The Journal of Open Source Education

LTPy - Learning tool for Python on Atmospheric Composition

Julia Wagemann[1], **Sabrina H. Szeto**[2], **Simone Mantovani**[3], and **Federico Fierli**[4]

1 Laboratory for Climatology and Remote Sensing (LCRS), Philipps University Marburg **2** Sabrina Szeto Consulting **3** Meteorological Environmental Earth Observation (MEEO) s.r.l. **4** European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT)

DOI: 10.21105/jose.00172

Software

- Review
- Repository
- Archive

Submitted: 29 January 2021
Published: 08 February 2023

Summary

Learning tool for Python (LTPy) is a Python-based series of notebooks on different open satellite- and model-based data on atmospheric pollution and climate. LTPy features data from six different satellites, including the Copernicus satellites Sentinel-3 and Sentinel-5 as well as the polar orbiting meteorological satellite series Metop and five different model-based product types from the two Copernicus services on Atmosphere Monitoring (CAMS) and Emergency Management (CEMS).

The course has the aim to facilitate the uptake and use of atmospheric composition data as well as showcasing possible application areas. LTPy is based on Jupyter notebooks, which allow for a high-level of interactive learning, as code, text description and visualisation is combined in one place. The structure of the course is aligned with a typical data analysis workflow and includes notebooks on data access, data exploration, case studies and exercises. LTPy consists of more than 50 Jupyter notebooks which are available on a Gitlab repository as well as on a dedicated LTPy JupyerLab platform, where the notebooks can be executed.

Statement of Need

The field of Earth Observations (EO) has experienced a series of disruptive changes. First, the adoption of open data policies in 2008 has led to an exponential growth in data uptake by users (Zhu et al., 2019). Second, technological advancements and new satellite constellations have led to an increase in data volume and variety. One main driver for this is Copernicus, the European Commission's Earth Observation programme. From Copernicus alone, it is estimated that currently every day 16 TBs of data are produced and disseminated (European Commission, 2022). These developments certainly continue, as many new programmes and satellite missions are in the pipeline. The second phase of Copernicus foresees the launch of six high-priority candidate missions till 2027 (Hebden, 2020). The European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) prepares the launch of the generation of its geostationary (Meteosat Third Generation (MTG)) and polar-orbiting (Metop-Second Generation (SG)) satellites (European Organisation for the Exploitation of Meteorological Satellites, 2022).

These disruptive changes lead naturally to an increase in number of users, but at the same time a much larger community of diverse users is interested in understanding and using the data (European Commission - Directorate General for Internal Market Industry, Entrepreneurship and SMES and PwC, 2019; Overpeck et al., 2011). In the past, scientists and researchers have been the main users of EO data, whereas now the data attracts non EO experts alike, coming from varied application domains (Sudmanns et al., 2020; Wagemann et al., 2021). This holds true especially for data on atmospheric composition, as the data is of interest for users working in different fields, such as fire monitoring, air quality, atmospheric composition monitoring and forecasting, modelling, climate assessment and dust monitoring. With the shift from EO expert users to multi-disciplinary non-expert users three main challenges arise: (i) users are unaware of the existence of specific data and data products, (ii) users fail to make use of the data due to diverse data formats and dissemination channels and (iii) users do not know potential application areas.

We aimed to address these challenges by creating a Jupyter-based training course with notebooks on data access, data exploration, case studies and exercises for different types of atmospheric composition data. The notebooks lead learners step-by-step through a typical data analysis workflow, including data loading, processing and visualization. The systematic use of the same Python libraries and a set of pre-defined functions makes the course accessible to learners with basic and more advanced programming skills alike. Additionally, the training course is available through a JupyterLab platform, where the required Python environment and data is already available. Instead of struggling with preparing the environment on their local machines, learners are able to directly start with the content-based training on atmospheric composition data.

Learning Objectives and Instructional Design

At the end of the LTPy course, learners will be able to:

- find satellite observations and model-based data on atmospheric composition that is relevant for their purposes
- load, process and visualize these data, and
- apply these data for specific applications domains, such as monitoring of dust, ozoene or wildfires.

Several design elements support the learner to easily navigate through the training modules and each notebook. Learners are invited to start with the 00_index.ipynb notebook, which introduces the training content and a proposed structure based on the four sections: 10 - Data Access, 20 - Data Exploration, 30 - Case Studies and 40 - Exercises. Each notebook has a navigation pane and a module overview, which allow learners to easily go to individual sections within a notebook and to navigate between notebooks. Further, any module prerequisites or additional content on the same data type are cross-referenced at the beginning of the notebook.

The course is designed for learners who have a basic understanding of geospatial data analysis in a programmatic way and accommodates learners with different levels of Python and coding literacy. An essential part of the LTPy course is a 'functions' library, which is a collection of 14 pre-defined functions that support the learner with data loading, pre-processing and visualization. Learners with basic Python knowledge learn Python by applying these functions, where only keyword arguments have to be provided. Learners with more Python experience can examine the functions in a separate notebook or build their own functions.

Course Content

The course is organised in four main parts:

10 - Data Access

Data access notebooks are in the folder 10_data_access and provide an overview of different data access systems and an example how different atmospheric composition data products can be accessed. The data access systems range from pure download services to cloud-based services where data is accessed in a programmatic way.

20 - Data Exploration

Data discovery notebooks are in the folder 20_data_exploration. These notebooks are organised per data product and aim at 'exploring' the respective dataset, including its file structure and internal organisation. The notebooks provide a step-by-step guide to load, pre-process and visualize each dataset.

30 - Case Studies

Case study notebooks are in the folder 30_case_studies. These notebooks are more advanced and have the aim to provide several atmospheric composition application examples. Case study modules often feature several atmospheric composition data types and may focus on different application areas, such as monitoring of fires, the analysis of stratospheric ozone or the analysis of air pollution. However, a case study can also feature a technical workflow, showcasing e.g. how to generate gridded climate records (Level 3 products) from Level 2 data.

40 - Exercises

Exercise notebooks are in the folder 40_exercises. These modules provide the learner an opportunity to actively practice the learned content. Exercise workbooks contain two types of exercises: (i) coding assignments and (ii) questions. Coding assignments ask to fill in an empty code cell based on the given instructions. Questions ask the learner to reflect on an output and encourage the learner to understand the results of code blocks. Exercise workbooks are based upon data discovery modules and learners are advised to go through the respective data discovery module before.

Teaching Experience

We started the training with LTPy in mid 2019 and since then, the notebooks have been used in more than 16 EUMETSAT training activities (in-person and virtual). Training events range from short courses spanning only several hours up to week-long training schools. By now, more than 1000 learners have been reached. The training modules have been used during in-person and online training events with an instructor, but are also developed in a way that they can be followed in a self-paced manner.

Each event included a feedback survey on the usefulness, clarity and ease of access of the training modules and platform. The general feedback is overall positive. More than

90% would either recommend or highly recommend the training and 40% particularly emphasized the benefits and usefulness of Jupyter notebooks. Suggestions for improvement and specific requests address the need for availability of the training material after the training and the possibility to use the training platform on a continued basis (Wagemann et al., 2022).

Acknowledgements

LTPy has been supported by the Copernicus Programme through EUMETSAT. We acknowledge contributions from Julien Chimot (EUMETSAT), Claudia Vitolo (formerly ECMWF, now ESA) and Miha Razinger (ECMWF).

References

European Commission. (2022). *Copernicus*. https://defence-industry-space.ec.europa.eu/eu-space-policy/copernicus_en

European Commission - Directorate General for Internal Market Industry, Entrepreneurship and SMES and PwC. (2019). *Copernicus Market Report - February 2019* (Vol. 2). https://doi.org/10.2873/011961

European Organisation for the Exploitation of Meteorological Satellites. (2022). *Meteosat series | EUMETSAT*. https://www.eumetsat.int/our-satellites/meteosat-series?sjid=future

Hebden, S. (2020). *Plans for a new wave of European satellites*. https://futureearth.org/wp-content/uploads/2020/01/issuebrief_04_03.pdf

Overpeck, J. T., Meehl, G. A., Bony, S., & Easterling, D. R. (2011). Climate Data Challenges in the 21st Century. *Science*, *331*. https://doi.org/10.1126/science.1197869

Sudmanns, M., Tiede, D., Lang, S., Bergstedt, H., Trost, G., Augustin, H., Baraldi, A., & Blaschke, T. (2020). Big Earth data: disruptive changes in Earth observation data management and analysis? *International Journal of Digital Earth*, *13/7*. https://doi.org/10.1080/17538947.2019.1585976

Wagemann, J., Fierli, F., Mantovani, S., Siemen, S., Seeger, B., & Bendix, J. (2022). Five Guiding Principles to Make Jupyter Notebooks Fit for Earth Observation Data Education. *Remote Sensing*, *14/14*. https://doi.org/10.3390/rs14143359

Wagemann, J., Siemen, S., Seeger, B., & Bendix, J. (2021). Users of open Big Earth data - An analysis of the current state. *Computers & Geosciences*, *157/104916*. https://doi.org/10.1016/j.cageo.2021.104916

Zhu, Z., Wulder, M. A., Woodcock, C. E., Hansen, M. C., Radeloff, V. C., Healey, S. P., Schaaf, C., Hostert, P., Strobl, P., Pekel, J. F., Lymburner, L., Pahlevan, N., & Scambos, T. A. (2019). Benefits of the free and open Landsat data policy. *Remote Sensing of Environment*, *224*. https://doi.org/10.1016/j.rse.2019.02.016