

floatCSEP: An application to deploy and conduct reproducible prospective earthquake forecasting experiments

Pablo Iturrieta                    <img alt="ORCID icon" data-bbox="584

experiments across diverse geographic regions, such as California, New Zealand, Italy, Japan, and globally (Bayona et al., 2021, 2022, 2023; Eberhard et al., 2012; Field, 2007; Gerstenberger & Rhoades, 2010; Iturrieta et al., 2024; Nanjo et al., 2011; Strader et al., 2018; Taroni et al., 2014, 2018; Tsuruoka et al., 2012; Werner et al., 2010; Zechar et al., 2013). These efforts have substantially advanced scientific rigor and established community standards in earthquake forecasting research, thereby contributing to better forecasts and seismic hazard assessments (e.g., Michael & Werner, 2018; Schorlemmer et al., 2018).

Statement of Need

Despite their achievements, the original CSEP Testing Centers used centralized, rigid software architectures that tightly coupled data management to local hardware, limiting reuse, scalability, and community engagement (e.g., Savran, Bayona, et al., 2022; Schorlemmer et al., 2018; Zechar et al., 2010). As noted by Mizrahi et al. (2024), there is broad consensus in the earthquake forecasting community that transparency and reproducibility are essential in forecast testing. However, due to the complexity of Testing Centers, independent researchers often require advanced technical expertise to access, reproduce, and analyze CSEP experiments. To overcome these limitations, the Python package pyCSEP (Graham et al., 2024; Savran, Bayona, et al., 2022; Savran, Werner, et al., 2022) was developed to provide core forecast evaluation routines that can be directly integrated into modelers' workflows.

However, pyCSEP alone lacks key features required to deploy and conduct entire forecasting experiments, such as interfacing with external models' source code, automating catalog access, managing data, standardizing workflow execution, and generating summary reports. Therefore, additional software is required that provides Testing Center capabilities while remaining decoupled from specific hosting hardware. To meet this need, we developed floatCSEP, which manages the entire experiment lifecycle, from model integration and initial deployment to the incremental updating of input data, forecasts, results, and reports as new observations become available. It is intended for earthquake forecast model developers, institutions that run CSEP-style forecasting experiments, and the broader statistical seismology community. To our knowledge, no existing software provides this complete end-to-end testing workflow; commonly used seismology tools instead address only individual steps, such as catalog queries (e.g., ObsPy, Beyreuther et al. (2010)) or forecast evaluation (e.g., pyCSEP).

Software Overview

The primary objective of floatCSEP is to provide a portable, automated, and reproducible Testing Center environment that runs on a standard workstation (subject to experiment-specific runtime and memory demands) and scales to HPC systems for computationally intensive experiments, such as time-dependent models producing daily forecasts. Experiments are defined through human-readable YAML (yaml.org) configuration files, which are processed through a simple command-line interface to ensure ease of use even for users without extensive computational expertise. This declarative approach simplifies the experiment setup, standardizes its workflow (Figure 1) and enhances its reproducibility.

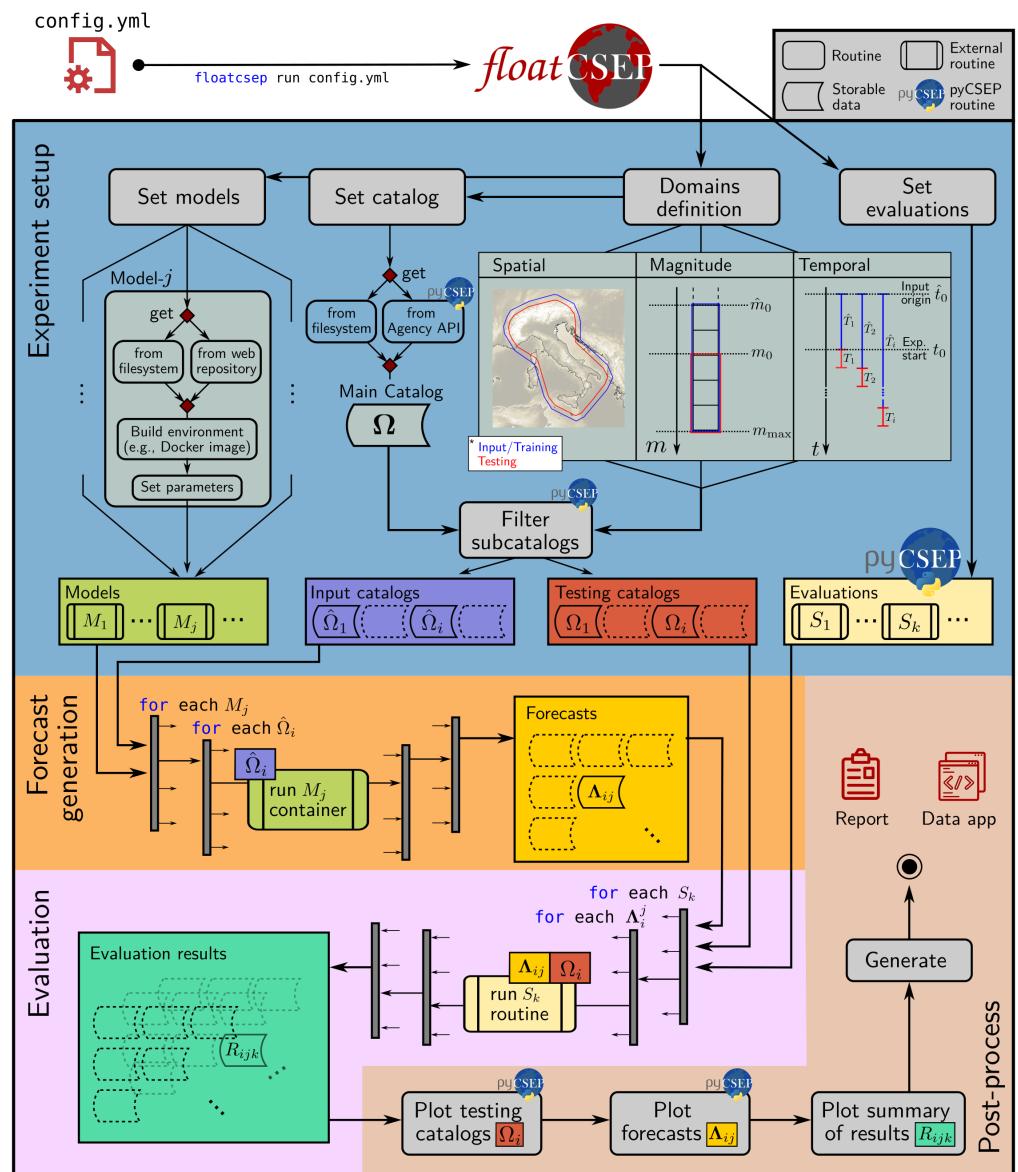


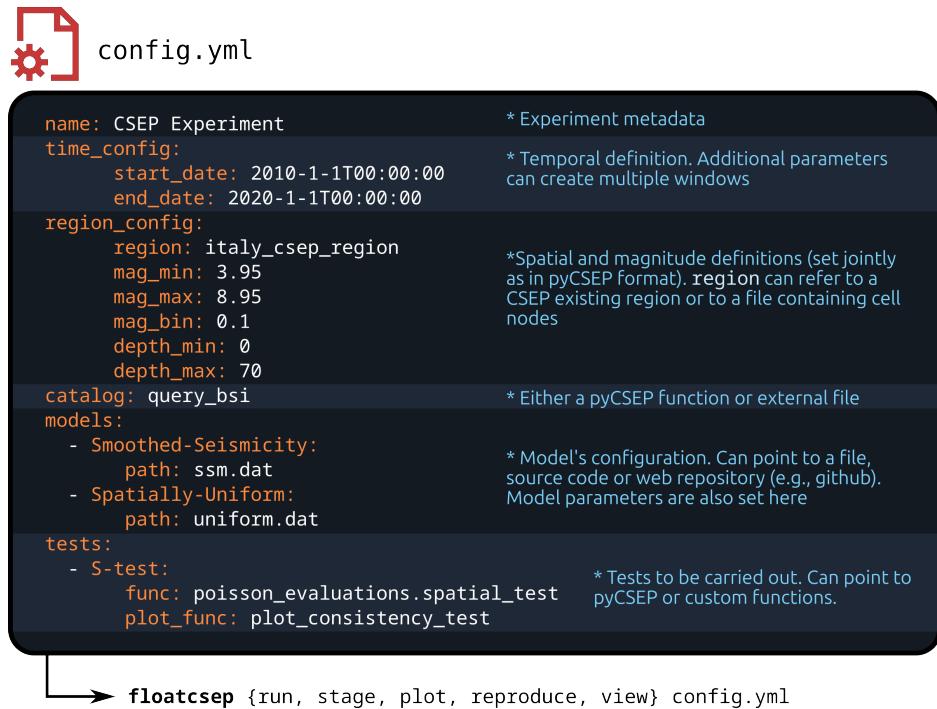
Figure 1: Workflow diagram of floatCSEP for a time-dependent experiment, which roughly consists of: 1) Defining time-space-magnitude ranges and discretizations for forecast generation and evaluation; 2) Querying and filtering earthquake catalogs (both for model input and evaluation); 3) Building the source code of external models, configuring its parameters and input data; 4) Generating forecasts by running each model's source code in a containerized environment; 5) Performing forecast evaluations and comparisons using pyCSEP's or user-implemented testing metrics; and 6) Producing reports including test results and visual representations.

floatCSEP uses pyCSEP as a dependency, incorporating its core functionality (forecast and catalog classes, and evaluation routines) alongside additional Testing Center operations, such as data management and computational containerization using Docker (Merkel, 2014). The application supports multiple forecast formats and accommodates both time-invariant and time-variant experiments. It handles forecasts produced either by models managed directly by floatCSEP or provided externally through raw files. Representative use cases are included as tutorials, which users can extend by incorporating new models, adding alternative evaluations, or replicating in different regions and timeframes. The software integrates

seamlessly with pyCSEP's existing testing routines and provides custom hooks for user-defined tests, visualizations, and reports.

Example Use

An example configuration file (`config.yml`) for a time-invariant, grid-based experiment in Italy with two models is shown in [Figure 2](#).



```

config.yml

name: CSEP Experiment          * Experiment metadata
time_config:
    start_date: 2010-1-1T00:00:00
    end_date: 2020-1-1T00:00:00
region_config:
    region: italy_csep_region
    mag_min: 3.95
    mag_max: 8.95
    mag_bin: 0.1
    depth_min: 0
    depth_max: 70
catalog: query_bsi             * Either a pyCSEP function or external file
models:
    - Smoothed-Seismicity:
        path: ssm.dat
    - Spatially-Uniform:
        path: uniform.dat
tests:
    - S-test:
        func: poisson_evaluations.spatial_test
        plot_func: plot_consistency_test      * Tests to be carried out. Can point to pyCSEP or custom functions.

floatcsep {run, stage, reproduce, view} config.yml

```

Figure 2: Simplest example of a configuration file for a time-invariant, grid-based experiment with two models. The run command executes the experiment end-to-end, generating and testing all missing forecasts from the last execution up to today. The stage command accesses and builds the models' source code, as well as preparing input and testing catalogs. The reproduce command re-executes an experiment and compares it with existing results using statistical and computational metrics. plot executes the post-process, image generation, and reporting of the experiment. view deploys a data app for visualizing the forecasts, catalogs, and test results.

Applications

floatCSEP is designed to support the following applications:

- Deploy and conduct new prospective experiments that incrementally incorporate new data, update forecasts, and provide evaluation results. While the CSEP community plans to use floatCSEP for new (official) experiments, we also encourage independent researchers to adopt floatCSEP to prospectively evaluate their models.
- Reproduce the results of completed prospective CSEP experiments within a containerized computational environment (e.g., [Iturrieta et al., 2024](#)).
- Create new retrospective or pseudo-prospective experiments for their easy reproduction and shareability.
- Plug new models into a completed or ongoing (float)CSEP experiment. Since CSEP experiments are clearly defined, they can be effectively used as benchmarks for comparing and developing new forecasting models (e.g., [Serafini et al., 2025](#)).

- Support continuous evaluation of Operational Earthquake Forecasting systems that provide authoritative, near-real-time forecasts (e.g., [Jordan et al., 2011](#); [Mizrahi et al., 2024](#)). However, most systems generate forecasts for overlapping windows (e.g., weekly forecasts updated daily) and evaluating the overall performance of such forecast collections remains an open methodological challenge (e.g., [Brehmer et al., 2025](#)).

floatCSEP contributes to a growing CSEP software ecosystem that, together with reproducibility packages (e.g., [Allison et al., 2018](#); [Bayona et al., 2022, 2023](#); [Graham et al., 2024](#); [Iturrieta et al., 2024](#); [Savran, Bayona, et al., 2022](#)), open-source forecasting models (e.g., [Mizrahi et al., 2023](#)), and long-term open-science repositories, could lay the foundation for building robust, collaborative benchmarks in earthquake forecasting.

Acknowledgements

Details of author contributions can be found in [CREDITS.md](#) in the code repository. The development of this software was supported and funded in part by (i) the European Commission under project Geo-INQUIRE <https://www.geo-inquire.eu/>, number 101058518 within the HORIZON-INFRA-2021-SERV-01 call; (ii) the European Union H2020 program, Grant number 821115, Real-time earthquake risk reduction for a reSilent Europe (RISE, <http://www.rise-eu.org/home/>); (iii) the Statewide California Earthquake Center (Contribution No. 14983). SCEC is funded by NSF Cooperative Agreement EAR-2225216 and USGS Cooperative Agreement G24AC00072-00; (iv) the U.S. Geological Survey Earthquake Hazards Program under Grant Nos. G24AP00059 and G25AP00379; and (v) the Leverhulme Trust through its Early Career Fellowship program.

References

- Allison, D. B., Shiffrin, R. M., & Stodden, V. (2018). Reproducibility of research: Issues and proposed remedies. *Proceedings of the National Academy of Sciences*, 115(11), 2561–2562. <https://doi.org/10.1073/pnas.1802324115>
- Bayona, J. A., Savran, W. H., Iturrieta, P., Gerstenberger, M. C., Graham, K. M., Marzocchi, W., Schorlemmer, D., & Werner, M. J. (2023). Are regionally calibrated seismicity models more informative than global models? Insights from California, New Zealand, and Italy. *The Seismic Record*, 3(2), 86–95. <https://doi.org/10.1785/0320230006>
- Bayona, J. A., Savran, W. H., Rhoades, D. A., & Werner, M. J. (2022). Prospective evaluation of multiplicative hybrid earthquake forecasting models in California. *Geophysical Journal International*, 229(3), 1736–1753. <https://doi.org/10.1093/gji/ggac018>
- Bayona, J. A., Savran, W. H., Strader, A., Hainzl, S., Cotton, F., & Schorlemmer, D. (2021). Two global ensemble seismicity models obtained from the combination of interseismic strain measurements and earthquake-catalogue information. *Geophysical Journal International*, 224(3), 1945–1955. <https://doi.org/10.1093/gji/ggaa554>
- Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). ObsPy: A Python toolbox for seismology. *Seismological Research Letters*, 81(3), 530–533. <https://doi.org/10.1785/gssrl.81.3.530>
- Brehmer, J. R., Kraus, K., Gneiting, T., Herrmann, M., & Marzocchi, W. (2025). Enhancing the statistical evaluation of earthquake forecasts—an application to Italy. *Seismological Research Letters*, 96(3), 1966–1988. <https://doi.org/10.1785/0220240209>
- Eberhard, D. A., Zechar, J. D., & Wiemer, S. (2012). A prospective earthquake forecast experiment in the western Pacific. *Geophysical Journal International*, 190(3), 1579–1592. <https://doi.org/10.1111/j.1365-246X.2012.05548.x>

- Field, E. H. (2007). Overview of the working group for the development of regional earthquake likelihood models (RELM). *Seismological Research Letters*, 78(1), 7–16. <https://doi.org/10.1785/gssrl.78.1.7>
- Geller, R. J. (1997). Earthquake prediction: A critical review. *Geophysical Journal International*, 131(3), 425–450. <https://doi.org/10.1111/j.1365-246X.1997.tb06588.x>
- Gerstenberger, M. C., & Rhoades, D. A. (2010). New Zealand earthquake forecast testing centre. *Pure and Applied Geophysics*, 167(8), 877–892. <https://doi.org/10.1007/s00024-010-0082-4>
- Graham, K. M., Bayona, J. A., Khawaja, A. M., Iturrieta, P., Serafini, F., Biondini, E., Rhoades, D. A., Savran, W. H., Maechling, P. J., Gerstenberger, M. C., & others. (2024). New features in the pyCSEP toolkit for earthquake forecast development and evaluation. *Seismological Research Letters*, 95(6), 3449–3463. <https://doi.org/10.1785/0220240197>
- Iturrieta, P., Bayona, J. A., Werner, M. J., Schorlemmer, D., Taroni, M., Falcone, G., Cotton, F., Khawaja, A. M., Savran, W. H., & Marzocchi, W. (2024). Evaluation of a decade-long prospective earthquake forecasting experiment in Italy. *Seismological Research Letters*, 95(6), 3174–3191. <https://doi.org/10.1785/0220230247>
- Jordan, T. H. (2006). Earthquake predictability, brick by brick. *Seismological Research Letters*, 77(1), 3–6. <https://doi.org/10.1785/gssrl.77.1.3>
- Jordan, T. H., Chen, Y.-T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., Papadopoulos, G., Sobolev, G., Yamaoka, K., & Zschau, J. (2011). Operational earthquake forecasting: State of knowledge and guidelines for utilization. *Annals of Geophysics*, 54(4), 315–391. <https://doi.org/10.4401/ag-5350>
- Kagan, Y. Y. (1994). Observational evidence for earthquakes as a nonlinear dynamic process. *Physica D: Nonlinear Phenomena*, 77(1-3), 160–192. [https://doi.org/10.1016/0167-2789\(94\)90132-5](https://doi.org/10.1016/0167-2789(94)90132-5)
- Kagan, Y. Y., & Jackson, D. D. (1994). Long-term probabilistic forecasting of earthquakes. *Journal of Geophysical Research: Solid Earth*, 99(B7), 13685–13700. <https://doi.org/10.1029/94JB00500>
- Merkel, D. (2014). Docker: Lightweight Linux containers for consistent development and deployment. *Linux Journal*, 2014(239), 76–91.
- 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. *Seismological Research Letters*, 89(4), 1226–1228. <https://doi.org/10.1785/0220180161>
- Mizrahi, L., Dallo, I., Elst, N. J. van der, Christoffersen, A., Spassiani, I., Werner, M. J., Iturrieta, P., Bayona, J. A., Iervolino, I., Schneider, M., & others. (2024). Developing, testing, and communicating earthquake forecasts: Current practices and future directions. *Reviews of Geophysics*, 62(3), e2023RG000823. <https://doi.org/10.1029/2023RG000823>
- Mizrahi, L., Schmid, N., & Han, M. (2023). *Imizrahi/etas: ETAS with fit visualization* (Version 3.2). Zenodo. <https://doi.org/10.5281/zenodo.7584575>
- Nanjo, K., Tsuruoka, H., Hirata, N., & Jordan, T. (2011). Overview of the first earthquake forecast testing experiment in Japan. *Earth, Planets and Space*, 63(3), 159–169. <https://doi.org/10.5047/eps.2010.10.003>
- Savran, W. H., Bayona, J. A., Iturrieta, P., Asim, K. M., Bao, H., Bayliss, K., Herrmann, M., Schorlemmer, D., Maechling, P. J., & Werner, M. J. (2022). pyCSEP: A python toolkit for earthquake forecast developers. *Seismological Society of America*, 93(5), 2858–2870. <https://doi.org/10.1785/0220220033>
- Savran, W. H., Werner, M. J., Schorlemmer, D., & Maechling, P. (2022). pyCSEP: A

- python package for earthquake forecast developers. *Journal of Open Source Software*. <https://doi.org/10.21105/joss.03658>
- Schorlemmer, D., & Gerstenberger, M. (2007). RELM testing center. *Seismological Research Letters*, 78(1), 30–36. <https://doi.org/10.1785/gssrl.78.1.30>
- 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., & others. (2018). The Collaboratory for the Study of Earthquake Predictability: Achievements and priorities. *Seismological Research Letters*, 89(4), 1305–1313. <https://doi.org/10.1785/0220180053>
- Serafini, F., Bayona, J. A., Silva, F., Savran, W., Stockman, S., Maechling, P. J., & Werner, M. J. (2025). A benchmark database of ten years of prospective next-day earthquake forecasts in California from the Collaboratory for the Study of Earthquake Predictability. *Scientific Data*, 12(1), 1501. <https://doi.org/10.1038/s41597-025-05766-3>
- Strader, A., Werner, M. J., Bayona, J. A., Maechling, P., Silva, F., Liukis, M., & Schorlemmer, D. (2018). Prospective evaluation of global earthquake forecast models: 2 yrs of observations provide preliminary support for merging smoothed seismicity with geodetic strain rates. *Seismological Research Letters*, 89(4), 1262–1271. <https://doi.org/10.1785/0220180051>
- 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. *Seismological Research Letters*, 89(4), 1251–1261. <https://doi.org/10.1785/0220180031>
- Taroni, M., Zechar, J., & Marzocchi, W. (2014). Assessing annual global $M6+$ seismicity forecasts. *Geophysical Journal International*, 196(1), 422–431. <https://doi.org/10.1093/gji/ggt369>
- Tsuruoka, H., Hirata, N., Schorlemmer, D., Euchner, F., Nanjo, K. Z., & Jordan, T. H. (2012). CSEP testing center and the first results of the earthquake forecast testing experiment in Japan. *Earth, Planets and Space*, 64(8), 661–671. <https://doi.org/10.5047/eps.2012.06.007>
- Werner, M. J., Zechar, J. D., Marzocchi, W., & Wiemer, S. (2010). Retrospective evaluation of the five-year and ten-year CSEP-Italy earthquake forecasts. *Annals of Geophysics*, 53(3), 11–30. <https://doi.org/10.4401/ag-4840>
- Zechar, J. D., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P. J., & Jordan, T. H. (2010). The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science. *Concurrency and Computation: Practice and Experience*, 22(12), 1836–1847. <https://doi.org/10.1002/cpe.1519>
- Zechar, J. D., Schorlemmer, D., Werner, M. J., Gerstenberger, M. C., Rhoades, D. A., & Jordan, T. H. (2013). Regional earthquake likelihood models I: First-order results. *Bulletin of the Seismological Society of America*, 103(2A), 787–798. <https://doi.org/10.1785/0120120186>