

OPEN DATA SCIENCE EUROPE

WORKSHOP

Introduction to ODSE datasets in Python

Sept 6, 2021: 9:00 - 10:30



Leandro Parente



leandro.parente@opengeohub.org



<https://opengeohub.org>



Introduction to ODSE datasets in Python - Outline

- VirtualBox, Docker, Conda...
- What Python libraries will we use?
- What is Open Data Science Europe?
- Which datasets are available?
- How access timeless and temporal datasets?

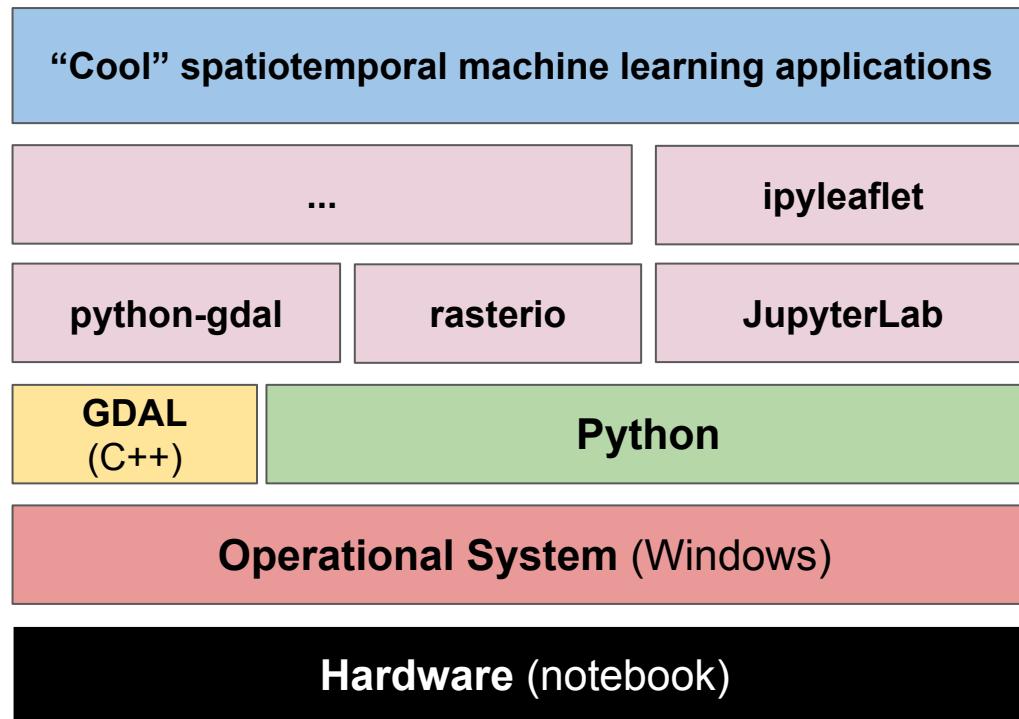


VirtualBox, Docker, Conda...



Why are we using VirtualBox?

VirtualBox, Docker, Conda...



VirtualBox, Docker, Conda...

“Cool” spatiotemporal machine learning applications

...

ipyleaflet

python-gdal

rasterio

JupyterLab

GDAL
(C++)

Python

Operational System (Windows)

Hardware (notebook)

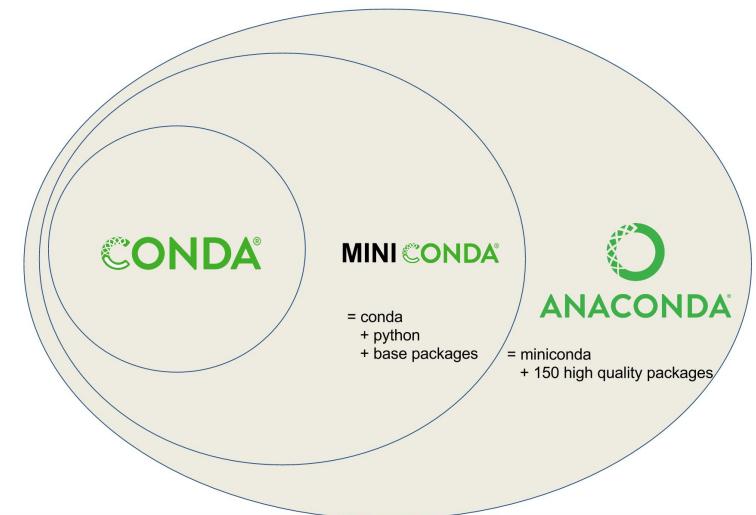
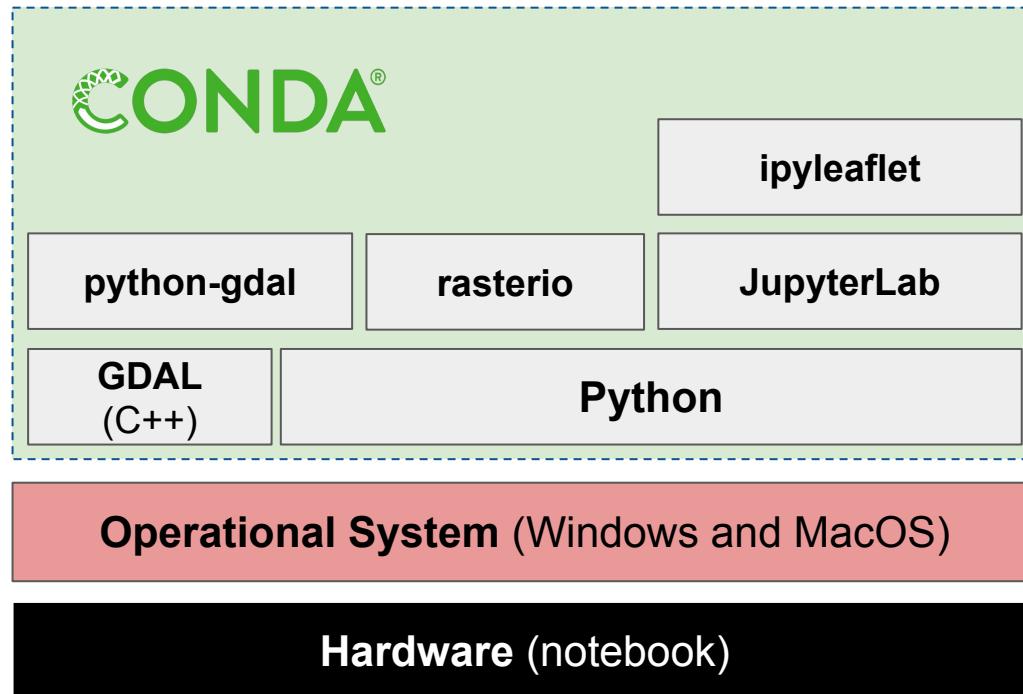
GDAL 3.1.0

 rouault released this on May 7, 2020

- Implement RFC 75: support for multidimensional arrays in MEM, VRT, netCDF, HDF4, HDF5 a MEM and netCDF. Read-only for others. Add gdalmdiminfo and gdalmdimtranslate utilities.
- Implement RFC 76: add capability of writing vector drivers in Python
- New GDAL drivers:
 - COG:** write-only, for Cloud Optimized GeoTIFF
 - EXR:** read/write driver, relying on OpenEXR library
 - ISG:** read-only, for geoid models of the International Service for the Geoid
 - RDB:** read-only, for RIEGL Database .mpx RDB 2 files (#1538) (needs proprietary SDK)
- New OGR drivers:
 - FlatGeoBuf:** read-support and creation (#1742)
 - MapML:** read/write driver for experimental web spec
- Improved drivers:
 - OAPIF driver (renamed from WFS3): updated to OGC API - Features 1.0 core spec
 - GTiff: improve performance of internal overview creation
 - GTiff: GeoTIFF 1.1 support
 - Shapefile driver: add read/creation/update support for .shz and .shp.zip
 - netCDF vector: read/write support for CF-1.8 Encoded Geometries (#1287)
 - VICAR: multiple improvements and write support (#1855)
 - DDS: add read support
- Other improvements:
 - gdalwarp: accept output drivers with only CreateCopy() capabilities
 - gdal_viewshed: new utility for viewshed algorithm
- Remove GFT driver now that the online service no longer exists (#2050)
- New Sphinx-based documentation
- Multiple security related fixes (ossfuzz)
- Continued code linting (cppcheck, CoverityScan, etc.)
- Compatibility with GDAL 3.0:
 - C and C++ API: backward compatible changes
 - C ABI: backward compatible changes
 - C++ ABI: modified
 - Functional changes: see [MIGRATION_GUIDE.TXT](#)

<https://github.com/OSGeo/gdal/releases>

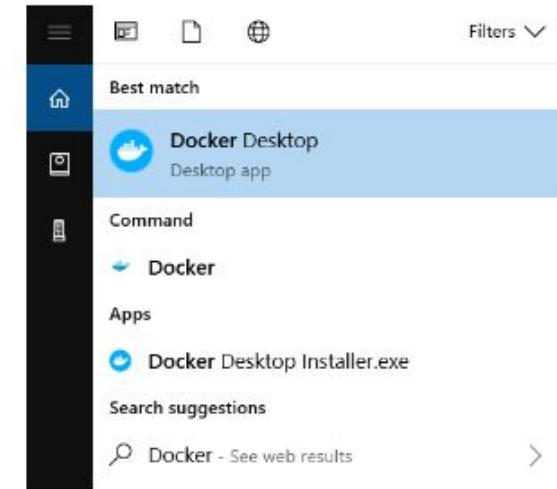
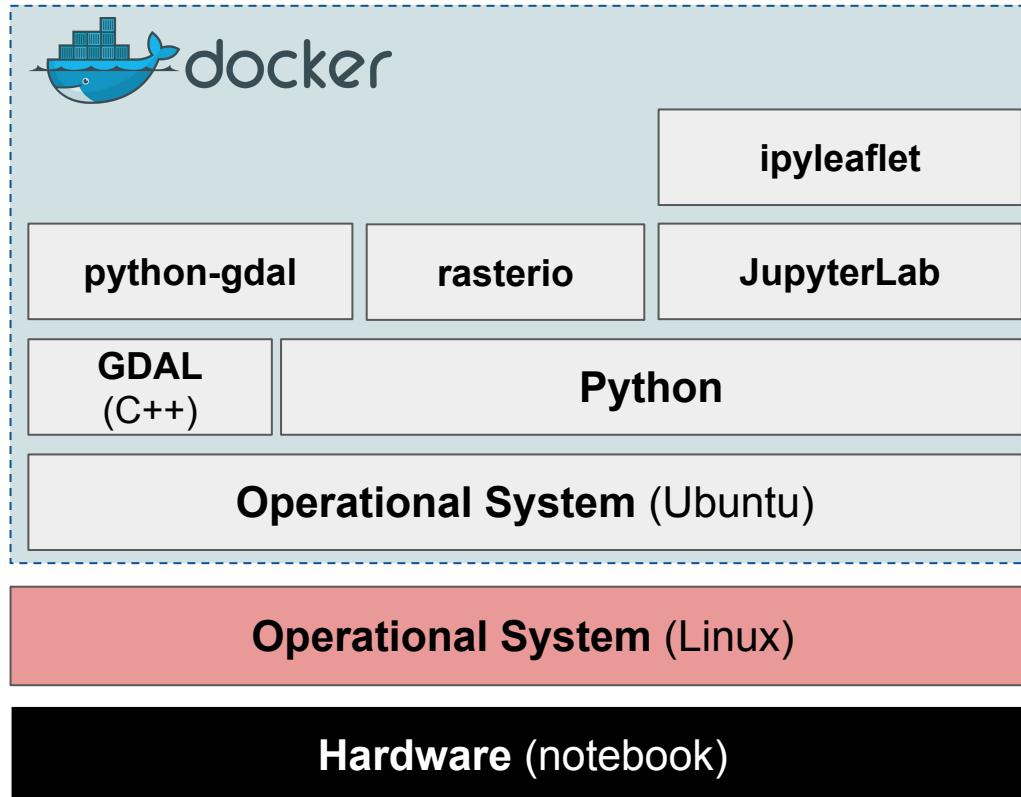
VirtualBox, Docker, Conda...



<https://docs.conda.io/en/latest/>

<https://anaconda.org/conda-forge/conda>

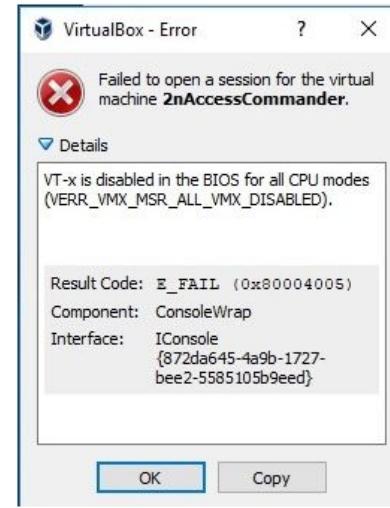
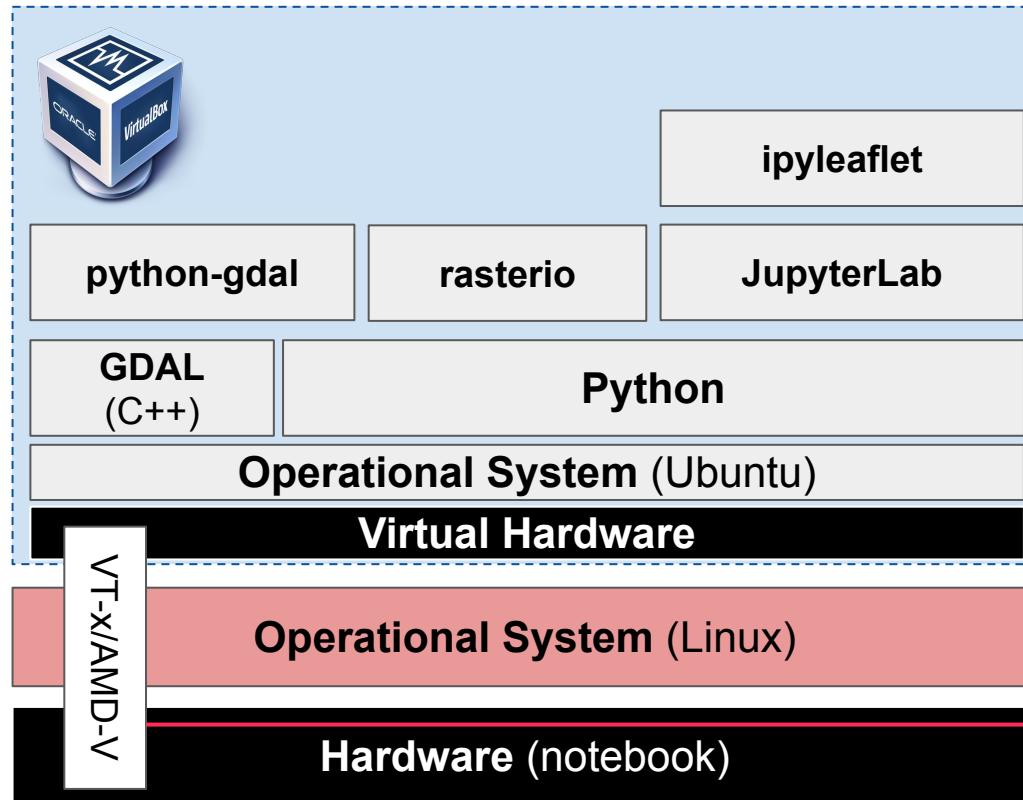
VirtualBox, Docker, Conda...



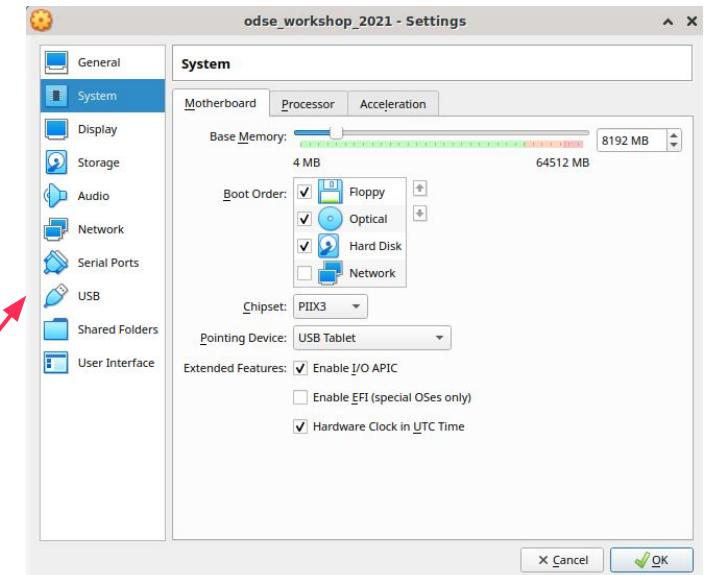
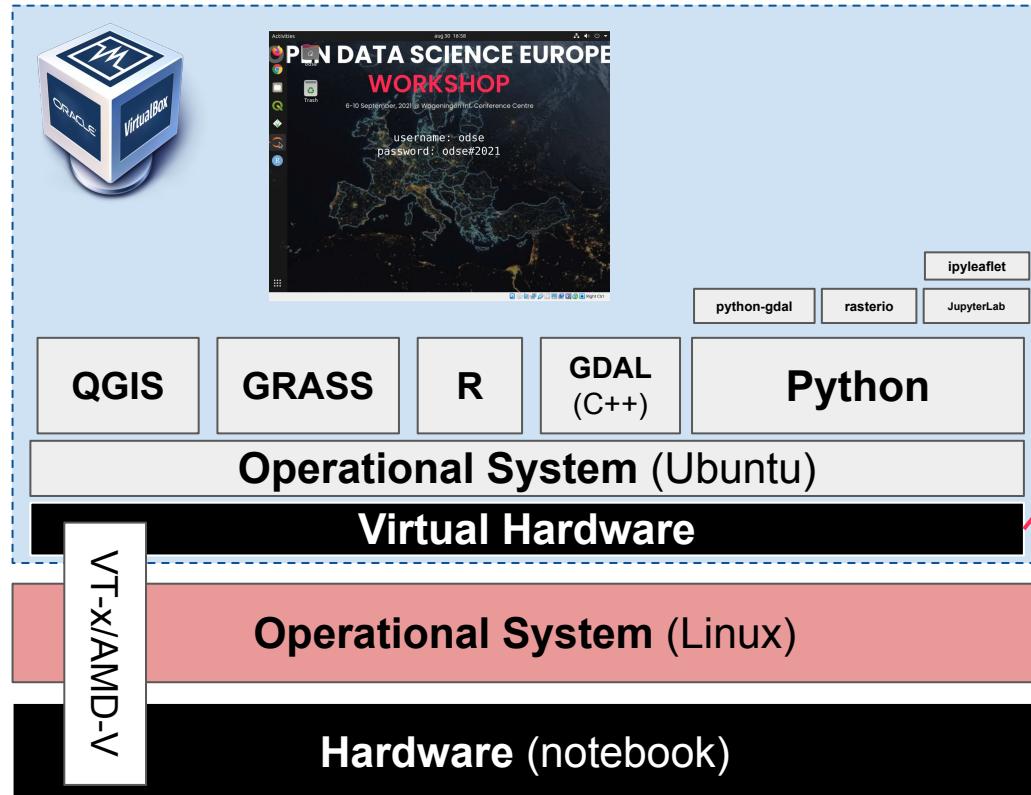
<https://docs.docker.com/desktop/windows/install/>

<https://hub.docker.com/repositories>

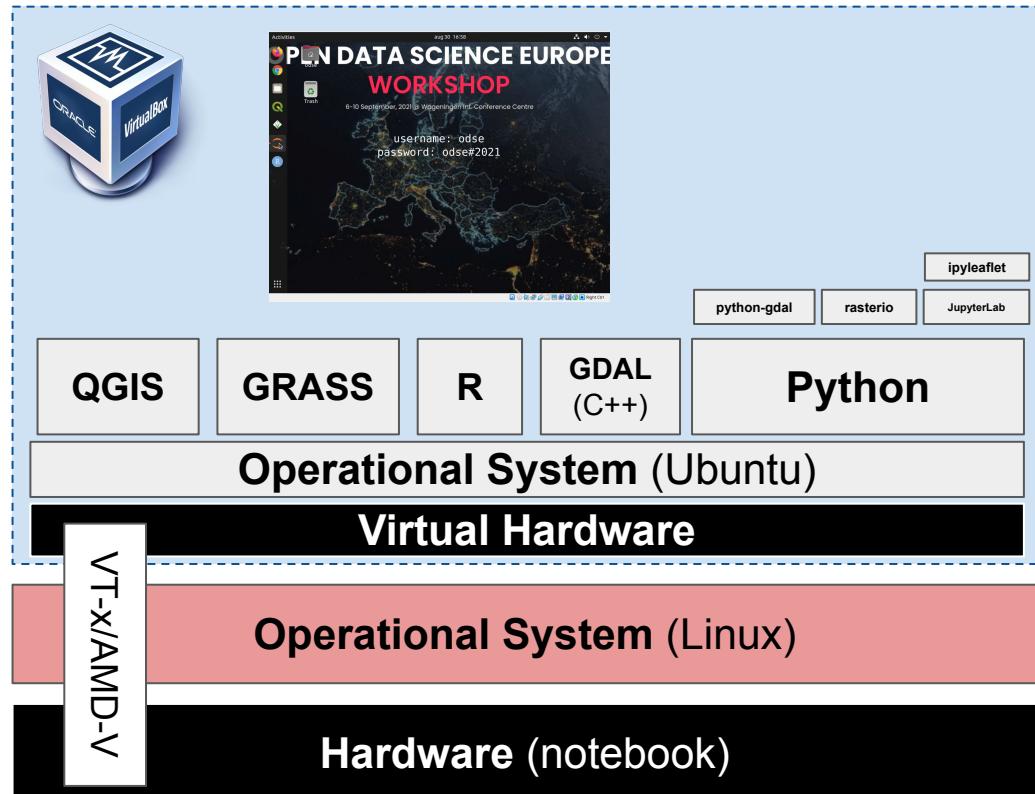
VirtualBox, Docker, Conda...



VirtualBox, Docker, Conda...



VirtualBox, Docker, Conda...



It is a machine inside a machine...



17 GB



VirtualBox, Docker, Conda...



- Easy way to share a complete environment, including all the softwares (GUI) and libraries
- A virtual machine is always slower than a real machine
- Adequate for training sessions

VirtualBox, Docker, Conda...



- Easy way to share a complete environment, including all the softwares (GUI) and libraries
- A virtual machine in general is slower than a real machine
- Adequate for training sessions

For production work



CONDA®



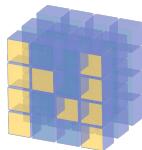
<https://hub.docker.com/r/opengeohub>

https://gitlab.com/geoharmonizer_inea/eumap/-/blob/master/conda_env.yml

What Python libraries will we use?

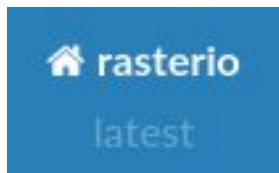


EUMAP
LIBRARY



NumPy

matplotlib



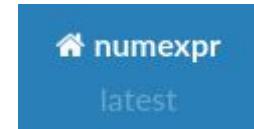
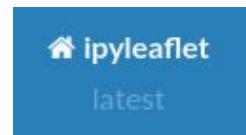
TensorFlow



Numba



GeoPandas



Bottleneck

...

OpenDataScience

What is Open Data Science Europe?

The Open Data Science Europe data portal / viewer aims at serving decision-ready layers such land cover, air quality and pollution, potential natural vegetation and similar.

What is Open Data Science Europe?

The Open Data Science Europe data portal / viewer aims at serving decision-ready layers such land cover, air quality and pollution, potential natural vegetation and similar.

Decision-ready layers

Spatiotemporal Machine Learning Algorithms

Data Harmonization

Open Source software

Publicly available data



What is Open Data Science Europe?

The Open Data Science Europe data portal / viewer aims at serving decision-ready layers such land cover, air quality and pollution, potential natural vegetation and similar.

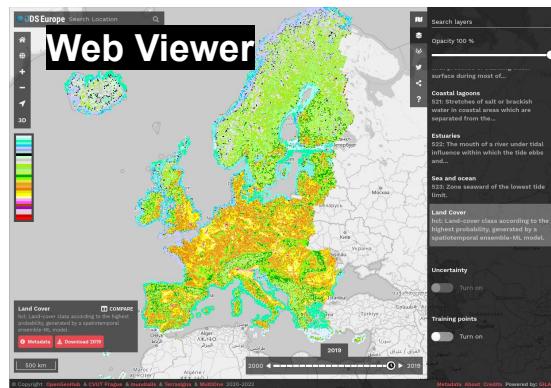
Decision-ready layers

Spatiotemporal Machine Learning Algorithms

Data Harmonization

Open Source software

Publicly available data



Metadata Catalog



S3 Service



Under construction

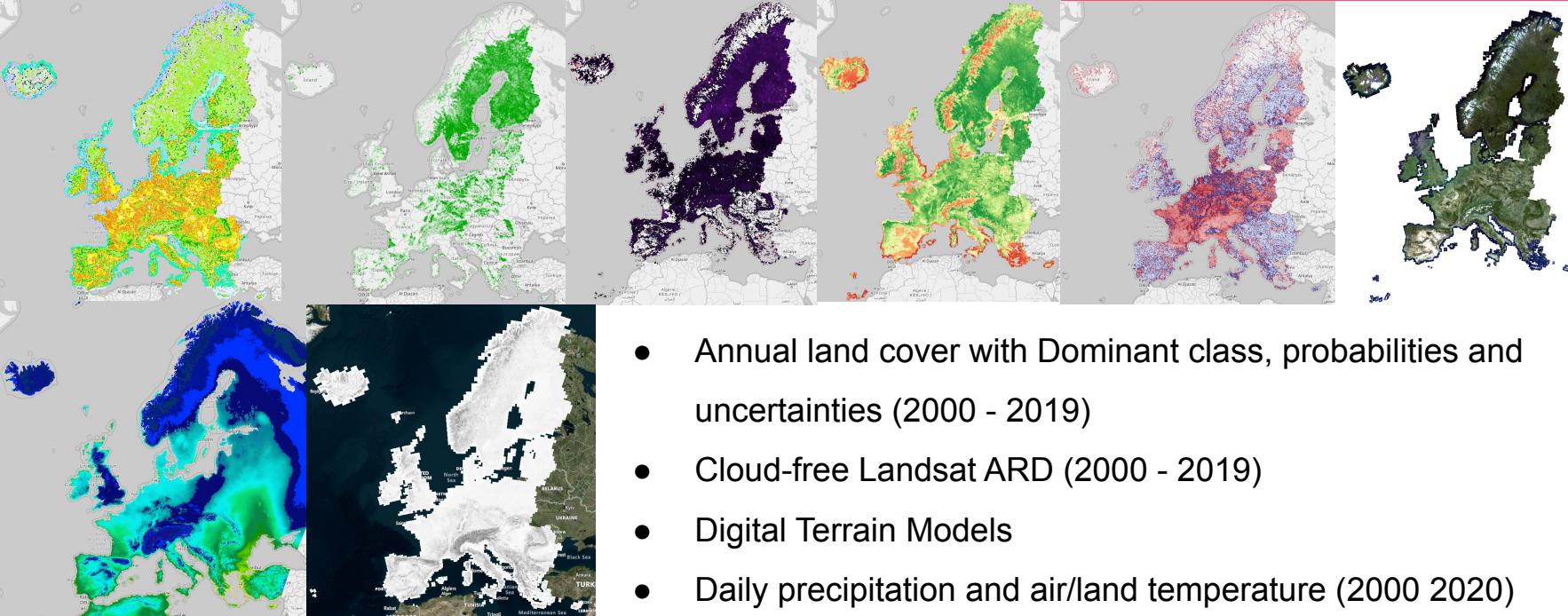


https://gitlab.com/geoharmonizer_inea

...

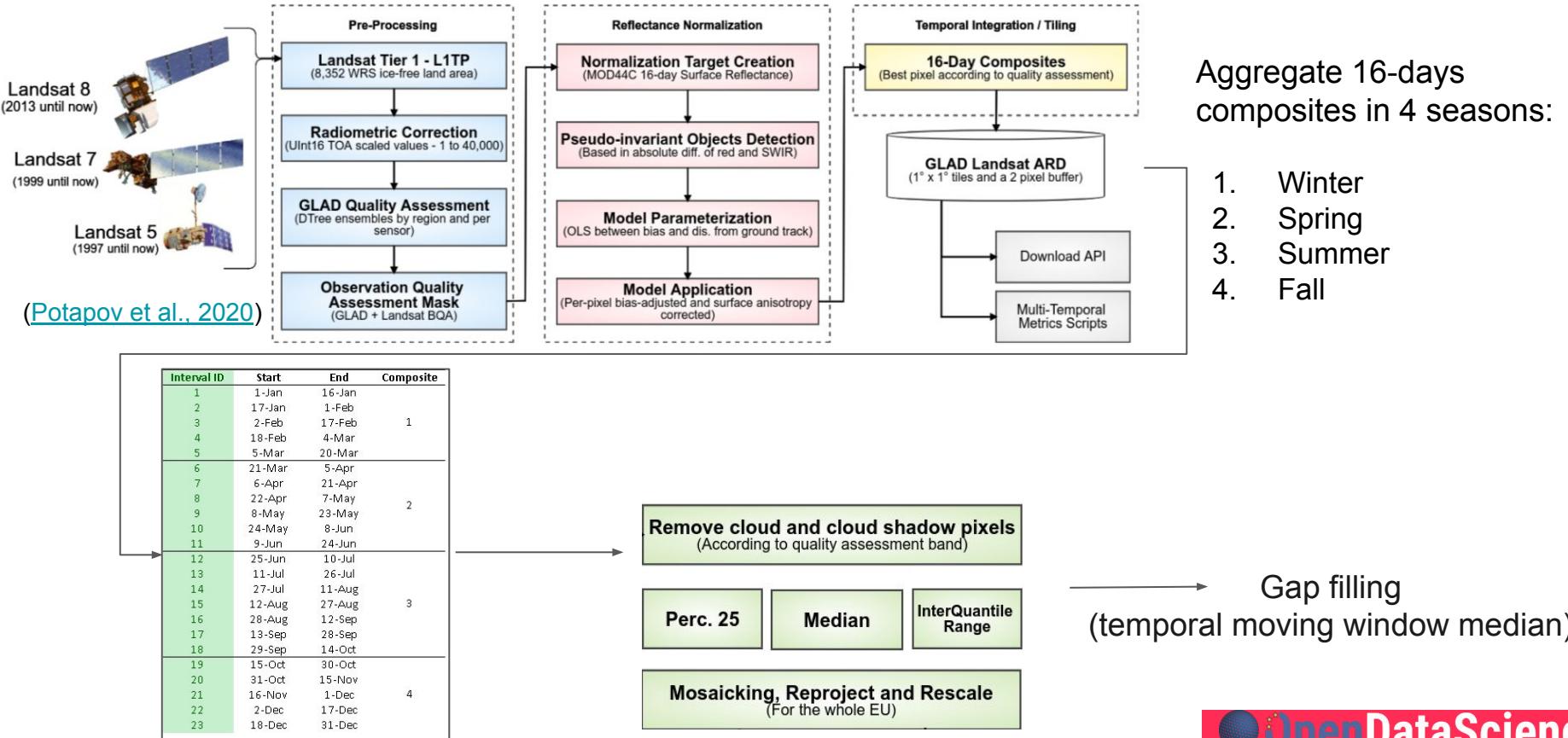
OpenDataScience

Which datasets are available?



Under preparation: Actual and potential natural vegetation, new version for the land cover maps, burned areas,

Cloud-free Landsat ARD (2000 - 2019)

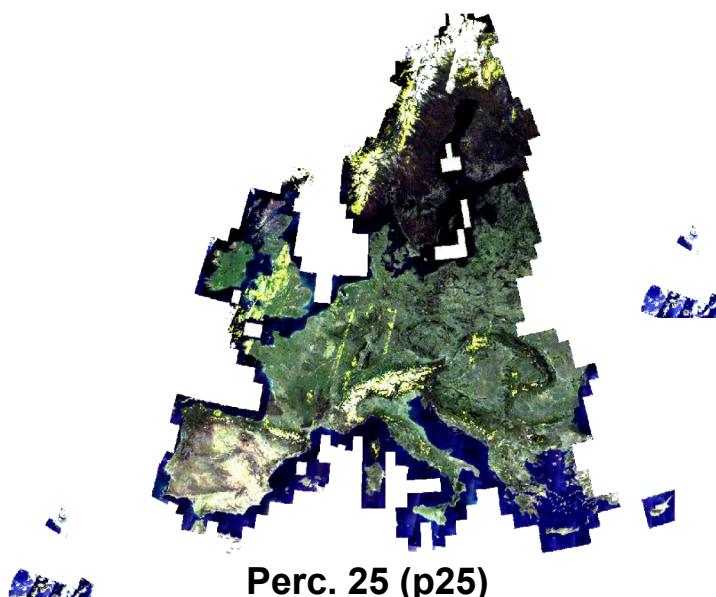


Cloud-free Landsat ARD (2000 - 2019)

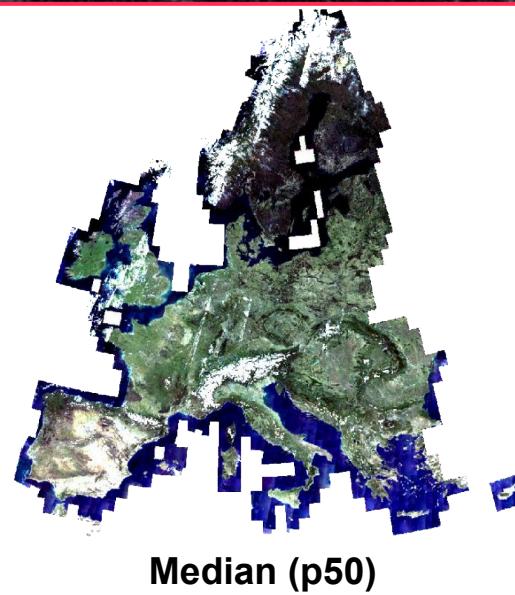
Red, Green, Blue

2018-03-21 to 2018-06-24 (spring)

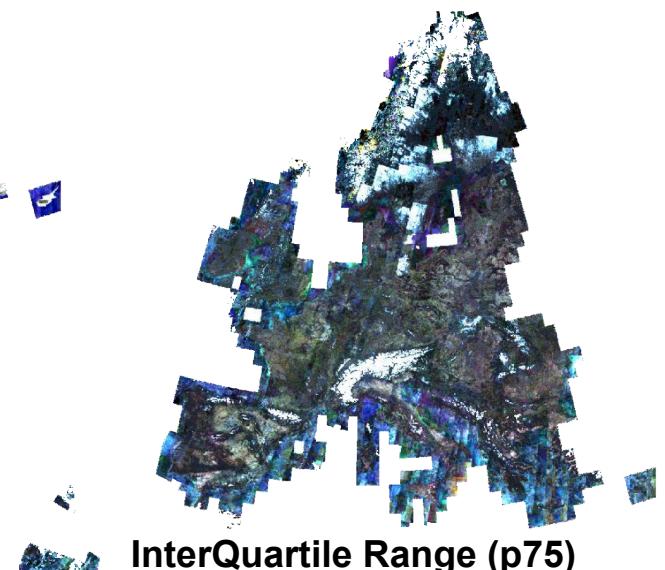
188,000 x 151,000 pixels



Perc. 25 (p25)

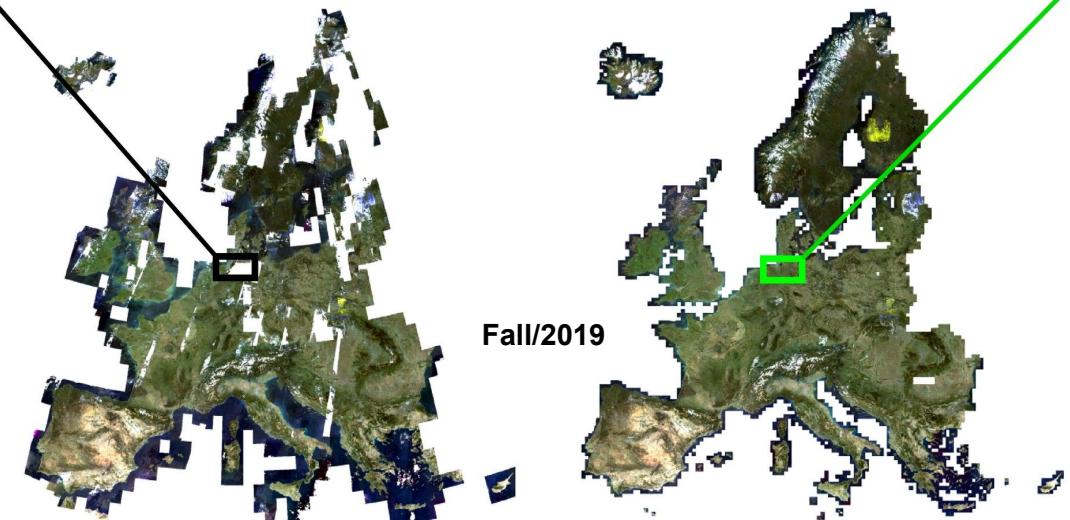
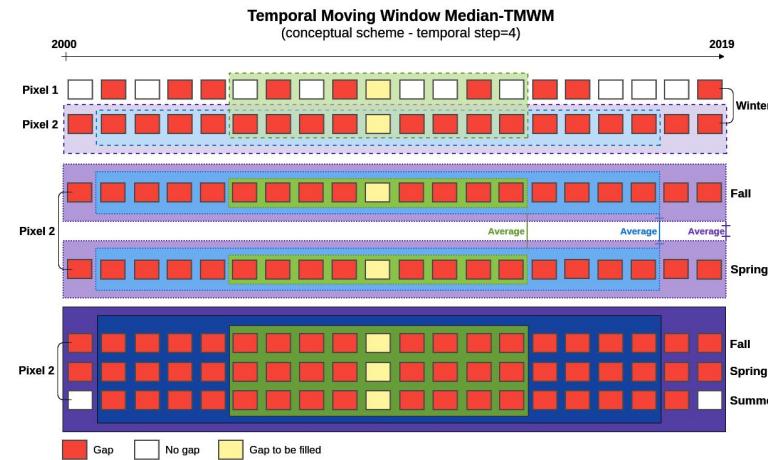


Median (p50)



InterQuartile Range (p75)

Cloud-free Landsat ARD (2000 - 2019)

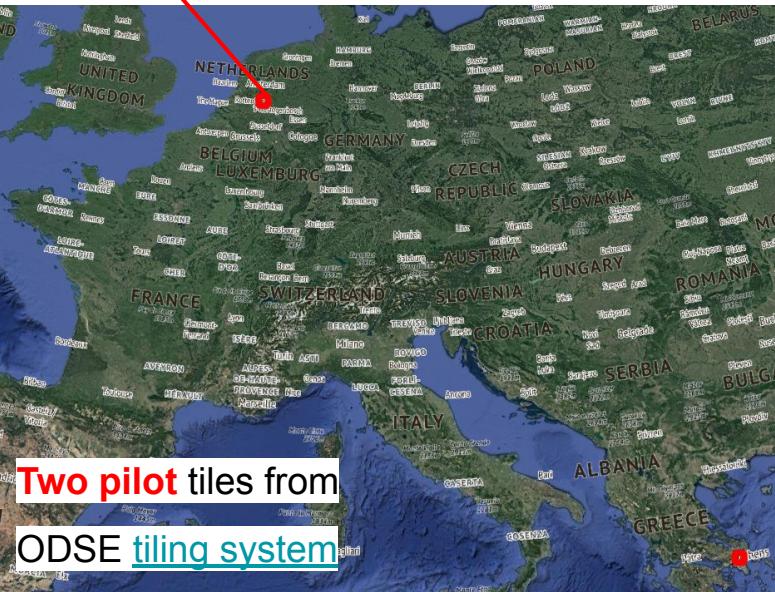


Total amount of yearly images: 84
(7 bands x 3 quantiles x 4 seasons)

Python training session datasets



30x30 km



Two pilot tiles from
[ODSE tiling system](#)

A screenshot of a file browser window titled "ODSE_workdir" with a dropdown menu set to "tile_5606". The left pane shows four folder icons: "adm", "dtm", "lcv" (which has a red arrow pointing to it), and "veg". The right pane lists 15 files under "Name", all of which are "lcv_blue_landsat.glad.ard_p25_30m_0..0cm_2000.06.25..2000.09.12_eumap_epsg3035_v1.1.tif" files with various sizes ranging from 486,4 kB to 414,7 kB.

| Name | Size |
|---|----------|
| lcv_blue_landsat.glad.ard_p25_30m_0..0cm_2000.06.25..2000.09.12_eumap_epsg3035_v1.1.tif | 486,4 kB |
| lcv_blue_landsat.glad.ard_p25_30m_0..0cm_2000.09.13..2000.12.01_eumap_epsg3035_v1.1.tif | 480,8 kB |
| lcv_blue_landsat.glad.ard_p25_30m_0..0cm_2000.12.02..2001.03.20_eumap_epsg3035_v1.1.tif | 494,0 kB |
| lcv_blue_landsat.glad.ard_p25_30m_0..0cm_2001.03.21..2001.06.24_eumap_epsg3035_v1.1.tif | 489,1 kB |
| lcv_blue_landsat.glad.ard_p25_30m_0..0cm_2001.06.25..2001.09.12_eumap_epsg3035_v1.1.tif | 485,7 kB |
| lcv_blue_landsat.glad.ard_p25_30m_0..0cm_2001.09.13..2001.12.01_eumap_epsg3035_v1.1.tif | 494,8 kB |
| lcv_blue_landsat.glad.ard_p25_30m_0..0cm_2001.12.02..2002.03.20_eumap_epsg3035_v1.1.tif | 481,2 kB |
| lcv_blue_landsat.glad.ard_p25_30m_0..0cm_2002.03.21..2002.06.24_eumap_epsg3035_v1.1.tif | 484,2 kB |
| lcv_blue_landsat.glad.ard_p25_30m_0..0cm_2002.06.25..2002.09.12_eumap_epsg3035_v1.1.tif | 497,8 kB |
| lcv_blue_landsat.glad.ard_p25_30m_0..0cm_2002.09.13..2002.12.01_eumap_epsg3035_v1.1.tif | 484,4 kB |
| lcv_blue_landsat.glad.ard_p25_30m_0..0cm_2002.12.02..2003.03.20_eumap_epsg3035_v1.1.tif | 414,7 kB |

Python training sessions



Day 1 Sept 6, 2021

9:00

Introduction to ODSE datasets in Python

11:00

Introduction to spatial and spatiotemporal data in Python

13:00

Spatiotemporal machine learning in Python pt. 1

15:30

Spatiotemporal machine learning in Python pt. 2



**Leandro
Parente**



**Chris van
Diemen**



Day 2 Sept 7, 2021

9:00

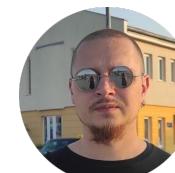
Working with harmonized LUCAS datasets

11:00

High performance computing in Python

13:30

Working with Cloud-Optimized GeoTIFFs in Python



**Luka
Antonić**



**Martin
Landa**

How access timeless and temporal datasets?

https://gitlab.com/geoharmonizer_inea/odse-workshop-2021

JupyterLab

File Edit View Run Kernel Tabs Settings Help

01_INTRODUCTION_TO_ODSE_DATASE

Introduction to ODSE datasets in Python

Raster Data

Raster Summary

Timeless Data

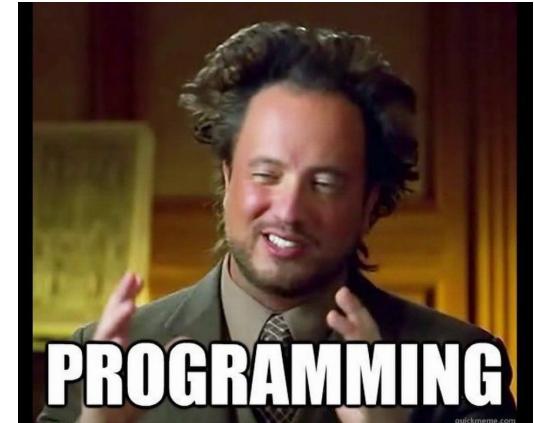
Temporal Data

01_introduction_to_odse_da

Python 3 (ipykernel)

Introduction to ODSE datasets in Python

In this training session you will learn how to access datasets produced within the scope [Open Data Science Europe \(ODSE\)](#), a initiative co-financed by the European Union under the Grant Agreement Connecting Europe Facility (CEF) Telecom project [2018-EU-IA-0095](#), which has the overall objective of producing and organizing various thematic geospatial data layers to support cross-border services over the entire continental Europe. The datasets produced/organized by ODSE are available through a [WebGIS viewer](#), a [metadata catalog](#) and [Cloud-Optimized GeoTIFF \(COG\) S3 service](#).



Installation



```
sudo pip install -e
```

```
'git+https://gitlab.com/geoharmonizer_inea/eumap.git#egg=eumap[full]'
```



```
sudo pip install tensorflow
```