Using Jupyter Notebooks in Earth Observation Education @EUMETSAT

Julia Wagemann, Federico Fierli, Simone Mantovani
CEOS WGISS-53 Meeting | 22 March 2022
• Earth Observation training @EUMETSAT has to cater for different levels of data, thematic and programming literacy

• EUMETSAT has developed a series of Jupyter notebook trainings on different application areas, e.g. atmospheric composition
  • Learning Tool for Python (LTPy) on atmospheric composition consists of over 70 notebooks related to data access, analysis, data discovery, case studies, exercises and thematic modules
  • TrainHub – EUMETSAT’s Jupyter notebook portal
  • Jupyterbook on ‘Dust Aerosol detection, monitoring and forecasting’

• 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

• Needed: Defining, sharing and implementing quality standards and best practices on how to make notebooks effective, reproducible and educational
  • Systematic use of recommended Python libraries
  • Modularisation of content, e.g. outsource functions
  • Use of instructional design patterns
  • Follow a naming nomenclature to order notebooks in a specific sequence
Earth Observation training @EUMETSAT

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

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
Jupyter notebooks – an introduction

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

- Became a popular tool for data exploration, analysis and training. In 2021, there are more than 10 million Jupyter notebooks available on Github
Beyond notebooks – the Project Jupyter ecosystem

Hosting of Jupyter notebooks

- GitHub
- GitLab

Static rendering / sharing of notebooks

- jupyter nbviewer

No-setup, customizable, Jupyter Notebooks environment

- colab
- binder

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

- jupyterhub
- lab

Kaggle
Jupyter notebook examples for EO training @EUMETSAT

Trainhub – EUMETSAT’s Jupyter notebook portal

https://trainhub.eumetsat.int/

Different domains

Search functionality based on notebook metadata
Overview of training notebooks available
Static rendering of the notebook

Login to a Jupyterhub-based processing platform

This notebook introduces you to monthly aggregated (level 3) Metop-A/B GOE-2 data. The notebook showcases the visualization of tropospheric Nitrogen Dioxide (NO2).

Search functionality based on notebook metadata
Overview of training notebooks available
Static rendering of the notebook
Jupyter notebook examples for EO training @EUMETSAT

Trainhub – EUMETSAT's Jupyter notebook portal

https://trainhub.eumetsat.int/

No extra database – all information required for the Trainhub portal are stored on notebook level as metadata

"metadata": {
"title": "Discover Sentinel-3 OLCI Level 1 data",
"description": "This notebook introduces you to …",
"author": "Julia Wagemann",
"image": "/img/img_04.png",
"link": "link_to_jupyterhub/notebook.ipynb",
"github": "link_to_github/notebook.ipynb",
"tags": {
"domain": "atmosphere",
"variable": "toa_reflectance",
"satellite": "Sentinel-3",
"sensor": "OLCI",
"level": "L1"}
},
Example: Learning Tool for Python on Atmospheric Composition

Jupyter-based training course on open satellite- and model-based data for atmospheric composition

- Data from 6 different satellites and 5 different model-based products
- Over 70 notebooks related to (i) data access, (ii) data discovery, (iii) case studies, (iv) exercises and (v) thematic modules
- A collection of 14 reusable functions for effective visualization and data handling

https://ltpy.adamplatform.eu
https://gitlab.eumetsat.int/eumetlab/atmosphere
Best practices: A need for ‘educational’ and ‘reproducible’ notebooks

- to make notebooks educational and reproducible
- To cater for different levels of data, thematic and programming literacy

Systematic use of recommended Python libraries for handling EO data

Modularisation of content, e.g. to outsource functions in a separate notebook

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

Follow a naming nomenclature, to order notebooks in an expected sequence. Include an index notebook.
Best practices: Making educational notebooks ‘accessible’

Module on ‘Dust Aerosol Detection, Monitoring and Forecasting’

Jupyterbook

Work in progress: https://dustbook.ltpy.adamplatform.eu/

Dust Aerosol Detection, Monitoring and Forecasting

Why this course?

Dust storms are common meteorological hazards in arid and semi-arid regions. They are usually caused by thunderstorms, or strong pressure gradients associated with cyclones, that increase wind speed over a wide area. Monitoring, forecasting and early warning systems for airborne dust are crucial to evaluate impacts and developing products to guide preparedness, adaptation and mitigation policies.

Interaction of airborne dust with weather and climate:

- **Airborne dust functions in a manner similar to the greenhouse effect:** it absorbs and scatters solar radiation entering Earth's atmosphere, reducing the amount reaching the surface, and absorbs long-wave radiation bouncing back up from the surface, re-emitting it in all directions.
- **Dust particles, especially if coated by pollution, act as condensation nuclei for warm cloud formation** and as efficient ice nuclei agents for cold cloud generation. Modification of the microphysical composition of clouds changes their ability to absorb solar radiation, which indirectly affects the energy reaching the Earth's surface. Dust particles also influence the growth of cloud droplets and ice crystals, thus affecting the amount and location of precipitation.
LTPy: Feedback from training participants

Since 2019, 1085 learners trained in 16 training events

<table>
<thead>
<tr>
<th>Type of training</th>
<th>Number of events</th>
<th>Number of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training schools</td>
<td>6</td>
<td>553</td>
</tr>
<tr>
<td>Thematic expert workshops</td>
<td>2</td>
<td>130</td>
</tr>
<tr>
<td>Short courses</td>
<td>8</td>
<td>402</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>1085</td>
</tr>
</tbody>
</table>

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

Needs

Training in line with user needs. More than half consider examples for basic processing, visualisation and analysis and training activities as helpful.
Lessons learned – Things that work well and don’t work so well

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

Challenges

• 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)
Conclusion & Outlook

• 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.