

For Python and R users: Citing your notebooks in your publication.

AGU Fall Meeting - 16 December 2021

Chris Erdmann
Assistant Director, Data Leadership
American Geophysical Union
0000-0003-2554-180X
@libcce | cerdmann@agu.org

Shelley Stall
Senior Director, Data Leadership
American Geophysical Union
0000-0003-2926-8353
@ShelleyStall | sstall@agu.org

Use of Notebooks



Peter Parente

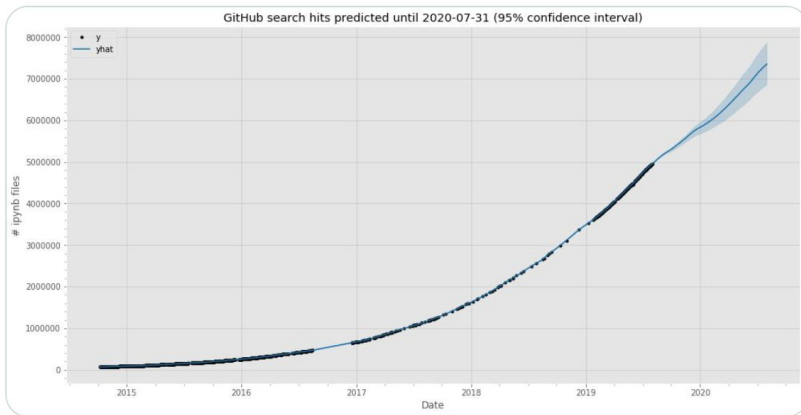
@parente

...

Coming up on 5 million public @ProjectJupyter notebooks on GitHub.

nbviewer.jupyter.org/github/parente...

📄 📄 📄 📄 📄 e6



9:26 AM · Aug 2, 2019 · Twitter Web App

[7.2M ipynb counts on GitHub](#)

(Dec 7, 2021) via Peter Parente's Notebook

Growth in Software Citation

DataCite Commons

client.uid:cern.zenodo



Pages ▾

Support

Sign In

Works

People

Organizations

216,936 Works

Publication Year

| | |
|-------------------------------|--------|
| <input type="checkbox"/> 2021 | 71,161 |
| <input type="checkbox"/> 2020 | 50,708 |
| <input type="checkbox"/> 2019 | 34,179 |
| <input type="checkbox"/> 2018 | 24,871 |
| <input type="checkbox"/> 2017 | 18,974 |
| <input type="checkbox"/> 2016 | 9,721 |
| <input type="checkbox"/> 2015 | 4,625 |
| <input type="checkbox"/> 2014 | 2,220 |
| <input type="checkbox"/> 2013 | 150 |
| <input type="checkbox"/> 2012 | 87 |
| <input type="checkbox"/> 2011 | 43 |
| <input type="checkbox"/> 2010 | 39 |

Work Type

| | |
|--|--------|
| <input checked="" type="checkbox"/> Software | 14,427 |
|--|--------|

Thousands Of Exon Skipping Events Differentiate Among Splicing Patterns In Sixteen Human Tissues

Liliana Florea, Li Song & Steven L Salzberg

Content published 2013 in [Zenodo](#)

Software and associated data from the paper, including: ASprofile software (ASprofile.tar.gz); Exon skipping events with annotations (BodyMap.exon_skipping.tbl); Concatenated GTF transcript file for the 16 tissues (BodyMap.ens61.0.0.gtf.gz); (All) Events extracted from pairwise transcript comparisons (BodyMap.ens61.0.0.as.gz); Expression (FPKM) values for exon skipping events in the 16 tissues (BodyMap.fpkms.tar.gz); and Supporting summary data (Excel) for the alternative splicing analyses (BodyMap.xls).

DOI registered August 31, 2013 via DataCite.

Software

<https://doi.org/10.5281/zenodo.7068>

Computational Model Of Arterial Blood Pressure

Daniel A. Beard, Klas H. Pettersen, Brian E. Carlson, Stig W. Omholt & Scott M. Bugenhagen

Content published 2013 in [Zenodo](#)

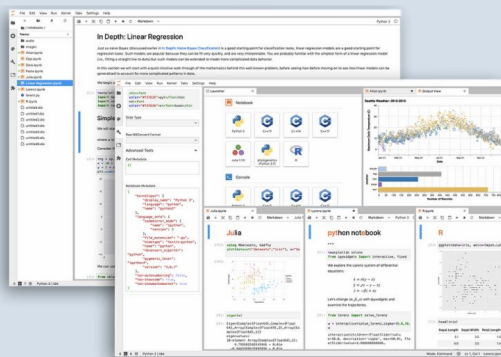
A computational model was developed and identified to serve as a phenomenological representation of the major physiological processes controlling arterial blood pressure.

FEEDBACK

<https://commons.datacite.org/doi.org?query=client.uid%3Acern.zenodo&resource-type=software>

NOTEBOOKS

New EarthCube Peer-Reviewed Jupyter Notebooks Now Available



A novel element of the **EarthCube 2020 Annual Conference** was a call for **notebooks**, which led to twelve peer-reviewed Jupyter notebooks that encompass an array of geoscience data tools, software, services, and libraries. Each notebook was reviewed by scholars within the geoscience and cyberinfrastructure community at the EarthCube 2020 Annual Conference.

These notebooks are now available on GitHub for interested researchers to view and execute, and they will soon be published in the Earth and Space Science Open Archive (ESSOAr).

<https://www.earthcube.org/notebooks>

Earth and Space Science Open Archive >

This is a preprint and has not been peer reviewed.

ESSOAr is a venue for early communication or feedback before peer review. Data may be preliminary. [Learn more about preprints](#)

Preprint • [Open Access](#) • You are viewing the latest version by default [v1]

PmagPy Online: Jupyter Notebooks, the PmagPy Software Package and the Magnetism Information Consortium (MagIC) Database

Authors

Lisa Tauxe 

Rupert Minnet 



Supplemental



Related



Details

Files

PmagPy_online_Earthcube_conference.ipynb

File Type: Jupyter Notebooks

File

Name: [PmagPy_online_Earthcube_conference.ipynb](#)

Links

https://github.com/earthcube2020/ec20_tauxe_etal

Link Type: Jupyter Notebooks

https://github.com/earthcube2020/ec20_tauxe_etal

<https://doi.org/10.1002/essoar.10504182.1>

JGR Space Physics

Research Article |  Full Access

Resonant Whistler-Electron Interactions: MMS Observations Versus Test-Particle Simulation

E. Behar , F. Sahraoui, L. Berčič



[Volume 125, Issue 10](#)

October 2020

e2020JA028040

Appendix A: The `aidaPy` Library and Available Notebooks and Scripts

All the tools developed for this study have been implemented in the open-source `aidaPy` python library, part of the European project AIDA: <https://aida-space.eu/>

The library and its installation instructions can be found at <https://gitlab.com/aidaspace/aidapy>.

The library documentation is maintained in <https://aidapy.readthedocs.io/en/latest/index.html>

The python script `vdf_mms.py` reproduces most of the article's VDF figures and is available at https://gitlab.com/aidaspace/aidapy/-/blob/master/examples/01_missions/vdf_mms.py

The more didactic notebook `4_whistler_cyclotron.ipynb`, with explanatory text, code samples and included figures, reproduces all the VDF results, and can be viewed on any web browser:

https://gitlab.com/aidaspace/notebooks_aida/-/blob/master/02_velocity_distribution_tool/04_whistler_cyclotron.ipynb

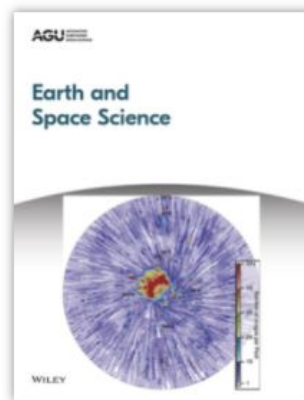
<https://doi.org/10.1029/2020JA028040>

Earth and Space Science

Technical Reports: Methods | [Open Access](#) |   

pyUserCalc: A Revised Jupyter Notebook Calculator for Uranium-Series Disequilibria in Basalts

Lynne J. Elkins , Marc Spiegelman



[Volume 8, Issue 12](#)

December 2021

e2020EA001619

To run the example code and use this article as a functioning Jupyter notebook, while in a web-enabled browser the user should select an embedded code cell and then simultaneously type the “Shift” and “Enter” keys to run the cell, after which selection will automatically advance to the following cell. The first cell below imports necessary code libraries to access the Python toolboxes and functions that will be used in the rest of the program:

```
[1]: # Select this cell with by mouseclick, and run the code by simultaneously  
    →typing the 'Shift' + 'Enter' keys.  
    # If the browser is able to run the Jupyter notebook, a number [1] will  
    →appear to the left of the cell.  
  
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
%matplotlib inline  
  
# Import UserCalc:  
import UserCalc
```

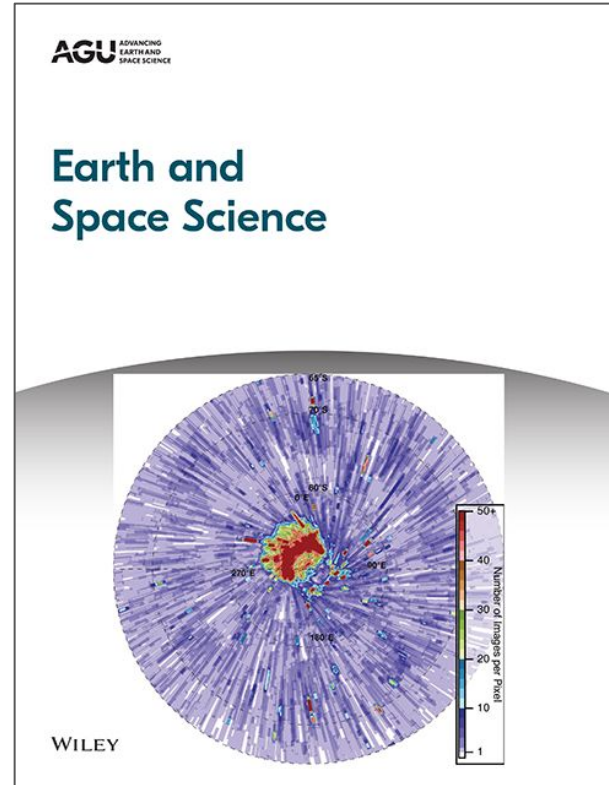
Access the full executable Jupyter notebook version of this manuscript here



<https://doi.org/10.1029/2020EA001619>

Notebooks @ AGU ESS

- Elevate notebooks
 - Scholarly objects
 - Teaching tools
- Give credit



Guidance for Authors - Jupyter Notebooks



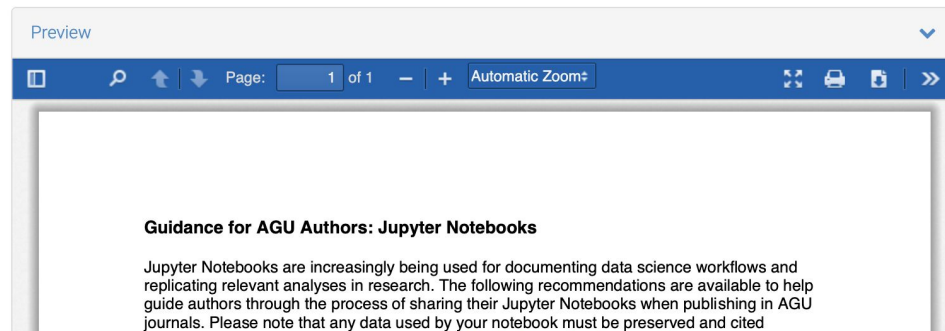
May 20, 2021

Other Open Access

Guidance for AGU Authors - Jupyter Notebooks

Erdmann, Christopher; Stall, Shelley

Jupyter Notebooks are increasingly being used for documenting data science workflows and replicating relevant analyses in research. The following recommendations are available to help guide authors through the process of sharing their Jupyter Notebooks when publishing in AGU journals. Please note that any data used by your notebook must be preserved and cited separately from the notebook to comply with [AGU's data and software guidance](#).



Erdmann, Christopher, Stall, Shelley, Gentemann, Chelle, Holdgraf, Chris, Fernandes, Filipe P. A., Gehlen, Karsten Peters-von, & Corvellec, Marianne. (2021). Guidance for AGU Authors - Jupyter Notebooks. Zenodo.

<https://doi.org/10.5281/zenodo.5651648>

Guidance for AGU Authors - Jupyter Notebooks

JUNE 14, 2021

Guidance For AGU Authors: Jupyter Notebooks

Jupyter Notebooks are increasingly being used for documenting data science workflows and replicating relevant analyses in research. The following recommendations are available to help guide authors through the process of sharing their Jupyter Notebooks when publishing in AGU journals. Please note that any data used by your notebook must be preserved and cited separately from the notebook to comply with [AGU's data guidance](#).

First, read. [Ten simple rules for writing and sharing computational analyses in Jupyter Notebooks](#) and the recommendations from [Jupyter Notebooks and reproducible data science](#) will help you with creating

- <https://data.agu.org/resources/jupyter-notebooks-guidance>
- https://github.com/AGU-Data/agu-data.github.io/blob/master/_resources/jupyter-notebooks-guidance.md

First, read. [Ten simple rules for writing and sharing computational analyses in Jupyter Notebooks](#) and the recommendations from [Jupyter Notebooks and reproducible data science](#) will help you with creating computational notebooks that are easier to read, run, and explore.

Second, store and preserve. Upload your notebook in a public repository that gives you a Digital Object Identifier (DOI), so that your notebook is preserved in the scholarly record and can be cited. We recommend using the following solution:

- *GitHub with documentation* (e.g., README), [\(open source\) license](#), [citation file](#) (e.g., CFF), [CodeMeta/JSON-LD](#) file. Manage dependencies (with [conda-lock](#)). Use the GitHub release bridge with [Zenodo](#) to get a DOI ([Making your code citable](#)).

Third, execute. Facilitate the rerunning of your notebook by readers through a zero-install environment in the cloud. We recommend [these instructions](#) and the following solution:

- [Binder](#) or [Pangeo](#), upload your notebook and an environment.yml file, run it, and share it via link

No matter which execution environment you choose, make sure you obtain a publicly accessible link to your notebook so that it can be executed by any reader.

Fourth, cite and describe. Establish the connection between your notebook and paper using these two steps.

- Cite your notebook, using your DOI, in the References section of your paper. Example citation below:

Fenner, M. (2019). Jupyter Notebook FREYA PID Graph Key Performance Indicators (KPIs) (Version 1.1.0). DataCite. <https://doi.org/10.14454/3BPW-W381>

- Describe the availability of your notebook by including the DOI of the preserved release used by your paper, the development platform where it is hosted, and the execution link in the Availability Statement of your paper. For AGU, the Availability Statement is located in the Open Research section. Statement template is below:

“The Jupyter Notebook to execute the analysis in the paper can be found at [include link to rerunnable notebook link in the cloud], is hosted at [development platform, e.g., GitHub] and is preserved at [include the link to the preserved notebook DOI, version, and license].”

Fifth, render (optional). Consider rendering your notebook using the [Executable Book Project](#) and [Jupyter Book](#). An example [project gallery](#) of Jupyter Books is available. The [Neuro Libre Project](#) features additional Jupyter Book examples.

Guidance for AGU Authors - R Scripts and Markdown



November 5, 2021

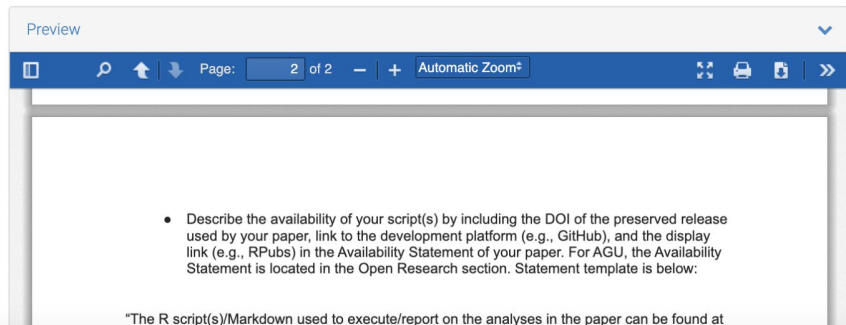
Other Open Access

Guidance for AGU Authors: R Script(s)/Markdown

Erdmann, Christopher; Meyer, Michael F.; Little, John R.; Hondula, Kelly; Stachelek, Jemma; Oleksy, Isabella; Brousil, Matthew R.; Claborn, Kelly; Mesman, Jorrit; Dennis, Tim

R is a popular programming language in the AGU community used for analyses and reporting. The following recommendations are available to help guide authors through the process of sharing their R script(s)/Markdown when submitting to AGU journals. Please note that any data used in your R script(s)/Markdown must be preserved and cited separately according to [AGU's Data and Software for Authors](#) guidance.

This work is supported by Accelerating Open and FAIR Data Practices Across the Earth, Space, and Environmental Sciences: A Pilot with the NSF to Support Public Access to Research Data project funded by the National Science Foundation, Grant 2025364.



Erdmann, Christopher, Meyer, Michael F., Little, John R., Hondula, Kelly, Stachelek, Jemma, Oleksy, Isabella, Brousil, Matthew R., Claborn, Kelly, Mesman, Jorrit, & Dennis, Tim. (2021). Guidance for AGU Authors: R Script(s)/Markdown. Zenodo.

<https://doi.org/10.5281/zenodo.5647998>

Guidance for AGU Authors - R Scripts and Markdown

NOVEMBER 05, 2021

Guidance for AGU Authors: R Script(s)/Markdown

R is a popular programming language in the AGU community used for analyses and reporting. The following recommendations are available to help guide authors through the process of sharing their R script(s)/Markdown when submitting to AGU journals. Please note that any data used in your R script(s)/Markdown must be preserved and cited separately according to [AGU's Data and Software for Authors](#) guidance.

First, read. The article, [There is an R in Reproducible](#), highlights five reproducible workflow principles to ensure your analyses produce consistent results. The [Tidyverse Style Guide](#) and Jenny Bryan's [naming things](#) are also nice places to

- <https://data.agu.org/resources/r-guidance-agu-authors>
- https://github.com/AGU-Data/agu-data.github.io/blob/master/_resources/r-guidance-agu-authors.md

Data & Software for Authors

WHAT IS NEEDED?

AGU requires that the underlying data needed to understand, evaluate, and build upon the reported research be available at the time of peer review and publication. Additionally, authors should make available software that has a significant impact on the research. This entails:

1. Depositing the data and software in a trusted repository, as appropriate, and preferably with a DOI
2. Including an [Availability Statement](#) as a separate paragraph in the Open Research section explaining to the reader where and how to access the data and software
3. And including [citation\(s\)](#) to the deposited data and software, in the Reference Section.

Click on the headings below for detailed information on:

- [Models & Simulations](#)
- [Journal-Specific Data Guidance](#)
- [International Geo Sample Numbers](#)

Most of your questions regarding data and software should be answered by the resources below. Just in case, if you still have questions, you can contact DataHelp@agu.org.

WHAT DATA NEEDS TO BE AVAILABLE?

Primary and processed data used for your research should be preserved and made available. Generally, the underlying data are considered to be the types of data usually preserved in domain repositories for each discipline. These may include raw data, but are usually the processed or refined data that support and lead to the described results and allow other readers to assess your conclusions and build off your work.

In your paper, cite these data, as well as any data you used from other sources, and include information about access to the data in the availability statement. For **model or simulation data**, follow [journal specific guidance](#) on prioritizing preserved output; in general, availability of software is most important.

Very large data (greater than 1 terabyte or TB) can be a challenge to preserve as there often fees and additional resources required. One option to consider, institutions often offer solutions for data preservation and compliance. Again, refer to the [journal specific guidance](#) for more information or email DataHelp@agu.org.



Fernando Pérez @fperez_org · 15h



Many thanks [@libcce](#) [@danielskatz](#) and team for organizing this excellent, timely session!

Now that some ground work about the value of these tools is done, I look forward to finding ways of making them smoothly running parts of the system for a variety of use cases.



Chris Erdmann @libcce · 16h

Advancing Notebooks as Scholarly Objects #AGU21 #DataHelpDesk
@ESIPfed @theAGU @EarthCube @EarthDataHelp @HughShanahan's
slides on Survey of the use cases for notebooks
doi.org/10.5281/zenodo... 1/n

[Show this thread](#)

A satellite image of Earth from space, showing the eastern coast of North America and the Atlantic Ocean. The text "AGU Fall Meeting 2022" is overlaid in white.

AGU Fall Meeting 2022

Thank you.



Chris Erdmann

Assistant Director, Data Leadership
American Geophysical Union
0000-0003-2554-180X
@libcce | cerdmann@agu.org

Shelley Stall

Senior Director, Data Leadership
American Geophysical Union
0000-0003-2926-8353
@ShelleyStall | sstall@agu.org