of 300 m, sea and land surface temperature and colour with a resolution of 1 km
• Measures water vapour, cloud water content and thermal radiation emitted by the Earth
• Determines global sea surface temperatures with an accuracy greater than 0.3 K
• Airbus Defence and Space supplies Microwave Radiometer
SENTINEL-5P
• Global observation of key atmospheric constituents, including ozone, nitrogen dioxide, sulphur dioxide and other environmental pollutants
• Improves climate models and weather forecasts
• Provides data continuously during five-year gap between the retirement of Envisat and the launch of Sentinel-5
• Airbus Defence and Space prime contractor for satellite and TROPOMI instrument
SENTINEL-4
• Provides hourly updates on air quality with data on atmospheric aerosol and traces gas concentrations
• Spatial sampling is 8 km and spectral resolution between 0.12 nm and 0.5 nm
• Airbus Defence and Space prime contractor for spectrometer
• Carried aboard EUMETSAT's Meteosat Third Generation (MTG) satellites
SENTINEL-5
• Measures air quality and solar radiation, monitors stratospheric ozone and the climate
• Global coverage of Earth's atmosphere with an unprecedented spatial resolution
• Airbus Defence and Space prime contractor for instrument
• Carried aboard EUMETSAT's MetOp Second Generation satellites
SENTINEL-6
• Observes changes in sea surface height with an accuracy of a few centimeters
• Global mapping of the sea surface topography every 10 days
• Enables precise observation of ocean currents and ocean heat storage; vital for predicting rises in sea levels
• Airbus Defence and Space prime contractor for satellite
2014
2020

Pansharpening …

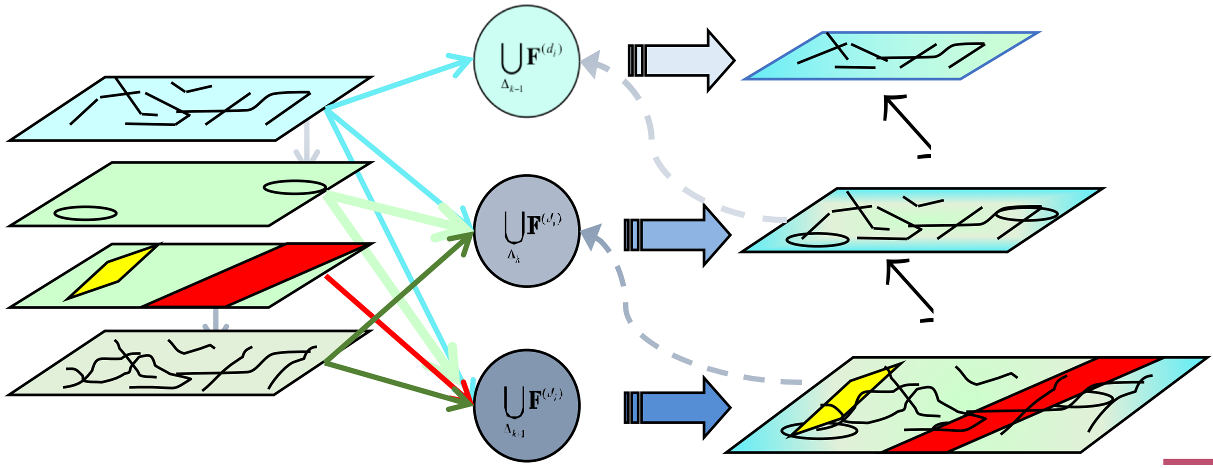


… towards hyperspectral pansharpening



Multiresolution/temporal/frequency/… fusion

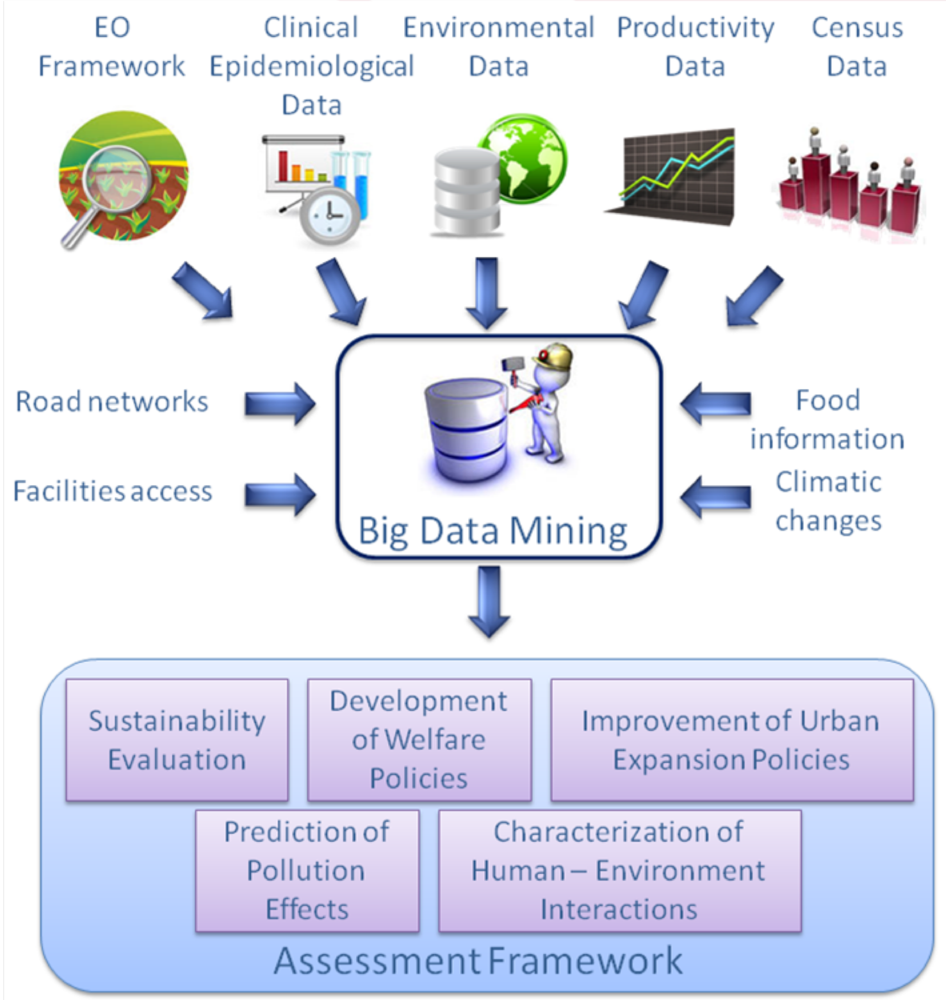
- Multiple scales, multiple time instances, multiple frequencies, multiple polarizations, … could and should be used
- The information from each data set reconciled so that features extracted from one data set match with the similar extractions using other data sets and, at the same time, help infer more refined features at other spatial resolutions.



And then ... non-EO data

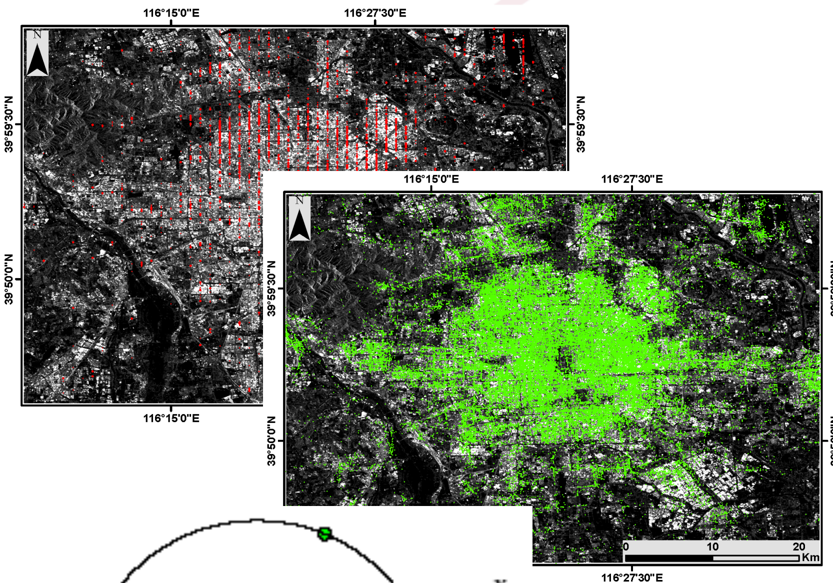
Images, social media

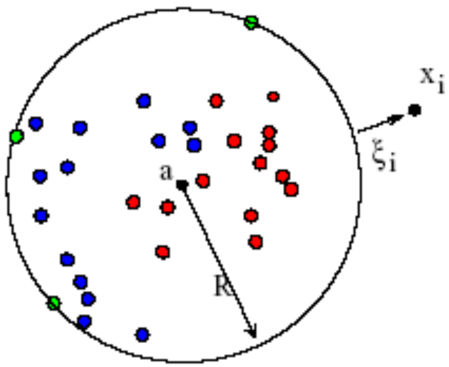
... but not only!

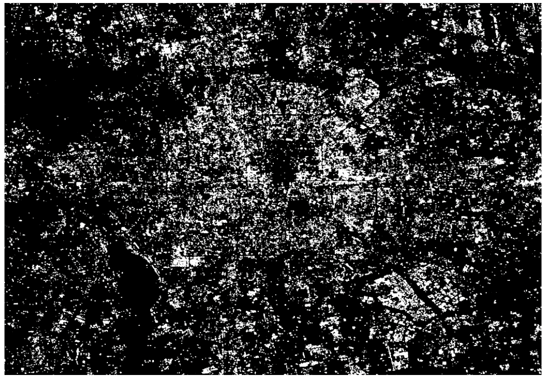


UNIVERSITÀ DI PAVIA

Urban extents from Tweets



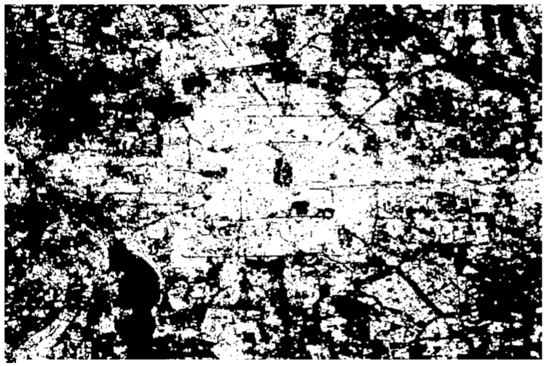




Impervious surface extraction using RF & Twitter.

Impervious surface extraction using RF & Weibo.

Impervious surface extraction using Clustering and Weibo.

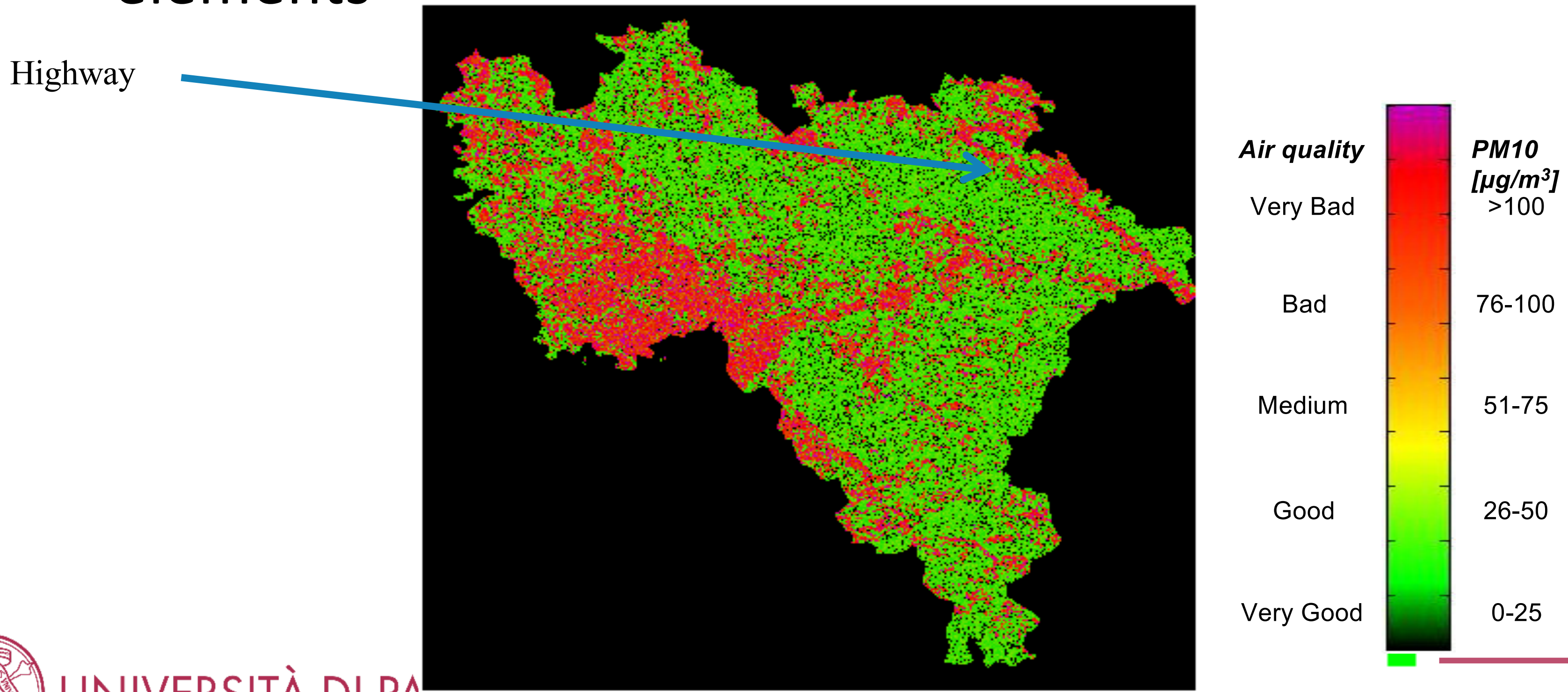


Impervious surface extraction using EO data only.

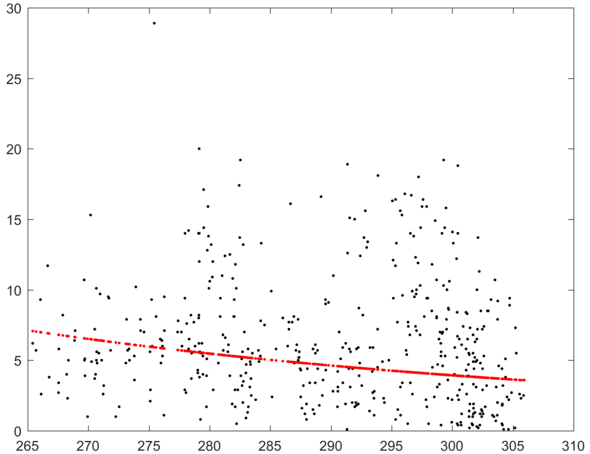
UNIVERSITÀ DI PAVIA

From impervious surfaces to pollution

- Using the thermal band, and considering artificial landscape elements



From Thermal IR to $PM_{2.5}$

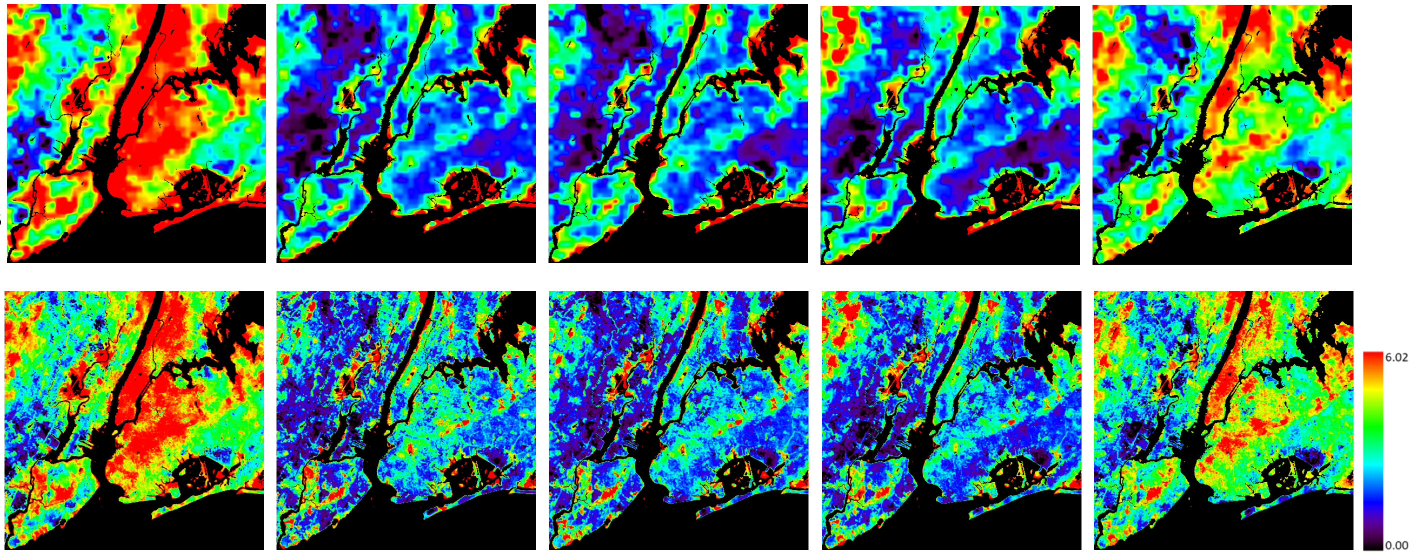


$PM_{2.5} = exp(28.7012) * (BT)^{-4.7920}$

The $PM_{2.5}$ concentration decreases with the increase of at-satellite BT, since more $PM_{2.5}$ impose more blocking effects on the energy transformation process from the land surface to the satellite.

The $PM_{2.5}$ measurements by 11 ground stations on 46 dates are used to model the regression between the MODIS data and the measured $PM_{2.5}$ concentrations observed at the closest time point to the MODIS imaging time on the 46 dates.

MODIS $PM_{2.5}$ (1000m)

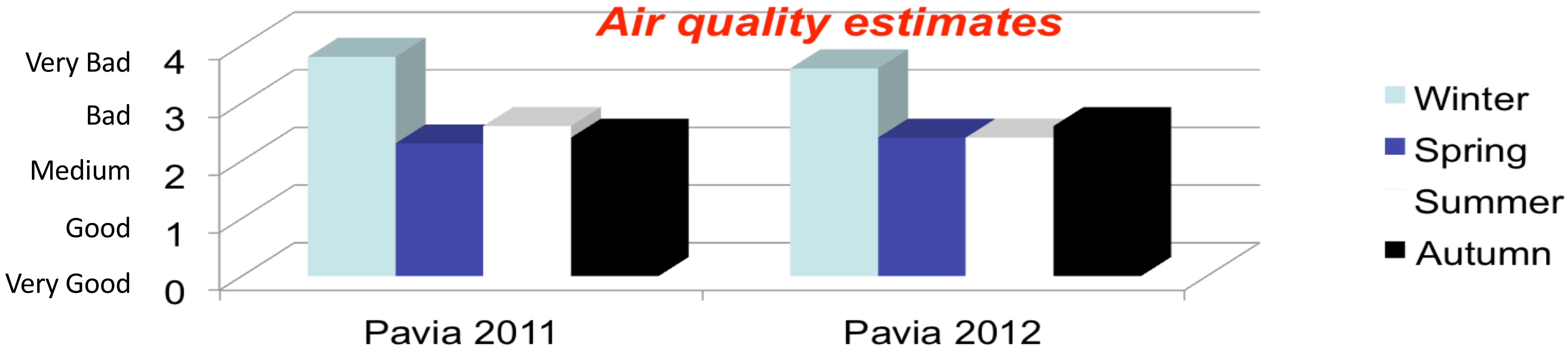


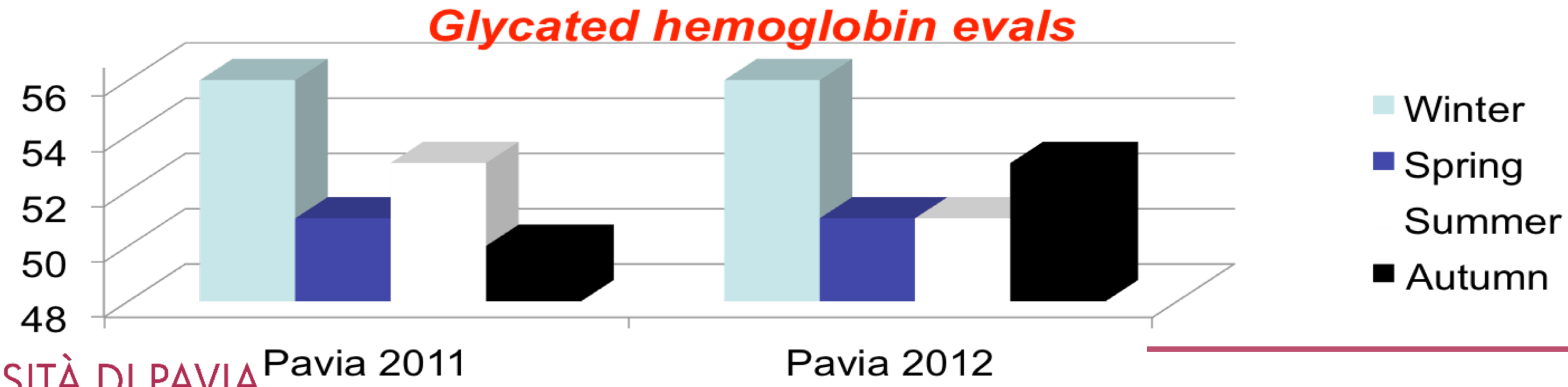
Landsat-like $PM_{2.5}$ (100m)

05 April | 10 June | 19 July | 22 August | 14 October

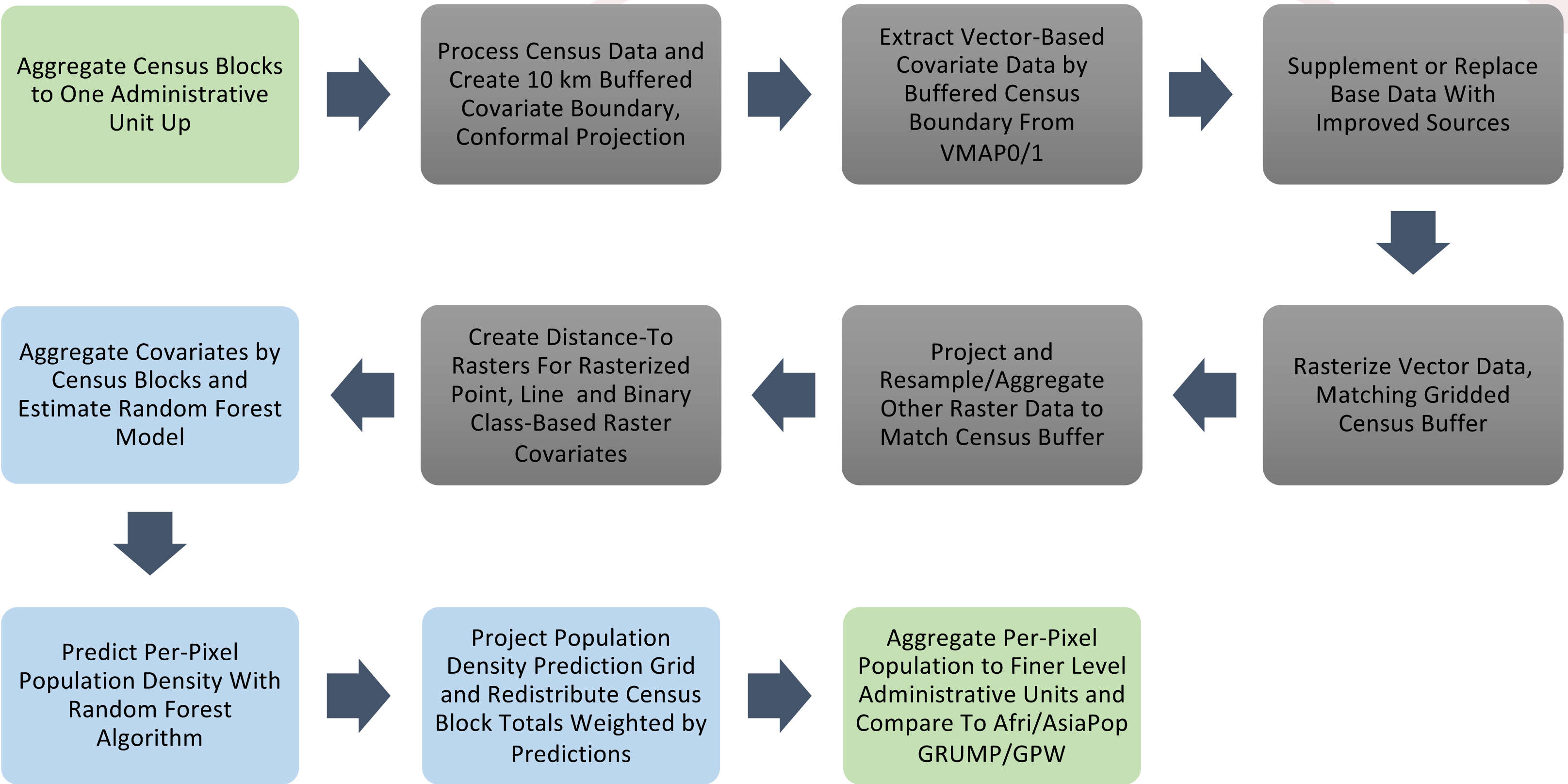
UNIVERSITÀ DI PAVIA

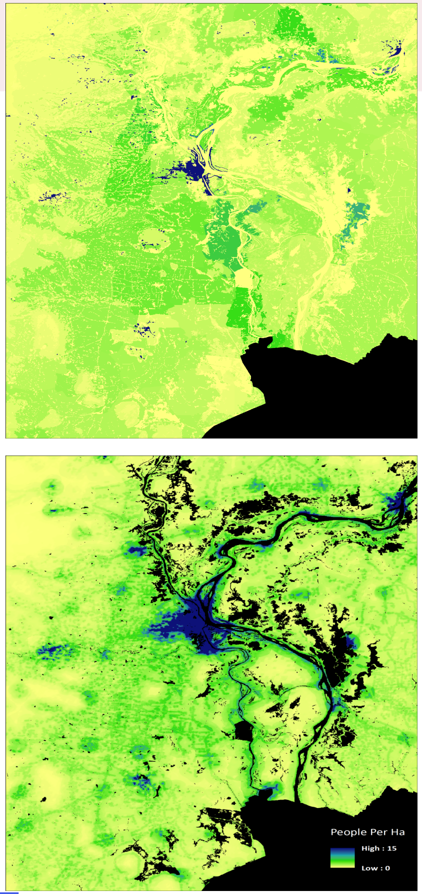
Air quality & medical records





Fine grid population maps

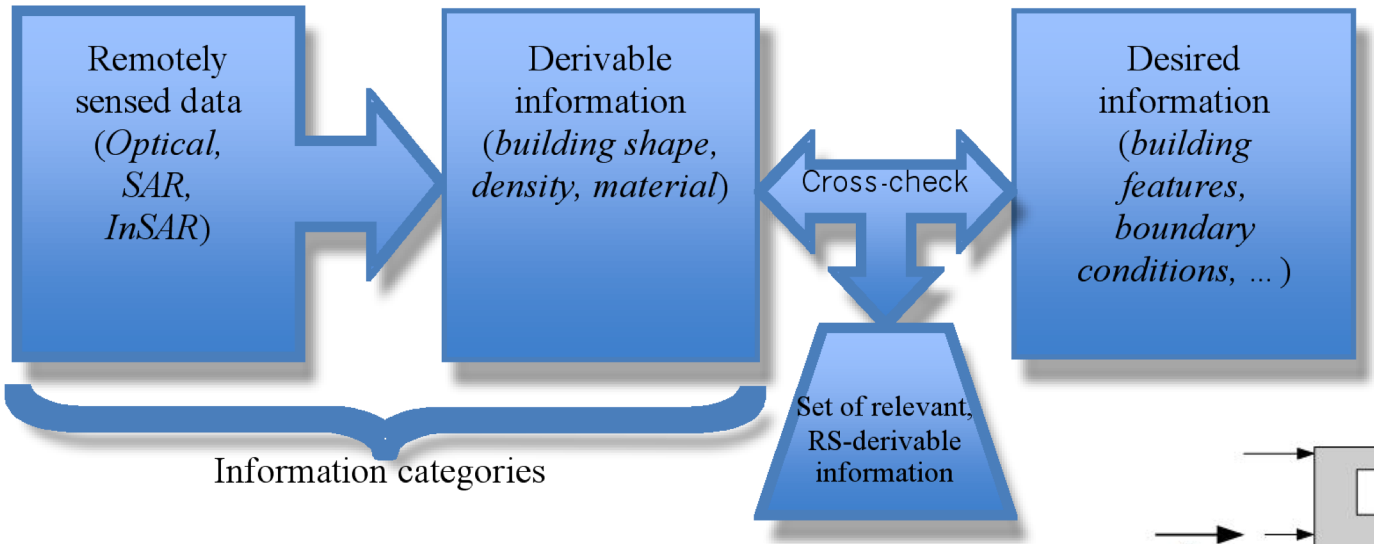


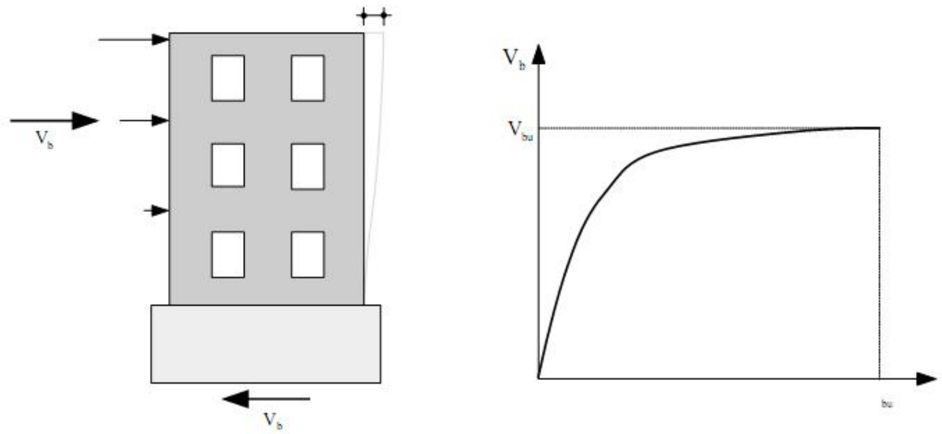


Stevens, F.R., Gaughan, A.E., Linard, C., and A.J. Tatem. Disaggregating census data for population mapping using Random Forests with remotely-sensed and other ancillary data. *In Review*: *Plosone*

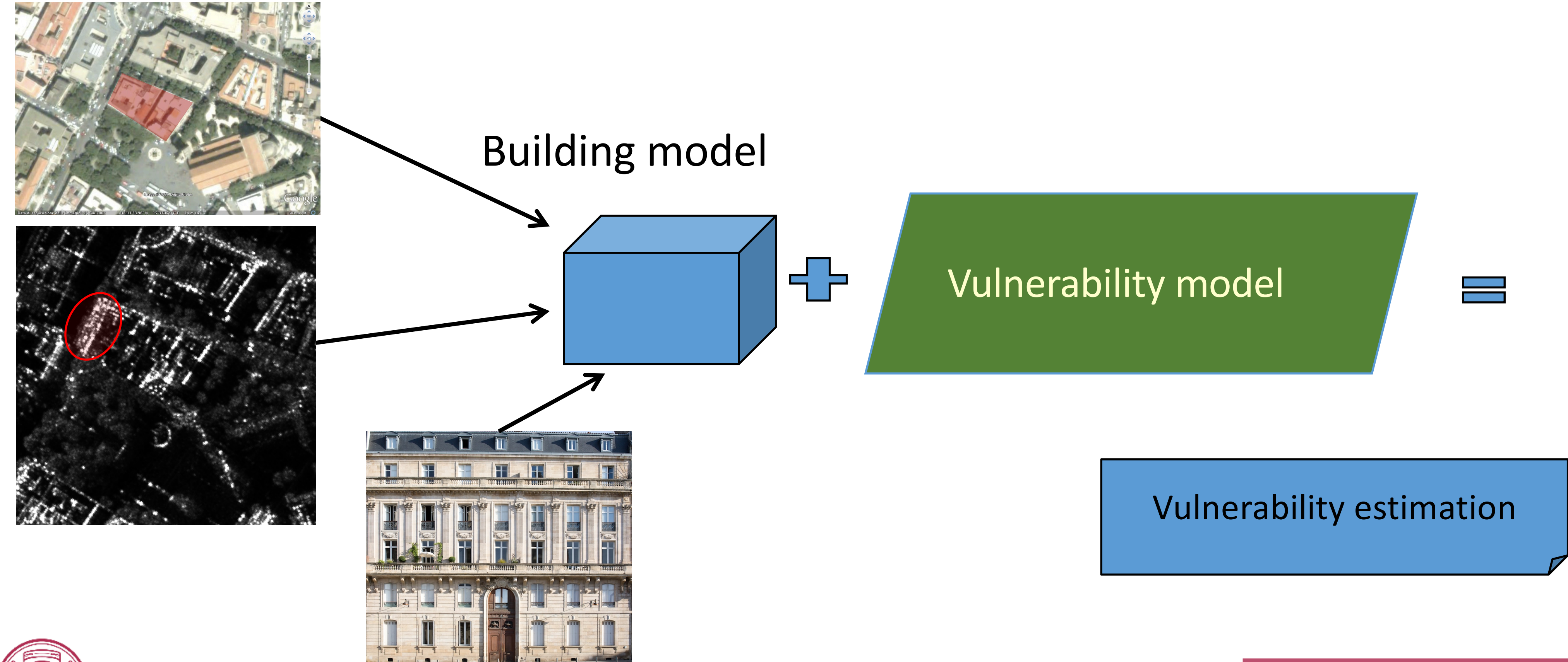
UNIVERSITÀ DI PAVIA

From buildings to risks





“Data fusion of all types” in the loop



Challenges

- Feature selection, dimensionality reduction starting from (very) heterogeneous data sets:
  - computationally intensive if performed by the system
- AI may help
  - but scene intelligence requires large training sets, currently unavailable
- Better using
  - semisupervised (quantum?) machine learning
    - for classification/mapping
  - Physics-based data fusion
    - to reduce (or understand) uncertainties (but solving complex equ.)

EO data mining

- Looking for local affinity patterns, i.e. “common behaviors” that a set of samples share over a finite set of attributes.

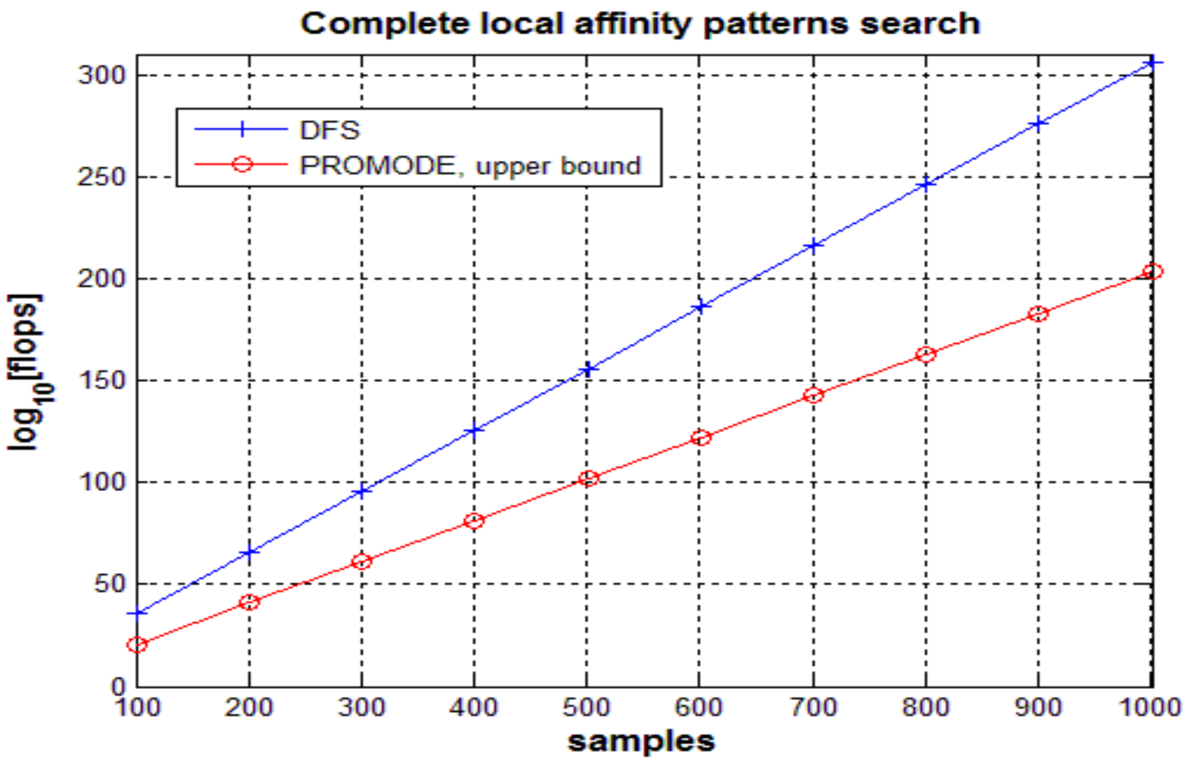
$$\underline{\underline{H}} = \begin{bmatrix} 0 & 1 & 1 & 2 & 0 & 0 & 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & 2 & 0 & 1 & 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 2 & 1 & 1 \\ 0 & 1 & 0 & 2 & 0 & 1 & 1 & 1 & 0 & 2 \\ 2 & 1 & 0 & 1 & 2 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

2nd sample

9th attribute

(Graph) Depth-first search

- Employed in many existing mining frameworks, but not very efficient



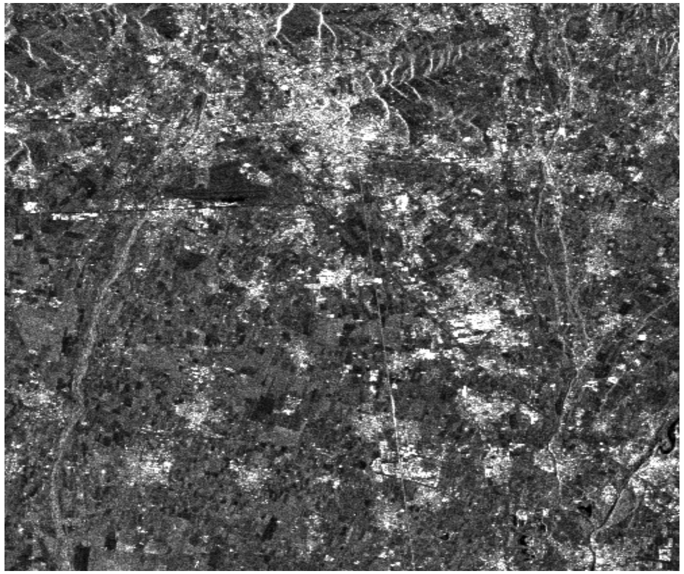
(samples, attributes) = (960, 225000)

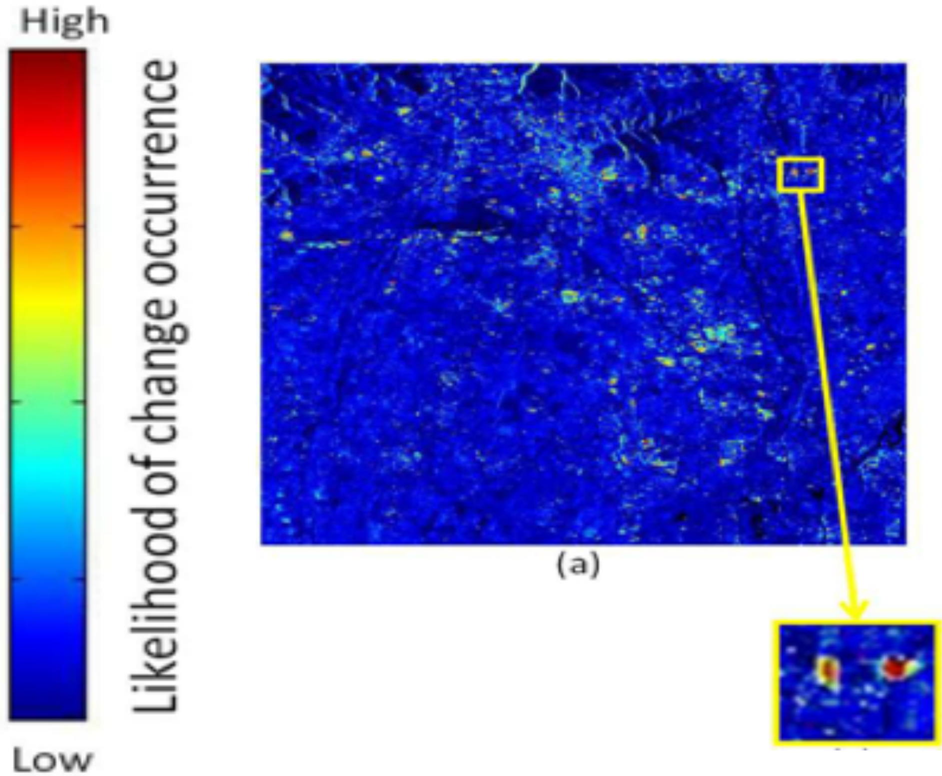
DFS 10^{289}

PROMODE 10^{8}

Temporal data mining using EO

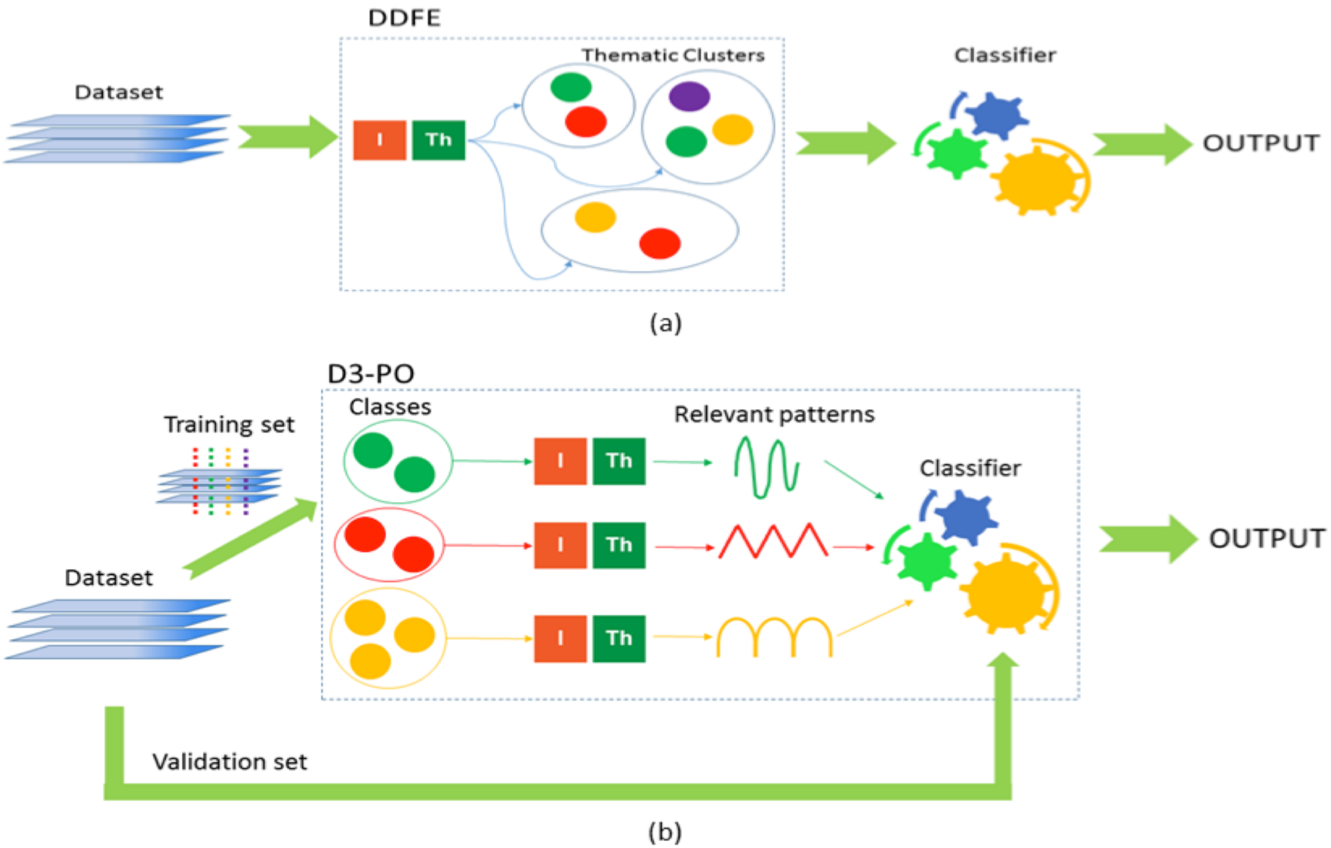
- A sequence of 73 SAR images
- LAP = similar radar properties over time for given space samples

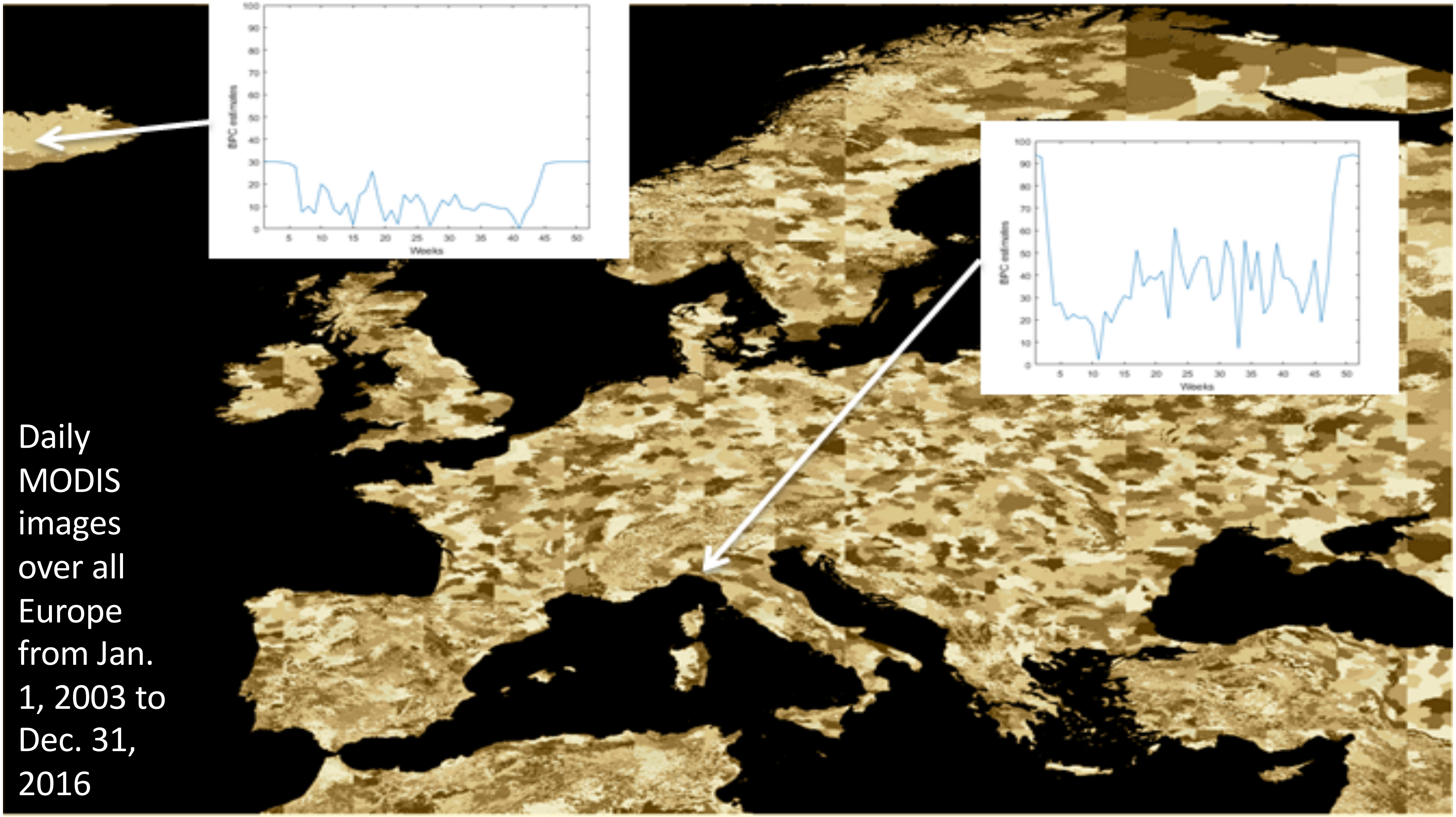




UNIVERSITÀ DI PAVIA

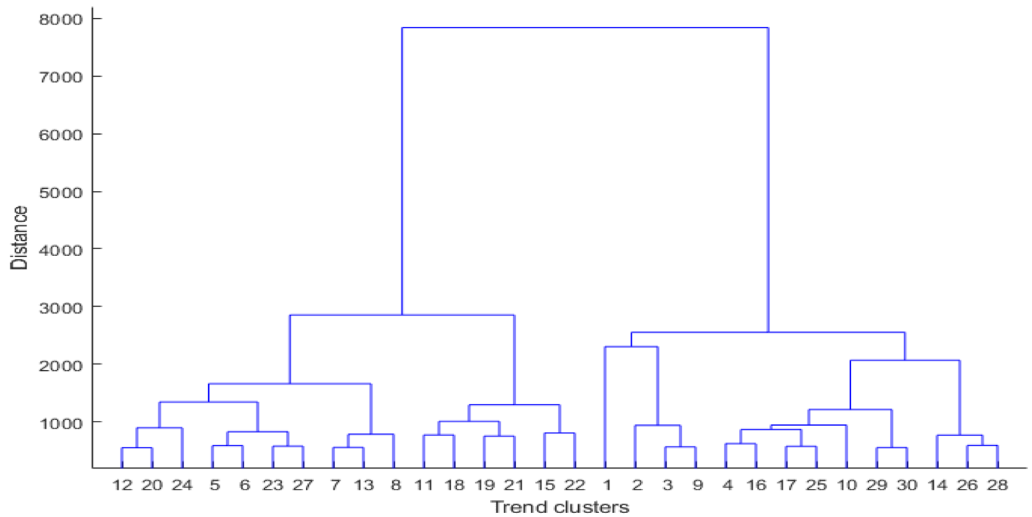
Clustering vs. classification




Weeks
Weeks
Daily MODIS images over all Europe from Jan. 1, 2003 to Dec. 31, 2016

Experimental results

- At the global (European) level, DDFE highlighted **nearly 10,000 different trends of air quality dynamics** that can be discriminated **in 2016**. Dependences and similarity links among significant patterns can be extracted by means of a linkage dendrogram.

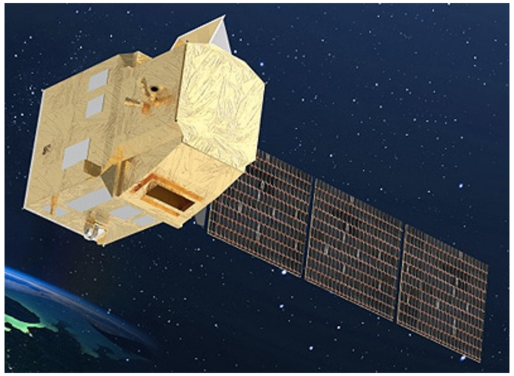


Challenges

- Efficient multi-clustering approaches
  - combining multiple clustering approaches
  - useful in case of known/unknown patterns
  - we still have the “name the pattern” issue
- Data mining using computationally efficient approaches in long temporal sequences
  - missing values in the temporal sequence
  - Effective approaches to select significant patterns

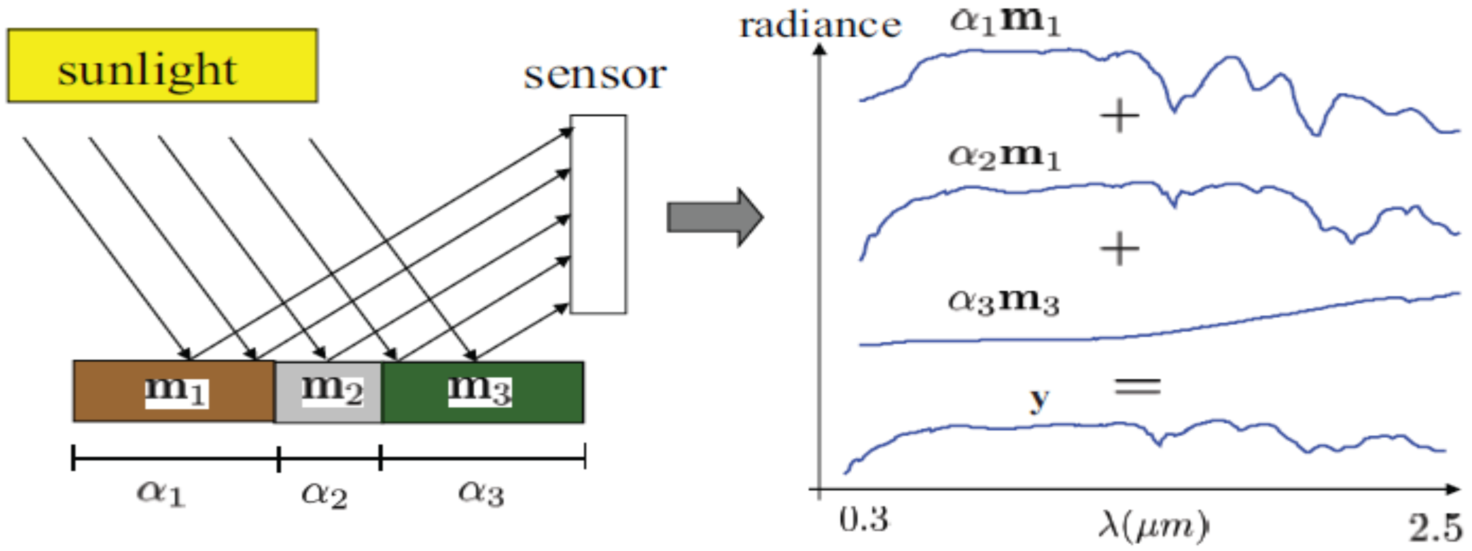
EO data mining: spatial dimension

- More and more EO data sets are currently used to extract data at a finer resolution than what they were meant for.
- Typical examples are hyperspectral sensors, which have many spectral bands (> 100) but relatively coarse spatial resolution.



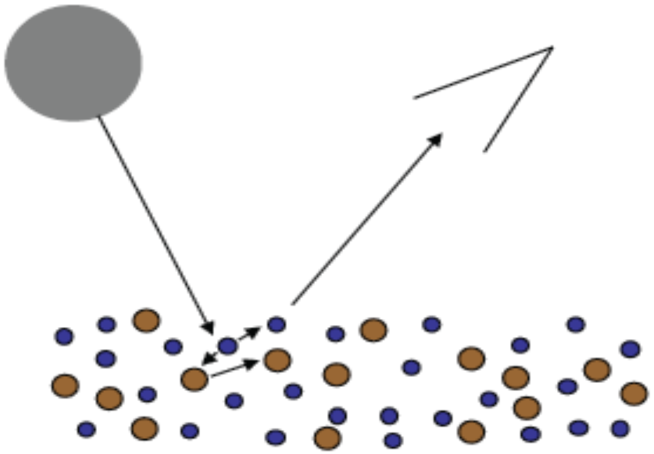
Pixel as a linear mixture

- For each pixel of a given image, the recorded signal is a mixture of light scattered by substances (endmembers) located in the field-of-view



Pixel as a non-linear mixture

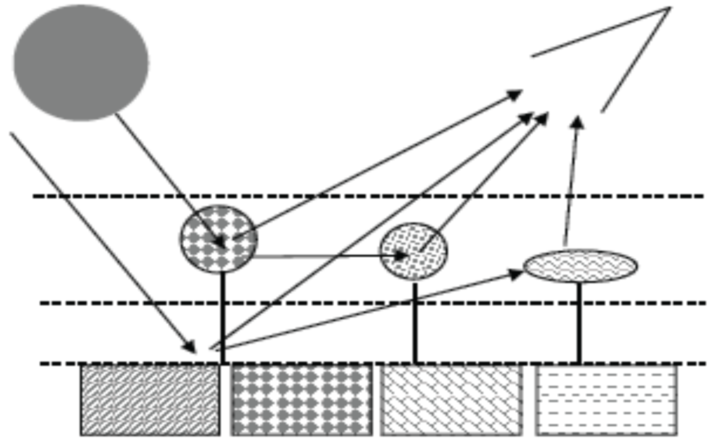
Intimate mixture (particulate media)



$$y = f(\alpha, \theta)$$

Material densities — Media parameters

Two-layers: canopies+ground



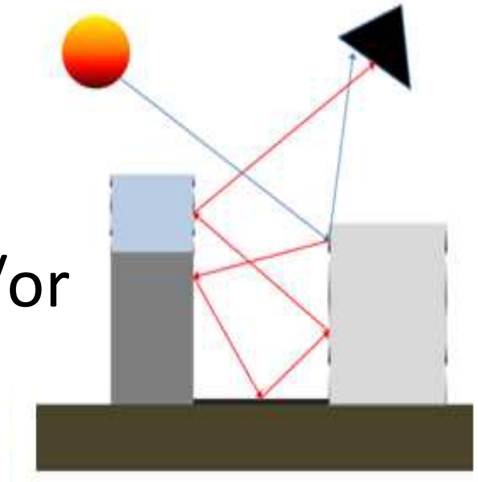
$$y = \underbrace{\sum_{i=1}^{p} \alpha_i \mathbf{m}_i}_{\text{Single scattering}} + \underbrace{\sum_{\substack{i,j=1 \\ i \neq j}}^{p} \alpha_{i,j} \mathbf{m}_i \odot \mathbf{m}_j}_{\text{Double scattering}}$$

UNIVERSITÀ DI PAVIA

Courtesy: J.M. Bioucas-Dias

Non-linear models

- Bilinear mixture models
  - easy to implement;
  - can describe very well 2-elements scenes;
  - might not efficiently track the endmember composition in case of multiple scattering and/or higher order non-linearities.



- Intimate mixture models
  - very accurate
  - quite cumbersome to invert and achieve abundances.

Polynomial models

- Properly extending linear models to polynomial (with non-linearity order $p \geq 2$) may help to have a more flexible solution.

Linear

$$\underline{y}_l = \sum_{r=1}^{R} a_{rl}\underline{m}_r + \sum_{k=2}^{p}\sum_{r=1}^{R} \beta_{rkl}\underline{m}_r^k$$

Order-p scattering and interferences

Challenges

- Selection of the *p*-linear model
  - per-pixel vs. per-object vs. per-scene
- Inversion of the *p*-linear model
  - parallel processing vs. local/multiscale refinements
  - transformation into a liner optimization model via auxiliary variable?
- Inversion of the "intimate mixture model" considering spatial constraints

UNIVERSITÀ DI PAVIA

References

- A. Marinoni, A. Plaza, J. Plaza, P. Gamba, "Nonlinear Hyperspectral Unmixing using Nonlinearity Order Estimation and Polytope Decomposition", IEEE J. of Selected Topics in Applied Earth Observation and Remote Sensing, doi: 10.1109/JSTARS.2015.2427517, vol. 8, no. 6, pp. 2644-2654, 2015.
- A. Marinoni, and P. Gamba, "A Novel Approach for Efficient p-linear Hyperspectral Unmixing", IEEE J. of Selected Topics in Signal Processing, doi: 10.1109/JSTSP.2015.2416693, vol. 9, no. 6, pp. 1156-1168, 2015.
- A. Marinoni, and P. Gamba, "A novel approach for efficient detection of affinity patterns in remotely sensed Big Data", IEEE J. of Selected Topics in Applied Earth Observation and Remote Sensing, doi: 10.1109/JSTARS.2015.2485401, vol. 8, n. 10, pp. 4622 – 4633, 2015
- A. Marinoni, and P. Gamba, "Efficient detection of anomaly patterns through global search in remotely sensed Big Data", Journal of Applied Remote Sensing, doi: 10.1117/1.JRS.10.045012, vol. 10, no. 4, 045012, 2016.
- A. Marinoni, P. Gamba, "Unsupervised data driven feature extraction by means of mutual information maximization", IEEE Trans. on Computational Imaging, doi: 10.1109/TCI.2017.2669731, vol. 3, no. 2, pp. 243-253, 2017.
- A. Marinoni, G.C. Iannelli, P. Gamba, "An information theory-based scheme for efficient classification of remote sensing data", IEEE Trans. on Geoscience and Remote Sensing, doi: 10.1109/TGRS.2017.2716187, vol. 55, no. 10, pp. 5864-5876, Oct. 2017.
- A. Marinoni, A. Dagliati, R. Bellazzi, P. Gamba, "Inferring air quality maps from remotely sensed data to exploit georeferenced clinical onsets: the Pavia 2013 case", Proc. of IGARSS'15, Milan, July 2015, pp. 3937-3940, doi: 10.1109/IGARSS.2015.7326686.
- Y. Zhao, B. Huang, A. Marinoni and P. Gamba, "High spatiotemporal resolution PM2.5 concentration estimation with satellite and ground observations: A case study in New York City," 2018 IEEE International Conference on Environmental Engineering (EE), Milan, 2018, pp. 1-5, doi: 10.1109/EE1.2018.8385255
- A. Marinoni, D. De Vecchi, D. Tuia, and P. Gamba, "Discovering temporal patterns of air quality in different parts of Europe with data driven feature extraction", Proc. of IGARSS'18, Valencia (Spain), July 22-27, 2018, pp. 2066-2069, doi: 10.1109/IGARSS.2018.8519052.
- G.C. Iannelli, P. Gamba, "Urban extent extraction combining Sentinel data in the optical and microwave range", IEEE J. of Selected Topics in Applied Earth Observation and Remote Sensing, doi: 10.1109/JSTARS.2019.2920678, vol. 12, no. 7, pp. 2209-2216, July 2019.
- Z. Miao, G.C. Iannelli, P, Gamba, "Using social media data to map urban areas: ideas and limits", Proc. of IGARSS 2019, Yokohama, Japan, July 2019, pp. 5800-5803, doi: 10.1109/IGARSS.2019.8898361.
- P. Gamba, "Image and data fusion in remote sensing of urban areas: status issues and research trends", International Journal of Image and Data Fusion, doi: 10.1080/19479832.2013.848477, vol. 5, no.1, pp. 2–12, 2014.
- M. Dalla Mura, F. Pacifici, S. Prasad, P. Gamba, J. Chanussot, "Challenges and Opportunities of Multimodality and Data Fusion in Remote Sensing", Proc. of IEEE, doi: 10.1109/JPROC.2015.2462751, vol. 103, no. 9, pp. 1585-1601, 2015