Deliverable 8.7

D8.7 EU forecast testing center operational

Deliverable information	
Work package	[WP8 - Impact: Exploitation, dissemination and services for securing a demonstrable societal, economic and scientific impact of RISE]
Lead	[GFZ]
Authors	[Pablo Iturrieta (GFZ), William Savran (USC), Danijel Schorlemmer (GFZ)
Reviewers	[Warner Marzocchi (UNINA)]
Approval	[Management Board]
Status	[Final]
Dissemination level	[Public]
Will the data supporting this document be made open access? (Y/N)	Y
If No Open Access, provide reasons	
Delivery deadline	[28.02.2023]
Submission date	[28.02.2023]
Intranet path	[DOCUMENTS/DELIVERABLES/]

Summary

The goal of the Collaboratory for the Study of Earthquake Predictability (CSEP) is to advance research on earthquake predictability by rigorously testing earthquake forecast models. This is achieved through prospective evaluation experiments, which require defining a-priori experiment parameters, rules and authoritative data sources before conducting any evaluations. These experiments are facilitated by a testing center, which used to rely on centralized, over-designed servers, hindering the reusability and replicability of testing source code. CSEP testing experiments are being modernized into open experiments, namely Floating Experiments, a format that decentralizes testing and promotes best practices in open science. Here, we develop the feCSEP application that has the purpose of deploying a Floating Experiment. Instead of using a physical, centralized server, feCSEP facilitates the containerization of all the experiment's artifacts required to run it, which can be hosted in a open data repository (e.g. Zenodo). This can eventually be downloaded and easily run on any machine, as long as it has the computational capacities. Thus, this format ensures that all experiment artifacts are publicly available and can be reproduced by other researchers. In this deliverable we formally define a Floating Experiment and detail the features of feCSEP. We run experiments with multiple characteristics to demonstrate the feCSEP features and capabilities: (i) the forecasts from the 2020 European Seismic Hazard Model, (ii) testing Global Quadtree Models, (iii) the Italy 2010 Forecasting Experiment and (iv) the new Earthquake Forecasting Experiment for Italy. The rapid and rigorous deployment of experiments in this format can serve as a basis for future benchmark problems developed by CSEP.

Introduction

The Collaboratory for the Study of Earthquake Predictability (CSEP) has been working since 2006 to promote and enable the rigorous testing of earthquake forecast models by deploying testing experiments in multiple regions (Schorlemmer et al. 2018 and references therein). Starting in California at the Southern California Earthquake Center (SCEC), CSEP developed the first version of a software system for testing experiments, a scheme that was replicated and further developed for new experiments in Japan, New Zealand, Italy and the globe.

The design of the testing system requires thorough management of the experiment's artifacts to ensure its reproducibility. An experiment must fully specify the authoritative input data (e.g. catalogs), forecasts and evaluation results, while using procedures agreed upon by the experiment stakeholders (i.e. modeler, testers, host institutions). Originally, the experiment's data were acquired, processed and stored by a single hosting server and software system, which kept track of the computed results. However, this led to the development of an overly complicated complex system that has proven difficult to maintain or interact with. This centralized structure of the CSEP operations worked well for each individual experiment, but it didn't provide the flexibility needed for the reusability and replicability of the research code, nor the code availability to get more people involved. The codes, as used in the testing center software system, could not easily be extracted to help modelers or users to explore their models or generate test results.

To overcome the limitations of the previous system versions, the CSEP community released the pyCSEP software package/toolkit (Savran et al., 2022). It contains the core functions, routines and procedures to evaluate earthquake forecasts, while releasing the source code for the community to review and/or implement novel routines within a common testing stage. However, maintaining experiments on a testing server requires large time-scales with continuous evaluations, and during this process, reproducibility must be safeguarded. The package pyCSEP provides the essential features of forecast testing, but it is meant to be included into the modeler/tester's workflow rather than to deploy an experiment. However, an experiment has an additional complexity due to file management routines, specification of the computational artifacts for reproducibility, and delivering the results in a comprehensive way.

Within the RISE project, we expand these ideas into a novel concept of experiment design, namely Floating Experiments, which is manifested through the feCSEP software application. This application has the explicit purpose of seamlessly deploying a testing experiment without the need for advanced computational skills. Floating Experiments are motivated by the belief that anyone with access to sufficient computing power to run an experiment should be able to reproduce it. It decouples the experiment system code from the experiment files, where the experiment results can be reproduced with simple code instructions contained in a simple text file. While official institutions should still be responsible for running and maintaining experiments, Floating Experiments would allow any researcher to have transparent and open access to source codes, forecasts and evaluations results, or even to design experiments of their own. We expect this framework to improve earthquake forecasting research by providing benchmarks for the development of new models.

Floating Experiments

A Floating Experiment represents an idealized experiment implementation organized by CSEP, where the results from the experiment are recorded and shared with the community. To operate official experiments and the authoritative results of a Floating Experiment, CSEP would operate the required hardware on-demand and publish results to a publicly available repository. This lowers the effective run-time and computational burden of the testing system and allows the experiment to be run essentially anywhere, hence the name Floating Experiment.

Floating Experiments rely heavily on Digital Object Identifiers (DOI), obtained from open data repositories (e.g. Zenodo), that ensure the immutability of the results and the prolonged experiment's reproducibility. Every experiment release (e.g. for every catalog update) has assigned a DOI version, containing all the necessary artifacts to run the experiment. Instead of storing all the results in a physical testing center, we are creating a reproducibility package (Krafczyk et al., 2021) that contains all the elements of an experiment.

An individual Floating Experiment can be as simple as downloading the package and running it with a few commands on a suitable computer. It should be provided as a "turn-key" product that can be executed with a single command by anyone with sufficient computing power (or time) to run the experiments. A similar approach is used in Openquake (Pagani et al., 2014), which runs complex seismic hazard calculations from input data (e.g. forecasts) and a configuration file that contains the instructions to be executed. In practice, a Floating Experiment can be allocated in an official repository (Gitlab & Zenodo), cloned to a local machine, run, and results being committed/pushed back into the repository (Figure 1).

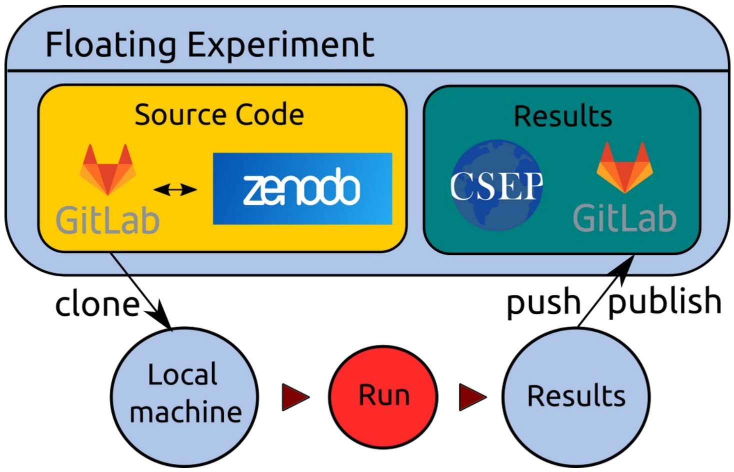

Figure 1. Scheme of how Floating Experiments are implemented.

The `feCSEP` application

The feCSEP application constitutes the system architecture of a testing experiment, curating the experiment's constituent artifacts. It satisfies the current open-source scientific standards (i.e. FAIR principles) as well as CSEP philosophy. The core functionality of feCSEP is based on the previously released package pyCSEP and has the following features:

- Manifest explicitly the experiments' rules and definitions: The system is flexible to define testing dates, regions, authoritative data, time-window, time-dependency, forecast formats, etc.
- Perform the fundamental tasks of a testing experiment, where the only limitation should be the computational time and storage limit. These tasks are:
  - Access, deploy, execute and store the competing forecast models and their results.
  - Access, retrieve and record the authoritative data set (e.g. ANSS Comcat, gCMT, INGV BSI)
  - Define a battery of tests, by interfacing with pycsep, and record their specifications (e.g. pseudo-random seeds, iterations, confidence intervals, etc.)
  - Evaluate the models' forecasts and store, register and summarize the results.
- Allow an Experiment to be run from scratch or from a previously executed state in the machine.
- Provide the book-keeping of numerous forecasts and result files/databases. It will contain the

required information to initialize an experiment from a middle-step, record the information for reproducibility, and be easily serialized or imported.

- Display the results comprehensibly, for example at the top-level of a git repository, which can be updated in every run. This will be extended to a static webserver/service (e.g. using Dash to visualize tests, forecasts, catalogs, etc.).

An additional difficulty to guarantee full reproducibility is to maintain the experiment's computational environment unmodified. The libraries and dependencies of the forecasting models (and the testing system as well) are being constantly updated, which may likely hinder the experiment to be executable in the future, even if was at its design. To address this, feCSEP allows the integration with Docker, which containerizes the computational environment of the experiment's and each forecasting model's. Thus, by using a simple text file, a researcher can re-instantiate the operational system, programming language, libraries and dependencies required to run the models and the experiment itself. A basic code diagram is given in Figure 2 that includes the basic classes and features of feCSEP. The application can be found in:

https://git.gfz-potsdam.de/csep-group/fecsep

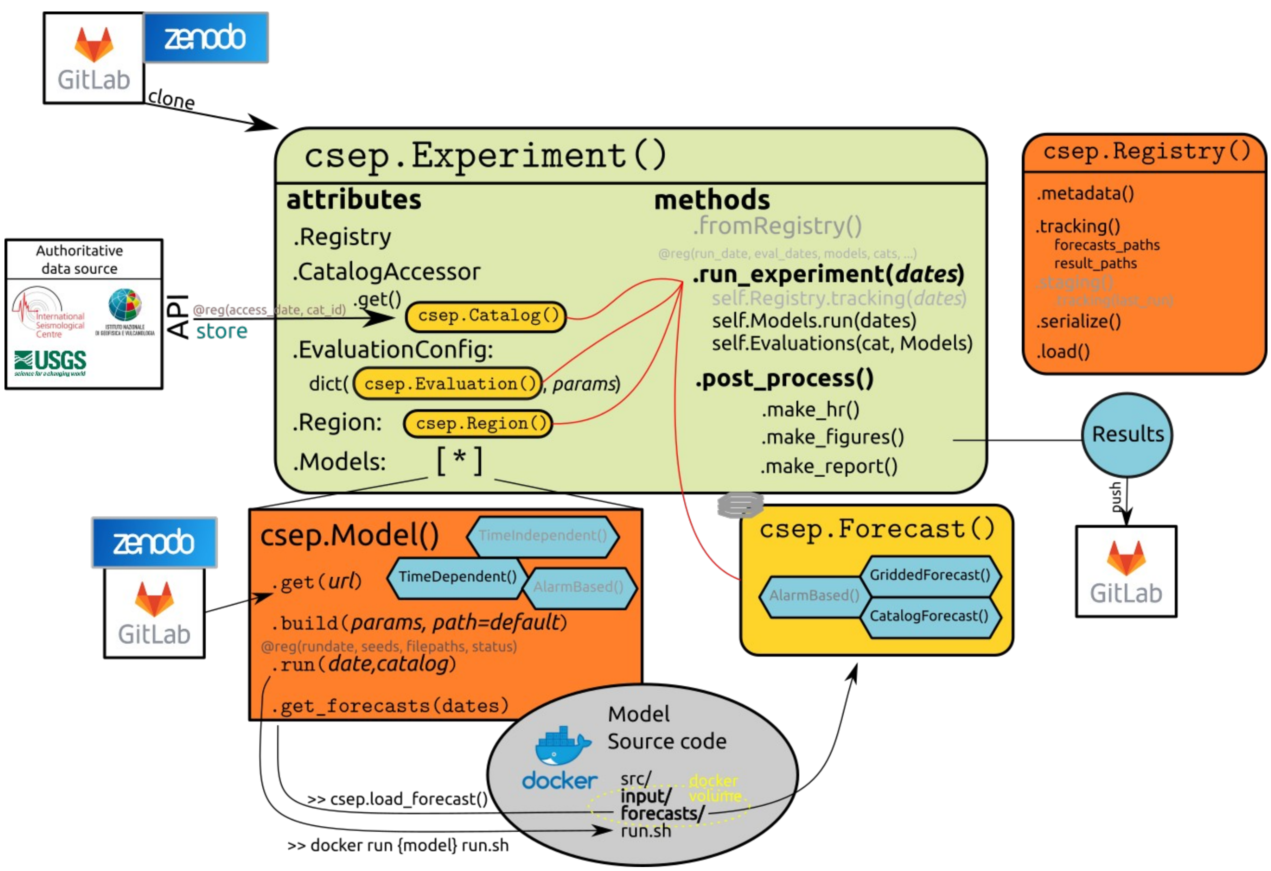

Figure 2. feCSEP code diagram. The Floating Experiment specifications and artifacts can be available in Zenodo/Gitlab, which generates an Experiment object in a local machine that: retrieve a authoritative catalog, build a model, generate a forecast, evaluate it and display the results.

Applications

The European Seismic Hazard Model 2020 (ESHM20)

The long-term forecasting components, i.e. source model logic-tree, of the ESHM20 are being tested pseudo-prospectively, using out-of-sample data from 2014. It evaluates time-independent forecast with M>4.8 observations within a 8-year time window. The authoritative data source is the European-Mediterranean earthquake catalogue (EMEC, Grünthal and Wahlström, 2012; Weatherill et al., in-prep). The experiment is contained in a repository open to the public, which can be easily reproduced in a local machine. It is being updated on-demand, until a continuous data-feed of the EMEC catalog is implemented. The floating experiment is available in:

https://git.gfz-potsdam.de/csep-group/fecsep-efehr20

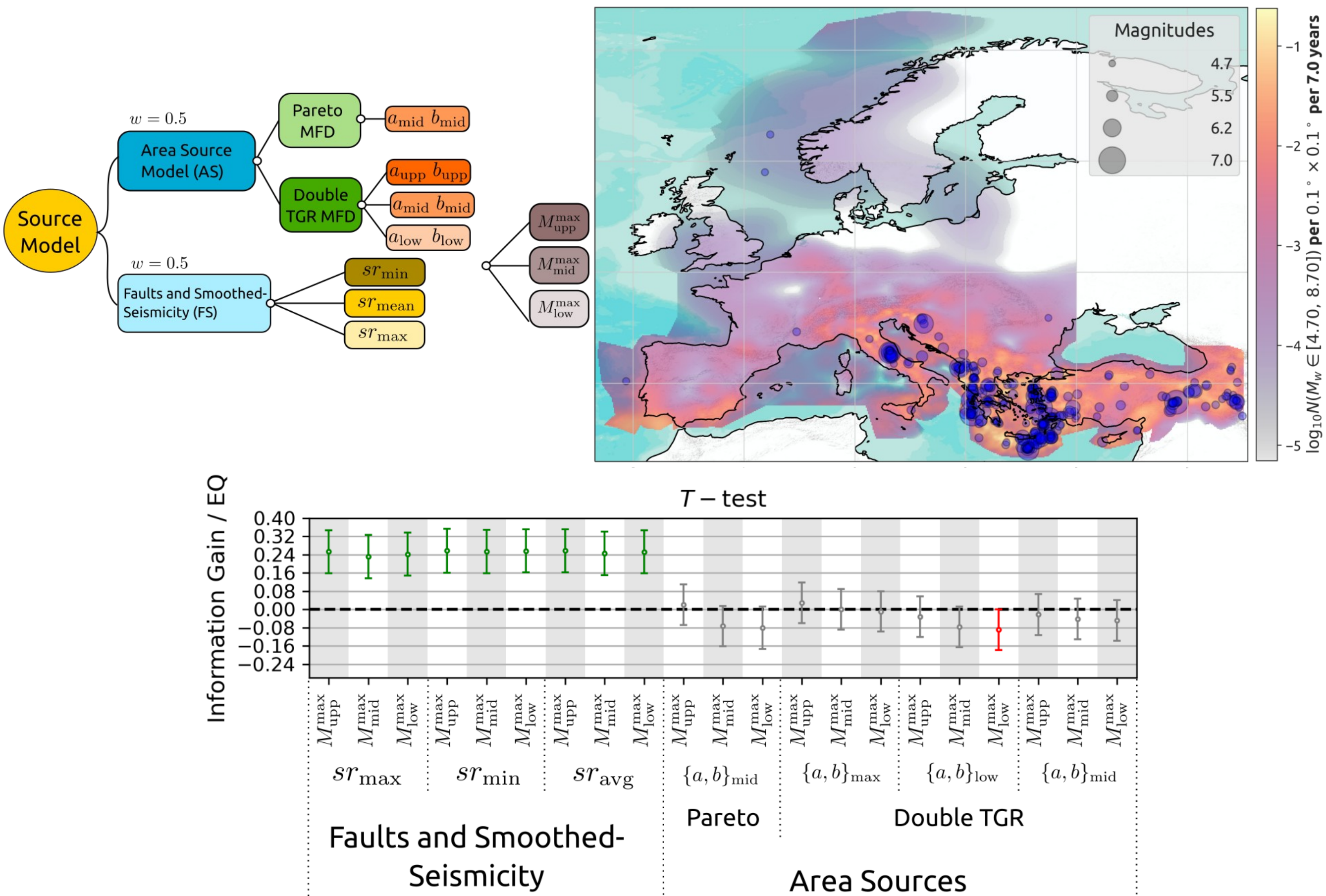

Figure 3. (top-left) The ESHM20 source model logic tree, composed of different model typologies, magnitude-frequency distributions and slip distributions. **(top-right)** The Faults and Smoothed-Seismicity model (FS) and the testing catalog. **(bottom)** The results of the T-test showing the information gain of each branch compared to the previous 2013 ESHM smoothed seismicity model. Only fault models present a significant improvement to the ESHM2013.

Global Earthquake Forecasting Experiment (GEFE) - Quadtree

Quadtrees are hierarchical tree structures to represent a spatial region with variable resolution, in opposition classic gridded spaces, where each node is allowed to have either zero or four child nodes (Khawaja et al., 2023). The main goals of this experiment are to understand the dependency of the testing region on the evaluation results and to understand whether forecasts defined on different spatial grids can be accurately compared against one another. The experiment is designed on a global region, for magnitudes greater than 6.0 and an time-independent window of 9 years, starting from 2014. This type of experiment provides an useful working example, because Quadtree regions drastically improve the computational performance of the evaluations enabling this experiment to be reproduced on any laptop computer. The floating experiment is available in:

https://git.gfz-potsdam.de/csep-group/gefe-quadtree/

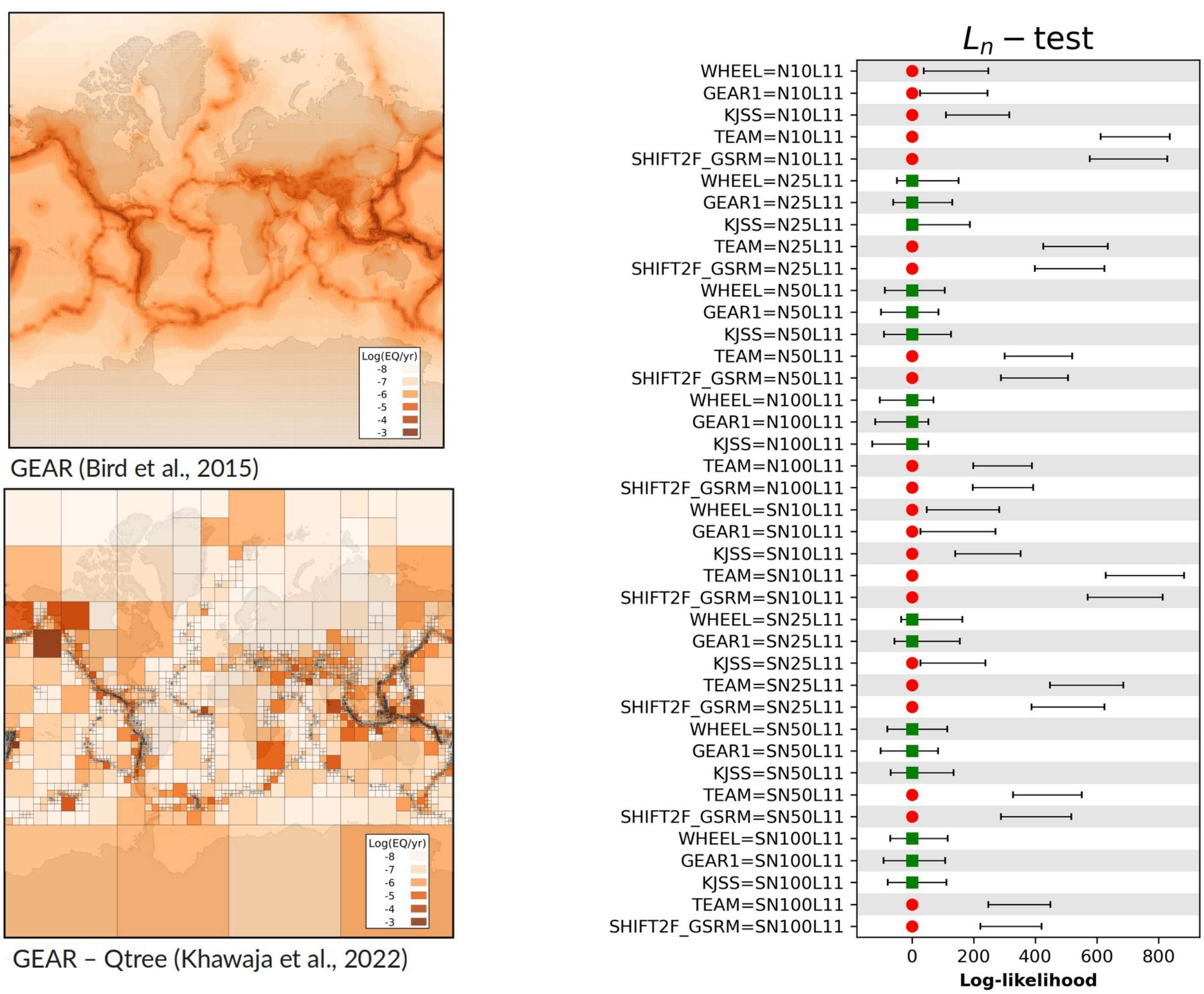

Figure 4. (left) A Quadtree model composed from a parent forecast (GEAR1, Bird et al., 2015) **(left)** Results of a conditional log-likelihood tests. Results show that GEAR1 outperforms other models at multiple resolutions. All the tests and figures are contained in the experiment repository.

The 2010 Italy Forecasting Experiment

The Italy region was submitted to a forecasting experiment in 2010 for a 10-year period of observation (Schorlemmer et al., 2010). However, unlike short-term operational earthquake forecasting, the scarcity of earthquakes targeted by long-term forecasts (e.g. magnitude larger than 5.0) requires an observation frame large enough to empirically validate a forecast. We continue the experiment, exploiting the forecasts' time-independence, to study the regularity of testing results in time. The floating experiment is available in:

https://git.gfz-potsdam.de/csep-group/fe_italy_ti

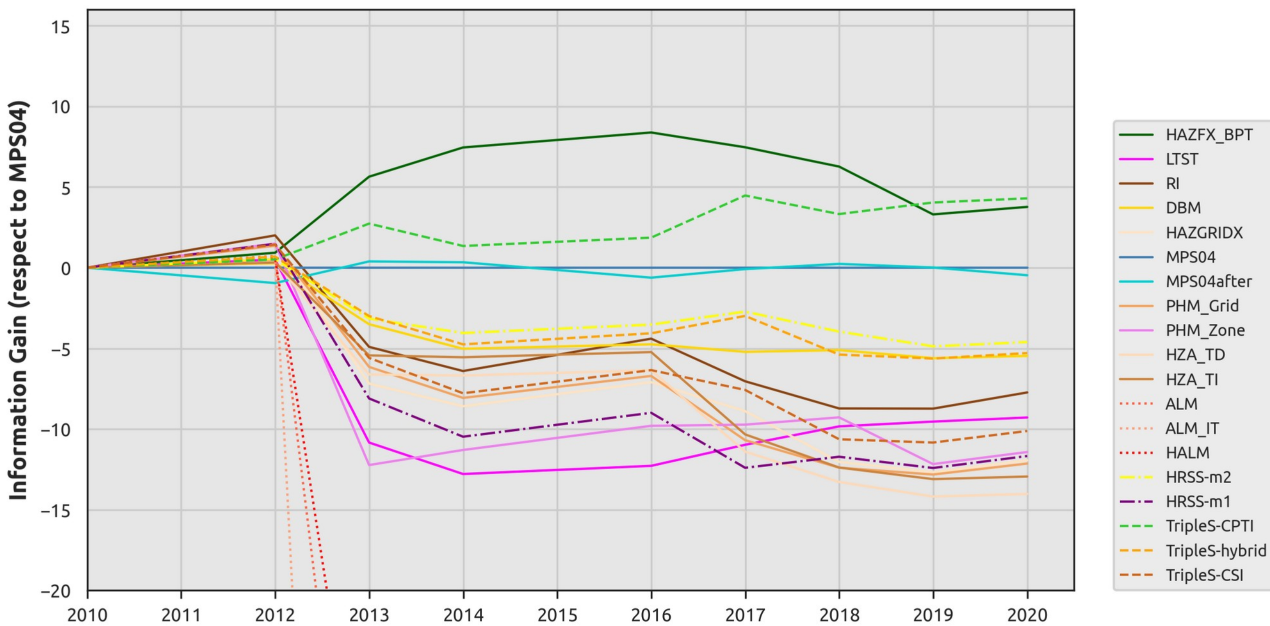

Figure 4. Information Gain (respect to the reference national model) evolution of the competing models during the Experiment. Models are colored by clustering their performance at 2015 (e.g. green are performing well, blue indifferent, red poorly). The groups' performance remain fairly constant (i.e.. the clusters maintain similar ranking) until the end of the experiment, vouching for the stability of the evaluation results.

The (new) Earthquake Forecasting Experiment for Italy

This Experiment has been design to test the performance of novel short-term (time-dependent) forecasting models for 1-day time windows. It expands the classical CSEP horse-race to inter-model and intra-model hypothesis testing. Here, the capacities of feCSEP are being explored due to the complexity arisen from operating and book-keeping time-dependent forecasts. The beta-phase of the experiment, although delayed, will begin end of April with opening the call to modelers external to RISE. The experiment rules, competing models and the beta implementation of this Floating Experiment can be found in:

https://git.gfz-potsdam.de/csep-group/rise_italy_experiment/experiment_setup
https://git.gfz-potsdam.de/csep-group/rise_italy_experiment/models
https://git.gfz-potsdam.de/csep-group/rise_italy_experiment/experiment_system

References

Bird, P., Jackson, D., Kagan, Y. , Kreemer, C., & Stein, R. (2015). GEAR1: A Global Earthquake Activity Rate Model Constructed from Geodetic Strain Rates and Smoothed Seismicity. *Bulletin of the Seismological Society of America*, *105*(5), 2538-2554.

Grünthal, G., & Wahlström, R. (2012). The European-Mediterranean earthquake catalogue (EMEC) for the last millennium. *Journal of seismology*, *16*, 535-570.

Khawaja, A., Schorlemmer, D., Hainzl, S., Iturrieta, P., Savran, W. H., Bayona, J. A., & Werner, M. J. (2023). Multi-resolution grids in earthquake forecasting: The Quadtree approach. *Bulletin of the Seismological Society of America*, *113*(1), 333-347.

Krafczyk, M. S., Shi, A., Bhaskar, A., Marinov, D., & Stodden, V. (2021). Learning from reproducing computational results: introducing three principles and the Reproduction Package. *Philosophical Transactions of the Royal Society A*, *379*(2197), 20200069.

Pagani, M., Monelli, D., Weatherill, G., Danciu, L., Crowley, H., Silva, V., ... & Vigano, D. (2014). OpenQuake engine: An open hazard (and risk) software for the global earthquake model. *Seismological Research Letters*, *85*(3), 692-702.

Savran, W. H., Bayona, J. A., Iturrieta, P., Asim, K. M., Bao, H., Bayliss, K., ... & Werner, M. J. (2022). pycsep: A python toolkit for earthquake forecast developers. *Seismological Society of America*, *93*(5), 2858-2870.

Schorlemmer, D., Christophersen, A., Rovida, A., Mele, F., Stucchi, M., & Marzocchi, W. (2010). Setting up an earthquake forecast experiment in Italy. *Annals of Geophysics*.

Schorlemmer, D., Werner, M. J., Marzocchi, W., Jordan, T. H., Ogata, Y., Jackson, D. D., ... & Zhuang, J. (2018). The collaboratory for the study of earthquake predictability: Achievements and priorities. *Seismological Research Letters*, *89*(4), 1305-1313.