AAXY tutorial

Contents

Introduction

AAXY stands for *associative analysis between X and Y*. The AAXY is a general tool allowing to perform associative analysis between any spatial data instances (continuous, or symbolic) X and abstract (symbolic) spatial data instances Y. In the AAXY the analytics task is done by Symbolic Machine Learning (SML) operating on the X data instances translated to data sequences and by implementation of an objective data association interestingness measure called Evidence-based Normalized Difference Index (ENDI). The AAXY tool is the core technology that was used to solve the problem to translate the global remote sensing data records from heterogeneous sensors available since the epoch 1975, to the abstract class of “built-up area” as required by the GHSL baseline data production. The SML approach was introduced in [1]. The SML was demonstrated superior to the state-of-the-art machine learning approaches in remote sensing data scenarios where large noise is present in the learning set and abstract classes are hill-defined in the feature space (see [3]). The SML is considered a suitable solution for spatial big data analytics tasks [2].

The aim of this tutorial is didactical: to provide a practical guide on the use of the AAXY tool with a specific set of data input. In order to run the tutorial, you will need to download the AAXY tools (http://ghsl.jrc.ec.europa.eu/documents/AAXY.zip), download the tutorial data (http://ghsl.jrc.ec.europa.eu/documents/Tutorial_AAXY.zip), follow step-by-step the instructions included in this document, observe the results and compare with the input data (image and land cover).

The input data of this tutorial is composed by X data collected from the Landsat satellite sensor over an area of 106x 86 km in the inland region of Venice, Italy. The Y data used in the tutorial are the abstract classes as reported by the CORINE Land COVER (CLC) classification schema (http://www.eea.europa.eu/publications/COR0-landcover) in the same spatial region.

In the tutorial, some CLC classes provided to the SML as abstraction to find in the X (image) data instances. Specifically:

1. Urban fabric (all)
2. Urban fabric (industrial and commercial)
3. Urban fabric (residential)
4. Forest Broad-leaved
5. Forest Coniferous
6. Water
7. Wetlands

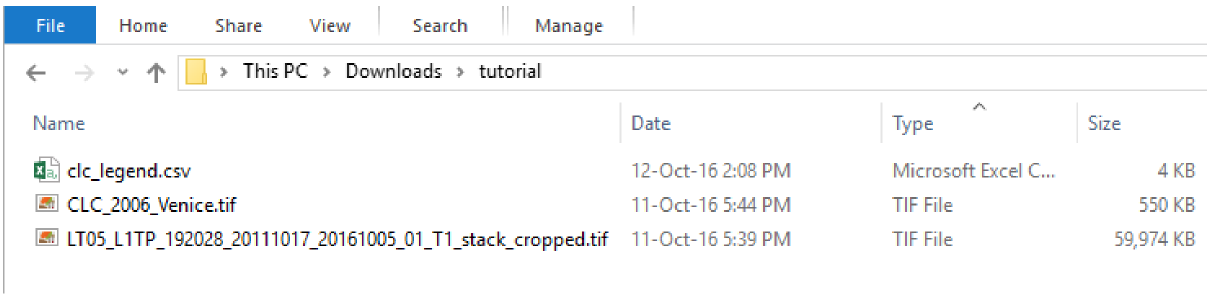
The AAXY tool will provide the output describing how the image data instances are associated to the abstract classes provided as example, using the ENDI as measure of the association. In this example, the X data instances are recorded in 2011 with 30 meters of spatial resolution, while the Y data instances are produced in 2006 at 100 metres of spatial resolution. The capacity to handle temporal, spatial, and abstraction thematic gaps and inconsistencies between the X and Y sets is a characteristic of the AAXY data analytic tool that made this tool suitable for big data analytics scenarios. The ENDI output can be used as input of successive steps of data binarization and fusion aiming to implement single or multiple abstraction classification models. They are not included in this tutorial but you may have a look in the examples provided in [1].

Get the data

All the data needed for this tutorial can be downloaded in a zip file from the GHSL website:

http://ghsl.jrc.ec.europa.eu/documents/Tutorial_AAXY.zip

Download the zip file and extract the content to a folder, e.g. *Downloads\tutorial*.



The folder contains 3 files:

1. **LT05_L1TP_192028_20111017_20161005_01_T1_stack_cropped.tif**: a multiband Landsat 5 cropped scene of the Venice area
2. **CLC_2006_Venice.tif:** a Corine land cover scene of the same area
3. **clc_legend.csv**: the legend of the Corine land cover classes

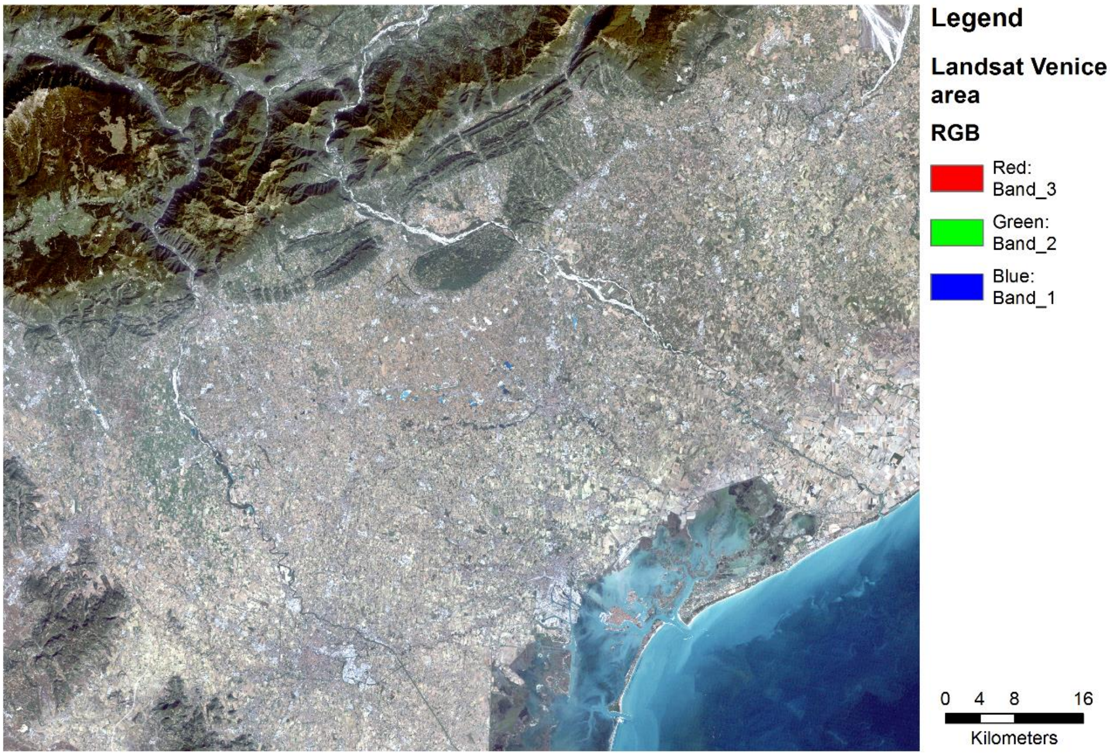
We will use AAXY tool to extract ENDI measure from the Landsat image using the Corine land cover as reference data.

To get the AAXY executable from the GHSL website go at this link:

http://ghsl.jrc.ec.europa.eu/documents/AAXY.zip

Download and unzip somewhere in your hard drive. In this tutorial we will copy the aaxy.exe file in the *Download/tutorial* folder.

Input: LT05_L1TP_192028_20111017_20161005_01_T1_stack_cropped.tif



This image is obtained from the Landsat 5 scene *LT05_L1TP_192028_20111017_20161005_01_T1*, which can be downloaded for free at USGS:

http://earthexplorer.usgs.gov/metadata/12266/LT51920282011290MOR00/

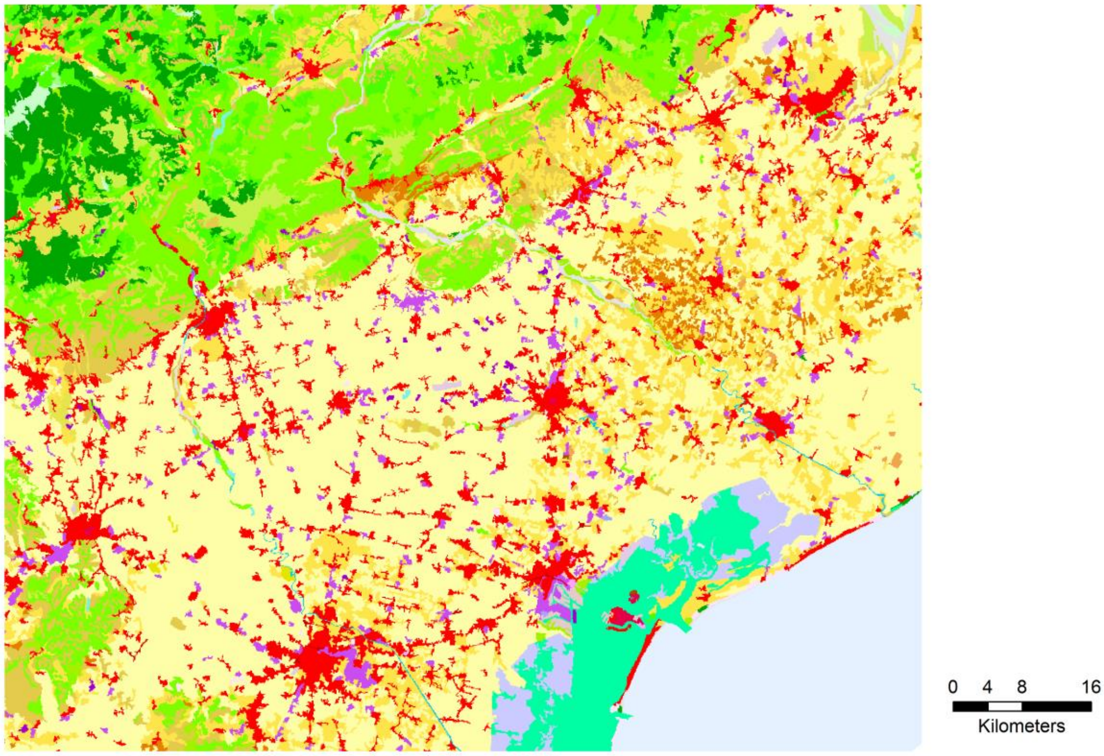
The file used here it's a multiband raster composed by bands 1,2,3,4,5,7 and clipped to extent:

Top	5107998.51413
Bottom	5021408.49693
Left	689147.886671
Right	795473.349426109

The image is an RGB bands image created using the following band combination:

RED	GREEN	BLUE
Band 3	Band 2	Band 1

Reference: CLC_2006_Venice.tif



This image is the Corine land cover of the same area that we will use as reference data for the associative analysis.

Legend: clc_legend.csv

This table describes the link between the values in the Corine land cover image and the classes that they represent.

For instance, value range from 1 to 11 corresponds to “Artificial surfaces”, while values above 39 regard “Water bodies”.

Workflow

The workflow of this tutorial is very simple: we compute the same operation using aaxy.exe with different parameters and we apply a very simple post processing to have same setup for all tests.

All command options are described in detail in the AAXY User Guide document [4], here we focus on how to use the tool to obtain different results using the same reference data.

The following tests report about

1. the classes used from the reference
2. the exact command used to extract the data
3. the results obtained.

If we break down the commands, the main parameters we are going to change are:

- **-py:** values corresponding to the positive reference values
- **-ny:** values corresponding to the negative reference values

These are used to specify which class we want to use to get the confidence measure from our input scene, and they are different for every test.

The test result is a raster file with floating point value with range -1, 1 which represent the ENDI absolute values extracted from the Landsat image. Several type of ENDI are available through the tool, to have the complete list of parameters please check again the AAXY User Guide [4].

The post processing performed to enhance visualization is made of 2 operations:

1. Apply a value stretch to Min Max of -1 and 1
2. Apply a color scheme ranging from blue (-1) to red (1)

Test 1: Urban fabric (all)

In this test we will extract confidence information about the *"Urban fabric"* and *"Industrial, commercial and transport units"* classes.

Classes

From the Corine land cover legend we see that these classes correspond to grid values 1 to 6. All other classes are used as negative reference.

Command

```
aaxy.exe -x LT05_L1TP_192028_20111017_20161005_01_T1_stack_cropped.tif -y
CLC_2006_Venice.tif -q 8 -py 1:6 -ny 7:48 -o CLC_Urban_fabric_all.tif
```

Result

Test 2: Urban fabric (industrial and commercial)

In this test we will extract confidence information about the *"Industrial, commercial and transport units"* class only.

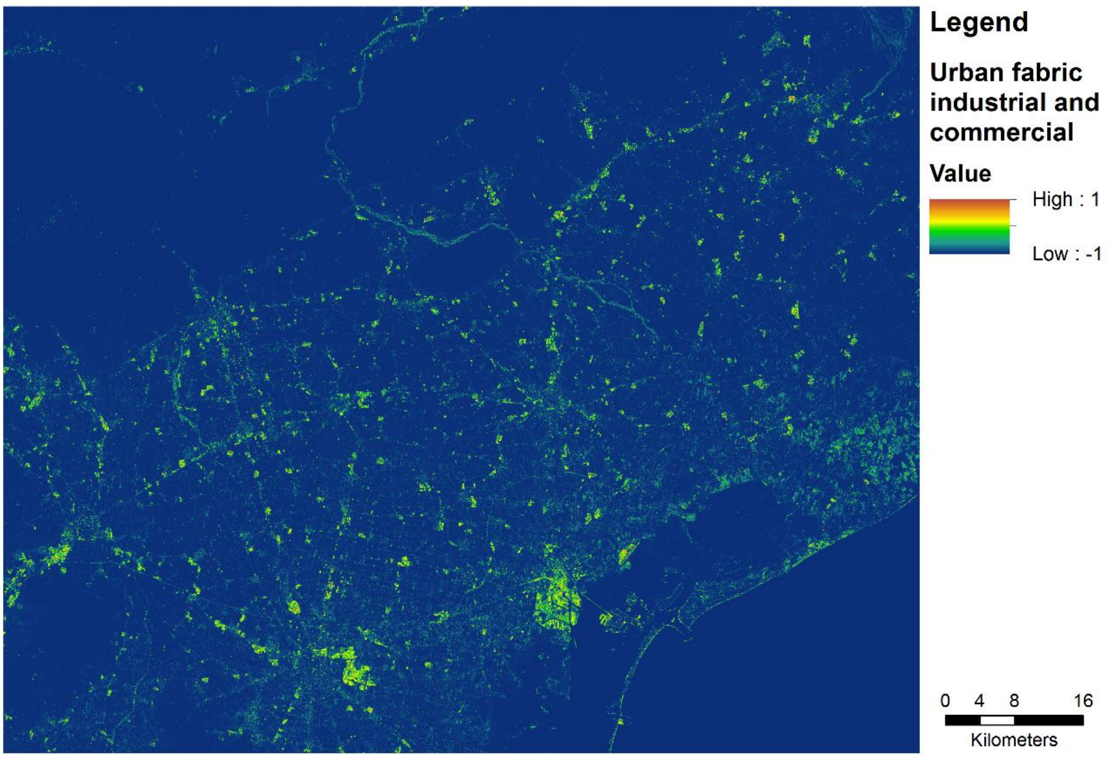
Classes

From the Corine land cover legend we see that these classes correspond to grid values 3 to 6. All other classes are used as negative reference.

Command

```
aaxy.exe -x LT05_L1TP_192028_20111017_20161005_01_T1_stack_cropped.tif -y
CLC_2006_Venice.tif -q 8 -py 3:6 -ny 1,2,7:48 -o
CLC_Urban_fabric_industrial_and_commercial.tif
```

Result



Test 3: Urban fabric (residential)

In this test we will extract confidence information about the *"Urban fabric"* class only.

Classes

From the Corine land cover legend we see that these classes correspond to grid values 1 to 2. All other classes are used as negative reference.

Command

```
aaxy.exe -x LT05_L1TP_192028_20111017_20161005_01_T1_stack_cropped.tif -y
CLC_2006_Venice.tif -q 8 -py 1,2 -ny 3:48 -o CLC_Urban_fabric_residential.tif
```

Result

Test 4: Forest Broad-leaved

In this test we will extract confidence information about the "*Broad-leaved forest*" class.

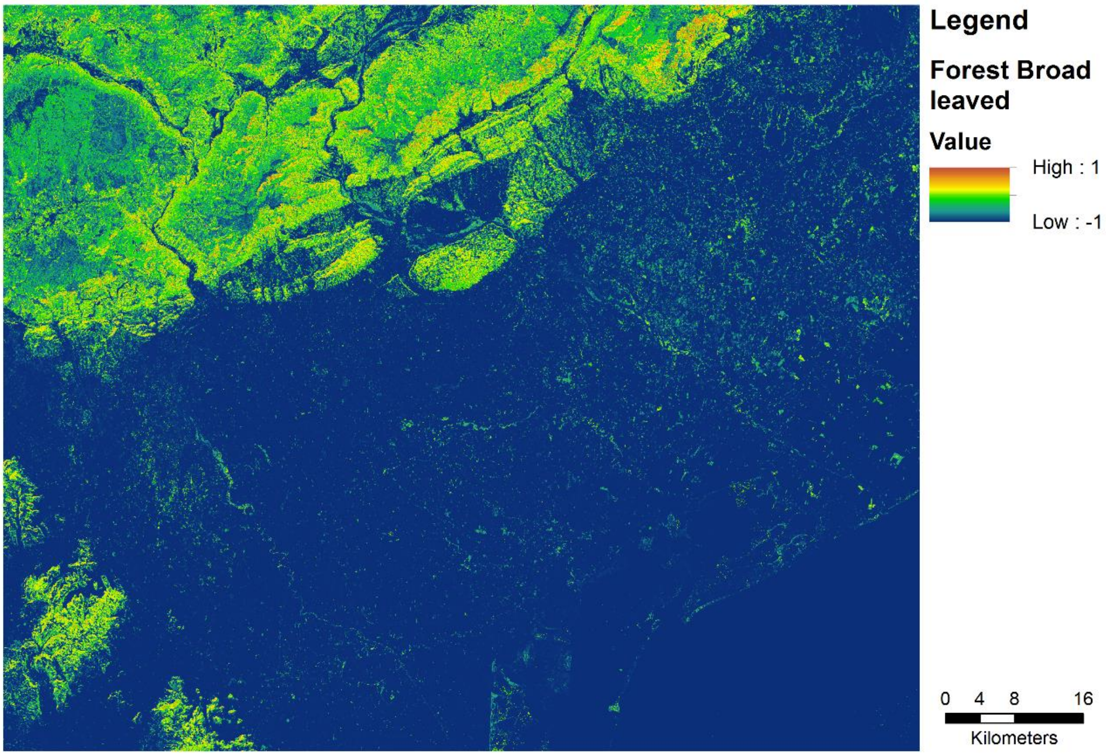
Classes

From the Corine land cover legend we see that these classes correspond to grid value 23. All other classes are used as negative reference.

Command

```
aaxy.exe -x LT05_L1TP_192028_20111017_20161005_01_T1_stack_cropped.tif -y
CLC_2006_Venice.tif -q 8 -py 23 -ny 1:22,24:48 -o CLC_Forest_Broad-leaved.tif
```

Result



Test 5: Forest Coniferous

In this test we will extract confidence information about the *"Coniferous forest"* class.

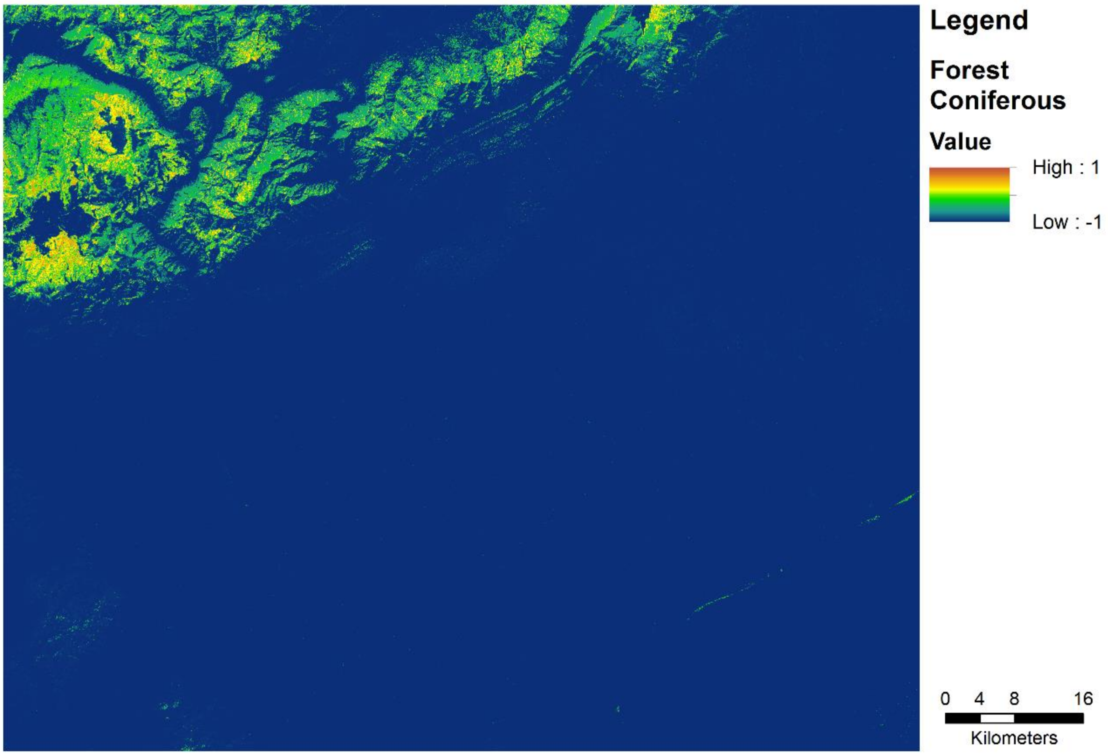
Classes

From the Corine land cover legend we see that these classes correspond to grid values 24. All other classes are used as negative reference.

Command

```
aaxy.exe -x LT05_L1TP_192028_20111017_20161005_01_T1_stack_cropped.tif -y
CLC_2006_Venice.tif -q 8 -py 24 -ny 1:23,25:48 -o CLC_Forest_Coniferous.tif
```

Result



Test 6: Water

In this test we will extract confidence information about the *"Water bodies"* classes.

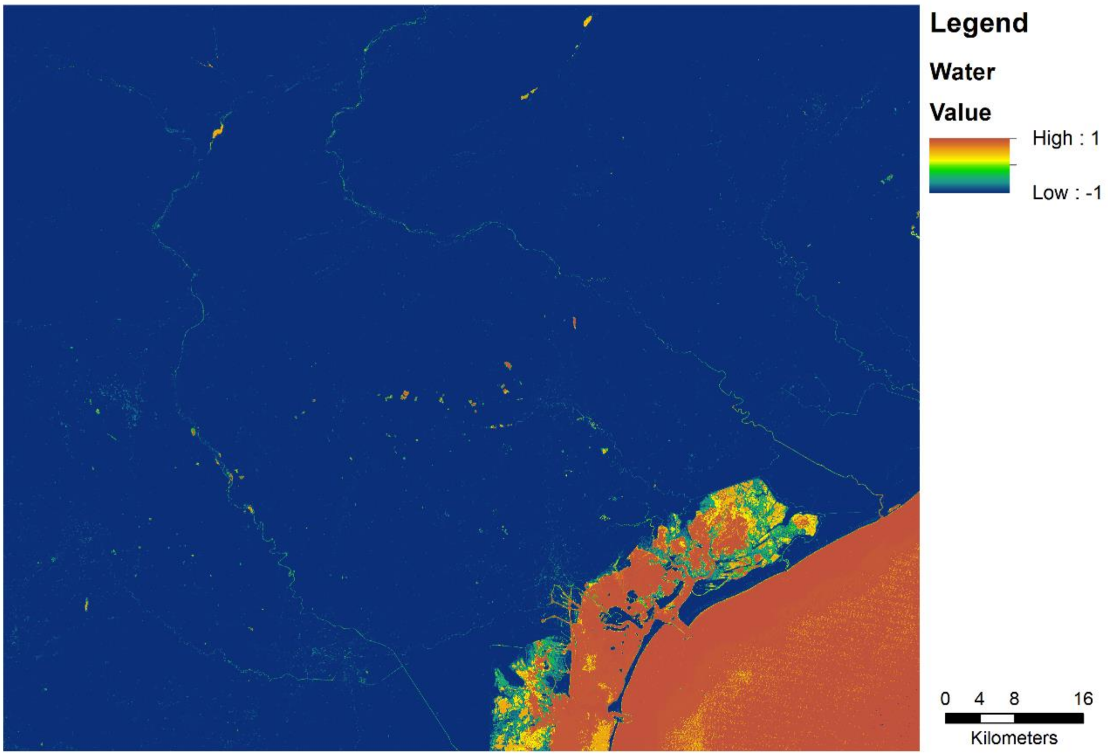
Classes

From the Corine land cover legend we see that these classes correspond to grid values 40 to 48. All other classes are used as negative reference.

Command

```
aaxy.exe -x LT05_L1TP_192028_20111017_20161005_01_T1_stack_cropped.tif -y
CLC_2006_Venice.tif -q 8 -py 40:48 -ny 1:39 -o CLC_Water.tif
```

Result



Test 7: Wetlands

In this test we will extract confidence information about the *"Wetlands"* class.

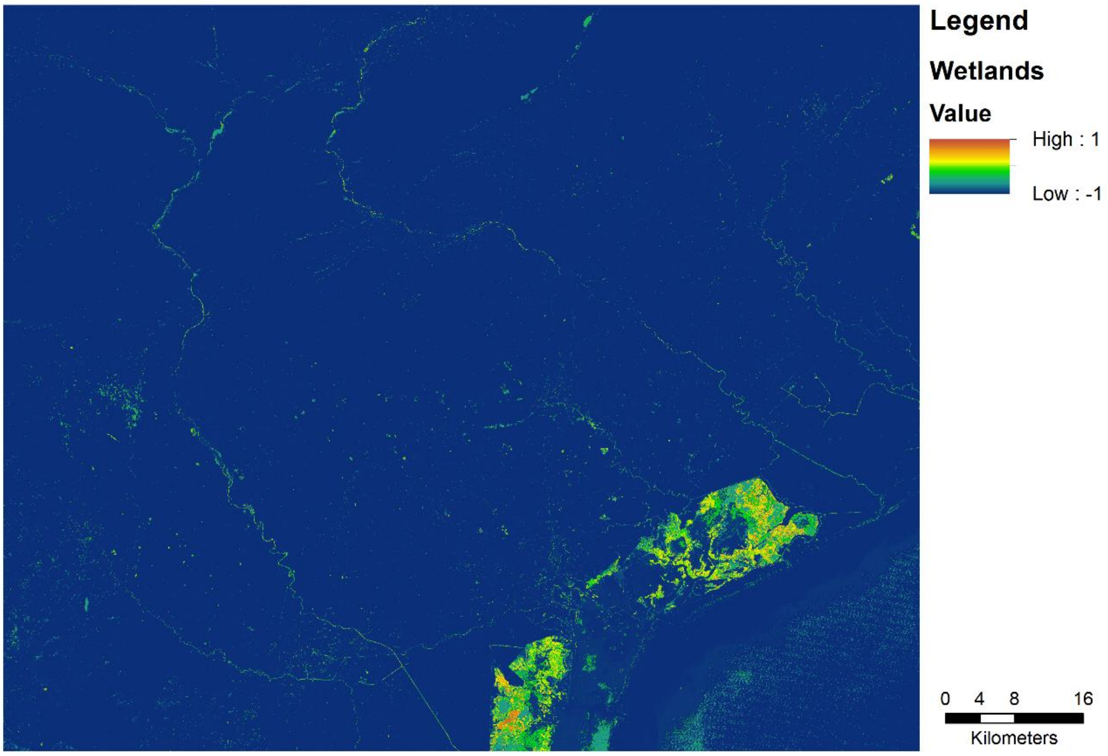
Classes

From the Corine land cover legend we see that these classes correspond to grid values 35 to 39. All other classes are used as negative reference.

Command

```
aaxy.exe -x LT05_L1TP_192028_20111017_20161005_01_T1_stack_cropped.tif -y CLC_2006_Venice.tif -q 8 -py 35:39 -ny 1:34,40:48 -o CLC_Wetlands.tif
```

Result



References

[1] A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning

Pesaresi, M.; Syrris, V. & Julea, A.

Remote Sensing, 2016 Remote Sens. 2016, 8(5), 399;

doi:10.3390/rs8050399

http://www.mdpi.com/2072-4292/8/5/399

[2] ANALYZING BIG REMOTE SENSING DATA VIA SYMBOLIC MACHINE LEARNING

Pesaresi, M.; Julea, A.; Syrris, V.

doi: 10.2788/854791

https://ec.europa.eu/jrc/en/publication/analyzing-big-remote-sensing-data-symbolic-machine-learning

[3] Benchmarking of the Symbolic Machine Learning classifier with state of the art image classification methods

Pesaresi, M.; Syrris, V. & Julea, A.

JRC Technical Report EUR 27518 EN;

doi:10.2788/638672

http://publications.jrc.ec.europa.eu/repository/handle/JRC97964

[4] AAXY User Guide

Pesaresi, M. ; Maffenini, L.

doi:10.2788/782191

http://publications.jrc.ec.europa.eu/repository/handle/JRC102959

GHSL project

http://ghsl.jrc.ec.europa.eu