José Antonio Bayona Viveros
School of Earth Sciences

ORCID: https://orcid.org/0000-0002-3026-0327

# Advancing Earthquake Predictability Research with Open Science Practices

submitted for the category of

Widening Reach

in the

Open Research Prize 2022

## SUMMARY

The Collaboratory for the Study of Earthquake Predictability (CSEP) has set an international standard for transparent, reproducible and prospective evaluation (i.e., against future earthquakes) of seismicity forecasts to alleviate the crisis of untestable predictions and unverifiable claims of success in earthquake predictability research. To elevate this standard, I provide data, documentation and code to fully reproduce the results of a prospective evaluation of earthquake forecasting models for California that were developed nearly a decade ago. This effort includes a repository with forecasts in Zenodo, a reproducibility package freely available on GitHub, and an open access journal publication.


## What did you do?

I prospectively evaluated the abilities of 22 seismicity models to forecast earthquakes with magnitude larger than 5 in California over the past decade. For this task, I used a set of existing and new methods from the pyCSEP software toolkit, a Python package that contains statistical tests, forecast visualisation routines and earthquake-catalogue processing utilities for evaluating seismicity forecasts (pyCSEP is presented in the Journal of Open Source Software https://doi.org/10.21105/joss.03658, and freely available at https://github.com/sceccode/pycsep). Furthering CSEP's principle of open science, I designed a reproducibility package, i.e., a structured set of data, code and other artefacts, to ensure the reproducibility, transparency and replicability of my test results. This reproducibility package is linked to a citable data repository in Zenodo, from which forecasts and the evaluation earthquake catalogue can automatically be downloaded. As of February 2022, 490 downloads have been made (see https://doi.org/10.5281/zenodo.5141567). The code is also freely available on GitHub at https://github.com/bayonato89/Reproducibility-hybrids, so users can easily run it by following the instructions listed on the documentation file and typing only three lines of code. Finally, I published this investigation in a peer-reviewed journal under an open access licence to reach a wider audience (see https://doi.org/10.1093/gji/ggac018).

## Why did you do it?

In retrospective evaluations, hybrid (or ensemble) seismicity models showed great predictive skills. However, only prospective forecast evaluations can be considered rigorous enough to objectively describe their capacity to forecast earthquakes, and therefore I conducted this study. The results of this prospective evaluation are contrary to the outcomes of retrospective analyses, justifying the need of this long-term open research. Additionally, I designed a reproducibility package to demonstrate how to create replicable results, build confidence among researchers in earthquake forecasting models, and elevate the standards of earthquake prediction research.

## How did you do it?

Hybrid earthquake forecasting models were created by a co-author of this work nearly ten years ago and stored within CSEP servers for prospective testing. To assess them, I created code and documentation using Python and Jupyter notebooks, which are open-source software tools. I then incorporated several open-source software best-practices reported in the literature to store the earthquake forecasts and evaluation catalogue in Zenodo, and the model software on GitHub. In this manner, all the figures presented within the manuscript can be recreated in a couple of hours (using a computer with an i7, 10$^{th}$ generation processor).

## What barriers / challenges did you have to overcome?

Traditionally, earthquake forecasts, software and input data were stored on CSEP servers that provided no community access to ensure the integrity of the experiments. Although beneficial, this controlled environment also made it difficult to share test results. Therefore, the first barrier to overcome was the implementation of reproducibility packages as a solution for data storage at CSEP, since this "change in research culture" from a tightly controlled testing environment to easily accessible forecasts and evaluations not only implies possible scrutiny by the scientific community, but also potential misuse of the results. The second major challenge for my open science goals was the programming barrier, as I needed to gain some experience working with Shell scripts and reproducibility packages to create a software product that was transparent, computationally cheap, and easy to use.

## What does it mean for you and your research?

It means a lot, because now I am ready to continue doing open science research and can train other researchers within the CSEP community to carry out and improve reproducibility practices in our field. In fact, several members of CSEP, including researchers at the University of Bristol, GFZ Potsdam, and the Southern California Earthquake Center, have embraced these practices, as discussed in a recently submitted community manuscript on the pyCSEP toolkit, which also includes a reproducibility package. Thus, I think that this concept of reproducibility packages will be the backbone for future CSEP forecast experiments and hopefully set a new community standard for open science.

## How might your findings / approach help other researchers?

The results of our prospective evaluation show that the hybrid earthquake forecasting models did not outperform their common individual "parent" model, which is based on smoothing the locations of small earthquakes. These results indicate that small-magnitude seismicity is useful in mapping the location of future larger earthquakes in California. Commonly, small earthquakes have been neglected in Probabilistic Seismic Hazard Analyses, because large earthquakes, although infrequent, account for most of the seismic energy released worldwide. However, there is increasing evidence, including this research, that small earthquakes can be informative in forecasting the occurrence of large damaging events. Hence, I believe that this publication and its reproducibility package could be beneficial for other researchers to formalise and test their hypotheses/models about the seismogenesis, explore further techniques to form ensemble earthquake models, and study potential relationships among geophysical datasets that could improve our understanding of the earthquake generation process.

## Additional Information

Open access publication:

Geophysical Journal International

*Geophys. J. Int.* (2022) **00,** 1 https://doi.org/10.1093/gji/ggac01

GJI Seismology

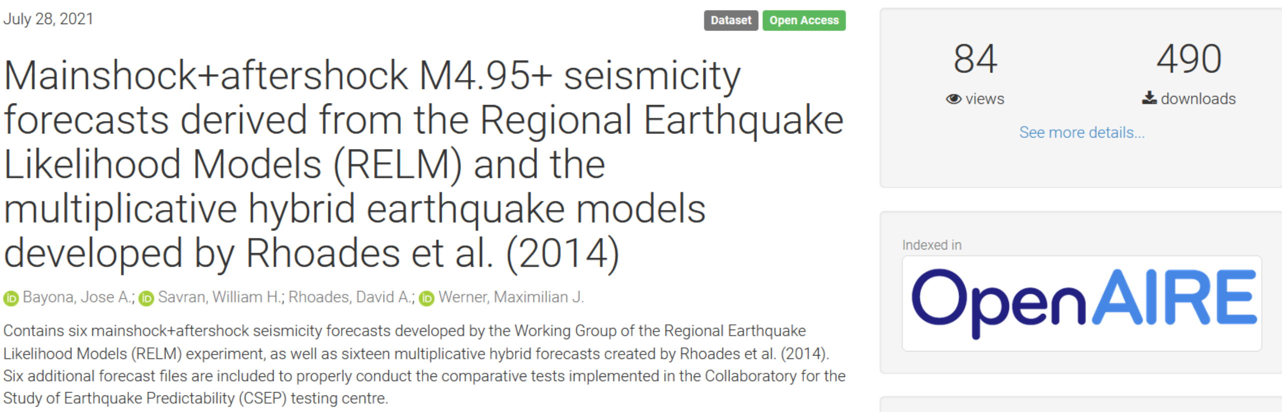
**Prospective evaluation of multiplicative hybrid earthquake forecasting models in California**

J.A. Bayona,[1] W.H. Savran,[2] D.A. Rhoades[3] and M.J. Werner[1]

[1]*School of Earth Sciences, University of Bristol, Queens Road, BS81QU, Bristol, UK. E-mail: jose.bayona@bristol.ac.uk*
[2]*Southern California Earthquake Center, University of Southern California, Los Angeles, CE 90089-0742, USA*
[3]*GNS Science, 1 Fairway Drive, Avalon 5010 PO Box 30-368, Lower Hutt 5040, New Zealand*

Accepted 2022 January 14. Received 2021 November 21; in original form 2021 July 2

Data repository:



July 28, 2021 Dataset Open Access

Mainshock+aftershock M4.95+ seismicity forecasts derived from the Regional Earthquake Likelihood Models (RELM) and the multiplicative hybrid earthquake models developed by Rhoades et al. (2014)

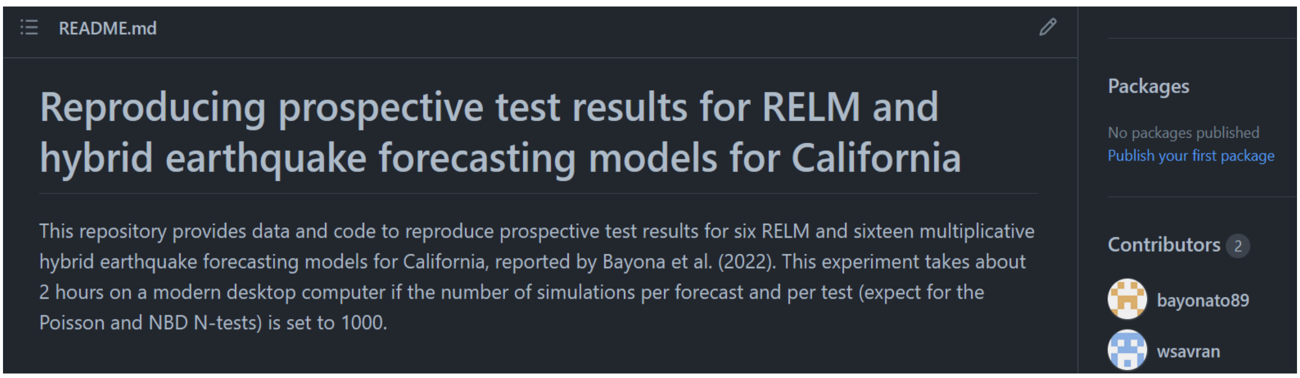
Bayona, Jose A.; Savran, William H.; Rhoades, David A.; Werner, Maximilian J.

Contains six mainshock+aftershock seismicity forecasts developed by the Working Group of the Regional Earthquake Likelihood Models (RELM) experiment, as well as sixteen multiplicative hybrid forecasts created by Rhoades et al. (2014). Six additional forecast files are included to properly conduct the comparative tests implemented in the Collaboratory for the Study of Earthquake Predictability (CSEP) testing centre.

84 views 490 downloads
See more details...

Indexed in
OpenAIRE

Reproducibility package:



README.md

Reproducing prospective test results for RELM and hybrid earthquake forecasting models for California

This repository provides data and code to reproduce prospective test results for six RELM and sixteen multiplicative hybrid earthquake forecasting models for California, reported by Bayona et al. (2022). This experiment takes about 2 hours on a modern desktop computer if the number of simulations per forecast and per test (expect for the Poisson and NBD N-tests) is set to 1000.

Packages
No packages published
Publish your first package

Contributors 2
bayonato89
wsavran