

A process for independent reproduction of computations underlying research

Eglen, S., & Nüst, D. (2019). CODECHECK: An open-science initiative to facilitate the sharing of computer programs and results presented in scientific publications. *Septentrio Conference Series*, (1). <https://doi.org/10.7557/5.4910>



Open Reproducible Research

Reproducible research refers to achieving the same results (e.g., tables, figures, numbers) as reported in the paper by using the same source code and data. In **Open reproducible research**, these materials are publicly accessible.

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. (2016). What does research reproducibility mean? *Science Translational Medicine* 8(341), 341ps12-341ps12.
<https://doi.org/10.1126/scitranslmed.aaf5027>

The CODECHECK philosophy

- Reproducibility tools like Code Ocean and Binder set the bar high by making code reproducible for everyone.
- CODECHECK simply asks: “Was the code reproducible once for someone else?”
- CODECHECK checks if the code runs and generates the expected number of output files.
- The contents of those output files are checked visually and available for others to see.
- The validity of the code is not checked.

CODECHECK process in four steps

Step 1: Provide me with access to code and data underlying the results in the manuscript.

CODECHECK process in four steps

Step 1: Provide me with access to code and data underlying the results in the manuscript.

Step 2: Reproducibility attempt: Executable? Same results?

CODECHECK process in four steps

Step 1: Provide me with access to code and data underlying the results in the manuscript.

Step 2: Reproducibility attempt: Executable? Same results?

Step 3: Create certificate, including how I ran the code, what was checked, and a copy of the outputs (plus differences), recommendations.

CODECHECK process in four steps

Step 1: Provide me with access to code and data underlying the results in the manuscript.

Step 2: Reproducibility attempt: Executable? Same results?

Step 3: Create certificate, including how I ran the code, what was checked, and a copy of the outputs (plus differences), recommendations.

Step 4: Publication of certificate on Zenodo and a badge for your code repository.

CODE WORKS 

Example

For more examples, check: <https://codecheck.org.uk/>



The image shows a CODECHECK certificate. At the top left is the ITC logo, and at the top right is the University of Twente logo. The text "CODECHECK certificate" is centered. Below it is a large green badge with "CODE CHECK" and a white checkmark, with the URL "https://codecheck.org.uk/" underneath. The main text of the certificate reads: "Reviewed paper: Egor Prikaziuk, Peiqi Yang, Christiaan van der Tol (2021): Google Earth Engine Sentinel-3 OLCI Level-1 Dataset Deviates from the Original Data: Causes and Consequences. *Remote Sensing*. 2021; 13(6):1098. <https://doi.org/10.3390/rs13061098>

To cite this report, use: Markus Konkol (2021, July): CODECHECK of *Google Earth Engine Sentinel-3 OLCI Level-1 Dataset Deviates from the Original Data: Causes and Consequences*. DOI: [10.5281/zenodo.5106408](https://doi.org/10.5281/zenodo.5106408)

Summary: The reproduction was successful. The authors followed good scientific practice and

<https://doi.org/10.5281/zenodo.5106408>

Sentinel-3 images download

This script automates the per-point search and download of ESA Sentinel-3 images.

An interactive documentation can be found on <https://s3-loader.readthedocs.io/>

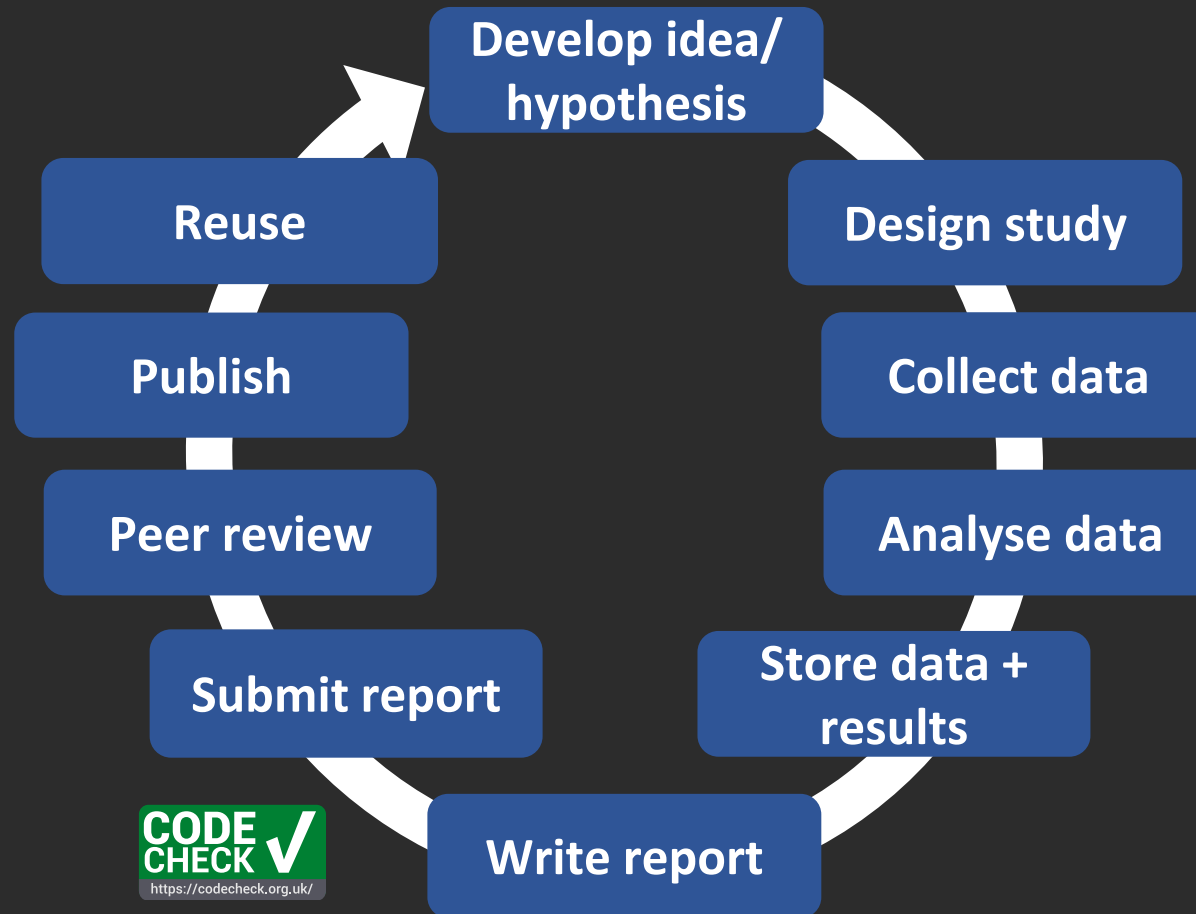
A CODECHECK certificate is available confirming that the computations underlying the article could be independently reproduced: **CODE WORKS** ✓

CODE AND DATA AVAILABILITY

Our code is openly available at the GitHub repository <https://github.com/tedinburgh/causality-review> and Ref 45. The data that support the findings of this study are openly available at the same repository. A CODECHECK certificate is available confirming that the computations underlying this article could be independently executed: <https://doi.org/10.5281/zenodo.4720843>.

Existing open-access code for some indices include repositories for information theory and transfer entropy: IDTxI⁴² v1.1, PyIF⁴⁶; and for convergent cross mapping: pyEDM⁴⁴ v1.7.4. We also adapted fuzzy *c*-means code based on Ref⁴⁷. We checked our results for transfer entropy and convergent cross mapping against those from IDTxI and pyEDM respectively. All code in our repository and in these others is Python.

When is the best time to CODECHECK?



How to prepare for a CODECHECK

At least: just write a bit of documentation.

A bit more? create a so-called Research Compendium

- Create a folder that includes all materials necessary to reproduce the results
 - Manuscript
 - Data
 - Source code
 - Ideally in a notebook format (e.g., R Markdown, Jupyter notebook)
- Add a license!

Further general remarks

- Everyone at ITC can request a CODECHECK
- Effort you need to invest is mainly about communication
- Unsuccessful CODECHECKs have no implications (“safe space”)
- Issues in the code will not damage your work
- Synthetic datasets can be used in case of privacy concerns
- Also possible with licensed software (if I can get a license)
- In case of big data, data subsets can be used to reduce computation time
- It is possible to have a CODECHECK after publication

Get started

- For more information, visit <https://www.itc.nl/research/open-science/codecheck/>
- You want to start your first CODECHECK? Drop me a message: m.konkol@utwente.nl