Lessons from the field
Developing the Tutorial:
TIME SERIES OF SATELLITE DATA USING PYTHON IN THE CLOUD

Marisol García-Reyes
Farallon Institute \ marisolgr@gmail.com
ROADMAP

• Motivation
• The tutorial
• Lessons learned
OPEN SCIENCE

• Sharing Data, Methods & Code

• Ensure accessibility to people that does not have it
  • Limited resources or
  • Limited expertise (not in the same field)

• Advancement & innovation by bringing diversity of ideas
Data and computing is migrating to the cloud

MAKE SATELLITE DATA & CLOUD COMPUTING MORE ACCESSIBLE TO NON-EXPERTS

One time series at a time
ROADMAP

• Scientific Motivation

• The tutorial itself

• Lessons learned
Many web-based user-friendly interfaces to acquire & plot images & short time series of satellite & satellite-based data.
Satellites around for >4 decades now

Great for Climate Studies!

Salomonson, 2015
CLIMATE RESEARCH

Global Sea Surface Temperature Anomaly

Trend: 1981-2020

Blended satellite and in situ sea surface temperature

https://www.hillaryscannell.com/research/
ECOSYSTEMS RESEARCH (OR ANY OTHER FIELD)

I’m an anchovy

Biological indicator

Climate indicator
ECOSYSTEMS RESEARCH
(OR ANY OTHER FIELD)

- Analysis where little/short in situ physical data is available
- Investigate local/regional mechanisms
But knowing where to start & discern the large (& changing) information available online can be intimidating & discouraging

This tutorial aims to provide an overview of what is possible, bypassing many of the background struggles we all had to surmount:

• Installing and updating python
• Having the correct libraries
• Learning the basic to arrive to the cool stuff
• Locating the data
• Learning how to get the data
Scientists

- with limited expertise: satellite data & computing
- with limited resources: storage, computing power & fast internet
- need an update on tools and/or skills
ROADMAP

• Motivation
• The tutorial
• Lessons learned
**Satellite Data Python Tutorials**

Tutorial to learn how to access and process Satellite Data using Python and JupyterLab in the Cloud

**Objective**

This tutorial aims to provide scientists who want to use satellite data with the necessary tools for obtaining, temporally analyzing, and visualizing these data using the Cloud. Note: This is not a tutorial on Python per se - there are a myriad of resources for that. The purpose of this tutorial is to learn, through examples, only the necessary Python code and tools required to do simple temporal analysis of satellite data. We want you to get your toes wet, get to see and use the power of Python, and then maybe you will want to learn more. For that, we encourage you to visit the links on the Resources section at the end of each chapter.

This project, supported by the Better Scientific Software foundation, and originally by NASA, aims to increase accessibility of satellite data and cloud technologies to a broad scientific community through easy-to-follow Python examples.
where to run it

tutorials available @

satellite data @

GitHub

binder

local

amazon web services
• Chapter 1 – What’s Python
• Chapter 2 – Jupyter Notebooks
• Chapter 3 – Basics of Python: Data structures
• Chapter 4a – xarray
• Chapter 4b – matplotlib
• Chapter 5 – Satellite data in the cloud
EXAMPLE CHAPTERS

• Chapter 6 – Ocean: SST
• Chapter 7 – Atmosphere: winds
• Chapter 8 – Land: Vegetation Index

• Plot data, basic time series analysis

Data:
- Satellite
- Satellite-based

Acquisition:
- Cloud
- Online
Tutorial for Acquisition & Analysis of **Time Series of Satellite Earth Data** in Python on the Cloud

github.com/marisolgr/python_sat_tutorials
ROADMAP

• Motivation

• The tutorial

• Lessons learned
PERSONAL PERSPECTIVE

• I am not an expert on technology or coding

• I am a user
  • one step removed from the end user who has limited coding expertise
LESSONS LEARNED

• Lessons learned:
  • Making the tutorials
  • Teaching the tutorials
TEACHING THE TUTORIAL

• Time series earth data acquisition is a challenge on its own
  • Gridded data formats vs. Time-step ‘image’ files
  • Generalizations are not simple because of diverse data types, but problems are similar

• Some coding expertise is necessary

• People that tried (and failed many times) to access similar data, find it most useful
MAKING THE TUTORIALS

• Challenges:
  • Data
  • Software
  • Bringing it all together
CHALLENGES WITH DATA

- Formats
- Storage
- Availability
CHALLENGES WITH SOFTWARE

• New coding language, tools, libraries
  • Updates and versions
  • Availability and stability of tools
  • Access/functions differ for each data/format

• Coding expertise
  • New tools/methods can be difficult to grasp quickly
BRINGING IT ALL TOGETHER

• Tools
  • Guidance for which one to use
  • Versions
BRINGING IT ALL TOGETHER

• Tools
  • Guidance for which one to use
  • Versions

• Coding expertise
  • Troubleshooting work flow problems
  • Steep & long learning curve
• Started with mybinder. Worked great!

• Few updates in the libraries were challenging when building in mybinder, and resulted in no access to data

• Since I couldn’t fix access to AWS data, I switched to pangeo binder (another platform)

• A change in the Pangeo server configurations and images led to loose access to AWS data (libraries problems that the user – me - couldn’t fix)

• Tried mybinder again, and worked!
• Started with mybinder. Worked great!

• Few updates in the libraries were challenging when building in mybinder, and resulted in no access to data

• Since I couldn’t fix access to data with mybinder, I switched to pangeo binder (another platform).

• A change in the Pangeo server configurations and images led to loose access to AWS data (libraries problems that the user – me - couldn’t fix)

• Tried mybinder again, and worked!

I had to get help from an expert!
THE GOOD PARTS

• Accessing the data for free, from a free online binder, was not possible when I started. Now you can access it from anywhere.

• New clever and promising ways to store and access data are emerging

• Things are changing, and fast, but seems like in a good direction

• Many fields have similar challenges, opportunities for collaborative work
Standard data formats that facilitate access & preprocessing in any axes
  ✓ zarr or similar chunking/mapping techniques

Maturity and stability of data & software
  ✓ new and upgraded versions at a fast pace is not conducive to broader use

More time to get there & to learn cloud computing
TAKE AWAY MESSAGES:

• Broad use of cloud data & resources - we are close but not there yet

• Access of data is not just availability of data – accessibility is important

• The computing expertise needed to access & acquire cloud data is very very underestimated