EPhysLab

CIGUS Network
Research Centres of the
Galician University System

Campus Auga. Ourense
Universidade Vigo

GNU

Scientific method and scientific publication in the age of AI

Replicability, code and data sharing

Juan A. Añel[1,2] and Robert Pinder[3]

[1]EPhysLab, CIM-UVigo, Universidade de Vigo, Ourense, Spain
[2]GNU Project
[3]PSE Healthy Energy, Oakland, CA, USA

Santiago de Compostela, 28 October 2026

Computational reproducibility

> Others should obtain the same results with the code and data of a study. This requires code and data that are available, documented and preserved over time. The scientific method relies on it.

Deposited code stops running

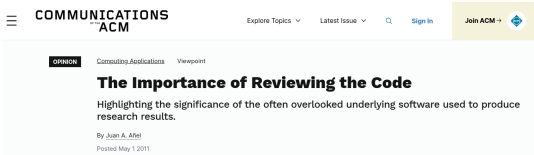
- Dependencies change.
- Computing environments are lost.
- Model input/output is too large to archive.

Specific problems from AI models

- Training is not deterministic.
- Training data are not available.
- Proprietary software/models.

Code and data must be in proper repositories

- All the code and data behind a paper should be deposited in a trusted repository.
- Each deposit needs a persistent identifier, such as a DOI, and a fixed version.
- **Origin:** the deposit records which version produced the published results.
- **Review:** referees and readers can inspect the code, not only the text.
- **Availability:** the material is still there years later, independent of personal or project web pages.

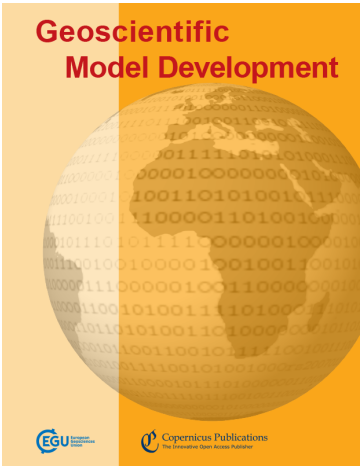


Añel (2011), Commun. ACM 54, 40–41



Añel (2025), Nature 641, 39

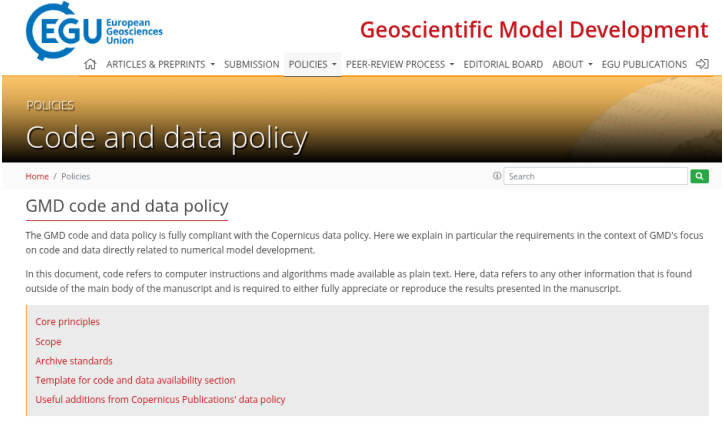
Geoscientific Model Development



A Case Study

- Open-access journal of the European Geosciences Union.
- Public code is required for model description papers since 2013.
- The policy has been revised several times since then.
- We look at it from our editorial work in the journal.

GMD code and data policy



gmd.copernicus.org/policies/code_and_data_policy.html

Editorials

v1.0 (2013)
v1.1 (2015)
v1.2 (2019)

- Code and data archived in long-term repositories.
- A “Code and data availability” section in every paper.

Four tiers of criteria

A **Availability.** Are the code and data available?

B **Completeness.** Does the deposit contain everything needed?

C **Portability.** Can the code run outside the author's environment?

D **Reproducibility.** When the code runs, do the results match the paper?

Corpus. 3,212 GMD papers, 2017–2026.

Tools. GEMMA4, an open multimodal language model, and a coding agent for tier D.

Result of each check. **PASS**, **FAIL** or **OPEN**.

Validation on 100 papers reviewed by hand: 99% agreement for tier A, about 75% for B and C.

A: Availability

> Is there a code and data availability statement? Do the links resolve to available deposits? Is at least one deposit the model code? Does running it require paid licences?

A0 Every cited deposit is listed. “Data provided on request” fails.

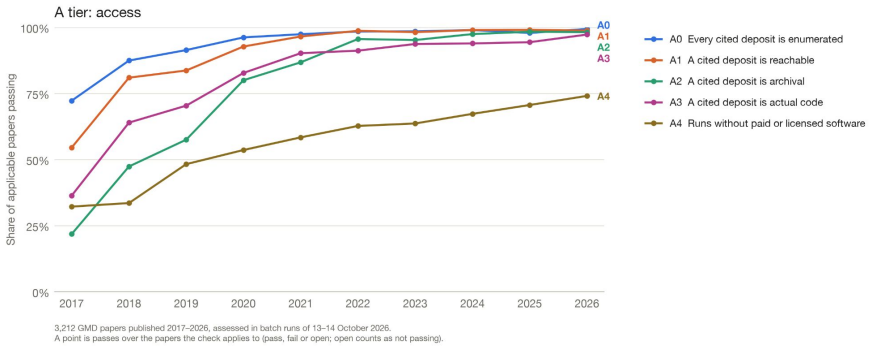
A1 At least one cited deposit can be reached.

A2 At least one deposit is archival: DOI, persistent and immutable. A GitHub link alone is not enough.

A3 At least one deposit is the model code, judged from its metadata.

A4 The main code runs without paid or licensed software such as MATLAB or IDL.

A tier: access



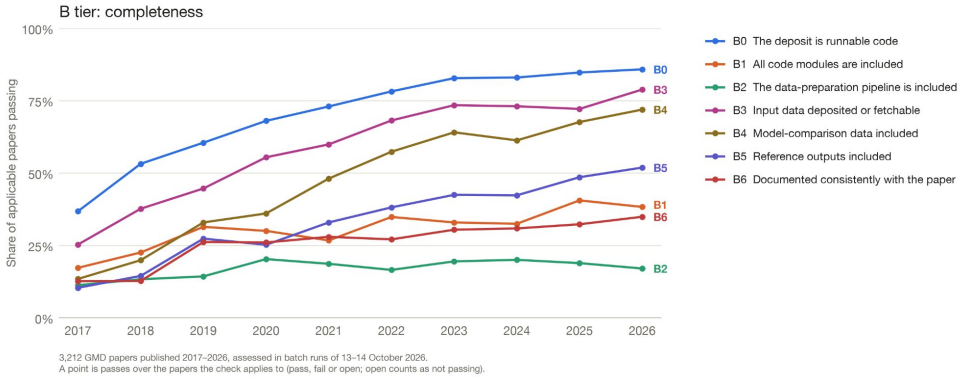
A4 is not a journal requirement. Even so, the share of papers whose code runs on open software keeps growing.

B: Completeness

> Is everything needed to run the code in the deposit? This covers the authors' code, libraries and dependencies, configuration, input data, and data to compare with for verification.

B0 The deposit contains runnable code, not only PDFs, figures or tables.
B1 All referenced modules, headers, scripts and functions are included or are standard packages.
B2 The data preparation steps described in the paper map to code in the deposit.
B3 All input data are deposited or can be downloaded from public repositories.
B4 The data used to evaluate the model are included or public.
B5 Reference outputs are provided to check that the model was run and configured correctly.
B6 The configuration in the deposit matches the one described in the paper.

B tier: completeness



C: Portability

> Can the model run independently of the authors' computing environment? Is there documentation on how to run it? Are dependencies available, library versions still usable, and the code free of machine-specific parts?

C1 A README or similar explains what to run, in what order, to produce each result.

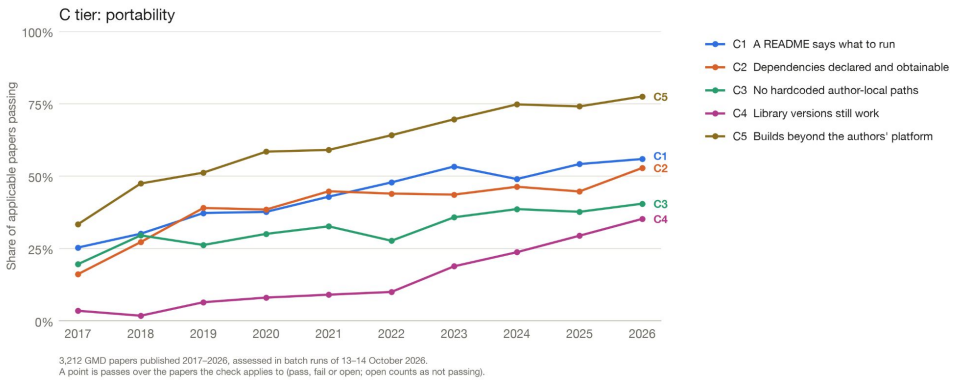
C2 Dependencies and build files are declared: a Makefile, a requirements.txt, etc.

C3 No hardcoded author-local paths such as `/home`, `/Users` or `D:/WORK`.

C4 The library and API versions used still work in current environments.

C5 Any binaries can be rebuilt from source on other platforms.

C tier: portability



D: Reproducibility

A coding agent tries to

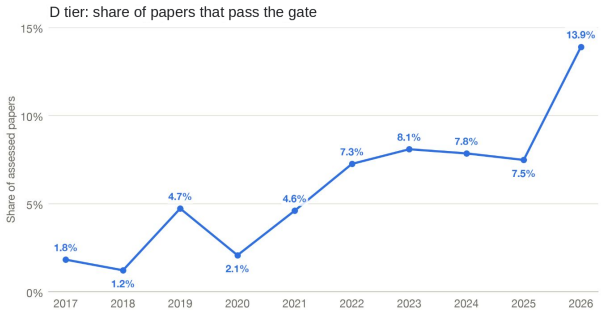
- **D1** build the executable,
- **D2** run it,
- **D3** produce output in the correct format,
- **D4** match the verification data or the claims of the paper.

> **Gate.** Tier D is only attempted on papers that pass A3, A4, B0, B1, B3, B5, C1 and C2.

≈8%

of all papers pass these checks and the D tier

D tier: papers that pass the gate



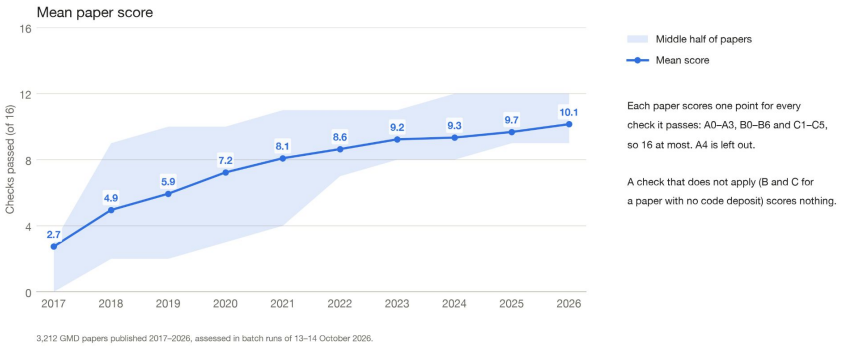
3,212 GMD papers published 2017–2026, assessed in batch runs of 13–14 October 2026.
204 papers pass the gate in all. Gate as of 12 October 2026.

A paper passes the gate when all of these hold:

- It is a modelling paper
- A3 pass: a cited deposit is code
- The code was actually read
- A4 pass: no paid or licensed software
- B0 not fail: runnable code
- B1 pass or open: all modules included
- B3 pass or n/a: input data available
- B5 pass: reference data, not only figures
- C1 pass: a README says what to run
- C2 pass: dependencies declared

So far, nearly all papers that pass the gate (204 in total) run and reproduce the claimed results.

Mean paper score



Published papers become more reproducible over time. The mean score rises from 2.7 to 10.1 out of 16.

Adapting to AI

- Requesting repositories for training and output data.
- Requesting model weights.
- Requesting trained models.
- Use of free software and free models.

Adapting to AI

- Requesting repositories for training and output data.
- Requesting model weights.
- Requesting trained models.
- Use of free software and free models.

> **In an AI era remember,**
> **results not understood or reproducible are not science**

Summary

1. A journal policy works where it sets explicit requirements. Almost all GMD papers pass checks A0 to A3 since 2022.
2. Completeness and portability lag behind. Data preparation code and working library versions are often missing.
3. Full reproduction is still possible for a small share of papers, about 14% of those published in 2026.
4. An open-weight large language model agrees with human review for availability. For completeness and portability it can support editors but not replace them yet.

References

Añel, J. A. (2011). The importance of reviewing the code. *Commun. ACM* 54, 40–41. doi:10.1145/1941487.1941502

Añel, J. A. (2025). Preserve access to crucial climate data. *Nature* 641, 39. doi:10.1038/d41586-025-01328-0

GMD Executive Editors (2013). Editorial: The publication of geoscientific model developments v1.0. *Geosci. Model Dev.* 6, 1233–1242. doi:10.5194/gmd-6-1233-2013

GMD Executive Editors (2015). Editorial: The publication of geoscientific model developments v1.1. *Geosci. Model Dev.* 8, 3487–3495. doi:10.5194/gmd-8-3487-2015

GMD Executive Editors (2019). Editorial: The publication of geoscientific model developments v1.2. *Geosci. Model Dev.* 12, 2215–2225. doi:10.5194/gmd-12-2215-2019

Añel, J. A., García-Rodríguez, M. and Rodeiro, J. (2021). Current status on the need for improved accessibility to climate models code. *Geosci. Model Dev.* 14, 923–934. doi:10.5194/gmd-14-923-2021

García-Rodríguez, M., Rodeiro-Iglesias, J. and Añel, J. A. (2026). Code accessibility and code quality across phases of the models of the Coupled Model Intercomparison Project. *Geosci. Model Dev.* 19, 7687–7704. doi:10.5194/gmd-19-7687-2026

Añel, J. A. (2017). Comment on “Most computational hydrology is not reproducible, so is it really science?” by Hutton et al. *Water Resour. Res.* 53, 2572–2574. doi:10.1002/2016WR020190

Thank you

Juan A. Añel · j.anhel@uvigo.gal
EPhysLab, CIM-UVigo, Universidade de Vigo, Ourense

Funded by the Xunta de Galicia (GRC-ED431C 2025/37).