The Journal of Open Source Software

pyCSEP: A Python Package For Earthquake Forecast Developers

William H. Savran*[1], Maximilian J. Werner[2], Danijel Schorlemmer[3], and Philip J. Maechling[1]

1 Southern California Earthquake Center, University of Southern California **2** University of Bristol **3** GFZ Potsdam

DOI: 10.21105/joss.03507

Software

- Review
- Repository
- Archive

Editor: Katy Barnhart

Submitted: 15 July 2021
Published: 26 July 2021

Summary

For government officials and the public to act on real-time forecasts of earthquakes, the seismological community needs to develop confidence in the underlying scientific hypotheses of the forecast generating models by assessing their predictive skill. For this purpose, the Collaboratory for the Study of Earthquake Predictability (CSEP) provides cyberinfrastructure and computational tools to evaluate earthquake forecasts. Here, we introduce pyCSEP, a Python package to help earthquake forecast developers embed model evaluation into the model development process. The package contains the following modules: (1) earthquake catalog access and processing, (2) data models for earthquake forecasts, (3) statistical tests for evaluating earthquake forecasts, and (4) visualization routines. `pyCSEP` can evaluate earthquake forecasts expressed as expected rates in space-magnitude bins, and simulation-based forecasts that produce thousands of synthetic seismicity catalogs. Most importantly, `pyCSEP` contains community-endorsed implementations of statistical tests to evaluate earthquake forecasts, and provides well defined file formats and standards to facilitate model comparisons. The toolkit will facilitate integrating new forecasting models into testing centers, which evaluate forecast models and prediction algorithms in an automated, prospective and independent manner, forming a critical step towards reliable operational earthquake forecasting.

Background

Successfully predicting the time, location, and size of future earthquakes would have immense societal value, and this quest underlies much of the research in seismology and earthquake geology. To date, however there have been no reliable earthquake predictions methods. An earthquake prediction is a deterministic statement about whether or not an earthquake will occur in a particular geographic region, time window, and magnitude range. On the other hand, an earthquake forecast provides the probability that such an earthquake will occur (Jordan et al., 2011). Most of the current research effort focuses on developing probabilistic earthquake forecasting models that encode empirical or physics-based hypotheses about the occurrence of seismicity. To what degree earthquakes can be predicted remains an open and important question.

As Schorlemmer et al. (2018) states "the fundamental idea of CSEP is simple in principle but complex in practice: earthquake forecasts should be tested against future observations to assess their performance, thereby ensuring an unbiased test of the forecasting power of a model." Practically, this requires a prospective evaluation of the earthquake forecasts. Prospective

*Corresponding author reachable at wsavran [at] usc [dot] edu

evaluation requires that model developers fully specify their models (with zero-degrees of freedom) before the experiment begins (Schorlemmer et al., 2018). Specific parameters for an experiment are determined through community consensus, such as the geographic testing region and magnitude range, authoritative data sets used to evaluate the forecasts, the evaluation metrics, and the precise specification of a forecast. These parameters are defined in full before the start of the experiment. This standardization ensures that any potential conscious or unconscious biases are reduced, because the evaluation data are collected after each model has been provided for evaluation.

Statement of need

Over the last decade, CSEP has led numerous prospective earthquake forecasting experiments (see, e.g., Michael & Werner, 2018). These experiments are formally conducted within testing centers (Schorlemmer & Gerstenberger, 2007) that contain the software required to autonomously run and evaluate earthquake forecasts in a fully prospective mode. The software design emphasized a carefully controlled computing and software environment which ensured integrity of testing results (Zechar et al., 2009). However, its monolithic software design made it difficult for researchers to utilize various routines in the testing centers in their own work without replicating the entire testing center configuration on their own system. In addition, software was developed by a single developer, leading to personnel risk and a lack of opportunities for others to contribute directly.

`pyCSEP` was designed to provide vetted methods to evaluate earthquake forecasts in a Python package that researchers can include directly in their research. The statistical tests and tools to evaluate earthquake forecasts are required by all model developers, and greatly benefit from open-source development practices by providing standardized, well-tested, and community-reviewed software tools. At the time of publication, `pyCSEP` has been used for two published articles (Bayona et al., 2020; Savran et al., 2020), and is being used by several research groups participating in the Real-time earthquake risk reduction for a resillient Europe (RISE) project and others.

pyCSEP Overview

`pyCSEP` provides an open-source implementation of peer-reviewed statistical tests developed for evaluating probabalistic earthquake forecasts (Rhoades et al., 2011; Savran et al., 2020; Schorlemmer et al., 2007; Werner et al., 2011; Zechar et al., 2013). In addition, `pyCSEP` provides routines for working with earthquake catalogs and visualizations. The core design of `pyCSEP` includes classes that represent earthquake forecasts, catalogs, and various spatial regions. Higher level functions are implemented using these classes to provide routines for common tasks in evaluating earthquake forecasts.

Earthquake forecasts can either be specified as expected earthquake rates over discrete space-magnitude-time regions (Schorlemmer et al., 2007) or as families of synthetic earthquake catalogs with each catalog representing a realization from the underlying stochastic model (e.g., Savran et al., 2020). Earthquake catalogs are row-based data sets that contain features of an earthquake. At a minimum, an earthquake must be defined by its geographical location (latitude, longitude), origin time, and magnitude. In addition, `pyCSEP` provides classes for working directly with forecasts from the Uniform California Earthquake Rupture Forecast with Epidemic-type Aftershock Sequences Version 3 (Field et al., 2017). `pyCSEP` also provides classes for interacting with earthquake catalogs and performing operations on them, such as filtering and binning events on the space-magnitude grids needed for evaluation. `pyCSEP` also includes numerous plotting utilities that interface directly with `matplotlib` and `Cartopy` (Hunter, 2007; Met Office, 2010 - 2015). Space-magnitude regions facilite gridding operations that are necessary for evaluating earthquake forecasts. These objects model regular latitude, longitude cells where earthquakes can be aggregated for evaluation and visualization purposes. `pyCSEP` provides pre-defined spatial regions that have been used in previous experiments (Field, 2007; Taroni et al., 2018).

`pyCSEP` interfaces directly with popular numerical and plotting libraries such as `Numpy`, `matplotlib`, and `pandas` (Harris et al., 2020; Hunter, 2007; McKinney, 2010). Users already familiar with these librarys can adapt `pyCSEP` directly into their code. `pyCSEP` provides file-formats for forecasts and earthquake catalogs, and can allow users to specify custom filetypes.

Acknowledgements

This research was supported by the Southern California Earthquake Center (Contribution No. 11030). SCEC is funded by NSF Cooperative Agreement EAR-1600087 & USGS Cooperative Agreement G17AC00047. Maximilian J. Werner and Danijel Schorlemmer received funding from the European Union's Horizon 2020 research and innovation program (Number 821115, RISE: Real-Time Earthquake Risk Reduction for a Resilient Europe).

References

Bayona, J. A., Savran, W. H., Strader, A., Hainzl, S., Cotton, F., & Schorlemmer, D. (2020). Two global ensemble seismicity models obtained from the combination of interseismic strain measurements and earthquake-catalogue information [Journal Article]. *Geophysical Journal International*, *224*(3), 1945–1955. https://doi.org/10.1093/gji/ggaa554

Field, E. H. (2007). Overview of the working group for the development of Regional Earthquake Likelihood Models (RELM) [Journal Article]. *Seismological Research Letters*, 1–10. https://doi.org/https://doi.org/10.1785/gssrl.78.1.7

Field, E. H., Jordan, T. H., Page, M. T., Milner, K. R., Shaw, B. E., Dawson, T. E., Biasi, G. P., Parsons, T., Hardebeck, J. L., Michael, A. J., Weldon, R. J., Powers, P. M., Johnson, K. M., Zeng, Y. H., Felzer, K. R., Elst, N. van der, Madden, C., Arrowsmith, R., Werner, M. J., & Thatcher, W. R. (2017). A synoptic view of the third Uniform California Earthquake Rupture Forecast (UCERF3) [Journal Article]. *Seismological Research Letters*, *88*(5), 1259–1267. https://doi.org/10.1785/0220170045

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(03), 90–95.

Jordan, T. H., Chen, Y. T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., & Papadopoulos, G. (2011). Operational earthquake forecasting. State of knowledge and guidelines for utilization [Journal Article]. *Annals of Geophysics*. https://doi.org/10.4401/ag-5350

McKinney, Wes. (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt & Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56–61). https://doi.org/10.25080/Majora-92bf1922-00a

Met Office. (2010 - 2015). *Cartopy: A cartographic python library with a matplotlib interface*. https://scitools.org.uk/cartopy

The Journal of Open Source Software

Michael, A. J., & Werner, M. J. (2018). Preface to the focus section on the Collaboratory for the Study of Earthquake Predictability (CSEP): New results and future directions [Journal Article]. *Seismological Research Letters*, *89*(4), 1226–1228. https://doi.org/10.1785/0220180161

Rhoades, D. A., Schorlemmer, D., Gerstenberger, M. C., Christophersen, A., Zechar, J. D., & Imoto, M. (2011). Efficient testing of earthquake forecasting models [Journal Article]. *Acta Geophysica*, *59*(4), 728–747. https://doi.org/10.2478/s11600-011-0013-5

Savran, W. H., Werner, M. J., Marzocchi, W., Rhoades, D. A., Jackson, D. D., Milner, K., Field, E. H., & Michael, A. J. (2020). Pseudoprospective evaluation of UCERF3-ETAS forecasts during the 2019 Ridgecrest sequence [Journal Article]. *Bulletin of the Seismological Society of America*, *110*(4), 1799–1817. https://doi.org/10.1785/0120200026

Schorlemmer, D., & Gerstenberger, M. C. (2007). RELM testing center [Journal Article]. *Seismological Research Letters*, *78*(1), 30–36. https://doi.org/10.1785/gssrl.78.1.30

Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D., & Rhoades, D. A. (2007). Earthquake likelihood model testing [Journal Article]. *Seismological Research Letters*, *78*(1), 17–29. http://srl.geoscienceworld.org/content/78/1/17.extract

Schorlemmer, D., Werner, M. J., Marzocchi, W., Jordan, T. H., Ogata, Y., Jackson, D. D., Mak, S., Rhoades, D. A., Gerstenberger, M. C., Hirata, N., Liukis, M., Maechling, P. J., Strader, A., Taroni, M., Wiemer, S., Zechar, J. D., & Zhuang, J. C. (2018). The collaboratory for the study of earthquake predictability: Achievements and priorities [Journal Article]. *Seismological Research Letters*, *89*(4), 1305–1313. https://doi.org/10.1785/0220180053

Taroni, M., Marzocchi, W., Schorlemmer, D., Werner, M. J., Wiemer, S., Zechar, J. D., Heiniger, L., & Euchner, F. (2018). Prospective CSEP evaluation of 1-day, 3-month, and 5-yr earthquake forecasts for Italy [Journal Article]. *Seismological Research Letters*, *89*(4), 1251–1261. https://doi.org/10.1785/0220180031

Werner, M. J., Helmstetter, A., Jackson, D. D., & Kagan, Y. Y. (2011). High-resolution long-term and short-term earthquake forecasts for California [Journal Article]. *Bulletin of the Seismological Society of America*, *101*(4), 1630–1648. https://doi.org/10.1785/0120090340

Zechar, J., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P. J., & Jordan, T. H. (2009). The collaboratory for the study of earthquake predictability perspective on computational earthquake science [Journal Article]. *Concurrency and Computation: Practice and Experience*, *22*(12), 1836–1847. https://doi.org/10.1002/cpe.1519

Zechar, J., Schorlemmer, D., Werner, M. J., Gerstenberger, M. C., Rhoades, D. A., & Jordan, T. H. (2013). Regional earthquake likelihood models i: First-order results [Journal Article]. *Bulletin of the Seismological Society of America*, *103*(2A), 787–798. https://doi.org/10.1785/0120120186