



a data editor's view of the state of data and
software citation in astronomy



The Pencil Code

a high-order finite-difference code for compressible MHD



topics

- ❖ who are the AAS Journal data editors?
- ❖ what are current models for software review and citation?
- ❖ where can open data linked to journal articles live (or die)?
- ❖ what would a review for “reproducibility” look like?

Data/Software Keywords:	Data/software review
<p>This manuscript contained the following Data/Software keywords that may require review:</p> <p>github.com, sourceforge.net</p>	<p>1) We recommend that living code on github repositories (e.g. Be-synthesis-with-moog) place a "frozen" version on Zenodo (or other 3rd party repositories that issue DOIs) and then cite them in the article. A tutorial on how to do this is available here:</p> <p>https://github.com/AASJournals/Tutorials/tree/master/Repositories</p> <p>[Edit]</p>

data review at submission: a quick(ish) review of >90% of all manuscripts

- run scripts to identify linked code repositories;
- review links for remote, unarchived data/code/figures;
- request data be included or archived & linked to final article;
- review tables, figures & animations for size or accessibility;
- submit data/code recommendations to scientific editor for review.

example: animations (dominates our time)

Data/Software Review

Editor:

Please direct the author to the AAS Journal's Graphics Guide: animations are no longer supplemental material. Per the Guide, the author should modify the videos and improve the text/caption descriptions of the animations. More information and an example edit for Figure 5 can be found at these links:

<https://journals.aas.org/graphics-guide/#animations>

<https://authortools.aas.org/AAS21849/>

PA: I put a pretty good concat of the 4 Figure 5 animations at the above link.

[\[Edit\]](#)

example: data problem (published w/o fix)

Data/software review

The manuscript says that the catalog as described in Table 2 will be "released on-line". Can the authors be more specific in the revised manuscript? Do they mean with the published article in machine readable format? If on an external site we recommend a copy also be placed in a 3rd party repository that issues DOIs like Zenodo. A tutorial on how to set this up is available here:

<https://github.com/AASJournals/Tutorials/blob/master/Repositories/UsingRepositories.md>

example: software citations (fixes of var types)

Data/software review

It is great the authors are using \software to highlight the code used in the manuscript but citation should be given when known, e.g. Scipy (Jones et al. 2001), emcee (Foreman-Mackey et al. 2013), CASA (v5.1.1; McMullin et al. 2007), MIRIAD (Sault, Teuben & Wright M., 1995), & radiobear (de Pater et al. 2019).

Data/Software Review

Editor:

Even though the version they use is a modified GOTPM code, the original developers deserve attribution via citation: <https://ui.adsabs.harvard.edu/abs/2004NewA....9..111D/abstract> . For the purposes of reproducibility the text would best list more details of the modified version of GOTPM, e.g., <https://astro.kias.re.kr/~kjhan/GOTPM/index.html>.



example: (no cost) data request

Data/software review

The authors should consider providing the spectra shown in Figure 1 as the Data behind the Figure (DbF). The readers would surely appreciate having this important data set available with the published article.

example: accessibility (mainly informative)

Data/Software Review

Editor: The authors should begin using tools such as Color Oracle to check their figures for color accessibility. In a number of Figures color is used as the sole discriminator of point data while different symbols are better. Red/Green points or solid lines are especially inaccessible.

PA: At submission the author's wrote (and we should followup on at PA):

After acceptance, I would like to provide most of the figures in machine-readable format (ascii or fits files), so that other scientists can easily retrieve this extensive data set and play with it. The data set is extremely rich, and allows for

full range of possible parameterizations for the systematics that we investigated are detailed in Section~\ref{sec_parameterization}. We used the `\textsc{batman}` **package** to model the transit \citep{2015PASP..127.1161K} using a linear limb-darkening prescription.

374-

375: We used the Python MCMC **package** `\textsc{emcee}` to fit the parameters describing the transit and the systematics simultaneously \citep{2013PASP..125..306F}. We fitted each observation separately. We used the **routines** of \citep{2010PASP..122..935E} to convert the calendar dates in the headers of each frame of each observation to BJD_{TDB} .

376-

377: Figure~\ref{whitelight_corrected} shows the white lightcurves for each observation after removing

at post-acceptance: data editing

- 15-20% of accepted manuscripts enter post-acceptance data editing;
 - tables are standardized; interactive/animated materials edited;
- data repositories or PID requested for some raw or submitted materials;
 - run scripts to pick out code mentions, check for references;
- request authors archive and cite their codes & find preferred citations.



The Pencil Code

a high-order finite-difference code for compressible MHD



models for software citation

- Journal article or Journal like citation;
 - Indexed in a Code Library;
- Direct Citation of Software (by version/release).

OPEN ACCESS

The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package*

The Astropy Collaboration, A. M. Price-Whelan¹ , B. M. Sipőcz^{4,4}, H. M. Günther², P. L. Lim³, S. M. Crawford⁴ , S. Conseil⁵, D. L. Shupe⁶, M. W. Craig⁷, N. Dencheva³ [+ Show full author list](#)

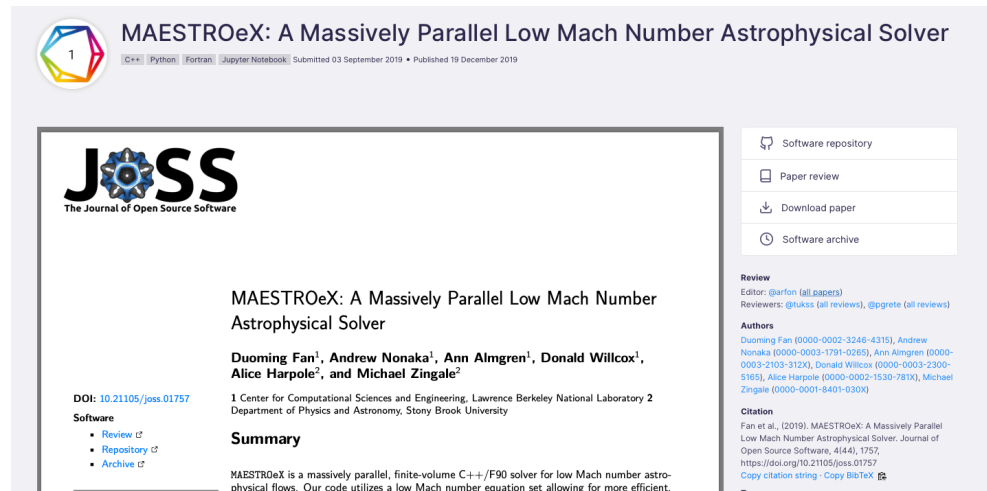
Published 2018 August 24 • © 2018. The American Astronomical Society.

[The Astronomical Journal](#), [Volume 156](#), [Number 3](#)



models for software citation: Journal article

- Publish a proxy for the software via a “Journal article.”
- Examples: Astropy Collaboration et al. (2013, 2018)
- Accumulates new authors/developers via new publication;
- “Astronomy” Journals do not directly review the software.
- <https://www.astropy.org/acknowledging.html>



The screenshot shows the article page for "MAESTROeX: A Massively Parallel Low Mach Number Astrophysical Solver" on the JOSS website. The page includes the JOSS logo, the article title, authors (Duoiming Fan, Andrew Nonaka, Ann Almgren, Donald Willcox, Alice Harpole, and Michael Zingale), and a summary. It also features a sidebar with links to the software repository, paper review, download paper, and software archive. The article is published in the Journal of Open Source Software, 4(44), 1757, with a DOI of 10.21105/joss.01757.

models for software citation: Journal “like” article

- The Journal of Open Source Software: <https://joss.theoj.org/>
- Scripted, detailed peer-review of code;
- Accumulates new authors/developers via new publication;
- A “science” submission to the AAS Journals can have a parallel “software” peer-review.
 - <https://doi.org/10.3847/1538-4357/ab4f75>
 - <https://doi.org/10.21105/joss.01757>



The screenshot shows the ASCL record for the 'Pencil' code. On the left is a sidebar with navigation links: 'VIEW', 'Abstract' (selected), 'Citations (9)', 'References', 'Co-Reads', 'Similar Papers', 'Volume Content', 'Graphics', 'Metrics', and 'Export Citation'. The main content area displays the title 'Pencil: Finite-difference Code for Compressible Hydrodynamic Flows', a 'Show affiliations' button, and the authors 'Brandenburg, Axel; Dobler, Wolfgang'. The abstract text describes the code as a high-order finite-difference code for compressible hydrodynamic flows with magnetic fields, designed for massively parallel shared- or distributed-memory computers. It mentions typical scientific targets like driven MHD turbulence, convection in a slab, accretion disc turbulence, and dust particle evolution. The code is written in well-commented Fortran90. At the bottom, the publication is listed as 'Astrophysics Source Code Library, record ascl:1010.060'. On the right, there are sections for 'FULL TEXT SOURCES' (with a 'Publisher' link) and 'ASSOCIATED WORKS (2)' (with a 'Described in Source Software' link).

models for software citation: Code Index

- List your software in the Astrophysics Source Code Library (ASCL);
- Provides links between the source codebase and documentation.
- Stores the “preferred” citation for a developer team;
- No versioning but author list can be modified at any time for any reason.
- ASCL Record for Pencil: <https://ascl.net/1010.060>

December 16, 2018 Software Open Access

pencil-code/pencil-code: v2018.12.16

Axel Brandenburg; aghah; wdboller; Philippe Bourdin; Sven Binger; mrheinhardt; Nils Erland L. Haugen; Antony Mee; Natalia Babikova; Tobias Heinemann; Boris Dintrans; Frederick Gent; Dhruvadiya MITRA; Simon Candelaesi; Jörn Warnecke; Petri Kärpälä; Andreas Schreiber; Jonas Krueger; Jørgen Røysland Aarnes; Piyali Chatterjee; Xiangyu Li; Graeme R Sarson; Joshua Boyd; Wladimir Lyra; rplasson; simo-tuomisto; alexanderhubbard; Luiz Felipe S. Rodrigues; Gustavo Guerrero; asnodin

This is the current version; see the "Scientific Usage of the Pencil Code" under <http://pencil-code.nordita.org/highlights/> for the currently over 400 publications using the code. The code has received 29,502 commits since 2001. The latest version is available under <https://github.com/pencil-code>

38 views 2 downloads [See more details...](#)

Available in **GitHub** Indexed in **OpenAIRE**

Publication date: December 16, 2018
DOI: [10.5281/zenodo.2314922](https://doi.org/10.5281/zenodo.2314922)
Related identifiers: Supplement to <https://github.com/pencil-code/pencil-code/tree/v2018.12.16>

Preview

pencil-code-v2018.12.16.zip

The previewer is not showing all the files

File	Size
pencil-code-pencil-code-4824f05	1.4 kB
• .gitignore	365 Bytes
• .travis.yml	1.7 kB
• README.md	1.7 kB
• bin	
• README	3.8 kB
• auto-test	54.0 kB
• build-testcase-cvs	2.6 kB
• combine_videosfiles	9 Bytes
• copy-snapshots	7.3 kB
• copy-snapshots_exp	6.5 kB
• cvs-add-rundir	1.3 kB
• cvsci_run	2.6 kB
• cvsci_run_bash	2.3 kB
• dx_get_dataidr	392 Bytes
• dx_list_varfiles	715 Bytes
• dx_table2csv	408 Bytes

models for software citation: Direct Citation of Software

- Citation of a digital object representing a persistent version of software
- Github is not considered a preservation platform.
- Examples:
 - Developer export code to a separate platform (e.g., [Zenodo](#))
 - Code is automatically archived (e.g., [Software Heritage](#))



December 16, 2018 Software Open Access

pencil-code/pencil-code: v2018.12.16

Axel Brandenburg; ajohari; wdoeller; Philippe Bourdin; Sven Binner; mrheinhardt; Nils Erlund L. Haugen; Antony Mee; Natalia Babikova; Tobias Heinemann; Boris Dintrans; Frederick Gent; Dhruvadiya MITRA; Simon Candelaesi; Jörn Warnecke; Petri Kärpälä; Andreas Schreiber; Jonas Krueger; Jørgen Røysland Aarnes; Piyali Chatterjee; Xiangyu Li; Graeme R Sarson; Joshua Boyd; Vladimir Lyra; rplasson; simo-tuomisto; alexanderhubbard; Luiz Felipe S. Rodrigues; Gustavo Guerrero; asnodin

This is the current version; see the "Scientific Usage of the Pencil Code" under <http://pencil-code.nordita.org/highlights/> for the currently over 400 publications using the code. The code has received 29,502 commits since 2001. The latest version is available under <https://github.com/pencil-code>

38 views 2 downloads [See more details...](#)

Available in  Indexed in 

Publication date:
December 16, 2018
DOI:
[DOI: 10.5281/zenodo.2314922](https://doi.org/10.5281/zenodo.2314922)
Related identifiers:
Supplement to
<https://github.com/pencil-code/pencil-code/tree/v2018.12.16>

Preview

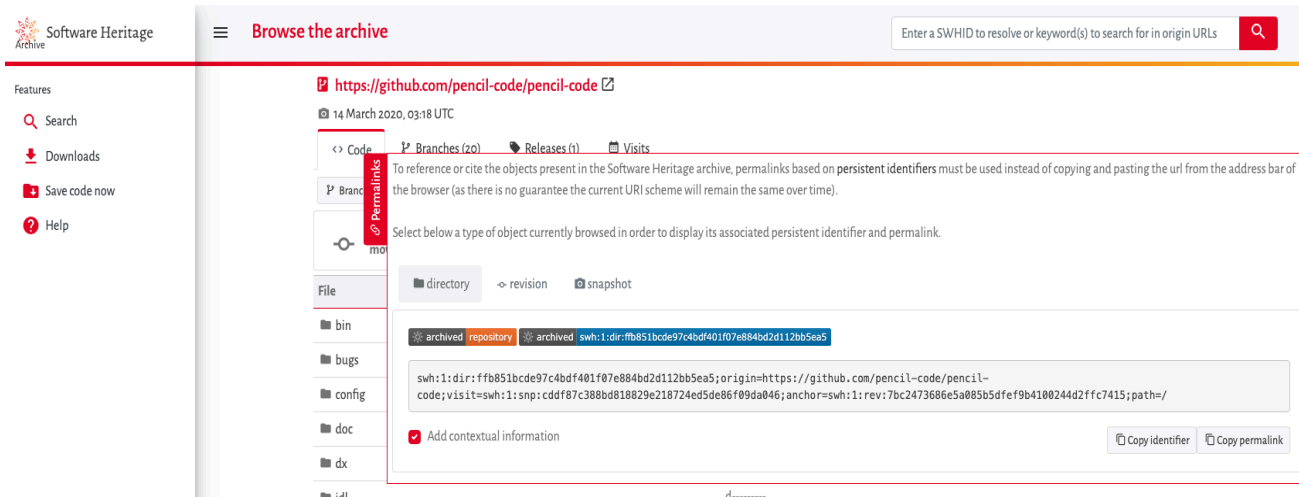
pencil-code-v2018.12.16.zip

The preview is not showing all the files

File	Size
pencil-code-pencil-code-4824f05	1.4 kB
• .gitignore	365 Bytes
• travis.yml	1.7 kB
• README.md	1.7 kB
• bin	
• README	3.8 kB
• auto-test	54.0 kB
• build-testcase-cvs	2.6 kB
• combine_videos	9 Bytes
• copy-snapshots	7.3 kB
• copy-snapshots_exp	6.5 kB
• cvs-add-rundir	1.3 kB
• cvsci_run	2.6 kB
• cvsci_run_bash	2.3 kB
• dx_get_data_dir	392 Bytes
• dx_list_varfiles	715 Bytes
• dx_table2csv	408 Bytes

models for software citation: Direct Citation of Software

- Zenodo archiving requires developers to:
 - Use versioning as a proxy for citation;
 - Review and maintain all the Zenodo object metadata for each version;
- ADS Indexes any cited version of a Zenodo software object:
 - [doctype:software and bibstem:"zndo"](#)
- Latest Release (v2018.12.16): <https://doi.org/10.5281/zenodo.2314922>



models for software citation: Direct Citation of Software

- Software Heritage automatically (not quite) indexes all software repositories
- Citations are granular (down to the commit);
- Citations have no authorship model (yet) nor can they be indexed in ADS.
- https://archive.softwareheritage.org/browse/origin/directory/?origin_url=https://github.com/pencil-code/pencil-code



models for software citation: Pencil Code

- There is no preferred citation model on the Pencil landing pages
- A user who goes to ASCL finds the ASCL entry is the “preferred” citation;
 - The Zenodo entry has never been cited and thus is not indexed in ADS;
- What does the Pencil Code community desire from these software citation models?



The Pencil Code

a high-order finite-difference code for compressible MHD



models for data citation and reproducibility: Pencil Code

- Examples of data citation: [Zenodo AAS Community](#), [MAST](#), [MESA inlists](#)
- What could the Pencil Code Community imagine from a reproducibility review?

