

SYMBOLIC MACHINE LEARNING: extracting information on human settlements

Dr Christina Corbane

Christina.corban@ec.europa.eu



Joint Research Centre

the European Commission's
in-house science service



ec.europa.eu/jrc



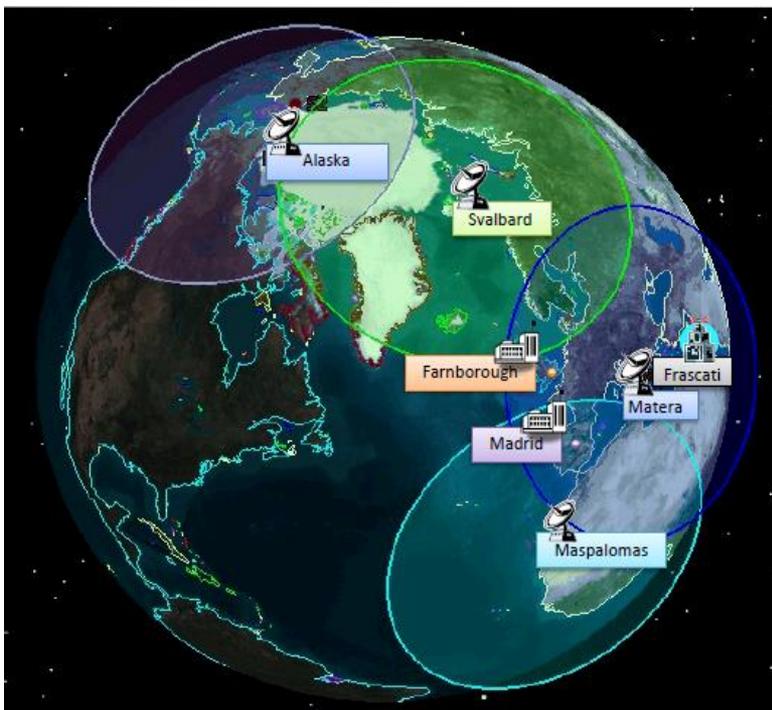
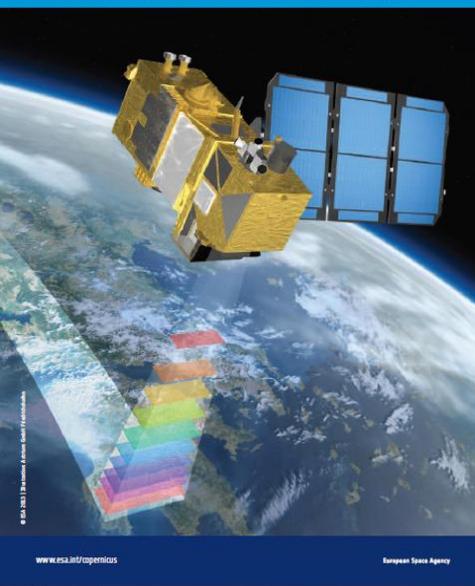
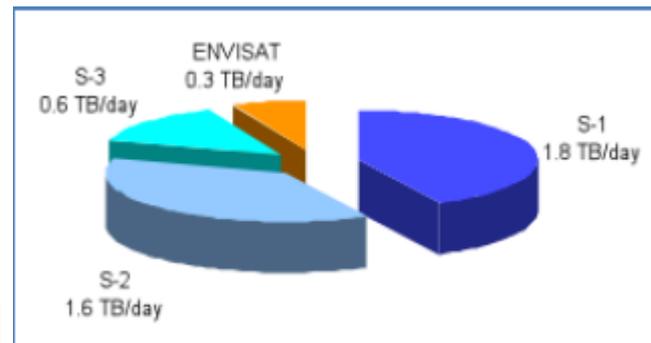
The Earth Observation & Big Data landscape

- 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

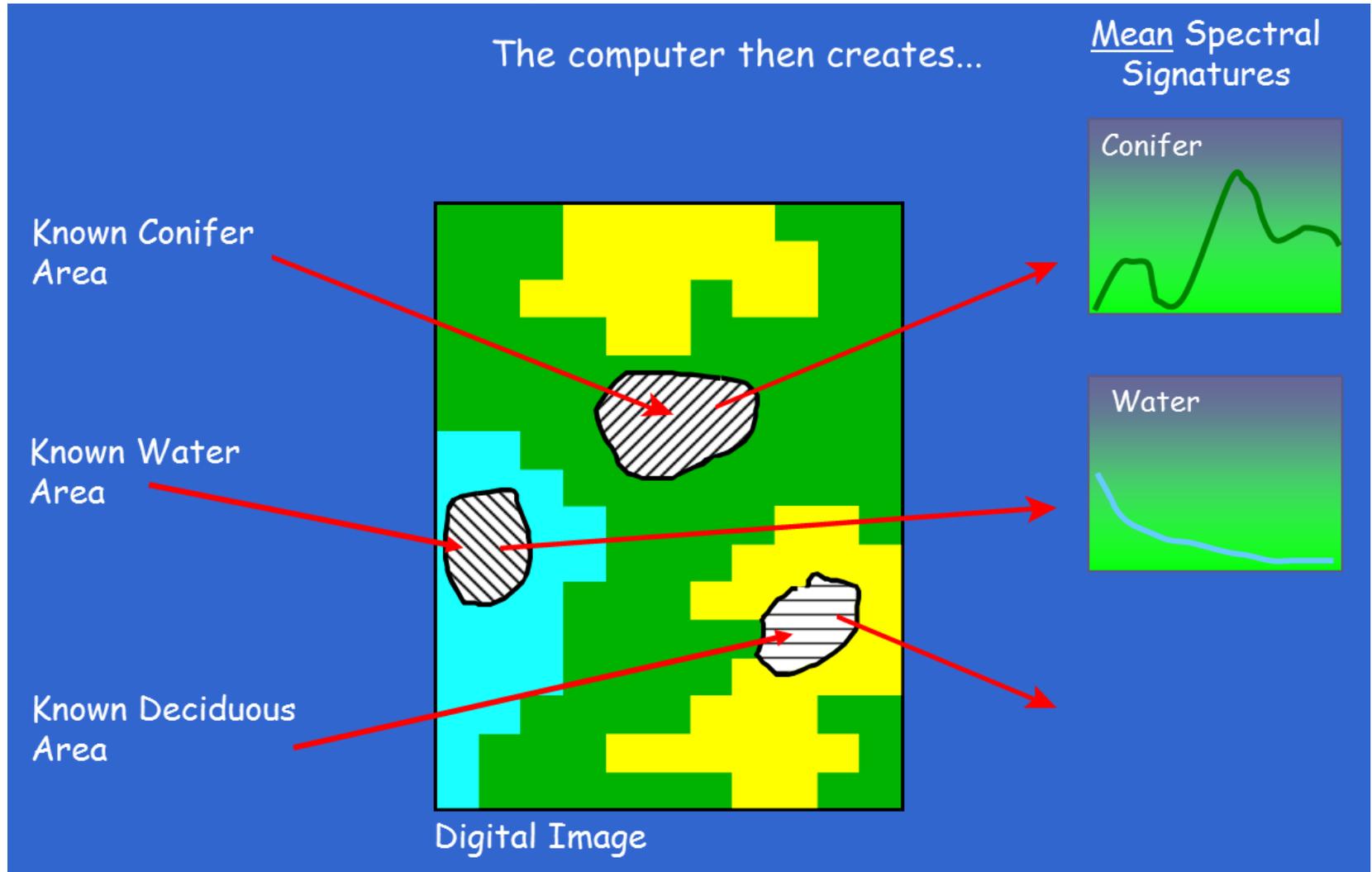
The problem

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.



European
Commission

SYMBOLIC MACHINE LEARNING



ec.europa.eu/jrc



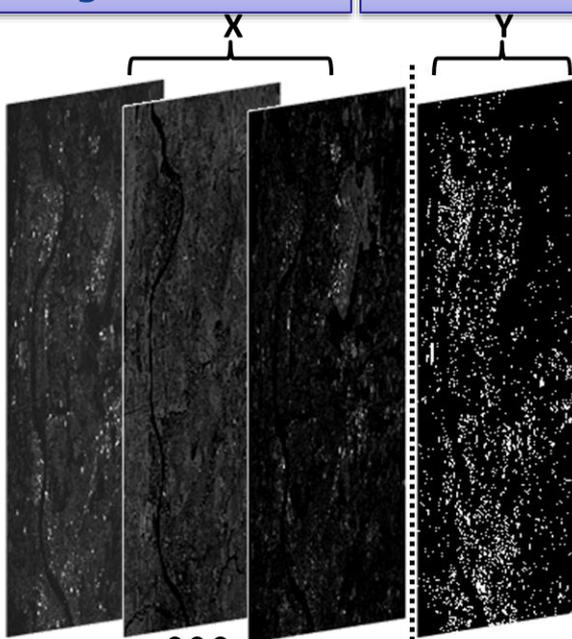
Joint Research Centre
the European Commission's
in-house science service

Joint
Research
Centre

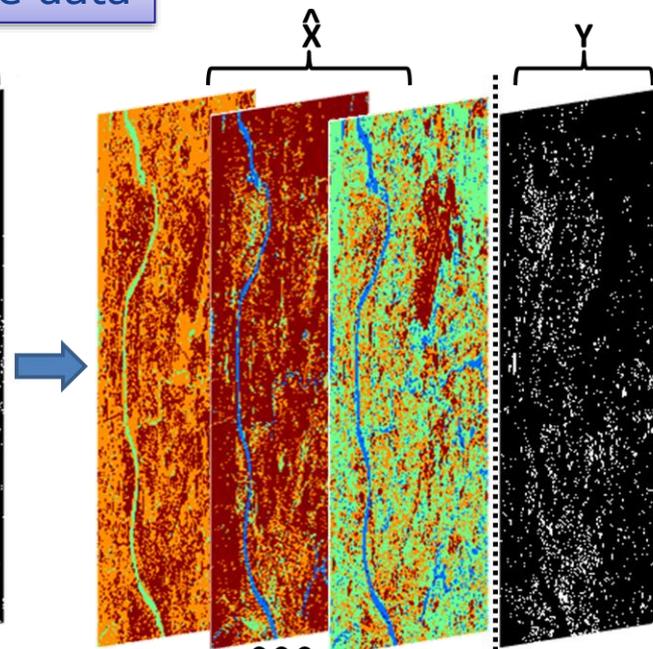
Symbolic Machine Learning (workflow)

Image Features

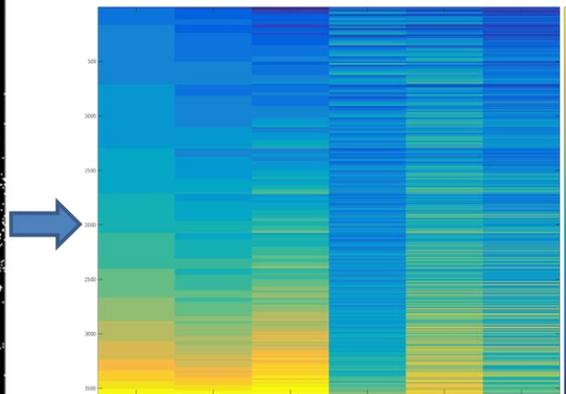
Reference data



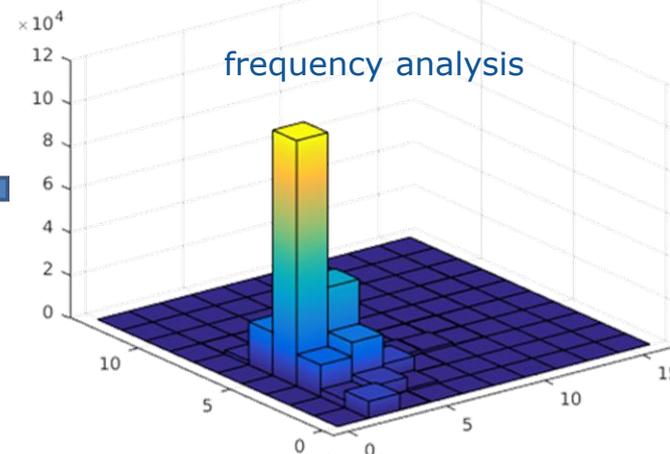
the input dataset



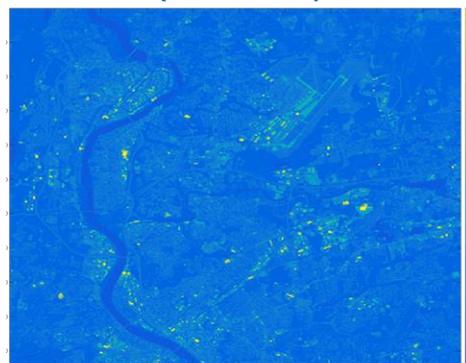
Data Reduction
(taxonomy scheme)



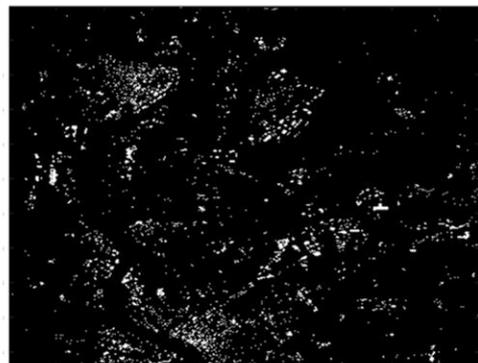
unique sequences



frequency analysis



Association Analysis
(Measure of Confidence)



classification

Symbolic Machine Learning (workflow)

Image Features

Spectral bands and indices

Textural features

Morphological features

the input dataset

Data Reduction
(taxonomy scheme)

unique sequences

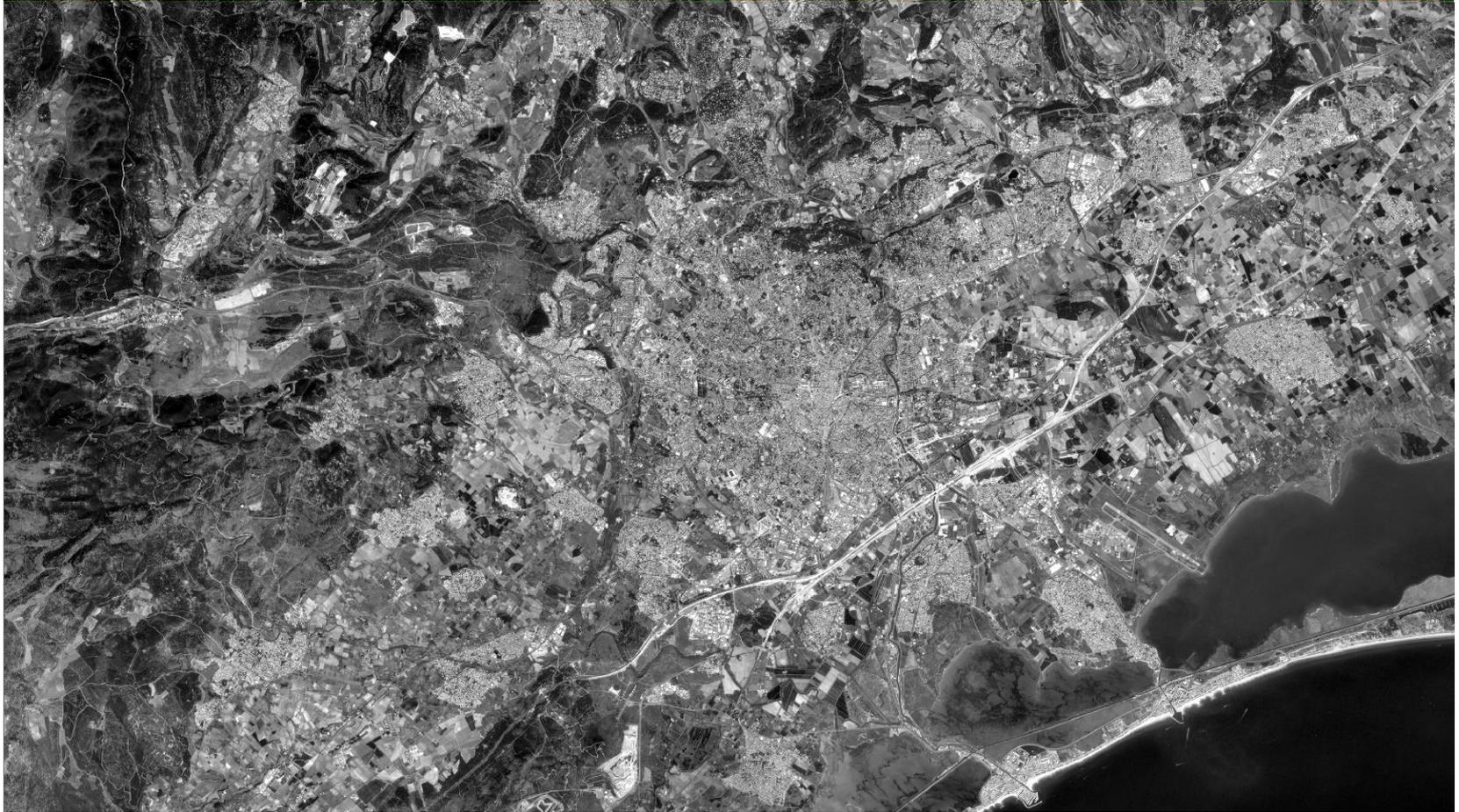
frequency analysis

Association Analysis
(Measure of Confidence)

classification

INPUT FEATURES :

Remote Sensing image bands (spectral features)

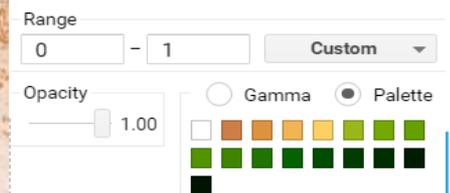


INPUT FEATURES :

Remote Sensing image bands (Indices, multitemporal features)

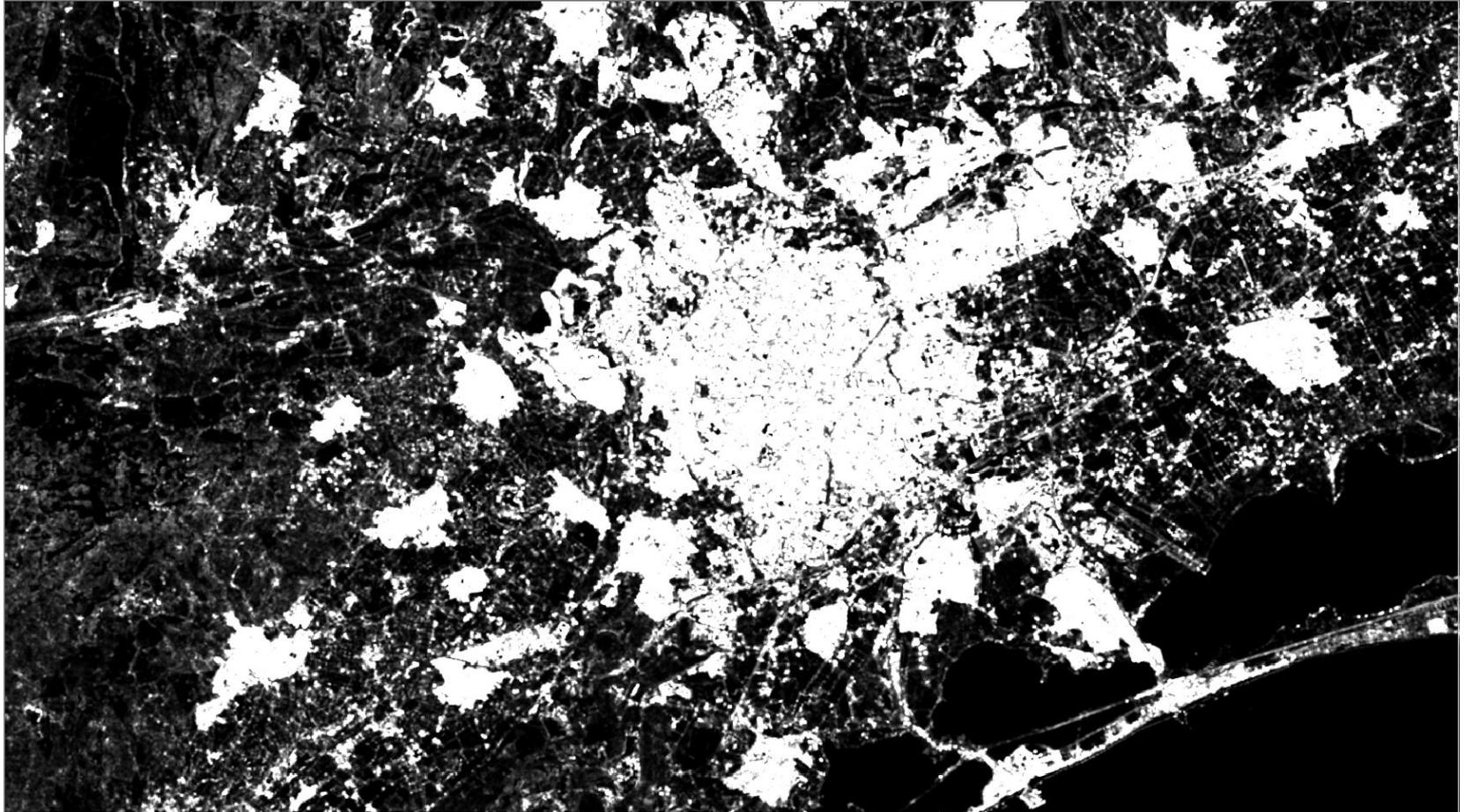
Sentinel-2

(Max NDVI calculated over Montpellier areas: 01 Dec 2015 –20 July 2016)



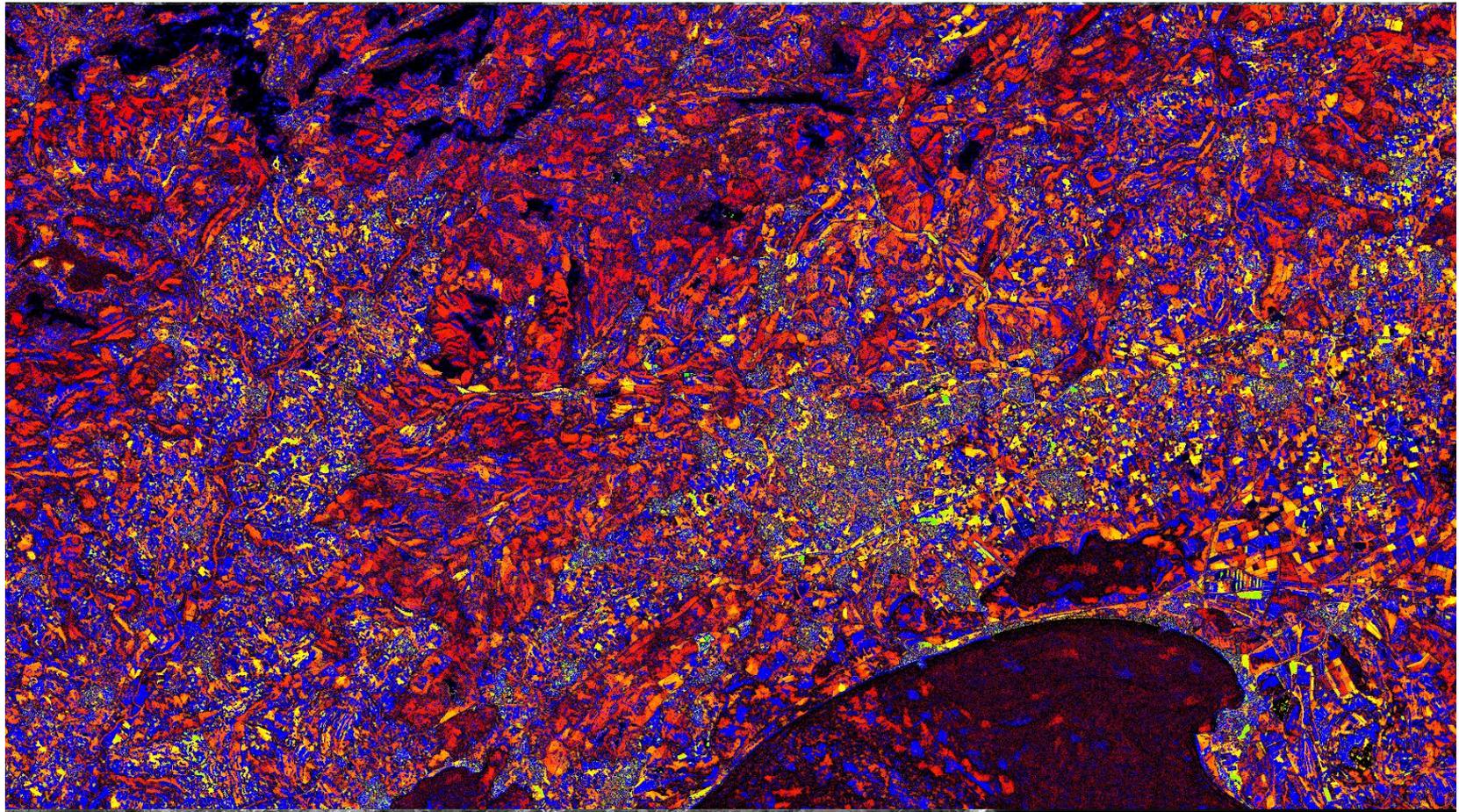
INPUT FEATURES :

Textural Features (e.g. derived from *PANTEX*)

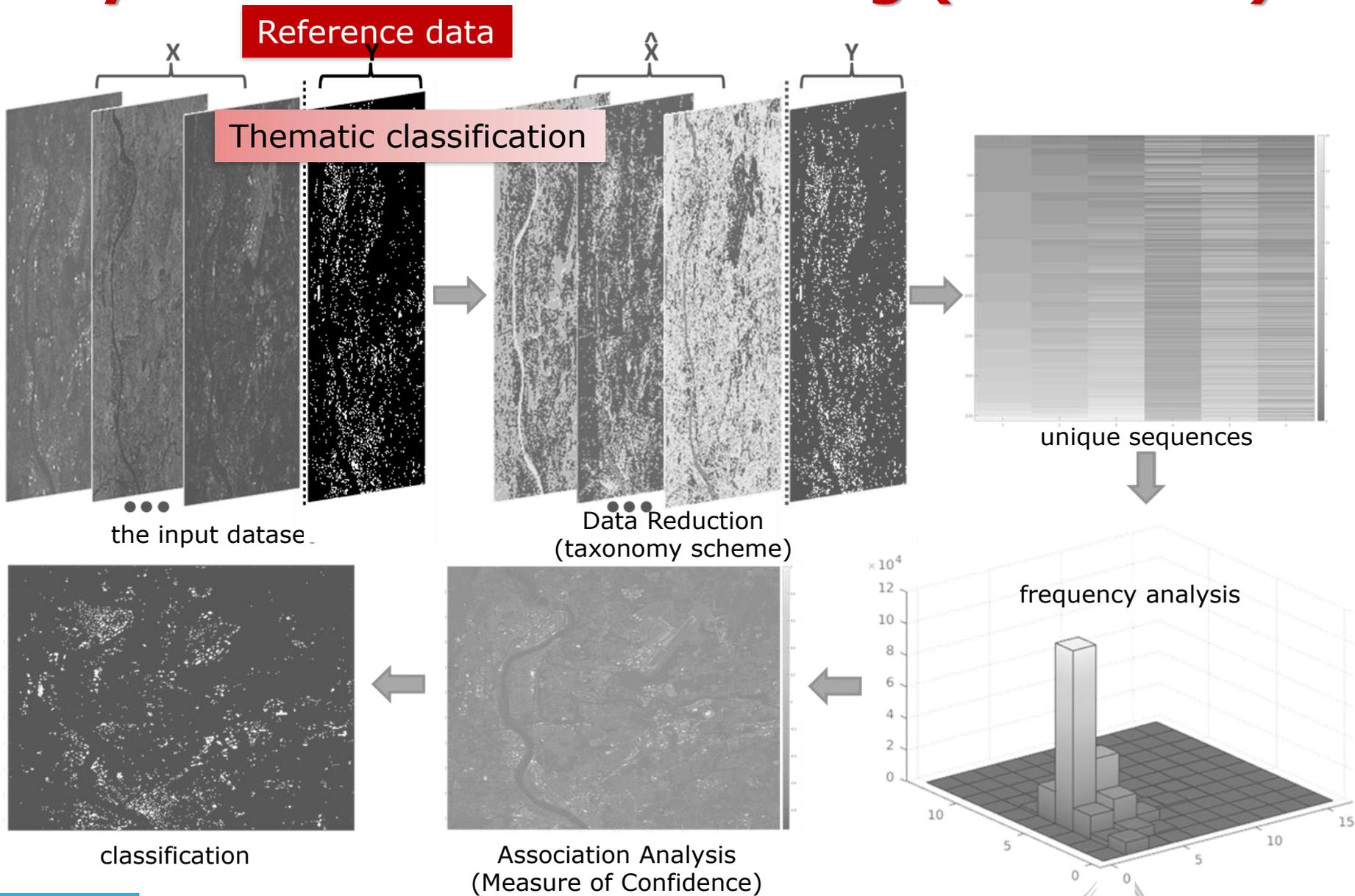


INPUT FEATURES :

Morphological Features (e.g. derived from *characteristic-saliency-level (CSL)*)

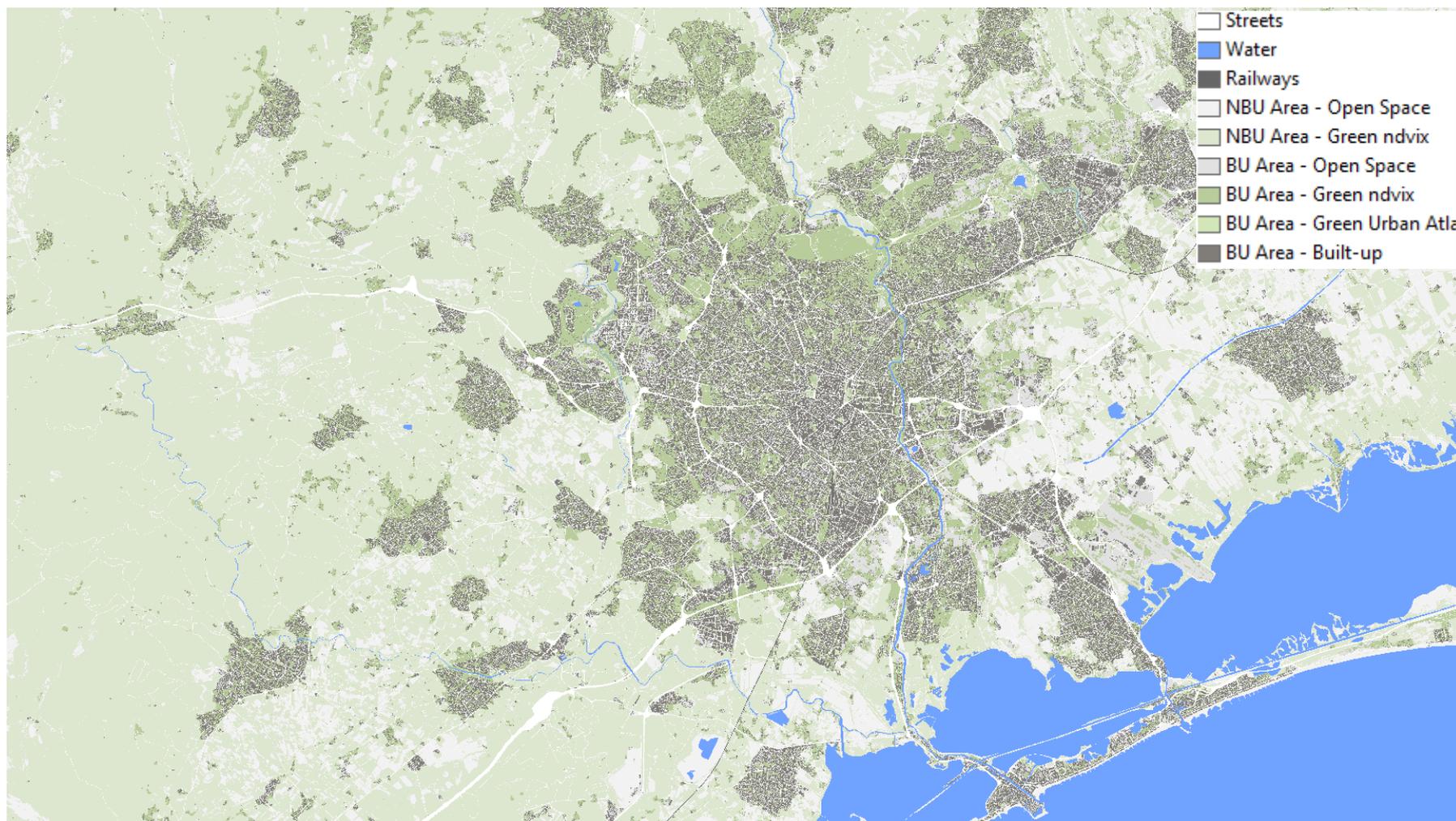


Symbolic Machine Learning (workflow)

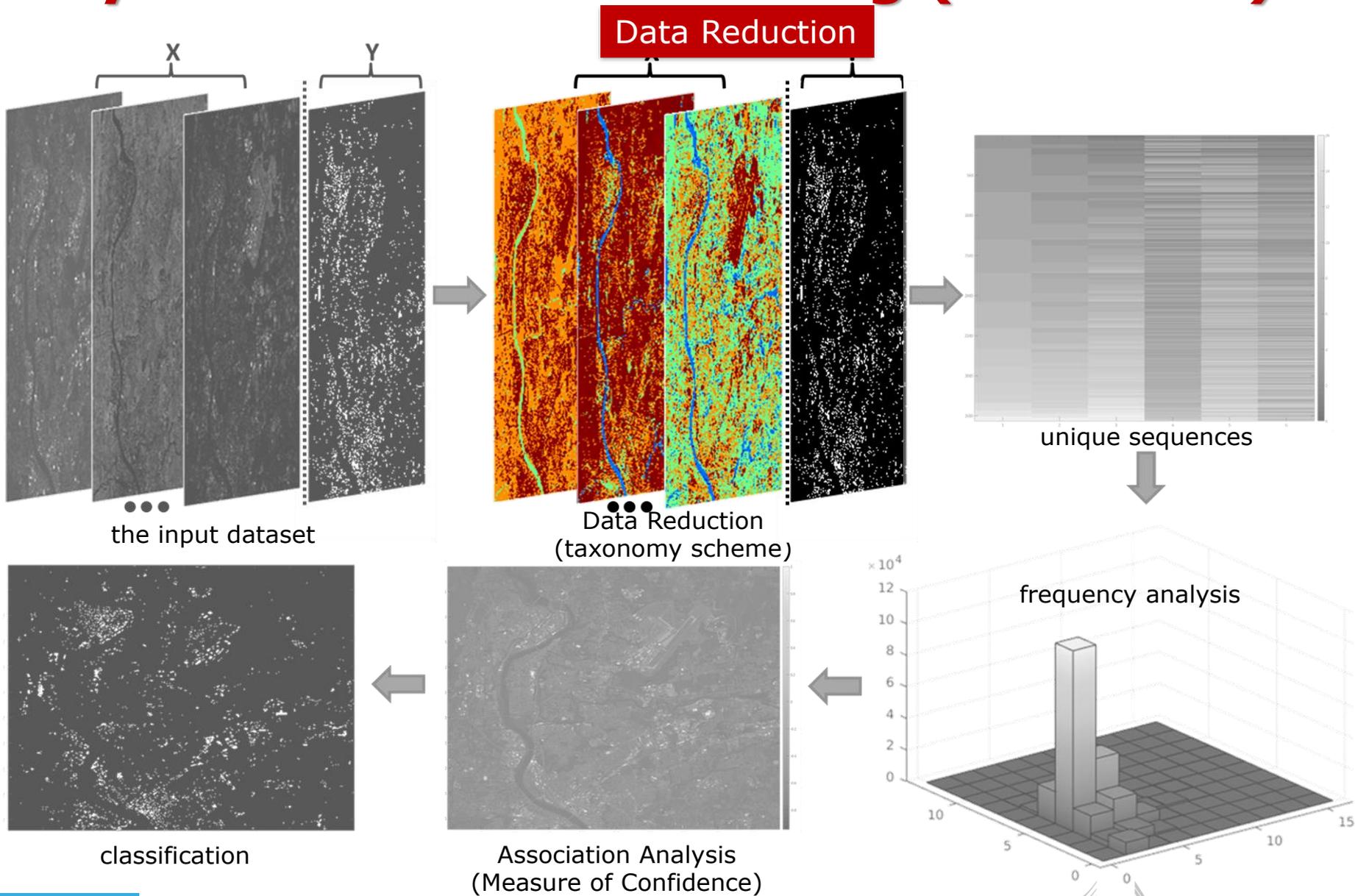


REFERENCE DATA:

Landcover classification : e.g. European Settlement Map – 2.5 and 10 m resolution



Symbolic Machine Learning (workflow)



DATA REDUCTION:

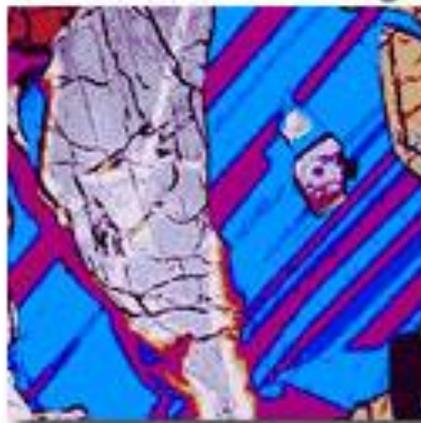
Median Cut Algorithm

Original Image

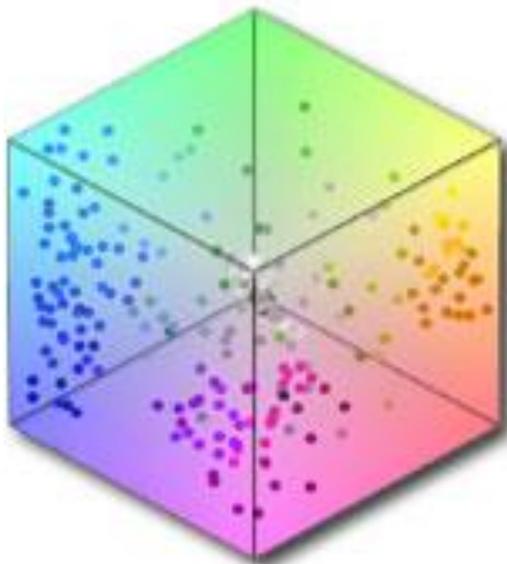


(23838 Colors)

Color Reduced Image



(16 Colors)

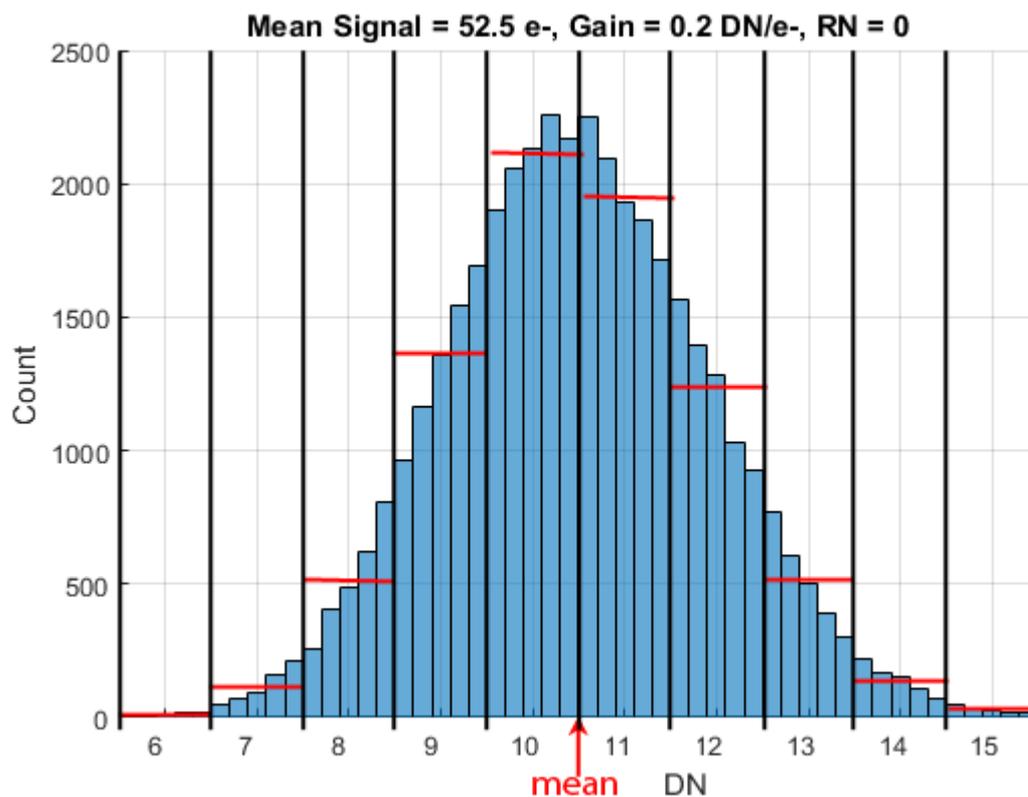


Color Sample
of Image

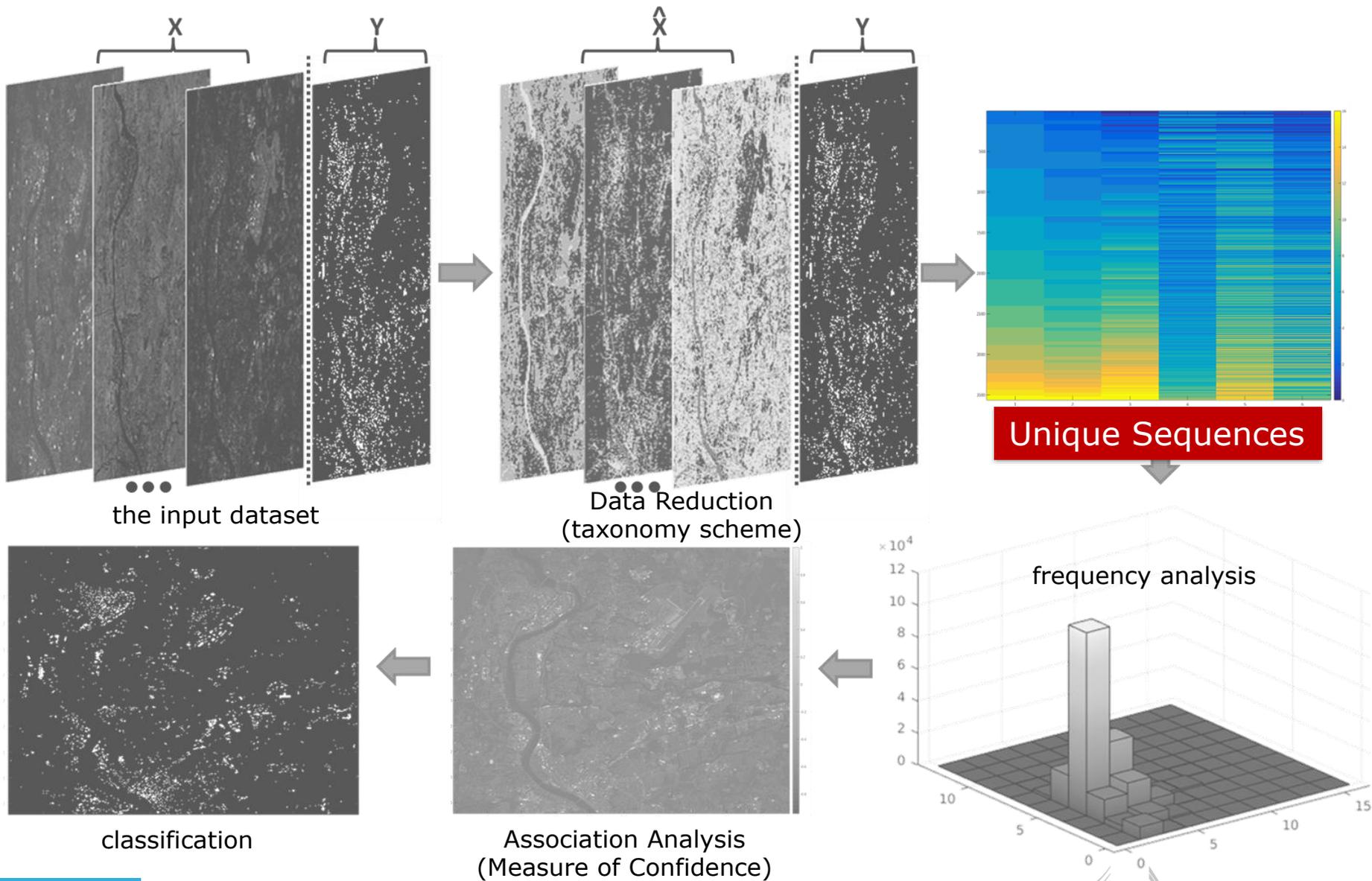


16 Color Palette

Uniform quantization



Symbolic Machine Learning (workflow)



UNIQUE SEQUENCES:

Feature 1

2	4	10	12	7
8	9	4	4	3
2	1	11	12	11
..
..

Feature 2

2	5	9	12	7
7	1	1	2	7
5	5	9	8	8
..
..

Feature 3

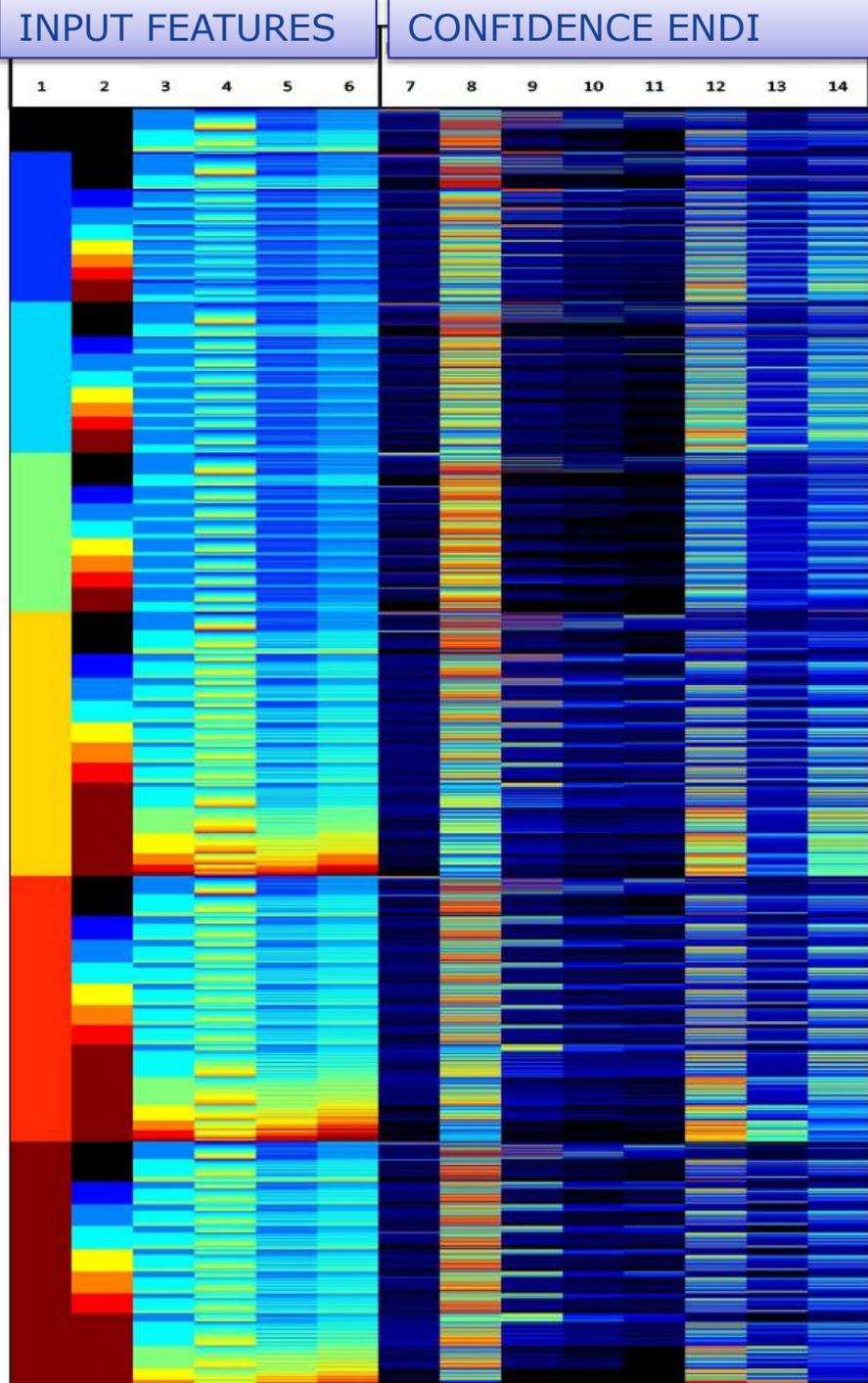
1	4	12	10	11
7	9	4	4	5
3	4	5	9	6
..
..

Sequence 1

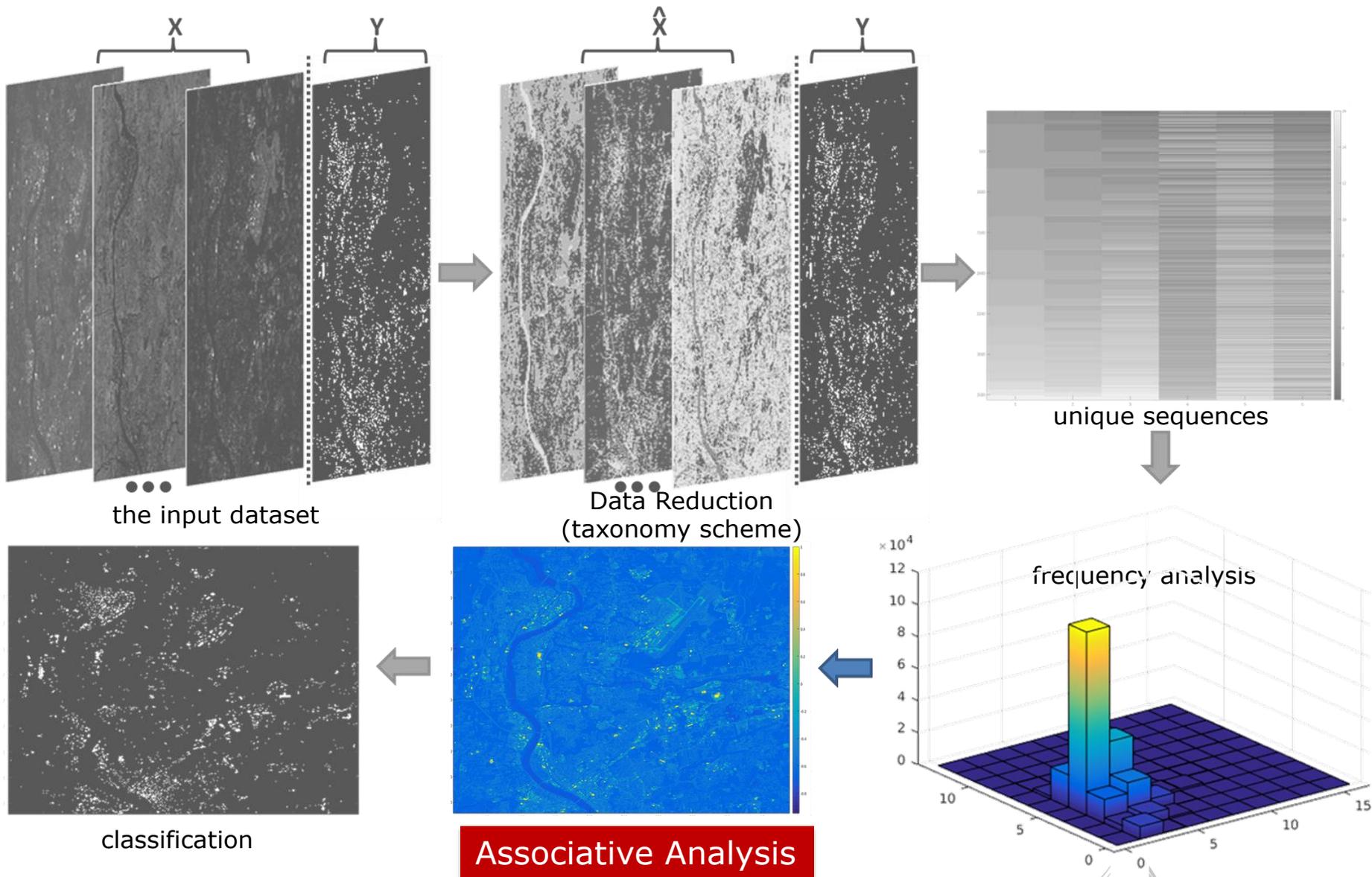
2
2
1
..
..

Sequence 2

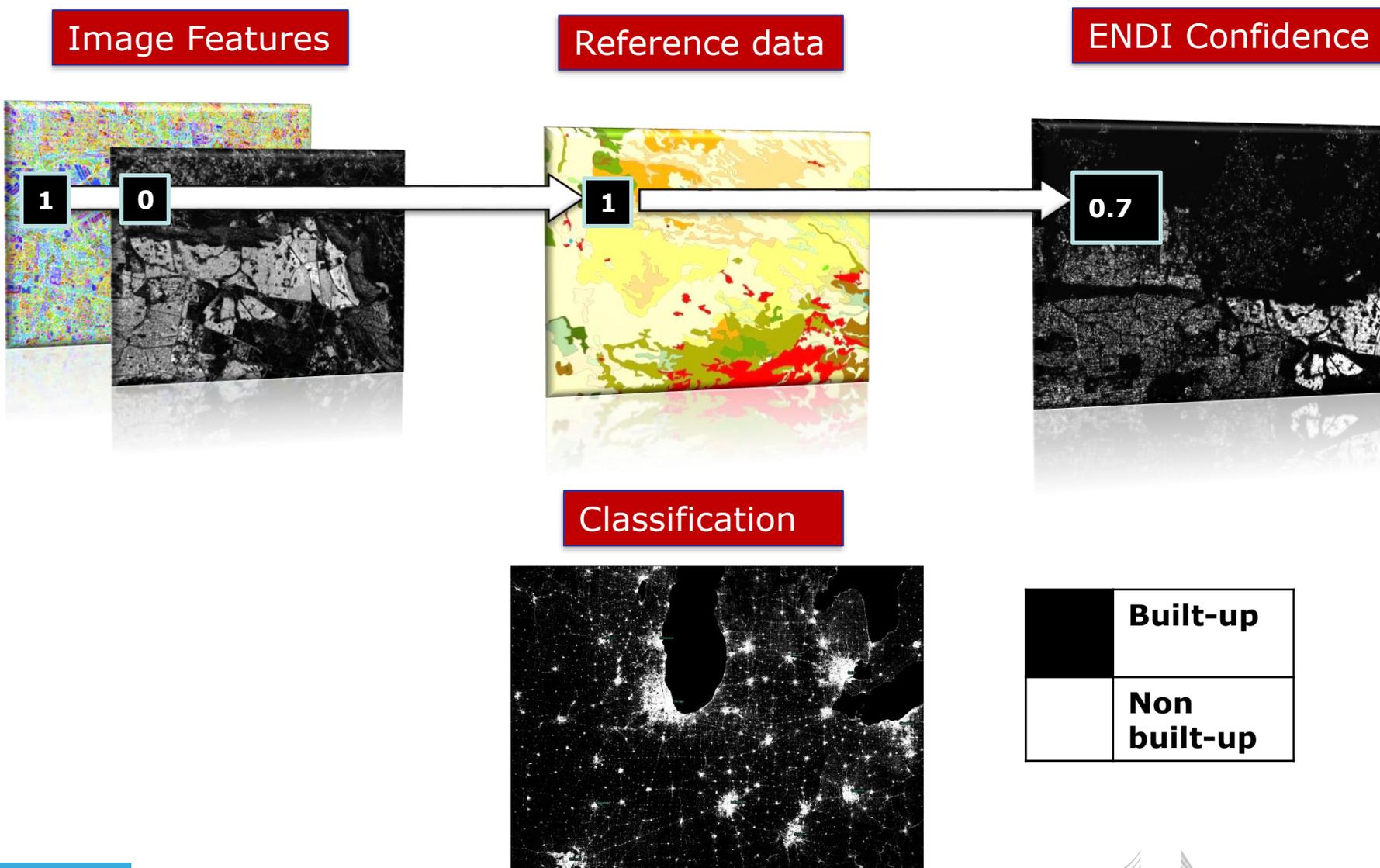
4
5
4
..
..



Symbolic Machine Learning (workflow)



ASSOCIATIVE ANALYSIS



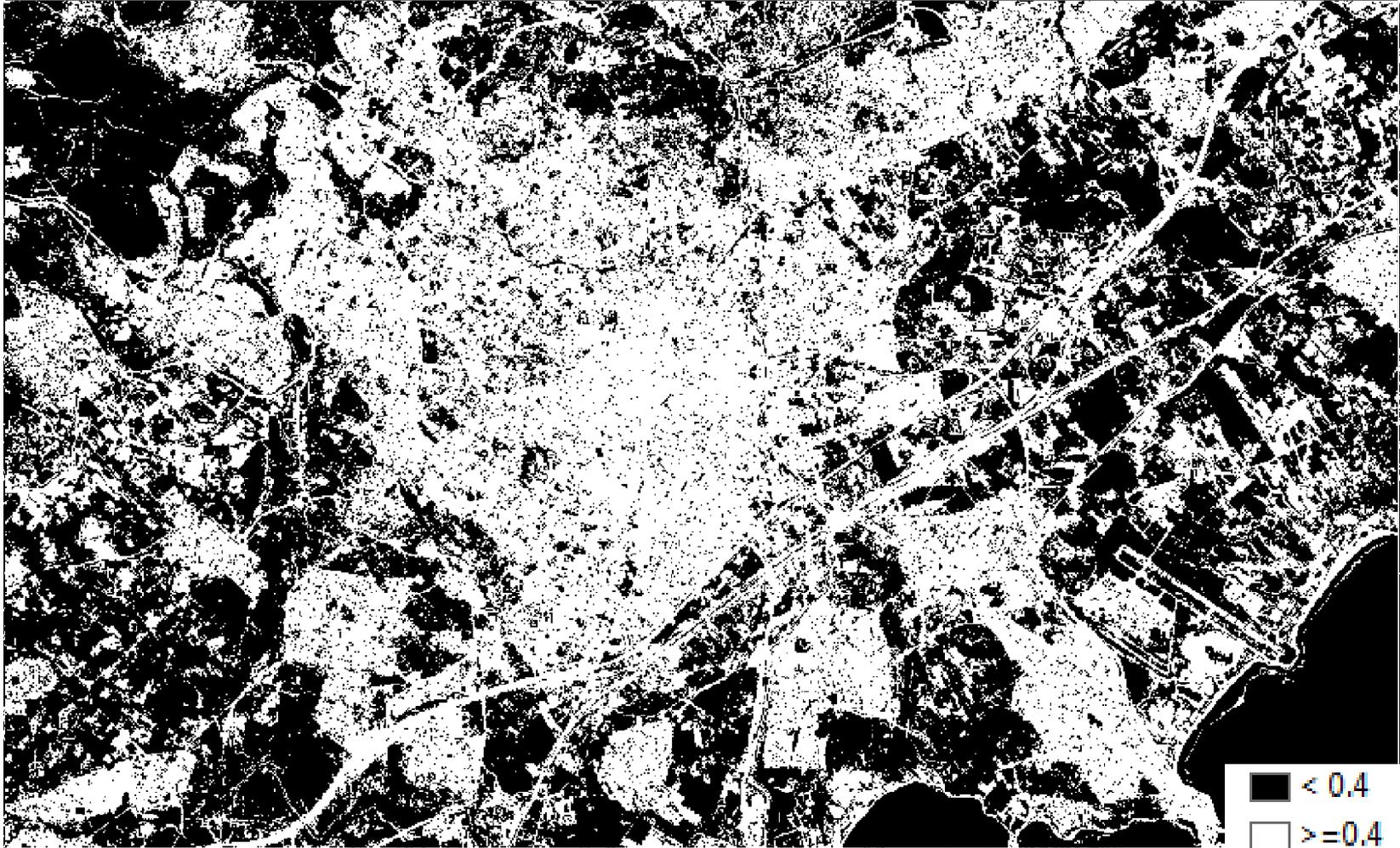
Output Confidence Measure ENDI

The ENDI confidence measure Φ_E^a of the (antecedent) data instances $X(= \bar{X}_{q_i F})$, provided the positive Y^+ and negative Y^- (consequent) data instances, is defined as follows:

$$\Phi_E^a(X, Y^+, Y^-) = \frac{f_{pos-} - f_{neg}}{f_{pos+} + f_{neg}}$$



OUTPUT: CONFIDENCE & CLASSIFICATION



**Built-up = CONFIDENCE
LAYER > 0.4**



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



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



Result of SML classification at 2.5 m spatial resolution.

*Pesaresi M., V. Syrris and Julea A.M. **Benchmarking of the Symbolic Machine Learning classifier with state of the art image classification methods - application to remote sensing imagery.** EUR 27518 EN, 2015. doi:10.2788/638672.*

PANTEX

*textural
characteristics of
panchromatic
satellite*



ec.europa.eu/jrc

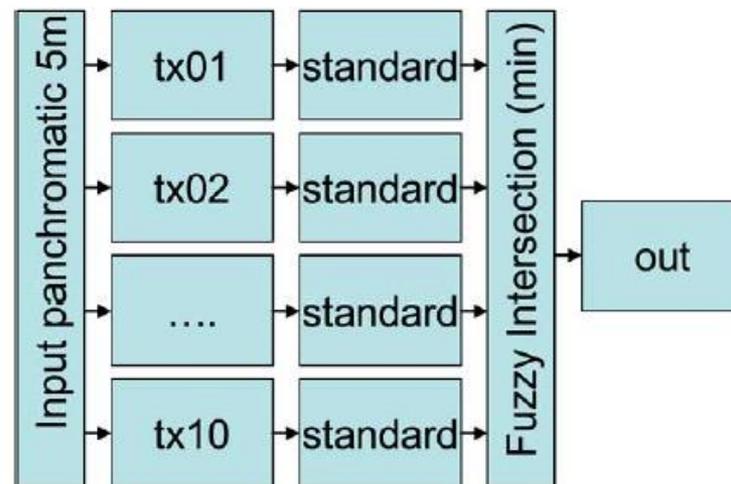


Joint Research Centre

the European Commission's
in-house science service

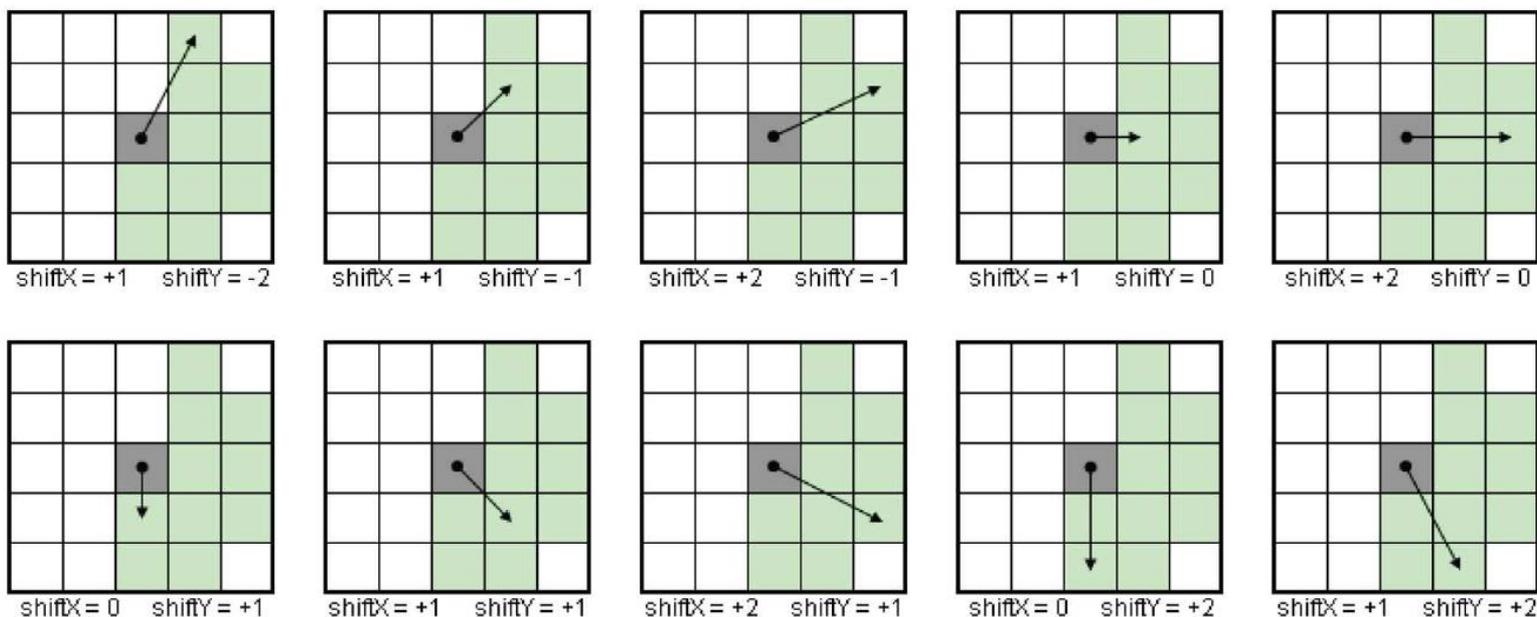
BASIC PRINCIPLES

- Based on the calculation of texture measures (**CONTRAST**) derived from the gray-level co-occurrence matrix (GLCM)
- Uses anisotropic rotation-invariant texture-derived measurement for discriminating the presence of built-up
- The use of rotation-invariant textural measures is derived by applying different direction and displacement parameters and based on $\min \cap$ and $\max \cup$ operators instead of the average



Martino Pesaresi, Andrea Gerhardinger, and François Kayitakire, A Robust Built-Up Area Presence Index by Anisotropic Rotation-Invariant Textural Measure, *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING*, VOL. 1, NO. 3, SEPTEMBER 2008, DOI: 10.1109/JSTARS.2008.2002869

BASIC PRINCIPLES



Ten GLCM displacement vectors

BASIC PRINCIPLES

A, B and C:

agricultural fields with different geometry and presence/absence of perimetric roads

E, F and G:

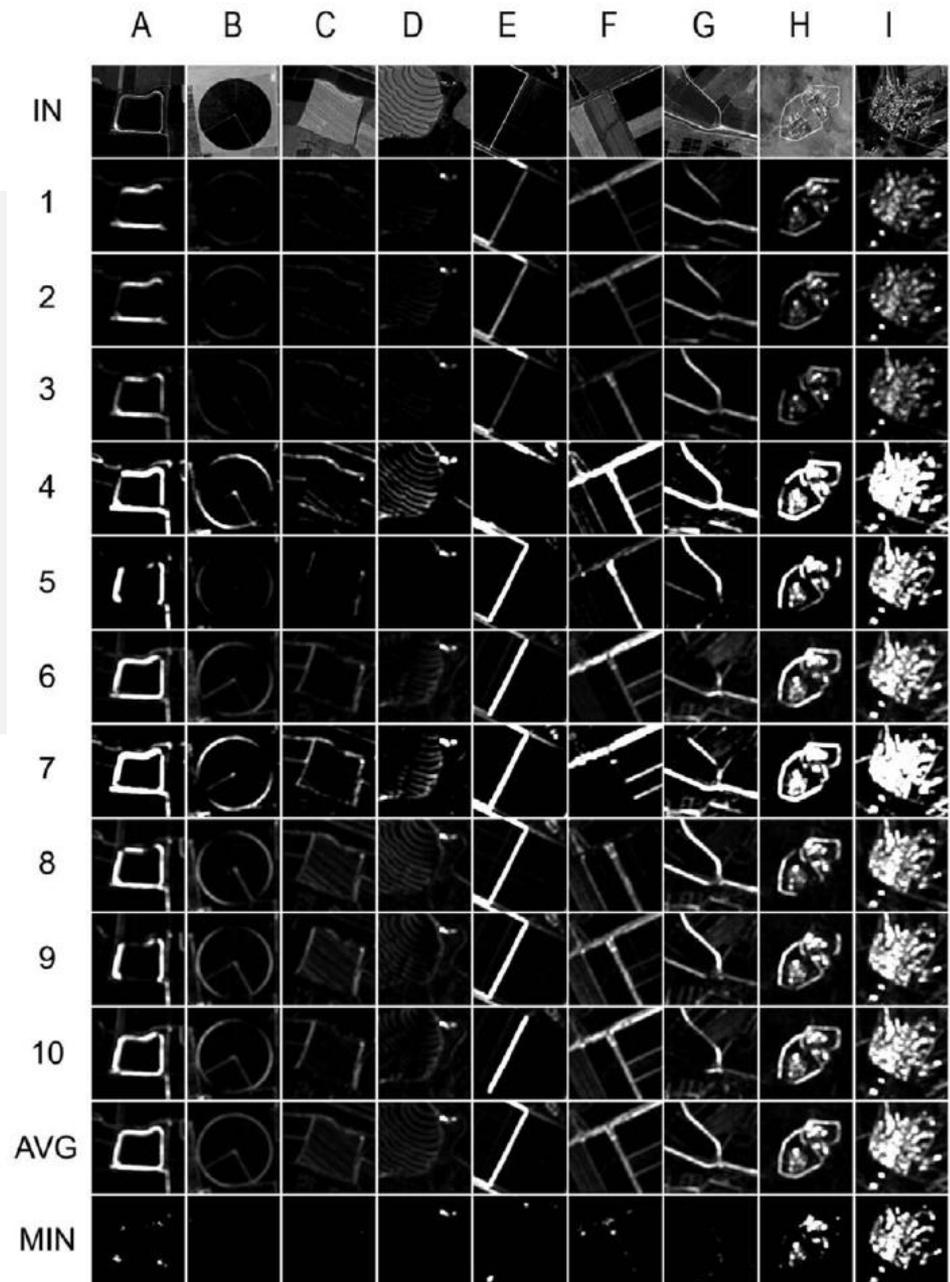
different types of roads in agricultural areas

H:

Built-up with roads

I:

Built-up with no roads





Characteristic-Saliency-Level CSL feature model



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

BASIC PRINCIPLES

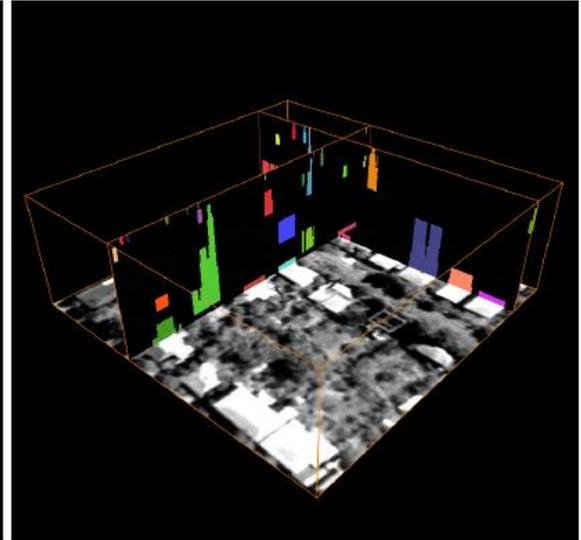
- Based on Differential Morphological Decomposition DMD (Gueguen et al., 2010)
- DMD: a multiresolution pyramid transform defined by series of morphological filters
- Segmentation of DMD using watershed algorithm



(a) Quickbird image



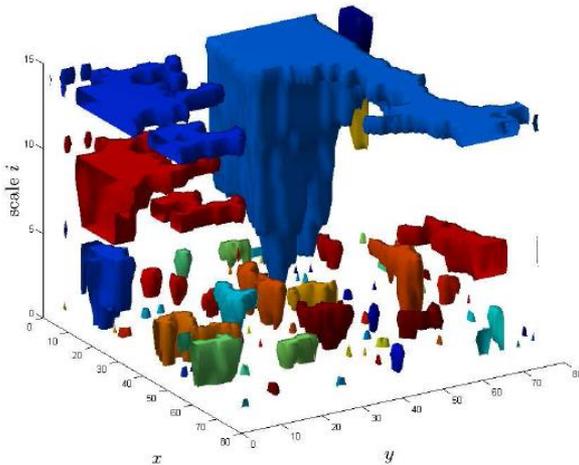
(b) segmentation \mathcal{O}



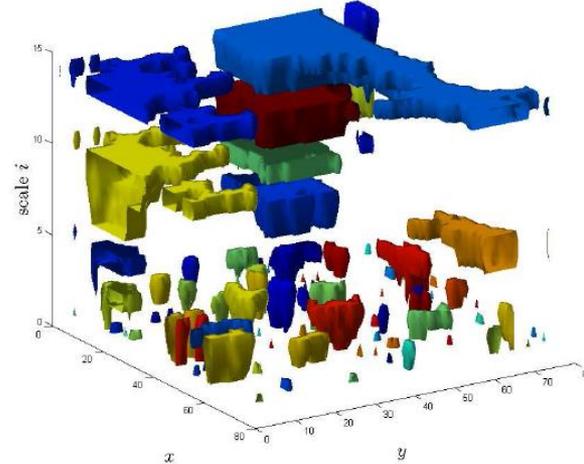
(c) transects of \mathcal{O}

BASIC PRINCIPLES

- Based on **Differential Morphological Decomposition DMD** (Guegen et al., 2010)
- DMD: a multiresolution pyramid transform defined by series of morphological filters
- Segmentation of DMD using watershed algorithm



**Connected components
segmentation of DMD**



Watershed segmentation of DMD

BASIC PRINCIPLES

- Differences between consecutive filtered images of DMPs are taken to create a stack of images containing different size or shape classes of details
- Characteristic-Saliency-Leveling (CSL) is a model allowing the compression and storage of the multi-scale information contained in the DMPs into raster data layers, used for further analytic purposes.
- It consists of three raster layers derived from the image multi-scale decomposition; the Characteristic scale (C), the Saliency (S) and the Leveling (L).



(a)



(b)



(c)

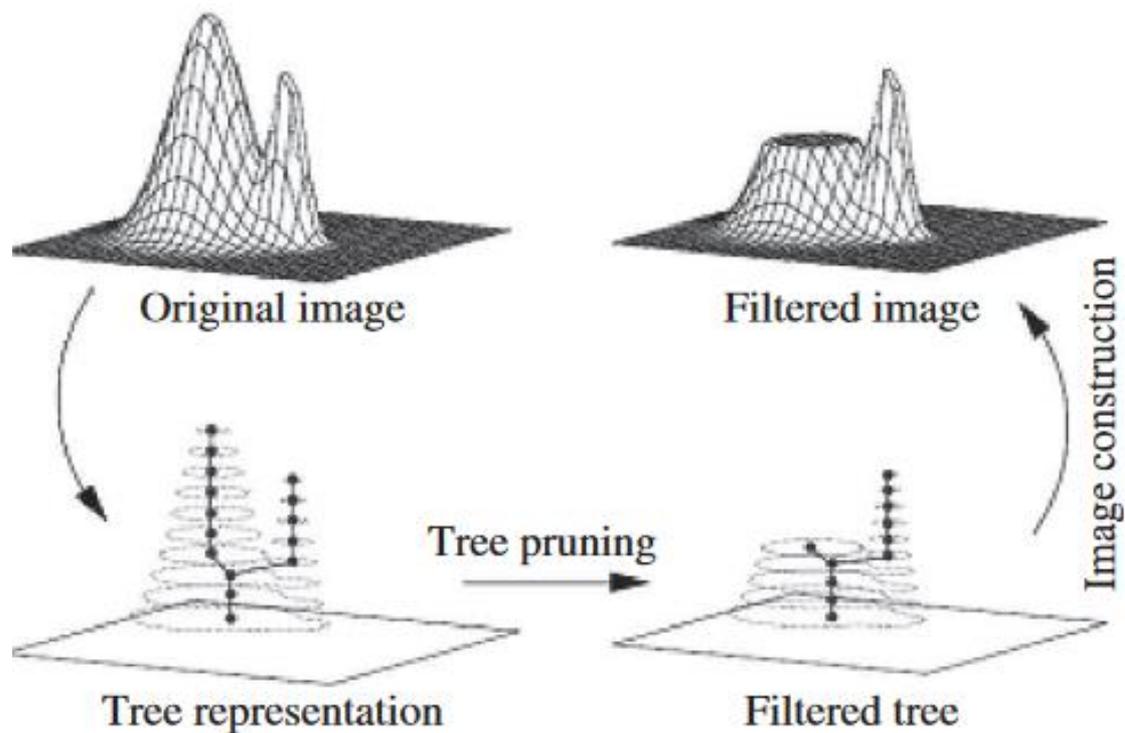
Source article: <http://www.mdpi.com/2220-9964/5/3/22/pdf>

THE CSL MODEL

- The winning scale is referred to as the Characteristic scale C .
- The response associated to the winning scale is called the Saliency S .
- The level of the pixel x before the iteration of the respective attribute filter at the reference scale i is recorded as L .

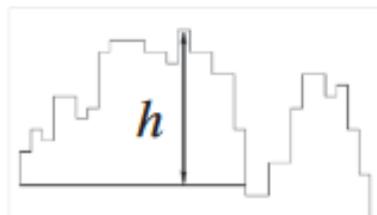
The 3 parameters constitute a non-linear mixture model called the CSL model

THE CSL MODEL

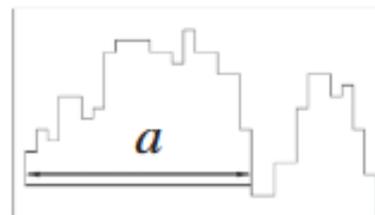


1 - Construction of a tree structure of the image (Max Tree) corresponding to the DMP

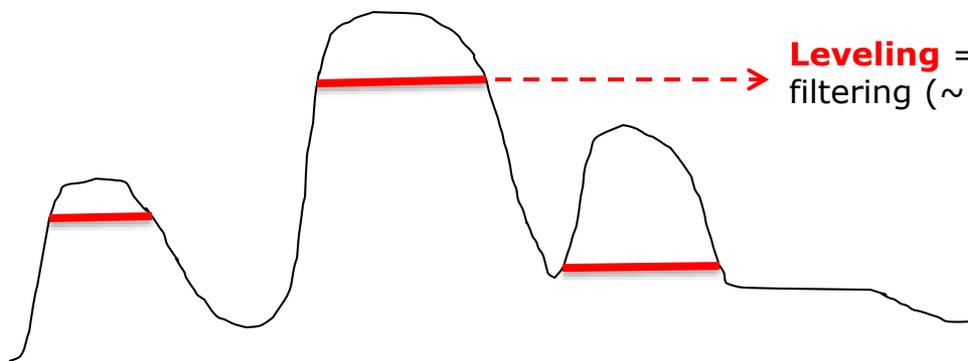
THE CSL MODEL



Saliency
(Closing /Opening)



Characteristic(Closing /Opening)

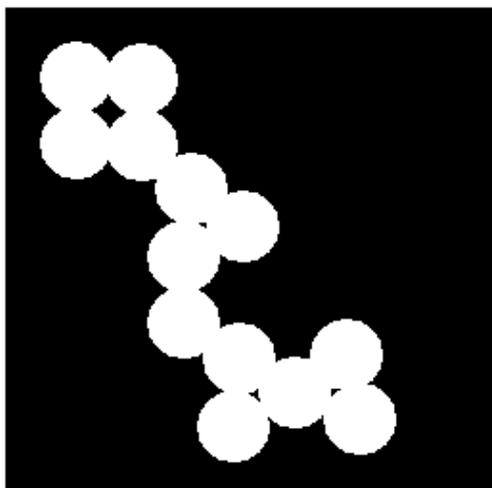


Leveling = Remaining signal after filtering (\sim contrast)

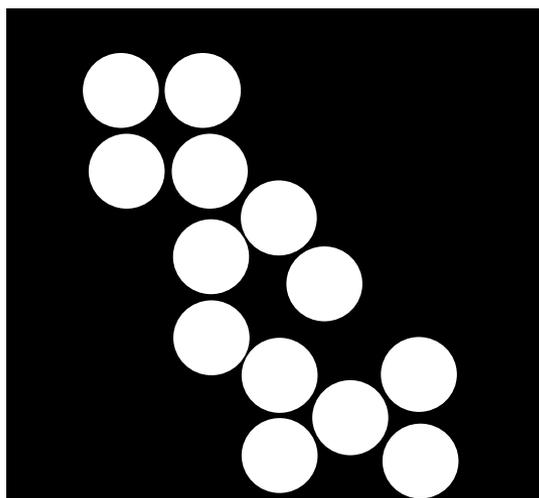
2 – Tree filtering and calculation of the CSL parameters

Morphological operators

- **Opening** : is erosion followed by dilation.
- **Closing**: is dilation followed by erosion.



Binary image

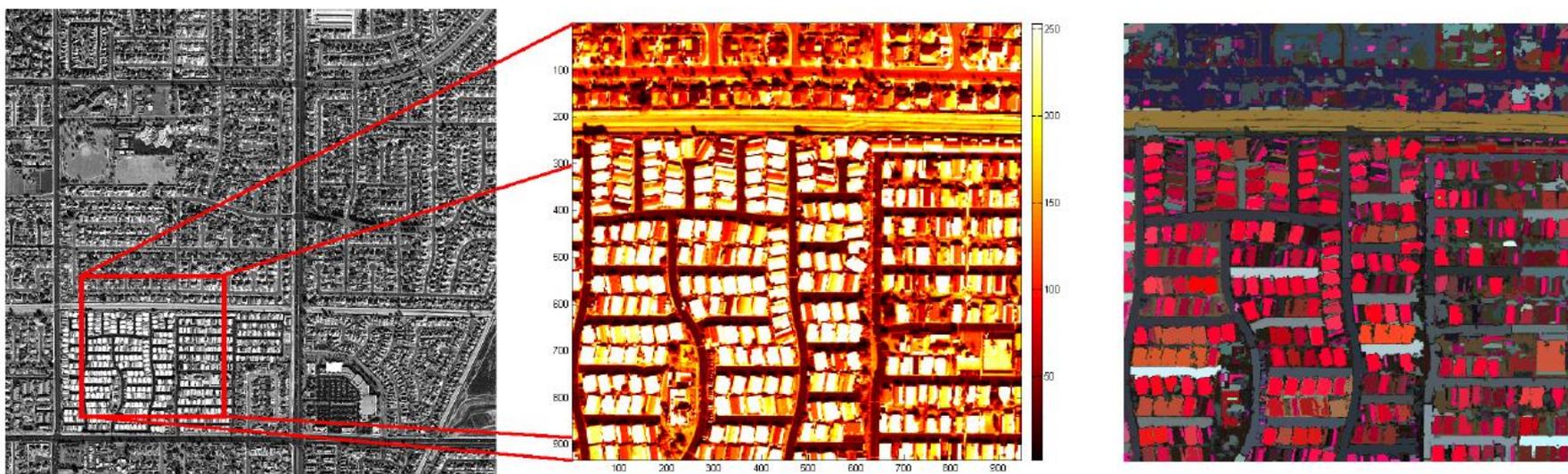


Opening



Closing

Example of Built-Up Segmentation



Ontario (LA), USA
WorldView 1@DigitalGlobe; 2008;
Distributed by Eurimage.

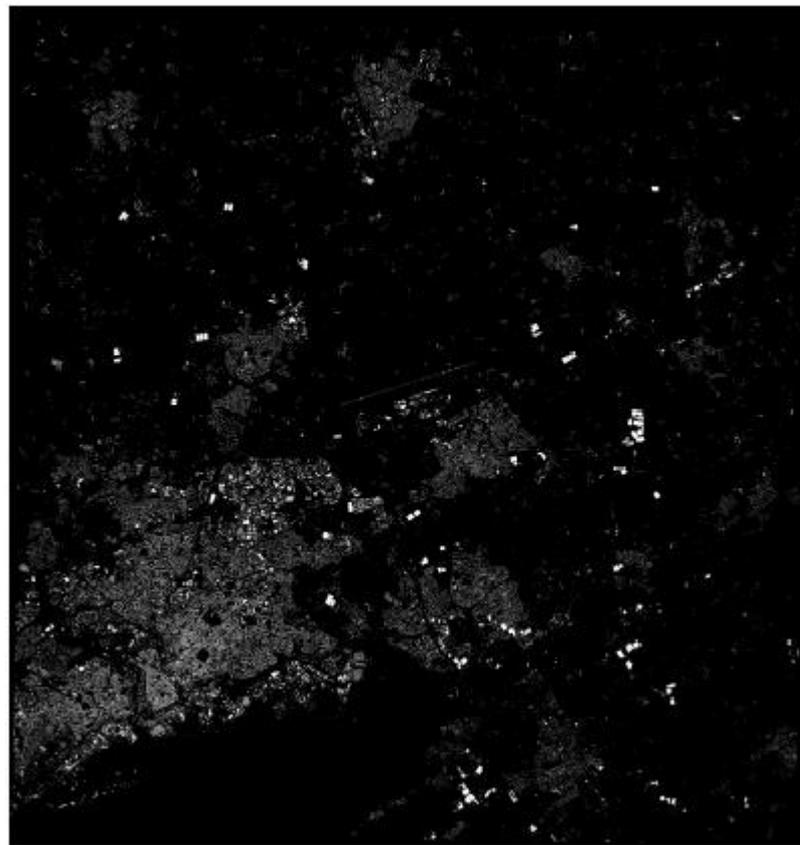
Low-level semantic layer
Warmer colors indicate
high reflectance

CSL2RGB layer
Warmer colors indicate
higher confidence in built-up detection

BASIC PRINCIPLES



WorldView 2 RGB image of the city and its surroundings at 1.6 m spatial resolution.

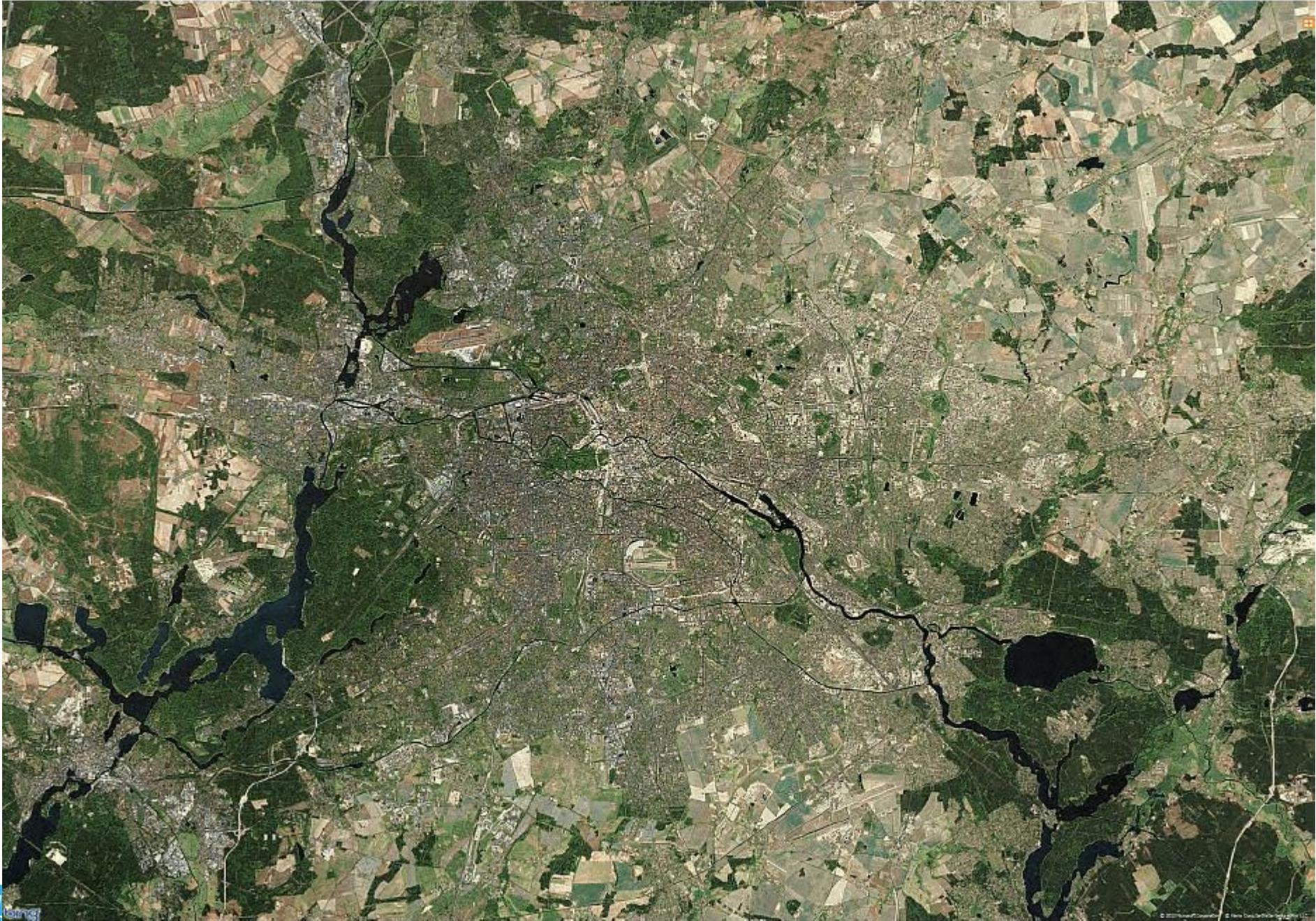


CSL-segmentation

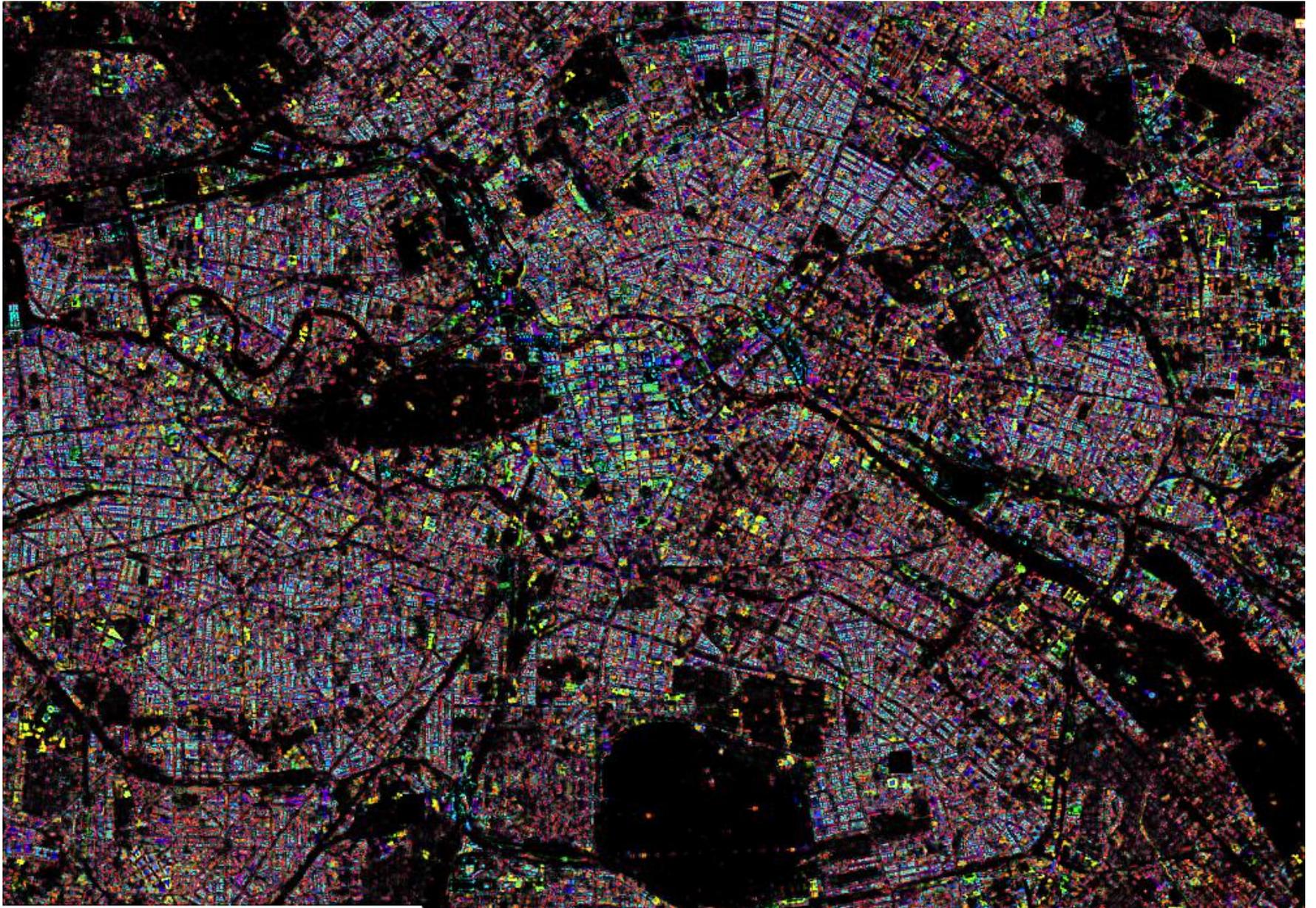
BASIC PRINCIPLES



**WorldView 2 RGB image of the city
and its surroundings at 1.6 m
spatial resolution.**









Stay in touch



JRC Science Hub:
ec.europa.eu/jrc



YouTube:
JRC Audiovisuals



Twitter: ***@EU_ScienceHub***



Vimeo:
Science@EC



Facebook: ***EUsciencehub***



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

