



International Ocean Colour Science  
Meeting 2019

Advancing Global  
Ocean Colour  
Observations

# Breakout Workshop: Scientific computing and the Open Source software revolution: Opportunities for ocean colour science

Joaquín E. Chaves (NASA GSFC/SSAI), Erdem M. Karaköylü (NASA GSFC/SAIC), Joel P. Scott (NASA GSFC/SAIC)

Tuesday, April 9

## Key questions:

- What recent shifts among Ocean Color community have there been towards open source tools?
- How and why is the community using these tools?
- How are teaching, mentoring, and research deliverables changing by open source?
- Are these tools promoting reproducibility, code preservation, accessibility, diversity & inclusion?

# What is Open Source/Science software?

“Open source software is software with source code that anyone can inspect, modify, and enhance”

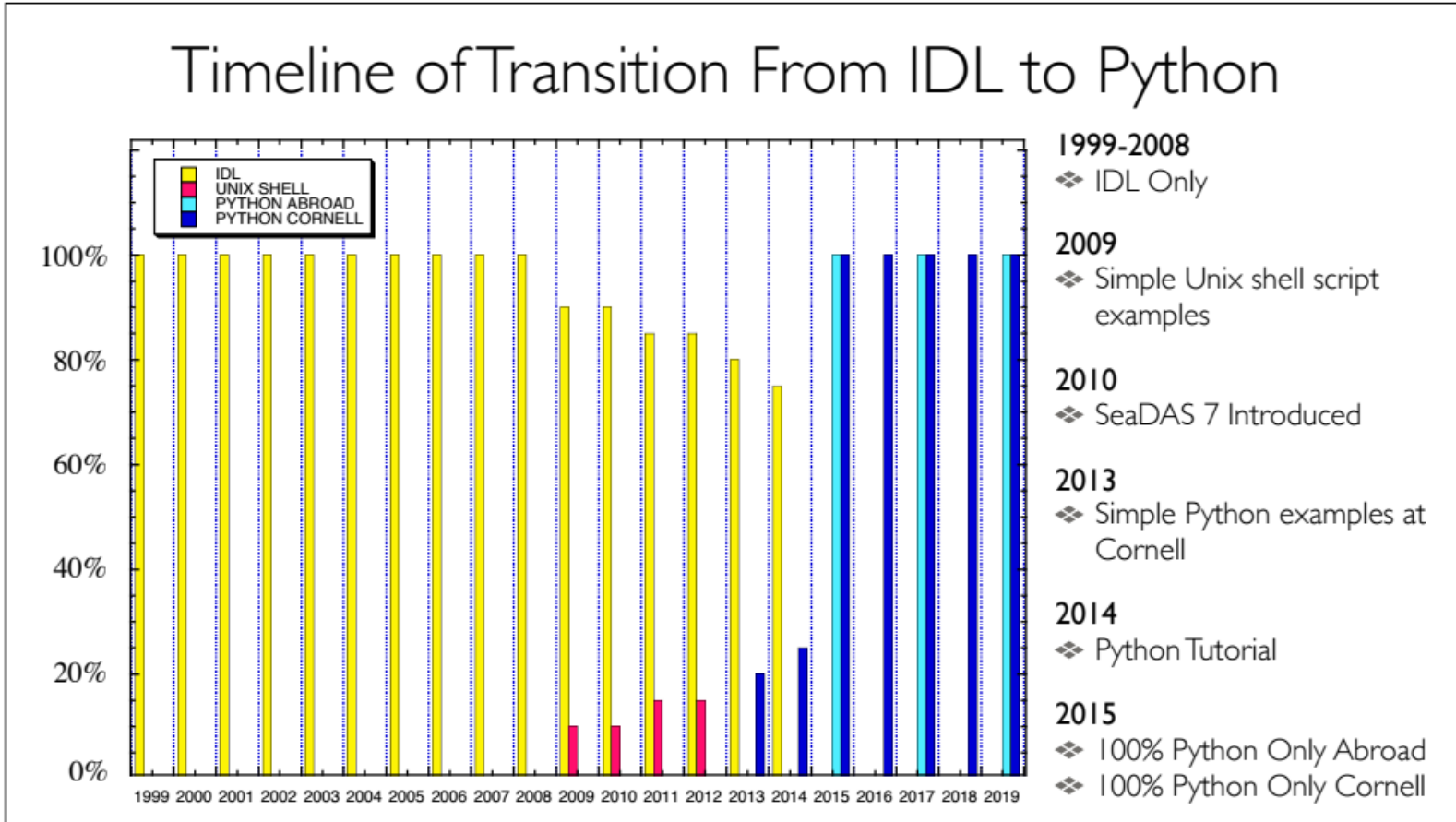


## **Take-aways:**

- 1) Everyone (including newcomers) have something to offer
- 2) Popular/emerging open source tools are Python3, Jupyter-Notebook, Xarray, CartoPy, Git versioning protocol, etc
- 3) Diverse factors have brought community members to open source tools
- 4) Lack of licensing support
- 5) Broader international reach
- 6) Vast number of resources out there to get started
- 7) Stackoverflow
- 8) Cory Schafer on Youtube
- 9) Scribd book resource website to read books on Python

## **Place to start**

- 1) Try simple task
- 2) Convert code from old language to Python
- 3) Play around with someone else's code
- 4) Google
- 5) Attend conferences, dataviz groups, PyLadies chapters, other local resources



Dr. Myung-Sook Park, KIOST:

*Python for multi-year GOCI ocean color products analysis:  
sharing the advantages and issues:*

IDL user for 15 years; switched to Python about a year ago:

Reasons:

--Freedom to use on multiple platforms, computing speed,  
multiple packages available

# Jupyter: Tools for the Life Cycle of a Computational Idea

Leveraging Python & Jupyter Notebook to validate satellite Ocean Color retrievals via SeaBASS data (co-chair: **Joel. Scott**)

Probabilistic Programming Workflow in Python (co-chair **Erdem. Karaköylü**, NASA GSFC/SAIC)

Browser-based environment to perform, document, annotate, & annotate computational ideas in various languages:

**JULia-Python-R**

Links to talks:

[https://github.com/jpscot/IOCS\\_2019\\_Busan\\_OpenScience](https://github.com/jpscot/IOCS_2019_Busan_OpenScience)

[https://github.com/madHatter106/IOCS\\_BUSAN\\_2019](https://github.com/madHatter106/IOCS_BUSAN_2019)

Project Jupyter exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages.

### The Jupyter Notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

[Try it in your browser](#) [Install the Notebook](#)

**Language of choice**  
The Notebook has support for over 40 programming languages, including Python, R, Julia, and Scala.

**Share notebooks**  
Notebooks can be shared with others using email, Dropbox, GitHub and the [Jupyter Notebook Viewer](#).

**Interactive output**  
Your code can produce rich, interactive output: HTML, images, videos, LaTeX, and custom MIME types.

**Big data integration**  
Leverage big data tools, such as Apache Spark, from Python, R and Scala. Explore that same data with pandas, scikit-learn, ggplot2, TensorFlow.

Open Source in education, mentoring, and knowledge distribution:

- **Develop a community ‘open science’ statement**
  - Good progress making data open, can we encourage open software too (e.g., processors).
  - Support for open source use in training
- **Expand great work of deriving community standards for protocols/providing consolidated reports.**
  - Can we do this for open source software use in our community? E.g. provide ocean colour examples with open source code/standards/versioning etc.
- **Raising use of community repositories**
  - Great history of doing this with data
  - Share useful snippets
  - Standards to be discussed! Licencing and attribution.
  - What are ‘common workflows’ for Ocean Colour (and beyond?)



## Actions:

Preliminary Working group?

Could we have a code repository that exists as a live IOCCG report  
Titled: Open Science Principles for Ocean Colour

Would we need a curator?

Do it as a <https://github.com/eumetsat-training>

[https://gitlab.com/ben.loveday/EUMETSAT\\_training\\_jupyter\\_notebooks](https://gitlab.com/ben.loveday/EUMETSAT_training_jupyter_notebooks)

Allow code submissions as demos/examples

— common workflows: plotting, matchups,  
adding cruise track to satellite data

Propose workshop to scope out what this would be.