

# Addressing emerging training needs in Earth Observation with Jupyter notebooks

Julia Wagemann<sup>1</sup>, Simone Mantovani<sup>2</sup>, Federico Fierli<sup>3</sup>,  
<sup>1</sup> Julia Wagemann Consulting, <sup>2</sup> MEEEO s.r.l., <sup>3</sup> EUMETSAT

*Living Planet Symposium | 26 May 2022*

Wagemann, J., Fierli, F., Mantovani, S., Siemen, S., Seeger, B. and J. Bendix (2022): Using Jupyter Notebooks for Earth Observation data education. *Remote Sensing*. (under review)



## CHALLENGES

A growing gap between the amount of EO data produced every day and the ability of users to find, access and process the data

Growing data volumes, data discovery and limited processing capacity are major challenges users face, which hinder data uptake and use<sup>1</sup>

## TRAINING NEEDS

To teach EO data users about existing and upcoming products and to teach them how to access and work with these data products

## EO TRAINING PROVIDER

EO data providers

vs.

Higher education system

<sup>1</sup> Wagemann et al. (2021): Users of open Big Earth data – An analysis of the current state. Computers and Geosciences.

Aim to increase use and overall uptake of EUMETSAT and Copernicus data

## Target group and audience



### Highly diverse group of EO practitioners

- Can be **'expert' users** who would like to learn more about upcoming missions and data
- Also **users who just started using EO data products** for a specific application
- Coming from **different geographical regions**
- **Access to a widely differing range of resources** e.g. own computers, high performance clusters, cloud.



## Training types and needs

### Different training types offered

- **Short courses**: light-weight introduction to a specific topic, usually 1 to 1.5 hours long
- **Thematic expert workshops / training schools**: multiple days, aimed at experts and graduate students

### Training has to

- cater for **different levels of data, thematic and programming literacy**
- Offer a **high degree of flexibility** in terms of how the courses and modules can be taught



# Beyond notebooks – the Project Jupyter ecosystem

copernicus.eumetsat.int

Version control, hosting  
and sharing



GitLab

Static rendering and  
sharing



Jupyter  
nbviewer

Jupyter

No-setup, customizable  
Jupyter notebook  
environment

colab



binder

kaggle

Group and multi-user  
access to notebooks on  
a remote server





# On which side are you? Like or dislike Jupyter notebooks?

copernicus.eumetsat.int



*Perkel (2018, 2021):*

Within 10 years, became the **de-facto standard** for data exploration, analysis and training.

In 2021, there are **more than 10 million Jupyter notebooks** available on Github

Interactive web application that combines code, computational output, explanatory text and multimedia resources



*Grus, J. (2018):*

Cells can be executed out of order and **fosters poor coding practices**

*Pimentel et al. (2019):*  
1 out of 4 notebooks could be re-executed and **only 4% produced the same results**

<sup>1</sup> Perkel (2018): Why Jupyter is data scientists' computation notebook of choice. Nature.

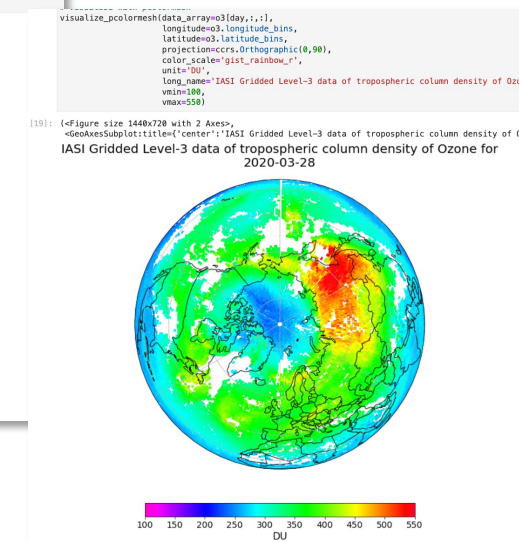
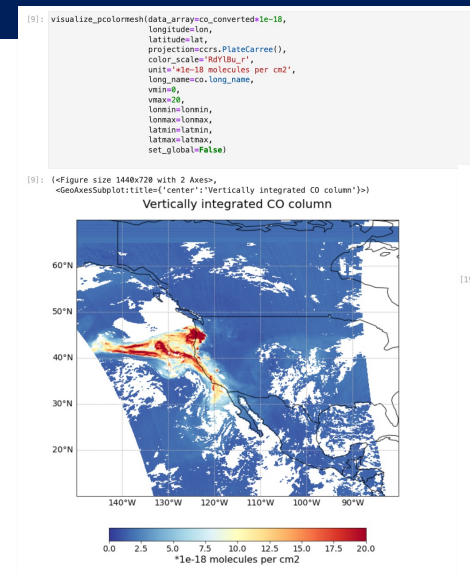
<sup>2</sup> Perkel (2021): Ten computer codes that transformed science. Nature.

<sup>3</sup> Pimentel et al. (2019): A Large-Scale Study About Quality and Reproducibility of Jupyter Notebooks. IEEE.

<sup>4</sup> Grus, J. (2018): I don't like Notebooks. JupyterCon 2018

## Jupyter-based training course on open satellite-, model- and ground-based data on atmospheric composition

- Provide a **general overview of different satellite-, model- and ground-based data** on atmospheric composition
- Provide **code examples and step-by-step guides** on how to load, process and visualize these data
- Provide **examples of thematic application areas**, e.g. fire monitoring, air quality, dust, or stratospheric ozone



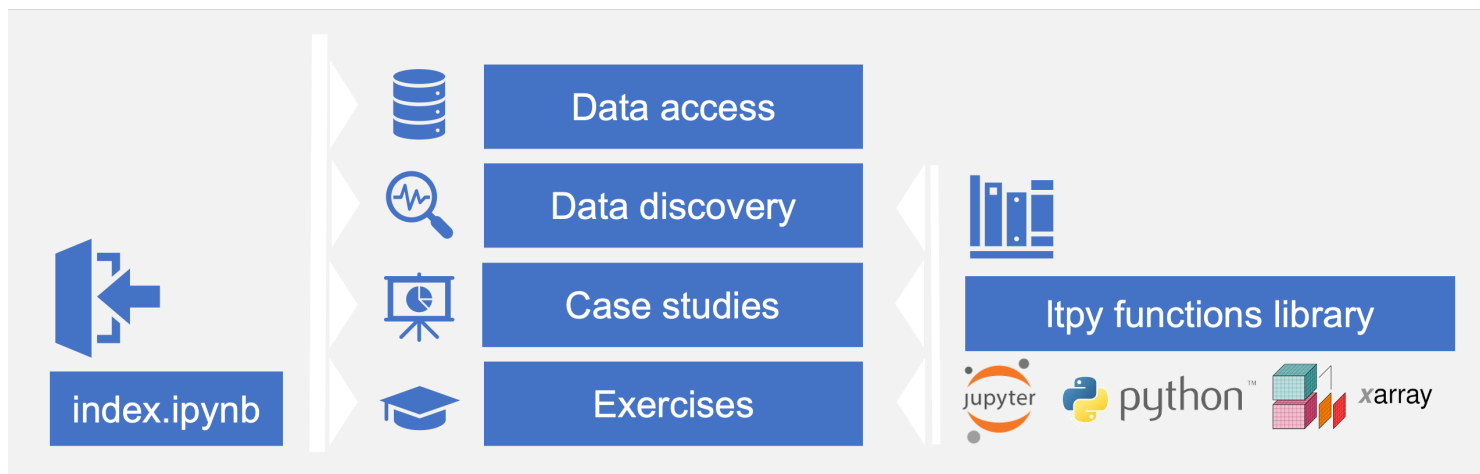
Main course

Thematic module on dust



# Example: LTPy main course

LTPy main course is aligned with a typical data analysis workflow



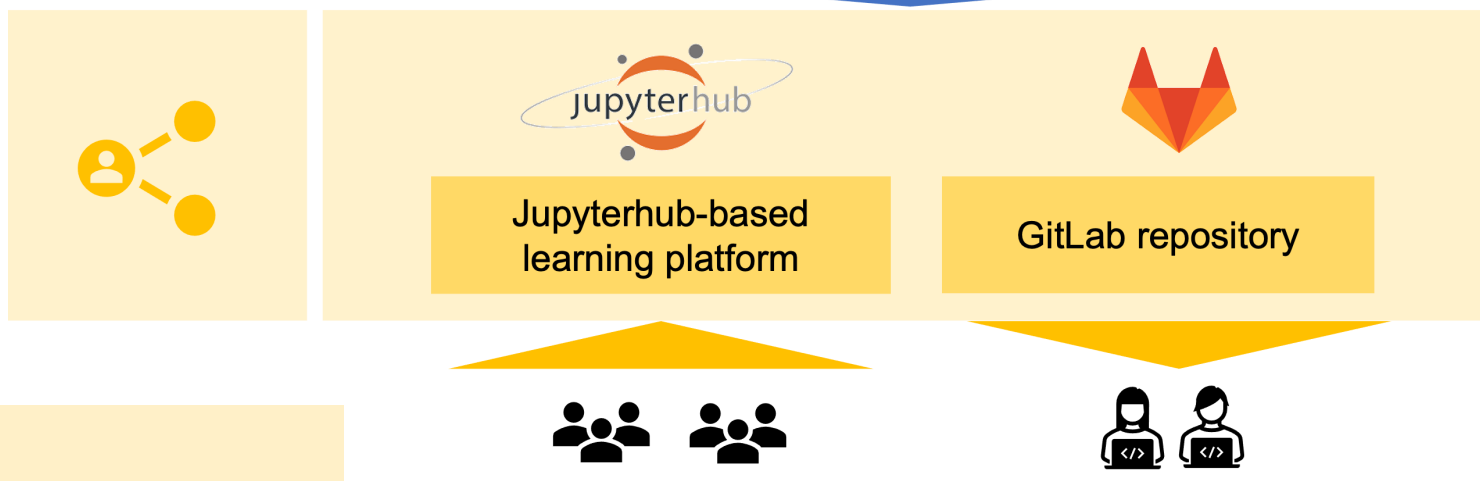
Data from **6 different satellites** and **5 different model-based products**



Over **50 notebooks** related to (i) data access, (ii) data discovery, (iii) case studies and (iv) exercises



A **collection of 14 reusable functions** for effective visualization and data handling



<https://ltpy.adamplatform.eu>

<https://gitlab.eumetsat.int/eumetlab/atmosphere/atmosphere>





# Example: LTPy thematic module on dust monitoring and forecasting

copernicus.eumetsat.int

Self-contained collection of notebooks related to dust monitoring and forecasting



Data from 6 satellites, 3 model-based products and 3 ground-based observations



18 notebooks related divided in two sections: (i) data discovery and (ii) practical assignments



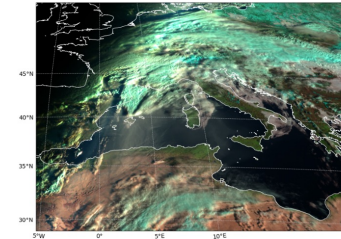
Accessible in form of a Jupyterbook and a dedicated Jupyterhub instance

<https://dust.trainhub.eumetsat.int>

<https://dust.ltpy.adamplatform.eu>

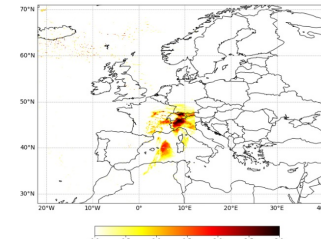
## SATELLITE-BASED DATA (selection)

Natural color composite of southern Europe, recorded by MSG at 2021-02-06 07:15



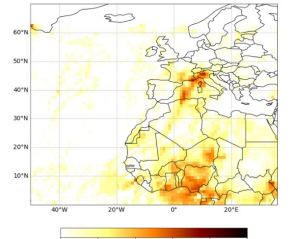
Meteosat Second Generation Level 1B RGB composite

Aerosol index from 388 and 354 nm 2021-02-06T00:00:00.000000000



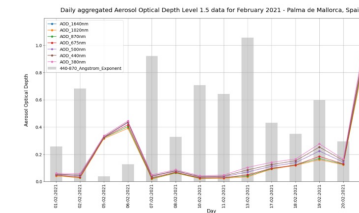
Sentinel-5P TROPOMI Aerosol Index Level 2 product

Absorbing aerosol index averaged for each grid cell 2021-02-05



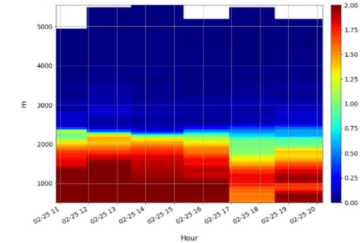
Metop-ABC GOME-2 Absorbing Aerosol Index Level 2 product

## GROUND-BASED OBSERVATIONS

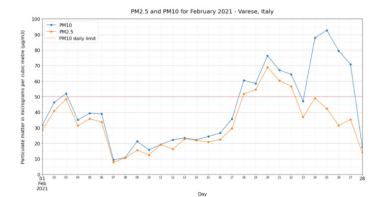


Aerosol Robotic NETwork (AERONET)

aerosol backscatter coefficient - Ispra, Italy on 25 February 2021



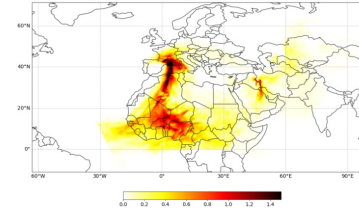
EARLINET Lidar backscatter profiles



European Environment Agency (EEA) Air Quality Data

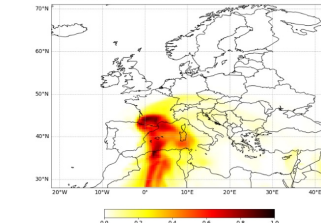
## MODEL-BASED DATA

Aerosol optical depth 2021-02-06T06:00:00.000000000



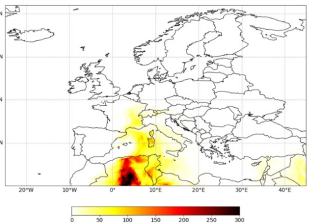
NMMB/MONARCH regional model - Aerosol Optical Depth

Dust Aerosol Optical Depth at 550nm 2021-02-06T06:00:00.000000000



Copernicus Atmosphere Monitoring Service (CAMS) global forecast - Dust Aerosol Optical Depth

mass\_concentration\_of\_dust\_in\_air 2021-02-06T15:00:00.000000000



Copernicus Atmosphere Monitoring Service (CAMS) European air quality forecasts - Dust concentration





# Jupyter notebook examples for EO training @EUMETSAT

copernicus.eumetsat.int

## Trainhub – EUMETSAT’s Jupyter notebook portal

<https://trainhub.eumetsat.int/>

Different domains

Login to a Jupyterhub-based processing platform

**EUMETSAT** EUROPEAN WEATHER CLOUD  
CLOUD-COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

ATMOSPHERE MLEO WEATHER SOUNDING DUST FIRE

Log In

**Advanced Search**

Search

Back to Default Search

**Filters**

**Domain**

- Atmosphere (38)

**Category**

- Data Discovery (15)
- Case Study (16)
- Exercise (7)

**Variable**

- Atmosphere Mole Content Of Nitrogen Dioxide (9)
- Absorbing Aerosol Index (3)
- Atmosphere Optical Thickness Due To Ambient Aerosol Particles (6)
- Mole Concentration Of Ammonia In Air (1)
- Atmosphere Mole Content Of Carbon Monoxide (6)
- Radiance (4)
- Fire Radiative Power (7)
- Fire Weather Index (3)
- Atmosphere Moles Of Formaldehyde (1)

**Get Started**

- Discover Metop-A/B GOME-2 - Tro...  
  
This notebook is the first of two 'da...
- Discover Metop-A/B GOME-2 - Tro...  
  
This notebook is the second of two...
- Discover Metop-A/B GOME-2 - Tro...  
  
This notebook introduces you to m...
- Discover Metop-A/B/C GOME-2 - ...  
  
This notebook introduces you to M...
- Discover PMAp - Aerosol Optical D...  
  
This notebooks introduces you to i...
- Discover Metop-A/B IASI - Ammon...  
  
This notebooks lets you discover L...

Search functionality based on notebook metadata

Overview of training notebooks available

Static rendering of the notebook

Discover Metop-A/B GOME-2 - Tropospheric Nitrogen Dioxide L3

Publisher: MEOO Created 18/01/2021

This notebook introduces you to monthly aggregated gridded (Level 3) Metop-A/B GOME-2 data. The notebook showcases the visualisation of tropospheric Nitrogen Dioxide (NO<sub>2</sub>).

#Atmosphere #Data Discovery #Atmosphere Mole Content Of Nitrogen Dioxide #Metop-Ab #Gome-2 #L3

Code Output

Logo EU Copernicus EUMETSAT

<< Index

<< 212 - Metop-A/B GOME-2 - Tropospheric NO<sub>2</sub> - Level 2 - Pre-processing

214 - Metop-A/B/C GOME-2 - Absorbing Aerosol Index - Level 3 >>

**20 - DATA DISCOVERY**

**SEE ALSO** \*\*30 - CASE STUDIES\*\* - [321 - Metop-A/B GOME-2 - Tropospheric NO<sub>2</sub> map and time-series - Level 3] [../30\_case\_studies/321\_air\_pollution\_map\_time-series\_Metop-AB\_GOME-2\_NO2Tropo\_L3.ipynb] - [323 - Air pollution - Tropospheric NO<sub>2</sub> anomaly map - Metop-A/B GOME-2 - Level 2] [../30\_case\_studies/323\_air\_pollution\_map\_europe\_2020\_Metop-AB\_GOME-2\_NO2Tropo\_L2.ipynb] - [324 - Air pollution - Tropospheric NO<sub>2</sub> time-series analysis - Metop-A/B GOME-2 - Level 2] [../30\_case\_studies/324\_air\_pollution\_time-series\_europe\_2020\_Metop-AB\_GOME-2\_NO2Tropo\_L2.ipynb]

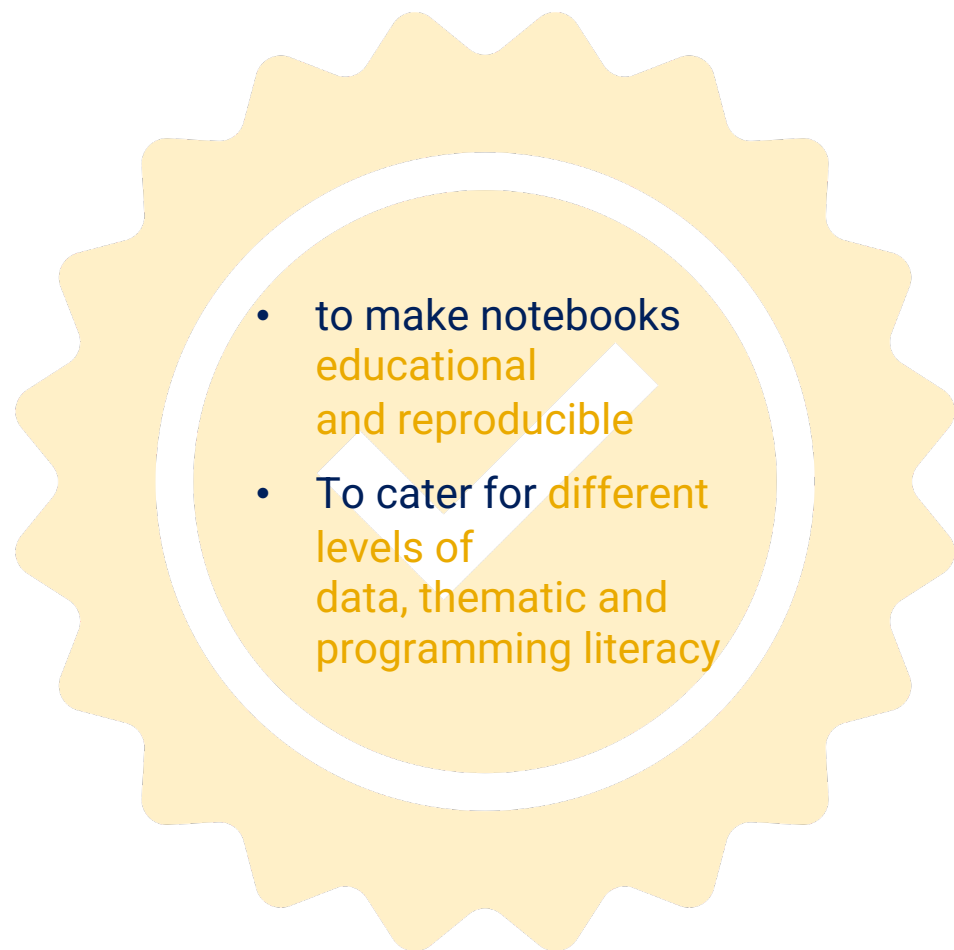
**2.1.3 Metop-A/B GOME-2 - Tropospheric NO<sub>2</sub> - Level 3**

Metop-A/B GOME-2 data records are monthly aggregated products (Level 3), regridded onto a regular latitude longitude grid. Data records are disseminated in the data format netCDF.

Data records are processed for three parameters:

- tropospheric ozone column,
- total and tropospheric nitrogen dioxide column, and
- total water vapour column.

This notebook guides you through how the data can be loaded and gives basic plotting examples.



Systematic use of recommended Python libraries for handling EO data



Modularisation of content, e.g. to outsource functions



Use of instructional design patterns, such as navigation pane or a table of contents



Provide guided learning pathways, e.g. follow a naming nomenclature to order notebooks in an expected sequence or include an index notebook.



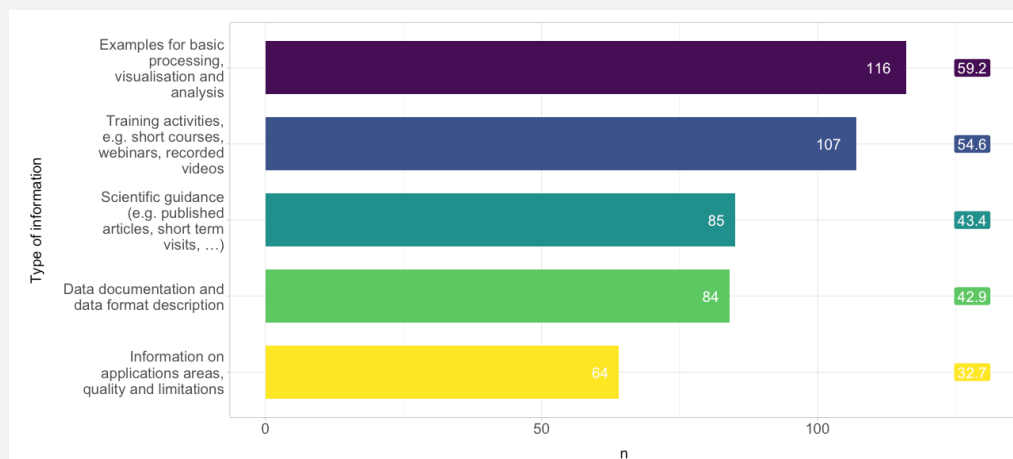
# LTPy: Feedback from training participants

Since 2019,  
1085 learners  
trained in 16  
training events

Type of training	Number of events	Number of participants
Training schools	6	553
Thematic expert workshops	2	130
Short courses	8	402
<b>Total</b>	<b>16</b>	<b>1085</b>

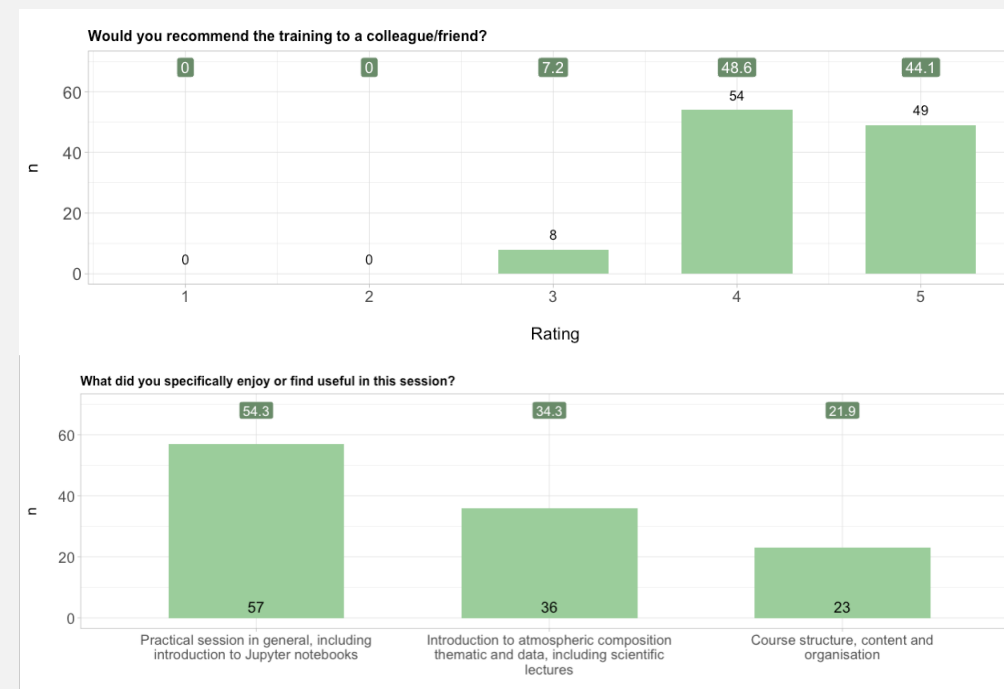
## Needs

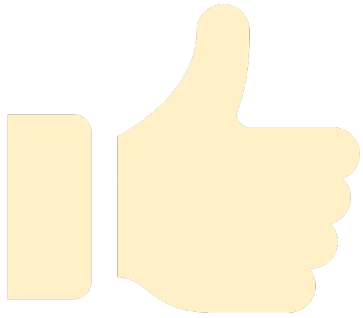
Training in line with user needs. More than half consider examples for basic processing, visualisation and analysis and training activities as helpful



## Training feedback

Feedback overall very positive. Practical training part with Jupyter notebooks is particularly useful.



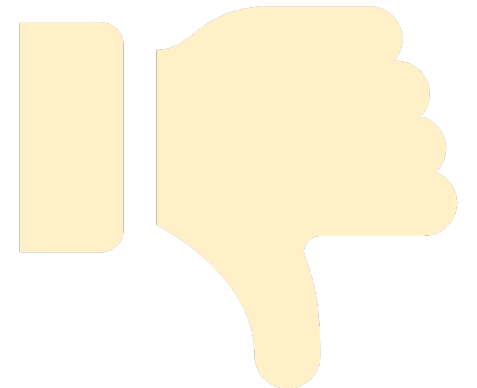


## Advantages

- **Notebooks remove elements of fear** – people are more familiar with a browser than a command line
- **Notebooks allow to break up the code into step-by-step workflows** and also to describe the workflow in richer ways
- **The wider ecosystem**, including sharing and hosting notebooks, make notebooks well suited as a teaching tool
- Useful as training material **in an instructor-led setting and for self-paced learning alike**

- Accessing data and using Python in the browser is a **paradigm-change**
- Compromise between having a **pre-installed Python environment setup vs. learning how to setup the environment**
- Compromise between **downloading data vs. making data already available on the training platform** (fragmentation of data landscape a barrier)
- **Maintenance of notebooks** – especially with many contributors.

## Challenges



- In less than a decade, **Jupyter notebooks became the de-facto standard to conduct data science**, including analysing Earth observation data.
- Jupyter notebooks and the Jupyter ecosystem, such as JupyterLab, **will play a significant role in how users of EO data will access and process data in the future**
- **Using computational notebooks for training/education brings in a set of additional requirements** that require the integration of didactical concepts, instructional design patterns and best practices for coding



Next step

Defining, sharing and implementing quality standards and best practices on how to make notebooks effective, reproducible and educational



# Thank you!

Wagemann, J., Fierli, F., Mantovani, S., Siemen, S., Seeger, B. and J. Bendix (2022): Using Jupyter Notebooks for Earth Observation data education. *Remote Sensing*. (under review)