



Opening My Science: A Jupyter Book on Analyzing Sea Level Variability with Xarray

Yuta Norden
SIParCS 2023

August 01, 2023

NCAR

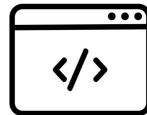


No Transparency

data



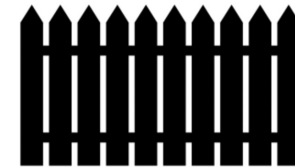
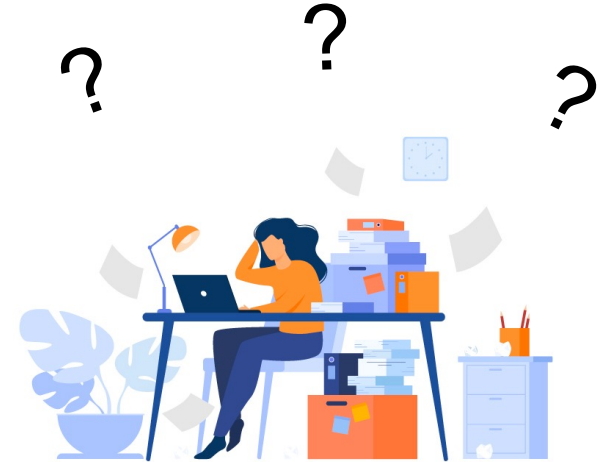
code



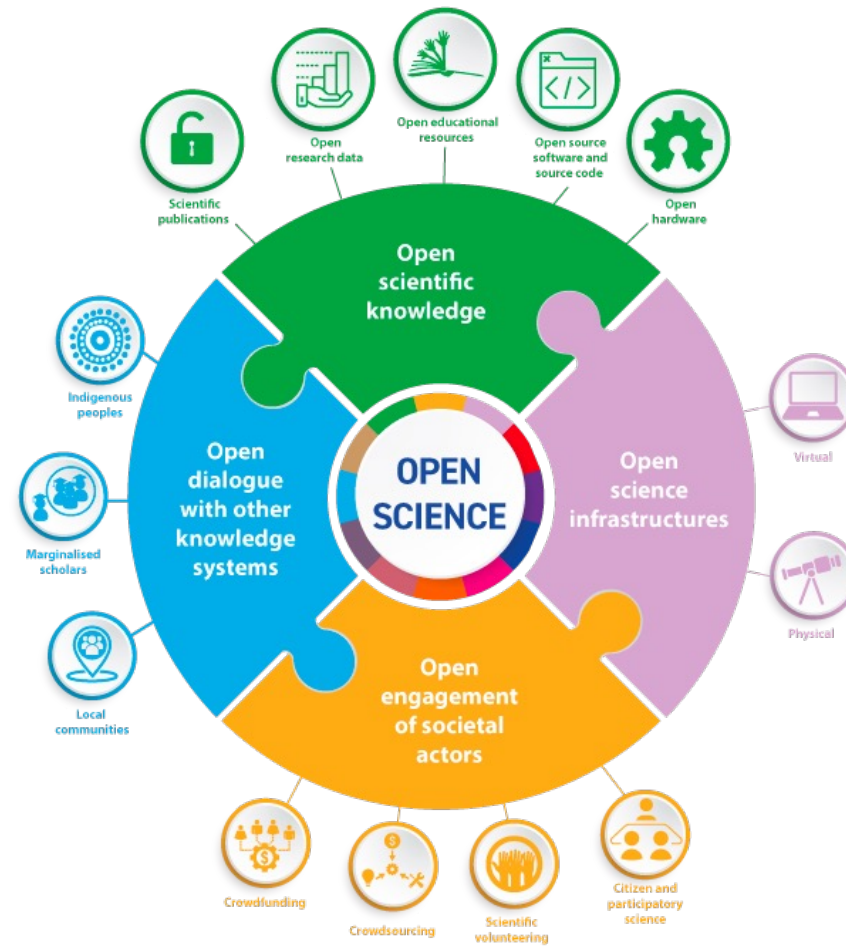
publications



No Reproducibility



The Age of Open Science

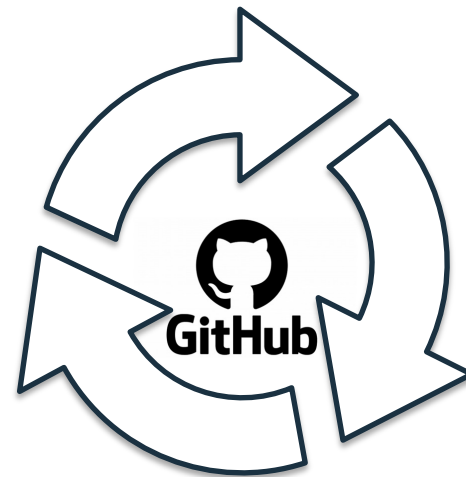


Internship Goal:

PRACTICE OPEN SCIENCE

Technical skills

- getting better at coding



Community

- shared values
- understanding of respect
- ethics



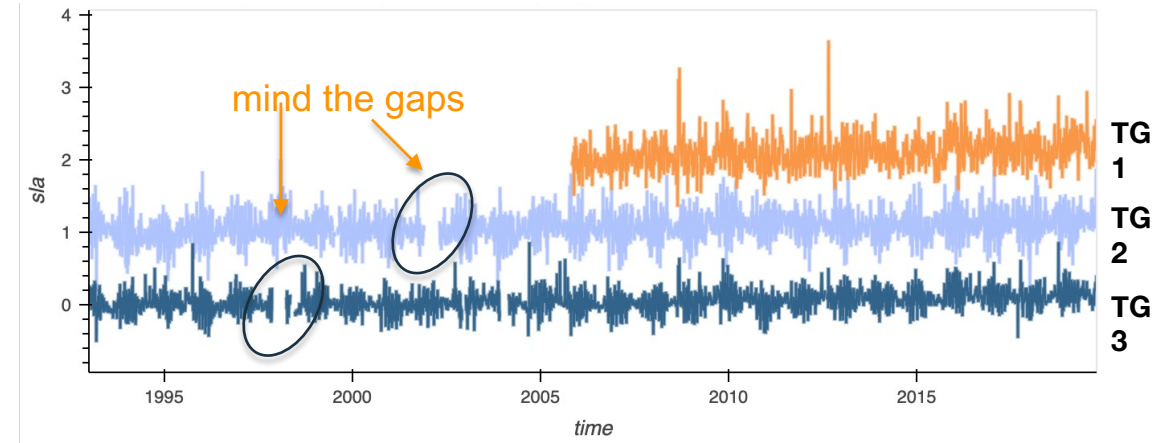
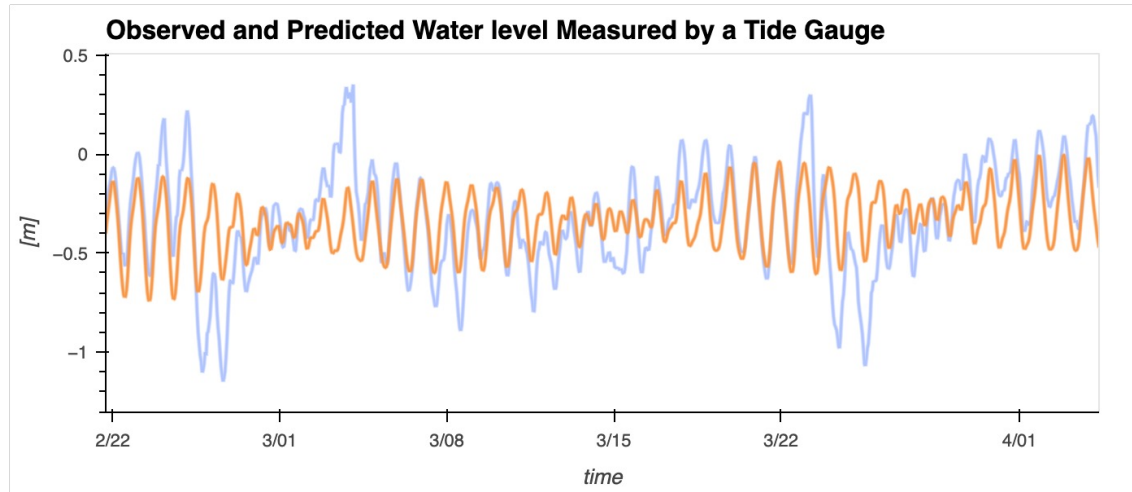
PANGEO



SciPy 2023

Do you know what the surface of the ocean looks like from space?

Tide Gauge Data Tidying



234 separate netCDF files



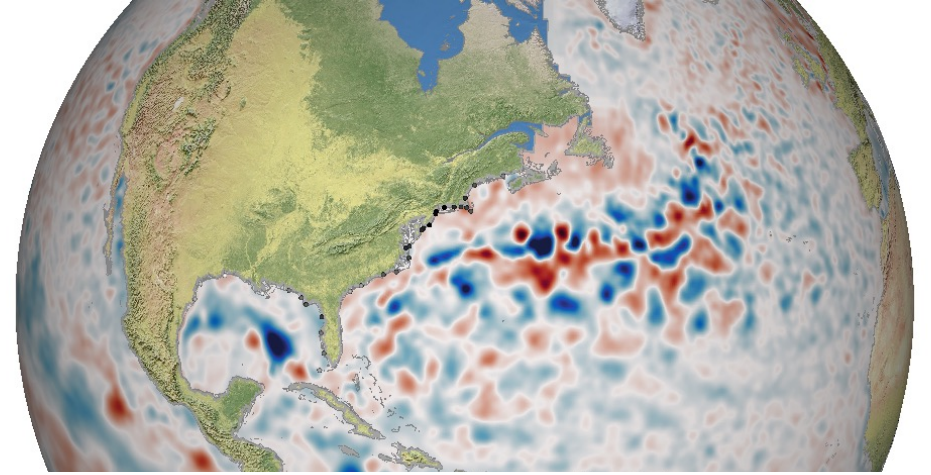
Creating a catalogue of metadata →

	id	city	latitude	longitude	start times	end times
0	1611400.nc	Nawiliwili	21.954400	-159.356100	1954-01-01 01:00:00	2021-05-31 23:00:00
1	1612340.nc	Honolulu	21.306694	-157.867000	1914-01-01 01:00:00	2021-05-31 23:00:00
2	1612480.nc	Mokuoloe	21.433056	-157.790000	1981-01-01 01:00:00	2021-05-31 23:00:00
3	1615680.nc	Kahului, Kahului Harbor	20.895000	-156.476694	1954-01-01 01:00:00	2021-05-31 23:00:00
4	1617433.nc	Kawaihae	20.036600	-155.829400	1990-01-01 01:00:00	2021-05-31 23:00:00

1.

From the DataFrame select :

- stations on the East Coast of the U.S.
by longitude
- select by record length : 1993 - present
available satellite altimetry record



SLA (sea level anomalies) from satellite altimetry

Now I have the selected stations IDs

calculate the anomalies
select the time slice
average hourly → daily
(if at least 18 measurements in a day)



```
xr.open_mfdataset ( Single Step preprocess )
```

single xarray Dataset object

dimensions: 'time' , 'station_id'



keep only time-series
that are 95%
complete



if gaps : interpolate

From a Scientific Question to a Coding Problem

2.

Empirical Orthogonal Functions (EOF) analysis

$$Z(x, y, t) = \sum_{k=1}^N PC(t) \cdot EOF(x, y)$$

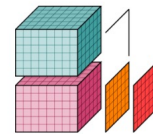
EOFs are spatial structures of variability

Principal components (PCs) are the temporal components that tell you how the amplitude of each EOF varies with time

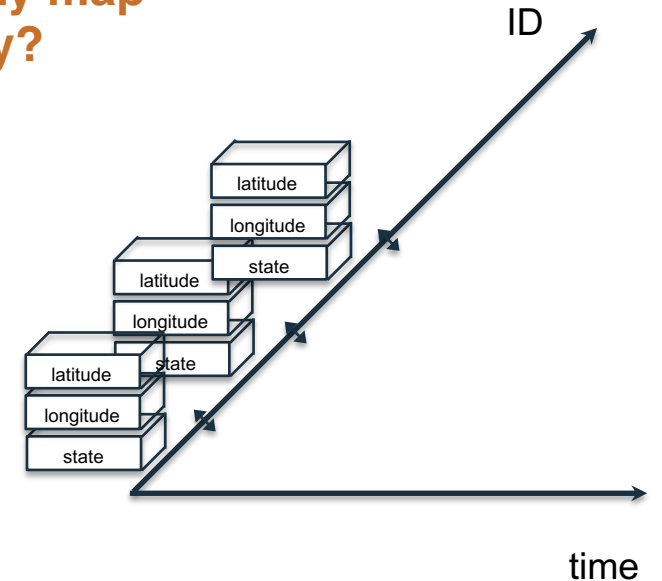
3.

How does the regional scale structure in coastal variability map onto large atmospheric and oceanic patterns of variability?

Multiple Regression Function



xarray.Dataset



Result

Research Efforts

HTF Dynamics

Tide Gauge Data Retrieval and Processing

EOF analysis of TG data

Multiple Linear Regression

Learning Goals

Import Packages

Multiple Linear Regression. Problem setup.

Altimetry file

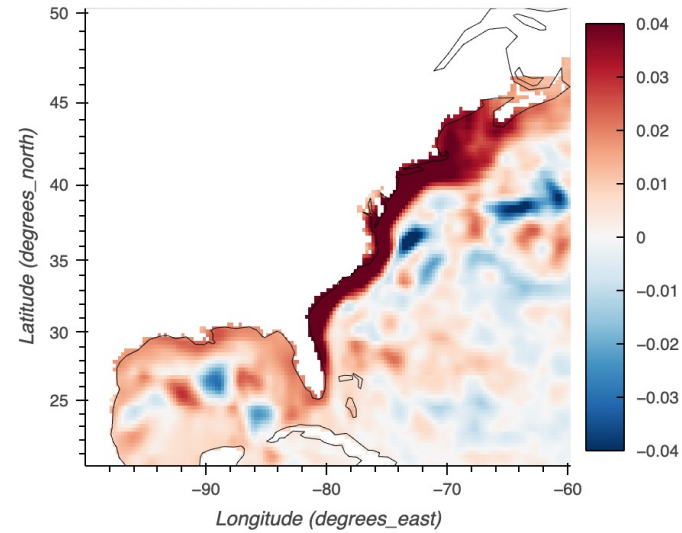
Multiple Linear Regression

apply_ufunc

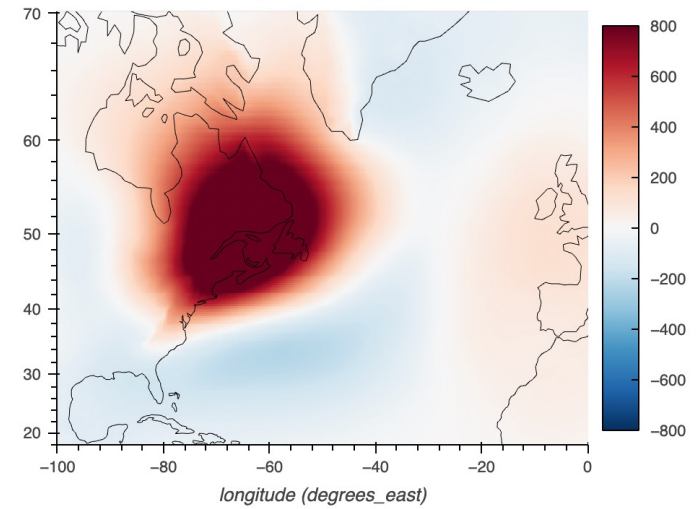
Altimetry Cloud

Non-ambitious .apply_ufunc

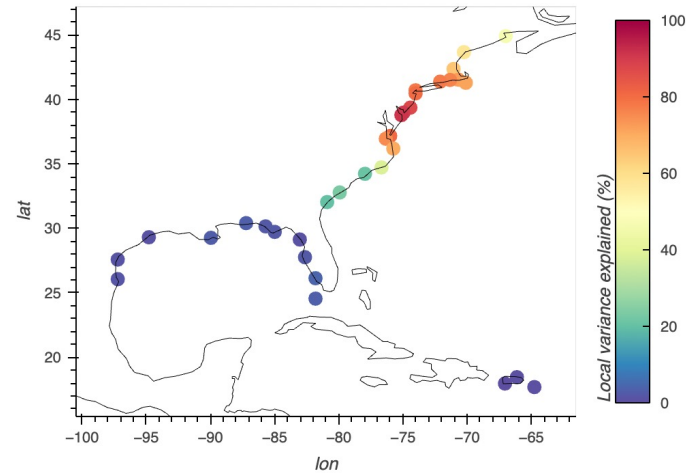
MLR Regression coefficients [m]
(altimetry regressed onto Principal Components)



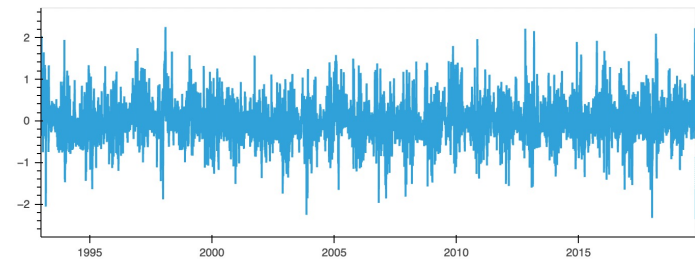
MLR Regression coefficients [Pa]
(sea level pressure regressed onto Principal Components)

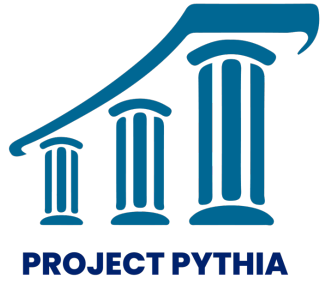


EOF1: 47.32%



Principal Component 1 (temporal pattern)

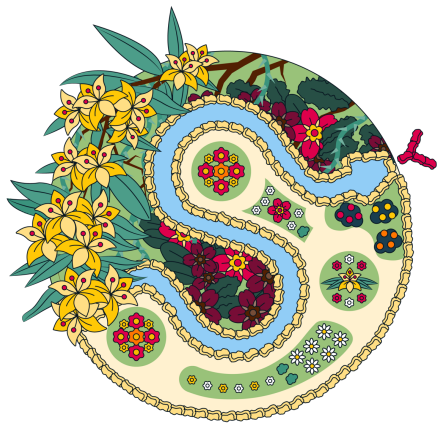




Pythia “Cook-Off” 2023

June 20-23



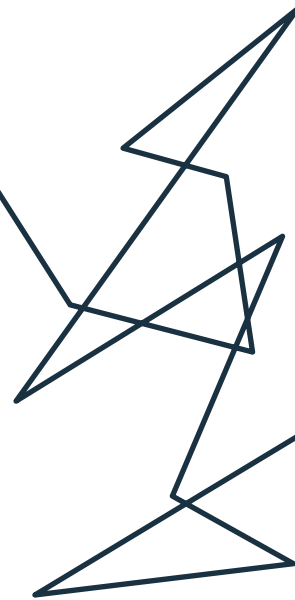


SciPy 2023



Breaking the barrier

lonely



community



I pledge to always do my research according to principles of open science: keep my data and analysis publicly available, help open source science community grow and thrive

Work in Progress

complete and update my Jupyter Book with my research

- **introduce more complicated use case for `.apply_ufunc`**

working with cloud data

- **create a pangeo-recipe for altimetry SSH (sea surface height)**

update a few Xarray docs

- **add examples to important `xr.` functions**

THANK YOU !



SIParCS 2023 cohort



MENTORS

Deepak

Julia

Jessica

Scott

Virginia

Jerry

NESSI cohort

My advisor Dr. P. Thompson for letting me participate

Everyone at NCAR for making me feel welcome and supported



Thank you NASA's Open Source Tools, Frameworks, and Libraries Program (award 80NSSC22K0345) for sponsoring my work !