Advancements in pyCSEP: Enhancing Earthquake Forecast Evaluation with New Features and Regional Applications

Kenny Graham, José A. Bayona, Pablo Iturrieta, Francesco Serafini, Emanuele Biondini, William H. Savran, Maximilian J. Werner and pyCSEP development team

Take home messages

- pyCSEP is an open-source toolkit for evaluating earthquake forecasts.
- It includes modules for accessing catalogs, visualizing models, conducting statistical evaluations and other utilities.
- Recent updates include:
  - Updates to New Regions and Official Catalogs
  - Projections of global GEAR1 forecasts unto custom regions
  - Multi-Resolution Grids in Earthquake Forecasting: The Quadtree Approach
  - Evaluation of Alarm-based forecasting models
  - Non-Poissonian tests and new ranking scores
- Researchers are encouraged to contribute ideas and methods to enhance the pyCSEP toolkit.

Outline

- Introduction to CSEP and pyCSEP
- Summary of Tests implemented in pyCSEP
- Update to the pyCSEP toolkit
- Future Developments (The floating experiment)
- Conclusion

The Collaboratory for the Study of Earthquake Predictability (CSEP) represents a worldwide, open community with a core mission of advancing earthquake predictability research.

This mission is achieved through the rigorous evaluation of probabilistic earthquake forecast models and prediction algorithms.

CSEP has operated international testing centers for more than 16 years, facilitating prospective experiments guided by the following foundational principles.

Guiding Principles

Transparency: Testing centers are committed to cataloging all data products provided by modelers, whether generated by the modelers themselves or by the testing center.

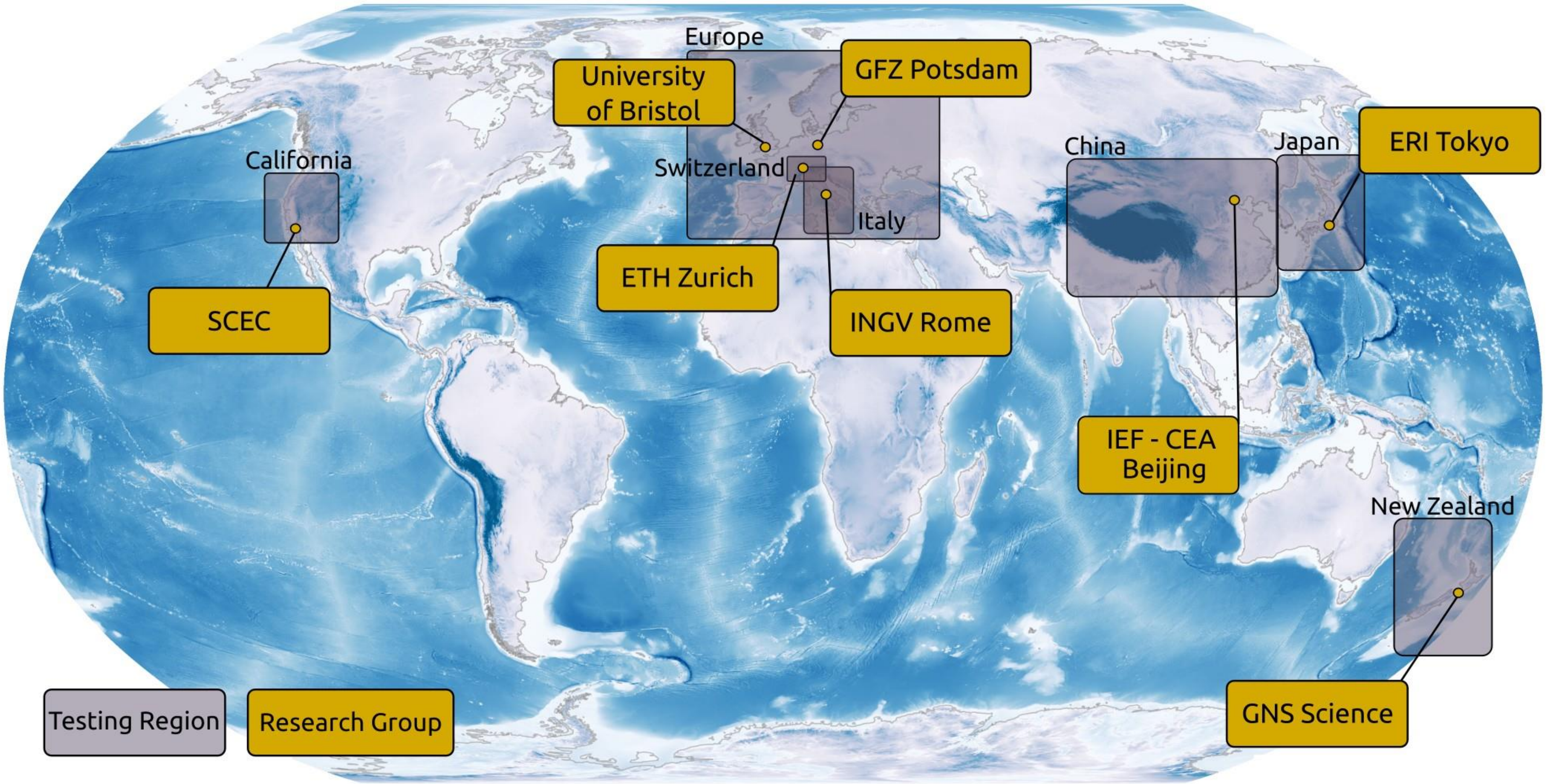
Controlled Environment: Testing centers are meticulously designed to eliminate bias. Once models are installed, modelers are strictly prohibited from accessing them to ensure a fair and controlled evaluation environment.

Comparability: Forecasts are systematically organized by testing class. Within each testing class, forecasts are thoroughly compared against one another and against the available data.

Reproducibility: Testing centers are equipped with the capability to re-run tests later, ensuring the reliability and reproducibility of the evaluation process.

These guiding principles underscore the rigorous and objective nature of the testing procedures employed by CSEP to evaluate earthquake forecast models and prediction algorithms.

CSEP Represents an International Effort



CSEP spearheads a global initiative to meticulously assess earthquake forecasting models through autonomous experiments on CSEP-managed servers, known as **testing centers.** Source: https://cseptesting.org/

CSEP2 Development (towards pyCSEP)

To address the limitations of CSEP1, CSEP2 development was initiated in 2018 with the goal of achieving both scientific and technical advancements..

- **Technically**, the CSEP1 software was difficult to extend, and the testing computer infrastructure was time consuming and expensive to operate. Providing new decentralized approach for conducting CSEP-coordinated experiments was needed.
- **Scientifically**, along with grid-based forecast (as tested in CSEP1) new types of earthquake forecasts (stochastic/catalog-based) needed to be added.

CSEP2 was designed to support new earthquake forecasts, simplify software, lower testing centre costs, and provide a flexible framework for new features, all while maintaining CSEP1's rigorous evaluation principles.

Overview of pyCSEP

- Modernize CSEP experiments using community software development and an open-experiment format
- Open experiment format ensures all data are available through FAIR practices and results can be easily reproduced on different computers
- pyCSEP provides a flexible framework for incorporating new features seamlessly while upholding the guiding principles from CSEP1 for rigorously evaluating
- More information about CSEP and pyCSEP are available at: https://cseptesting.org

- pyCSEP software available on at: https://github.com/SCECcode/pycsep

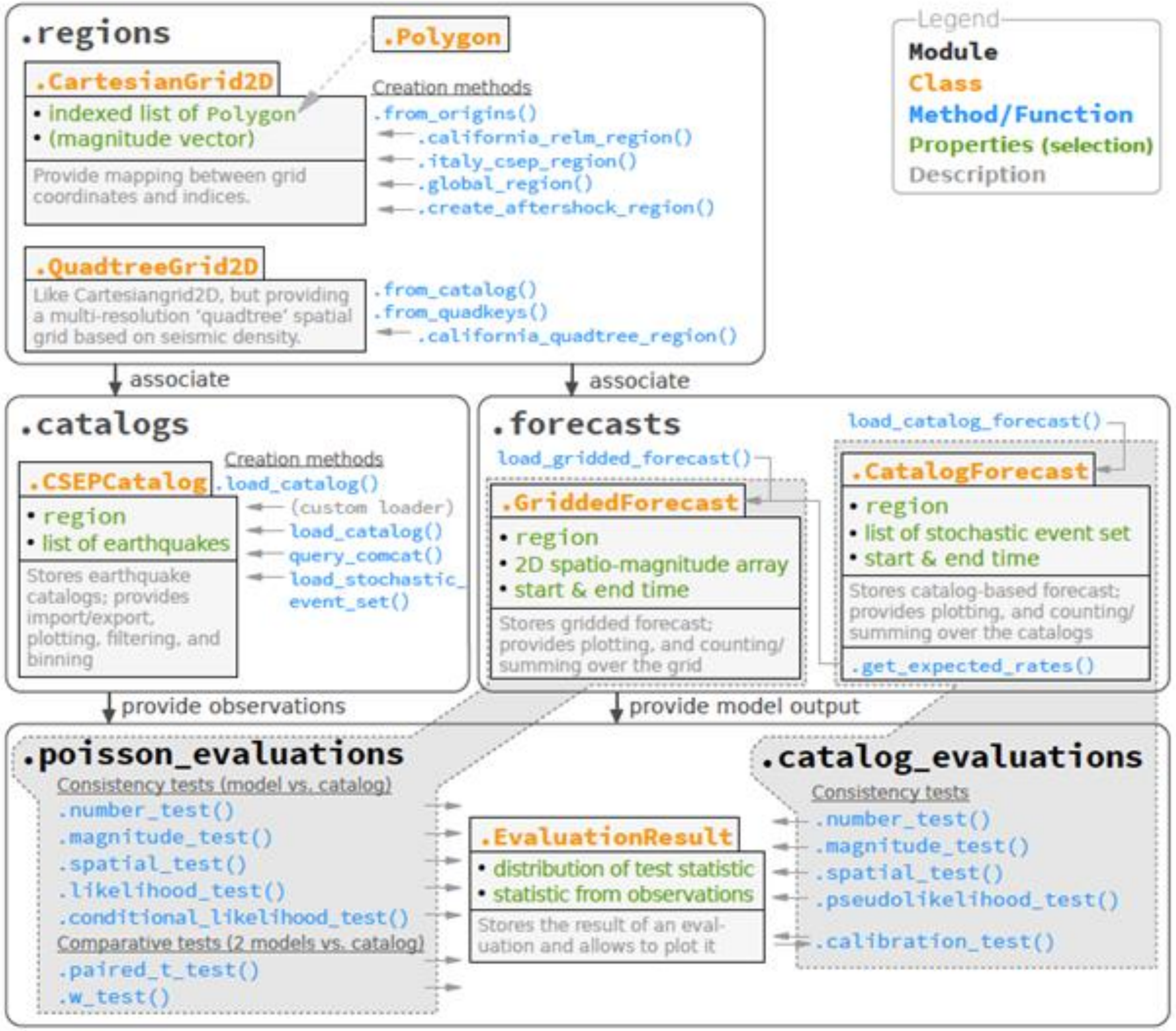


Figure from **Savran et al., 2022**

GETTING STARTED with pyCSEP

Installing pyCSEP is as simple as:

conda install --channel conda-forge pycsep

In-depth instructions can be found at:

https://cseptesting.org/getting_started/installing.html

- ❑Documentation: https://cseptesting.org
- ❑Examples: https://cseptesting.org/tutorials/
- ❑Report issues: https://github.com/SCECcode/pycsep/issues
- ❑To Contribute: https://github.com/SCECcode/pycsep

pyCSEP, Example of accessing catalog

```
import csep
from csep.core import regions
from csep.utils import time_utils

# Set start and end time
start_time = time_utils.strptime_to_utc_datetime('2008-01-01 00:00:00.0')
end_time = time_utils.strptime_to_utc_datetime('2017-01-31 23:59:59.0')

# Magnitude bins properties
min_mw = 4.95
max_mw = 8.95
dmw = 0.1

# Create space and magnitude regions.
magnitudes = regions.magnitude_bins(min_mw, max_mw, dmw)
region_nz = regions.nz_csep_region()

# Read New Zealand catalog
nz_catalog = csep.query_gns(start_time, end_time,max_longitude=179.8)
nz_catalog = nz_catalog.filter('magnitude >= 4.95')
nz_catalog = nz_catalog.filter_spatial(region_nz)
print(f'\n\n this is the catalog *** {nz_catalog}')
```

Summary of Tests implemented in pyCSEP

Grid-based Forecast:

- ❑ Consistency tests:
  - ➢ N(umber) Test
  - ➢ Negative binomial Number Test
  - ➢ M(agnitude) Test
  - ➢ S(patial) Test
  - ➢ Binary Spatial Test
  - ➢ L(ikelihood) / C(onditional) L(ikelihood) Test
  - ➢ Binary C(onditional) L(ikelihood) Test
  - ➢ Brier Score
- ❑ **Comparative tests:**
  - ➢ T-Test
  - ➢ W(ilcoxon Signed-Rank) Test
  - ➢ Kagan Information Gain (Kagan I1 score)
- ❑ **Diagnostic tests:**
  - ❖ K-Ripley functions' Residual (in development)

Catalog-based Forecast:

- Consistency tests:
  - Number Test
  - Magnitude Test
  - Spatial Test
  - Calibration Test
  - (Pseudo)-Likelihood Test
  - Spatial Event Occurrence Probability Test
  - Total Earthquake Rate Distribution
- Comparative tests:
  - None
- Diagnostic tests:
  - None

Alarm-based Forecast:

- Evaluating a model's success rate versus its false alarm rate
  - Molchan Diagram
  - Receiver Operating Characteristic (ROC)
  - Concentration ROC curves
- Receiver Operating Characteristic (ROC) for Gridded forecasts (in development)

New Features in the pyCSEP Toolkit for Earthquake Forecast Development and Evaluation

Kenny M. Graham*[1], José A. Bayona[2], Asim M. Khawaja[3], Pablo Iturrieta[3], Francesco Serafini[2], Emanuele Biondini[4], David A. Rhoades[1], William H. Savran[5], Philip J. Maechling[6], Matthew C. Gerstenberger[1], Fabio Silva[6], and Maximilian J. Werner[2]

Abstract

The Collaboratory for the Study of Earthquake Predictability (CSEP) is a global community dedicated to advancing earthquake predictability research by rigorously testing probabilistic earthquake forecast models and prediction algorithms. At the heart of this mission is the recent introduction of pyCSEP, an open-source software tool designed to evaluate earthquake forecasts. pyCSEP integrates modules to access earthquake catalogs, visualize forecast models, and perform statistical tests. Contributions from the CSEP community have reinforced the role of pyCSEP in offering a comprehensive suite of tools to test earthquake forecast models. This paper builds on Savran. et al. (2022), where pyCSEP was originally introduced, by describing new tests and recent updates that have significantly enhanced the functionality and user experience of pyCSEP. It showcases the integration of new features, including access to authoritative earthquake catalogs from Italy (Bolletino Seismico Italiano), New Zealand (GeoNet), and the world (Global Centroid Moment Tensor), the creation of multi-resolution spatial forecast grids, the adoption of non-Poissonian testing methods, applying a global seismicity model to specific regions for benchmarking regional models and evaluating alarm-based models. We highlight the application of these recent advances in regional studies, specifically through the New Zealand case study, which showcases the ability of pyCSEP to evaluate detailed, region-specific seismic forecasts using statistical functions. The enhancements in pyCSEP also facilitate the standardization of how the CSEP forecast experiments are conducted, improving the reliability and comparability of the earthquake forecasting models. As such, pyCSEP exemplifies collaborative research and innovation in earthquake predictability, supporting transparent scientific practices and community-driven development approaches.

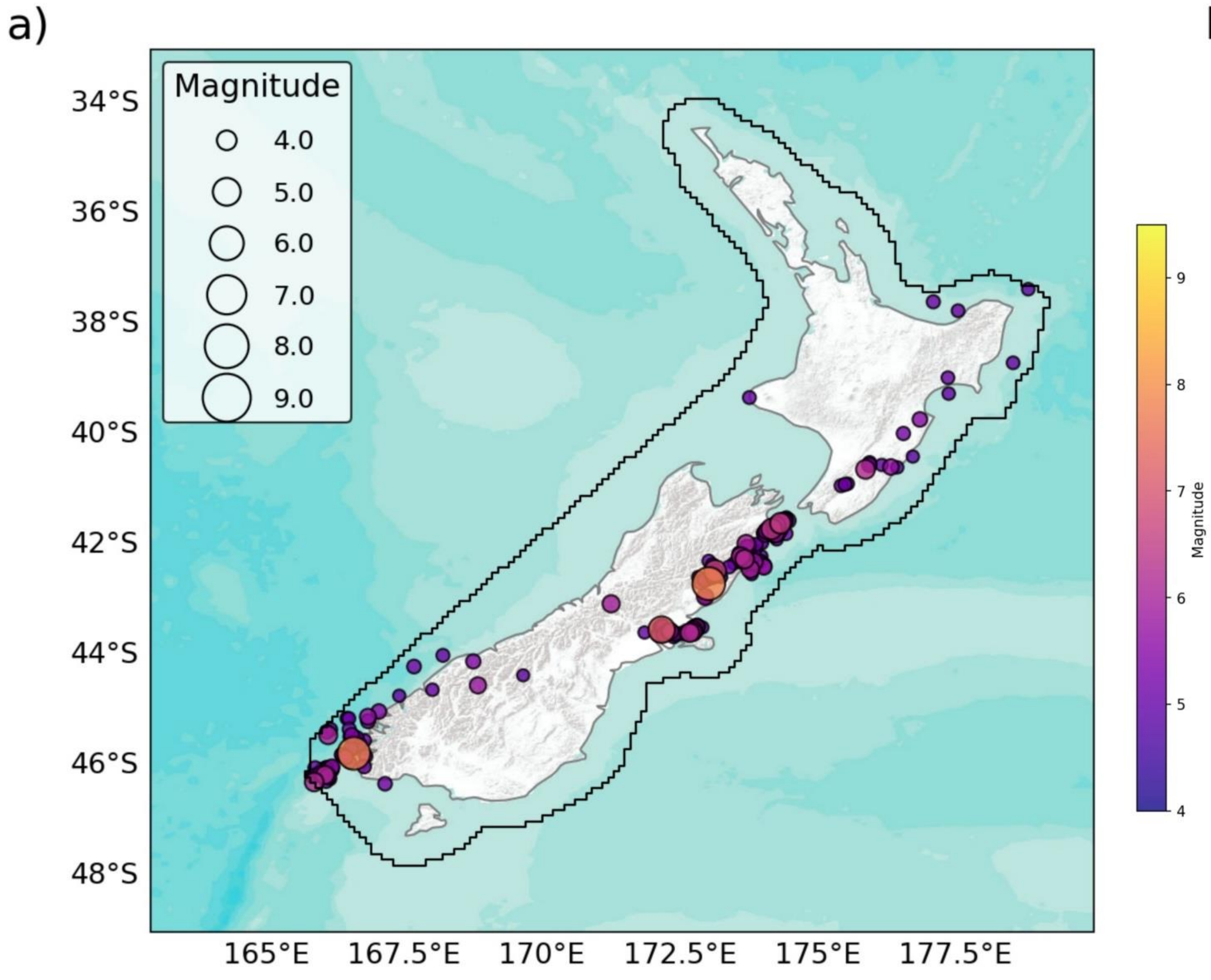
Cite this article as Graham. K.M., et al., (2022). New Features in the pyCSEP Toolkit for Earthquake Forecast Development and Evaluation, *Seismol. Res. Lett.* **XX**, 1–28, doi: 00.0000/000000000.

Supplemental Material

Accepted for publication

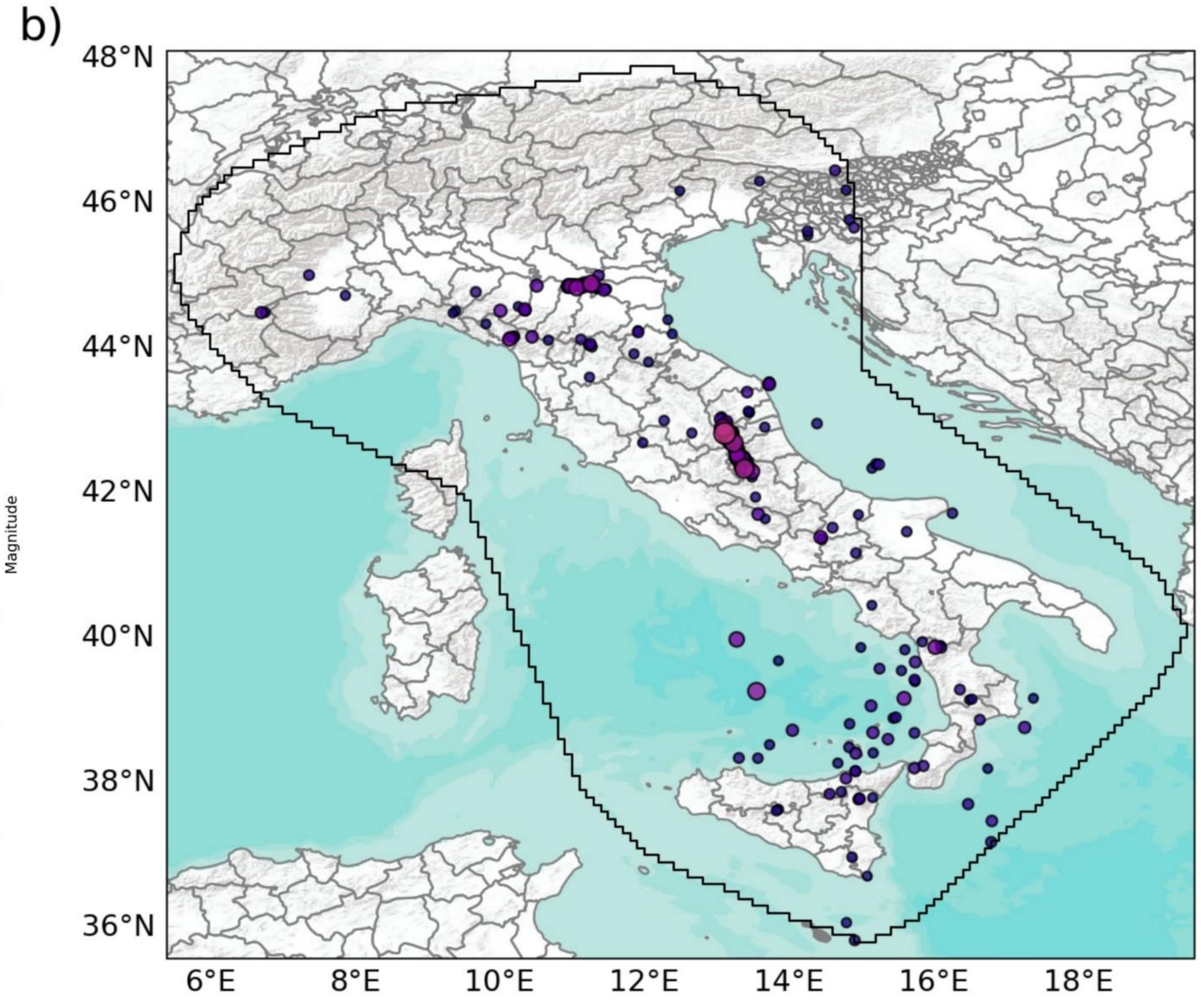
Updates to New Regions and Official Catalogs

New Zealand (GeoNet)



`csep.regions.nz_csep_region()`: NZ region
`csep.query_gns()` downloads NZ catalog

Italy (Bollettino Seismico Italiano)

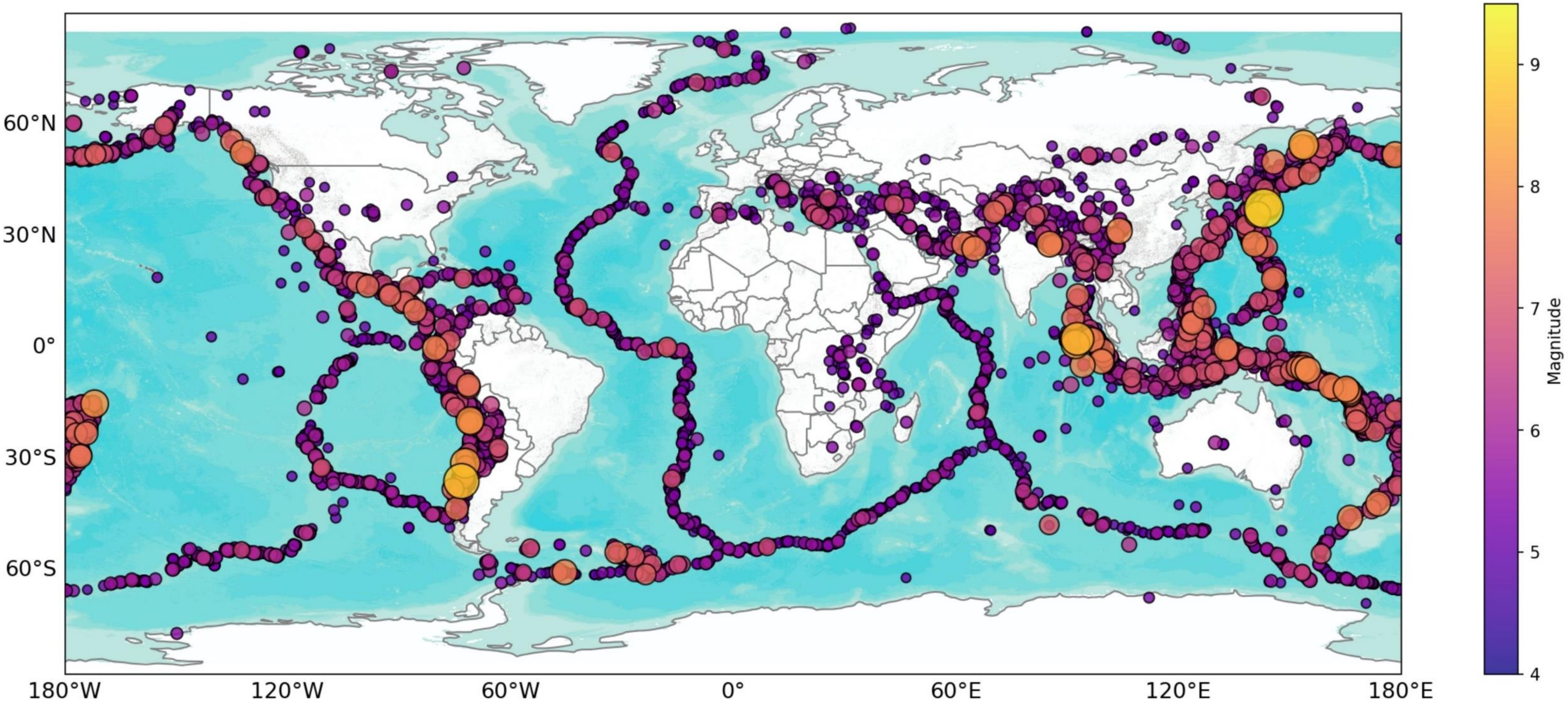


`csep.regions.italy_csep_region()`: Italy region
`csep.query_bsi()`: Download Iataly catalog

pyCSEP, Example of accessing catalog

```
import csep
from csep.core import regions
from csep.utils import time_utils

# Set start and end time
start_time = time_utils.strptime_to_utc_datetime('2008-01-01 00:00:00.0')
end_time = time_utils.strptime_to_utc_datetime('2017-01-31 23:59:59.0')

# Magnitude bins properties
min_mw = 4.95
max_mw = 8.95
dmw = 0.1

# Create space and magnitude regions.
magnitudes = regions.magnitude_bins(min_mw, max_mw, dmw)
region_nz = regions.nz_csep_region()

# Read New Zealand catalog
nz_catalog = csep.query_gns(start_time, end_time,max_longitude=179.8)
nz_catalog = nz_catalog.filter('magnitude >= 4.95')
nz_catalog = nz_catalog.filter_spatial(region_nz)
print(f'\n\n this is the catalog *** {nz_catalog}')
```

World (Global Centroid Moment Tensor [GCMT] catalog)



access the GCMT catalog through the csep.query_gcmt() function

Integration of multi-resolution; QuadTree spatial grids

The Quadtree Approach

- Typically, a single-resolution Cartesian grid with 0.1° × 0.1° cells results in 6.48 million global cells.
- While single-resolution grids are easy to store and reuse due to their regularity, they can be resource-intensive and may have reduced test power (Khawaja et al., 2022).
- The actual earthquake data used for forecasting is much less, creating a significant disparity between the number of earthquakes and grid cells (Khawaja et al., 2022).
- Quadtree multi-resolution grids provide optimal data-driven resolution for forecast generation and testing.
- Quadtree multi-resolution grids, though requiring more initial programming, use fewer computational resources and enhance test power. It provides efficient and data-driven resolution.

The Quadtree Approach

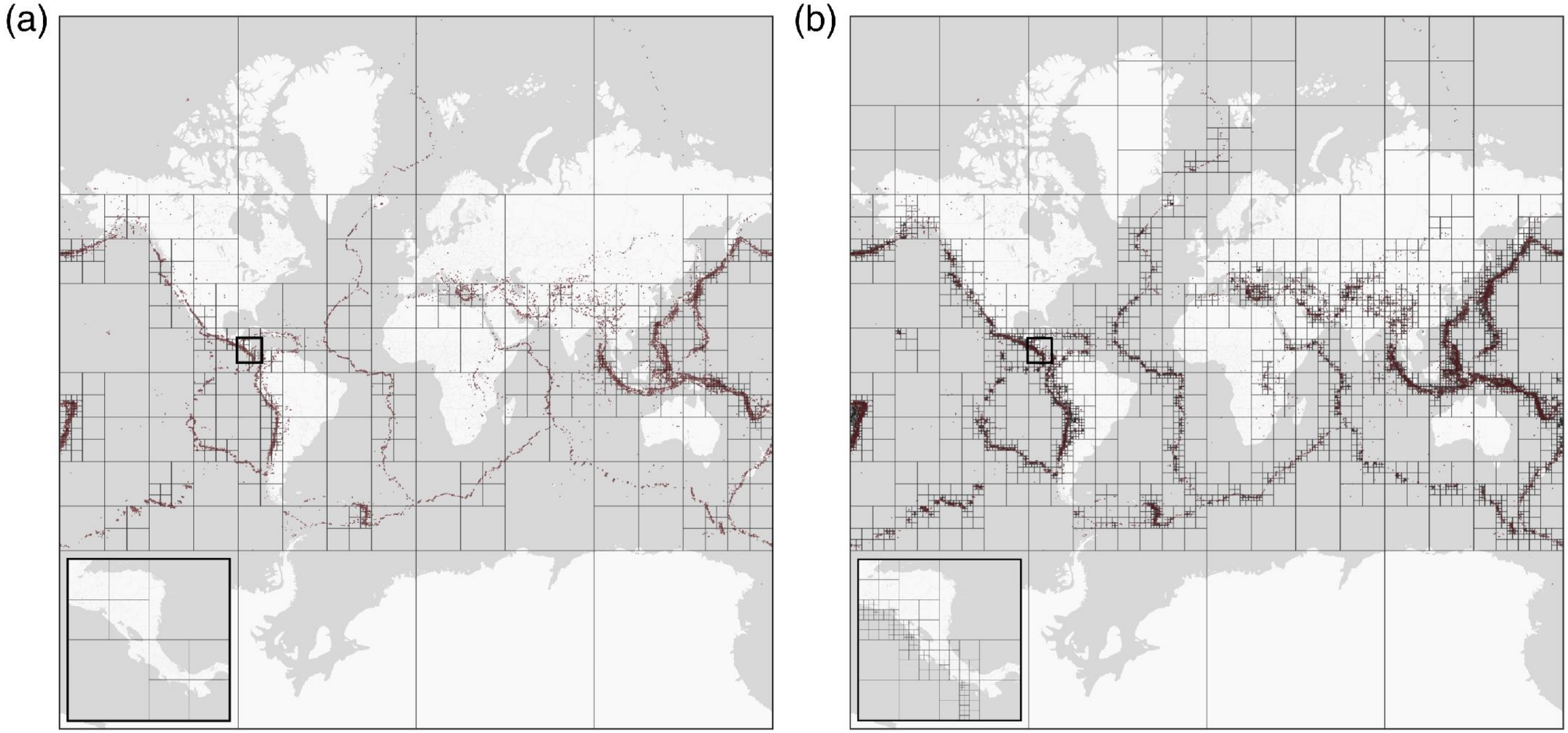
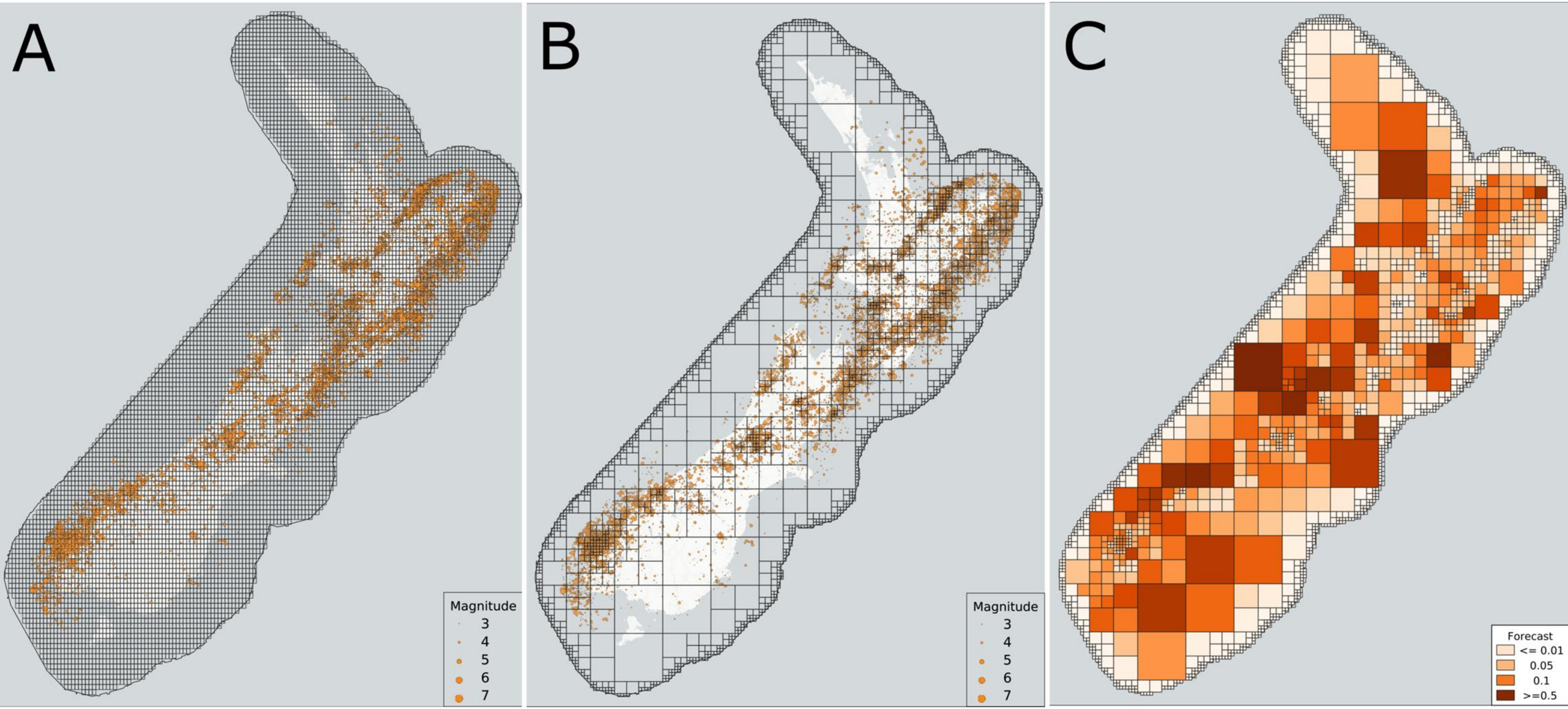
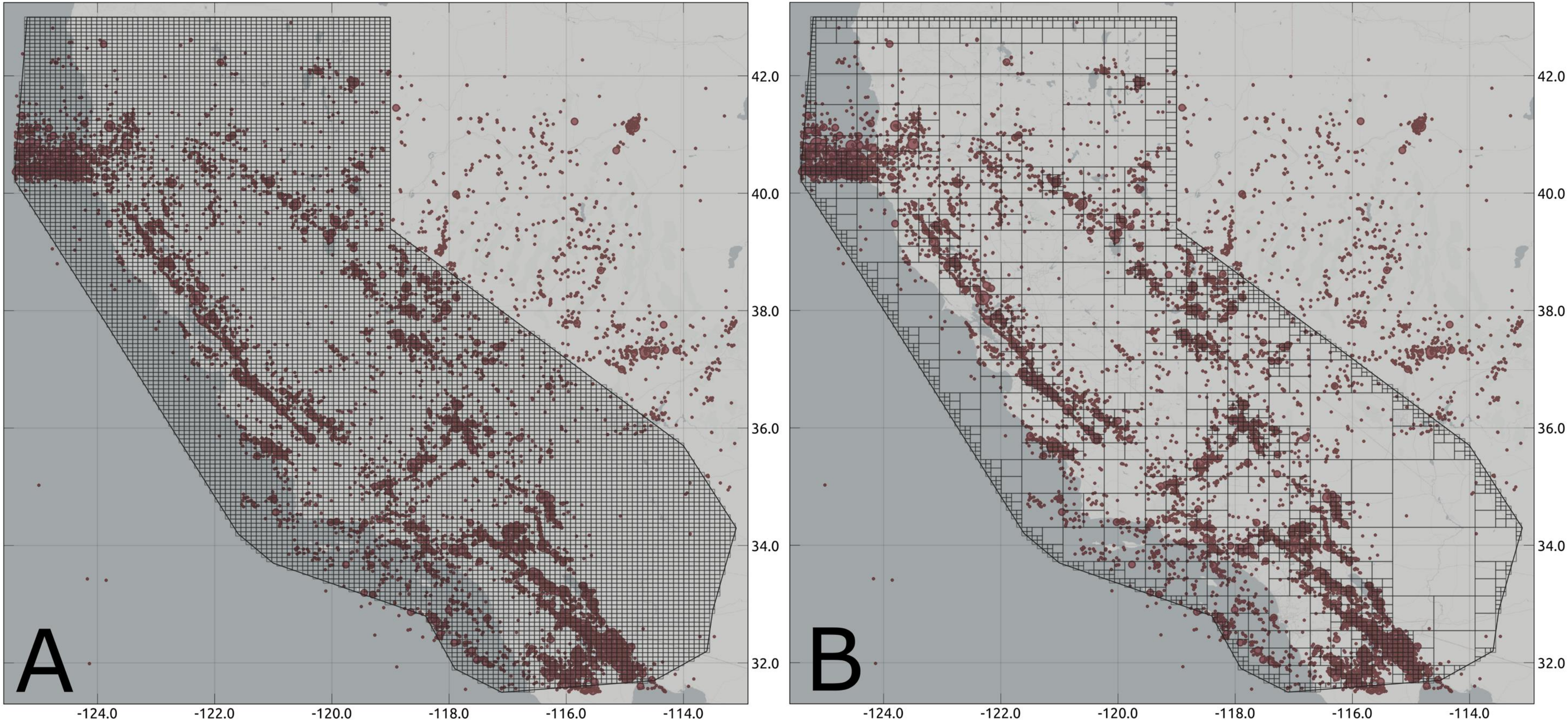


Figure from Khawaja et al., 2022

Data-driven generation of multi-resolution Quadtree grids based on earthquake catalogs. The grid resolution is determined by two conditions: the maximum number of earthquakes allowed per grid cell Nmax and the maximum zoom level Lmax allowed for every cell. Grids of the global testing region: (a) Nmax=100 and Lmax=11 (N100L11), (b) Nmax=10 and Lmax=11 (N10L11).

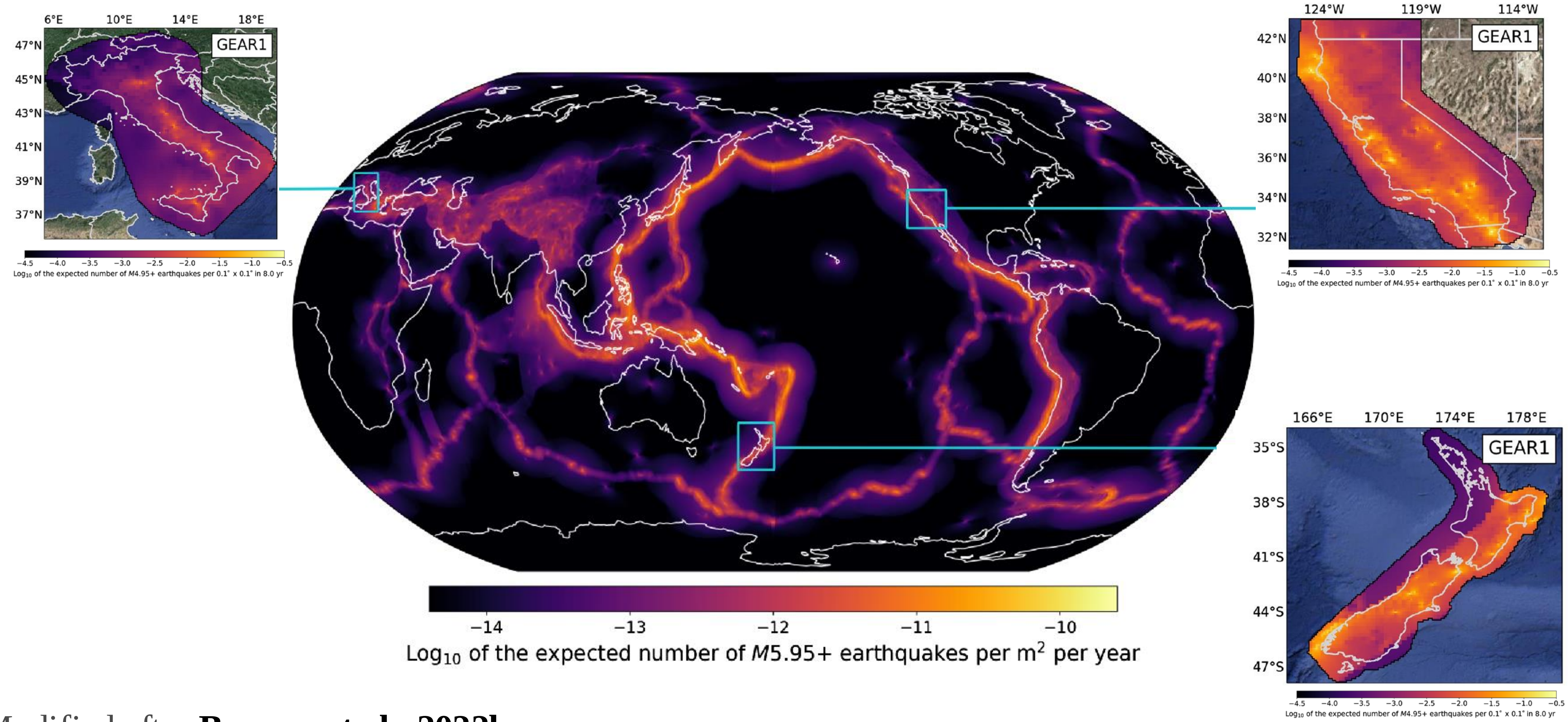


A) A 0.1° × 0.1° grid within the CSEP New Zealand region, showing M ≥ 3.95 earthquakes from 1985 to 2006. B) A Quadtree grid defined by N_max = 100 earthquakes per cell, L_max = 12 zoom levels, and the NZ region boundary. B) A one-year M ≥ 4.95 earthquake forecast for New Zealand using the multi-resolution Quadtree grid, with expected earthquakes per cell.


A
B
42.0
40.0
38.0
36.0
34.0
32.0
-124.0
-122.0
-120.0
-118.0
-116.0
-114.0

Projecting Global GEAR1 model to custom regions

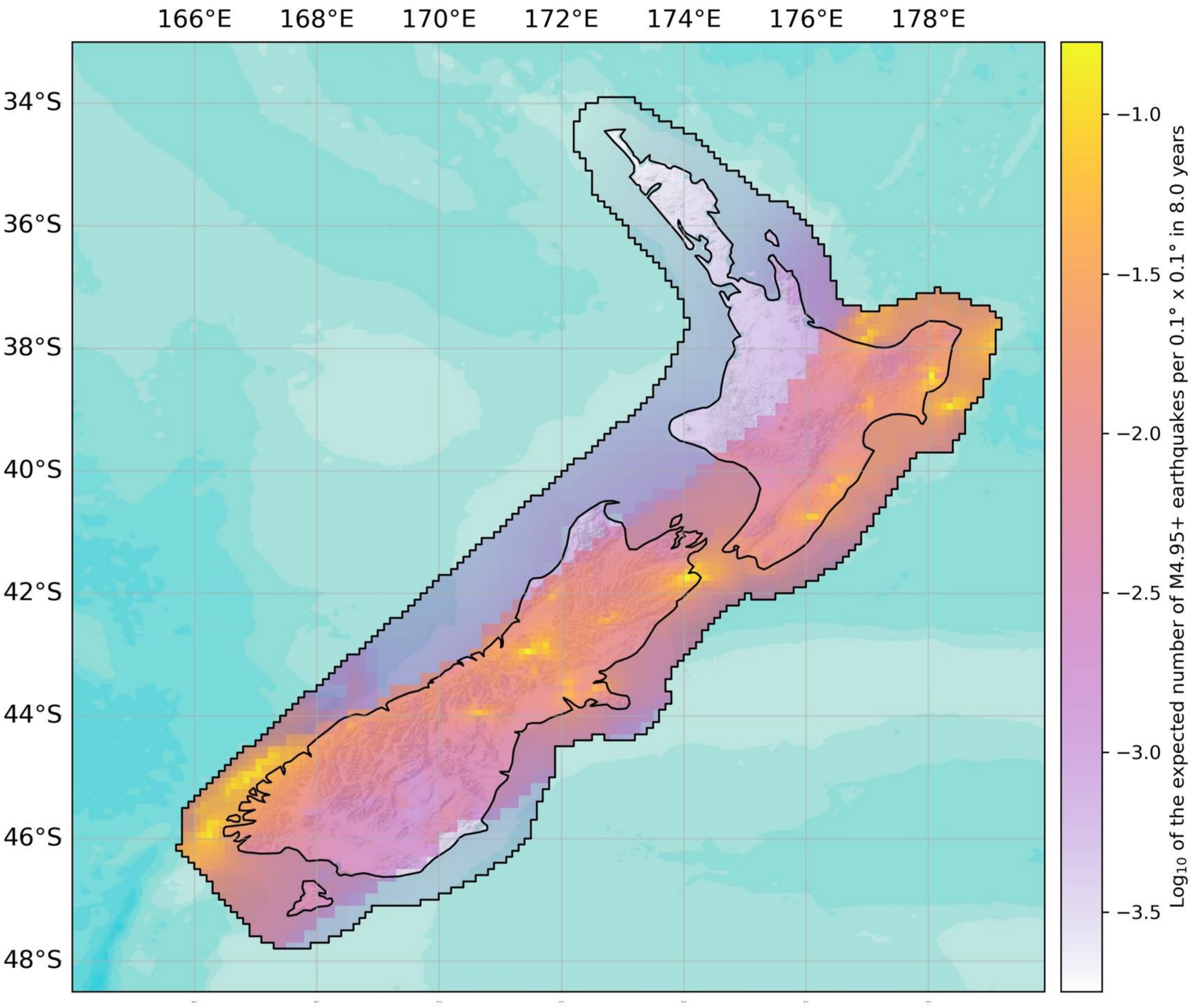
The Global Earthquake Activity Rate (GEAR1; Bird et al., 2015) is a time-invariant seismicity model combining smoothed seismicity and inter-seismic strain data through empirical calibration.



Modified after **Bayona et al., 2022b**

- pyCSEP's new function, mapping_GEAR1(), enables users to project and extrapolate annual global rates to any specified geographic region Graham et al., 2024.

- Projection of the GEAR1 model onto the CSEP New Zealand testing region (black polygon), showing the expected number of M 4.95+, depth ≤ 40 km earthquakes per spatial cell over eight years.



Evaluating Alarm-Based Models

Evaluating Alarm-Based Models

pyCSEP provides tools for evaluating alarm-based earthquake forecasting models. The current functions evaluates a model's success rate versus its false alarm rate.

The functions includes:

- ❖ generating ROC plots: `csep.utils.plots.plot_ROC_diagram()`
- ❖ concentrated ROC plots: `csep.utils.plots.plot_concentration_ROC()`
- ❖ Molchan diagrams using: `csep.utils.plots.plot_Molchan_diagram()`

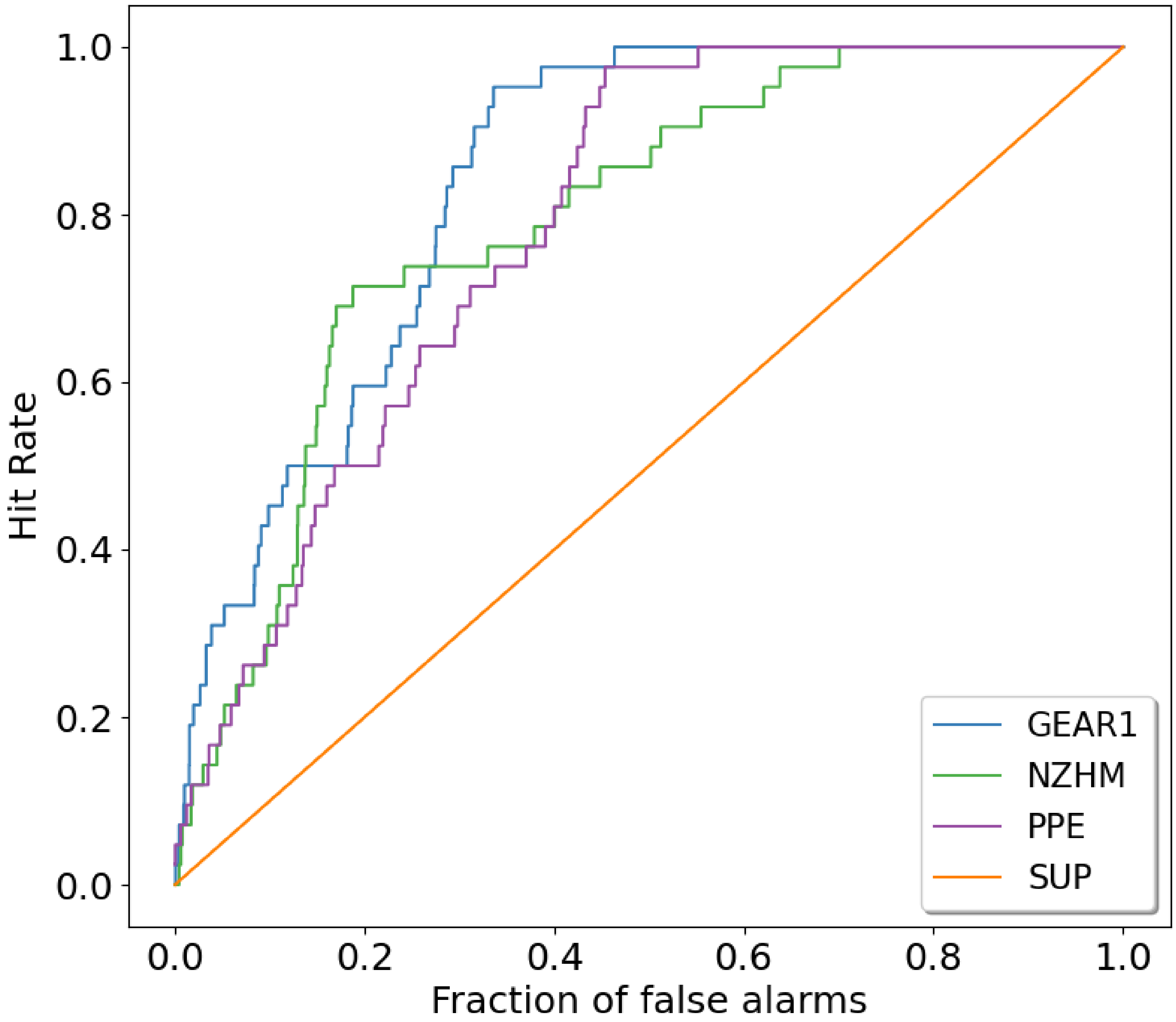
Receiver Operating Characteristic (ROC) for Gridded forecasts (in development)

Receiver Operating Characteristic (ROC)

Receiver Operating Characteristic (ROC) curves, obtained using the alarm-based approach, comparing a Poisson uniform (SUP) seismicity model (orange line) and different spatially specific time-invariant earthquake forecasting models for New Zealand.

ROC plots:

`csep.utils.plots.plot_ROC_diagram()`

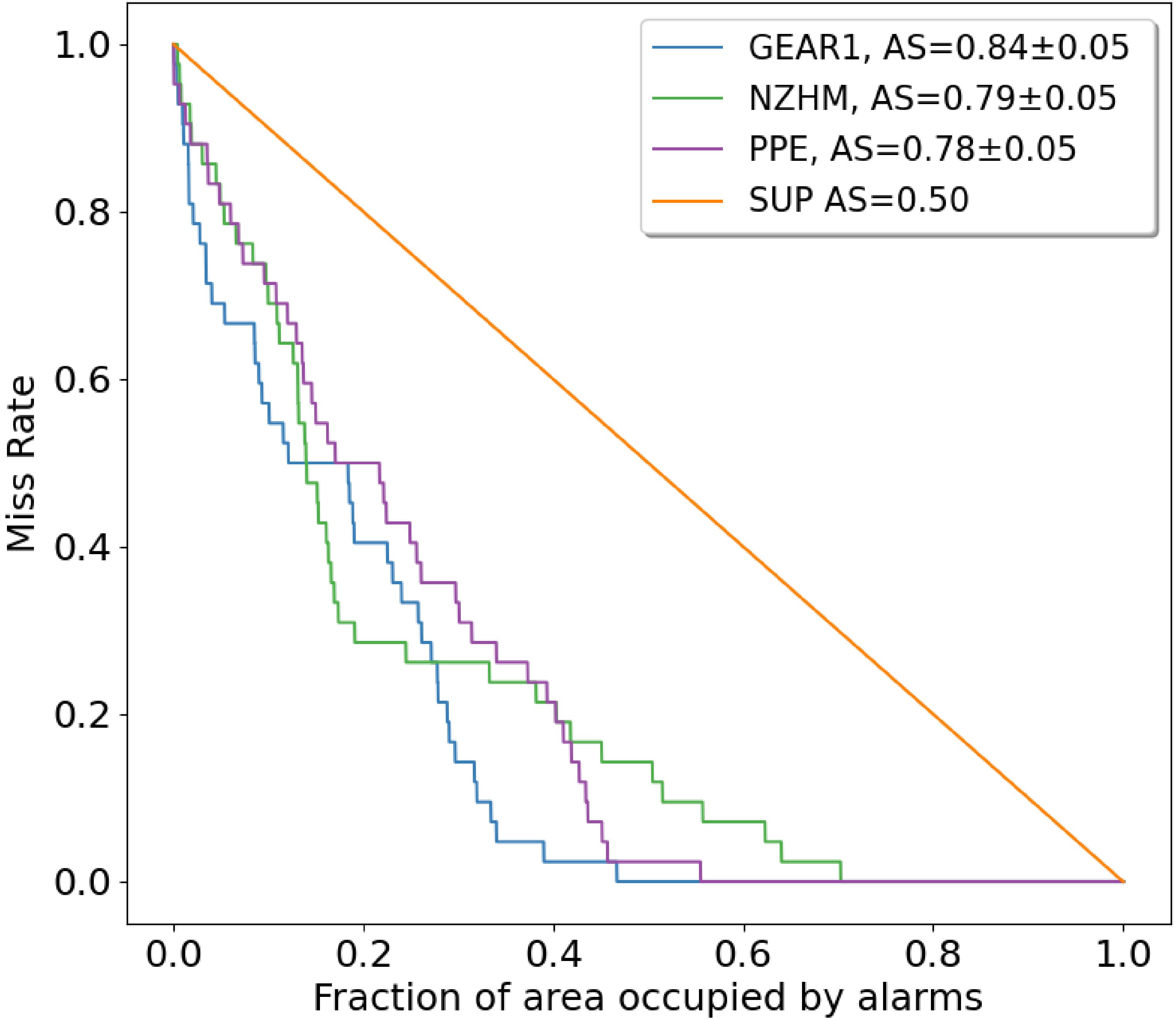


Molchan diagrams

Comparison of Molchan diagrams depicting the predictive performance of models for New Zealand, including comparisons among themselves and with the random reference model (SUP, orange curve). The legend includes the corresponding Area Skill (AS) score values for each model. According to this statistic, the higher the Area Skill (AS) score of a model, the better its predictive ability.

Molchan diagrams:

csep.utils.plots.plot_Molchan_diagram()



New metrics and tests

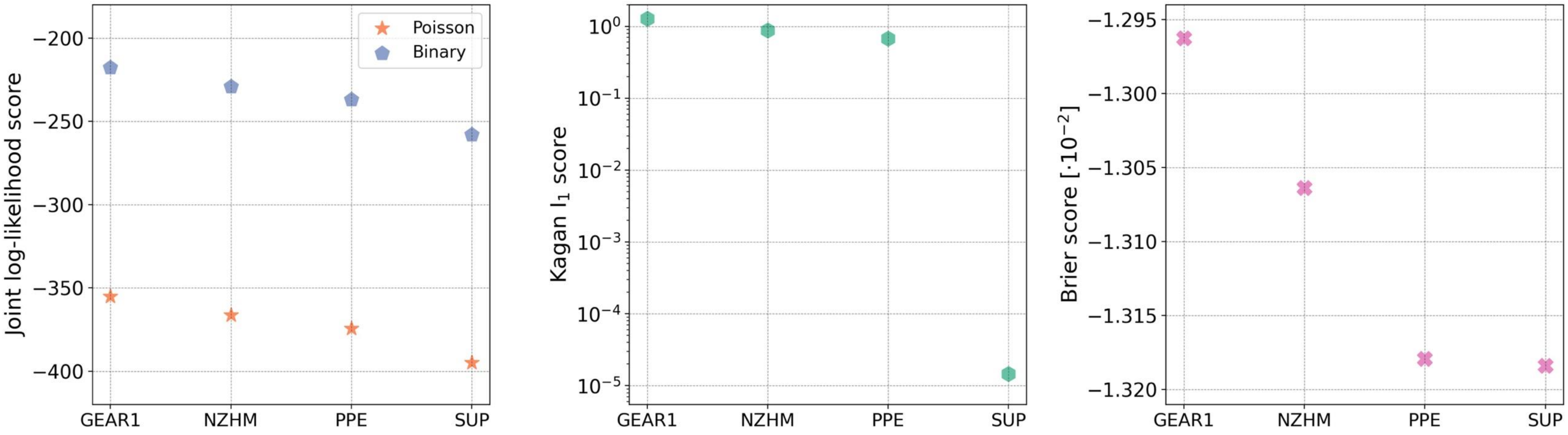
New metrics and tests

We've introduced few new tests, including the binary spatial (S) and conditional likelihood (CL) tests.

We've also added new scores: the Kagan information score [5] and the Brier score [6,7].

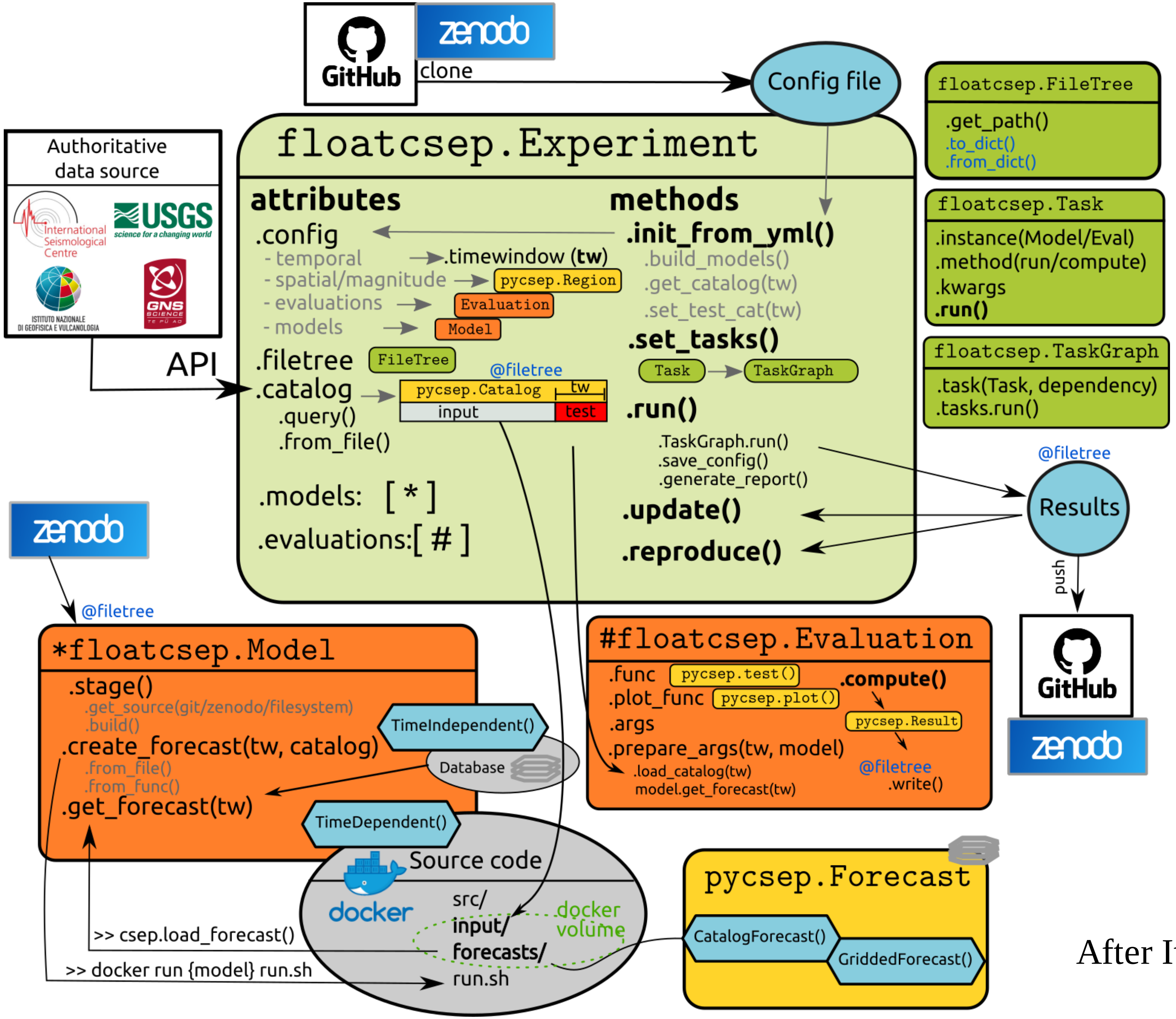
Brier score ranks probabilistic forecasts,

Kagan information score evaluates a model's performance against a Poisson-based reference model [4].



Quantitative comparisons of earthquake forecasting models for New Zealand based on a) Poisson and binary joint log-likelihood scores, b) Kagan I1 information scores and c) Brier scores. We use M≥ 4.95 earthquakes observed between 2014 and 2022 in New Zealand.

The Floating Experiment



After Iturrieta et al., in prep

Conclusion

- The Collaboratory for the Study of Earthquake Predictability (CSEP) is focusing on advancing the evaluation of earthquake forecasts.
- Recent updates include support for new catalogs, projecting GEAR1 forecasts onto any region, defining and evaluating forecasts on QuadTree grids, performing non-Poissonian tests, and evaluating alarm-based forecasting models.
- The CSEP research community is an active international collaboration with participants from the US, UK, Italy, New Zealand, Switzerland, China and other.
- The pyCSEP software development is managed by SCEC as a community software development efforts with contributions for an international group of developer.
- We welcome your participation in CSEP activities. Please contact Max Werner or Philip Maechling or any of the pyCSEP development team

References

1. **Schorlemmer** et al., 2010. First results of the regional earthquake likelihood models experiment. Pure and Applied Geophysics
2. **Schorlemmer** et al., 2018 Earthquake Predictability: Achievements and Priorities. Seismological Research Letters
3. **Savran** et al., 2022. pyCSEP: A Python Toolkit for Earthquake Forecast Developers. Seismological Research Letters
4. **Graham** et al., 2024 New Features in the pyCSEP Toolkit for Earthquake Forecast Development and Evaluation Seismological Research Letters
5. **Kagan**, Y. Y. 2009. Testing long-term earthquake forecasts: likelihood methods and error diagrams. Geophysical Journal International
6. **Brier**, G.W. 1950. Verification of forecasts expressed in terms of probability. MonthlyWeather Review
7. **Serafini** et al., 2022. Ranking earthquake forecasts using proper scoring rules: binary eventsin a low probability environment. Geophysical Journal International.
8. **Bayona** et al., 2022a. Prospective evaluation of multiplicative hybrid earthquake forecasting models in california. Geophysical Journal International.
9. **Bayona** et al., 2022b. Are Regionally Calibrated Seismicity Models More Informative than Global Models? Insights from California, New Zealand, and Italy. The Seismic Record
10. **Bird** et al., 2015. Gear1: A global earthquake activity rate model constructed from geodetic strain rates and smoothed seismicity. Bulletin of the Seismological Society of America
11. **Khawaja** et al., 2022. Multi-Resolution Grids in Earthquake Forecasting: The Quadtree Approach. Bulletin of the Seismological Society of America

Thank you