European Commission

SYMBOLIC MACHINE LEARNING: extracting information on human settlements

Dr Christina Corbane
Contact: Dr Martino Pesaresi
Martino.pesaresi@jrc.ec.europa.eu

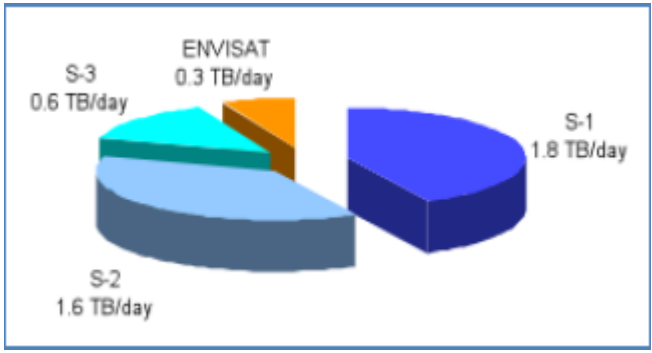
Joint Research Centre
the European Commission's
in-house science service

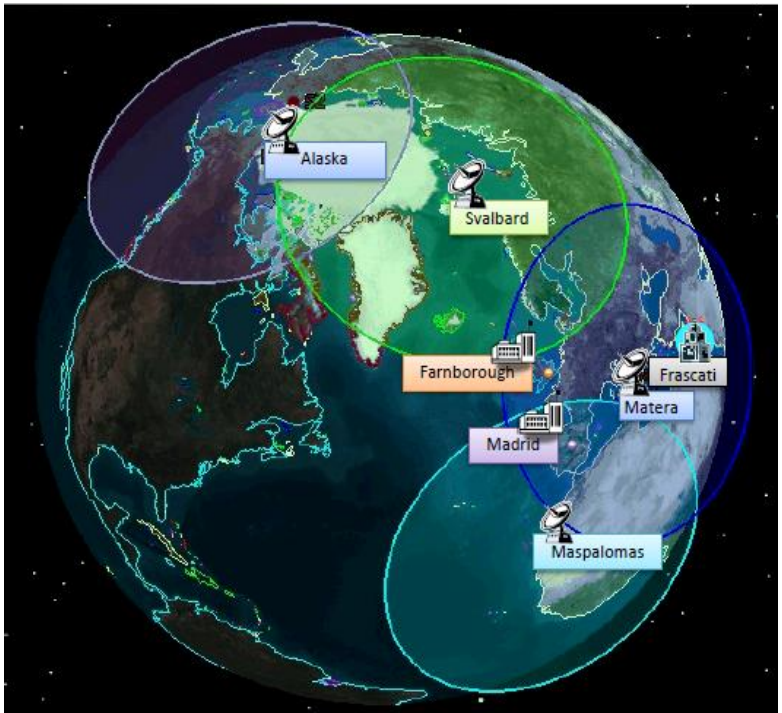
Joint Research Centre

The Earth Observation & Big Data landscape

- New era of data abundance, exponentially growing in volume, but only partially structured and harmonized.
- Copernicus Sentinel missions and the NASA Landsat mission provide daily TB of data, setting new standards in large-scale data management (storage, retrieval, maintenance, delivery, communication).
- The political decision for free and open access to these data constitutes a landmark in the history of Earth Observation and data exploration;

Sentinel 1,2
Data Volume & Streaming





Continuous raw data supply rate of circa 4Tbyte/day

Petabyte scale mission lifetime archive

European Commission

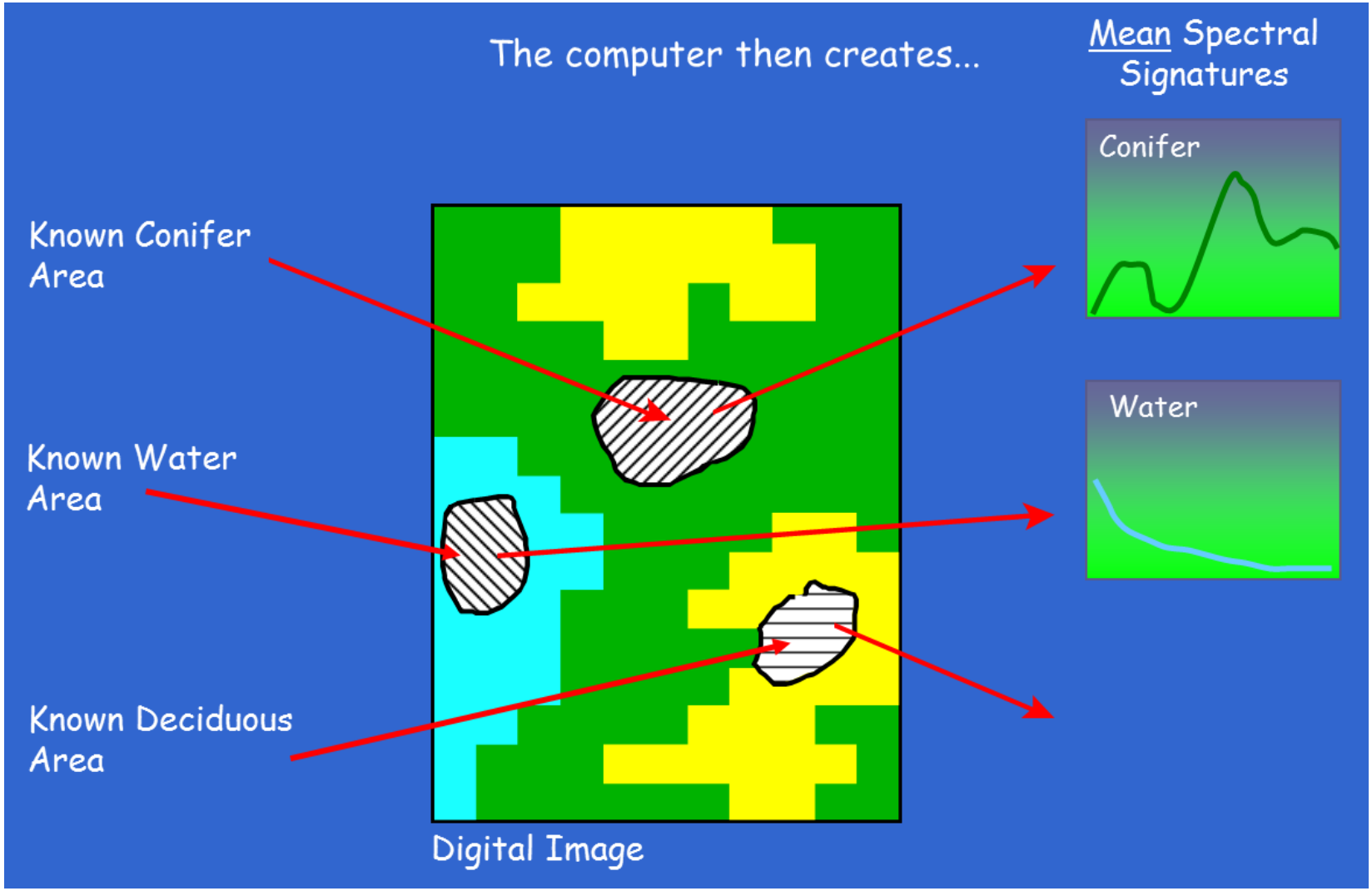
The problem

Standard paradigm for extracting information from Earth Observation data relies on physical explicit modeling of the relationships between target's energy absorption-reflection-emitting properties and sensor technical characteristics.

Difficult to apply in Big data landscape due to:

- High requirements of input data in terms of quality and standardization (stability, calibration);
- Cost for the collection of necessary ancillary data;
- Cost of porting the model in different sensors

Traditional image classification



Our proposal: Symbolic Machine Learning

Data-driven exploratory approach, where the machine learns automatically statistical relationships among features/variables based on similarities

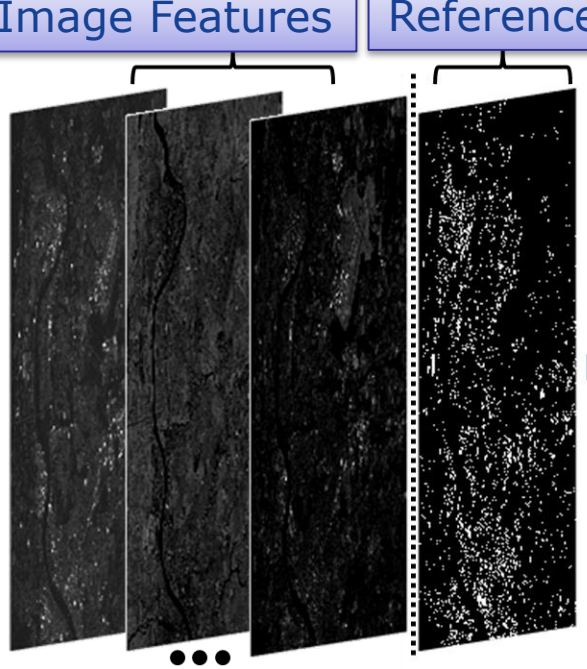
Includes: statistical learning, machine learning, data mining

Supervisory signals: positive and negative examples

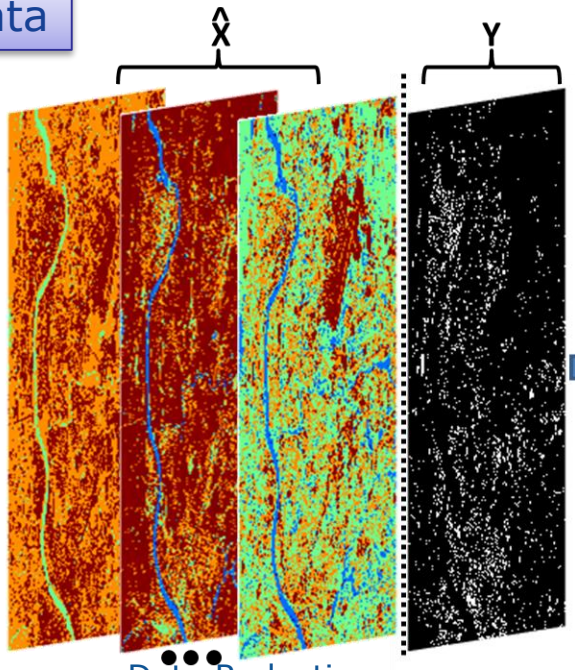
- Most often derived from human-assisted annotations;
- Conditionally by low-resolution geographical thematic maps.

Symbolic Machine Learning (workflow)

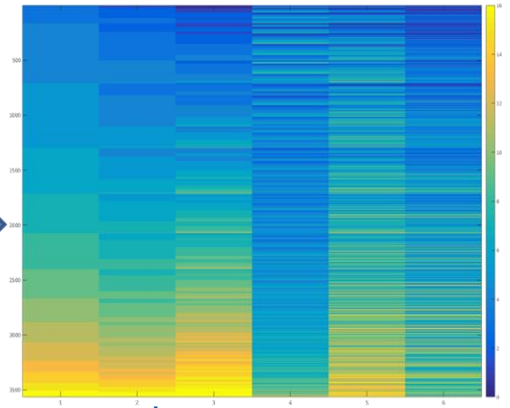
Image Features | Reference data



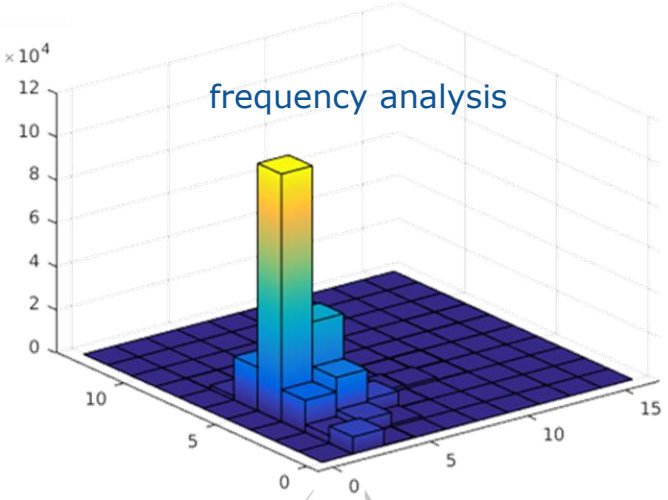
the input dataset



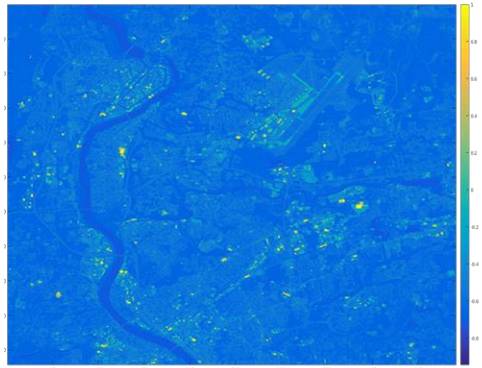
Data Reduction
(taxonomy scheme)



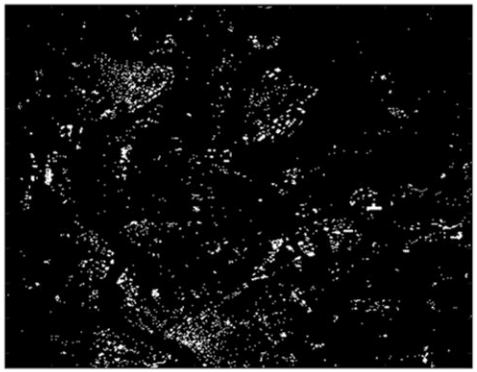
unique sequences



frequency analysis



Association Analysis
(Measure of Confidence)

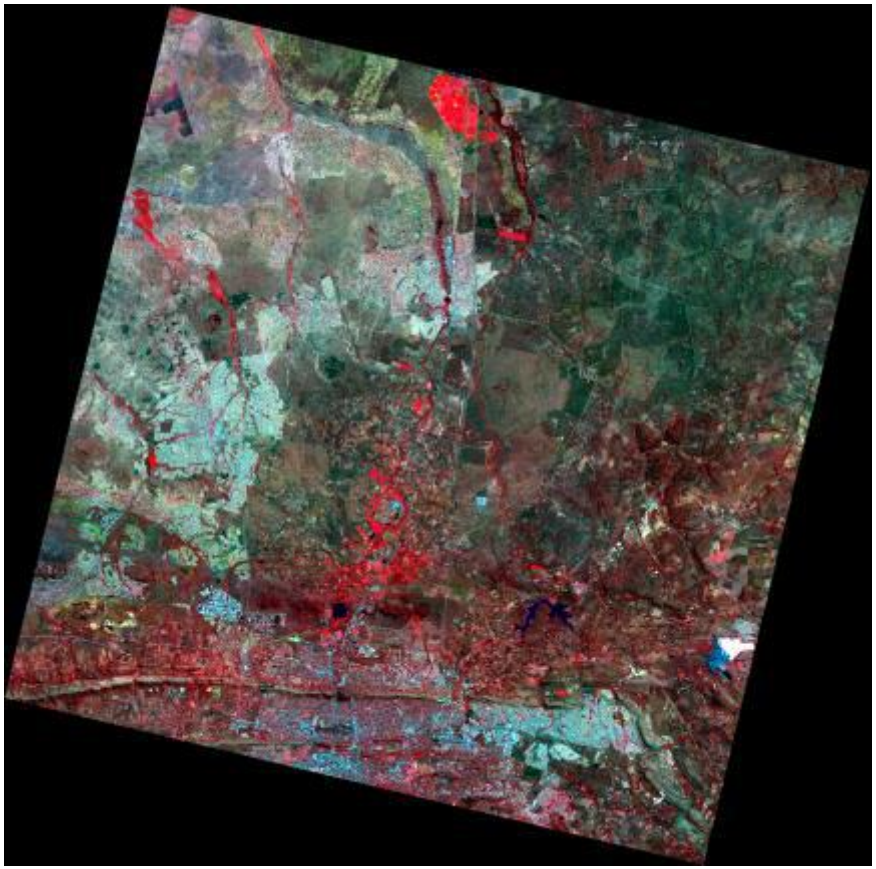


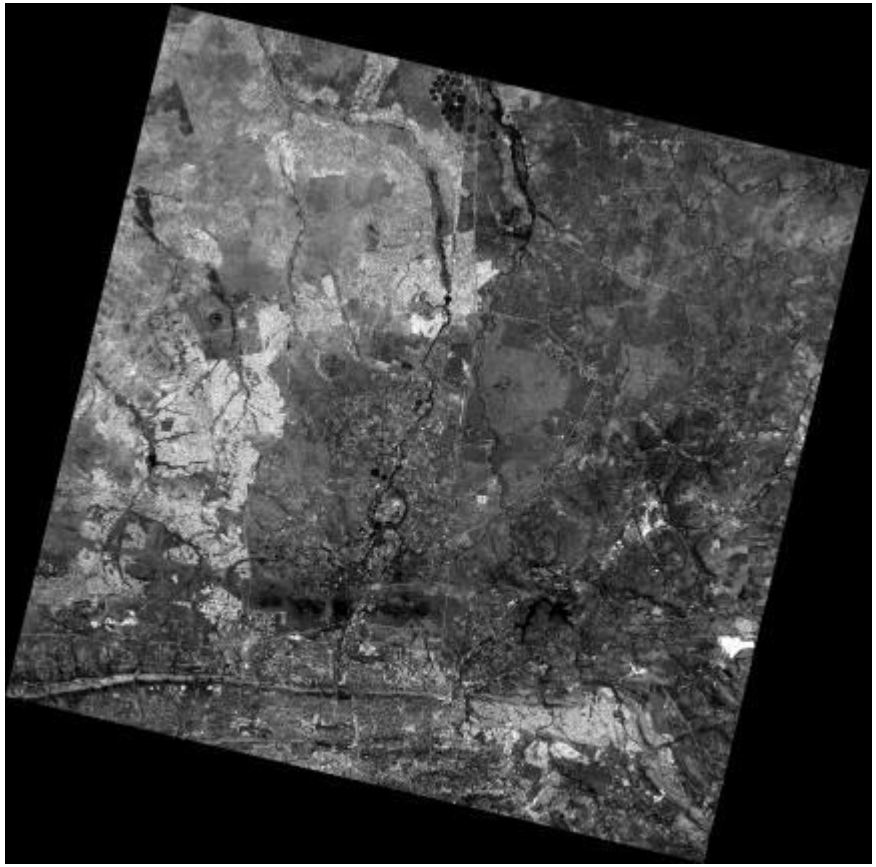
classification

European Commission

Symbolic Machine Learning (workflow)

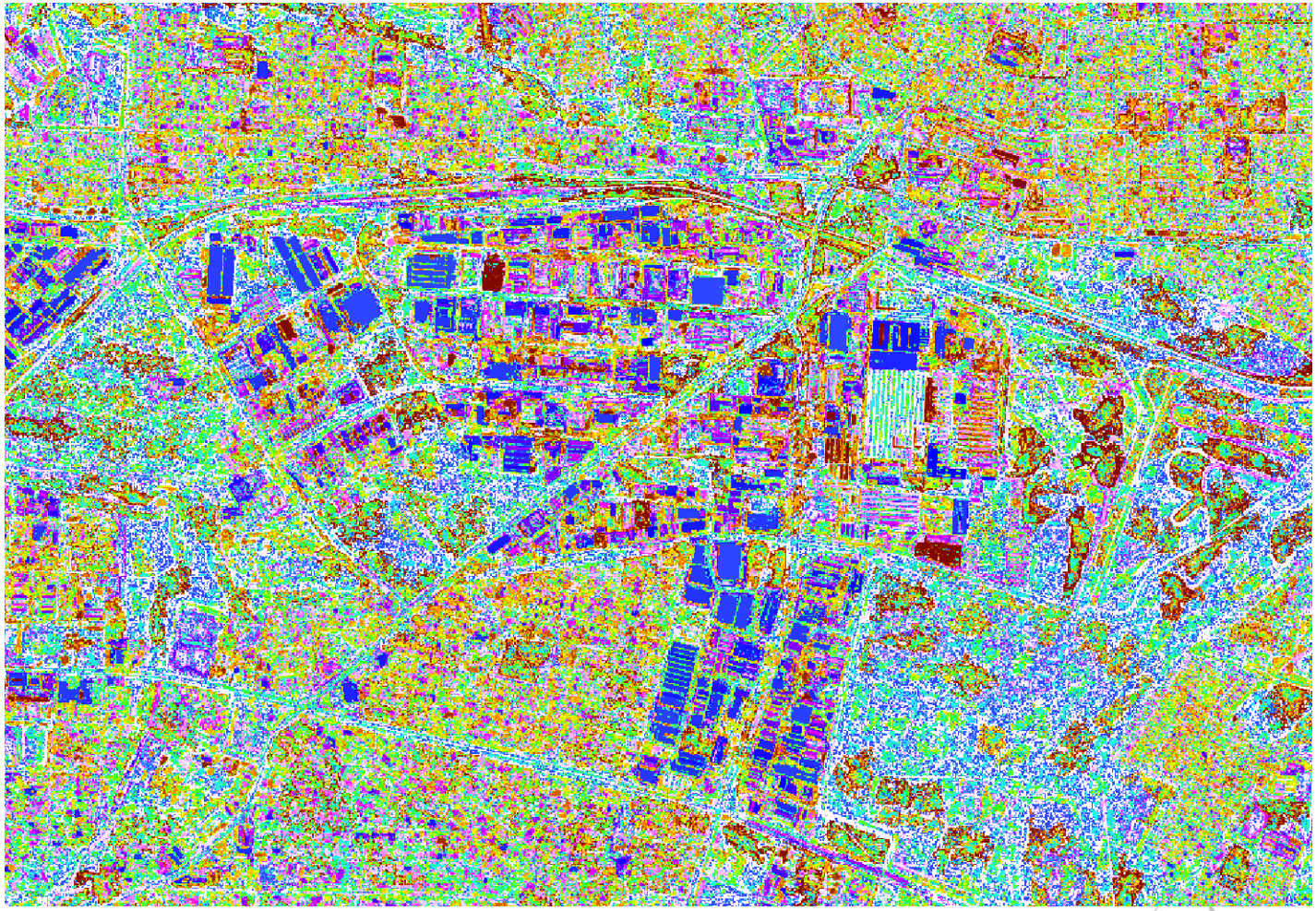
INPUT DATASETS : Remote Sensing image bands (spectral features)





Symbolic Machine Learning (workflow)

INPUT DATASETS : Remote Sensing image bands (spectral features)

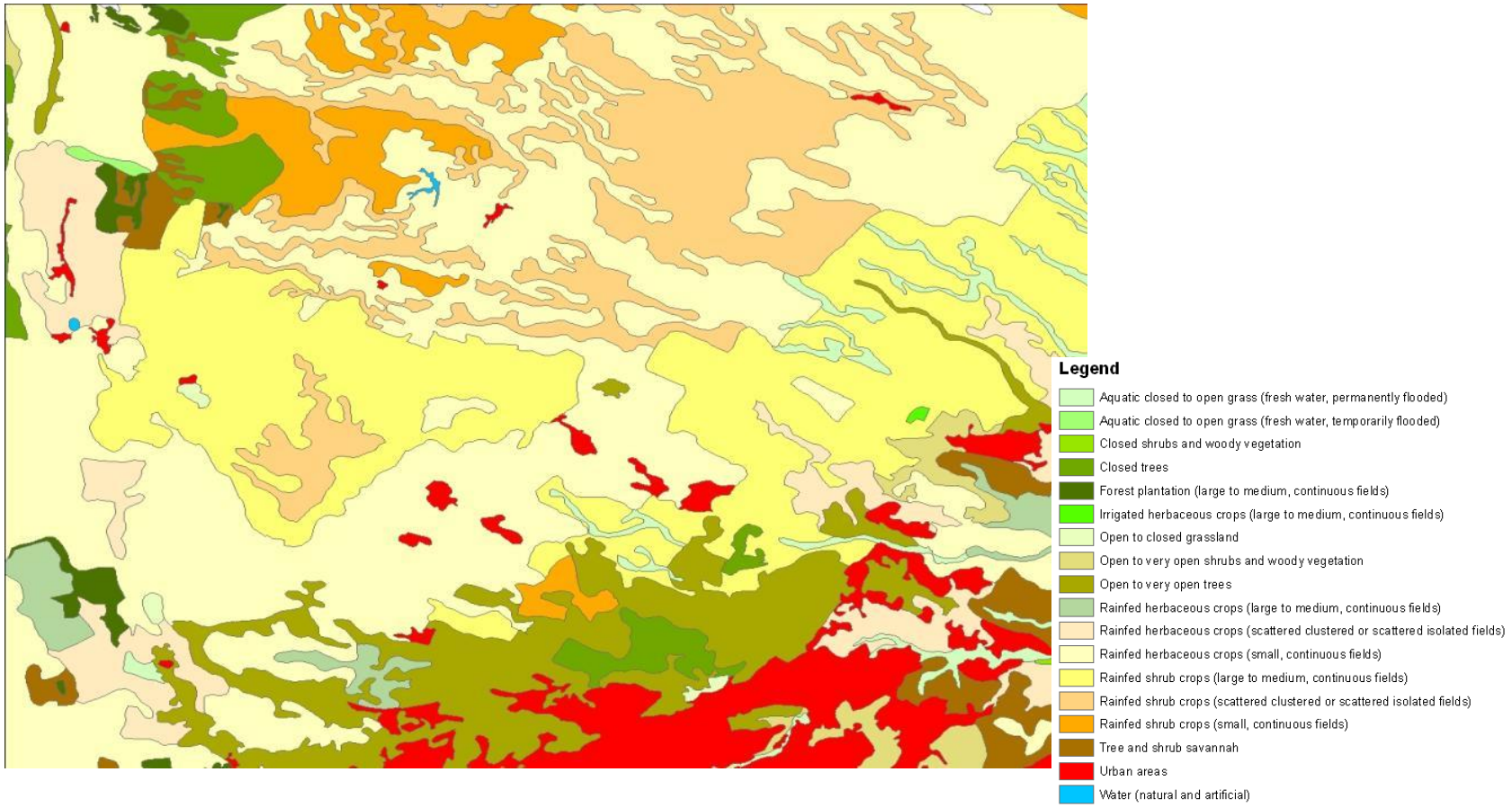


Symbolic Machine Learning (workflow)

INPUT DATASETS : Textural features

Symbolic Machine Learning (workflow)

INPUT DATASETS : Reference data (e.g. Landcover)



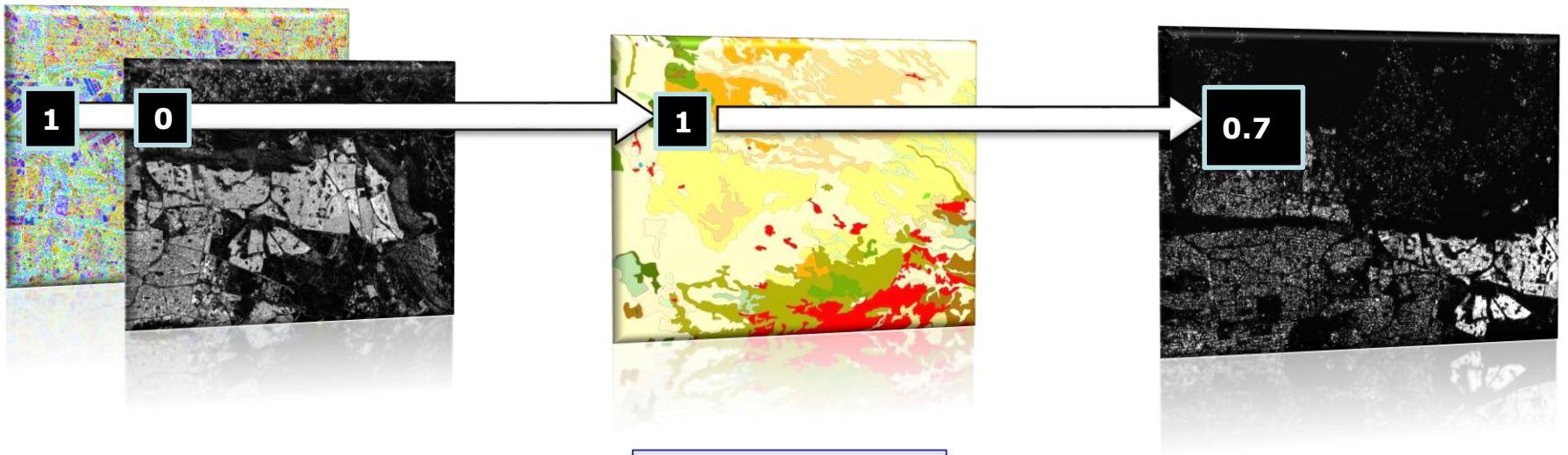
Symbolic Machine Learning (workflow)

OUTPUT: CONFIDENCE & CLASSIFICATION

Image Features

Reference data

Confidence



Classification

	Built-up
	Non built-up

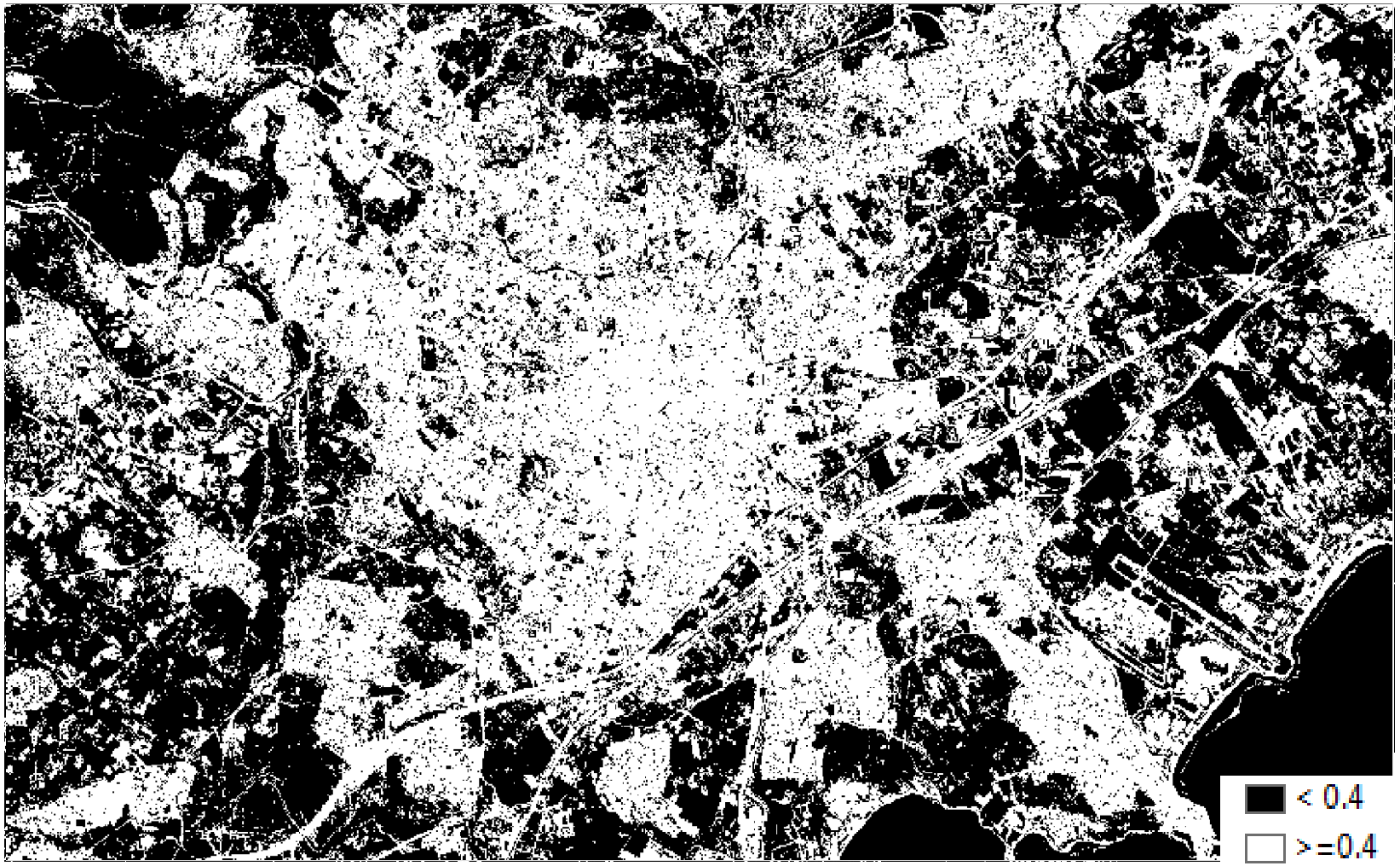
European Commission

Input data in "false color" composite at a 2.5-m spatial resolution

Classes of the reference set extracted from the Land Cover at a 100-m resolution

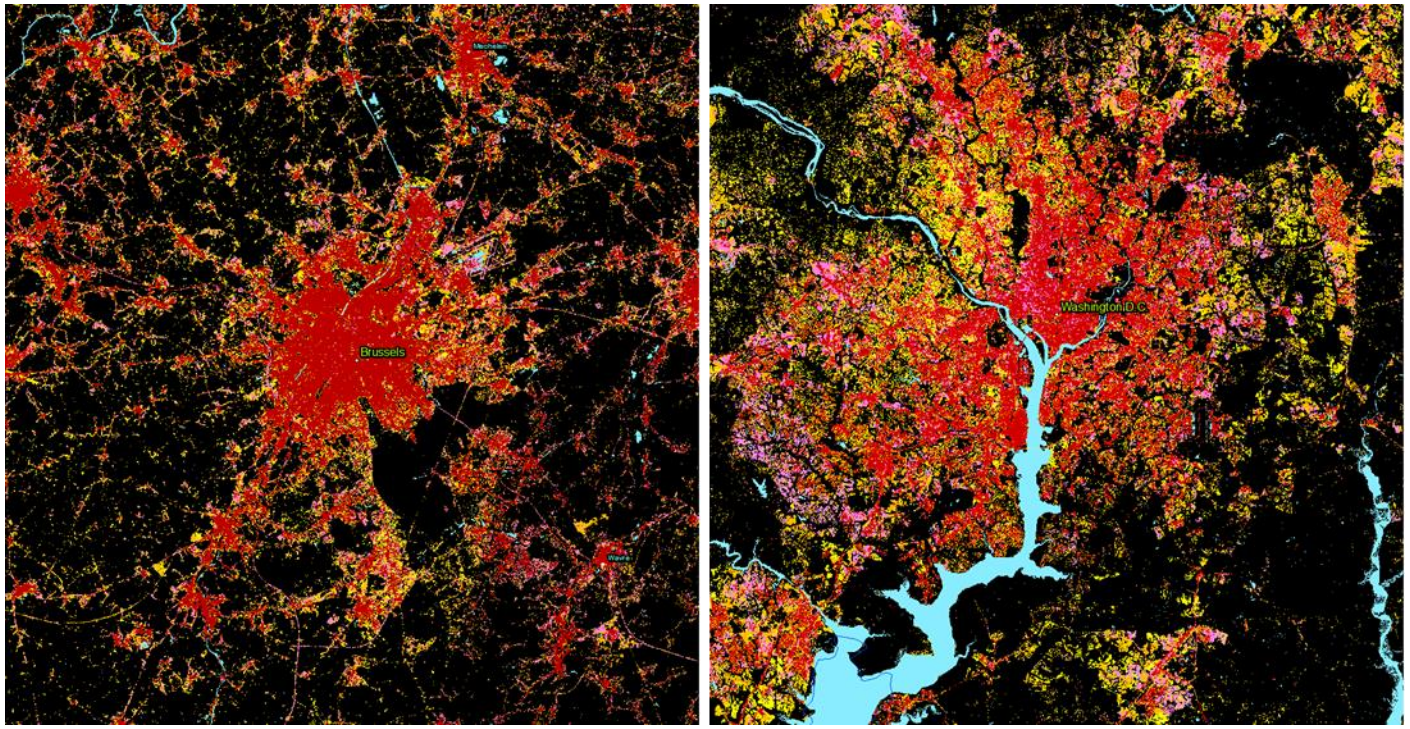
Result of SML classification at a 2.5-m spatial resolution.

OUTPUT: CONFIDENCE & CLASSIFICATION

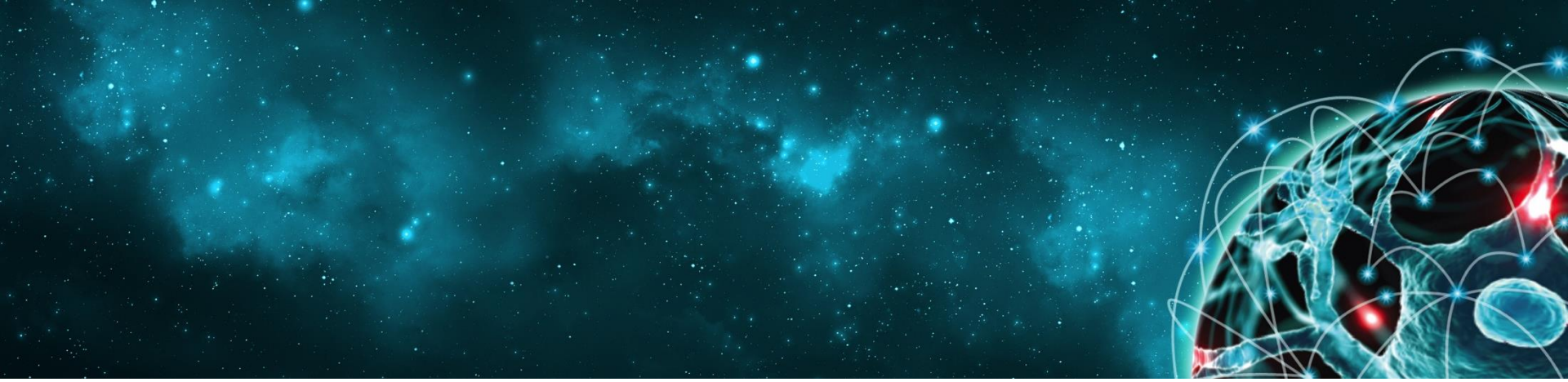


Built-up = CONFIDENCE LAYER > 0.4

Application: multi-temporal medium resolution Global Human Settlement Layer



Reference: M. Pesaresi, D. Ehrlich, S. Ferri, A. J. Florczyk, S. Freire, M. Halkia, A. Julea, T. Kemper, P. Soille, and V. Syrris, "Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014," JRC Technical Report EUR 27741 EN, EC, JRC, IPSC, 2016, http://publications.jrc.ec.europa.eu/repository/handle/JRC97705.



Stay in touch

JRC Science Hub:
ec.europa.eu/jrc

Twitter: ***@EU_ScienceHub***

Facebook: ***EUsciencehub***

LinkedIn: ***european-commission-joint-research-centre***

YouTube:
JRC Audiovisuals

Vimeo:
Science@EC