Developing an Executable Paper With the Python In Heliophysics Community (PyHC)

Steven Petrasi (1), Nicholas Ringuette (2,3), Luis Kamamizu (1), Eric Grimes (4), Jonathan Niehof (5), Kristina Oliveira (6), Brian Mac (1), Ryan Bunch (1)

(1) UCLA, (2) Johns Hopkins University, (3) Applied Physics Laboratory, (4) University of Iowa, (5) Aerospace, (6) Center for Astrophysics | Harvard & Smithsonian

AGU Fall Meeting
December 9–13, 2024
Washington, D.C.

WHAT ARE EXECUTABLE PAPERS?

"Scientific publication of software that combines code, data, and prose in a form ready for reproducible-science use by others in a computational environment."

- Source: **Ragan-Kelley's review of papers where the code is broken but still in the mix** (Ragan-Kelley et al. 2014)
  - Ragan-Kelley's review of papers where the code is broken
- This year's talk from **Lorena Barba** (prof. from University of Illinois)
- **Data and reproducibility** (and **open science**) (Jupyter Notebook, Pluto.jl, Quarto, etc.)

Nature 533, 452–454 (2016) doi:10.1038/533452a (1,576 PEOPLE SURVEYED ON REPRODUCIBILITY)

- **Big Ideas in Computational Science**
  - Computational reproducibility
  - Computational reproducibility
  - Computational reproducibility

Yihua Zheng:	**Science expertise**

FUTURE WORK

- Incorporate Kamodo
- Evolve this prototype into **two separate papers**:
  - **One about our adaptable Python workflow and how we make papers executable**
    - Use of Python packages described in-depth
    - More emphasis on using Deepnote and making executable papers in general
  - **One about magnetopause locations, leveraging our Python workflow**
    - More models compared against more sections of data
    - Comprehensive findings described
- Publish them with links to their executable versions
- **Advertise our approach** 📢

DISCLOSURES

PyHC's work is supported by NASA under award 80NSSC20K1443, managed by the Goddard Space Flight Center.

PlasmaPy development is supported by the U.S. National Science Foundation (NSF) and NASA, with many contributions from the broader plasma physics and open source communities.

AUTHOR INFORMATION

- Shawn Polson: Shawn Polson graduated from the University of Colorado Boulder in 2020 with his Masters in Computer Science. His grad school research focused on detecting anomalies in spacecraft with machine learning techniques. He works as a software engineer at the Laboratory for Atmospheric and Space Physics (LASP) in Boulder and he is the Technical Lead of the Python in Heliophysics Community (PyHC).
- Rebecca Ringuette: Rebecca Ringuette graduated in 2014 with her doctoral degree in physics from Louisiana State University on detecting terrestrial gamma-ray flashes from the rooftops of LSU. After teaching physics and astronomy at the high school and university levels for a few years, Rebecca transitioned into X-ray astrophysics in 2018, studying astrophysical observations of solar wind charge exchange and developing pipeline software in Python for the HaloSat CubeSat mission. She is now contracted as a scientific software developer through ADNET Systems, Inc, for the Community Coordinated Modeling Center at NASA GSFC, building Kamodo-based tools for easier interaction with space weather model data. Rebecca is also contributing to the improvement of the heliophysics infrastructure through publications and workshops.
- Lutz Rastaetter: Lutz Rastaetter received his Ph.D. in Physics from the Ruhr University Bochum, Germany, in 1997 on numerically modeling the magnetic environment of a rotating star loaded by an accretion disk. He joined NASA Goddard Space Flight Center in Greenbelt, Maryland as a post-doctoral researcher working on current sheet thinning and reconnection in the Earth's magnetosphere. Since the inception of the Community Coordinated Modeling Center (CCMC) in 2000, he has been working on model onboarding, online visualization, model-data validation, and the development of postprocessing tools for use by the science community. He leads the Kamodo model access software development at the CCMC.
- Eric Grimes: Eric Grimes received his M.S. in Physics from Auburn University in 2012. Since 2012, he has been a Programmer/Analyst at the UCLA Institute of Geophysics and Planetary Physics, specializing in scientific software development. He has developed analysis and visualization tools for several NASA Heliophysics missions, including Magnetospheric Multiscale Mission (MMS), Time History of Events and Macroscale Interactions during Substorms (THEMIS), and more, and currently leads development of the pySPEDAS project.
- Jonathan Niehof: Jonathan Niehof earned his PhD from Boston University in 2011 with investigations of energetic ions in Earth's magnetospheric cusps. Since then, he has lead data processing and analysis for several missions, including the ECT suite on the Van Allen Probes mission, the ISOIS suite on Parker Solar Probe, and the upcoming Helioswarm mission. He is currently at the University of New Hampshire.
- Nicholas Murphy: Nick Murphy is an astrophysicist in the Solar & Stellar X-Ray Group at the Smithsonian Astrophysical Observatory (SAO) and a Lecturer in Harvard University's Department of Astronomy. He is also a developer of PlasmaPy, a community-developed core Python package for plasma physics.
- Yihua Zheng: Yihua Zheng received her Ph.D in 2001 from University of New Hampshire, on the field-aligned currents in the auroral zone using sounding rocket measurements. She is a space science and space weather scientist & a forecaster who has been involved in providing space weather services to NASA robotic missions for over 10 years. She studies the myriad ways that space weather can manifest in near-Earth space. Yihua is the lead at the Community Coordinated Modeling Center (CCMC) for inner magnetosphere models, and for coupling space environment models with engineering impacts.

ABSTRACT

The reproducibility crisis was sparked five years ago by a Nature survey that revealed most researchers have tried and failed to reproduce another scientist's experiments (Baker, 2016). Although the severity of the crisis is still being debated, it has spurred researchers to include reproducibility and open science practices in their work. To that end, the Python in Heliophysics Community (PyHC) encourages members of the heliophysics community to create "executable papers," which are interactive pieces of software that combine text, data, and code to enable readers to reproduce every step taken to arrive at a publication's conclusions. We describe the development of an executable paper—in this case, one centered around magnetopause boundary crossings in satellite data—our coordination efforts between scientists and software developers, and the current technology that makes this possible. We aim to promote this novel paper format and be an example to the heliophysics community of how scientists and software developers can work together to produce research that is transparent and reproducible.

REFERENCES

Baker, M. 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454 (2016). https://doi.org/10.1038/533452a

https://commons.wikimedia.org/wiki/File:Magnetopause.jpg (modified)

https://github.com/EnsembleGovServices/kamodo-core

https://github.com/MAVENSDC/PyTplot

https://github.com/nasa/Kamodo

https://github.com/PlasmaPy/PlasmaPy

https://github.com/spacepy/spacepy

https://github.com/spedas/pyspedas

https://mms.gsfc.nasa.gov

Lasser, Jana, et al. "Salt polygons are caused by convection." https://github.com/JanaLasser/salt-polygons-are-caused-by-convection (2019)

Pembroke, A., DeZeeuw, D., Rastaetter, L., Ringuette, R., Gerland, O., Patel, D., and Contreras, M. (2022). Kamodo: A functional api for space weather models and data. Journal of Open Source Software, submitted